

# **A Mood-Based Music Classification and Exploration System**

by

Owen Craigie Meyers

B.Mus., McGill University (2004)

Submitted to the Program in Media Arts and Sciences,  
School of Architecture and Planning,  
in partial fulfillment of the requirements for the degree of

Master of Science in Media Arts and Sciences

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2007

© Massachusetts Institute of Technology 2007. All rights reserved.

Author\_\_\_\_\_

Program in Media Arts and Sciences  
May 11, 2007

Certified by\_\_\_\_\_

Barry Vercoe  
Professor of Media Arts and Sciences  
Program in Media Arts and Sciences  
Thesis Supervisor

Accepted by\_\_\_\_\_

Andrew B. Lippman  
Chair, Departmental Committee on Graduate Students  
Program in Media Arts and Sciences



# **A Mood-Based Music Classification and Exploration System**

by

Owen Craigie Meyers

Submitted to the Program in Media Arts and Sciences,  
School of Architecture and Planning,  
on May 11, 2007, in partial fulfillment of the  
requirements for the degree of  
Master of Science in Media Arts and Sciences

## **Abstract**

Mood classification of music is an emerging domain of music information retrieval. In the approach presented here features extracted from an audio file are used in combination with the affective value of song lyrics to map a song onto a psychologically based emotion space. The motivation behind this system is the lack of intuitive and contextually aware playlist generation tools available to music listeners. The need for such tools is made obvious by the fact that digital music libraries are constantly expanding, thus making it increasingly difficult to recall a particular song in the library or to create a playlist for a specific event. By combining audio content information with context-aware data, such as song lyrics, this system allows the listener to automatically generate a playlist to suit their current activity or mood.

Thesis Supervisor: Barry Vercoe

Title: Professor of Media Arts and Sciences, Program in Media Arts and Sciences



**A Mood-Based Music Classification  
and Exploration System**

by

Owen Craigie Meyers

The following people served as readers for this thesis:

Thesis Reader\_\_\_\_\_

Henry Lieberman  
Research Scientist  
MIT Media Laboratory

Thesis Reader\_\_\_\_\_

Emery Schubert  
Australian Research Fellow & Lecturer  
University of New South Wales



## Acknowledgments

I would like to thank the following people for their support and contributions with respect to this thesis:

First and foremost, I give thanks to my advisor, Barry Vercoe, for his direction, inspiration and encouragement.

I thank my thesis readers, Henry Lieberman and Emery Schubert, for the generous use of their time.

Brian Whitman, Tristan Jehan, and Scotty Vercoe have been especially helpful in providing a strong foundation of research in the areas of music recommendation, audio feature extraction, and mood classification without which the success of my work would not have been possible.

Many thanks to my fellow Music, Mind and Machine group members: Mihir Sarkar, Wu-Hsi Li, and Cheng-Zhi Anna Huang for their constructive criticism and comments.

Lastly, and most importantly, I thank my family for their constant support throughout this entire experience.





# Contents

<b>Abstract</b>	<b>3</b>
<b>1 Introduction</b>	<b>15</b>
1.1 Motivation . . . . .	15
1.1.1 Context-Aware Playlist Generation . . . . .	16
1.1.2 Retrieving Lost Music . . . . .	16
1.1.3 Music Classification . . . . .	16
1.1.4 Music Recommendation . . . . .	17
1.2 Contributions & Approach . . . . .	18
1.3 Thesis Structure . . . . .	19
<b>2 Background</b>	<b>21</b>
2.1 Psychology of Music . . . . .	21
2.1.1 Definition of Emotion . . . . .	22
2.1.2 Emotional Models . . . . .	23
2.1.3 Music and Emotion . . . . .	24
2.2 Music Theory . . . . .	27
2.3 Feature Extraction . . . . .	30
2.3.1 CLAM . . . . .	31
2.3.2 Creating Music by Listening [23] . . . . .	32
2.4 Emotion Detection in Music . . . . .	33
2.5 Music Classification Frameworks . . . . .	37
2.6 Playlist Generation Tools . . . . .	39
2.7 Music Recommendation Systems . . . . .	40
2.8 Natural Language Processing . . . . .	43
2.8.1 Commonsense Reasoning . . . . .	43
2.8.2 Lyrics . . . . .	46
<b>3 Design &amp; Implementation</b>	<b>47</b>
3.1 Emotional Model . . . . .	47
3.2 Audio Analysis . . . . .	51
3.2.1 Mode and Harmony . . . . .	52
3.2.2 Tempo, Rhythm, and Loudness . . . . .	53
3.3 Lyric Analysis . . . . .	55
3.4 Classification of Music . . . . .	57

3.4.1	Classification of Audio Features . . . . .	57
3.4.2	Classification of Lyrics . . . . .	60
3.5	Mood Player Interface . . . . .	60
3.5.1	Features . . . . .	60
3.5.2	Frameworks, Software, & Technical Implementation . . . . .	63
<b>4</b>	<b>Evaluation</b>	<b>67</b>
4.1	System Performance and Accuracy . . . . .	67
4.1.1	Classification of Music Database . . . . .	68
4.1.2	Classification of Lyrics . . . . .	70
4.2	Vs. Music Classification Experts . . . . .	71
4.3	Vs. Social Tagging Services . . . . .	74
4.4	User Evaluation . . . . .	77
<b>5</b>	<b>Conclusion</b>	<b>81</b>
5.1	Future Work . . . . .	83
5.1.1	Improvements . . . . .	83
5.1.2	Applications . . . . .	85
<b>A</b>	<b>User Evaluation</b>	<b>87</b>
	<b>Bibliography</b>	<b>89</b>

# List of Figures

2-1	Thayer’s two-dimensional model of emotion . . . . .	24
2-2	Multidimensional scaling of Russell’s circumplex model of emotion [44], p. 1168 . . . . .	25
2-3	Hevner’s adjective circle [21], p. 249 . . . . .	26
2-4	Schubert’s two-dimensional emotion space (2DES) [46], p. 564 . . . . .	28
2-5	iLike Page for Radiohead’s Kid A Album . . . . .	42
2-6	MOG Page for Radiohead’s Kid A Album . . . . .	44
3-1	Color mapping of Russell’s circumplex model of emotion used in this system	50
3-2	Simplified Decision Tree Schema . . . . .	59
3-3	Mood Player Environment . . . . .	61
4-1	Results for the five features: Mode, Harmony, Tempo, Rhythm, Loudness .	69
4-2	Popular Tags from Last.FM for Radiohead’s Idiotique . . . . .	76
4-3	Popular Tags from Qloud for Radiohead’s Idiotique . . . . .	76



# List of Tables

2.1	Hevner’s weighting of musical characteristics in 8 affective states [22], p. 626	27
3.1	Comparison of three emotional models, in terms of valence and arousal . . .	49
3.2	Mapping of musical features to Russell’s circumplex model of emotion . . .	51
4.1	Eight songs classified by the system . . . . .	68
4.2	Results from the lyric analysis of eight songs . . . . .	71
4.3	All Music Guide’s classification of Radiohead’s album Kid A . . . . .	73
4.4	Audio and lyric mood classification of songs from Radiohead’s Kid A album and their respective features from Pandora . . . . .	73
4.5	Our mood classification of eight songs versus AMG and Pandora . . . . .	75
4.6	Comparison of musical qualities of eight songs . . . . .	75
4.7	Tagging of Radiohead’s Idioteque . . . . .	77
4.8	Social tagging of eight artists . . . . .	78



# Chapter 1

## Introduction

Today's music listener faces a myriad of obstacles when trying to find suitable music for a specific context. With digital music libraries expanding at an exponential rate, the need for innovative and novel music classification and retrieval tools is becoming more and more apparent. Music listeners require new ways to access their music, such as alternative classification taxonomies and playlist generation tools. The work presented here is a music classification system that provides the listener with the opportunity to browse their music by mood, and consequently allows the listener to generate context-aware playlists, find lost music, and experience their music in an exciting new way.

### 1.1 Motivation

The applications and implications of this work are wide and varied. Mood classification can be applied to such scenarios as a DJ choosing music to control the emotional level of people on the dance floor, to a composer scoring a film, to someone preparing the soundtrack for their daily workout. Each of these situations relies heavily on the emotional content of the music.

### 1.1.1 Context-Aware Playlist Generation

The system presented here is intended to aid in the creation of playlists that suit a particular mood or context. With the advent of portable music players and the growth of digital music libraries, music listeners are finding it increasingly difficult to navigate their music in an intuitive manner. The next generation of playlist creation tools need to address this issue in order to provide listeners with the proper means with which to experience their music.

### 1.1.2 Retrieving Lost Music

With digital music libraries reaching sizes on the order of 10's of thousands of songs, many songs are lost in the masses. As a result, many people have large amounts of music that they never listen to. This phenomenon has been labeled The Long Tail [3], which describes the statistical distribution of a high-frequency population that immediately trails off to a low-frequency population. In this case, the high-frequency population refers to a small subset of songs that are played most often, while the remainder is rarely, or never played. Thus, it has been the mandate of music recommendation systems to explore, recommend, and retrieve lost music from this long tail.

### 1.1.3 Music Classification

The current state of music classification leaves something to be desired. Music is generally classified according to genre, such as Alternative, Christian/Gospel, Country/Folk, Dance/Electronic, Eclectic, Indie, Jazz/Blues, Pop, Rock, or World<sup>1</sup>. This type of classification is extremely limiting in that it attempts to enforce a broad organization that encompasses artist, album, and song. When the more general genre classification fails to categorize an artist, sub-genres are introduced. Though sub-genres are more specific and often describe more precisely the music of an artist, the list of sub-genres is constantly expanding and evolving, and as such, it lacks any sense of coherence or regularity. Another

---

<sup>1</sup><http://nest.echonest.com/post/950784>



issue is that an artist's music may evolve from album to album, or even from song to song. Moreover, a broad genre classification for a particular artist can be misleading. Ultimately, genre classification is a tool used for commercialization and marketing of music, and as a result is an unsuitable taxonomy for describing musical content<sup>2</sup>.

The music experts at All Music Guide (AMG) have devised several alternative classifications such as style and theme. These taxonomies are helpful to listeners choosing music for a particular event, such as exercising or a road trip, but again are not standardized and are accessible only from the AMG website.

A popular, yet relatively uncultivated classification method is that of mood. The experts at AMG have created a large mood taxonomy and digital media players such as MoodLogic and Sony's StreamMan also implement elements of mood classification. Apart from these few implementations, however, mood-based music systems have not been developed to their full potential despite the popularity of mood and emotion as a means of describing a song or musical context [7]. Thus, the goal of this thesis is to address the issue of music mood classification and its role in the listener's musical experience.

#### **1.1.4 Music Recommendation**

It is imperative to provide music listeners with the proper tools with which to access their music. Many of the current recommendation systems lack a thorough understanding of the content and context of a song. The Echo Nest, whose work will be built upon in the research presented here, is currently developing a system that takes both the content and context of music into account. Tools such as this are part of the next generation of the music recommendation movement, which focuses on providing the user with a more enjoyable listening experience.

---

<sup>2</sup>[http://en.wikipedia.org/wiki/Music\\_genre#Subjectivity](http://en.wikipedia.org/wiki/Music_genre#Subjectivity)

## 1.2 Contributions & Approach

Music psychology forms the basis of the work presented here. In order to classify music by mood and emotion one must first have a clear understanding of how both music and emotions are represented in the human mind. There have been many psychological studies performed throughout the past century relating to music and emotion, and as a result there exist many different representations and interpretations of human emotion and its relation to music. The objective of this thesis research is to utilize the best possible psychological model of emotion and to incorporate the findings of these studies into an innovative front-end for a digital music library, where music can be queried, browsed, and explored by mood, rather than by artist, album, or genre.

Both categorical and dimensional models of emotion have been applied to mood classification of music. The work of several prominent psychologists and music psychologists will be reviewed and evaluated in this thesis. It has been found that the dimensional approach to emotional modeling is best suited and most commonly used in musical applications [17], and therefore this model of emotion will be implemented in the mood classification system outlined in this document.

With the emotional model in place, the next step is to determine how this model relates to individual musical features. Many studies have been conducted with respect to the emotional effects of music and musical parameters [17]. This information is assessed in the context of our classification system and is put into practice through the use of state of the art audio and textual analysis tools, which extract relevant features from a piece of music. The musical features of a song are then mapped onto the emotional space using data from the psychological studies. The emotional response of the lyrics, obtained through natural language processing and commonsense reasoning, contributes to both the context and mood classification of the song. Several machine learning algorithms are applied to the song using its features as input data. The result is a mood classification that is indicative of the song's musical characteristics and lyrical content.

It is the hypothesis of this work that the combination of the three elements of emotional

modeling, audio feature extraction, and lyrical analysis will result in a novel and intuitive tool that music listeners may use in their daily activities for the generation of playlists and the overall enjoyment of their music. The goal of this work is to enable music listeners to easily navigate their digital music collection through the use of mood classification.

## 1.3 Thesis Structure

The background portion of this thesis is presented in Chapter 2. In this section the psychology of music is explored with respect to models of human emotion and the relationship between music and emotion. In addition, the music theory pertaining to relevant music features used in this system is defined. Audio feature extraction frameworks related to this thesis are surveyed and several music classification and recommendation services are reviewed to provide a basis from which to evaluate our mood classification system. Lastly, several textual analysis tools, including natural language processing and commonsense reasoning, are presented.

The design and implementation of the mood classification system described here is outlined in Chapter 3. This chapter details the emotional model used in this system as well as how individual music features can be mapped onto this model. An explanation follows regarding both the process of extracting the musical features from the audio signal and computing the affective value of song lyrics. The resulting audio and textual features are then used in the classification engine of the thesis work. The chapter concludes with an overview of the user interface and its features.

In Chapter 4 we evaluate the performance and accuracy of the system. The results from the mood classification system are tested against popular social and expert music tagging and classification services, including All Music Guide, Pandora, and Last.FM. Lastly, the results of a user evaluation are presented.

In the final chapter the work presented here is concluded and insight is provided into future directions and developments.



## Chapter 2

# Background

In this section various techniques and frameworks will be reviewed with respect to their relevance to this project. An accurate understanding of how emotions are represented both in the human mind and in the computer is essential in the design of a mood classification system. The relationship between emotion and music is important when mapping various musical parameters to an emotion space. Music theory is required to make informed decisions and observations regarding the extraction of salient music features and the classification of such features. By extracting and analyzing the appropriate audio features and textual metadata, the classification system will perform more efficiently and precisely.

### 2.1 Psychology of Music

The field of music psychology dates back to the 18th century, beginning with J.P. Rameau in 1722<sup>1</sup>. Since this time, scholars, researchers, and scientists have been studying how the human mind interprets and experiences music. The psychologically based fields of music perception and cognition explore how scientific representations of audio signals in the environment differ from representations within the mind. This includes the representation of

---

<sup>1</sup>Huron D., History of Music Psychology. Taken from <http://music-cog.ohio-state.edu/Music829F/timeline.html> on 02/01/2007.

the frequency of a sound, which is perceived by humans as pitch, and the sound’s intensity, which corresponds to loudness. Similar to how music perception seeks to find a model of music, psychologists have attempted to find a model of human emotion in the mind. Emotions can manifest themselves in a variety of different ways, and psychologists have found several approaches that model emotion intuitively, including the dimensional model, which classifies emotions along several axes such as pleasure, arousal, and dominance, and the categorical model, which consists of several distinct classes that form the basis for all other possible emotional variations. After clearly defining an emotional model, it is then possible to attribute specific musical features, such as pitch and loudness, to the different affective states of this model.

### **2.1.1 Definition of Emotion**

Emotion is a complex set of interactions among subjective and objective factors, mediated by neural/hormonal systems, which can (a) give rise to affective experiences such as feelings of arousal, pleasure/displeasure; (b) generate cognitive processes such as perceptually relevant effects, appraisals, labeling processes; (c) activate widespread physiological adjustments to the arousing conditions; and (d) lead to behavior that is often, but not always, expressive, goal-oriented, and adaptive [25].

Humans are capable of experiencing a vast array of emotional states. As such, there exist many terms and definitions of emotion as it relates to everyday life. Affect, mood, emotion, and arousal are often used interchangeably though each is unique and differentiable from each other. Emotional states can be broken down into various categories based on how they manifest and exhibit themselves in the individual. They range from the basic primitive affective instincts to more specific moods, which tend to have a prolonged effect and are indirectly influenced by the individual’s surroundings. An affective state, which is the broadest of emotional states, may have some degree of positive or negative valence, which is the measure of the state’s emotional charge. Moods are slightly narrower but provide the basis for more specific emotional states that are typically much shorter in duration and can generally be attributed to a particular stimulus. An emotional state is largely influenced

by the underlying mood and affective state of the individual. Thus, an emotional state is often the result of many interrelated and underlying influences, which ultimately manifest themselves visually, through facial expressions, or audibly, through vocalizations and vocal expressions. Lastly, arousal relates to the intensity of an emotional state, similar to how affect and valence trigger positive or negative states in the individual. A highly aroused emotional state will be very apparent in the individual [49].

The subjectivity of emotions creates ambiguity in terminology, and thus emotion remains a vague and relatively undefined area of the human experience. For the most part, everyone understands what an emotion is, and can differentiate amongst them, but when asked to define what an emotion is they hesitate [15].

### **2.1.2 Emotional Models**

Models of human emotion are diverse and varied owing to the subjective nature of emotion. The two major approaches to emotional modeling that exist in the field today are categorical and dimensional. Each type of model helps to convey a unique aspect of human emotion and together such models can provide insight into how emotions are represented and interpreted within the human mind.

The most common of the categorical approaches to emotion modeling is that of Paul Ekman's basic emotions, which encompasses the emotions of anger, fear, sadness, happiness, and disgust [11]. A categorical approach is one that consists of several distinct classes that form the basis for all other possible emotional variations. Categorical approaches are most applicable to goal-oriented situations.

A dimensional approach classifies emotions along several axes, such as valence (pleasure), arousal (activity), and potency (dominance). Such approaches include James Russell's two-dimensional bipolar space (valence-arousal) [43], Robert Thayer's energy-stress model [51, 52], where contentment is defined as low energy/low stress, depression as low energy/high stress, exuberance as high energy/low stress, and anxious/frantic as high energy,

high stress (see Figure 2-1), and Albert Mehrabian’s three-dimensional PAD representation (pleasure-arousal-dominance) [35]. The basic two and three-dimensional models have also been expanded to circular models, such as Russell’s circumplex model (see Figure 2-2) [44] and Kate Hevner’s adjective circle (see Figure 2-3) [21]. In these approaches, a list of adjectives, 28 terms for Russell, and 67 terms for Hevner, are mapped to their respective quadrant or octant.

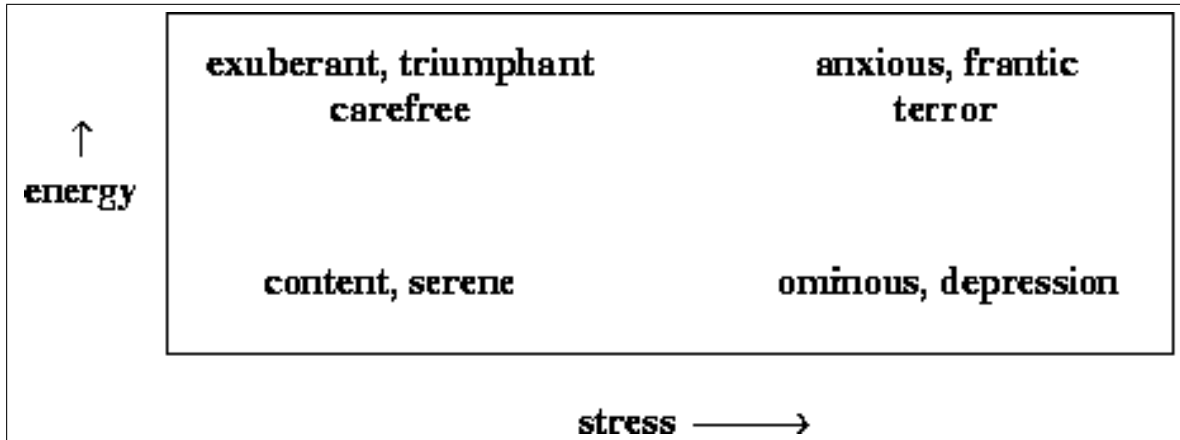


Figure 2-1: Thayer’s two-dimensional model of emotion

### 2.1.3 Music and Emotion

In the mid 20th century, scholars and researchers such as Hevner, Melvin Rigg, and Karl Watson began to make progress in relating specific musical features, such as mode, harmony, tempo, rhythm, and dynamics (loudness), to emotions and moods.

Hevner’s studies [19, 20, 21, 22] focus on the affective value of six musical features and how they relate to emotion. The results of these studies are summarized in Table 2.1. The six musical elements explored in these studies include mode, tempo, pitch (register), rhythm, harmony, and melody. These features are mapped to a circular model of affect encompassing eight different emotional categories (Figure 2-3). The characteristic emotions of each of the eight categories are dignified, sad, dreamy, serene, graceful, happy, exciting, and vigorous. Each category contains anywhere from six to eleven similar emotions for a total of 67 adjectives. This model is closely related to that of Russell [44], and thus provides



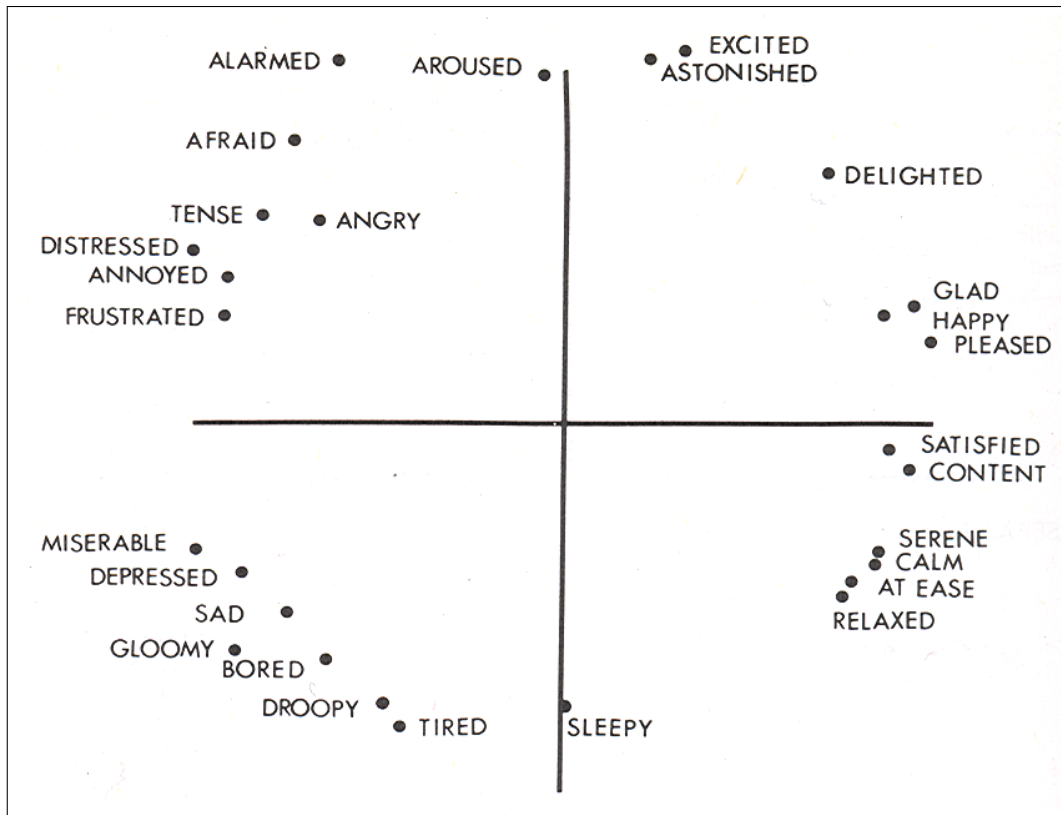


Figure 2-2: Multidimensional scaling of Russell's circumplex model of emotion [44], p. 1168

further validity for the circumplex model of emotion.

Rigg's experiment includes four categories of emotion; lamentation, joy, longing, and love. Categories are assigned several musical features, for example 'joy' is described as having iambic rhythm (staccato notes), fast tempo, high register, major mode, simple harmony, and loud dynamics (forte) [41, 42].

Watson's studies differ from those of Hevner and Rigg because he uses fifteen adjective groups in conjunction with the musical attributes pitch (low-high), volume (soft-loud), tempo (slow-fast), sound (pretty-ugly), dynamics (constant-varying), and rhythm (regular-irregular). Watson's research reveals many important relationships between these musical attributes and the perceived emotion of the musical excerpt [58]. As such, Watson's contribution has provided music emotion researchers with a large body of relevant data that they can now use to gauge the results of their experiments.

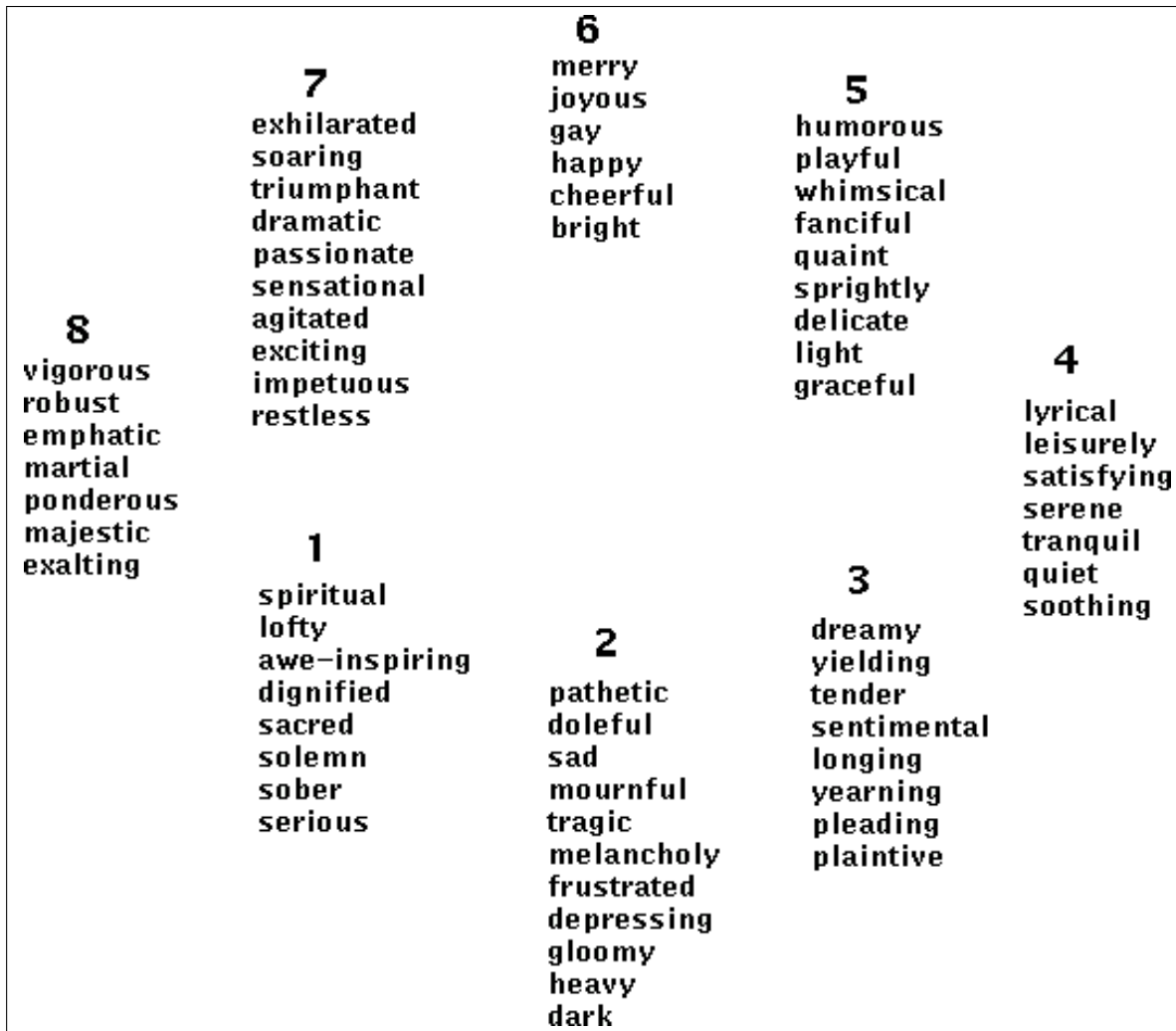


Figure 2-3: Hevner's adjective circle [21], p. 249

Since these initial ground-breaking studies, the field of music and emotion has blossomed into a thriving community whose current researchers include Paul R. Farnsworth, Leonard B. Meyer, Patrik N. Juslin, John A. Sloboda, Emery Schubert, Alf Gabrielsson, Erik Lindström, and David Huron, to name a few. The work of these scholars varies from emotional analysis of short musical excerpts to continuous measurement of emotion and its evolution throughout an entire piece of music.

Both Farnsworth [12, 14] and Schubert [45] have provided amendments of Hevner's original model. In these adjustments, Hevner's adjective groups are scrutinized and rearranged to form more precise adjective groups. Farnsworth modified Hevner's eight adjective groups

<b>Musical element</b>	<b>dignified/ solemn</b>	<b>sad/ heavy</b>	<b>dreamy/ sentimental</b>	<b>serene/ gentle</b>
<b>Mode</b>	major 4	minor 20	minor 12	major 3
<b>Tempo</b>	slow 14	slow 12	slow 16	slow 20
<b>Pitch</b>	low 10	low 19	high 6	high 8
<b>Rhythm</b>	firm 18	firm 3	flowing 9	flowing 2
<b>Harmony</b>	simple 3	complex 7	simple 4	simple 10
<b>Melody</b>	ascend 4	–	–	ascend 3
	<b>graceful/ sparkling</b>	<b>happy/ bright</b>	<b>exciting/ elated</b>	<b>vigorous/ majestic</b>
<b>Mode</b>	major 21	major 24	–	–
<b>Tempo</b>	fast 6	fast 20	fast 21	fast 6
<b>Pitch</b>	high 16	high 6	low 9	low 13
<b>Rhythm</b>	flowing 8	flowing 10	firm 2	firm 10
<b>Harmony</b>	simple 12	simple 16	complex 14	complex 8
<b>Melody</b>	descend 3	–	descend 7	descend 8

Table 2.1: Hevner’s weighting of musical characteristics in 8 affective states [22], p. 626

to form a new categorization of nine groups. This alteration strengthened the relationship between adjectives within each group and it weakened the circular aspect of the model [45]. Schubert’s work is an extension of Farnsworth’s revision, and involves a two-dimensional emotion space (2DES) with valence (pleasure) along the x-axis and arousal (activity) on the y-axis (see Figure 2-4). Building on Farnsworth’s findings, Hevner’s original eight groups are mapped to nine new groups (A through I) with a more precise subset of 46 of the 67 original adjectives. Several adjectives were also added from Russell’s circumplex model of emotion [44] and Whissell’s dictionary of affect [59]. These newly mapped emotional adjectives were then plotted in the 2DES based on their geometric position with respect to these two axes [45].

## 2.2 Music Theory

Mode is “a set of musical notes forming a scale and from which melodies and harmonies are constructed” [34]. Early Greek modes include Ionian, Dorian and Hypodorian, Phrygian and Hypophrygian, Lydian and Hypolydian, Mixolydian, Aeolian, and Locrian. The Ionian, Lydian, and Mixolydian modes are of major flavor and the Dorian, Phrygian, Ae-

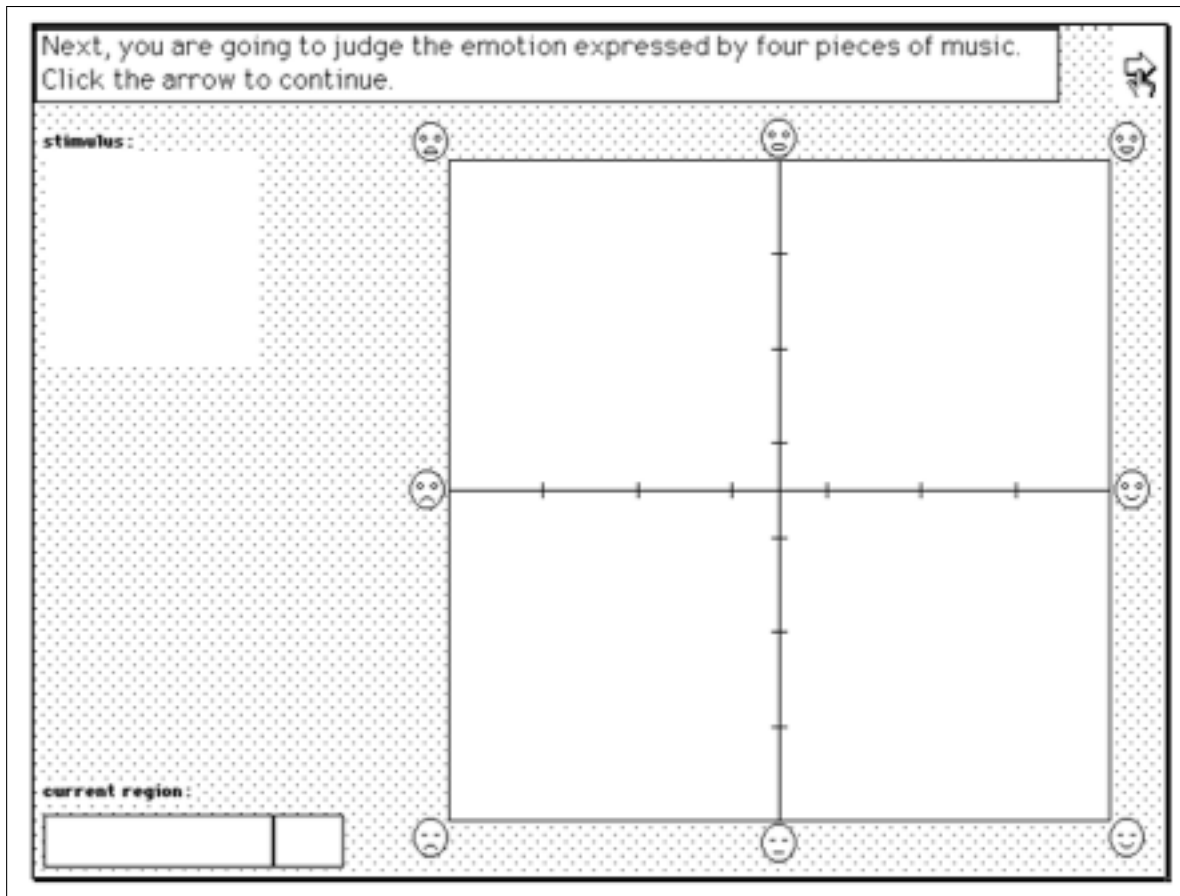


Figure 2-4: Schubert’s two-dimensional emotion space (2DES) [46], p. 564

olian, and Locrian modes are of minor descent. Major modes are often associated with happiness, gracefulness and solemnity while minor modes are related to the emotions of sadness, dreaminess, disgust, and anger [17].

Harmony is “the combination of simultaneously sounded musical notes to produce chords and chord progressions having a pleasing effect” [34]. Simple harmonies, or consonant chords, such as major chords, are often pleasant, happy, and relaxed. Complex harmonies contain dissonant notes that create instability in a piece of music and activate emotions of excitement, tension, anger, and sadness [17].

Tempo is defined as “the speed at which a passage of music is or should be played” [34], and is typically measured in beats per minute (bpm). A fast tempo falls into the range of 140 to 200 bpm (allegro, vivace, presto) and a slow tempo could be anywhere between 40 and

80 bpm (largo, lento, adagio). Fast tempi are generally considered lively and exciting, while slow and sustained tempi are majestic and stately. Depending on other musical factors, a fast tempo can trigger such emotions as excitement, joy, surprise, or fear. Similarly, a slow tempo is typical of calmness, dignity, sadness, tenderness, boredom or disgust [17].

The definition of rhythm with respect to emotion is not consistent among various authors, but the most common distinctions include regular/irregular (Watson) [58], smooth/rough (Gundlach) [18], firm/flowing (Hevner) [21], and simple/complex (Vercoe) [56]. Rhythm is officially defined as “the systematic arrangement of musical sounds, principally according to duration and periodic stress” [34]. The features proposed by the aforementioned researchers suggest that variations of the regularity or complexity of a rhythmic pattern in a piece of music trigger emotional responses. Regular and smooth rhythms are representative of happiness, dignity, majesty, and peace, while irregular and rough rhythms pair with amusement, uneasiness, and anger.

Loudness relates to the perceived intensity of a sound, while dynamics represent its varying volume levels. The dynamics of a piece of music may be either soft or loud. A loud passage of music is associated with intensity, tension, anger, and joy and soft passages are associated with tenderness, sadness, solemnity, and fear. Large dynamic ranges signify fear, rapid changes in dynamics signify playfulness, and minimal variations relate to sadness and peacefulness [17].

Other features that convey emotional responses include pitch (high-low), intervals, melody (melodic direction, pitch contour, and melodic motion), tonality (tonal-atonal-chromatic), timbre (number of harmonics), articulation (staccato-legato), amplitude envelope (round-sharp), musical form (complexity, repetition, new ideas, disruption, etc.), and interaction between factors [17]. These features are not being used in this context as they are either difficult to extract from an acoustic audio signal or they have not been thoroughly studied from the psychological perspective.

## 2.3 Feature Extraction

In the past several years the disciplines of digital signal processing (DSP) and music information retrieval (MIR) have evolved considerably. Early MIR systems processed symbolic data, such as MIDI and Csound orchestras and scores. These symbolic representations provide accurate pitch and rhythm information about the piece of music, which is often hard to extract from a raw audio signal. Only recently have DSP techniques advanced far enough to provide the MIR community with tools to extract such features as pitch, rhythm, and timbre. These tools are still in their infancy, but the implications are vast. Music Information Retrievalists can now use the actual audio data in their analyses instead of having to rely on the symbolic representation of a song.

The digital representation of an acoustic audio signal can be described either spectrally (frequency domain) or temporally (time domain). In the frequency domain, spectral descriptors are often computed from the Short Time Fourier Transform (STFT). By combining this measurement with perceptually relevant information, such as accounting for frequency and temporal masking, one can produce an auditory spectrogram which can then be used to determine the loudness, timbre, onset, beat and tempo, and pitch and harmony [23]. In addition to spectral descriptors, there also exist temporal descriptors, which are composed of the audio waveform and its amplitude envelope, energy descriptors, harmonic descriptors, derived from the sinusoidal harmonic modeling of the signal, and perceptual descriptors, computed using a model of the human hearing process. The items listed here are examples of low-level audio descriptors (LLD), which are used to depict the characteristics of a sound. Examples of spectral descriptors include the spectral centroid, spread, skewness, kurtosis, slope, decrease, rolloff point, and variation. Harmonic descriptors include the fundamental frequency, noisiness, and odd-to-even harmonic ratio. Finally, perceptual descriptors include Mel-frequency cepstral coefficients (MFCC), loudness, sharpness, spread, and roughness [39]. From these LLD's a higher-level representation of the signal can be formed.

### 2.3.1 CLAM

The C++ Library for Audio and Music (CLAM) is an open source framework written in C++. CLAM takes an object-oriented approach to research and development in the field of audio and music, including digital signal processing and audio synthesis and analysis. It implements a Digital Signal Processing Object Oriented Metamodel (DSPOOM), which includes the concepts of Inheritance Hierarchies, Polymorphism, and Late Building. This model enables the inclusion of semantic meaning of an audio sample in its feature set. The CLAM framework allows for flexibility and efficiency with respect to feature extraction and music information retrieval [1, 2].

A useful implementation of the CLAM framework is the CLAM Music Annotator<sup>2</sup>. This tool allows one to analyze and visualize a piece of music. The annotator extracts LLD's, as well as high-level features like the roots and modes of chords, note segmentation, and song structure. CLAM's annotator can be customized to extract any set of features based on an XML description schema definition. For example, one could create a general schema that extracts a wide range of LLD's, or a specific schema could be designed for chord extraction. In terms of visualization, the annotator includes a Tonnetz visualization for tone correlation display and a key space, courtesy of Jordi Bonada and Emilia Gomez at the University of Pompeu Fabre, for major and minor chord correlation.

The ChordExtractor tool, a schema included in the CLAM framework, extracts both the root and the mode of the chords in a piece of music. It implements a slightly modified version of the Christopher Harte algorithm<sup>3</sup>. This tool was developed at the Queen Mary University of London, University of Pompeu Fabra, and the Semantic Interaction with Music Audio Contents (SIMAC) group, and was ported to C++ by David Garcia Garzon and Katy Noland. The data is output in XML format, which can then be parsed to use for key and mode estimation.

---

<sup>2</sup>[http://iua-share.upf.es/wikis/clam/index.php/Music\\_Annotator](http://iua-share.upf.es/wikis/clam/index.php/Music_Annotator)

<sup>3</sup><http://www.aes.org/e-lib/browse.cfm?elib=13128>

### 2.3.2 Creating Music by Listening [23]

Loudness is a perceptual description of the intensity of a sound. In Jehan's work it is computed via critical band reduction. The critical Bark bands are a psychoacoustic measure that corresponds to the critical bands of hearing. The critical bands are a series of frequency regions along the basilar membrane, each with a unique response to the amplitude of incoming signals<sup>4</sup>. Extracting loudness measurements from the critical band representation of the audio signal is both computationally efficient and perceptually accurate.

Similar to loudness, timbre information of an audio signal is reasonably well represented in the perceptual description of the critical band. The first twenty-three bands that span the human range of hearing are used. Timbre is a measure of the tonal quality of a sound. It includes information about everything but pitch, loudness, duration, or location [6]. The important LLD's that relate to timbre are the temporal envelope (attack or articulation), spectral flux (variation of harmonics), and spectral centroid (brightness) of a sound.

Jehan represents rhythm using the loudness curve of the audio signal. This curve is the combination of the loudness at onset, maximum loudness, loudness at offset, length of the segment, and time location of the maximum loudness relative to the onset. The loudness values are measured in decibels and the time and length values are measured in milliseconds. The onset of each segment corresponds with significant variations in the loudness level. The valleys and peaks in the loudness value of the signal correspond to the attack and decay of the musical instruments, such as a snare drum hit or the pluck of a guitar string, and as such, the rhythm of each audio segment can be accurately represented with these five values.

The tempo and beat of a song can be derived from the rhythmical information. Jehan's analysis includes a bank of comb filters that are logarithmically distributed throughout the range of 60 bpm to 240 bpm. Each filter resonates at a specific bpm value, where the highest peak is representative of the tempo. This value is also given a confidence rating based on its relation to other peaks in the filter bank.

---

<sup>4</sup>Scavone, G. Hearing. Taken from <http://ccrma.stanford.edu/CCRMA/Courses/152/hearing.html> on 02/01/2007.



Lastly, pitch and harmony information are described using a 12-class chroma vector. This vector is representative of the twelve pitch classes (C, C#/Db, D, E, F, F#/Gb, G, G#/Ab, A, A#/Bb, B), regardless of register. To compute the chroma vector, a Fast Fourier Transform (FFT) of an audio segment is computed after which the energy distribution of the power spectrum is folded down from six octaves to one octave containing the twelve pitch classes.

## 2.4 Emotion Detection in Music

Automatic emotion detection and extraction in music is growing rapidly with the advancement of digital signal processing, audio analysis and feature extraction tools. As a fledgling field, the feature extraction methods and emotional models used by its proponents are varied and difficult to compare; however, these first small steps are important in forming a basis for future research. Moreover, mood detection in music is beginning to be seen as a relevant field of music information retrieval and promises to be an effective means of classifying songs. Recent publications of Lu et al. and Skowronek et al., among others, will be discussed below.

One of the first publications on emotion detection in music is credited to Feng, Zhuang, and Pan. They employ Computational Media Aesthetics to detect mood for music information retrieval tasks [16]. The two dimensions of tempo and articulation are extracted from the audio signal and are mapped to one of four emotional categories; happiness, sadness, anger, and fear. This categorization is based on both Thayer’s model [51] and Juslin’s theory [24], where the two elements of slow or fast tempo and staccato or legato articulation adequately convey emotional information from the performer to the audience. The time domain energy of the audio signal is used to determine articulation while tempo is determined using Dixon’s beat detection algorithm [10].

Another integral emotion detection project is Li and Ogihara’s content-based music similarity search [28]. Their original work in emotion detection in music [27] utilized Farnsworth’s ten adjective groups [13]. Li and Ogihara’s system extracts relevant audio descriptors using

MARSYAS [54] and then classifies them using Support Vector Machines (SVM). The 2004 research utilized Hevner’s eight adjective groups to address the problem of music similarity search and emotion detection in music. Daubechies Wavelet Coefficient Histograms are combined with timbral features, again extracted with MARSYAS, and SVMs were trained on these features to classify their music database.

Implementing Tellegen, Watson, and Clark’s three-layer dimensional model of emotion [50], Yang and Lee developed a system to disambiguate music emotion using software agents [62]. This platform makes use of acoustical audio features and lyrics, as well as cultural metadata to classify music by mood. The emotional model focuses on negative affect, and includes the axes of high/low positive affect and high/low negative affect. Tempo is estimated through the autocorrelation of energy extracted from different frequency bands. Timbral features such as spectral centroid, spectral rolloff, spectral flux, and kurtosis are also used to measure emotional intensity. The textual lyrics and cultural metadata helped to distinguish between closely related emotions.

Alternatively, Leman, Vermeulen, De Voogdt, and Moelants employ three levels of analysis, from subjective judgments to manual-based musical analysis to acoustical-based feature analysis, to model the affective response to music [26]. Their three-dimensional mood model consists of valence (gay-sad), activity (tender-calm), and interest (exciting-boring). The features of prominence, loudness, and brightness are present along the activity axis, while tempo and articulation contribute to varying degrees of valence. The axis of interest is not clearly defined by any features.

Wang, Zhang, and Zhu’s system differs slightly from the aforementioned models in that it analyzes symbolic musical data rather than an acoustic audio signal [57]. However, the techniques used are still relevant with respect to emotion detection in music. The user-adaptive music emotion recognition system addresses the issue of subjectivity within mood classification of music. This model employs Thayer’s two-dimensional model of emotion with some modifications. Both statistical and perceptual features are extracted from MIDI song files, including pitch, intervals, tempo, loudness, note density, timbre, meter, tonality (key and mode), stability, perceptual pitch height, and the perceptual distance between two

consecutive notes. SVMs were then trained to provide personally adapted mood-classified music based on the users opinions.

Another implementation of Thayer’s dimensional model of emotion is Tolos, Tato, and Kemp’s mood-based navigation system for large collections of musical data [53]. In this system a user can select the mood of a song from a two-dimensional mood plane and automatically extract the mood from the song. Tolos, Tato, and Kemp use Thayer’s model of mood, which comprises the axes of quality (x-axis) and activation (y-axis). This results in four mood classes, aggressive, happy, calm, and melancholic. A twenty-seven-dimension feature vector is used for the classification of the audio data. This vector contains cepstral features, power spectrum information, and the signal’s spectral centroid, rolloff, and flux. The authors conclude from the results of their studies that there are strong inter-human variances in the perception of mood and different perceptions of mood between cultures. They also deduce that the two-dimensional model is well suited to small portable devices as only two one-dimensional inputs are required.

Building on the work of Li and Ogihara, Wiczorkowska, Synak, Lewis, and Ras conducted research to automatically recognize emotions in music through the parameterization of audio data [61]. They implemented a k-NN classification algorithm to determine the mood of a song. Timbre and chords are used as the primary features for parameterization. Their system implements single labeling of classes by a single subject with the idea of expanding their research to multiple labeling and multi-subject assessments in the future. This labeling resulted in six classes: happy and fanciful; graceful and dreamy; pathetic and passionate; dramatic, agitated, and frustrated; sacred and spooky; and dark and bluesy.

A third system to employ the psychological findings of Thayer is that of Lu, Liu and Zhang, who introduced a method for automatically detecting and tracking mood in a music audio signal [32]. They created a hierarchical framework to extract features from the acoustic music data based on Thayer’s psychological studies [51]. This model classifies an emotion on a two-dimensional space as either content, depressed, exuberant, or anxious/frantic. Intensity, timbre, and rhythm are extracted and used to represent the piece of music. The first feature, intensity, is measured by the audio signal’s energy in each sub-band. Timbre

is represented by the spectral shape and contrast of the signal, and rhythmic information is gauged by its regularity, intensity, and tempo. This model also implements mood tracking, which accounts for changing moods throughout a piece of music. In the hierarchical model, intensity is first used as a rough measure of mood in each section of a piece of music, and timbre and rhythm are later applied to more accurately define each section’s mood space.

Less focused on the issue of the actual emotion detection in music, Skowronek, McKinney, and van de Par focused on discovering a ground truth for automatic music mood classification [48], which classified musical excerpts based on structure, loudness, timbre, and tempo. Russell’s circumplex model of affect was used in conjunction with nine bipolar mood scales. They found that “easy to judge” excerpts were difficult to determine, even by an experienced listener. In terms of affective vocabulary, they surmised that the best labels were tender/soft, powerful/strong, loving/romantic, carefree/lighthearted, emotional/passionate, touching/moving, angry/furious/aggressive, and sad.

Lastly, an emerging source of information relating to emotion detection in music is the Music Information Retrieval Evaluation eXchange’s (MIREX) annual competition, which will for the first time include an audio music mood classification category<sup>5</sup>. This MIR community has recognized the importance of mood as a relevant and salient category for music classification. They believe that this contest will help to solidify the area of mood classification and provide valuable ground truth data. At the moment, two approaches to the music mood taxonomy are being considered. The first is based on music perception, such as Thayer’s two-dimensional model. It has been found that fewer categories result in more accurate classifications. The second model comes from music information practices, such as All Music Guide and MoodLogic, which use mood labels to classify their music databases. Social tagging of music, such as Last.FM, is also being considered as a valuable resource for music information retrieval and music classification.

---

<sup>5</sup>[http://www.music-ir.org/mirex2007/index.php/Audio\\_Music\\_Mood\\_Classification](http://www.music-ir.org/mirex2007/index.php/Audio_Music_Mood_Classification)

## 2.5 Music Classification Frameworks

There exist many music classification frameworks and music information resources. This work is particularly relevant in the organization of music and the creation of playlists. A number of taxonomies, techniques, and frameworks are explained in the following section.

All Music Guide<sup>6</sup> (AMG) is an extensive online music database containing metadata for artist, albums, and songs. In addition to basic metadata, such as names, credits, copyright information, and product numbers, music is also classified by descriptive content such as style, tone, mood, theme, and nationality. Style is defined as sub-genre categories [33]. The site also provides relational content like similar artists and albums and influences. Lastly, biographies, reviews, and rankings are an integral part of the site's metadata as it allows one to gain cultural and social information about a piece of music.

MoodLogic<sup>7</sup> is an application that provides the listener with a multitude of alternative classification tools to better explore their music collection. These tools include the standard genre and year classification, but also explore tempo (slow/fast) and mood (aggressive/upbeat/happy/romantic/mellow/sad) classification. More obscure features range from danceability, energy level, and memorability of melody, to the topic of the lyrics, general type of instruments, and sound quality.

Corthaut, Govaerts, and Duval presented three categories of metadata classification [7]. Editorial metadata, such as music reviews, are manually input by music experts. Acoustic metadata provides an objective set of musical features obtained through analysis of an audio file. Thirdly, cultural metadata is the result of emerging patterns evident in documents produced by a specific society or environment. Within these three categories exist a multitude of metadata schemas. The most popular taxonomy is genre, followed by mood and context/purpose. Other common features include tempo, instrumentation/orchestration, and similar artists. Corthaut et al. use a metadata schema that includes version, language, bpm, voice, genre, instrument, mood, rhythm style, danceability, party level, global popu-

---

<sup>6</sup><http://www.allmusic.com>

<sup>7</sup><http://www.moodlogic.com>

larity, actuality, geographical region, occasion (Christmas, Halloween, etc.), type, loudness, dance style, continent, and sub-genre.

Tzanetakis' Music Analysis, Retrieval and Synthesis for Audio Signals (MARSYAS) software framework for audio processing and analysis [54, 55] is an excellent set of tools for music classification. This framework allows one to perform audio analysis, such as FFT, LPC, or MFCC that can then be stored as a feature vector for use in a classification and segmentation system. Both top-down flow and hierarchical classification and traditional bottom-up processing are supported by the framework, which results in a flexible architecture suited to most applications of music information retrieval and classification.

Zhang's semi-automatic approach for music classification addresses issues of subjectivity, clustering and user feedback [63]. The author's system classifies music as either vocal or instrumental: chorus, male solo, and female solo define the former category, while the latter is grouped according to its instrumentation (string, wind, keyboard, percussion). Vocal content is determined by the average zero-crossing rate and frequency spectrum, and instrumentation is the result of harmonic, timbral, and rhythmic feature extraction and correlation. The user is then able to manually correct any errors in the system and then re-classify.

Pohle, Pampalk, and Widmer presented an evaluation of frequently used audio features for classification of music into perceptual categories [40]. The perceptual categories are tempo, mood (happy/neutral/sad), emotion (soft/neutral/aggressive), complexity, and vocal content. The features extracted from the audio signal include timbral texture (spectral centroid, spectral rolloff, spectral flux, zero crossing rate, and the first five MFCCs), beat histogram, pitch histogram, MPEG-7 LLDs (audio power, audio spectrum centroid, audio spectrum spread, audio spectrum flatness, and audio harmonicity), and various other features such as spectral power and monophonic melody estimation. They found that the overall results were barely above baseline for the examined categories. As well, genre and emotion categorization worked the best, with emotion displaying a strong correlation to timbre. The authors believe that in conjunction with audio feature extraction there needs to be an inclusion of cultural data, usage patterns, listening habits, and song lyrics.

## 2.6 Playlist Generation Tools

Two approaches to playlist generation exist, each with their own positive and negative aspects. Automatic playlist generation attempts to create a mix of songs based on a set of criteria that is either set forth by the listener or deduced by the algorithm from the user’s listening habits. The opposing view to automatic playlist generation focuses on creating a set of tools with which the user may create their own music mix.

Mandel, Poliner, and Ellis employed SVMs in an active learning algorithm for music retrieval and classification [33]. This system was tested against AMG moods, styles, and artists. The most popular moods were rousing, energetic, playful, fun, and passionate, while popular styles include pop/rock, album rock, hard rock, adult contemporary, and rock & roll. The authors found MFCC statistics performed best, followed by Fisher Kernel and single Gaussian Kullback-Leibler divergence. Also, performance was more dependent on the number of rounds of active learning and less on the number of labeled examples. This active learning algorithm is well suited to the application of automatic playlist generators, based on a specific mood or style.

Andric and Haus introduced a system to automatically generate playlists by tracking a user’s listening habits [4]. Their algorithms produce playlists according to the listener’s preferences. The emphasis is on listening habits, not metadata. Though somewhat unsuccessful in its evaluation, the system provided useful insight into the relationship between listening habits and playlist generation.

Cunningham et al.’s work tackles music classification from the perspective of organization of personal music collections and the art of playlist creation [9]. The authors survey a number of classification techniques used by listeners with respect to organizing their personal music collections. These traditional techniques include organizing music by date of purchase, artist, genre or favorability, characterizing music by its intended use, such as an event or occasion, and the use of metadata and extra-musical documents. The survey also addresses issues such as distributed music collections, sharing of music, archiving of music, and the varying size of personal music collections. The authors conclude that music classification

is often a personalized activity and is more accurate and applicable when the classification methods can be tailored to the individual. In a second publication [8], Cunningham et al. look at unconventional paradigms for playlist creation as a way to create a playlist for an intended purpose or occasion. The authors identify the need for extensive browsing structures and an interactive browsing environment as key features in the creation of a good mix or playlist. Another important element is the ability to dynamically create playlists on portable music devices, such as mp3 players. This work presents a contrasting view to automatic playlist generation tools, and places more emphasis on providing listeners with relevant tools with which they may create their own playlists.

## 2.7 Music Recommendation Systems

Several music recommendation systems are described below. The services presented here represent the most popular and successful tools currently available to music listeners. They provide many useful services in addition to music recommendation, such as descriptive tagging and playlist generation, which are relevant to this thesis.

The Echo Nest<sup>8</sup> offers a new breed of music recommendation service. The founders, Brian Whitman and Tristan Jehan, address the issue from both a cultural and technical perspective. Cultural metadata is gathered from textual data on the web, such as blogs, music reviews, and chat rooms. This data provides a cultural and social context for the music being recommended [60]. Audio analysis is also performed to extract salient features from the audio signal, such as pitch, rhythm, loudness, and timbre [23]. The combination of these two sources of information is essential to the Echo Nest's recommendation service as it allows them to surpass the basic recommendation technique of collaborative filtering. The Echo Nest is able to create seamlessly beat-matched playlists, organize a personal music collection, as well as provide recommendations based on the cultural and audio data of a song.

---

<sup>8</sup><http://www.echonest.com>



Pandora Internet Radio<sup>9</sup> is built on the Music Genome Project and provides music recommendations through an online streaming music service. Pandora's music discovery engine is based on human analysis of musical content. Expert musicologists have spent years analyzing music from the past six decades in order to come up with a sophisticated taxonomy of musical information. Over 400 musical features are hand-labeled in order to accurately describe a piece of music. Several of these features include rhythm syncopation, key tonality, vocal harmonies, and displayed instrumental proficiency<sup>10</sup>. Users are able to find new music and listen to custom Internet radio stations based on their music tastes.

Last.FM<sup>11</sup> and Audioscrobbler<sup>12</sup> together constitute one of today's most popular social tagging and music recommendation systems. Audioscrobbler is a cross-platform plug-in that monitors a user's listening habits. This plug-in sends the user's currently playing track to the Last.FM site where it is stored in the user's profile. The Last.FM website then allows users to view their listening history and patterns, as well as those of their friends and others who have similar tastes. The site also enables users to tag artists, albums, and songs with descriptive words.

Goombah<sup>13</sup> is a music search and discovery tool that works with iTunes. By analyzing the user's iTunes library (metadata such as artists, albums, songs, play counts, etc.), Goombah is able to provide recommendations based on their listening habits. The service also offers free music downloads and the ability to browse the music collections of other users with similar music tastes.

iLike<sup>14</sup> is a social music discovery service that combines an iTunes plug-in with a web interface. The plug-in tracks the user's listening habits, similar to Audioscrobbler, and allows users to view similar artists and songs to the currently playing song. It also allows you to see what people in your social network are currently listening to. A sample page for Radiohead's Kid A album is shown in Figure 2-5. The iLike service is linked with

---

<sup>9</sup><http://www.pandora.com>

<sup>10</sup>[http://en.wikipedia.org/wiki/Pandora\\_\(music\\_service\)](http://en.wikipedia.org/wiki/Pandora_(music_service))

<sup>11</sup><http://www.last.fm>

<sup>12</sup><http://www.audioscrobbler.com>

<sup>13</sup><http://www.goombah.com>

<sup>14</sup><http://www.ilike.com>

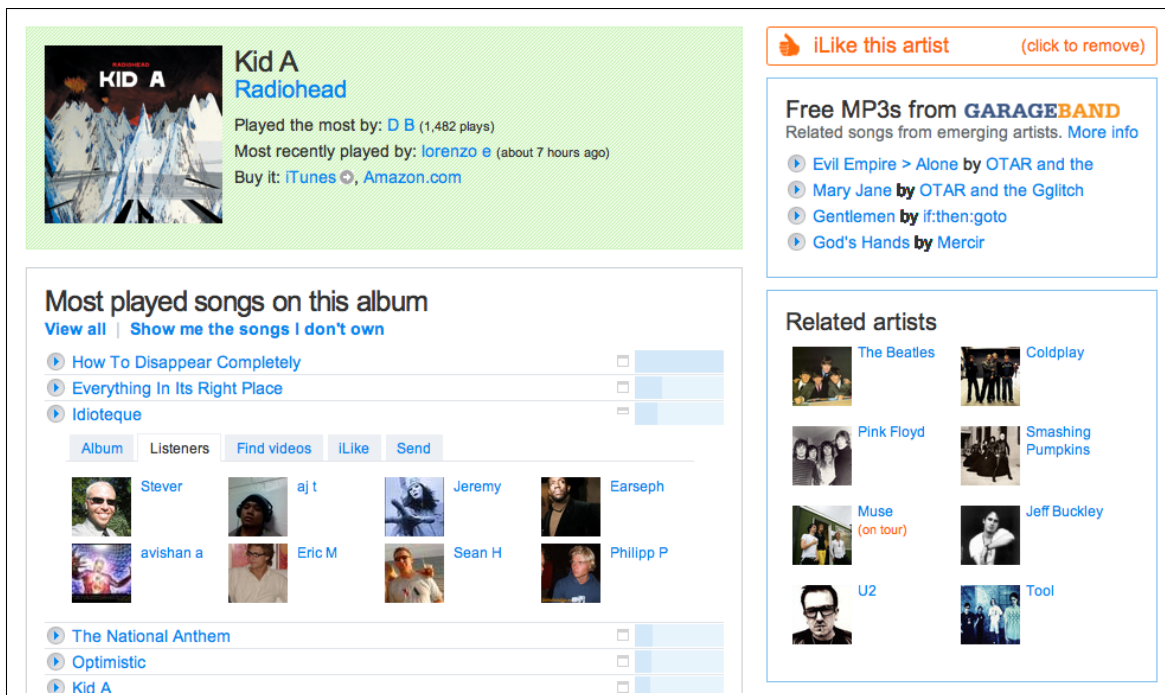


Figure 2-5: iLike Page for Radiohead's Kid A Album

<http://www.ilike.com/artist/Radiohead/album/Kid+A>

GarageBand<sup>15</sup>, an artist community that promotes independently produced music in order to provide free music downloads and recommendations from emerging artists.

Qcloud<sup>16</sup>, pronounced 'cloud', is a music search service with social networking features such as tagging. This "people powered music search" consists of a plug-in for the user's digital music player (iTunes, Songbird) in order to organize their library, and an online service that aggregates the listening habits and thoughts (via tags) of its community to provide recommendations. On the website one can search for music by artist or song title, by tag, or by browsing through user profiles. The tagging aspect of Qcloud is similar to Last.FM in that it allows users to input free text descriptions of a song's characteristics and qualities.

MyStrands<sup>17</sup> is another popular social recommendation and music discovery service. This product offers personalized playlist creation, tagging, and real-time recommendations by

<sup>15</sup><http://www.garageband.com>

<sup>16</sup><http://www.qcloud.com>

<sup>17</sup><http://www.mystrands.com>

providing the user with a plug-in for their digital music player. MyStrands also offers a web service with tools to promote social interaction, such as information on who is currently listening to the same music as you, or updates about your friends' music tastes. There is also a well-developed public API, OpenStrands, which provides developers with a plethora of web services and access to tag and community data.

MOG<sup>18</sup> is an online community of music listeners who share their listening habits and musical tastes with personal blogs. The MOG plug-in tracks a user's listening patterns and displays it in their online profile. Users can also discover new music through like-minded community members and tag their music with descriptive words. Music discovery and recommendation is influenced largely by user recommendations, as seen in Figure 2-6. MOG implements Gracenote Music ID technology and is built using Ajax and Ruby on Rails.

## 2.8 Natural Language Processing

The areas of natural language processing and textual analysis are relevant to the field of music recommendation and classification in that they provide tools with which to extract meaning and context from cultural metadata, such as music reviews or collaborative content websites. A valuable natural language processing tool is commonsense reasoning, which is particularly suited to the analysis of song lyrics as it enables the mining of key concepts and contexts from the lyrics.

### 2.8.1 Commonsense Reasoning

Common sense is defined simply as “good sense and sound judgment in practical matters” [34]. This is general knowledge that a society or culture shares, including intuitions and basic instincts, believed to be a common, natural, and basic understanding. This includes information that does not require esoteric knowledge or detailed research and study.

---

<sup>18</sup><http://mog.com>

# Idioteque


Artist: Radiohead > Album: Kid A

[buy from itunes](#)
[listen to this song](#)


## Reviews

[add post](#)


### Who's Listening




kingwun




Gordon Ma...



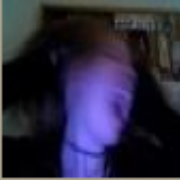
zackpe




Steven La...





mauerzahn



Liopleurodon







LIKE IT?

4

LIKE IT

### Tickle me emo!

Posted December 20, 2006 at 07:59 AM

Artist: Radiohead Album: Kid A Track: Idioteque


jonnyalmagest says:

Hahahahaha, that's funny.

Oh and by the way, according to my iTunes, I'm listening to the actual version of Idioteque, not some weird string quartet tribute...

Permalink | Write Comment | [Read All Comments \(5\)](#)

### Comments



Marta says:

That is soooo good ~ Thanks for sharing!

Posted 128 days ago | Delete

[+ SHARE](#)

Figure 2-6: MOG Page for Radiohead's Kid A Album

[http://mog.com/music/Radiohead/Kid\\_A/Idioteque](http://mog.com/music/Radiohead/Kid_A/Idioteque)

Marvin Minsky describes common sense as:

The mental skills that most people share. Commonsense thinking is actually more complex than many of the intellectual accomplishments that attract more attention and respect, because the mental skills we call “expertise” often engage large amounts of knowledge but usually employ only a few types of representations. In contrast, common sense involves many kinds of representations and thus requires a larger range of different skills [38].

Open Mind Common Sense (OMCS) [47] is a knowledge acquisition tool designed to build a large corpus of commonsense knowledge from community input data related to common everyday activities and knowledge. This project is similar to Mindpixel and Cyc. Since its conception in 1999 the corpus has grown to include over 700,000 facts. Data is input through a web interface that poses questions to the contributor such as “A hammer is for -----”, or “The effect of eating a sandwich is -----”. The system allows users to contribute facts, descriptions, and stories through natural language templates. It also enables community monitored validation and moderation. This system has been implemented in a number of projects, several of which are described below:

Emotus Ponens/EmpathyBuddy [29] is a textual affect-sensing tool that is modeled on real-world knowledge and commonsense reasoning. This application classifies sentences into the six different emotional categories, happy sad, angry, fearful, disgusted, and surprised in order to provide users with a way in which to emotionally evaluate their document. The commonsense knowledge is taken from the OMCS corpus.

ConceptNet [30] is an open source commonsense reasoning tool-kit developed in the Commonsense Computing Research Group at the MIT Media Laboratory. It provides tools for affect sensing, concept identification, topic and context awareness, concept projection, and analogies. The OMCS database contributes to this natural-language processing system. Knowledge is represented semantically through 1.6 million assertions about everyday life. These tools can be used to translate natural language textual input into meaningful concepts for use in practical contexts.

### 2.8.2 Lyrics

Baumann and Klüter presented a system for music information retrieval based on natural language textual input [5]. Non-musicians can use this system to retrieve music by forming queries using common everyday terms. The system is able to process song lyrics using a vector space model (VSM), which extracts and relates the most relevant terms in the document. Baumann and Klüter’s “super-convenient” method for music retrieval and classification shows progress in the area of semantic textual analysis with respect to song lyrics.

Logan, Kositsky, and Moreno employed semantic textual processing to automatically index music based on its lyrical content [31]. When applying this technique to artist similarity detection it was found to be most useful in conjunction with the analysis of the acoustic audio signal, though it performs reasonably well on its own. The authors used Probabilistic Latent Semantic Analysis (PLSA) to semantically analyze the song lyrics. This method classifies text documents by topics derived from frequently occurring words. Each document is represented by its characteristic vector, which can then be compared to other document vectors based on topic similarity and distance in the vector space. It was found that certain genres of music exhibited unique vocabularies. For example, Reggae music contains words like ‘girl’, ‘lover’, and ‘shout’, while words commonly found in Newage music include ‘ergo’, ‘day’, and ‘night’. These results show that song lyrics are an important source of metadata for music classification and retrieval.

## Chapter 3

# Design & Implementation

A mood classification system requires a strong understanding of emotional models and relevant audio and textual features with which to classify the song in question. In the previous section these models and features were described in detail. In this section these three key elements will be combined in an effort to design and implement a mood-based music classification system.

### 3.1 Emotional Model

Given that there exists a multitude of emotional models in the field of psychology, it is necessary to select one that is both applicable to musical contexts and is known and accepted in the fields of psychology and music psychology. Kate Hevner's circular model of emotion is one such case. The model was designed specifically for musical applications and has been thoroughly researched and studied with respect to the emotional effects of individual musical parameters. In Hevner's analysis emotions are grouped into eight categories, each with a distinct set of musical characteristics. The features that Hevner studied include mode, tempo, pitch, rhythm, harmony, and melody. The value of these six features in each of the eight affective groups is weighted according to its contribution to the group. For example, the group containing sad and heavy emotions is characterized by the strong

presence of a minor mode and low pitch content. This group is also distinguished by a slow tempo, complex harmony, and firm rhythm, though these three features are less influential than the mode and pitch content.

Since Hevner’s initial studies in the 1930’s, there have been several revisions to her model. Paul Farnsworth conducted the first study and revision in 1954. This study analyzed Hevner’s adjective clusters using a wide array of classical music excerpts. Farnsworth found that several of the 67 adjectives were inconsistent with the fellow members of its cluster. Another criticism of Hevner was her circular model, as Farnsworth states “There appears to be little empirical justification for the rationale of arranging the clusters in circle or clock-face form” [12]. Farnsworth presents his revision of Hevner’s model with 50 adjectives arranged in nine new clusters. These new clusters fit somewhat more accurately into a clock-face model; however, he admits that there are still some discrepancies. In 1969 Farnsworth added a tenth cluster, which included only one additional adjective, ‘frustrated’.

Emery Schubert is the most recent author to provide an update of Hevner’s adjective checklist [45]. In Schubert’s study a number of musically experienced people submitted their opinions regarding a list of 91 musical adjectives. The list of words includes Hevner’s 67 adjectives with additions from Russell’s circumplex model of emotion [44] and Whissell’s dictionary of affect [59]. The findings of the study reveal a slight correlation with Farnsworth’s results, but ultimately uncover nine new clusters arranged in a two-dimensional emotion space. A total of 46 emotional adjectives were kept from the original 91. Of these 46 terms, three originate from Russell’s circumplex model; relaxed, angry, and tense. The nine clusters can be described using the highest scoring adjective of the group, which are bright, lyrical, calm, dreamy, dark, majestic, tragic, tense, and dramatic. Of these nine clusters, all except ‘majestic’ can be arranged in a circular model, and because this study was based on a two-dimensional emotion space it can easily be mapped to other spaces of similar dimension, such as that of Russell’s model.

A mapping of Schubert’s update of Hevner’s adjective checklist to Russell’s circumplex model of emotion is relatively straightforward, considering the fact that both models are represented in a two-dimensional emotion space. Russell presents several scaling techniques



as a way to position a word in the circumplex space. The multidimensional scaling (found in [44], p.1168) resembles most closely the model used by Schubert in his update (see Figure 2-2), and as such is used as a guide to project other emotional adjectives onto the space. Table 3.1 compares the emotional models of Russell, Schubert, and Hevner and details their position in both the circumplex space and the two-dimensional valence-arousal space. Several of Russell’s emotions translate to more than one of Hevner and Schubert’s adjective groups. The values of two or more groups in Hevner or Schubert’s models are averaged to correlate with the new group in Russell’s model. For example, Hevner’s groups 4) Serene and 5) Graceful are averaged to produce Pleasure. Similarly, the musical feature weightings for the emotional space of Depression span Schubert’s groups F) Dark and G) Majestic, and are ultimately the average of the corresponding affect groups of Hevner’s model, 1) Dignified, 2) Sad, and 8) Vigorous. Also of note, cluster G) Majestic from Schubert’s model does not map directly to Russell’s circle and so values from this cluster are incorporated into cluster F) Dark because they both contain terms from the same groups in Hevner’s adjective circle. The resulting two-dimensional emotion space can be seen in Figure 3-1, which also attributes specific colors to each class of emotion for a more intuitive means of navigating the space.

Degree	Russell	Schubert	Hevner	Valence	Arousal
0°	Pleasure	B) Lyrical	4) Serene 5) Graceful	+	o
45°	Excitement	A) Bright	6) Happy	+	+
90°	Arousal	H) Dramatic	7) Exciting	o	+
135°	Distress	I) Tense	7) Exciting	-	+
180°	Displeasure	E) Tragic	2) Sad 3) Dreamy	-	o
225°	Depression	F) Dark G) Majestic	1) Dignified 2) Sad 8) Vigorous	-	-
270°	Sleepiness	D) Dreamy	3) Dreamy	o	-
315°	Relaxation	C) Calm	4) Serene	+	-

Table 3.1: Comparison of three emotional models, in terms of valence and arousal

The classification portion of our system implements Russell’s model in conjunction with Hevner’s original mapping of musical features to an emotional space. Of the original six musical parameters in Hevner’s studies, melody and pitch information are discarded because

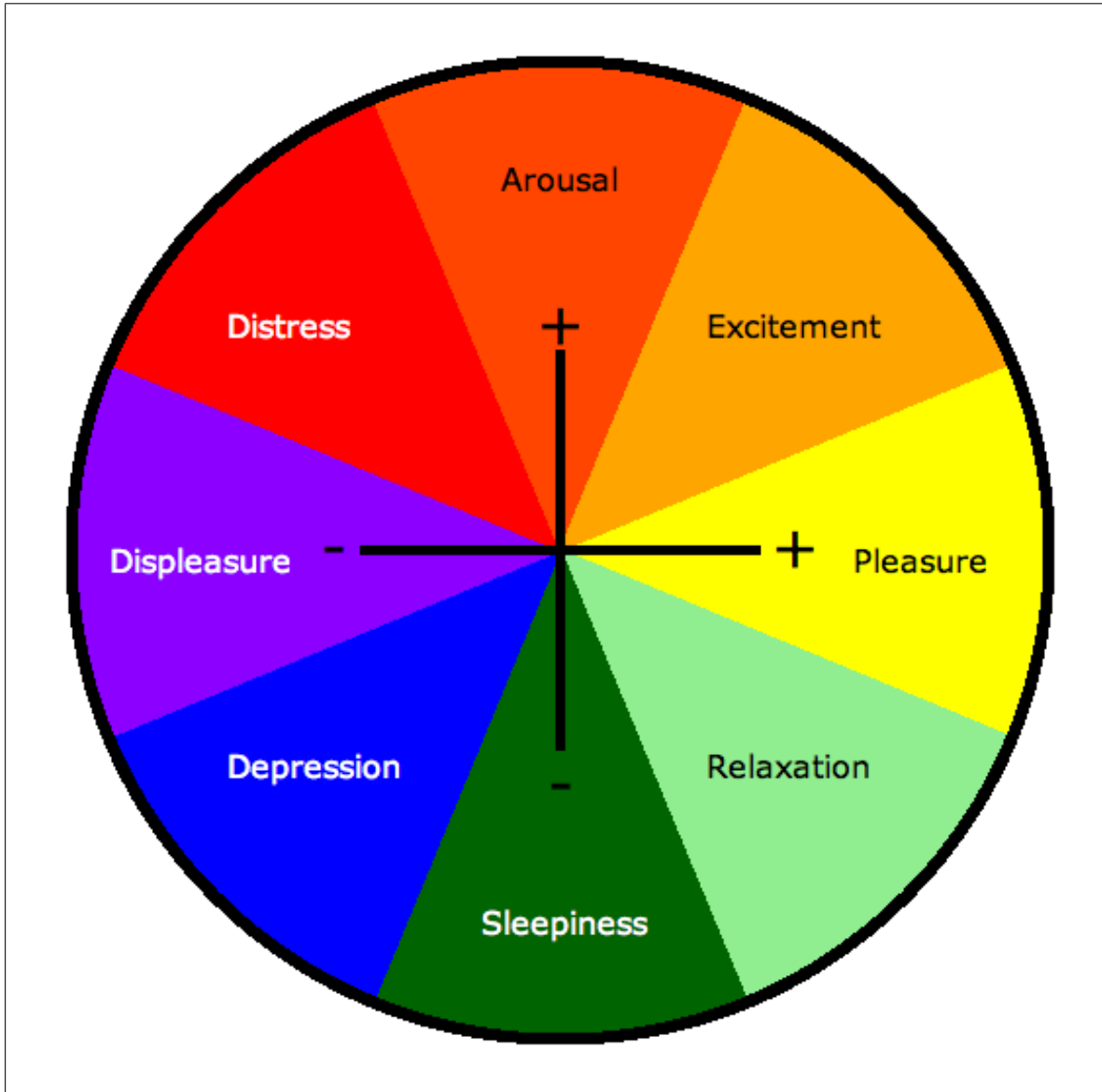


Figure 3-1: Color mapping of Russell's circumplex model of emotion used in this system

melody proves to be one of the more difficult features to extract from an audio signal, and the feature extraction tools used in this application distort the pitch material of the audio signal and destroy any sense of range and register by condensing all pitches into the span of an octave. The four remaining features employed here are mode, harmony, tempo, and rhythm. Additional data is used for the mapping of loudness to an emotional space, as this is not included in Hevner’s studies. This information is taken from Gabrielsson’s survey of musical factors and their effects on mood [17]. Loudness aligns itself roughly along the y-axis of arousal. High arousal and excitement are generally the result of loud music and peaceful and delicate emotions are triggered by soft music. The weightings for each feature and emotion are shown in Table 3.2. Positive values translate to major mode, simple harmony, fast tempo, regular rhythm, and high loudness, while negative values translate to minor mode, complex harmony, slow tempo, irregular rhythm, and low loudness.

<b>Mood</b>	<b>Mode</b>	<b>Harmony</b>	<b>Tempo</b>	<b>Rhythm</b>	<b>Loudness</b>
Pleasure	12	11	-7	-5	0
Excitement	24	16	20	-10	10
Arousal	0	-14	21	2	20
Distress	0	-14	21	2	10
Displeasure	-16	-2	-14	-3	0
Depression	-8	-4	-7	10	-10
Sleepiness	-12	4	-16	-9	-20
Relaxation	3	10	-20	-2	-10

Table 3.2: Mapping of musical features to Russell’s circumplex model of emotion

## 3.2 Audio Analysis

Through the extraction of relevant audio features one can gain pertinent knowledge of a song’s musical structure and its emotional characteristics. The five audio features presented here, mode, harmony, tempo, rhythm, and loudness, are extracted from the audio file using the C++ Library for Audio and Music (CLAM), an open source framework written in C++, and analysis software developed by Tristan Jehan and Brian Whitman of The Echo Nest. The ChordExtractor tool, a schema included in the CLAM framework, is used to extract the mode and harmony of a song and ENCLIAalyzer, an API to The Echo Nest’s analysis

tools, extracts features related to tempo, rhythm, and loudness.

### 3.2.1 Mode and Harmony

The ChordExtractor tool divides an audio file into small segments, on the order of 100 to 500 milliseconds, and assigns each segment a root (C, C#, etc.) and chord quality (Major, Minor, etc.). This data is stored in an Extensible Markup Language (XML) file with additional information regarding tuning position and strength, first and second chord candidate relevance, first and second chord index, energy, Harte pitch class profile, and Harte chord correlation. The chord quality may be one of Major, Minor, Major7, Minor7, Dominant7, MinorMajor7, Diminished, Diminished7, or Augmented, and could also appear as an open fifth dyad.

This selection of harmonies ranges from two-note dyads to four-note seventh chords. The open fifth dyad is constructed using two notes separated by a perfect fifth, while the four triads, Major, Minor, Diminished, and Augmented, consist of the root, third, and fifth. The Major and Augmented chords have a major third, but the Augmented chord contains an augmented fifth in place of the Major chord's perfect fifth. Similarly, the Minor and Diminished chords both include a minor third but with a differing fifth; the Minor triad uses a perfect fifth, while the Diminished triad contains a diminished fifth. The seventh chords are built using these four triads with the addition of a fourth note. The Major7 chord is a Major chord with an added major seventh. Likewise, a Minor7 chord fills out the Minor triad with a minor seventh note. The three remaining chords are constructed as follows: the Dominant7 chord, also known as a MajorMinor7, contains the chromatic pitch classes 0-4-7-10; the MinorMajor7 chord can be described using the pitch classes 0-3-7-11; and the Diminished7 chord is represented chromatically as 0-3-6-9.

In traditional functional harmony, where the stability, or consonance of a chord is determined by its position relative to the tonic, the chords mentioned above are used in various progressions to create tension and resolution in a piece of music. Consequently, emotional inflections become apparent in the music. Though functional harmony is largely abandoned

in popular Western music, the quality of a chord (Major, Minor, Diminished, etc.) is often used to create color effects in the music. These effects convey their own unique emotional qualities. For example, the Diminished7 chord's dissonant qualities evoke emotions of distress and anguish, while the Major chord triggers emotions of happiness and elation. The open fifth has a unique effect of its own, creating the illusion of openness and space.

As described earlier, the major and minor modes are important when used in an emotional context. In this system simple Major triads and complex Major7 and Dominant7 chords categorize major modes. The augmented chord represents a special case. This chord is somewhat dissonant, yet it maintains a major modality. To simplify the coding of harmony in this work the augmented triad is considered to be of major modality. Contrastingly, minor modes constitute minor and diminished triads, as well as the remaining seventh chords; Minor7, MinorMajor7, and Diminished7. The chord qualities of the three most frequently occurring roots in the song are collected and the majority of either major or minor chords determine the song's overall mode.

Similar to mode, the majority of simple or complex chords will determine the song's harmonicity. The harmonic complexity, and consequently the dissonance, of a chord is increased with the addition of more notes. A simple harmony consists of the four triads, Major, Minor, Diminished, and Augmented, and the open fifth dyad. Complex harmonies include Major7, Minor7, Dominant7, MinorMajor7, and Diminished7 chords.

### **3.2.2 Tempo, Rhythm, and Loudness**

The ENCLIANalyzer stores tempo, rhythm, and loudness data as specific tags in an XML file, similar to that of the ChordExtractor. Tags are associated both with global song information, such as the tempo and average/overall loudness, and with the individual segments of the audio file, including loudness, pitch, and timbre information. Global values give an impression of the general song characteristics and are good for classifying an entire song. Local segment data are useful for time-varying classifications where the evolution of these features is measured. The global information is used with respect to this system because

the goal is to assign a single mood to a piece of music.

The first and simplest feature extracted using the ENCLIAalyzer tool is tempo, which is measured in beats per minute (bpm). Values below 100 bpm, approximately, are classified as slow and values above are considered fast. However, bpm values are stored as a continuous range from 0 to 100, where 0 equals 40 bpm and 100 equals 200 bpm, so that tempo is not an absolute factor in the classification of the song. A song's XML analysis file output by the ENCLIAalyzer application contains global information related to tempo, such as the actual tempo value in bpm, the tempo confidence, beat variance, tatum length, tatum confidence, number of tatums per beat, time signature, and time signature stability. The tempo of the song is directly related to the tempo and tempo confidence values. If the tempo confidence is high then the tempo value is taken as is, otherwise the time signature and number of tatums per beat are consulted.

Secondly, the rhythmic values of a song consist of several rhythm-related features extracted by the ENCLIAalyzer. These are tempo confidence, beat variance, tatum confidence, and time signature stability. The beat variance is weighted highest of the four parameters, followed by tempo confidence, tatum confidence, and time signature stability. An average of these four values, weighted accordingly, results in a useful measure of rhythm regularity or irregularity. In the case of tempo confidence, this value indicates the amount of surety that exists with respect to tempo. A low confidence suggests a less stable tempo, one that is irregular. A regular tempo would have a high tempo confidence. Beat variance is another measure of regularity of rhythm in a song. A higher beat variance is equivalent to an irregular rhythm. In order to clearly describe a song's tatum confidence, the tatum must first be clearly identified. Jehan defines the tatum as follows

The tatum, named after jazz pianist "Art Tatum" in (J. A. Bilmes. Timing is of essence. Master's thesis, Massachusetts Institute of Technology, 1993.) can be defined as the lowest regular pulse train that a listener intuitively infers from the timing of perceived musical events: a time quantum. It is roughly equivalent to the time division that most highly coincides with note onsets: an equilibrium between 1) how well a regular grid explains the onsets, and 2) how well the onsets explain the grid [23].

If the tatum confidence is high then the presence of beats in the music is high, which also means that the rhythm of the music is more apparent and therefore a rough measure of a song’s rhythmic content. Moreover, if the time signature is relatively stable then the rhythm is regular.

A final feature extracted by the ENCLIANalyzer is loudness. The loudness values extracted from the song are measured on a scale from 20 dB (soft) to 80 dB (loud). These values can be broken down into measurements of the mean and variance of the maximum loudness, the mean and variance of the loudness at the beginning of a segment, and the mean and variance of the dynamics of the loudness. These six values provide information about the intensity and variation of the dynamics in the music.

### 3.3 Lyric Analysis

The lyrics of a song often contain a great deal of contextual information, including emotional content. By extracting the affective value of the lyrics one can gain additional information related to the mood of a piece. The `guess_mood` function, included in ConceptNet’s natural language tools, computes the affective value of raw text input. Our system uses the `guess_mood` function to extract salient emotional concepts and words from a song’s lyrics. Its output has been modified to reflect Russell’s dimensional emotional model rather than `guess_mood`’s original categorical model.

The system presented here uses natural language to classify music. The lyrics of a song are affectively analyzed to extract their emotional content. Keywords are extracted from the song’s lyrics and are mapped back to an emotional model consistent with that used by the audio analysis portion of the system. The affective value of the lyrics is then used in conjunction with the audio features to classify the song by mood.

Portions of the Lyricator software are used to analyze the lyrical content of a song and classify its emotional value [36]. With this software, the user begins by inputting the name of an artist or band that they would like to search for, followed by the name of a specific

song by this artist or band. Lyricator then searches a database of files containing the lyrics of popular songs and analyzes the emotional content of the chosen song. Mehrabian’s PAD representation is used as the emotional model for the classification and the OMCS corpus of raw sentences is employed for spreading activation and context mapping. Each word in the song is assigned a PAD value, which is then added together to give a total PAD score for the song. Using Scott Vercoe & Jae-woo Chung’s “Affective Listener” as a reference, the song is classified as engaging (+P, +A), soothing (+P, -A), boring (-P, -A), or annoying/angry (-P, +A). Though the PAD emotional model is not used in this mood classification system, the concepts relating to commonsense reasoning and emotional mapping are quite relevant.

In addition to accepting natural language as a classification tool, the system will accept natural language input as a form of playlist generation. In order to effectively map textual input to musical output, the text must be processed to extract its affective value. This is done by identifying keywords in the text document as in a blog entry, for example, and by applying spreading activation on these keywords to gain a broader perspective of their context and meaning. This web of words is then weighted according to its relation to a set of affective terms, in this case the emotional model of the system. With this information in place the system can then select emotionally appropriate music to accompany the text query.

The text analysis features of this system build on the previous work of Mood:Machine (formerly mySoundTrack) [37]. This is a piece of software developed to convert natural language input to music in the form of a playlist, or mix of songs. The taxonomy is based on AMG’s mood classification; where artist, album and song are each assigned several moods that reflect the affective response of the music. Mood:Machine uses a modified version of the `guess_mood` function to reflect the moods present in AMG’s taxonomy. This allows it to search a large corpus of commonsense knowledge to better connect the user’s query to one of the moods in Mood:Machine. The program then generates an iTunes playlist containing a list of songs that are based on the top moods returned by ConceptNet.

ConceptNet provides many of the text processing and affect sensing tools used in this system. The values computed by ConceptNet’s `guess_mood` function are classified according to



Ekman’s basic emotions, happy, sad, angry, fearful, disgusted, and surprised. Each emotion is weighted according to the relative presence of that emotion in the text document. However, since Ekman’s basic emotions do not correspond directly with the emotional model used by our mood classification system, the `guess_mood` function has been modified to output emotions resembling Russell’s circumplex model. In place of Ekman’s six emotions are pleasure, excitement, arousal, distress, displeasure, depression, sleepiness, and relaxation. These represent the eight octants of Russell’s circumplex model, which are defined by the axes of positive and negative valence and arousal. The presence of one or many of these emotions in the lyrics of a song will influence its overall mood classification. For example, if the lyrics exhibit a very excited emotion while the audio features map to a more relaxed emotion then the compromise between the two would be in the area of pleasure, depending on the weight of either factor.

## **3.4 Classification of Music**

The classification of a song incorporates the five musical features: mode, harmony, tempo, rhythm, and loudness, extracted from the audio signal and the affective value of the song’s lyrics. A decision tree is used for preliminary classification of the song database. Next, a k-nearest neighbor (k-NN) classification algorithm is applied to the database using the weights outlined in Table 3.2 as initial training data. For example, a song with a major mode, simple harmony, fast tempo, irregular rhythm, and moderate loudness will be considered ‘exciting’. The result of the k-NN classifier is then combined with the lyric’s affective value to provide a global emotion for the song.

### **3.4.1 Classification of Audio Features**

Decision trees provide a useful means of classifying data based on Boolean functions. This classification method assigns an initial mood to a song that is representative of its five musical features. The decision tree takes as input the floating-point values of the song’s

features and based on a predetermined threshold classifies the song accordingly. A song may fall into one of 32 ( $2^5$ ) classes depending on its musical features (see Figure 3-2). Each class is assigned a mood from Schubert’s updated list of Hevner’s adjectives [45]. The moods are indicative of the path that the decision tree has taken. As such, a song with a major mode, simple harmony, slow tempo, regular rhythm, and soft loudness will result in a ‘delicate’ mood. This corresponds to the octant of Relaxation in Russell’s model of emotion because ‘delicate’ belongs to group C) Calm in Schubert’s model. As each of the 32 moods in the decision tree are derived from Schubert’s nine groups of emotional adjectives, the results from the decision tree can be simplified to correlate with the eight classes of Russell’s circumplex model (see Table 3.1).

The second iteration of classification implements a k-NN algorithm. The k-NN classification algorithm classifies an object based on its proximity to neighboring objects in a multidimensional space. The algorithm is trained with a corpus of approximately 360 songs, which are initially classified using the decision tree algorithm described above. The result of the k-NN training is a multidimensional feature space that is divided into specific regions corresponding to the eight class labels of the training set; pleasure, excitement, arousal, distress, displeasure, depression, sleepiness, and relaxation.

The number of nearest neighbors,  $k$ , used in this system is determined through cross-validation of the data, in which an initial subset of the data is analyzed and later compared with the analysis of subsequent subsets. Specifically, a leave-one-out cross-validation (LOOCV) method is implemented where each observation in the training set is validated against the remaining observations. This process is then repeated for each sample in the dataset. Through this process it was determined that the best number of nearest neighbors for this application is  $k=5$ .

When a new song is loaded into the system its class is unknown. Its position in the multidimensional space is compared to that of its nearest neighbors using the measure of Euclidean distance, and the song is assigned a class based on its proximity to its five nearest neighbors.

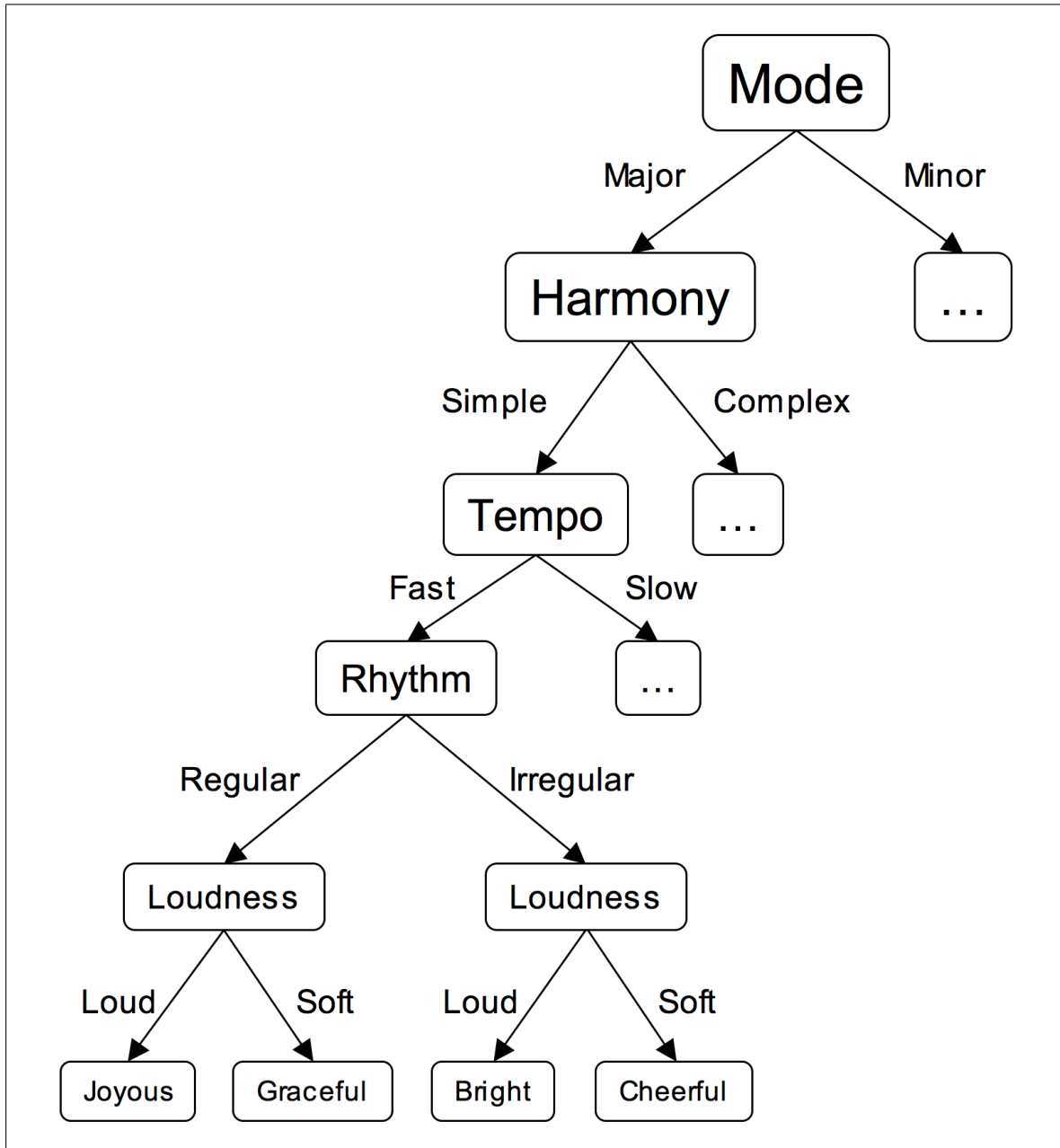


Figure 3-2: Simplified Decision Tree Schema

### **3.4.2 Classification of Lyrics**

The affective value of the song lyrics returned by ConceptNet’s analysis is used by our system as an additional measure of the song’s mood. By combining the classification of the audio features with the lyric analysis, a more accurate assignment of mood may be performed. The query to ConceptNet returns a list of several moods that are present in the song lyrics. These moods are each assigned a weight according to their relevance in the document. If there is a correlation between the top moods returned by ConceptNet and the resulting mood from the audio feature classification then our system will retain the original mood classification from the audio portion of the classification process. However, if there exists a discrepancy between the two results, which may occur if the songwriter intended for this incongruence, then the more predominant value, from either the audio or the lyrics, will determine the overall mood of the song.

## **3.5 Mood Player Interface**

The mood classification and exploration environment provides a similar functionality to most common media players, such as iTunes and Windows Media Player, but with the addition of several features relevant to mood and playlist generation. The interface is unique in that it allows one to view the values of individual music features and their effect on the mood of a song.

### **3.5.1 Features**

The Mood Player interface, shown in Figure 3-3 provides the user with a list of the songs in their music library and information pertaining to the song including ‘Song Title’, ‘Artist’, ‘Album’, ‘Duration’, ‘Genre’, ‘BPM’, and ‘Mood’. These fields provide relevant information about the song without cluttering the interface and are the most common features used in the generation of playlists. After loading a song into the library of the Mood Player, the user is able to view the song’s features and lyrics. The musical characteristics of the songs

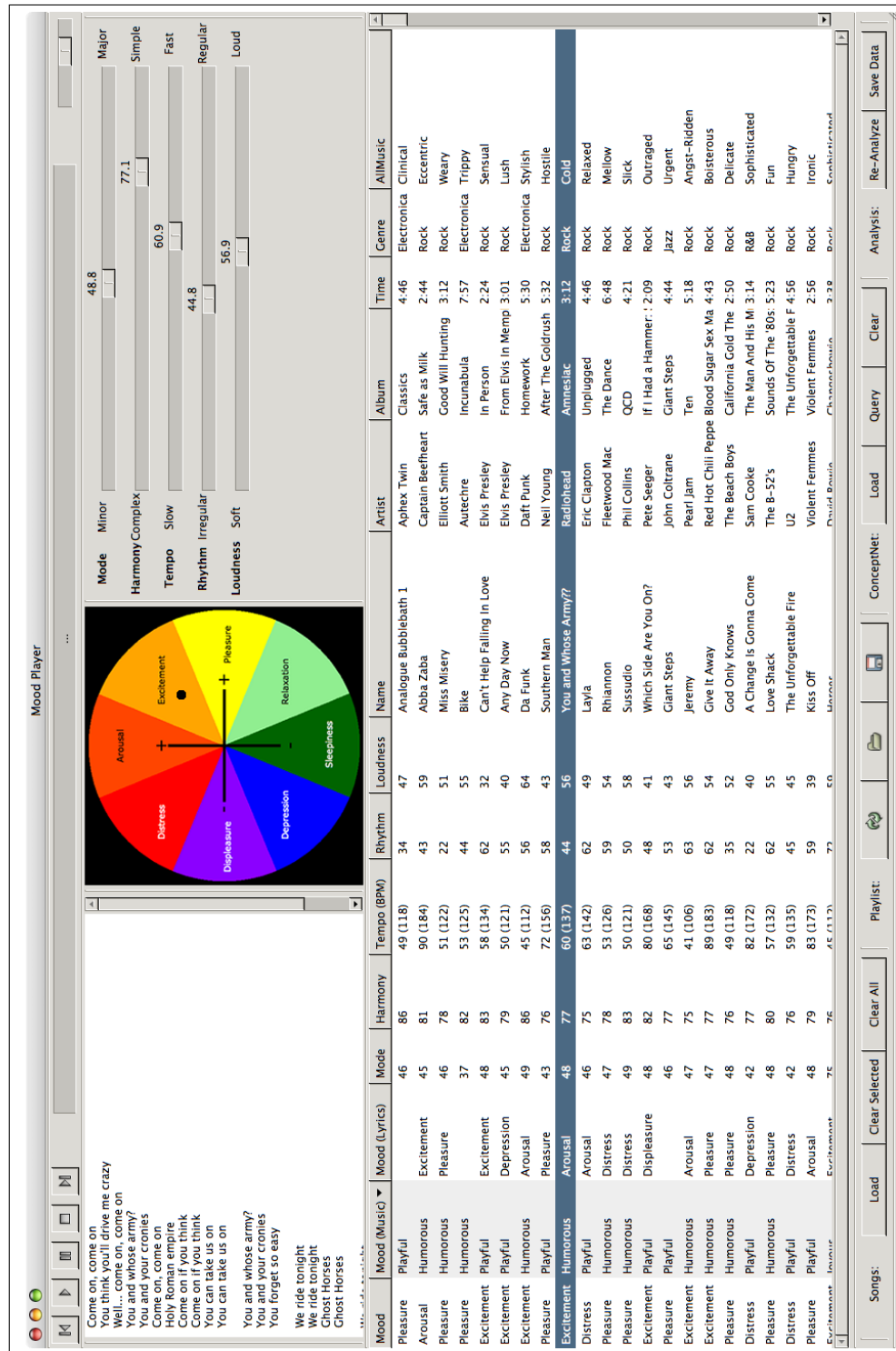


Figure 3-3: Mood Player Environment

in the library can then be browsed and compared by selecting a song. When the user finds a song that suits their mood they can then create a playlist using this song as a model. The user may also provide a textual query, such as the lyrics of a song or a blog entry. This free text input allows the user to quickly find music and create playlists based on the affective value of the document. A list of moods is returned by the query, which can then be used to focus the resulting playlist.

The novel features of the user interface are most visible in its playlist generation tools. The playlist options available to the user include a textual query option, the ability to set the position of the musical parameter sliders, and generation based on the Mood field. A playlist generated from text is useful for non-musicians and people who are unfamiliar with the technical theory of music. With this tool they may input any form of text that either conveys their current mood or describes the context in which they will be listening. Thus, no musical experience is necessary to find the correct song to fit a desired mood or context. In the case of slider-generated playlists, a user familiar with music theory may position the sliders in an arrangement to suit their current listening needs. These sliders may also be used in relation to a seed song. The user may choose a song that suits their mood, and then search for similar songs by finely adjusting one or more of the sliders. For example, if the current song has a fast tempo, but the user is searching for music that is slightly more relaxed, the tempo slider can be adjusted to a slower tempo value. As each song's features contribute to its global mood, the user may also use the Mood field to design a playlist. When the user has successfully created a playlist they may save their custom playlist for use in future contexts.

Visualization of music and mood is implemented through the use of sliders, representative of the individual musical parameters, and through a two-dimensional circular mood space. The sliders indicate the intensity of a musical parameter present in a song. Each of the five parameters is on a continuous bi-polar scale with negative values on the left, such as minor mode and complex harmony, and major values on the right, such as fast tempo and regular rhythm. The two-dimensional visualization allows the user to view the song's location in the circumplex model emotion. If the current playlist contains many songs with the same

mood, then these songs will cluster together in a specific location of the two-dimensional space. Playlists with varying moods will appear more loosely scattered throughout the space.

### 3.5.2 Frameworks, Software, & Technical Implementation

The mood classification system is implemented primarily in the Python programming language. This allows for cross-platform compatibility and easily maintainable code, as Python is a widely used language with a variety of available libraries. The graphical front-end to the classifier is written in PyGTK, a graphical toolkit for Python. PyGTK is a set of Python wrappers for the GTK+ GUI library. