An In-Situ Study of Real-Life Listening Context

Diane Watson
University of Saskatchewan
diane.watson@usask.ca

Regan L. Mandryk
University of Saskatchewan
regan.mandryk@usask.ca

ABSTRACT

Current models of musical mood are based on clean, noiseless data that does not correspond to real-life listening experiences. We conducted an experience-sampling study collecting in-situ data of listening experiences. We show that real-life music listening experiences are far from the homogeneous experiences used in current models of musical mood.

1. INTRODUCTION

Music recommender systems take information about music that users listen to, and then recommend other music that listeners are likely to enjoy. For example, some music recommender systems, such as Last.fm, group similar music together into playlists. Building good music recommender systems requires knowledge of a track's '*musical mood*', which can be described as the emotion that is expressed by a piece of music. Different musical moods are evoked through differing musical cues in a track. These musical cues, or auditory features, have previously been used to model musical mood with fairly good classification results. These results, however, are based on narrow data sets. Previous work that models musical mood consists of constrained and handpicked representative instances of a single musical genre (usually Western classical or Western popular)[1]. Furthermore, listening experiences are captured in laboratory settings. As such, researchers modeling musical mood know very little about what actual listening experiences look like and how variety in musical genre and listening context will affect modeling results.

It is important to understand real-life music listening experiences because they do not conform to data gathered in laboratory contexts in a number of ways. First, real-life listening experiences contain music from more than a single genre. Second, they happen in a variety of locations, such as in a car, at work, or in a public venue. Third, real-life listening experiences happen for many different reasons, such as to relax, to entertain, or to motivate. Finally, there are many different listening contexts; for example, one can listen to music as part of a social activity, or through a headset as a solitary activity. High performing models based on clean data, gathered in a laboratory and using a single musical genre, may fail when implemented in real-life contexts. We know that context is important but we do not understand how it affects musical experiences [2] or how to use real-life listening data to automatically classify musical mood.

As an initial step toward building musical mood classifiers that are effective for real-life listening experiences, we conducted an experience-sampling study where we collected data on real-life listening experiences, capturing listening context in-situ. Listening context included musical mood, affective state of the listener, reason for listening, activity, location, social company, and level of choice over the song. Genre, title, artist and any associations with the music were optionally captured.

We created an experience sampling application running on Android smartphones. Phones randomly polled the participants over a period of two weeks and asked them to fill out a short survey about the music they were listening to and the context of the listening experience. Our analysis of the data set showed that real-life music listening experiences are far from the homogeneous experiences used in current models of musical mood. In particular, listening experiences are not constrained to a single language or genre. Furthermore listening contexts are varied and music is usually a secondary activity, unlike in a laboratory context where listening to music is the primary activity. Our work shows how real-life musical listening experiences, gathered in situ during a user's daily life, differ from the previous data sets used in modeling musical mood.

2. RELATED WORK

2.1 Affective State

It is well documented that music can induce specific affective experiences in the listener. Affective state, or the emotion or mood a person is experiencing, can be described using either a categorical or dimensional approach. The categorical approach breaks emotions into discrete labeled categories (e.g., happiness, fear, joy) [3]. In contrast, the dimensional approach, which we use in this paper, represents affective state using two orthogonal dimensions: arousal and valence [4]. Arousal can be described as the energy or activation of an emotion. Low arousal corresponds to feeling sleepy or sluggish while high arousal corresponds to feeling frantic or excited. Valence describes how positive or negative an emotion is. Low valence corresponds to feeling negative, sad or melancholic and high valence to feeling positive, happy or joyful. Most categorical emotions can be described by Arousal-Valence (A-V) space (e.g., angry in Figure 1).

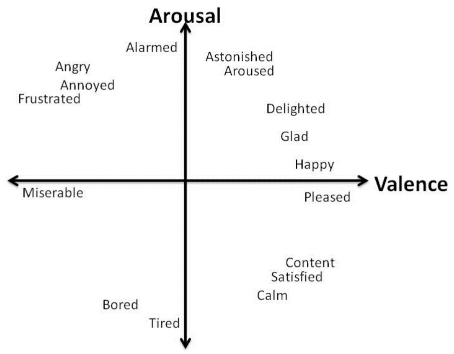

Figure 1 shows A-V space labeled with several of the categorical emotions.

2.2 Musical Mood

Musical mood, the emotion expressed by a piece of music, is to some degree perceived consistently across different listeners and even different cultures. Studies by Juslin and Sloboda have shown that listeners of different musical training classify musical mood into the same categories [5]. Fritz et al. found that the Mafa natives of Africa – without any exposure to Western music – categorized music into the same three basic emotional categories as Westerners [6]. Musical mood is frequently measured in arousal and valence [2] and we have used this approach in this paper. It should be noted that the affective state induced in the listener is not necessarily the same as the musical mood of the music [7], [8]. For example, an individual who is herself feeling frustrated (i.e., mood of the listener) can still perceive a piece of music as calm (i.e., musical mood).

2.3 Musical Mood Classification

Musical mood can be manually categorized by the listener, but researchers have also algorithmically classified musical mood using audio features extracted from the musical track. Work by Juslin [9] has identified seven musical features that are important in the interpretation of musical mood. He asked performers to play the same musical scores in such a way as to express four different musical moods (anger, sadness, happiness and fear) and then had listeners rate the strength of each mood. He found that performers and listeners used the same features to identify each mood, but weighted their importance differently. These features are:

- *Mode:* Mode refers to the key of the music. (e.g. A-)
- *Tempo / Rhythm:* Rhythm is the pattern of strong and weak beat. It can be described through speed (tempo), strength, and regularity of the beat.
- *Articulation:* Articulation refers to the transition and continuity of the music. It ranges from legato (connected notes) to staccato (short abrupt notes).
- *Intensity / Loudness:* Intensity is a measure of changes in volume.
- *Timbre / Spectrum:* Timbre describes the quality of the sound. It is often defined in terms of features of the spectrum gathered from the audio signal.

Musical mood has previously been modeled using only audio features of the music. Lu et al. classified classical music into the four quadrants of A-V space using audio features with an accuracy of 86.3% [1]. Their algorithm also detected places within the song where the mood changed. Experts specified musical mood. Feng et al. classified Western popular music into four moods using only two features: tempo and articulation. They achieved a precision of 67% and a recall of 66% [10]. They do not specify how they gathered musical mood.

Some effort has been made to incorporate other musical context with audio features to improve classification. Yang et al., working with a set of Western rock music, made small gains in their classification rates by adding lyrics to the audio features (from 80.7% to 82.8%) [11]. Musical mood was gathered in a laboratory study. Bischoff et al. integrated socially created tags with audio features, and while their classification rates were low due to problems with their ground truth data, they achieved better results using tags and audio features than audio features alone [12]. Their poor results may be due to the fact they were using a diverse, online, data set with multiple genres. Users of the AllMusic site specified musical mood in this data set.

2.4 Music Recommenders

Many commercial music recommender systems exist (e.g., Last.fm, Pandora, Apple's Genius). In 2010, Han et al. created COMUS, a context-based music recommender system that accounts for mood, situation and musical features [13]. Their system was limited to recommending music for only one listening purpose – to transition between emotional states – and assumed a prior explicit knowledge about how a specific individual changes their music habits depending on listening context.

2.5 Experience Sampling Methods in Music

Experience sampling methods (ESM) have been used in the past to reliably collect music listening experiences [14]. Past research has focused on the musical engagement and the reasons for listening [15] or the emotions induced by music [16] rather than the musical mood of the music.

3. METHODS

We surveyed participants using ESM to gather an in-situ data set related to affective state, musical mood, and listening context. We created an application that ran on Android 2.1 that generated custom surveys from XML files. Participants were asked to carry the phone with them at all times. While it would be possible to create a plug-in for an existing computer media player such as iTunes, we wanted to capture listening experiences in all contexts. For example, some activities, such as exercising, do not usually occur simultaneously with computer use. Participants were not required to use the phone as a media player as this would further limit listening contexts (e.g., music playing in the background at a restaurant). The tradeoff is that we could not automatically capture song title, artist, genre, or audio features such as tempo.

The program would query the user randomly (approximately hourly) by vibrating the phone. A participant could fill out a survey or dismiss the program by indicating they were too busy. Surveys were completed in less than five minutes and were filled out regardless of whether participants were listening to music. Four types of information were collected: musical mood, affective state, artist, title and genre, and listening context.

Musical Mood*:* Participants were asked to describe the musical mood of the song they were listening to using two five-point differential scales. They were asked to rate the arousal of the music by selecting one of five radio buttons between low arousal and high arousal. Similarly, they rated the valence of the music on a scale between sad and happy. A definition of arousal and valence was given to participants before the study and available from a help menu.

Affective State: Participants were asked to describe their personal arousal and valence using five-point differential scales similar to musical mood.

Artist, Title and Genre: Artist and title could optionally be entered in free-text fields that autocompleted to previously entered answers. A genre field was provided that autocompleted to a list of common genres taken from Wikipedia, but also allowed participants to enter their own genre.

Listening Context: Participants were asked questions describing their current listening context. Participants selected their current activity from a list (waking up, bathing, exercising, working, doing homework, relaxing, eating, socializing, romantic activities, reading, going to sleep, driving, travelling as a passenger, shopping, dancing, getting drunk, other). These activities were taken from [2], which lists the most common activities to occur in conjunction with music.

Participants also selected their location (home, work, public place, other) and social company (by myself, with people I know, with people I do not know). Participants selected their reason for listening (to express or release emotion, to influence my emotion, to relax, for enjoyment, as background sound, other) as well as whether or not they choose the song (yes, yes as part of a playlist, no). A text field was provided for participants to enter any terms or phrases they associated with the song.

3.1 Experience Sampling Study

Twenty participants, (14 male) with an average age of 25, were given an Android phone running the software for two weeks. They were paid per number of surveys completed, between \$5 and \$40 CAD. To obtain the maximum payout, 112 surveys were required, which is roughly eight surveys per day. A progress bar provided feedback about the number of completed surveys.

4. RESULTS

In total 1803 surveys were filled out, 610 of those surveys were completed when the participant was listening to music. Only the results of the music surveys are included in this paper.

4.1 Musical Mood

Participants tended to listen to music with high musical arousal and high musical valence. The music they were listening too had an average arousal of 2.64 (SD=1.05) and average valence of 2.66 (SD=1.14) on our five-point scale (0 is low, 2 is neutral, 4 is high). See Figure 2 for the distribution of musical mood. Larger circles correspond to a higher number of responses.

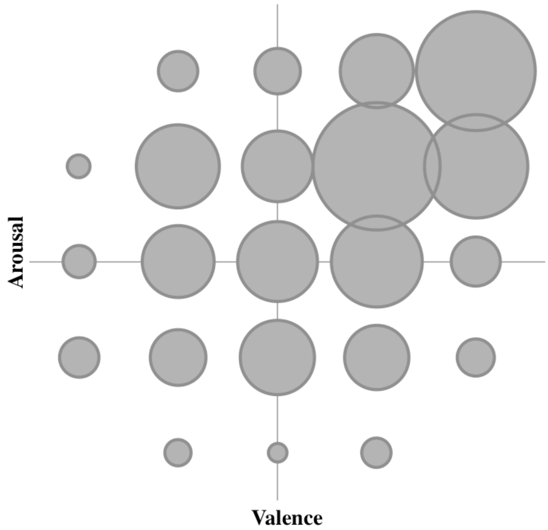

Figure 2 shows the distribution of musical mood on our five-point scale. Larger circles correspond to a higher number of responses. In general people listened to music with high valence and high arousal.

4.2 Affective State

Participants had an average arousal of 2.28 (SD=0.92) on our five-point scale (0 is low, 2 is neutral, 4 is high) and average valence of 2.64 (SD=0.90). See Figure 3 for the distribution of personal affect. Larger circles correspond to a higher number of responses.

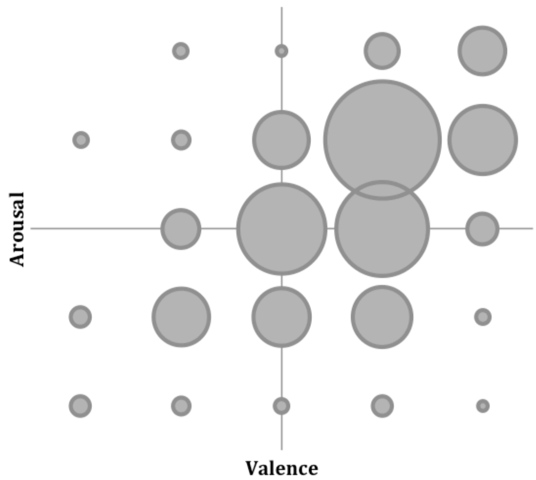

Figure 3 shows the distribution of personal affect on our five-point scale. Larger circles correspond to a higher number of responses.

4.3 Language

Songs were not limited to Western genres or even the English language. Using the artist and title provided by participants, lyrics were downloaded so that the distribution of languages could be examined. At least as 14% of the songs with artist and title specified were non-English. Some of the languages encountered were Persian/Iranian (4%), Japanese (4%), Chinese (2%), French (1%), Korean (1%), Bangladeshi (<1%), and Swedish (<1%). 2% of songs were written in an unidentified non-English language. All participants indicated they listened to at least some English music.

4.4 Artist, Title, Genre

Participants entered an artist 510 times, 325 of those were unique. Artists were encountered on average only once. 501 different song titles were entered, 423 were unique. 102 unique song genres were entered a total of 486 times.

Genres were placed into genre categories, namely the parent genre as listed on Wikipedia. For example, 'heavy metal' is a type of rock music. Only one genre category was possible, if a participant had listed two genres the first was chosen (e.g. "pop-rock" was coded as pop.) The most common genre categories were pop (28%), rock (23%), electronic (14%), jazz (7%), hip-hop & rap (6%), other (5%), modern folk (4%) and country (3%). The remaining genres fell into the categories of classical, traditional/indigenous music, soundtrack, blues, easy listening and R&B (see Figure 4).

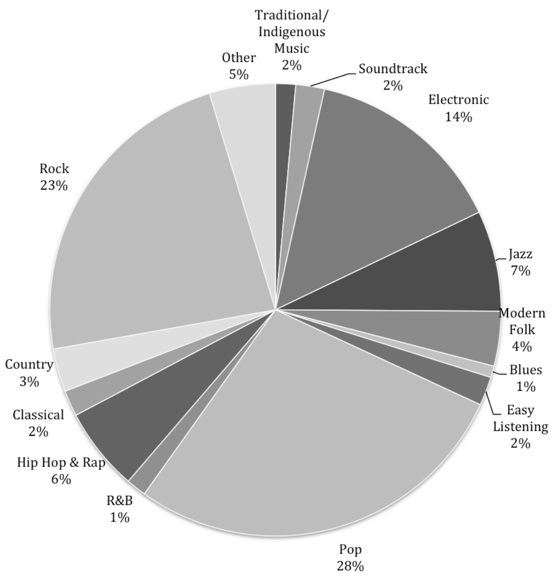

Figure 4 shows the different genre categories encountered in the study. Pop made up only 28% of total songs; classical was rarely encountered.

4.5 Listening Context

Unlike in laboratory contexts, music listening in real life occurs mainly as a secondary activity. The most common activities while listening to music were working (37%), relaxing (21%), eating (6%), driving (5%), travelling as a passenger (5%), other (5%) and socializing (4%) (see Figure 5).

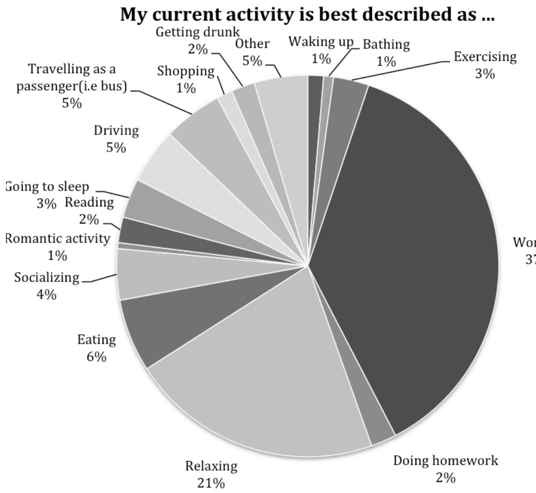

Figure 5 shows the different activities that happened in conjunction with music. Unlike laboratory contexts, listening to music in real life occurs mainly as a secondary activity.

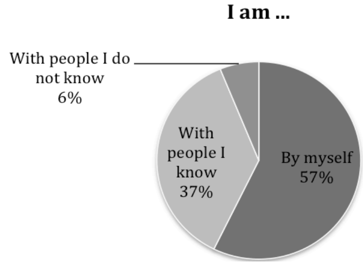

Figure 6 shows the social company of the listening experience. Music is often, but not always a solitary activity.

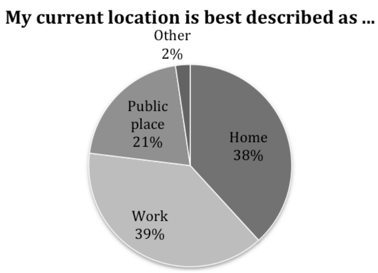

Figure 7 shows the types of locations listening experiences occurred in. Music was often listened to at home or work.

Music is often a solitary activity, taking place in home or work locations. Participants were by themselves 57% of the time, with people they knew 37% and with people they did not know 6% (see Figure 6). They were at work 39% of the time, at home 38%, in a public place 21% and in other locations 2% (see Figure 7).

The most common reason for listening was to use the music as background sound (46%) or enjoyment (25%). 10% of participants used the music to relax, 13% to influence their emotion, 4% to express or release emotion and 2% for other reasons (see Figure 8). Participants indicated they chose the song 74% of the time; 50% of the time it was as part of a playlist. While it may be possible to listen for background sound in a laboratory context, some of the other reasons (e.g., to express or release emotion) may be difficult to simulate.

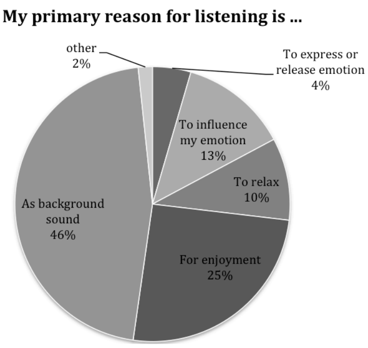

Figure 8 shows the reasons people listen to music. The main reason for listening was background sound. It is difficult to recreate some of these reasons for listening in a laboratory context.

4.6 Associations

Participants entered phrases and terms they associated with the music for 335 songs. These were then coded into themes, a list partially taken from [5] (see Figure 9). 45% of the time participants described an emotion or mood (e.g., "excitement," "depressing"). 20% of the time participants indicated lyrical or musical features of the song itself, such as a phrase from the lyrics (e.g., "baby I like it"), or specific instruments or descriptions of the music (e.g., "piano, sax",). Imagery (e.g., "space, planets, stars", "explosions") made up 15% of all associations. 7% of the time participants listed a specific person, location or memory (e.g., "my daughter," "winter in the cabin"). Other associations included religion and cultures (e.g., "Japan", "Satan"). Nostalgia (e.g., "the eighties, gospel style") was expressed 4% of the time.

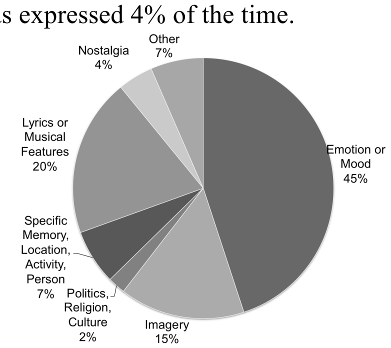

Figure 9 shows the distribution of association themes. The most common type of association was a description of an emotion or mood.

