2007-01-01

Towards a Personal Automatic Music Playlist Generation Alogorithm: the need for Contextual Information

Gordon Reynolds
Dublin Institute of Technology

Dan Barry
Dublin Institute of Technology, dan.barry@dit.ie

Ted Burke
Dublin Institute of Technology

Eugene Coyle
Dublin Institute of Technology, Eugene.Coyle@dit.ie

Towards a Personal Automatic Music Playlist Generation Algorithm: The Need for Contextual Information

Gordon Reynolds, Dan Barry, Ted Burke and Eugene Coyle

The Audio Research Group,
School of Electrical Engineering Systems,
Dublin Institute of Technology, Kevin St, D8, Ireland

{gordon.reynolds | dan.barry | ted.burke | eugene.coyle} @dit.ie

Abstract. Large music collections afford the listener flexibility in the form of choice, which enables the listener to choose the appropriate piece of music to enhance or complement their listening scenario on-demand. However, bundled with such a large music collection is the daunting task of manually searching through each entry in the collection to find the appropriate song required by the listener. This often leaves the listener frustrated when trying to select songs from a large music collection. In this paper, an overview of existing methods for automatically generating a playlist is discussed. This discussion outlines advantages and disadvantages associated with such implementations.

The paper then highlights the need for contextual and environmental information, which ultimately defines the listener's listening scenario. Environmental features, such as location, activity, temperature, lighting and weather have great potential as meta-data. Here, the key processes of a basic system are outlined, in which the extracted music features and captured contextual data are analysed to create a personalised automatic playlist generator for large music collections.

1. Introduction

With the aid of cost affordable storage and greater device inter-connectivity, a listener's personal music collection is capable of growing at an extraordinary rate. When faced with such large music collections, listeners can often become frustrated when trying to select their music. Hence, it becomes increasingly difficult for a listener to find music suited for a particular occasion.

To further the problem of music selection, today's culture of mobile technology enables the listener to transport an entire music collection in the pocket. Mobile music players now boast of song storage of up to 40,000 songs. As a result, many listeners will plan and prepare playlists for mobile activity that corresponds to a specific activity or mood, such as travelling and exercising. However, according to Suchman [1], plans alone do not dictate actions but only provide a framework that individuals can use to organise action. This implies that the listener attempts to execute previously prepared plans while continuously adapting their actions to the environment [2]. This scenario has led to a study of context-aware music devices [2] and the examination of the role of emotion in music selection.

This paper discusses a design proposal to further the research area of context-aware and emotion-aware music devices. In particular, how environmental data may be used to infer a listener's mood and how such information may integrate into the process of automatically generating a music playlist.

2. Overview of Existing Playlist Methodologies

This section provides a definition of a playlist and presents playlist attributes asscociated with such a definition. The automatic playlist generation process is then discused with an overview of its major themes.

2.1. Defining a Playlist

A playlist may be defined as a finite sequence of songs which is played as a complete set. Based upon this definition there are three important attributes associated with a playlist. These attributes are: 1) the individual songs contained within the playlist, 2) the order in which these songs are played and 3) the number of songs in the playlist.

The Individual Songs in the playlist are the very reason for generating such a playlist. It is therefore essential that each song contained within the playlist satisfies the expectations of the listener. These expectations are formed based upon the listener's mood, which in turn is influenced by the environment.

The Order in which the songs are played provides the playlist with a sense of balance which a randomly generated playlist can not produce. In addition to balance, an ordered playlist can provide a sense of progression such as, a playlist progressing from slow to fast or a playlist progressing from loud to soft.

The Number of Songs in a playlist determines the time duration of the playlist. An understanding of the length of a playlist is important, as song ordering and song balancing of the playlist is unachievable otherwise.

2.2. Playlist Implementations

As catagorised by Vossen [3], the current status of research involving automatic playlist generation is portrayed under two major types of impletmentations. These implementations are 1) Recommender Based Playlists and 2) Constraint Based Playlists.

2.2.1. Recommender-Based Playlists

A Recommender-Based System estimates the user's music preference from a localised music collection and then generates a set of songs based on these estimates from a wider music collection. There are two common approaches to implement a Recommender-Based System, these are 1) Content-Based Learning and 2) Collaborative Filtering.

Content-Based Learning analyses each song in the music collection and then matches songs which have musically similar attributes, such as tempo, instrumentation or genre. If a listener likes a particular song, usually indicated by the user listening to the entire song, then the Content-Based System will recommend

songs that are similar to that song. Figure 1 outlines the Content-Based Playlist generation procedure.
As shown in Figure 1, the user is required to specify a *seed song* and the number of songs required in the playlist. The seed song represents the type of music that the listener wants to listen to. The system then filters the music collection based on similarity to the seed song. A similarity song space is hence created from which a playlist is generated.

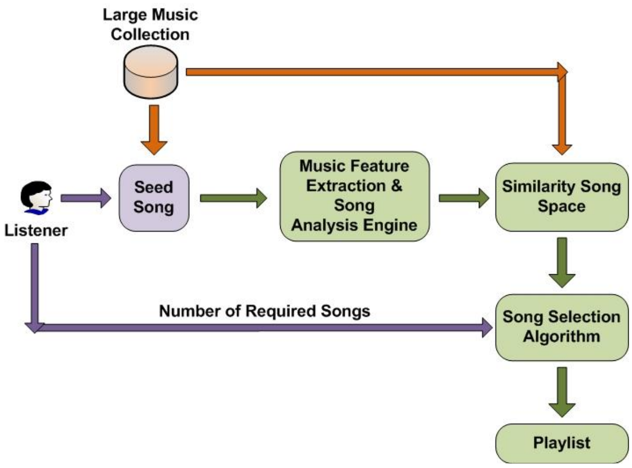

Figure 1: The Content-Based Playlist Generation Procedure.

In a Content-Based System, a significant disadvantage is that song order has no meaning since all the songs are similar. This may also suggest that the playlist may seem dull due the lack of song variation. However, such a system may be useful in circumstances where a themed playlist is required.

Collaborative-Filtering is a community process, as it employs a multi-user approach that uses explicit preference to match songs to a specific user. The system then expands this set of songs by finding another user with a similar taste in music. The system then recommends songs from this user back to the original user [4].
Figure 2 outlines the basic principle of Collaborative-Filtering in a Venn diagram. With Collaborative-Filtering, song order is not taken into account. However, Collaborative-Filtering does provide a varied playlist which may be more interesting to listen to when compared to a Content-Based Learning system.

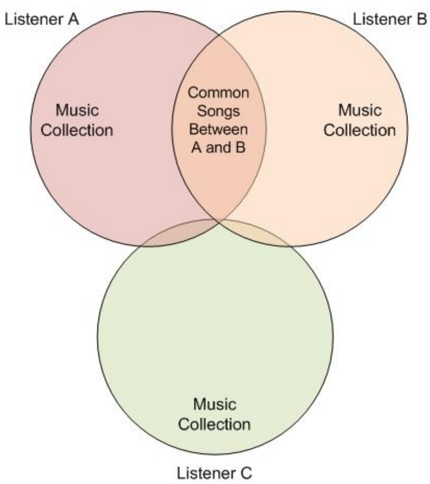

Figure 2: A Venn Diagram Indicating Collaborative-Filtering.

From Figure 2, Listener A has a lot in common with Listener B in terms of their music collections, compared to Listener C. As a result, the Collaborative-Filtering System will recommend songs from Listeners B collection to listener A and recommend songs from Listeners A collection to listener B. Nothing is recommended from Listeners C collection to either Listener A or Listener B. The system assumes that since Listener A and Listener B have so much music in common that their preferences must be the same, i.e. they have the same musical taste. Hence, Listener A would enjoy Listener B's music collection and vice-versa.

2.2.2. Constraint Based Playlist

In the Constraint-Based approach, song order in the playlist becomes a primary focus and hence to date, the only systems that consider the three requirements for a playlist, these are 1) Songs, 2) Order and 3) Length. This is achieved by forming a rule set which defines the song order in a playlist. An overview of such a procedure is given in Figure 3.

