MODELING MUSICAL MOOD FROM AUDIO FEATURES, AFFECT AND LISTENING CONTEXT ON AN IN-SITU DATA SET

A Thesis Submitted to the College of

Graduate Studies and Research

In Partial Fulfillment of the Requirements

For the Degree of Master of Science

In the Department of Computer Science

University of Saskatchewan

Saskatoon, CANADA

By

Diane Watson

PERMISSION TO USE

In presenting this thesis in partial fulfilment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professor or professors who supervised my thesis work or, in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any material in my thesis.

Requests for permission to copy or to make other use of material in this thesis in whole or part should be addressed to:

Head of the Department of Computer Science
176 Thorvaldson Building
110 Science Place
University of Saskatchewan
Saskatoon, Saskatchewan
Canada
S7N 5C9

ABSTRACT

Musical mood is the emotion that a piece of music expresses. When musical mood is used in music recommenders (i.e., systems that recommend music a listener is likely to enjoy), salient suggestions that match a user's expectations are made. The musical mood of a track can be modeled solely from audio features of the music; however, these models have been derived from musical data sets of a single genre and labeled in a laboratory setting. Applying these models to data sets that reflect a user's actual listening habits may not work well, and as a result, music recommenders based on these models may fail. Using a smartphone-based experience-sampling application that we developed for the Android platform, we collected a music listening data set gathered in-situ during a user's daily life. Analyses of our data set showed that real-life listening experiences differ from data sets previously used in modeling musical mood. Our data set is a heterogeneous set of songs, artists, and genres. The reasons for listening and the context within which listening occurs vary across individuals and for a single user. We then created the first model of musical mood using in-situ, real-life data. We showed that while audio features, song lyrics and socially-created tags can be used to successfully model musical mood with classification accuracies greater than chance, adding contextual information such as the listener's affective state and or listening context can improve classification accuracies. We successfully classified musical arousal in a 2-class model with a classification accuracy of 67% and musical valence with an accuracy of 75%. Finally, we discuss ways in which the classification accuracies can be improved, and the applications that result from our models.

ACKNOWLEDGMENTS

I would like to acknowledge my supervisor Regan Mandryk for her support and direction during my time at the University of Saskatchewan. I would also like to thank the members of the Interaction Lab in the Department of Computer Science for making my time enjoyable and memorable. Finally, I would like to thank Brett Watson letting me follow my dreams.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS/ACRONYMS

A-V	Arousal-Valence
EEG	Electroencephalography
EKG	Electrocardiography
EMG	Electromyography
ESM	Experience Sampling Methodology
GSR	Galvanic Skin Response
LIWC	Linguistic Inquiry Word Count
M	Mean
MFCC	Mel-frequency cepstrum
MIR	Music Information Retrieval
PAM	Photographic Affect Meter
PANAS	Positive and Negative Affect Scale
PID	Participant Identification Number
SAM	Self Assessment Mannequin
SD	Standard Deviation

CHAPTER 1

INTRODUCTION

Musical mood is the emotion expressed by a piece of music. Listeners interpret the musical mood of a musical track through a variety of cues in the form of audio features (e.g., mode, rhythm, articulation, intensity and timbre) [31]. Composers and musicians use these cues to convey a specific musical mood to listeners; however, music can have a musical mood even if it is not intentionally created (e.g., in the case of algorithmically-composed music). Although there is a large variation in styles of music around the world, musical mood is perceived the same across listeners of different cultures [19] and with different musical training [29].

1.1 MOTIVATION AND PROBLEM

Musical mood is an important aspect of a listener's experience, but can also be useful for systems that require knowledge about the music they are playing. One example is music recommenders, which are systems that recommend music that listeners might enjoy. A listener who enjoys a particular musical track can be recommended with similar tracks based on artist, genre, or style; however, knowledge about the musical mood of a track can help a music recommender make salient recommendations [2] that will be appreciated by a listener. To allow a music recommender to base its decisions on musical mood, we can algorithmically model the musical mood of a track – using the audio features previously described – and achieve fairly good classification results [6,41,66].

These current high-performing models of musical mood have main two problems that prevent their integration into music recommender systems: first, the musical tracks used in the models are constrained to a single genre or culture (Western Classical or Western Popular); second, the data for the models is collected in laboratory contexts and may not transfer to real-world listening contexts. Previous studies have suggested that constraining music to a single genre may

not be in line with what people actually listen to [20,28,30,59], although none of these studies have specifically investigated musical mood. In real life listening experiences, people listen for many different reasons and during many activities. As such, previous models (based on data gathered from a single genre or culture and gathered in a laboratory setting) may fail when applied to data sets gathered in real-life. Systems implementing these models, such as music recommender systems, may then fail when applied to data from real-life experiences. This failure may result in poor recommendations, and ultimately a negative user experience with the system.

1.2 SOLUTION

To solve the problem of building good musical mood classifiers that are effective for real-life listening experiences, we first need to understand whether real-life listening experiences are as homogenous as previous models have assumed. Second, we need see if previous models of musical mood apply to data that is collected in-situ, during a listener's daily life. Finally, we need to find a way to model musical mood that is effective on an in-situ and heterogeneous data set.

1.1.1 Steps to the Solution

To understand whether real-life listening experiences are homogenous, we gathered a data set of real-life musical listening experiences. To gather real-life listening experiences, we employed the Experience Sampling Methodology (ESM). ESM has previously been used for gathering real-life listening experiences [20,28,30,59]; however, most previous studies have been conducted with pagers and paper diaries. We created experience-sampling software, called Music Survey, which runs on Android smartphones and generated custom surveys from xml files. This software deployed surveys about once per hour that collected the musical mood of the music a user was listening to, the user's affective state, the user's listening context, and information (e.g., song title, artist, genre) about the music they were listening too.

We then deployed Music Survey in a field study. Twenty participants were given Android smartphones to carry with them for two-weeks. Participants were paid per number of surveys completed to encourage participation. In total 1803 surveys were completed; in 610 instances, participants were listening to music.

To determine whether the approaches used in previous models apply to our in-situ data set, we gathered the necessary information to model the musical mood of the musical tracks in our data set. First, we downloaded the songs for instances that had the artist and title provided. Matlab and MIRtoolbox [35] were then used to extract audio features describing the mode, rhythm, articulation, and timbre of the music. Lyrics for songs were also collected and analyzed using a textual analysis tool called Linguistic Inquiry Word Count (LIWC) [48]. We also collected publically-entered tags from Last.fm, a music recommender website. These were also analyzed using LIWC.

We then created Bayes Net models using data mining software called Weka [21]. We separated the features into three feature sets: musical information (audio features, lyrics and tags), musical information + affective state, and musical information + listening context. We modeled musical arousal and musical valence using each set. The first set was used to replicate the approach of previous models that have been applied successfully to data sets of a single genre, collected and analyzed in laboratory settings. The remaining two feature sets were used to improve upon the results of the first in order to find a way to model musical mood that is effective on an in-situ and heterogeneous data set.

1.3 CONTRIBUTIONS

We make three contributions in this thesis. First, we show how real-life musical listening experiences, gathered in situ during a user's daily life, differ from the previous homogenous data sets used in modeling musical mood. In particular, participants listened to music with a generally happy musical mood. Second, we successfully model musical mood from a data set gathered in-situ during a user's daily life; we are the first to do so. Finally we show that while musical

features (i.e., audio features, song lyrics and socially created tags) can successfully model musical mood with classification accuracies better than chance, adding contextual information, such as the listener's affective state or the listening context of the musical experience, can further improve classification accuracies. We successfully classify two states of musical arousal with a classification accuracy of 67% and musical valence with an accuracy of 75% when using both musical features and listening context.

1.4 THESIS OUTLINE

The remainder of this thesis presents the related work, methodology, results, and implications of our research.

Chapter 2 presents the literature in the field relating to affect and affective state, music as inducing emotion, music as expressing emotion (i.e., musical mood), automatically classifying musical mood, and music recommender systems that use affect or listening context to make recommendations.

Chapter 3 describes the experience-sampling software used to collect survey responses as well as the surveys themselves.

Chapter 4 covers the details of the ESM study, the results of the pre-study survey and the data collected during the ESM study.

Chapter 5 explains the classification features. In addition to information gathered during the survey (see Chapter 4), we extracted additional features for use in classification. These additional features fall into two categories: audio features and textual features. The additional classification features could not be gathered for all instances. The chapter ends with a table of all classification features.

Chapter 6 presents the methods and techniques employed in modeling musical mood, and the results of all models.

Chapter 7 discusses the results and offers design implications. We show that people listen to music with a generally happy mood (higher than neutral arousal and valence), that real-life data gathered in situ may cause previous models to fail, that our classification results are limited by a sparse data set, and that context is important when modeling musical mood. We also discuss future work.

Chapter 8 summarizes our research and discusses the contributions.

CHAPTER 2

RELATED WORK

In this section we present an overview of the areas of affect and musical emotion for the purposes of directing this thesis. We discuss defining and collecting personal emotion using both subjective and non-subjective measures. We delve into the theory behind the music as a means of inducing and expressing emotion and explore approaches to automatic musical mood classification.

2.1 PERSONAL EMOTION: AFFECT

Emotion, mood, and affect are terms that are often used interchangeably. For the purposes of this thesis, we will use the following distinction between the terms. Emotions are fleeting and reactionary experiences while moods are longer lasting and can influence shorter-term emotions [4]. Affect, or affective state, is the measureable state of physiological experience of emotion or mood experienced cognitively by a sentient being [18]. We use the term affect, or affective state, to mean a human being's measurable state of experiencing an emotion and/or mood.

2.1.1 Categorical vs. Dimensional

There are two main approaches to representing affective state: categorical and dimensional. The categorical approach breaks emotions into discrete labeled categories (e.g., happiness, fear, joy) [15]. Human beings often describe affect using categorical words and thus this system of describing affect is natural. However, determining core categories is a problem; unique emotional states may be left out of a comprehensive list (e.g., the emotion of schadenfreude – taking pleasure in the misery of others – is not generally included in lists of emotional states), or similar concepts may be represented with multiple categories (e.g., content/ satisfied or joy/elation). Organization of this affective space is key, but this organization varies from researcher to researcher. Hevner, for example, organized affective words into eight

clusters: Happy, Exciting, Vigorous, Dignified, Sad, Dreamy, Graceful, Serene [27]; while Shaver organized affective words using six categories: Love, Joy, Surprise, Anger, Sadness, Fear [56].

Russel's circumplex model of affect offers a dimensional approach. Here affect is described using continual dimensions of arousal and valence (and occasionally dominance, tension, or kinetics) [52]. Arousal can be described as the energy or activation of an emotion or mood. Low arousal corresponds to feelings such as "sleepy" or "sluggish" while high arousal corresponds to feelings such as "frantic" or "excited." Valence describes how positive or negative an emotion is. Low valence corresponds to negative feelings like "sad" or "melancholic" and high valence to positive feelings such as "happy" or "joyful." Dominance describes how controlled or controlling the emotion is. At the low end of the scale are feelings like controlled, influenced, cared for, awed, submissive, and guided. At the high end are feelings like controlling, influential, in control, important, dominant, and autonomous**.** Because arousal and valence are orthogonal dimensions, many of the categorical emotions can be labeled in Arousal-Valence (A-V) space. See Figure 1 for a standard mapping of A-V space. Although dominance can be included as a third dimension, it is often ignored, as it does not provide the same differentiating power as either arousal or valence.

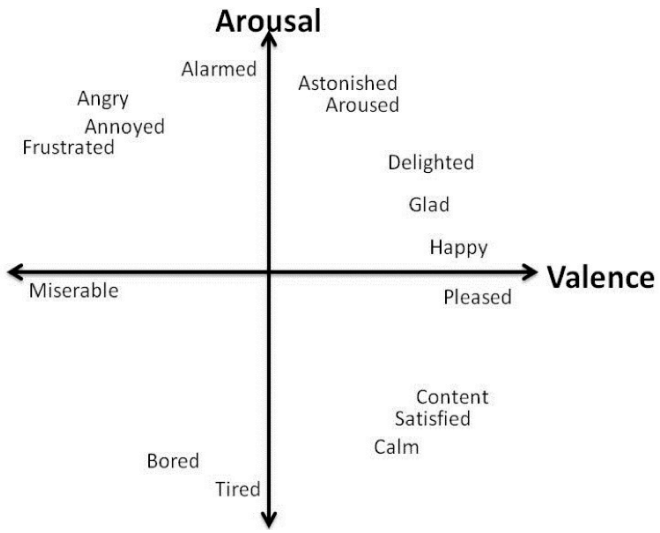

Figure 1 A-V space labeled with some of the categorical emotions

2.1.2 Gathering Affect

To make use of affect, it must be measured. This section discusses how affect can be collected, both in laboratory and natural contexts using both subjective and objective measures.

2.1.2.1 Subjective Measures

Subjective measures involve asking participants to describe their own affective state, either in retrospect or in the moment. The simplest methods involve asking participants to rate their arousal and valence on scales. There are a variety of standardized and validated scales, which we describe next.

Self-Assessment Manikin (SAM): SAM [7] measures arousal, valence, and dominance using a series of representational diagrams (Figure 2). Each scale consists of five pictures representing the intensity of the dimension. Verbal descriptions of the concepts are often included as well. Participants are asked to select an image or a box in between the images giving 9 total levels. SAM allows for quick assessments even if the concepts are confusing to participants, and has been used extensively in the subjective reporting of emotion (e.g., [38,39]).

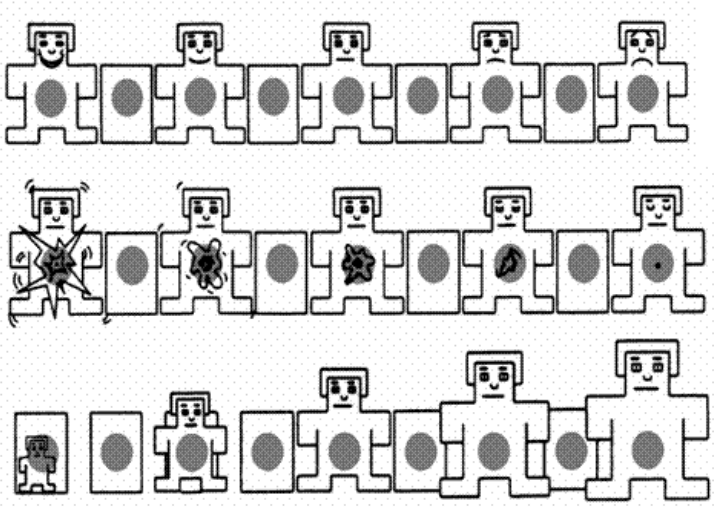

Figure 2 shows the self-assessment manikin (SAM), a scale that measures arousal, valence, and dominance by asking participants to select a manikin for each.

Positive and Negative Affect Scale (PANAS): PANAS [64], or the extended version PANAS-X, presents a participant with a list of mood terms and asks participants to rate how they are experiencing each between 1 (not at all) and 5 (extremely). Rather than focusing on arousal and valence, PANAS measures positive affect and negative affect with high reliability through a series of categorical terms. PANAS-X also provides a measure of fear, sadness, guilt, hostility, shyness, fatigue, surprise, joviality, self-assurance, attentiveness, and serenity.

Photographic Affect Meter (PAM): PAM [49] is a test meant to work directly on mobile phones despite the problem of limited screen real estate. Participants are shown a grid of images depicting an emotional state and asked to select one. A button is included that selects a new set of photos. These photos were collected from a set of photos on Flicker tagged with mood words and each photo has been rated with corresponding levels of arousal and valence. PAM has been shown to collect arousal and valence in-situ reliably. See Figure 3 for an example of a PAM test. At this time, PAM is not available for public use.

Figure 3 shows the photographic affect meter (PAM) that collects arousal and valence by asking users to select a picture that best captures how they feel.

Semantic Differential Scale: The semantic differential scale lists two bipolar adjectives on either side of a set of check boxes [10]. The participant is asked to select a box between the terms. To measure arousal one would select a box between "high arousal" and "low arousal." To indicate valence one selects a box between "positive" and "negative." Often verbal descriptions of the terms are provided.

Likert Scale: A Likert scale is similar to a semantic differential scale, except that participants are asked to rate their agreement with a statement (e.g., I feel happy today) on a scale from "strongly

disagree" to "strongly agree" [51]. It is generally good practice to present both sides of a statement in separate questions to avoid biasing participant responses (e.g., I feel happy today and I feel sad today).

The main disadvantage to subjective measures is that they are subject to problems with introspection and recall. Asking a participant to reflect on their affective state may in fact influence their affective state. Questionnaires of any sort are interruptive to the current task and what is being measured may be the affective state induced by the questionnaire. Despite these limitations, subjective measures are the most widely used technique for gathering affect because they are fast, easy to deploy, easy to understand, and participants are good at reporting on their own attitudes.

2.1.2.2 Physiological Measures

In terms of objective methods for gathering emotion, using physiological changes as indicators of affect is a widely used and reliable method (e.g., [44]). This approach requires physical sensors to gather a number of salient metrics, including:

Electromyography (EMG): EMG sensors are applied directly on the face to sense muscle movement. In particular the zygomaticus major muscle and corrugator supercilii are of interest. Zygomatic muscle activity corresponds to positive valence (smiling) and corrugator muscle activity corresponds to negative valence (frowning) [47].

Galvanic skin response (GSR): GSR sensors on the hands or feet measure changes in the conductivity of the skin due to sweat gland activation. Changes in GSR correspond to changes in arousal [60].

Electrocardiography (EKG): EKG sensors applied to the chest and stomach measure electrical activity of the heart. Research has shown that EKG can be used to differentiate between positive and negative valence [65].

Electroencephalography (EEG): An array of EEG sensors applied to the scalp measure surface electrical activity that corresponds to underlying neural activity. Studies have linked changes in EEG readings to both valence [61] and arousal [8].

These physical sensors are sensitive to even minor physiological changes. For example, EMG will often detect muscle activation without any noticeable physical changes to the face [43]. However, these sensors, are intrusive (i.e., applied directly onto the skin) and are only really suited to a laboratory setting. Measurements collected may not be entirely reliable as the physical nature of the sensors may make participants aware that they are being monitored and cause them to experience emotions related to their instrumentation rather than the intended stimuli.

2.1.2.3 Naturalistic Settings

While affect is often collected using the previously described methods in laboratory settings, the data gathered in laboratory settings may not generalize to daily life. There is often a need to collect affect in more natural settings. It is often important to collect affect in-situ – in the present situation and context in which it is experienced – rather than asking participants to reflect on past events. Musical experiences, for example, are so commonplace in daily life that they may be unmemorable; this may make retrospection difficult and inaccurate [58].

Experience sampling methods (ESM): ESM are one class of methods to capture user data in-situ [26]. Using ESM to gather affect could involve methods such as keeping a mood diary or filling out surveys repeatedly throughout a normal day. Participants are reminded to fill out the survey by use of a signaling device such as a pager or mobile phone. Reminders often occur randomly throughout the day. Sloboda and Neil for example, collected information about musical experiences, listening habits, and changes in affect due to music using ESM with book of surveys and a pager [58]. ESM is useful for detecting changes in affect over time or in respect to the context of the experience, but may be intrusive and subject to the limitations of subjective measures.

Non-invasive sensor instrumentation: Recent research has shown that affect can be modeled using keyboard usage patterns. The work in this area has been concerned with detecting the

subtle changes in behavior that arise through interaction with the current task. Zimmerman et al. found significant differences between neutral and other emotional states using keystrokes in a laboratory setting [69]. Epp et al. later collected in-situ keystroke data using ESM and successfully modeled classifiers for confidence, hesitance, nervousness, relaxation, sadness, and tiredness with accuracies ranging from 77% to 88% [16]. Keystroke dynamics modeling is an objective data source and is less intrusive than other objective methods as affect is modeled using subtle changes in behavior occurring as part of task completion. However it is limited to tasks that involve typing on a keyboard and results are still quite preliminary.

2.2 MUSIC AND EMOTION

This section describes how music can both express and induce emotion and how these concepts differ. Listening context, everything about a listening experience other than the music, is also defined.

2.2.1 Listening Context

The context of a listening experience can define that musical experience. Context is a broad term defined by Juslin as: "everything from the situation in which the musical activity takes place to the wider socio-cultural context" [28,29]. This context can include the location the listening experience takes place in, the reason one is listening to music, the activity being undertaken and any mental associations or imagery associated with the music.

Reason For Listening: People listen to music for several reasons. Previous research has identified reasons such as positive and negative mood management, diversion and distraction, personal identity and expression, background noise, to reflect on past memories, nostalgia, and enjoyment [40,58].

Activity: Certain activities are more likely to occur in conjunction with music than others. According to Juslin and Luakka, these activities include waking up, bathing, exercising,

working, doing homework, relaxing, eating, socializing, romantic activities, reading, sleeping, driving or cycling or running, and travelling as a passenger on a bus or plane [28].

Associations: Humans will often associate particular events or imagery with music. These may involve memories of certain people, nostalgia, or relaxing or violent imagery [28,57]. For example, the "Darling, they're playing our tune" phenomenon occurs when a person learns to associate a specific song with strong positive memories of their significant other [13].

2.2.2 Music as Expressing Emotion

We use the term *musical mood* to describe the emotion expressed by a piece of music. Musical mood may differ in terms of the composer, performer and listener; we are interested in the mood perceived by the listener. We should make it clear that the mood expressed is different than the affect induced [54,68]. Musical mood is frequently described in terms of A-V space [41] and this is the mapping we use here.

The existence of musical mood poses a puzzle to philosophers. Sentient beings express emotions in such a way to convey their current affective state. Tears express sadness; laughter expresses mirth. As music is not sentient, it can therefore not express emotion, yet the phenomenon of musical mood clearly exists [14]. It is possible the mood expressed is the mood of the composer, however, this does not account for music composed algorithmically. Likewise, the mood expressed may be that of the performer, but this does not account for music played automatically by a computer. As a solution to the problem of sentience (or rather a lack of sentience), it is thought that music acts a symbol or sign. These signs may have historical associations or merely mimic how humans express a particular emotion. The idea of music as a symbol gives way to the contour theory [14]. This theory states that some traits can be *happy-looking* without being *happy*. A weeping willow may look sad without being sad. Music, therefore, is a universal language of emotions created through symbols and signs that can express a musical mood without this mood ever being experienced by any sentient being.

Humans are very good at reading musical mood, which they do by interpreting musical cues. In 1959 Deryk Cooke published *The Language of Music,* an extensive lexicon describing specific

patterns and characteristics of music that translate to different musical moods [11]. A common example is the use of the minor key in sad (e.g., funeral) music. While Cooke focused on predominantly western music, the interpretation of these cues appears to be a universal phenomena. Categorical studies have shown that listeners of different musical training classify musical mood in the same way [50]. Furthermore, Fritz et al. found that the Mafa natives of Africa – without any exposure to Western music – categorized the same music into the same three basic emotional categories as Westerners [19].

Work by Juslin [31] 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, 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 varies between legato (smooth, connected notes) and 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 of the audio signal of the music.

Automatic classification of mood is possible through machine learning. Lu et al. classified western classical music using solely acoustic features describing the rhythm, timbre and intensity of the music. Musical experts labeled the musical mood ground truth. Music was classified into the four quadrants of A-V space with an accuracy of 86.3%. Their algorithm also detected mood boundaries, places where musical mood changed within a single selection of music, with an average recall of 84.1% [41]. In a more simplistic approach, Feng et al. attempted to classify

western popular music into the four moods used by Juslin using only two features: tempo and articulation. They achieved a precision of 67% and a recall of 66% [17]. They do not specify how they gathered musical mood.

Some effort has been made to incorporate additional context with audio 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 (80.7% to 82.8%) [66]. Musical mood was gathered in a laboratory setting. Birschoff et al. integrated socially created tags from the website Last.fm 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 [6]. 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 labeled musical mood in this data set. Schuller at al. used textual features discerned from song lyrics, audio features and high level features such as the corresponding ballroom dance style and genre to model arousal and valence separately. They modeled arousal into low and high with overall accuracy of 77.4% and valence with an accuracy of 72.8% [55]. Their dataset was taken from MTV's list of top ten songs (taken over 20 years) and covered multiple genres. Musical mood was gathered from young people in a laboratory setting. For further review of the area of automatic musical mood classification, see [33].

2.2.3 Music as Inducing Emotion

Not only can music express an emotion, but it is also known to induce affect in listeners. The difference between felt and perceived emotion is empirically valid and researchers must be careful to make this distinction in subjective surveys [54]. Induced emotions are not dominated by a particular type of music [28], despite most research being focused on classical genres.

Induced affect is often measured subjectively [28] in laboratory settings where participants are asked to listen to music and rate how it makes them feel. Induced affect can also be measured sensitively with more objective physiological measures [32]. For a comprehensive review of physiological responses to emotionally charged music see [42]. Both subjective and objective

methods are limited as the affect induced by particular piece of music may be highly influenced by context [3]. There is also not a 1-1 correlation with musical mood. The affect induced by sad music is not necessarily sadness though it may be a common reaction. As music is a form of entertainment, a response one would expect to any musical mood is enjoyment [28]. For example, after a bad breakup many people choose to listen to sad music as a way to vent their emotions and improve their mood.

There are several theories as to how music can induce affect in listeners. One such theory is that human beings are reacting to the *Arousal Potential* of the music [5]. According to Berlyne, listeners choose music that will keep them in an optimally-rewarding level of arousal. Music can increase arousal through use of novelty, expectation, complexity, conflict, and instability (dissonance). Music can decrease arousal through predictability, familiarity, grouping and patterns and dominance. For example, music of the Romantic era increases arousal through ”complex texture, large orchestras, frequent chromatic notes and dissonances, sudden jumps into remote keys, long melodic lines with protracted development and upward progression of pitch, and shear length [5].”

Another theory of how music produces an affective response in the listener is Meyer’s theory of *Musical Expectancy* [45]. According this theory, affect is induced through the listener’s expectations of the music. Each phrase of musical notes has a specific set of notes and harmonics that one has learned to expect should follow. The probabilistic distribution of these expectations varies with composer, genre, style and culture. Emotional responses arise through the tension caused and released when these expectations are created and then eventually resolved. This process must be learned and a listener’s expectations are shaped relative to all music they have heard previously.

It is also possible that music induces affect through the well-documented process of *Emotional Contagion* [24]. Emotional contagion is as simple as smiling when one sees someone else smiling. This mimicking of other’s emotional cues leads itself to the spread of affect. This synchronization is well documented in speech [24] and even text-based communication [23]. It is thought that emotional contagion occurs in music as music often features acoustic patterns

similar to speech. For example, angry music is loud and bursty; sad music is slow and lethargic. These patterns are similar to the way an angry or sad person would respond if asked a question. Happy music has been shown to increase zygomatic muscle activity (smiling) in listeners and sad music increases corrugator muscle activity (frowning) [42].

It is also possible that music induces emotions through associations and mental imagery. A common example is nostalgia or the "Darling they're playing our tune" phenomena [13]. The affect experienced is the combination of a reaction to both the music and the images/associations experienced [28]. This may explain the phenomenon of making a "visual" soundtrack, where one puts a series of picture to music or vice versa.

2.2.4 ESM and Music

ESM is an established methodology for collecting real-life listening experiences in the area of music. Several studies have been conducted. Slaboda et al. first established ESM as a way to explore musical experiences [58,59]. Juslin et al. used ESM to explore the emotions induced by music (i.e., affect) [30]. Greasley and Lamont later used ESM to investigate both engagement with music and the function it provides (i.e. reason for listening) [20]. All of these studies, to some extent gathered listening context; however, most were conducted using pagers and paper diaries. To our knowledge, no literature exists where ESM methods are used to gather musical mood.

2.2.5 Music Recommender Systems

Our motivation for understanding musical mood is that it is a potentially important feature for success in music recommender systems. Music recommender systems suggest music to listeners that is likely to be of interest. These suggestions can be based on socially created tags, clustering of similar artists or songs, or audio features such as tempo and articulation. The suggestions are generally static and most do not adjust to the affective state of the listener, their listening context or socio-cultural context of the music. There are a few notable recommender systems that attempt to take into account context while making suggestions.

Last.fm: Last.fm [70] is a commercial music recommender that uses socio-cultural content in the form of socially created tags and groups musicians with similar tags into Internet radio stations. These tags are unlimited and often include descriptions of the music (e.g., rock), mood words (e.g., chill) and mental associations or imagery (e.g., 90's, gangsters).

COMUS: COMUS [22] uses a support vector regression music mood classification system that attempts to combine listening context with musical mood. The listening context incorporated is limited; however, as it is supplied to the system in the form of pre-defined user preferences (e.g., when Bob is at home, working, he likes rock music).

Affective Remixer: The Affective Remixer [9] attempts to recommend music while responding to real time changes in affective state. The system suggests music in segments, which are mixed together, rather than creating a playlist of entire songs as to adapt more quickly to changes in affective state. The goal of the Affective Remixer is to transition a listener to a specific affective state.

Flytrap: Flytrap [12] is a group music recommender system that recommends music based on the personal libraries of a group of listeners. It uses a network of genre similarity to find music that all listeners in the room will enjoy, even if it does not exist in their specific music library.

MoodMusic: MoodMusic [2], like Flytrap, incorporates group context. However, rather than use group listening tastes it takes a measure of group mood and then chooses music with a similar musical mood.

Using musical mood as a feature in a recommender system can be advantageous as a single genre or artist can vary significantly in their musical style. Musical mood can be similar between songs in a way that transcends genre or artist; furthermore it is a concept easily understood by human listeners.

CHAPTER 3

METHODOLOGY

This chapter will cover the ESM software, called Music Survey, used to collect music listening experiences in-situ during a user's daily life. We will also discuss the subjective surveys used by the software..We will start by discussing the preliminary survey.

3.1 PRE-STUDY SURVEY

A preliminary survey was presented to participants before they were handed the phones. The survey asked questions about their music listening habits as well as how often they felt they experienced the same emotion as the music they were listening to.

3.2 EXPERIENCE SAMPLING

We used experience-sampling methods (ESM) to capture real-life in-situ musical experiences. We presented participants with quick short surveys on mobile smartphones that participants carried with them at all times for two weeks. These devices both notified participants that a survey was to be completed as well as administered the survey. Participants were paid per survey they completed to encourage participation. To receive the maximum payout, participants had to fill out an average of eight surveys per day over a two-week period (see Chapter 4 for details about participant payouts).

3.3 SOFTWARE

The experience sampling software was written to run on Android smartphones running Android 2.1. This application was developed by modifying the open source *XMLGui* [1] code to generate custom surveys on an Android phone from easily modified XML files. The software was created as free running application. While it would have been possible to create a plug-in for an existing computer media player such as iTunes, we wanted to capture listening experiences in contexts other than sitting at a desk. For example, some activities, such as exercising, often occur in conjunction with music, but not usually simultaneously with computer use. For the same reason, the music survey application did not function as a music player. The mobile nature of the Android phone makes it possible to capture a wide selection of listening contexts (e.g., music playing in the background at a restaurant) and we did not wish to limit it to only music playing on the device. The trade-off is that we could not automatically capture song title, artist, genre or acoustic features such as music tempo.

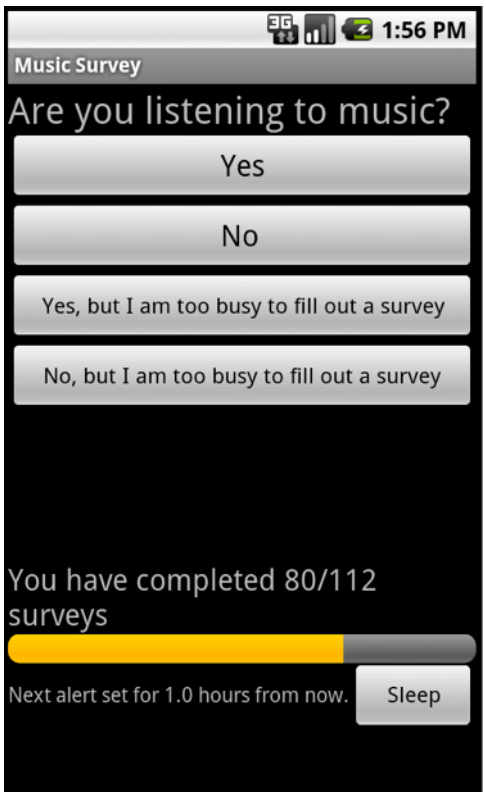

Figure 4 shows the main opening screen of the music survey application. Included are the buttons that will bring up a survey, as a progress bar showing the number of surveys completed and a sleep button that will delay the next alert by any number of hours.

3.3.1 Interface

The phones would vibrate approximately once per hour to indicate that there was survey to be filled out. The time between survey notifications was handled with an Android alarm and was not guaranteed to be exactly one hour. Users could either ignore the vibrations or unlock the phone and be presented with an opening screen (see Figure 4). See Figure 5 for the application flow chart. Users first begin by indicating if they are listening to music and have time for a survey. The software then prompted users for their affective state, location, activity, song information, song musical mood and listening context when the users were listening to music. Only affect and location and activity were collected when the user was not listening to music.

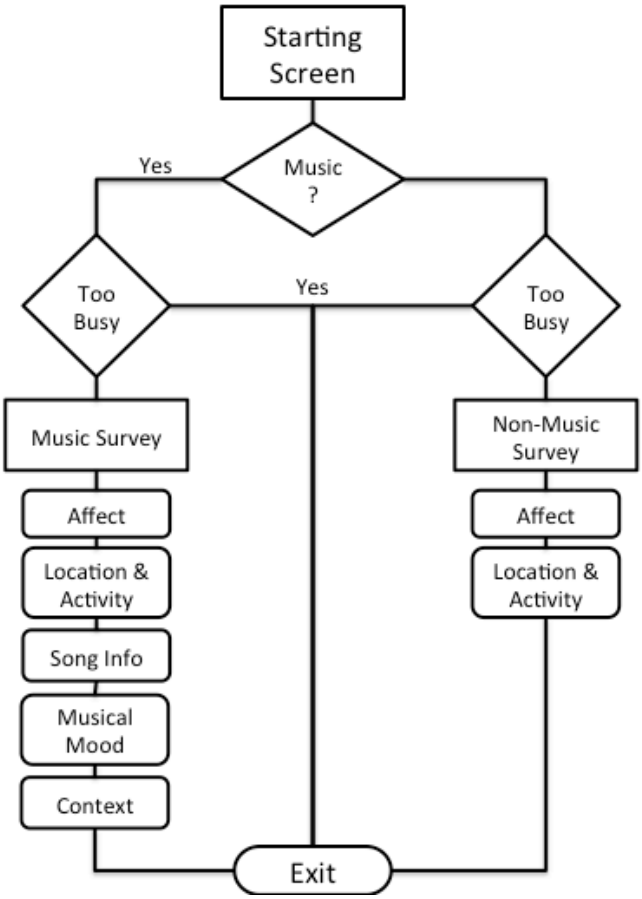

Figure 5 shows the application flow.

3.3.1.1 Opening Screen

This opening screen would present the users with four buttons to indicate if they were listening to music. Tapping "Yes" would open a survey where they would be asked about their affect, musical mood and listening context. Selecting "No" would open the same survey but with no questions about musical mood or listening context. Both other buttons indicated the participant was too busy to fill out a survey and would close the application. The number of times each button was selected was logged. See Figure 5 for an application flow chart.

Also included on the opening screen were a progress bar and a sleep button. The progress bar showed how many surveys had been completed out of the minimum required to receive

maximum payout for the study (eight per day x fourteen days). This was included to encourage participants to fill out more surveys. The sleep button allowed participants to delay the next alert for up to 12 hours if they did not want to be disturbed. The time until the next alert was indicated next to the sleep button.

3.3.1.2 Menu Button

Pressing the Menu button on the Android phone opened a help menu. Here participants had access to a tutorial and definitions. The tutorial survey included instructions on how to interact with the survey on the phone. The definitions included were the same as the ones handed out on paper to the participants at the beginning of the study. These definitions were:

> *Personal mood:* This is the current mood (or emotion) you personally are experiencing. You will be asked to rate your mood on two scales, arousal and valence.
>
> *Musical mood:* This is the mood or emotion a piece of music expresses. You will also be asked to describe musical mood on two scales, arousal and valence.
>
> *Arousal:* Arousal describes the energy of an emotion. Low arousal corresponds to feelings like relaxed, calm, sluggish, dull, sleepy, and unaroused. High arousal corresponds to feelings like stimulated, excited, frenzied, jittery, wide awake, and aroused.
>
> *Valence:* Valence describes how positive or negative an emotion is. Negative valence corresponds to feelings like unhappy, annoyed, unsatisfied, melancholic, despairing, and bored. Positive valence corresponds feelings like happy, pleased, satisfied, contented, hopeful, and relaxed.