5. DISCUSSION

Our experience-sampling study collected in-situ data that reflects real-life music listening experiences. This study shows that unlike data usually used for modeling musical mood, listeners rarely limit themselves to a single genre or even language of music. Participants did not limit their music in any way to Western popular music; classical music was rarely encountered. In general, the listening contexts captured in our study also differ greatly from a laboratory context. Like in a laboratory context, music is often, but not always a solitary activity. However, unlike laboratory contexts, music occurs mainly as a secondary activity. Music was also listened to for a variety of reasons that cannot be easily mimicked in a laboratory.

Only one category of our study showed quite homogenous answers – musical mood. Many previous studies have assumed that people listen to music with four emotional categories (happy, sad, fear, anger) [9]; however, in our study we found that people tended to listen to music with high arousal and high musical valence (i.e., happy). The other three emotions may simply not be equally represented when capturing in-situ data. It may be that people listen to music in these three emotions only in certain circumstances (e.g., sad music after a breakup) and that these circumstances simply never occurred during the study. Participants were polled about once per hour, and the timing of the polls may have missed specific contexts, but the study suggests that musical mood clusters around positive arousal and valence.

Although our data set reveals a wide variety of listening contexts due to the in-situ nature of our data collection method, our study has several limitations that we wish to address. Participants are unlikely to answer a survey during some activities (e.g., driving) and may be more likely to answer during others (e.g., working), which could skew our sample. Also, some categories may overlap in our data. For example, one could be reading and relaxing at the same time, but only the primary context was collected. Although we captured music from a two-week listening period, the number of participants and length of the study may have been too small to collect a fully representative sample of listening context. Two weeks is not long enough to capture possible seasonal patterns (e.g., Christmas music); the study was done in a time period in July that does not intersect with any seasonal music.

It may be possible that musical mood is not invariant to listening context (i.e., that participants rate the same song with different musical moods depending on listening context). We cannot investigate this in our data set, however, as it is possible that participants are choosing music with different musical moods in different contexts; one would need the same songs played in a variety of listening contexts – in our study, songs and artists were only encountered once on average. To examine the relationship between listening context and musical mood, one could provide participants with representative samples in a music library for use in their musical listening.

5.1 Modeling Musical Mood with In-Situ Data

Previous attempts to model musical mood were based on a single genre and a selection of samples within that genre. Our data shows that the genres vary, and models of musical mood developed for one genre may not transfer well to other genres or multiple genres. The next step is to determine whether previous models of musical mood based on auditory features apply to a data set gathered in situ with a wide variation in song and genre; however, there are two main issues that must be addressed when modeling data gathered in situ.

The first issue is class skew. When modeling musical mood one wants all moods to be equally represented in the training data to avoid inflated classification accuracy. In our data, few songs existed with low arousal and valence. It may be difficult to find songs representative of low musical arousal or low musical valence in real-life listening situations; people may simply not listen to this type of music often. Models of musical mood should account for class skew using methods such as undersampling, a technique in which random instances are chosen such that there is an equal spread of classes.

The second issue that must be dealt with is that the data being modeled is noisy. Models will be based on data covering multiple genres and languages. Previous work has often been based on Western popular or Western classical music and often used handpicked representative instances. Listening experiences happened only in laboratory settings. The noisy data provided from real-life listening experiences may produce upper limits on classification accuracy, which are likely to be much lower than in previous research.

5.2 A Context-Aware Music Recommender

Our data shows variation in musical song and genre, but also shows that the reasons for listening to music, and the context of the listening situation vary. Bringing this user-centric data into a model of musical mood may help solve the problems created from applying models across genres. This type of *context-aware model of musical mood* would include listening context in order to predict the musical mood that a person is likely to listen to. These models would then be implemented into a *context-aware music recommender system* that would recommend music by predicting a musical mood and then creating a playlist from the user's music library that matches this mood.

There are two types of recommendation systems that could be created. One would focus on personalized music recommendations, suggesting music from a person's own music library to match their current situation. The second type of recommendation system would focus on contextual music recommendations, making general suggestions for specific situations, such as background music at a restaurant.

Because most music in this study had high musical arousal and high musical valence, it may be that in the general case people want to listen to music with this specific mood. A music recommender may only have to recommend happy music most of the time. Therefore, predicting the outlier instances where this is not the norm becomes more important.

6. CONCLUSIONS

We conduct an in-situ experience-sampling study of real-life music listening experiences. We show that real-life music listening experiences are far from the homogeneous experiences used in current models of musical mood. In particular, listening experiences are not constrained to a single language or genre. Furthermore the context of the listening experience is highly variable and music is usually a secondary activity.

Acknowledgments
This work was funded in part by the GRAND NCE.

7. REFERENCES

[1] L. Lu, D. Liu, and H.-J. Zhang, "Automatic mood detection and tracking of music audio signals," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 14, no. 1, pp. 5 – 18, Jan. 2006.

[2] P. Juslin and P. Laukka, "Expression, Perception, and Induction of Musical Emotions: A Review and a Questionnaire Study of Everyday Listening" *Journal Of New Music Research*, vol. 33, no. 3, pp. 217–238, 2004.

[3] P. Ekman, *Basic Emotion*. John Wiley & Sons, Ltd., 2005.

[4] J. A. Russell, "Core affect and the psychological construction of emotion," *Psychological Review*, vol. 110, no. 1, pp. 145–172+, 2003.

[5] P. Juslin and J. Sloboda, *Music and Emotion: Theory and Research*. Oxford University Press, 2001.

[6] T. Fritz, S. Jentschke, N. Gosselin, D. Sammler, I. Peretz, R. Turner, A. D. Friederici, and S. Koelsch, "Universal Recognition of Three Basic Emotions in Music," *Current Biology*, vol. 19, no. 7, pp. 573 – 576, 2009.

[7] K. . Scherer and M. R. Zentner, "Emotional Effects Of Music: Production Rules," in *Music and Emotion:Theory and Research*, New York, NY, USA: Oxford University Press, 2001, pp. 361–392.

[8] M. R. Zentner, S. Meylan, and K. . Scherer, "Exploring 'musical emotions' across five genres of music.," ICMPC, Keeyle, UK, 2000.

[9] P. N. Juslin, "Cue utilization in communication of emotion in music performance: Relating performance to perception.," *Journal of Experimental Psychology: Human Perception and Performance*, vol. 26, no. 6, pp. 1797–1812, 2000.

[10] Y. Feng, Y. Zhuang, and Y. Pan, "Popular music retrieval by detecting mood," in *ACM SIGIR conference on Research and development in information retrieval*, New York, NY, USA, 2003, pp. 375–376.

[11] D. Yang and W. Lee, "Disambiguating Music Emotion Using Software Agents," in *ISMIR*, Barcelona Spain, 2004.

[12] K. Bischoff, C. Firan, R. Paiu, W. Nejdl, C. Laurier, and M. Sordo, "Music Mood and Theme Classification-a Hybrid Approach," in *ISMIR*, Kobe, Japan, 2009.

[13] B.-J. Han, S. Rho, S. Jun, and E. Hwang, "Music emotion classification and context-based music recommendation," *Multimedia Tools Appl.*, vol. 47, pp. 433–460, May 2010.

[14] J. A. Sloboda, S. A. O'Neill, and A. Ivaldi, "Functions of music in everyday life: An exploratory study using the Experience Sampling Method.," *Musicae Scientiae*, vol. 5, no. 1, pp. 9–32, 2001.

[15] A. E. Greasley and A. Lamont, "Exploring engagement with music in everyday life using experience sampling methodology," *Musicae Scientiae*, vol. 15, no. 1, pp. 45 –71, Mar. 2011.

[16] P. N. Juslin, S. Liljeström, D. Västfjäll, G. Barradas, and A. Silva, "An experience sampling study of emotional reactions to music: Listener, music, and situation.," *Emotion*, vol. 8, no. 5, pp. 668–683, 2008.