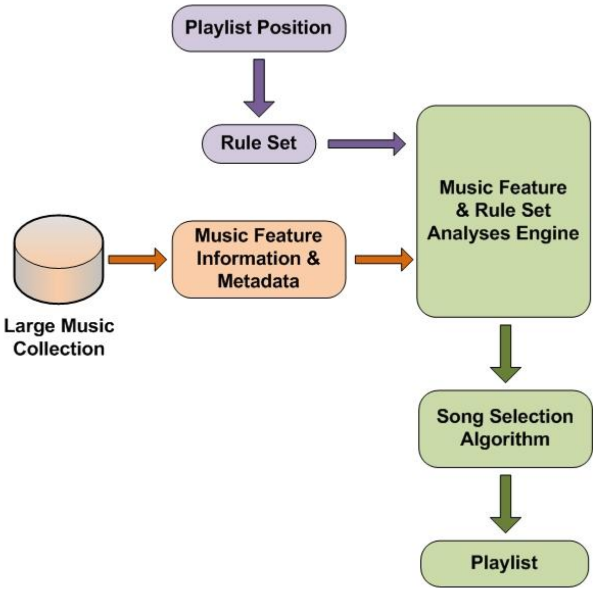

Figure 3: Overview of the Constraint-Based Process.

The rule set provides a set of definitions to which a song must adhere to before being selected. An example of a rule set, as implemented by Vossen in [3], defines a global rule by the requirement that the tempo for each song in the playlist must be above a desired value. Other examples of constraint rules include, that no two adjacent songs in the playlist can be from the same artist or same album. Once the appropriate song is found in the music collection it is inserted into the suitable playlist location. The system then searches for a song to fit the rule set of the next playlist location.

Based on the previously presented playlist implementations, this project is currently investigating the compatability of a Constraint-Based approach in its implementation. The Constraint-Based approach provides the most flexibility yet strict framework for creating an algorithm for automaticlly generating a music playlist.

3. The Need for Contextual/Environment Data

The selection process is dominently ruled by the emotional state and attitude of the individual. Individuals are a function of mood [5] and music selection is no different. Therefore, to provide a listener with a meaningful personalised automatic playlist generation system, it is ideal for the system to consider the listeners mood. Measuring such a parameter directly from the listener borders on impossibility. However, with the establishment of attitude theory in the 1930's, strong links have been forged between an individuals environment and attitude, which in turn defines mood and behaviour [6]. The experience of an individual in the outside world reflects how they feel on the inside [5].

With such strong defined theoritical links between an individuals environemnt and their behaviour, it may be possible to reduce the need to infer mood from a listener in order to create an automatic playlist of songs to suit that mood. Such an approach may circumvent problems raised by *Tolos et al.* in [7], such as defining a set of moods that is relatively unabigous, widely accepted and useful for the average user.

It is proposed to design a system that will monitor a listeners environment and observe their choices in music selection. Analogous to a basic input/output black box system, given the inputs (environmental features) and the outputs (the selected songs) one is required to reconstruct the transfer process, i.e the listeners mood or behaviour, Figure 4.

.

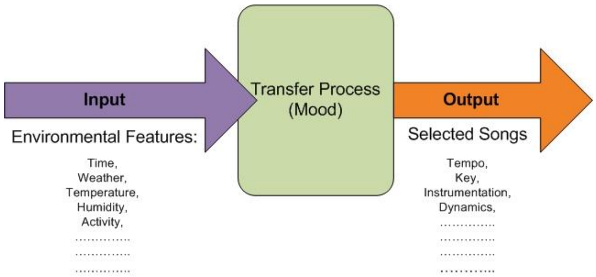

Figure 4: A black box approach, given the inputs and the outputs, one is required to reconstruct the transfer process.

It should be noted that this process is not equivelent to extracting the musical mood or the musical emotional state of a composition as implemented by *Liu et al.* [8]. With the aid of music theory, *Liu et al.* showed with the implementation of an algorithm, that a composition may occupy a particular emotional space, musically speaking.

However, each individual listener subjectively interperates the emotional or mood space of a composition based on their experience. Hence, the same composition is capable of causing a diverse array of emotions within the listening community. Such an example of this is a composition by *Carl Orff* titled *Carmina Burana*. A classical enthusiast and non-horror movie viewer may recognise and perceive the piece in the appropriate classical context it was written in. However, the non-classical lover and horror movie fanatic will recognise the piece as the theme tune to the horror movie '*The Omen*'. In this case, hearing the music piece out of context may induce a sense of fear, uneasiness and terror for the listener. This is because the listener only associates this theme with horror.

The long term consistancy and reliability of using envionmental data in the selection process of automatic playlist generation is founded upon the habitual qualities of human nature. Covey explains, that an individual's action and re-action is pre-conditioned by their environment [5]. Also, as outlined by Ostrom in [6], an individuals attitude, which is a description of their behaviour and formed through experience in their environemnt, operates to make the individuals world predictable and orderly.

3.1. Choosing Appropriate Environmental Features

It is required to identify and categorise environmental features that may affect a listener's mood or music selection process. As an example of how environmental features can affect mood, an investigation into to the effect of lighting on office workers [9] discovered that natural lighting reduces stress and promotes a general sense of *good being* compared to artificial lighting. It is also suggested that the type of lighting combined with the intensity of the artificial light may determine the level of negative effects experienced. In addition, it has also been documented that the weather appears to influence mood and productivity [10].

To commence, it is proposed to consider seven environemnetal features. These features are 1) Time and Date, 2) Weather, 3) Lighting Conditions, 4) Humidity Conditions, 5) Temperature Conditions, 6) Noise Conditions and 7) the Listener's Activity.

3.2. Capturing Environment Data

A brief outline on how each of the environmental features may be captured is given in this chapter.

3.2.1. Time and Date

It is possible to capture time and date using the system clock of the proposed music player, which is a PC based device. With the availability of time and date, it is possible to expand the analysis to include day of the week (Monday, Tuesday, ...), month of the year (Janruary, February, ...), time of the day (morning, afternoon, ...) and season (Winter, Summer, Autumn and Spring).

3.2.2. Weather

It is proposed to obtain weather data through an available METAR service online due to its strict and compact data format. METAR data is available from all airports and it is regularly updated on a thirty minute schedule.

3.2.3. Lighting, Temperature, Humidity and Noise Conditions

With the use of an appropriate sensing device, lighting, temperature, humidity and noise conditions can be monitered and captured. An array of hardware devices exist to capture such parameters. Hardware considerations are discussed further in Chapter 5.

3.2.4. Activity

It is proposed to determine a listeners activity in two forms, these are 1) social scheduling and 2) using an accelerometer. Social scheduling is based upon calender events which involves taking advantage of predictable behavour such as working schedules, travelling schedules, excercising schedules and relaxing schedules. Further information on a listener's activity may be captured electronically with the use of an accelerometer, such as the E-LIS3L02AS4 from STMICROELECTRONICS. An accelerometer is capable of measuring a listener's physical movement, such as walking, running and jumping.

To summerise, environmental parameters have a significant affect on mood and hence influences the music selection process. Therefore, environmental features have great potential as meta-data to allow the listener greater flexibility when searching or accessing a music collection. In addition, environmental features may provide a valuable source of information for an automatic playlist generation algorithm in the generation of playlists to suit a listener's mood.

4. Integrating Contextual/Environment Data into an Automatic Playlist Generation System

Textual meta-data such as artist's name, song title and music genre were initially the only mechanism a listener had for indexing their music collection. However, in recent years, musicians and technologists established the research field of Music Information Retrieval (MIR). One of the principle achievements of MIR was to extend the available meta-data to include musical features extracted directly from acoustic signals. These musical features include tempo, key and timbre. These features allow the user to express music selection and indexing based on actual acoustic information rather than tagged textual information.

It is proposed to develop a system that will extend the range of meta-data further. It proposes to use environmental information to represent the listeners listening scene and mood. With the consideration of such environmental features, Figure 5 outlines three unique feature spaces which may be used to represent a music collection, namely textual descriptions, musical features and environmental features.

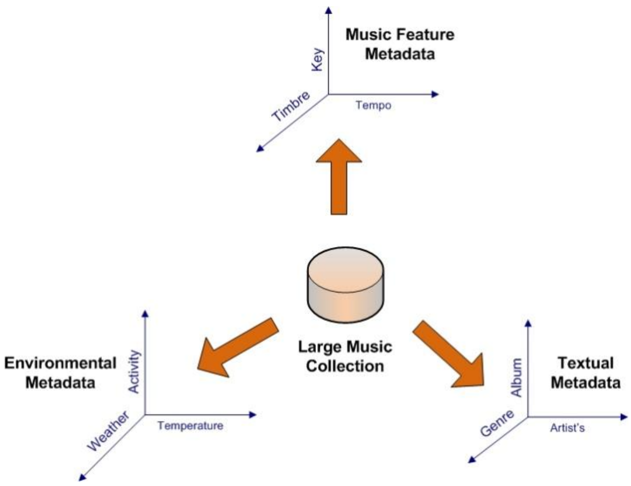

Figure 5: Outlines three unique feature spaces in which a listener's music collection may be indexed.

4.1. Existing Meta-Data

Meta-data is described as everything that is not essence, that is, it is data about data [11]. In audio terms, it usually means data that describes, relates to, or structures essence. To date, the use meta-data is the most dominant means that allows a listener to search a music collection. The search criteria can be specified by artist's name, album name, album genre and release date – only to name a few.

Given the vast range and availability of meta-data, the proposed audio system will integrate such information into its music selection process in combination with environmental features and music features.

4.2. Music Feature Extraction

The proposed system will initially consider three music features. These features are 1) Global Timbre – generally used to describe the similarity of two compositions, 2) Tempo – which describes the speed of a compoition and 3) Musical Key – which describes the relative pitches between notes.

However, as the system develops additional music features will be examined, for example music dynamics, temporal features and spectral features.

4.2.1. Global Timbre

Timbre is a percepual audio characteristic which allows listeners to perceive or distinguish between two sounds with the same pitch and same intensity [12]. The term *Global Timbre* refers to the timbre description covering the full duration of a composition and not just at a particular instant in time nor particualr instrument.

In [13], *Aucoutrier* implements a music similarity technique that employs the use of Mel Frequency Cepstrum Coefficients with Gausian Mixture Models. Based on subject evaluation, Aucoutrier found 80% of the songs suggested by the system as being similar was also identified as being similar by the test users. Logan has also used this type of method to indicate music similarity with similar positive results.

4.2.2. Tempo

Tempo is defined as the speed at which a musical composition is played at [12]. Experiments concerning musical tempo have conveyed its' effects on people in the areas of performance of track athletes [14] and on general spatial awareness, arousal and mood [15].

Leue uses Spectral Energy Flux in combination with a Comb Kernal Filter Bank to deduce tempo from a composition [16]. With such a system, *Leue* acheived accuracies of 80% for pop music and 63% for classical music. Pop music generally produces the more accurate result with tempo detection due to the heavey percussive nature of pop music compared to classical music. *Alonso et al.* has also investigated tempo tracking and extraction using a similar method and has obtained accuracies of up to 89.7%.

4.2.3. Musical Key

Musical key may be defined as the relative pitches or notes contained within a composition [12]. Determining the key of a composition has several applications including mood induction. They mode of the key is deemed to provide a specific emotional connatation [17].

Using a Chroma-based estimation technique, Peeters uses Harmonic Peak Subtraction with a Hidden Markov Model to extract a key from a composition [18]. Confined to the following classical categories, keyboard, chamber and orchrastra an accuracy of 90.3%, 94% and 85.3% was obtained respectively. Pauws also introduced a key extraction algorithm based on chromagrams with an accuracy of 75.1% for classical music [19].

4.3. Using Intelligence

The heart of the proposed system requires the implementation of an intelligent engine capable of learning and applying the learned information to intelligently make a decision. Intelligent systems currently being investigated are Artificial Neural Networks and Hidden Markov Models.

However, due to the complexity of the system it is expected that several intelligent systems are required. For example, in music feature extraction, Gaussian Mixture Models are reported to provide the most efficient performance in calculating song similarity were as, Hidden Markov Models are noted for their performance in segmentation [20].

In the context of an automatic playlist generation algorithm, several different intelligent models have been used. *Jin* describes a process using Hidden Markov Models and experiences an 83% improvement in retrieval time when compared to a forward searching algorithm [21].

4.4. System Process Overview

This section provides a high-level overview of the processes required within the proposed system. This overview is described in two parts, 1) the Learning Process and 2) the Operational Process.

4.4.1. The Learning Process

Figure 6 overviews the learning process of the proposed playlist generation system with integrated environmental meta-data. As the listener selects the required music manually, the system monitors all events in the background which involves the capturing of environmental features. These environmental features are then added to the chosen songs as new meta-data. As different songs are selected, the system identifies and quantifies how each selected song is similar and how they differ. MIR algorithms then use this information to find other songs deemed musically similar to the chosen songs within the bounds of a similarity threshold. Once identified, these songs are also tagged with the same environmental meta-data. The identification process is based upon existing meta-data and music features extracted from each song as previously described in Section 4.2. All results are then catalogued within the system for future reference.

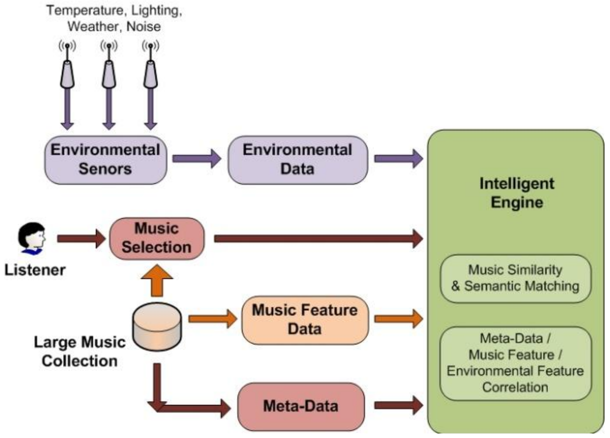

Figure 6: Learning Process of the Proposed Playlist Generator.

Once the system has gained enough experience of the listener's selection process, the system is then capable of automatically generating a meaningful music playlist to suit the listeners listening environment and hence their listening needs. Such a process is proposed in Figure 7.

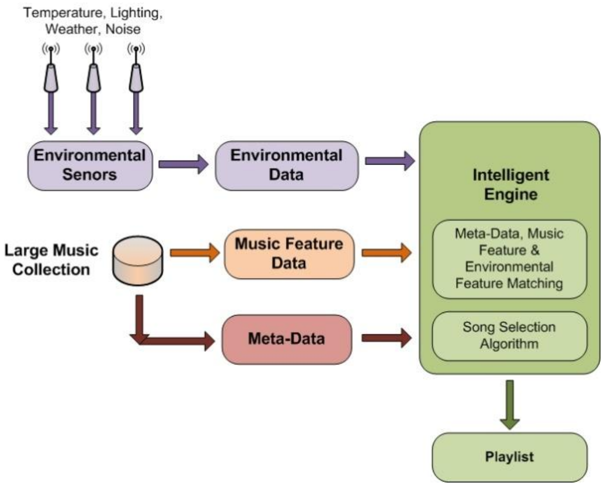

Figure 7: Operational Process of the Playlist Generator.

4.4.2. Operational Process

The trained system, Figure 7, operates without the need for the listener's interaction, since the intelligent engine functions under the same set of parameters and definitions in which it was trained. When the listener requires to listen to music, the trained system will analyse the current listening environment through a sensor array. Based upon the systems previous training, the captured environmental data is matched and assigned to appropriate music features. The music collection is then filtered according to the required features and a song selection algorithm generates the appropriate playlist.

5. Hardware Considerations

It is important that the development and test hardware platform for the proposed music player is mobile, unobtrusive and does not contribute to the listener's mood. In addition, the music device must have appropriate storage capacities and processing power. As a result, a small form factor PC such as the Samsung Q1b – Ultra Mobile PC is currently being used as a testbed. This device is portable and highly inter-connectable with the availability of onboard LAN, WLAN, Bluetooth and USB services. The system uses a 7" touch screen and operates under Windows XP Tablet Edition or Windows Vista rather than a scaled version of Windows such as PocketPC.

A hardware consideration for sensors to allow the capture of environmental data includes the HOBO U12 data logger which is currently being used. This device is compact, portable and self-contains temperature, light and humidity sensors. The unit also has an available external data channel which allows the connection of an external noise level sensor. The data logger is accessible through a standard USB connection and is compatible with the Keyspan USB Server allowing access via Ethernet or WiFi.

To detect and monitor a listener's movement, an Olimex accelerometer (MOD-MMA7260Q) is currently being used. This is a 3-axis device and is pre-mounted on a development board which includes the appropriate support ICs. A mini-USB connection is required to interface with the accelerometer.

6. Conclusions

The concept of an automatic music playlist generation system is presented in this paper. An overview of existing playlist generation techniques is discussed were advantages and disadvantages are outlined.

A Constraint-Based approach was identified as an appropriate playlist generation method. This is due to the fact that it completely encapsulates the entire definition of a playlist and it provides a flexible yet strict framework to work within.

In addition, the key processes of a proposed system were outlined, in which meta-data, the extracted music features and captured environmental data are analysed to create a personalised automatic playlist generator for large music collections.

To further research in this area, a survey has been created to gather appropriate information. This survey can be found on the Audio Research Group's website at www.audioresearchgroup.com/survey. All participation is welcomed.

To conclude, this paper has shown that mood determines the listeners music selection process. Also in reverse, it was shown how music may induce mood in a listener. But more importantly, this paper has discussed and demonstrated that an individual's environment strongly influences mood and hence the listeners music selection process. Based upon these strong influences, it is concluded that environmental features pertaining to a listeners environment has significant potential as meta-data and may provide a valuable resource in the automatic generation of music playlists.

References

[1] L. Suchman. *Plans and Situated Actions, The Problem of Human-Machine Communication.* Cambridge University Press (1987).

[2] S. Tamminen, A. Oulasvirta, K. Toiskallio and A. Kankainen. *Understanding Mobile Contexts*. Proc. Mobile HCI, 17-31 (2003).

[3] M.P. Vossen. *Local Search for Automatic Playlist Generation.* Masters Thesis, Technische Universiteit Eindhoven (2006).

[4] N. Kravtsova, G. Hollemans, T.J.J. Denteneer, and J. Engel. *Improvements in the Collaborative Filtering Algorithms for a Recommender System*. Technical Note NL-TN-2001/542, Philips Research, Eindhoven, (2002).

[5] S. R. Covey. *The 7 Habits of Highly Effective People.* Simon&Schuster UK LTD, (2004)

[6] A.G. Greenwald, T.C. Brock, T.M. Ostrum. *Psychological Foundatins of Attidue.* Academic Press, New York. (1973).

[7] M. Tollos, R. Tato, T. Kemp. *Mood-Basd Navigation Through Large Collection of Musical Data.*. The 5th International Conference on Music Information Retrieval, Barcelona (Spain), ISMIR (2004).

[8] D. Liu, L. Lu and H-J. Zhang. *Automatic Mood Detection from Acoustic Data*. John Hopkins University, (2003).

[9] L. Edwards and T. Torcellini. *A Literature Review of the Effects of Natural Light on Building Occupants*. Technical Note NREL/TP-550-30769, National Renewable Energy Laboratory, Colorado, (2002).

[10] A.G. Barnston. The *Effect of Weather on Mood, Productivity, and Frequency of Emotional Crisis in a Temperate Continental Climate*. International Journal of Biometeorology. Vol. 32, No. 2, (1998).

[11] AES, *Demystifying Audio Metadata*, Journal of the Audio Engineering Society, Vo. 51, No.7/8, 744-751 (2003).

[12] Orio. *Music Retrieval: A Tutorial and Review.* Foundations and Trends in Information Retrieval. Vol. 1, No. 1, 1-90 (2006).

[13] J-J. Arcourtier and F. Pachet. *Music Similarity Measures: What's the Use?* The 3rd International Conference on Music Information Retrieval, Paris (France), ISMIR (2002).

[14] J.R. Brown. *The Effects of Stressed Tempo Music on Performance Times of Track Atheletes*. Florida State University. Florida, (2005).

[15] G. Husain, W. F. Thompson and E. G. Schellenberg. *Effects of Musical Tempo and Mode on Arousal, Mood and Spatial Abilities. Music Perception*, Vol. 20, No. 2, 151-171, (2002).

[16] I. Leue and O. Izmirli. *Tempo Tracking with a Periodicity Comb Kernel.* The 7th International Conference on Music Information Retrieval, Victoria, BC (Canada), ISMIR (2006).

[17] Kastner. *Perception of Major / Minor Distinction: OIV Emotional Connections in Young Children.* Music Perception, Vol. 8, No.2, 189-202.

[18] G. Peeters. *Chroma-Based Estimation of Musical Key from Audio-Signal Analysis.* The 7th International Conference on Music Information Retrieval, Victoria, BC (Canada), ISMIR (2006).

[19] S. Pauws. *Musical Key Extraction from Audio.* The 5th International Conference on Music Information Retrieval, Barcelona (Spain), ISMIR (2004).

[20] J.C. Platt, C.J.C. Burges, S. Swenson, C. Weare, and A. Zheng. *Learning a Gaussian Process Prior for Automatically generating music playlists.* In proc. Of the 14th Conference on Advances in Neural Information Processing Systems, Volume 14,(2001)

[21] H. Jin and H.V. Jagadish. *Indexing Hidden Markov Models for Music Retrieval*. IRCAM-Centre Pompidon, Michigan, (2002).