3.3.1.3 Survey

The surveys were presented to the user a few questions at a time, with similar questions being grouped together. Users could swipe to scroll through the text if necessary. Questions that were

required were marked with an '*.' A 'Next' button would move to the next set of questions. There was no way to change previous answers once 'Next' had been pressed.

There were three types of question interactions. The first was a horizontal set of radio buttons with a set of semantically opposed terms at either end. Users indicated their choice by tapping a radio button. Only one response was allowed per question (See Figure 6(a)). The second type of question involved a drop down list. Tapping a bar next to the question would bring up a list of possible answers. Users tapped on an answer to select it. (See Figure 6(b)). The last type of question provided users with a free form text field that they could fill in with the on-screen keyboard on the phone. Suggestions of possible answers would be made to the user as they typed (see Figure 6(c)). For most text fields, suggestions were answers from previous surveys although some were populated from lists gathered from websites and Wikipedia. Users did not have to select a suggestion; they could enter into the text field any possible response by using the keyboard.

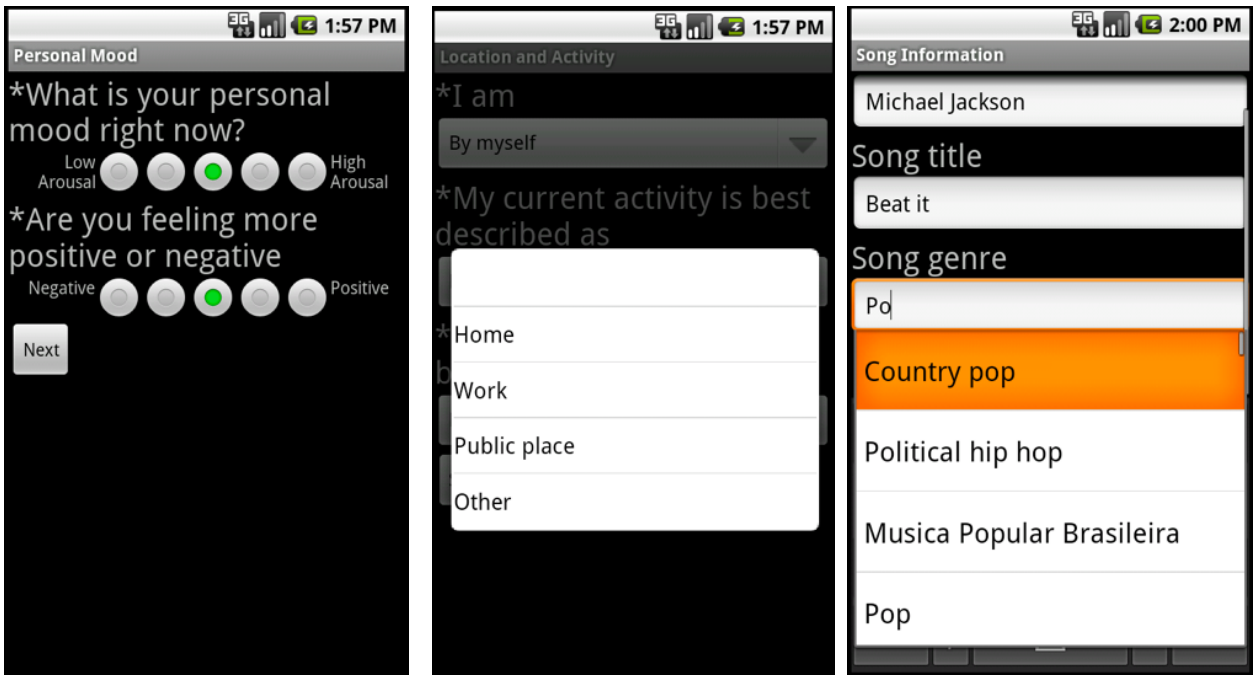

Figure 6 shows all three types of question interactions possible. (a) shows the radio buttons used in the semantic differential scales collecting affect and musical mood. (b) shows an example of a drop down menu. (c) shows a sample of suggestions that occur while typing "Pop" into the genre field.

3.4 SURVEY QUESTIONS

Four types of information were collected through the survey: affective state, musical information, musical mood, and listening context. For a full list of survey questions, see the appendices. Age, sex, and participant identification number (PID) were entered before the phones were handed to participants to anonymize the data as it was collected.

3.4.1 Affective State

Affective state was collected subjectively. Participants were asked to their arousal and valence on a five-point semantic differential scale by selecting a radio button between sad and happy and

low arousal and high arousal. Five-point differential scales were used as this was the maximum number of radio buttons that would fit on the screen width, and because we did not expect to model more than five categories of mood. A long definition of the terms used in the semantic differential scale was give to participants before the study and available under a help menu within the application.

3.4.2 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 [28], which lists the most common activities to occur in conjunction with music. Participants also selected their location (from home, work, public place, other) and social company (by myself, with people I know, with people I do not know). The participants were also asked to select 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 chose the song (yes, yes as part of a playlist, no). Finally, a text field was provided for participants to optionally enter any terms or phrases that they associated with the song.

3.4.3 Music Mood

Musical mood was gathered subjectively using five-point semantic differential scales similar to those for gathering affective state.

3.4.4 Artist, Title, Genre

Artist and title could optionally be reported in free-text fields. To make sure the survey time remained short, these fields would auto-complete to any answers from previously completed surveys. A genre field would similarly auto complete to a list of common genres taken from Wikipedia, but also allowed participants to enter their own genre. Entering an artist, title, or genre was optional.

CHAPTER 4

SURVEY RESULTS

This chapter outlines the results of the ESM study, the results of the pre-study surveys and the data collected during the ESM study.

4.1 STUDY

Twenty participants were given an Android phone running the ESM software for two weeks. Before the study they were given an instruction sheet containing definitions of key terms and a short pre-study survey collecting demographics and information about their normal music listening habits. See the appendices for the information form (Appendix A), consent form (Appendix B) and all survey questions (Appendix C and D).

Participants were paid between 5 dollars and 40 dollars CAD for participating in the study; payment depended on the number of surveys completed. To obtain the maximum payout, an average of eight surveys per day or 112 total surveys were required. On average participants were paid $31.77 CAD. In total 1803 surveys were filled out; 610 of those surveys were done when the participant was listening to music. Only the results of the surveys completed while listening to music are included in the results.

4.2 PRE-STUDY SURVEY RESULTS

Of the 20 participants, 14 were male and 6 were female. Ages ranged between 19 and 30 (M=25). Most participants (90%) listened to music several times a day; they listened primarily on a computer (60%), iPod/iPhone (25%), cellphone (10%) and portable media player (5%).

While computers were most common, participants used on average 3 different types of devices to listen to their music.

In general participants seemed to be aware of a link between musical mood and the affect it induces. When asked whether they perceive themselves to be experiencing the same emotion that a piece of music expresses, 10% of participants indicated that this is always the case, 75% indicated this happened often, whereas 15% said this occurred occasionally. Participants also indicated affect was a reason to listen to music – 50% of participants sometimes listened to music to influence their emotions and 45% indicated they listen to music to express emotion. This is not surprising given the number of participants who indicated that they often feel the same emotion a piece of music expresses; however, this was not the main reason that participants listen to music. Only 10% of participants indicated that influencing emotion was their primary reason for listening, and only 5% indicated that expressing emotion was their primary reason for listening. This is interesting as it indicated that participants often experience the emotion music expresses, but they do not always intend to. Participants also listened for several other reasons – 90% of participants listened to music sometimes for enjoyment and 40% of participants indicated this was the primary reason they listened. Participants sometimes used music as background sound (80%), and 30% of participants indicated that this was the main reason they listened to music. Music was used as a way to relax (60%), and 15% of participants indicated that this was the primary reason they listen to music.

The most common activities to take place in conjunction with music were working, exercising, relaxing, travelling (as a passenger), driving, dancing and doing homework with over 50% of participants indicating that they sometimes combine each of these activities with music. Almost half of participants (45%) listed working as the primary activity they combined with music, whereas 15% of participants listed relaxing. Other activities included doing homework, travelling (as a passenger), eating, driving, dancing, socializing, getting drunk, and going to sleep. In general, participants use music in the background while performing other activities – 80% of participants said that they usually consider music a background activity.

4.3 ESM SURVEY RESULTS

This next section presents the data collected by the Music Survey software during the ESM study.

4.3.1 Affective State

In our ESM application, participants rated their current arousal and valence on five-point semantic differential scales (between 0 and 4). In general participants were experiencing high arousal (M=2.28, SD=0.92) and high valence (M=2.64, SD=0.90) when notified by our application. This is not surprising, as music is a form of entertainment, and therefore we would expect people listening to music to often be in a good mood. See Figure 7 for a two dimensional distribution of affect in A-V space. Larger circles correspond to a higher frequency of responses.

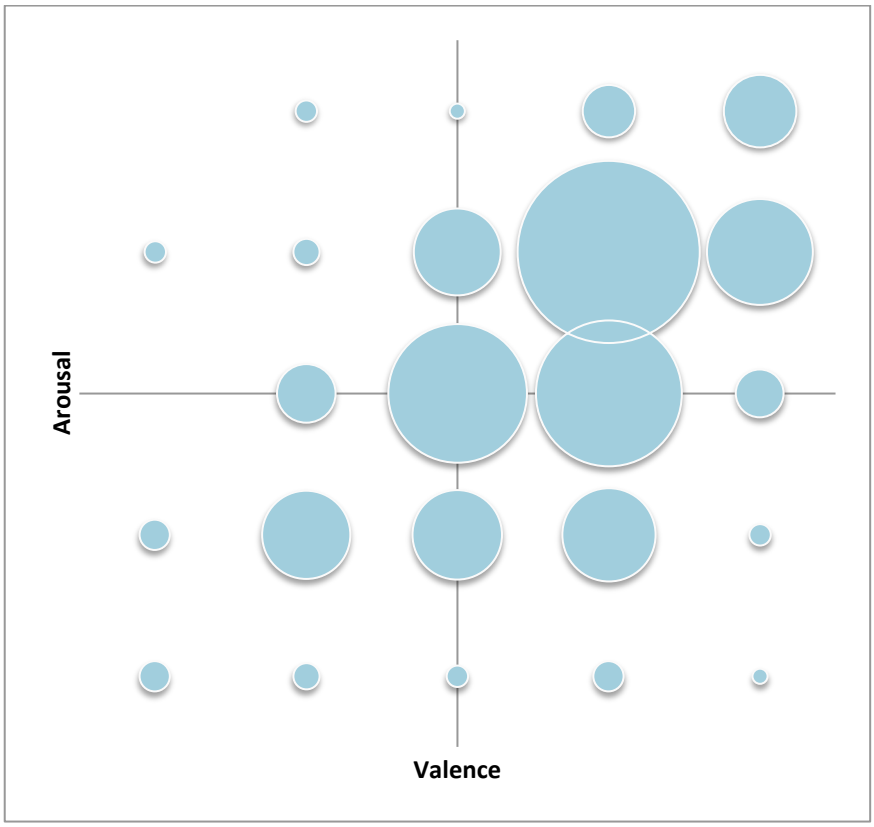

Figure 7 shows affective state in A-V Space. Larger circles correspond to a higher number of responses.

4.3.2 Musical Mood

Participants rated musical arousal and musical valence on five-point semantic differential scales (between 0 and 4). In general the music had high musical arousal (M=2.64, SD=1.05) and high musical valence (M=2.66, SD=1.14). See Figure 8 for a two dimensional distribution of musical mood in A-V space. Larger circles correspond to a higher frequency of responses.

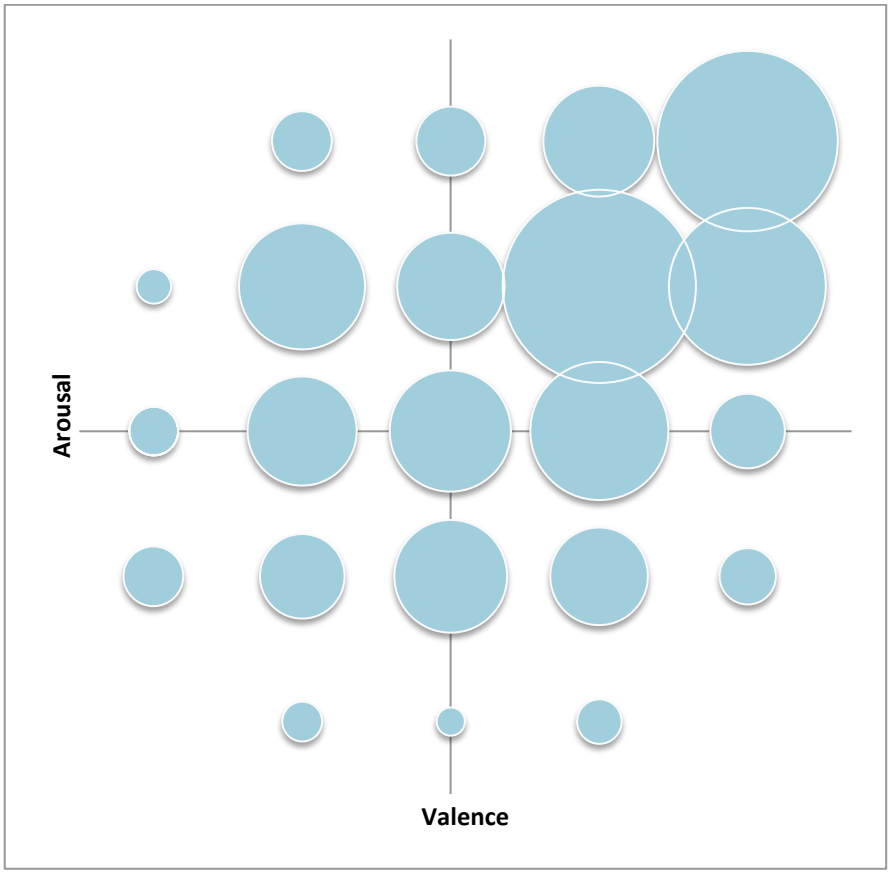

Figure 8 shows a two dimensional distribution of musical mood in A-V space. Larger circles correspond to a higher number of responses. Musical mood tended to be high arousal high valence.

4.3.3 Listening Context

Many combinations of activity, location, and social company were collected. Overall, participants replied at least once with every activity (with the exception of dancing and shopping). The most common activities that occurred in conjunction with music were working (37%), relaxing (21%), eating (6%), driving (5%), travelling as a passenger (5%), and other (5%). See Figure 9 for a graph of activity prevalence. Participants were mostly alone. They were by themselves 57% of the time, with people they knew 37%, and with people they did not know 6%. See Figure 10(b) for social company. Participants were at work 39% of the time, at home 38%, in a public place 21% and in other locations 2% of the time (see Figure 10(a)). In general, participants indicated a level of choice over the music they were listening to, suggesting most music came from their own music libraries. Participants chose their song 24% of the time, as part of a playlist 50% of the time, and did not choose the song 26% of the time (see Figure 11).

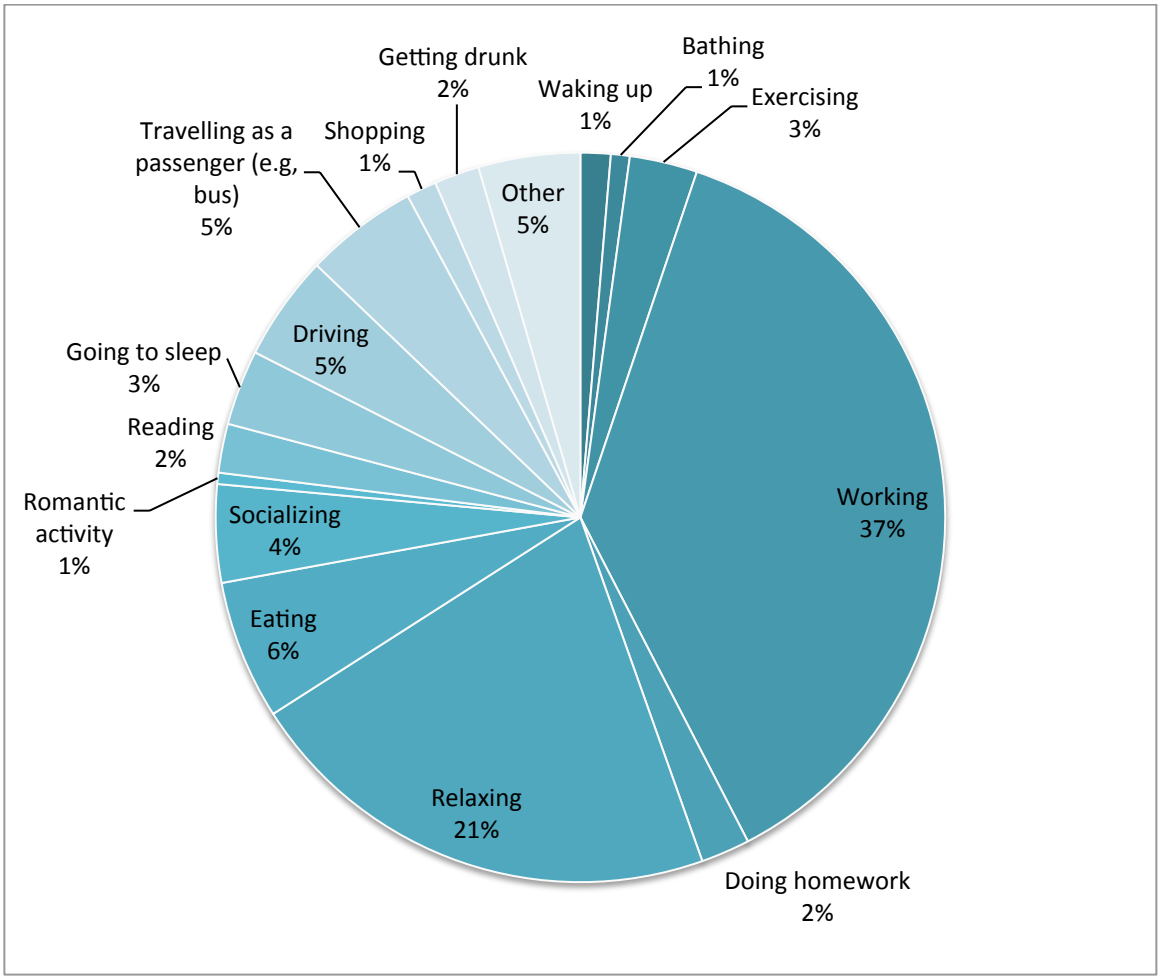

Figure 9 shows the most common activities to occur in conjunction with music.

4.3.4 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.

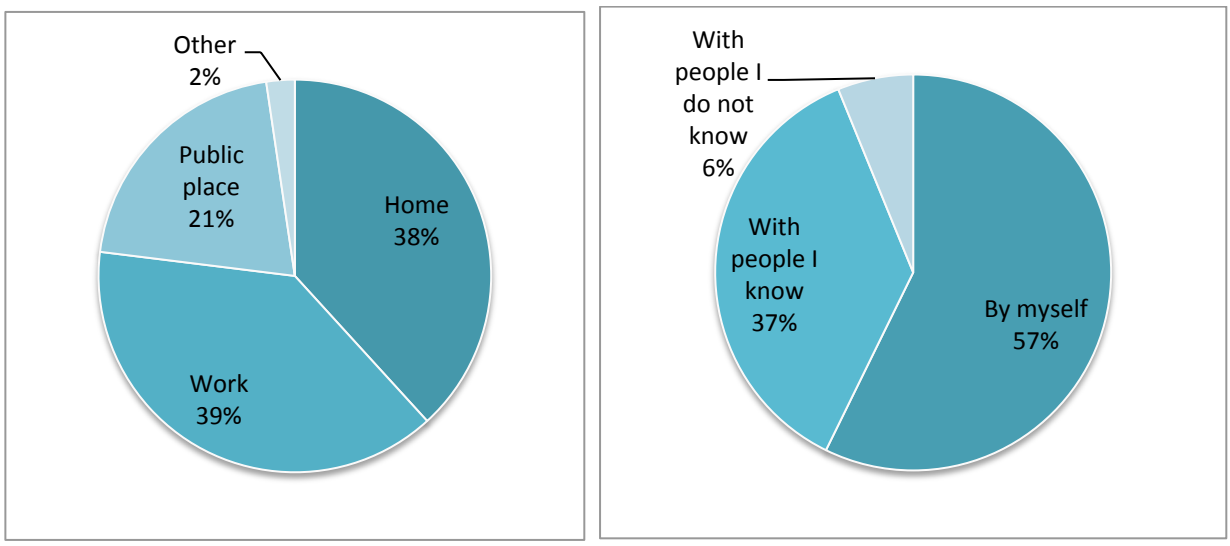

Figure 10 shows (a) the location and (b) the social company of participants while they were listening to music.

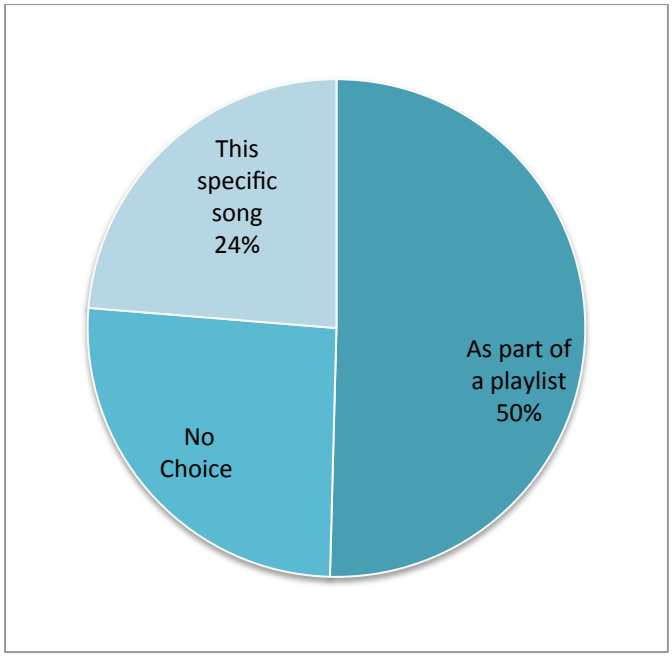

Figure 11 shows the level of choice over the song.

Participants entered phrases and terms they associated with the music for 335 songs. These were then coded into association categories starting with common mental associations with music

from [28]. A total of 23 categories were created. Participants indicated lyrical or musical features of the song itself 20% of the time, 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," "good beats"). Participants described a specific emotion or mood 45% of the time; this list was further broken down into 11 categories (see Table 1). Imagery (e.g., "Calm trees swaying in wind, peaceful) made up 15% of all associations. Participants indicated a specific person, memory, location or activity 7% of the time (e.g., "my daughter," "winter in the cabin"). A feeling of nostalgia was indicated 4% of the time (e.g., "echoes, high school, nostalgia") and religion, politics or culture were mentioned 2% of the time (e.g., "satan," "Japan"). See Table 1.

Table 1 Association Categories

Association Category	**Sub Category**		
Lyrics or Musical Features		20%	
Emotion or Mood		45%	
	Energy, Excitement		9%
	Sadness, Loss or Pain		9%
	Love and Romance		6%
	Empowerment, Strength		5%
	Calm, Peaceful, Sleepy		4%
	Happiness		3%
	Hope		3%
	Humor		3%
	Anger		1%
	Fun or Enjoyment		1%
	Sex		1%
Imagery		15%	
	Other		7%
	Relaxing		4%
	Violent		3%
	Dark or Fearful		2%
Religion, Politics, Country or Culture		2%	
Memory		7%	
	Location or Activity		5%
	Person		2%
Nostalgia		4%	
Other		7%	

4.3.5 Song Information

Participants entered an artist 510 times; 325 of those were unique. There were 501 different song titles entered; 423 were unique. There were 102 unique song genres entered a total of 486 times. These genres were placed into genre categories, namely the parent genre as listed on the Wikipedia hierarchy of popular music genres. For example, 'heavy metal,' alternative rock,' and 'punk' are all types 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 genres were pop (28%), rock (23%), electronic (14%), jazz (7%), hip hop & rap (6%), other (5%), modern folk (4%) and country (3%)[1]. See Figure 12 for the genre distribution.

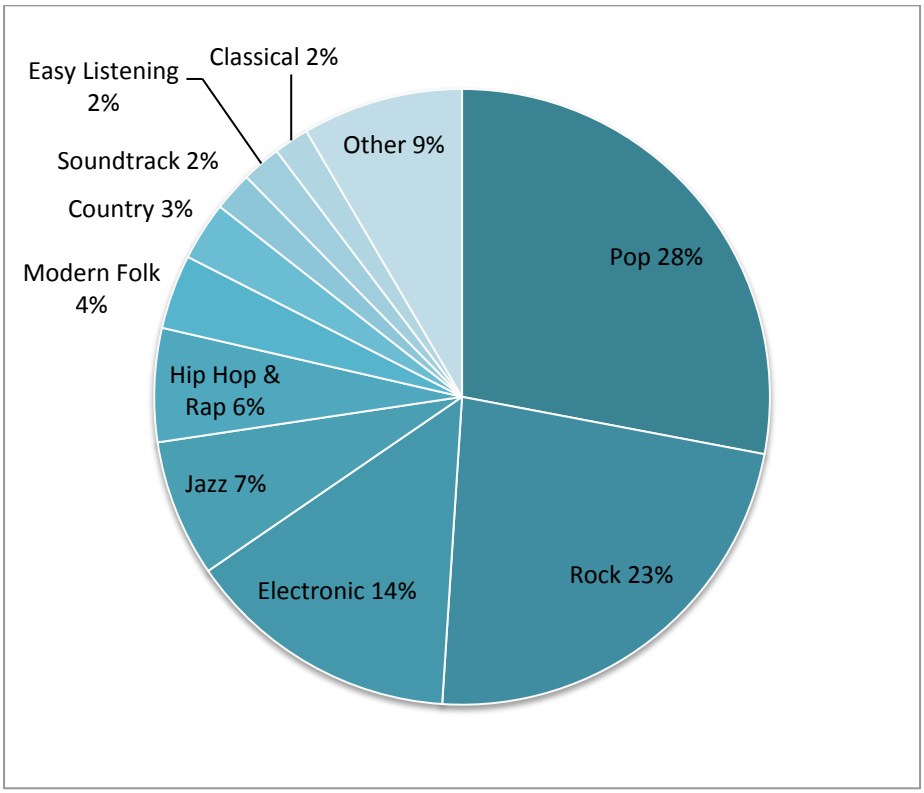

Figure 12 shows the distribution of encountered genres

[1] To see the full Wikipedia popular music genre hierarchy, go to http://en.wikipedia.org/wiki/List_of_popular_music_genres. This hierarchy does not include user-supplied genres such as 'alternative' or genres not considered popular such as 'classical'.

CHAPTER 5

ADDITIONAL CLASSIFICATION FEATURES

In this chapter we discuss the additional classification features that were gathered from the song audio files, lyrics, and tags. The chapter ends with a table of all classification features. In addition to information gathered during the survey (see Chapter 4), we extracted additional features for use in classification. These additional features fall into two categories: audio features and textual features. The additional classification features could only be gathered for some instances (i.e., songs with an artist and title or associations); these instances are described in the following sections.

5.1 AUDIO FEATURES

Songs were downloaded using the specified artist and title from legitimate sources such as iTunes. Songs were downloaded only when no ambiguity existed in the artist and title information entered by the participant. For example, if no artist was listed, the song was only downloaded when there was only one song with that title. The researcher handled spelling mistakes in the artist and title. Sometimes popular lyrics from the song were used as a title incorrectly but it was still clear which song it was (e.g., “Baby I like it” instead of “I Like It”). Songs were downloaded in various formats and converted to mp3 for processing. Songs that were longer than 10 minutes were not included for processing. Songs longer than 10 minutes tended to be DJ compilations including multiple songs and there is no way to tell which exact part of the compilation participants were listening to while completing the survey. Based on song title and artist information entered, 343 songs were downloaded and successfully analyzed.

Audio features were then extracted using the MIRToolbox [35] in MATLAB. Work by Juslin [31] has identified five types of features as important to the expression of emotion in music: mode, rhythm, articulation, intensity/loudness and timbre/spectrum (See Chapter 2). Of these five features we could analyze all but intensity/loudness, mainly because have no information

about how loud the music was playing when our participants were listening to it. For a detailed description of each function see the MIRtoolbox user manual [36]. These audio features were chosen because of their use in automatic genre classification and automatic musical mood classification [17,41,63].

Mode: We collected two features that describe the mode, specifically the key (e.g., C, B flat) and the modality (i.e., whether the key is major or minor). To compute the key we used *mirkey,* which determines the probable strengths of all possible keys and returns the one with the highest probability. To compute whether the key is major or minor we used *mirmode.* This function estimates the modality and returns a number, with negative numbers corresponding to minor modes and positive numbers corresponding to major modes. A number close to zero indicates an ambiguity between major and minor.

Rhythm: We collected two features that describe the rhythm of the music: tempo and pulse clarity. Tempo was collected using *mirtempo()*, which returns an estimation of the number of beats per minute throughout the song. In the case of songs with changing tempo, it returns the average. *mirtempo()* detects note onsets (i.e., where each note begins) and then looks for periodicities from the corresponding onset detection curve. Pulse clarity, using *mirpulseclarity()*, describes how easily a listener can perceive the tempo by detecting the relative strength of the beat (i.e. the relative strength of the pulsations within the music). For more information on how pulse clarity is modeled, see [34].

Articulation: Articulation is captured through two features: attack slope and average silence ratio (ASR). The attack slope, using *mirattackslope()*, detects the slope in the spectrum of the beginning of each note. It is an indicator of how aggressively each note is played. The overall average attack slope was used as a feature. ASR detects the ratio of silence in the music by detecting the number of frames with less than average energy within a certain threshold. This was done with the function *mirlowenergy('ASR').* See Feng [17] for more information on how ASR is calculated.

Timbre: Timbre features are descriptions of the sound spectrum of the music. The sound spectrum shows the distribution of energy frequencies of the sound. Seven features were used: brightness, rolloff, spectral flux, spectral centroid, average sensory roughness, mel-frequency cepstrum (MFCC), and low energy. Brightness, using *mirbrightness(),* is a measure of the amount of energy above a specific cutoff point in the spectrum. Rolloff, using *mirrolloff(),* estimates the amount of high energy in the spectrum by finding a specific frequency such that 85% of the energy is contained above that frequency. For more information about rolloff, see [63]. Spectral flux, using *mirflux(),* divides the spectrum into frames and then calculates the difference between the spectrum in each successive frame. We used the average of this value taken across the duration of the song. Spectral centroid, using *mircentroid(),* calculates the geographical center of the spectrum. Sensory roughness, calculated using *mirroughness(),* corresponds to when several sounds of nearly the same frequency are heard, causing a "beating" phenomenon. High roughness corresponds to harsher music with more "beating." Average sensory roughness taken over the duration of the entire song was used as a feature. MFCC, using *mirmfcc()*, describes the sound by separating it into several bands. A full description is out of the scope of this work; see [46,53]. Low energy, using *mirlowenergy(),* separates the spectrum into frames and then calculates the percentage of frames with less than average energy. Low energy is similar to ASR but it does not use a threshold; see [63].

5.2 TEXTUAL FEATURES

Three sets of textual features were collected and analyzed with Linguistic Inquiry and Word Count (LIWC) [48] – a textual analysis program. These textual features included lyrics, socially created tags and mental associations. LIWC calculates the percentage of words in a block of text that fall into any of 80 word categories. The output percentages were used as features. Previous work achieved improvements to the classification of musical mood by using lyrics and tags [6,66]; however, our use of an affective text analysis tool is novel.

5.2.1 Lyrics

Lyrics were collected from various online sources for 270 songs (recall that 343 songs were downloaded). Some songs did not contain lyrics. Other songs were written in a foreign language; only English lyrics were analyzed. Songs that were mainly English but contained a few foreign words were included in our analysis. There were sometimes differences between the lyrics collected and the words heard when listening to the song itself. For example, some written lyrics contained notations indicating that certain verses or phrases were repeated (e.g., "x2," "chorus," "bridge"). These were manually removed and replaced with the repeated text. Some lyrics included markup on non-word sounds (e.g., "oooo") or names indicating the singer responsible for the lyrics. These were left as is in the text. As some of the lyrics came from websites that allowed user-entered lyrics, spelling mistakes existed. These were corrected as they were found. Also words ending prematurely with a "'" (e.g., "singin'") to indicate a slight different in pronunciation were corrected to the full word (e.g., "singing") to ensure that the affective nature of the text was captured by LIWC. The song lyrics were analyzed using LIWC and the output percentages for each category were used as features.

5.2.2 Tags

Tags were downloaded using the artist and title specified from last.fm, a site that combines similar artists into radio stations based on socially created tags. These tags can be added by any of the site's many listeners. The complete set of all tags associated with each song was analyzed using LIWC and the output percentages for all categories were used as features. Tags were collected for 291 instances; some instances may not have been in Last.fm's database as they were too obscure or were foreign music. Other instances may have been listed under a slightly different title (e.g., with additional information in the song title such as featured artists).

5.2.3 Mental Associations

Mental associations were collected during the in-situ study. These were analyzed using LIWC and the output percentages were used as features; 335 instances had mental associations.

5.2.4 LIWC

LIWC is a tool that counts the number of words that are present in a number of categories [48]. Only words in the LIWC dictionary are counted in a category; the dictionary consists of 4500 words, and words may fall into multiple categories. Words are captured by word stems; "hungr*" will capture any words beginning with "hungr" and having any combinations of letters, numbers, hyphens or punctuation. So "hungry," "hungriest" and "hungrier" will all be counted.

All 80 categories produced by a LIWC analysis fall into four different types of words: linguistic processes, psychological processes, spoken categories and personal concerns. Linguistic processes include standard linguistic definitions: pronouns, articles, verbs, adverbs, prepositions, conjunctions, quantifiers, negations numbers, and swear words. Psychological processes are further broken down into social processes (family, friends, humans), affective processes (positive affect, negative affect, anxiety, anger, sadness), cognitive processes (insight, causations, discrepancy, certainty, inhibition, inclusive, exclusive), perceptual processes (seeing, hearing feeling), biological processes (body, health, sexuality), and relativity (motion, space, time). Personal concern includes categories for work, achievement, leisure, home, money, and religion. Spoken categories include assent, non-fluencies, and fillers. See Appendix E for a breakdown of the LIWC dictionary.

5.3 SUMMARY

There were 293 features used in total. This included 240 textual features (Section 5.2) derived through LIWC, 2 additional features describing associations (Section 4.3.3), 34 audio features (Section 5.1), 4 features describing mood (Section 3.4.1 and Section 3.4.3), and 13 features collected through the music survey (Section 3.4). See Table 2 for a summary of classification features.

Table 2 Summary of Classification Features

Features	**Feature Type**	**Number of Features**	**Description**
Lyrics	Musical Features	80	LIWC output
Tags	Musical Features	80	LIWC output
Audio Features	Musical Features	34	Assorted features describing rhythm, articulation, timbre and mode
Affective State	Affective Feature	2	Personal arousal and personal valence.
Associations	Listening Context	82	LIWC output, orig. association, association category
Survey Questions	Listening Context	13	Location, activity, company, choice, reason, in background, title, artist, genre, genre category, age, sex, PID

CHAPTER 6

MODELING RESULTS

This chapter will discuss the methods employed in modeling musical mood, namely the methods for dealing with sparse data and class skew. We split the features into three different feature sets and present the results of our models of musical arousal and musical valence.

6.1 FEATURE SETS

We used a number of feature combinations in creating our models, which can be summarized as three feature sets (see Chapters 4 and 5 for details on the features):

Musical Features: Our first feature set used acoustic features, lyrical features, and tag features, as these features were used in previous models based on laboratory-gathered data sets of a single genre. There were 198 different features in this set. Note that we are not actually testing previous models, only similar feature sets.

Musical Features + Affective Features: Our second feature set used all the musical features but added personal arousal and valence for a total of 200 different features.

Musical Features + Listening Context: Our third feature set combined musical features with the listening context collected in our study for a total of 296 features.

6.2 DATA

Several steps were taken to clean the data for modeling purposes. There were two issues that needed to be dealt with: sparse data and class skew.

6.2.1 Sparse Data

Because users optionally specified artist and title, many instances in our data set contained blank entries for these fields, making it impossible to download the song and extract audio features or collect lyrics. Furthermore, because many songs were foreign, it was difficult to find them on iTunes, it was impossible to analyze the lyrics with LIWC (which only works on English words), and the songs were likely not available on Last.fm, meaning that no tags could be gathered. In addition to these problems, some songs did not have lyrics and some were too obscure to find accurate lyrics for without access to the original sources (e.g., track information in CD booklets). Furthermore, the audio processing scripts for some song files failed during processing due to unknown errors in the files. We could only find and process the song to collect audio features for 45% of the data; this led to the musical feature part of the data being sparse.

Because we are using musical features to build models, this sparse data is a problem. Sparse data can mean lower classification rates. We could solve this problem by collecting more data with these features included, but this was not possible without running the study with a longer timeframe and a larger number of participants. To improve coverage of the data set, we restricted the instances to only songs with a title specified. This meant that some songs would still not have any data, but the percentage of instances with audio features, lyrics, and tags would be greatly improved. We did not restrict the data to only points with audio features because some songs with titles had lyrics or tags but no audio features. Furthermore, restricting the data set further would greatly reduce the number of point used after undersampling (see 6.2.2). Audio features only covered 45% of the original data set but covered 66% of the data set restricted to titles only; the percentage of instances with lyrics went from 37% to 55% and tag information went from 38% to 55%. See Figure 13 for the improvements.

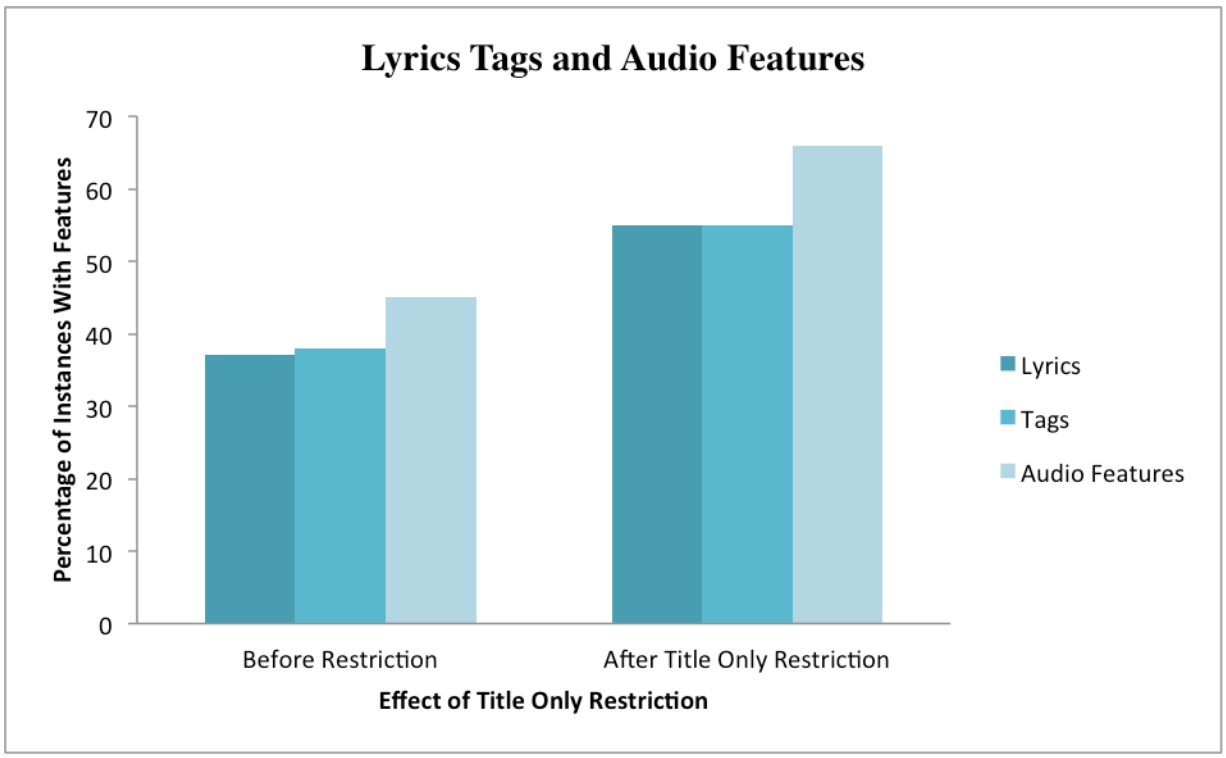

Figure 13 Shows the improvement in coverage of the audio features, lyrical features and features based on tags.

6.2.2 Class Skew

Due to the "in the wild" nature of the study, musical arousal and musical valence had an uneven distribution of responses – an effect known as class skew. Participants generally listened to songs with high arousal and high valence and other musical moods were not as well represented in our data set. To prevent over-fitting of the model due to class skew, musical arousal and musical valence were binned into two levels: low, and high. Neutral instances were ignored. See Figure 14 for the original distribution of responses before binning and 6.3 and 6.4 for the distribution of responses after binning responses into low and high. By binning responses into low and high we are effectively modeling results into each of the quadrants of A-V space as previous work has done – although we are modeling arousal and valence separately.

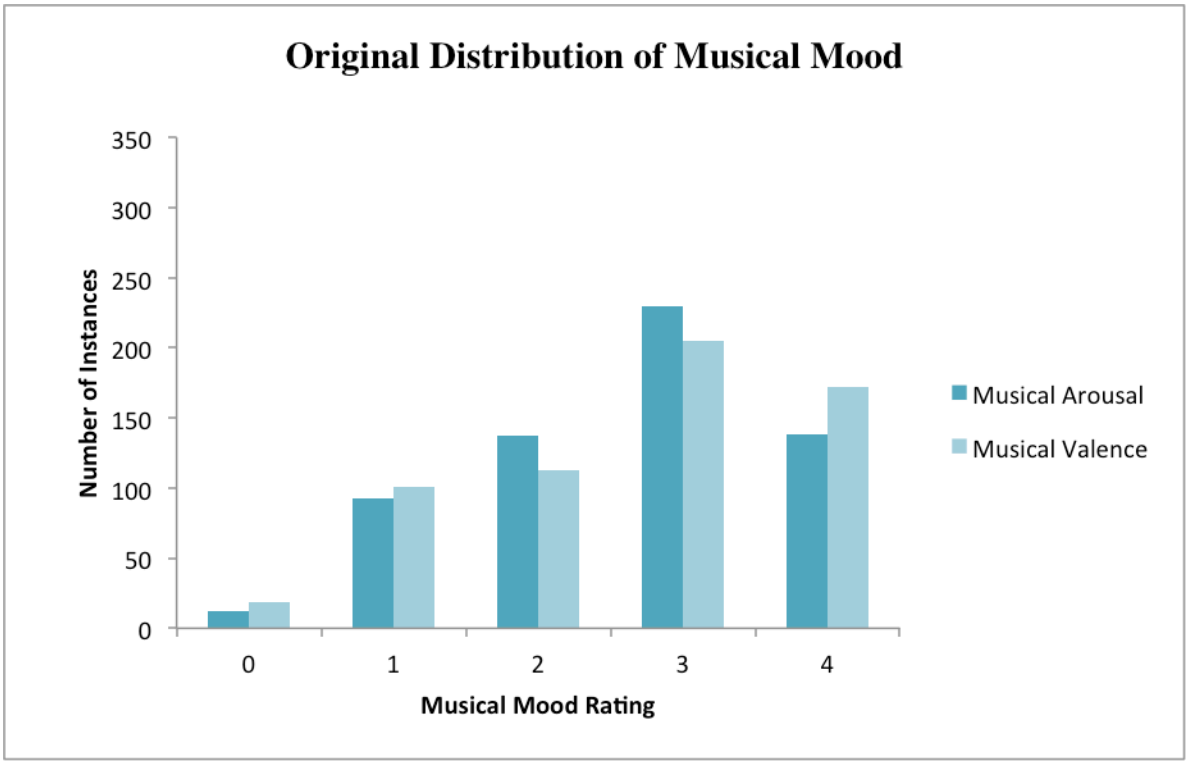

Figure 14 Shows the original distribution of responses for musical arousal and musical valence. 0 is low and 4 is high. Notice that both musical arousal and musical valence are clustered around mid-high -- high and very few participants responded with low arousal or low valence.

Undersampling, a technique that selects a random number of instances to obtain an equal distribution and eliminate class skew, was used. This lowered the total number of instances from 610 to 122 when modeling musical arousal and 156 when modeling musical valence. To avoid any effects caused by the specific set of random instances chosen, this process was completed five times, and the average accuracies of all runs are reported. See Figures 15 and 16 for the distribution of responses after undersampling.

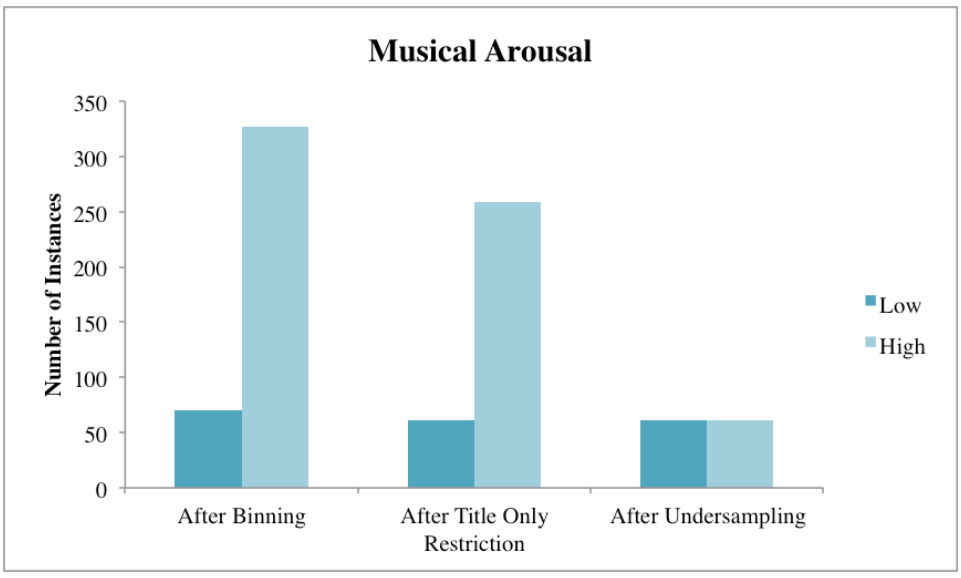

Figure 15 Shows the distribution of musical arousal after binning into low and high, after title only restriction and after undersampling.

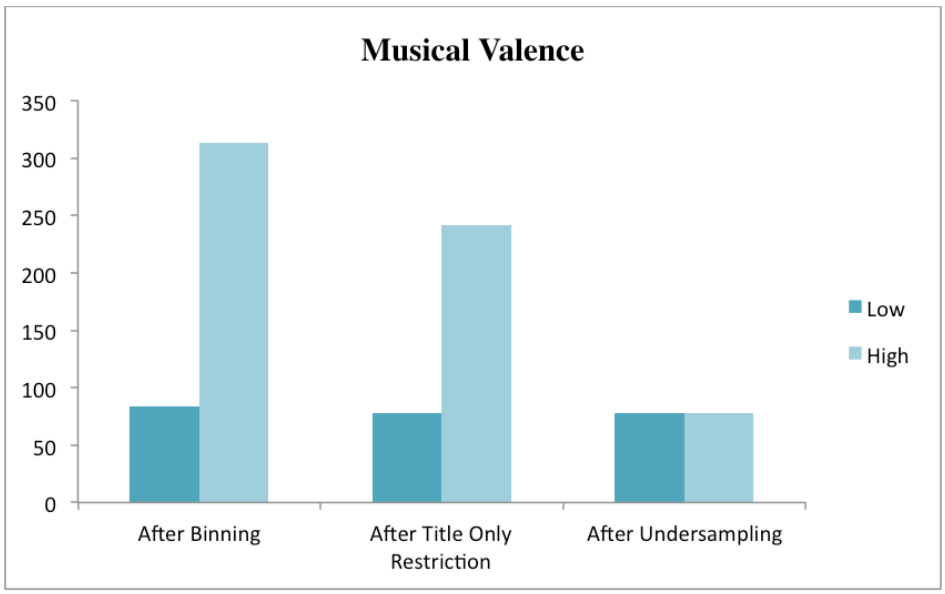

Figure 16 Shows the distribution of musical valence after binning, restriction and undersampling.

6.3 RESULTS

All models were created in Weka [19] using Bayesian Network classifiers, Markov Blankets and tenfold cross-validation. We used Bayesian Networks because they readily handle missing or sparse data [25] and were used in previous work [55]. We used Markov blankets to help find an optimal feature set for classification. Markov blankets assume that every node in the network is only affected by the other nodes in the network on which it is dependent [67]. Finally, we use tenfold cross-validation to avoid any effects caused by the set of training and testing instances chosen in the model. We modeled musical arousal and musical valence separately, using each feature set. See Figure 17 for classification accuracies.

Musical Features: Using only musical features (audio features, lyrics, and tags), musical arousal had an average classification accuracy (over the five runs of undersampled data, see section 6.2.2) of 59.5% (SD=3.1, kappa[2]=0.1984). Musical valence had a classification accuracy of 53.0% (SD=6.8, kappa=0.0604). Using a one-sample t-test on the results of the five runs with the undersampled set, we show musical arousal (t_4=6.74, $p<0.005$) significantly exceeds the performance of a random classifier (chance = 50%), whereas musical valence does not (t_4= 0.99, $p>0.377$). However, the results for both arousal and valence are lower than previously reported classification accuracies of homogenous laboratory data sets.

Musical Features + Affective State: When we combined affective features (personal arousal and valence) with musical features, musical arousal had a classification accuracy of 60.3% (SD=4.6, kappa=0.2121). Musical valence had a classification accuracy of 60.2% (SD=4.6, kappa=0.2074). Using one-sample t-tests, we show that both musical arousal (t_4=4.99 $p<0.01$) and musical valence (t_4=4.70, $p<0.01$) performed significantly better than chance (50%), and we achieved improved classification accuracies of musical arousal and musical valence by using a combination of affective and musical features; however, they were not statistically-significant improvements (music arousal t_8=0.32, $p>.75$, musical valence t_8= 1.92, $p>0.09$)

---

[2] Kappa is a statistic that describes the inter-rate agreement of the model (i.e., how much can be attributed to chance). Kappa varies between 0 and 1, a kappa of 0 means that the agreement is equivalent to chance.

Musical Features + Listening Context: When we combined musical features with listening context features, musical arousal has a classification accuracy of 67.4% (SD=1.7, kappa=0.3437). Musical valence has a classification accuracy of 75.7% (SD=1.5, kappa=0.5133). Using one-sample t-tests, we show that the models of both musical arousal (t_4=23.54, p <0.0001) and musical valence (t_4=38.75, p<0.0001) performed significantly better than chance (50%). We achieved significant gains in classification accuracy in both musical arousal (t_8=4.96, p<0.01) and musical valence (t_8=7.29, p<0.0001) over using only musical features. Similarly, we achieved significant gains in musical arousal (t_8=3.23, p<0.01) and musical valence (t_8=6.85, p<0.0001) over musical features and affective features combined.

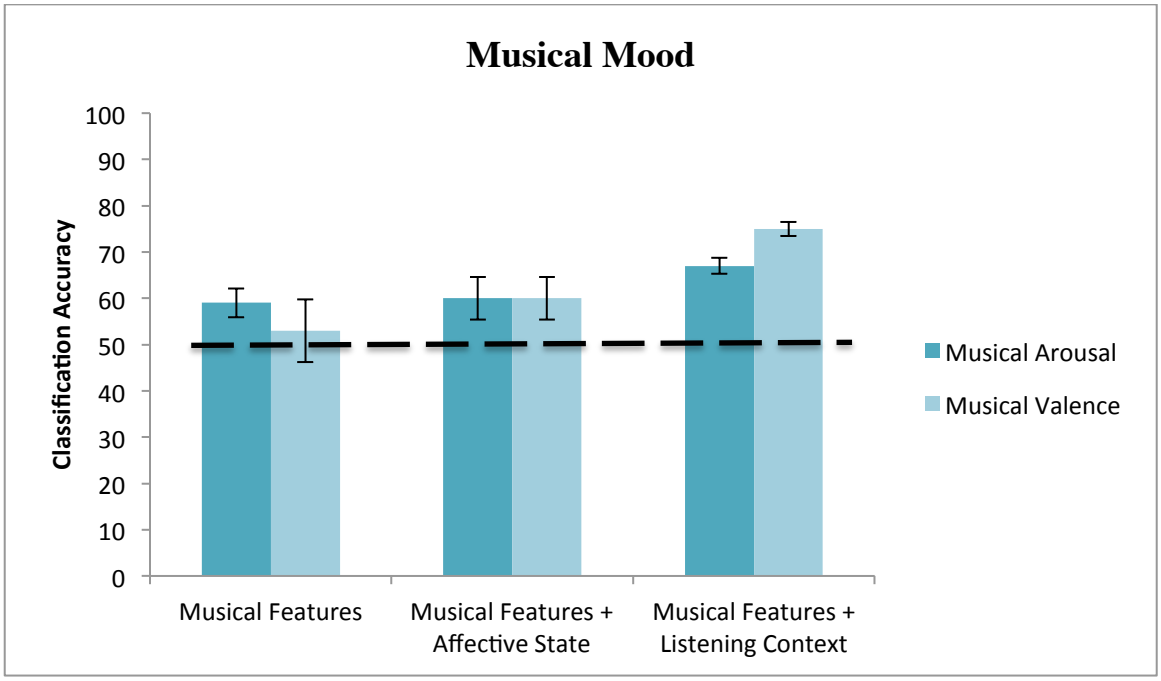

Figure 17 Shows the classification accuracies for each feature set. The dotted line is chance. All models performed better than chance; however, the valence model using musical features alone was not significantly better than chance.

CHAPTER 7

DISCUSSION

Our results showed that people listen to music with a generally happy mood (higher than neutral arousal and valence), that applying previous models to real-life data gathered in situ may cause the approaches of previous models to fail, that the performance of our models is limited by sparse data, and that context is important when modeling musical mood. In this chapter, we discuss these results and offer design implications.

7.1 LIMITATIONS

In this section, we describe the limitations of the ESM methodology, of our study, and of our models.

7.1.1 ESM Methodology

ESM methods have weaknesses. We could not capture music, audio features of the music (such as tempo), or physiological measures of affect as easily or as accurately as we could in a laboratory setting. In addition any data collected via ESM will be somewhat messy. For example, we cannot control the sound quality, volume or amount of background noise, as we could in a laboratory with headphones and a quiet room. There is no way to tell what part of the song participants were listening to when the survey was answered the survey, or even if they listened to the entire song.

Furthermore, ESM does not necessarily capture a representative selection of responses. First, ESM surveys are interruptive to the current task. Music listening during some activities, such as driving, is difficult to capture due to the interruptive nature of the survey (driving while operating a mobile device is also illegal.) Some activities, such as socializing, are unlikely to be interrupted by participants in order to complete a survey, as this would be considered rude. Likewise, activities such as bathing while listening to music, may not take place in the vicinity of

the phone no matter how diligent the participant is in carrying the phone with them. We would expect that these activities would therefore be underrepresented. Other activities may be overrepresented. For example, it may be easier for participants to keep the phone stationary at their desk as they work and participants may enjoy filling out the survey as a distraction to working. Due to the interruptive nature of the responses, there is no way to produce an evenly distributed or fully representative set of responses.

7.1.2 Study

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. For example, 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.

Only one category of responses in our study yielded homogenous answers – musical mood. Many previous studies have assumed that people listen to music with four emotional categories (happy, sad, fear, anger)[31]; 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 emotion categories only in certain circumstances (e.g., sad music after a breakup) and that these circumstances occurred rarely 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 the musical mood of music from real-life listening clusters around positive arousal and valence. So while previous work modeled a single musical genre with multiple moods, our real-life data set shows music from many genres, but clustered around a happy musical mood.

Timestamps were collected for each survey with the intent to study effects of time of day on listening context and musical mood. However, these were collected in the wrong format. While moving the data from SPSS to Weka, this information was lost. As a result we were not able to use at the time the survey was collected as a feature.

7.1.3 Model

While a classification accuracy of 75% is much improved over a two-state random classifier, or one based on auditory features alone, a music recommender suggesting songs with the wrong mood a quarter of the time may result in a negative user experience. This can be circumvented in a few ways. First a music recommender can select tracks from a personal music library; users are more likely to enjoy their own music even if the recommendation is off. Second, a playlist rather than a single song could be recommended so that a majority of the music recommended is suitable. Third, combining this model with existing recommendation systems that use clustering of similar genres and artists could further improve existing prediction rates. We could also improve the classification rates of our models by collecting a more comprehensive data set, as our low number of instances may have contributed to lower classification accuracies. Conducting a longer study with more participants could collect a more comprehensive data set. Finally, 75% is the maximum accuracy of the model, achieved when only predicting musical arousal. Prediction of both musical valence and musical arousal at the same time would lower this maximum accuracy.

Furthermore, the models based on musical features alone may have had low classification accuracy due to a low coverage of features. Gathering these features is a problem that must be addressed. We did not make artist and title required fields in the survey as we did not want participants to quit the survey because they did not know the artist or title or to merely fill in unusable information such as "?." In the future, we could reward participants for filling in these fields in the same manner that we rewarded them for completing surveys (i.e., increasing the study payout for each artist and title completed). Participants could be expected to fill out a certain percentage of artists and titles. This would require feedback (such as a progress bar) so that participants would know if they were meeting the study requirements. Some mobile devices also have applications that will identify songs by using the microphone on the device to listen to a small portion of the song. Using a free API called Echoprint, it may be possible to integrate this ability into the ESM software used in the study. Gathering the artist and title may not be enough – in our study, we encountered many instances where the song was not on popular music stores such as iTunes. Another option would be to create a plug-in for an existing media player

(e.g., iTunes), which would automatically gather the song and send it to a server for the processing required to extract acoustic features. The problem with this method is it severely restricts context to only instances in which participants are listening using iTunes, on their computer, or with an Internet connection; however, it would enable 100% coverage of audio feature across all instances.

7.2 STRENGTHS

Despite the limitations, ESM captures a wide range of real-life experiences in naturally occurring contexts. It would be impossible to artificially recreate this wide range of contexts in a laboratory setting. For example, one cannot force someone to listen to a song "to express or release emotions" in a lab. Every context captured would require a different experimental setting. For example, to capture different locations would require several experiments in the participant's home, work place and different public places. Furthermore popular music is time dependent. The music captured by our study is a snapshot of what our participants listened to in the same two-week period.

We captured a dataset very different from the homogenous data used in previous work. Participants listened mainly to their own music libraries. This means the music captured is not limited to a representative set of Western classical or Western popular as most previous studies have been. Many participants listened to some foreign non-English music. This may have been due to the specific set of participants used in the study (i.e., some of them may have not been native English speakers), but we did not capture language information in our demographics survey so we have no way to control for language. However, no participant listened to only foreign music; they all listened to a large selection of English, Western music. The selection of music captured by our study was far from homogenous.

Our model is the only model of musical mood to adapt to listening context using in-situ data. It was formed on a data set with sparse coverage of some features, suggesting that creating Bayes Net models on data sets with more complete coverage would increase classification accuracy.

We also use two techniques, binning and undersampling, that deal with the class skew encountered when using in-situ data. These techniques may be unavoidable on this type of data set. Even if we provided a music library with an equal number of instances in each class, if the participant has a choice over the songs, they may still choose only happy music. In fact, it is possible that participants had a music library with songs in every combination of arousal and valence, but generally chose the happy songs to listen to during the study.

We chose to follow the approach used by Schuller et al. [55] (i.e., model binary levels of arousal and valence separately), rather than the approach used by Lu et al. [41] (i.e., modeling four quadrants of A-V space) for two reasons. The first is that due to unequal distribution of responses, we simply did not have enough data points representing each of the four quadrants. For example, the low-arousal–low-valence quadrant had only 31 instances before applying the title only restrictions and undersampling, which would likely reduce it to fewer than 10 usable instances. The second reason is that we wanted any future music recommender using this model to make predictions based on either musical arousal or musical valence or a combination of both. Our classification accuracy for musical arousal are 9% lower than Schuller et al., however musical valence was 2% higher, suggesting that our results are at least comparable. The contribution we make, over and above Schuller et al.'s work, is that our models are applied to heterogeneous data sets based on real-life listening experiences rather than a set of songs in the Western popular genre that is labeled in a laboratory. We also show the importance of listener affective state and listener context in modeling musical mood.

7.3 IMPORTANT OF CONTEXT

Our work has shown that listening context is important to models of musical mood. It may be possible that context is important when modeling musical mood because participants rate musical mood differently depending on their context. For example, a user may rate the same song differently depending on whether they are working alone or cooking with friends. We cannot confirm or refute this with our data set, as one would want the same songs played in a

variety of listening contexts – in our study, songs were only encountered once on average. It may be possible to use statistical tests such as an ANOVA on our dataset to determine if there is a difference in musical mood depending on context, however these tests will not tell us if this difference occurs because participants are rating the songs differently in difference contexts or merely choosing music with different musical moods based on listening context. For example, a user may generally choose to listen to music with high arousal when exercising and low arousal when eating dinner. In that case our model predicts the type of musical mood listeners want to listen to, based on listening context, which is very useful for automatically generating playlists.

Similarly, affective state of the user was shown to be important in our models of musical mood. Participants may rate musical mood differently depending on their affective state. This is a tricky relationship to investigate as the music itself has a hand in inducing an affective state in a listener. Any correlation found between musical mood and affective state does not show directionality of the relationship. To examine the relationships between listening context, musical mood, and affective state, we could provide users with representative samples in a music library. By listening to (and rating) the same song in a variety of contexts and affective states, the relationship between these three factors might be made clear.

7.4 IMPLICATIONS FOR DESIGN

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 by applying models across genres. Music recommenders may want to collect and use listening context in their recommendations. This can be done using several methods. The first method is to use a short survey much as was done in our ESM study. The surveys can be short as some of the information could be inferred (e.g., if the user is socializing one can make a general assumption that the location is in some degree public and the user is with people they know). Alternatively, if the music recommender operates on a mobile device, sensors available on mobile devices can be used.. We could use

these sensors to gather information about the ambient noise (using the microphone), ambient light (using the camera), participant location (using the GPS), or the amount of activity the participant is engaging in (using the accelerometer). We could use the information to infer context. If they are connected to a public Wi-Fi, or if we know that their location is public (e.g., on the University Campus) we can assume their listening experience is taking place in a public location. If the application detects voices (using the microphone), we could assume that the user is not alone. We could also use sensors to shorten the survey questions. For example, if we determine that the user is active (using accelerometer activity), we modify the questions to ask them if they are exercising or dancing. If we notice they are moving large distances quickly (using the GPS), we can ask them if they are driving or travelling as a passenger. Finally sensors could be added to the participant's environment that we could use to infer context. We could, for example, place a pressure sensor underneath a participant's couch. If they are on their couch they are most likely relaxing. An infrared sensor could be used to count the number of people in a participant's house, allowing us to determine if the participant is alone or with people (that they most likely know if they have allowed them into the house).

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 want 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. Both of these recommendation systems require musical mood to be pre-calculated. Some older versions of MIRtoolbox contain an automatic classification function that returns an arousal and valence value for a song. Musical mood could also be queried from users and the average value stored. Or models such as ours could be refined and extended to multiple contexts, so that they are robust enough to be used in a context-aware music recommender system.

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.

7.5 FUTURE WORK

Based on the results of this work, the next step is to create a context-aware music recommender system. This system could take in the context of the listening experience and use this context to compile a playlist. Based on our models, the system would recommend a musical mood listeners are likely to enjoy, and then create playlists of songs with this specific musical mood, (based on a data set labeled from musical features). The system could also make suggestions of songs for purchase the user might enjoy. Predictions would need to be evaluated through a user study, conducted in-situ, to preserve the importance of context.

To create the underlying model for this music recommender, a larger in-situ data set must be collected. The study would need to run for a longer time period (i.e., months) with a larger pool of participants. If participants received bonuses for filling out genre, title and artist, and were asked to provide a copy of their music library at the end of the study for audio feature processing, then we could achieve better coverage of features. This larger, more comprehensive data set would help to improve classification accuracies.

CHAPTER 8

CONCLUSION

Current models of musical mood have two major failings that prevent their applicability. First the data used is of a single genre. Second, the music listening experiences are generally collected in a laboratory setting. Real-life listening experiences are not in line with the data sets used in previous models. Models based on these clean, noiseless data sets may fail when applied to naturally gathered in-situ data.

To solve these two problems, we collected listening experiences (including listening context) using experience-sampling methodology (ESM), and used this data to model musical mood. We first created Music Survey – ESM software that ran on Android smartphones. This software deployed surveys about once per hour. We then handed these phones to 20 participants for a period of two weeks. Participants were paid per number of surveys they filled out to encourage participation. We downloaded the song, lyrics and socially created tags for each song where possible. Lyrics and tags were analyzed with the Linguistic Inquiry Word Count Tool (LIWC).

We created Bayes Net Models using data mining software called WEKA. We separated the features into three feature sets: music information (audio features, lyrics, tags), musical information + affective state, and musical information + listening context. We modeled musical arousal and musical valence for each feature set.

8.1 CONTRIBUTIONS

In this section, we discuss our three main contributions

Our first contribution is that real life listening experiences are heterogeneous and not in line with previous data sets used in modeling musical mood. In particular, a large number of genres and languages (and therefore cultures) were encountered. Unlike listening to music in a laboratory,

music listening was usually secondary to other activities. Finally, music was usually happy, unlike previous work that has assumed people listen to music with four different musical moods [31].

The second contribution we make is that we modeled musical mood from a data set collected in-situ during a user's daily life. We are the first to do so.

Finally, we showed that we can successfully model musical mood on in-situ data with classification accuracies higher than chance; however, classification accuracies were lower than in some previous models. We successfully classified musical arousal with a classification accuracy of 60% and musical valence with an accuracy of 53% when using only musical features (audio features, lyrics and tags). Adding affective state or listening context further improved classification accuracies. We successfully classified musical arousal with a classification accuracy of 60% and musical valence with an accuracy of 60% when using both musical features and affective state. We successfully classified musical arousal with a classification accuracy of 67% and musical valence with an accuracy of 75% when using both musical features and listening context. These classification results may be lower than in prior work for three reasons: the heterogeneous nature of the data set, low coverage of some features (i.e., no song title, song not on iTunes, failures during processing), or because of noise introduced into the data set by participants (e.g., mistakes while using the touch screen on the device or a misunderstanding of key terms).

8.2 SUMMARY

Musical mood, the emotion expressed by a piece of music, can be automatically classified from in-situ data with fairly good results when listening context is included as a classification feature. Our work has two implications. First researchers creating models of musical mood should include listening context as a feature as it may improve classification accuracies. Second, designers of systems using these models of musical mood, such as music recommenders, may

also want to include listening context where possible as it may improve system performance by making better predictions.

REFERENCES

1. Ableson, F. Build dynamic user interfaces with Android and XML. *IBM developerWorks*, 2010. http://www.ibm.com/developerworks/xml/tutorials/x-andddyntut/index.html.

2. Bauer, J., Jansen, A., and Cirimele, J. MoodMusic. *Proceedings of the 24th annual ACM symposium adjunct on User interface software and technology*. UIST'11

3. Becker, J. Anthropological Perpectives on Music and Emotion. In *Music and Emotion:Theory and Research*. Oxford University Press, New York, NY, USA, 2001, 135–160.

4. Beedie, C., Terry, P., and Lane, A. Distinctions between emotion and mood. *Cognition & Emotion 19*, 6 (2005), 847–878.

5. Berlyne, D.E. *Aesthetics and Psychobiology*. Appleton Century Crofts, New York, NY, USA, 1971.

6. Bischoff, K., Firan, C.S., Paiu, R., Nejdl, W., Laurier, C., and Sordo, M. *Music Mood and Theme Classification - A Hybrid Approach*. ISMIR 2009

7. Bradley, M.M. and Lang, P.J. Measuring emotion: The self-assessment manikin and the semantic differential. *Journal of Behavior Therapy and Experimental Psychiatry 25*, 1 (1994), 49 – 59.

8. Chanel, G., Kronegg, J., Grandjean, D., and Pun, T. Emotion Assessment: Arousal Evaluation Using EEG's and Peripheral Physiological Signals. In B. Gunsel, A. Jain, A. Tekalp and B. Sankur, eds., *Multimedia Content Representation, Classification and Security*. Springer Berlin / Heidelberg, 2006, 530–537.

9. Chung, J.-W. and Vercoe, G.S. The affective remixer: personalized music arranging. *CHI '06 extended abstracts on Human factors in computing systems*, ACM (2006). CHI 2006

10. Coan, J.A. and Allen, J.J.B., eds. *Handbook of emotion elicitation and assessment.* Oxford University Press, New York, NY, US, 2007.

11. Cooke, D. *The Language of Music*. Oxford University Press, London, 1959.

12. Crossen, A., Budzik, J., and Hammond, K.J. Flytrap: intelligent group music recommendation. *Proceedings of the 7th international conference on Intelligent user interfaces*, ACM (2002), 184–185.

13. Davies, J.B. *The Psychology of Music*. Hutchinson, London, 1978.

14. Davies, S. Philosophical Perspectives on Music's Expressiveness. In *Music and Emotion:Theory and Research*. Oxford University Press, New York, NY, USA, 2001, 23–44.

15. Ekman, P. *Basic Emotion*. John Wiley & Sons, Ltd., 2005.

16. Epp, C., Lippold, M., and Mandryk, R.L. Identifying emotional states using keystroke dynamics. *Proceedings of the 2011 annual conference on Human factors in computing systems*, ACM (2011), 715–724.

17. Feng, Y., Zhuang, Y., and Pan, Y. Popular music retrieval by detecting mood. *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval*, ACM (2003), 375–376.

18. Forgas, J.P. Mood and judgment: The affect infusion model (AIM). *Psychological Bulletin 117*, 1 (1995), 39–66.

19. Fritz, T., Jentschke, S., Gosselin, N., et al. Universal Recognition of Three Basic Emotions in Music. *Current Biology 19*, 7 (2009), 573 – 576.

20. Greasley, A.E. and Lamont, A. Exploring engagement with music in everyday life using experience sampling methodology. *Musicae Scientiae 15*, 1 (2011), 45 –71.

21. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I.H. The WEKA Data Mining Software: An Update. *SIGKDD Explorations 11*, 1 (2009).

22. Han, B.-J., Rho, S., Jun, S., and Hwang, E. Music emotion classification and context-based music recommendation. *Multimedia Tools and Applications 47*, (2010), 433–460.

23. Hancock, J.T., Gee, K., Ciaccio, K., and Lin, J.M.-H. I'm sad you're sad: emotional contagion in CMC. *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, ACM (2008), 295–298.

24. Hatfield, E., Cacioppo, J.T., and Rapson, R.L. Emotional Contagion. *Current Directions in Psychological Science 2*, 3 (1993), pp. 96–99.

25. Heckerman, D. *A Tutorial On Learning with Bayesian Networks*. Microsoft Research, 1995.

26. Hektner, J.M., Schmidt, J.A., and Csikszentmihalyi, M. *Experience sampling method: Measuring the quality of everyday life.* Sage Publications, Inc, Thousand Oaks, CA, US, 2007.

27. Hevner, K. Experimental studies of the elements of expression in music. *The American Journal of Psychology 48*, (1936), 246–268.

28. Juslin, P. and Laukka, P. Expression, Perception, and Induction of Musical Emotions: A Review and a Questionnaire Study of Everyday Listening - Journal of New Music Research. *Journal Of New Music Research 33*, 3 (2004), 217–238.

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

30. Juslin, P.N., Liljeström, S., Västfjäll, D., Barradas, G., and Silva, A. An experience sampling study of emotional reactions to music: Listener, music, and situation. *Emotion 8*, 5 (2008), 668–683.

31. Juslin, P.N. Cue utilization in communication of emotion in music performance: Relating performance to perception. *Journal of Experimental Psychology: Human Perception and Performance 26*, 6 (2000), 1797–1812.

32. Khosrowabadi, R., Wahab, A., Ang, K.K., and Baniasad, M.H. Affective computation on EEG correlates of emotion from musical and vocal stimuli. *Neural Networks, IEEE - INNS - ENNS International Joint Conference on 0*, (2009), 1590–1594.

33. Kim, Y.E., Schmidt, E.M., Migneco, R., et al. Music Emotion Recognition: a State of the Art Review. *11th International Society for Music Information and Retrieval Conference*, (2010).

34. Lartillot, O., Eerola, T., Toiviainen, P., and Fornari, J. Multi-Feature Modeling of Pulse Clarity: Design Validation and Optimization. ISMIR 2008 (2008).

35. Lartillot, O. and Toiviainen, P. A Matlab Toolbox for Musical Feature Extraction from Audio. *Proceedings of the 10th International Conference on Digital Audio Effects*, (2007).

36. Lartillot, O. MIRtoolbox Users Manual 1.3.4. 2011.

37. Lavy, M.M. Emotion and the Experience of Listening to Music A Framework for Empirical Research. 2001.

38. Livingston, I., Nacke, L., and Mandryk, R. Influencing Experience: The Effects of Reading Game Reviews on Player Experience. In J. Anacleto, S. Fels, N. Graham, B. Kapralos, M. Saif El-Nasr and K. Stanley, eds., *Entertainment Computing – ICEC 2011*. Springer Berlin / Heidelberg, 2011, 89–100.

39. Livingston, I.J., Nacke, L.E., and Mandryk, R.L. The impact of negative game reviews and user comments on player experience. *ACM SIGGRAPH 2011 Game Papers*, ACM (2011), 4:1–4:5.

40. Lonsdale, A.J. and North, A.C. Why do we listen to music? A uses and gratifications analysis. *British Journal of Psychology (London, England: 1953)*, (2010).

41. Lu, L., Liu, D., and Zhang, H.-J. Automatic mood detection and tracking of music audio signals. *Audio, Speech, and Language Processing, IEEE Transactions on 14*, 1 (2006), 5 – 18.

42. Lundqvist, L.-O., Carlsson, F., Hilmersson, P., and Juslin, P.N. Emotional responses to music: experience, expression, and physiology. *Psychology of Music 37*, 1 (2009), 61 –90.

43. Mandryk, R.L., Atkins, M.S., and Inkpen, K.M. A continuous and objective evaluation of emotional experience with interactive play environments. *Proceedings of the SIGCHI conference on Human Factors in computing systems*, ACM (2006), 1027–1036.

44. Mandryk, R.L. and Atkins, M.S. A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. *International Journal of Human-Computer Studies 65*, 4 (2007), 329–347.

45. Meyer, L.B. *Emotion and Meaning in Music*. Chicago University Press, Chicago, IL, 1956.

46. Müller, M. *Information Retrieval for Music and Motion*. Springer, 2007.

47. Partala, T., Surakka, V., and Vanhala, T. Real-time estimation of emotional experiences from facial expressions. *Interact. Comput. 18*, 2 (2006), 208–226.

48. Pennebaker, J.W., Francis, M.E., and Booth, R.J. Linguistic Inquiry and Word Count LIWC 2001. *Word Journal Of The International Linguistic Association*, (2001), 1–21.

49. Pollak, J.P., Adams, P., and Gay, G. PAM: a photographic affect meter for frequent, in situ measurement of affect. *Proceedings of the 2011 annual conference on Human factors in computing systems*, ACM (2011), 725–734.

50. Robazza, C., Macaluso, C., and D'Urso, V. Emotional Reactions to Music by Gender, Age, and Expertise. *Perceptual and Motor Skills 79*, 2 (1994), 939–944.

51. Rogers, Y., Sharp, H., and Preece, J. *Interaction Design: Beyond Human - Computer Interaction*. Wiley Publishing, 2011.

52. Russell, J.A. Core affect and the psychological construction of emotion. *Psychological Review 110*, 1 (2003), 145–172+.

53. Sahidullah, M. and Saha, G. Design, analysis and experimental evaluation of block based transformation in MFCC computation for speaker recognition. *Speech Communication 54*, 4 (2012), 543 – 565.

54. Scherer, K.. and Zentner, M.R. Emotional Effects Of Music: Production Rules. In *Music and Emotion:Theory and Research*. Oxford University Press, New York, NY, USA, 2001, 361–392.

55. Schuller, B., Hage, C., Schuller, D., and Rigoll, G. 'Mister D.J., Cheer Me Up!': Musical and Textual Features for Automatic Mood Classification. *Journal of New Music Research 39*, 1 (2010), 13–34.

56. Shaver, P., Schwartz, J., Kirson, D., and O'Connor, C. Emotion knowledge: Further exploration of a prototype approach. *Journal of Personality and Social Psychology 52*, 6 (1987), 1061–1086.

57. Sloboda, J. and Juslin, P.N. Psychological Perspective on Music and Emotion. In *Music and Emotion:Theory and Research*. Oxford University Press, New York, NY, USA, 2001, 71–104.

58. Sloboda, J. and O'Neill, S. Emotions in Everyday Listening to Music. In *Music and Emotion:Theory and Research*. Oxford University Press, New York, NY, USA, 2001, 415–430.

59. Sloboda, J.A., O'Neill, S.A., and Ivaldi, A. Functions of music in everyday life: An exploratory study using the Experience Sampling Method. *Musicae Scientiae 5*, 1 (2001), 9–32.

60. Stern, R.M., Ray,, W.J., and Quigley, K.S. *Psychophysiological recording*. Oxford University Press, New York, NY, USA, 2001.

61. Tomarken, A.J., Davidson, R.J., and Henriques, J.B. Resting frontal brain asymmetry predicts affective responses to films. *Journal of Personality and Social Psychology 59*, 4 (1990), 791–801.

62. Trochidis, K., Delbe, C., and Bigand, E. Investigation of the relationships between audio features and induced emotions in Contemporary Western music. .

63. Tzanetakis, G. and Cook, P. Musical genre classification of audio signals. *Speech and Audio Processing, IEEE Transactions on 10*, 5 (2002), 293 – 302.

64. Watson, D., Clark, L.A., and Tellegen, A. Development and Validation of Brief Measures of Positive and Negative Affect: The PANAS Scales. *Journal of Personality and Social Psychology 54*, 6 (1988), 1063 – 1070.

65. Winton, W.M., Putnam, L.E., and Krauss, R.M. Facial and autonomic manifestations of the dimensional structure of emotion. *Journal of Experimental Social Psychology 20*, 3 (1984), 195 – 216.

66. Yang, D. and Lee, W. Disambiguating Music Emotion Using Software Agents. *Proc. of Intl. Society for Music Information Retrieval*, Universitat Pompeu Fabra (October2004).

67. Yaramakala, S. and Margaritis, D. Speculative Markov blanket discovery for optimal feature selection. *Data Mining, Fifth IEEE International Conference on*, (2005), 4 pp.

68. Zentner, M.R., Meylan, S., and Scherer, K.. Exploring ‘musical emotions’ across five genres of music. (2000).

69. Zimmermann, P., Guttormsen, S., Danuser, B., and Gomez, P. Affective computing--a rationale for measuring mood with mouse and keyboard. *International Journal of Occupational Safety and Ergonomics: JOSE 9*, 4 (2003), 539–551.

70. Last.fm. http://www.last.fm/.

APPENDICES

APPENDIX A – ADDITIONAL INFORMATION SHEET

The Study

This study is concerned with understanding the relationship between a person's mood, the musical mood of piece of music and the listening context. During the study you will be asked to carry an Android phone on your person at all times. The session will run over the course of two weeks, during which time you will be asked periodically to fill out a questionnaire on your current mood and information about the music (if any) you are listening to. You have the option to opt out of any specific questionnaire: **Do not fill out a questionnaire in situations where it is unsafe to do so (for example, driving, operating machinery or while completing other tasks which require your attention).** The study pays a minimum of 5$ and a maximum of 40$ depending on the number of surveys filled out. To receive the maximum honorarium one must fill out 112 surveys (8 per day).

We are interested in music you are listening too from all sources and all activities. This included music you are listening to on your computer, stereo, iPod, or even playing in the background at a restaurant. While the Android phone will vibrate once an hour, you can also fill out a survey at any time, and we encourage you to do so.

Using The Phone

Feel free to use the android phone for more than just the survey. It includes a media player if you wish to use if for music, and can use nearby wireless Internet hotspots.

- Clicking the back button will close an open window. The back button looks like an arrow.

- Clicking the home button will close the application completely and take you to the "home" screen. The home button looks like a house.

The phone has an application on it called *Music Survey*. Once an hour, the phone will vibrate, and open this application that will prompt you to fill out a survey. If you wish to fill out a survey at any time, open this application. There is no limit to how often or how many surveys you can fill out.

- Clicking the menu button while in the *Music Survey* application will bring up a menu including these options:
    - Admin Controls: Do not worry about this, as this is for researchers only and password protected.
    - Sound: This allows you to turn on (or off) a sound notification. When turned on, the phone will both vibrate AND make a noise.
    - Definitions: This will bring up a list of important definitions on the phone.
    - Tutorial: This is a tutorial that explains how to fill out the survey. You can do the tutorial as many times as you wish.
- Clicking the menu button while filling out a survey will bring up the Definitions and Tutorial menu options.
- Sleep: If you are busy, and do not wish the phone to bother you, click the "sleep" button on the main page. Then select the number of hours you wish the phone to sleep and select "OK". The phone will not vibrate for the specified number of hours. If you open the application again before it vibrates it will cancel sleep mode.

Definitions

Below are some terms that are important you understand. These definitions are also on the device and can be accessed at any time.

Personal mood: This is the current mood (or emotion) you personally are *experiencing*. You will be asked to rate your mood on two scales, arousal and valence.

Musical mood: This is the mood or emotion a piece of music *expresses*. You also be asked to describe musical mood on two scales, arousal and valence.

Arousal: At the low arousal corresponds to feelings like relaxed, calm, sluggish, dull, sleepy, and unaroused. High arousal corresponds to feelings like stimulated, excited, frenzied, jittery, wide awake, and aroused.

Valence: Valence describes how positive or negative an emotion is. Negative valence corresponds to feelings like unhappy, annoyed, unsatisfied, melancholic, despairing, and bored. Positive valence corresponds feelings like happy, pleased, satisfied, contented, hopeful, and relaxed.

Questions?

If you have any question or problems during the study, please email diane.watson@usask.ca at any time during the study.

APPENDIX B – CONSENT FORMS

DEPARTMENT OF COMPUTER SCIENCE 

UNIVERSITY OF SASKATCHEWAN
INFORMED CONSENT FORM

Research Project: **Music, Mood and Listening Context**

Investigators: Dr. Regan Mandryk, Department of Computer Science (966-4888)
Diane Watson, Department of Computer Science (966-2327)

This consent form, a copy of which has been given to you, is only part of the process of informed consent. It should give you the basic idea of what the research is about and what your participation will involve. If you would like more detail about something mentioned here, or information not included here, please ask. Please take the time to read this form carefully and to understand any accompanying information.

This study is concerned with understanding the relationship between a person's mood, the mood of piece of music and the listening context. During the study you will be asked to carry an Android phone on your person at all times. The session will run over the course of two weeks, during which time you will be asked periodically to fill out a questionnaire on your current mood and information about the music (if any) you are listening to. You have the option to opt out of any specific questionnaire: **Do not fill out a questionnaire in situations where it is unsafe to**

do so (for example, driving, operating machinery or while completing other tasks which require your attention). The study pays a minimum of 5$ and a maximum of 40$ depending on the number of surveys filled out. To receive the maximum honorarium one must fill out 112 surveys (8 per day).

At the end of the session, you will be given more information about the purpose and goals of the study, and there will be time for you to ask questions about the research.

The data collected from this study will be used in articles for publication in journals and conference proceedings.

As one way of thanking you for your time, we will be pleased to make available to you a summary of the results of this study once they have been compiled (usually within two months). This summary will outline the research and discuss our findings and recommendations. If you would like to receive a copy of this summary, please write down your email address here.

Contact email address:_____

All personal and identifying data will be kept confidential. If explicit consent has been given, textual excerpts, photographs, or video recordings may be used in the dissemination of research results in scholarly journals or at scholarly conferences. Anonymity will be preserved by using pseudonyms in any presentation of textual data in journals or at conferences. The informed consent form and all research data will be kept in a secure location under confidentiality in accordance with University policy for 5 years post publication. Do you have any questions about this aspect of the study?

You are free to withdraw from the study at any time without penalty and without losing any advertised benefits. Withdrawal from the study will not affect your academic status or your access to services at the university. If you withdraw, your data will be deleted from the study and destroyed.

Your continued participation should be as informed as your initial consent, so you should feel free to ask for clarification or new information throughout your participation. If you have further questions concerning matters related to this research, please contact:

- Dr. Regan Mandryk, Assistant Professor, Dept. of Computer Science, (306) 966-4888, regan@cs.usask.ca

Your signature on this form indicates that you have understood to your satisfaction the information regarding participation in the research project and agree to participate as a participant. In no way does this waive your legal rights nor release the investigators, sponsors, or involved institutions from their legal and professional responsibilities. If you have further questions about this study or your rights as a participant, please contact:

- Dr. Regan Mandryk, Assistant Professor, Dept. of Computer Science, (306) 966-4888, regan@cs.usask.ca
- Office of Research Services, University of Saskatchewan, (306) 966-4053

Participant's signature:_____

Date:_____

Investigator's signature:_____

Date:_____

A copy of this consent form has been given to you to keep for your records and reference. This research has the ethical approval of the Office of Research Services at the University of Saskatchewan.

APPENDIX C – PRE-STUDY QUESTIONAIRRE

PID _____

Age _____

Sex

- ○ Male
- ○ Female

I listen to music

- ○ Several times a day
- ○ Once a day
- ○ A couple times a week
- ○ Once a week
- ○ A couple times a month
- ○ Once a month

I find myself experiencing the same emotion a piece of music is trying to express

- ○ Always
- ○ Often
- ○ Occasionally
- ○ Rarely
- ○ Never

I mainly listen to music as my (Choose the best answer)

- ○ Primary activity
- ○ Background activity

I listen to music on my (check all that apply)

- □ Stereo system
- □ Portable media player
- □ Cellphone

- ☐ Computer
- ☐ Radio
- ☐ TV radio station
- ☐ iPod/iPhone
- ☐ Other: _____

I listen to music primarily on my (choose the best answer)

- ○ Stereo system
- ○ Portable media player
- ○ Cellphone
- ○ Computer
- ○ Radio
- ○ TV radio station
- ○ iPod/iPhone
- ○ Other: _____

In general, I listen to music to (check all that apply)

- ☐ To express or release emotion
- ☐ To influence my emotion
- ☐ To relax
- ☐ For enjoyment
- ☐ As background sound
- ☐ Other: _____

I primarily listen to music to (choose the best answer)

- ○ To express or release emotion
- ○ To influence my emotion
- ○ To relax
- ○ For enjoyment
- ○ As background sound
- ○ Other: _____

In general I listen to music when I am (check all that apply)

- ☐ Waking up
- ☐ Bathing
- ☐ Exercising
- ☐ Working

- ☐ Doing homework
- ☐ Eating
- ☐ Relaxing
- ☐ Socializing
- ☐ Romantic activity
- ☐ Reading
- ☐ Going to sleep
- ☐ Driving
- ☐ Travelling as a passenger
- ☐ Shopping
- ☐ Dancing
- ☐ Getting drunk
- ☐ Other: _____

I primarily listen to music when I am (choose the best answer)

- ○ Waking up
- ○ Bathing
- ○ Exercising
- ○ Working
- ○ Doing homework
- ○ Eating
- ○ Relaxing
- ○ Socializing
- ○ Romantic activity
- ○ Reading
- ○ Going to sleep
- ○ Driving
- ○ Travelling as a passenger
- ○ Shopping
- ○ Dancing
- ○ Getting drunk
- ○ Other: _____

APPENDIX D – ESM SURVEY QUESTIONS

What is your personal mood right now?

Low Arousal ○ ○ ○ ○ ○ High Arousal

Are you feeling more positive or negative?

Negative ○ ○ ○ ○ ○ Positive

I am

- ○ My myself
- ○ With people I know
- ○ With people I do not know

My current activity is best described as

- ○ Waking up
- ○ Bathing
- ○ Exercising
- ○ Working
- ○ Doing homework
- ○ Eating
- ○ Relaxing
- ○ Socializing
- ○ Romantic activity
- ○ Reading
- ○ Going to sleep
- ○ Driving
- ○ Travelling as a passenger (i.e. bus)
- ○ Shopping
- ○ Dancing
- ○ Getting drunk
- ○ Other

The location I am in is best described as

- ○ Home

- ○ Work
- ○ Public place
- ○ Other

Song Information:

Artist:_____

Title:_____

Genre:_____

What musical mood do you think the music is trying to express?

Low Arousal ○ ○ ○ ○ ○ High Arousal

Is it positive or negative?

Negative ○ ○ ○ ○ ○ Positive

I chose this song

- ○ Yes
- ○ Yes as part of a playlist
- ○ No

I am listening to music as my

- ○ Primary activity
- ○ Background activity

The primary reason I am listening is

- ○ To express or release emotion
- ○ To influence my emotion
- ○ To relax
- ○ For enjoyment
- ○ As background sound
- ○ Other

Some words phrases or images I associate with this song are:

APPENDIX E – LIWC DICTIONARY

Category	Abbrev	Examples	Words in category
Linguistic Processes			
Word count	wc		
words/sentence	wps		
Dictionary words	dic		
Words>6 letters	sixltr		
Total function words	funct		464
Total pronouns	pronoun	I, them, itself	116
Personal pronouns	ppron	I, them, her	70
1st pers singular	i	I, me, mine	12
1st pers plural	we	We, us, our	12
2nd person	you	You, your, thou	20
3rd pers singular	shehe	She, her, him	17
3rd pers plural	they	They, their, they'd	10
Impersonal pronouns	ipron	It, it's, those	46
Articles	article	A, an, the	3
Common verbs	verb	Walk, went, see	383
Auxiliary verbs	auxverb	Am, will, have	144
Past tense	past	Went, ran, had	145
Present tense	present	Is, does, hear	169
Future tense	future	Will, gonna	48
Adverbs	adverb	Very, really, quickly	69
Prepositions	prep	To, with, above	60
Conjunctions	conj	And, but, whereas	28
Negations	negate	No, not, never	57
Quantifiers	quant	Few, many, much	89
Numbers	number	Second, thousand	34
Swear words	swear	Damn, piss, fuck	53
Psychological Processes			
Social processes	social	Mate, talk, they, child	455
Family	family	Daughter, husband, aunt	64
Friends	friend	Buddy, friend, neighbor	37
Humans	human	Adult, baby, boy	61
Affective processes	affect	Happy, cried, abandon	915
Positive emotion	posemo	Love, nice, sweet	406
Negative emotion	negemo	Hurt, ugly, nasty	499
Anxiety	anx	Worried, fearful, nervous	91

Category	Abbrev	Examples	Words in category
Anger	anger	Hate, kill, annoyed	184
Sadness	sad	Crying, grief, sad	101
Cognitive processes	cogmech	cause, know, ought	730
Insight	insight	think, know, consider	195
Causation	cause	because, effect, hence	108
Discrepancy	discrep	should, would, could	76
Tentative	tentat	maybe, perhaps, guess	155
Certainty	certain	always, never	83
Inhibition	inhib	block, constrain, stop	111
Inclusive	incl	And, with, include	18
Exclusive	excl	But, without, exclude	17
Perceptual processes	percept	Observing, heard, feeling	273
See	see	View, saw, seen	72
Hear	hear	Listen, hearing	51
Feel	feel	Feels, touch	75
Biological processes	bio	Eat, blood, pain	567
Body	body	Cheek, hands, spit	180
Health	health	Clinic, flu, pill	236
Sexual	sexual	Horny, love, incest	96
Ingestion	ingest	Dish, eat, pizza	111
Relativity	relativ	Area, bend, exit, stop	638
Motion	motion	Arrive, car, go	168
Space	space	Down, in, thin	220
Time	time	End, until, season	239
Personal Concerns			
Work	work	Job, majors, xerox	327
Achievement	achieve	Earn, hero, win	186
Leisure	leisure	Cook, chat, movie	229
Home	home	Apartment, kitchen, family	93
Money	money	Audit, cash, owe	173
Religion	relig	Altar, church, mosque	159
Death	death	Bury, coffin, kill	62
Spoken categories			
Assent	assent	Agree, OK, yes	30
Nonfluencies	nonflu	Er, hm, umm	8
Fillers	filler	Blah, Imean, youknow	9

APPENDIX F – MODELING MUSICALLY INDUCED AFFECT

Just as music can express an emotion, it can also induce an affective state. Emotional experiences to music as so well documented that music is often used an emotional stimuli in laboratory studies (e.g., [32]). These studies make the assumption that people experience happiness when listening to happy music and sadness when listening to sad music, and while this may often be the case [28], musical mood and affective state are two separate, but related concepts [54].

To investigate our data set further, we modeled the affect induced by music by using the methods previously discussed in the main body of this thesis. This chapter will discuss the literature related to modeling affective state induced by music as well as the results of our models. For the methods employed see Chapters 4 and 6. A description of the data set is seen in Chapter 3.

15.1 RELATED WORK

Very little literature exists about predicting induced emotion, or affect, from the music being listened to[3]. Preliminary work by Trochidis et al. [62] suggests that many of the same musical features that are used to classify expressed emotion (i.e., musical mood) are important to classifying induced emotion as well. However, according to Lavy [37], the context of the listening experience is as important to how the listener reacts to the music as the music itself.

[3] This gap in the literature exists not necessarily because no one has looked at the issue, but because those that have fail to indicate specifically that they are looking at induced rather than expressed emotion. This ambiguity is unfortunate as these are two clearly different concepts [68][54], although they may be closely intertwined. Some researchers simply assume previous work discusses the concept most favorable to their own research (e.g., [41] is discussed correctly as expressed emotion in [33] and incorrectly as induced emotion in [62]).

15.2 AFFECT AND MUSIC

We collected affect both when the participants were listening to music and when they were not (See Figure 18). When participants were not listening to music, their arousal and valence centered around neutral (neutral arousal and neutral valence). When they were listening to music, their arousal and valence were shifted up and to the right (higher arousal and higher valence). This shift suggests that they were experiencing enjoyment when listening to music. Over all instances, both arousal (t_{1772}=8.93 $p<0.001$) and valence (t_{1772}=8.05 $p<0.001$) were significantly different when listening to music than when not listening to music. This suggests that music does have an effect on affective state.

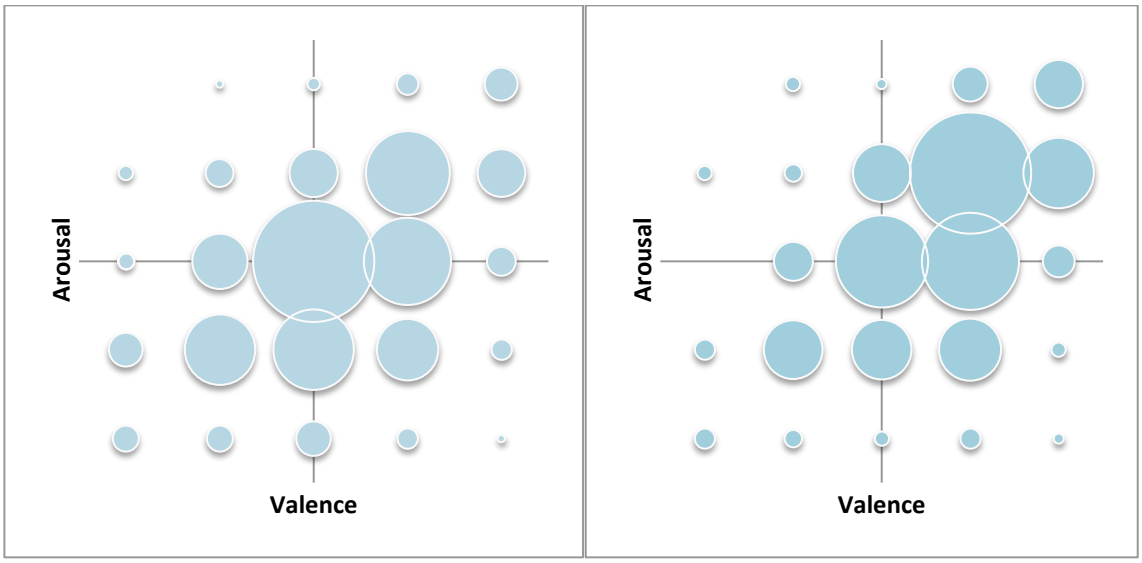

Figure 18 shows affective state without music (on the left) and affective state with music (on the right).

15.3 MODEL

We created Bayes Net classifiers for induced affect.

15.3.1 Feature sets

We used different combinations of features that can be summarized as two feature sets (See Chapter 3 and 5 for a description of the features).

Musical Features: Our first feature set used audio features, lyrical features, and tag features, as these features were used in previous models of musical mood. There were 198 different features in this set.

Musical Features + Musical Mood: Our second feature set uses musical features combined with musical mood. There were 200 different features in this set.

Musical Features + Listening Context + Musical Mood. Our third feature set combined musical features with the listening context collected in our study for a total of 296 features.

15.4 DATA

15.4.1 Sparse Data

Because users could optionally specify an artist and title, some features, (namely audio features, lyrics and tags) were sparse. To improve the percentage of instances covered by these features, we restricted our data set to only those instances in which a title was provided. See Section 6.2.1 for more information. See Figure 20 and 21 for the distribution of responses after the title only restriction.

15.4.2 Class Skew

To account for class skew, results were binned into low and high; neutral instances were ignored. Also, undersampling, a technique that selects a random selection of instances such that a uniform distribution is achieved, was applied to the data. This process was completed five times and the

average of all runs is reported. See Figure 19 for the original distribution of responses and Figure 20 and 21 for the distribution of responses after binning and undersampling.

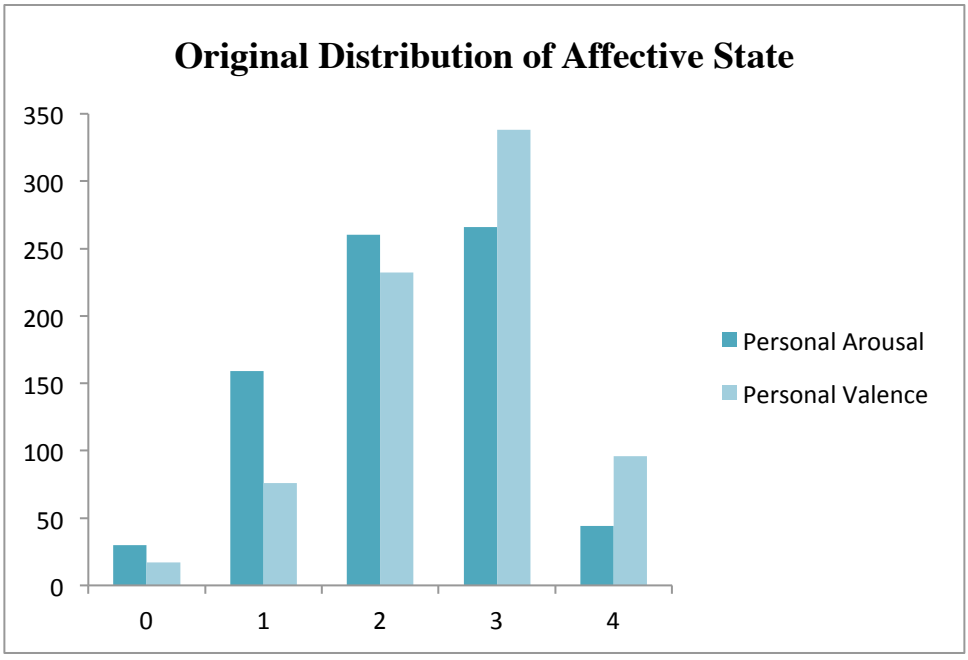

Figure 19 Original Distribution of Affective state

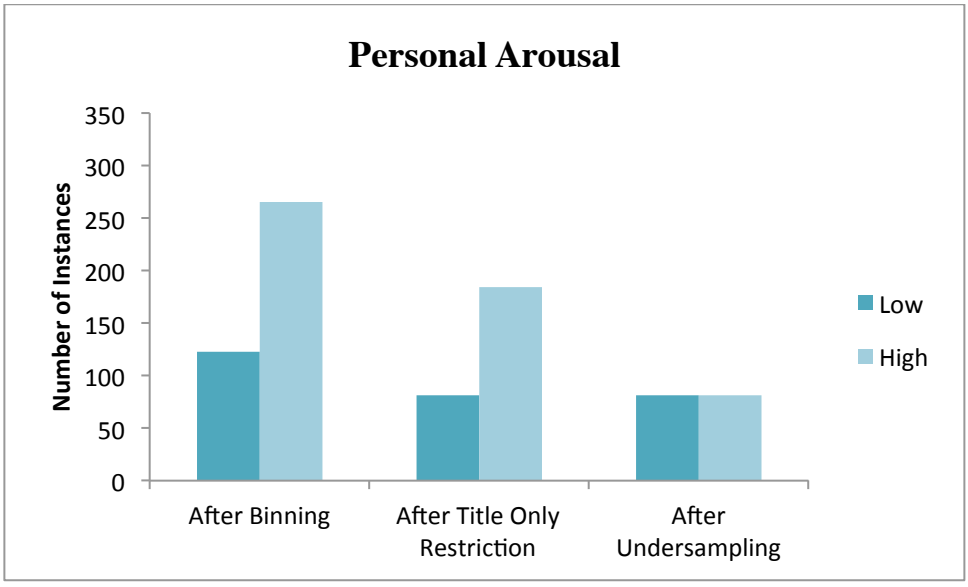

Figure 20 Personal arousal after binning, restriction and undersampling

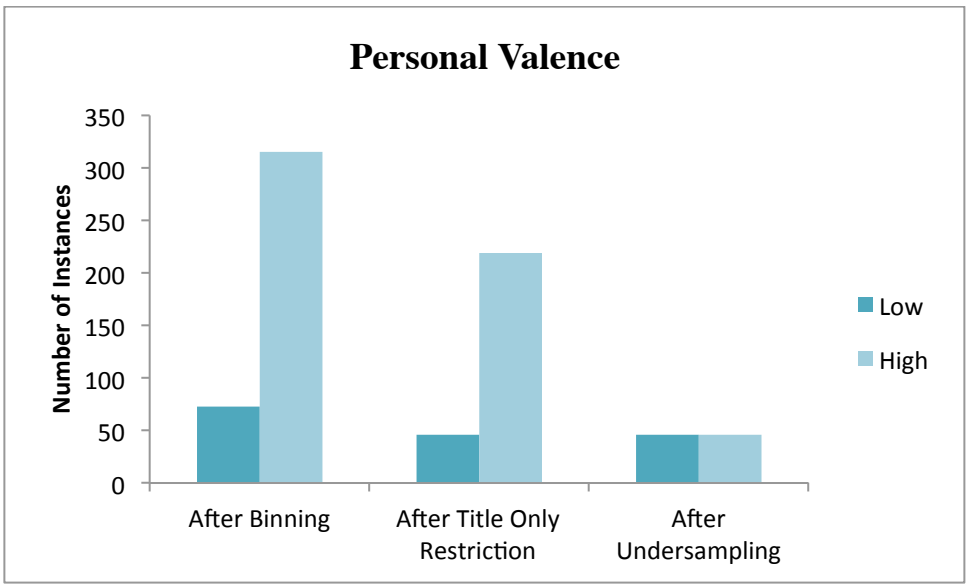

Figure 21 Personal valence after binning, restriction and undersampling

15.5 RESULTS

We present the results of our models in the next section. See Figure 22 for a graph of classification accuracies for each model.

Musical Features: Using only musical features (audio features, lyrics and tags), personal arousal has an average classification accuracy (over the five runs of undersampled data, see section 6.2.2) of 48.2% (SD=0.78, kappa[4]=-0.0351) Personal valence has a classification accuracy of 48.9% (SD=3.27, kappa=-0.0232). Neither personal arousal nor personal valence was classified higher than a random classifier (chance=50%).

Musical Features + Musical Mood: When we combined musical features + musical arousal and valence, personal arousal has an average classification accuracy of 53.6% (SD=2.64, kappa=0.0731) Personal valence has a classification accuracy of 48.0% (SD=5.87, kappa=-

.0349). Personal arousal (t_4=3.06 $p<0.05$) was significantly higher than chance (50%) and also we achieved a significant improvement in personal arousal (t_8=4.35 $p<0.001$) over using just musical features alone.

Musical Features + Listening Context: When we combined musical features with listening context, personal arousal has a classification accuracy of 64% (SD=5.16, kappa=0.2816). Personal valence has a classification accuracy of 52% (SD=5.83, kappa=0.0562). Using one-sample t-tests we show that personal arousal (t_4=6.09 $p<0.01$) performed significantly better than chance (50%); personal valence (t_4=1.08, $p>0.3$) did not. We achieved improved classification accuracies of personal arousal and personal valence by using a combination of listening context and musical features; however only musical arousal was a significant improvement (personal arousal t_8=4.03, $p<0.01$, personal valence t_8=1.29, $p>0.2$).

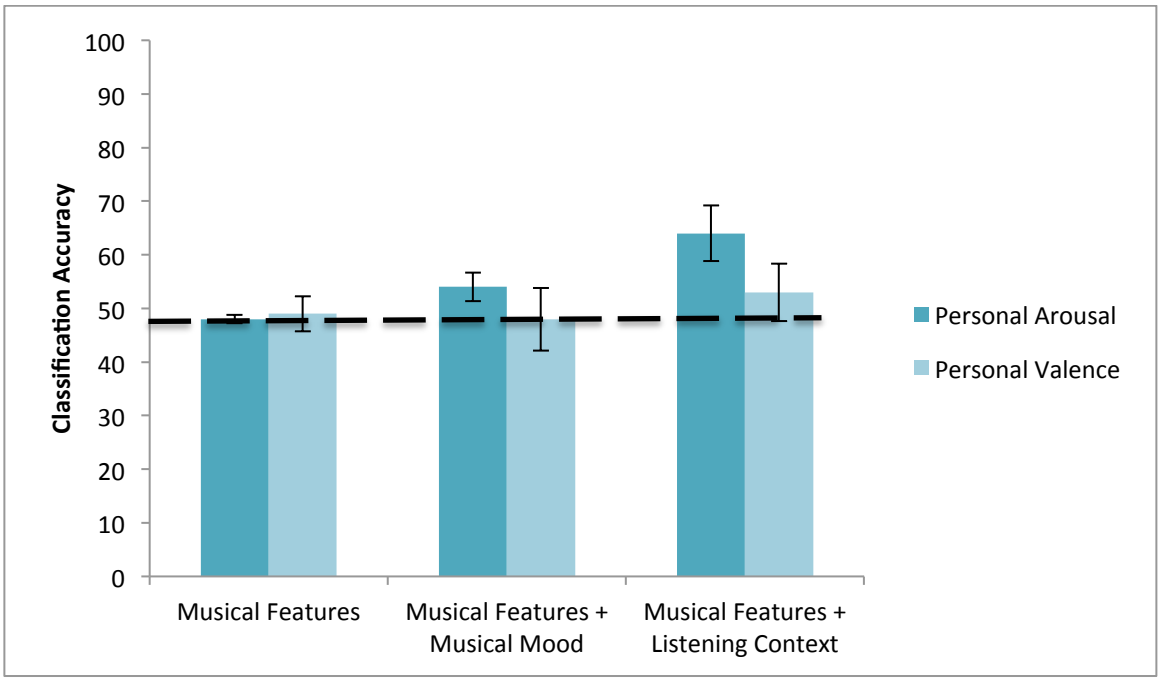

Figure 22 Classification accuracies of affect. The dotted line is chance.

15.6 DISCUSSION

We predict induced affect correctly with a maximum classification accuracy of 64% for personal arousal and 52% for personal valence when using musical features and listening context. We also predict personal arousal significantly more often than chance when using musical features and musical mood with a classification accuracy of 53%.

Although studies using music as emotionally charged stimuli assume that people experience the same emotion as the music they are listening to, our low classification accuracies when using musical mood have shown that this may not always be the case in naturalistic settings. The context appears to be more important in predicting induced affect, at least in terms of predicting musical arousal. Valence may be more difficult to predict. This does not mean that using music to induce emotion in laboratory settings is flawed; it only means that it may not work as well in real life contexts.

It is very difficult to draw many conclusions from this work, in part due to the low classification accuracies. Even for the higher classification accuracies, the low kappa values suggest that the performance of the models is not reliable. Music is known to induce emotion, however users may also be making song choices based on the emotion they experience at the time. Therefore, this relationship may work in both directions. To examine the relationships between listening context, musical mood, and affective state, we could provide users with representative samples in a music library. By listening to (and rating) the same song in a variety of contexts and affective states, the relationship between these three factors might be made clear.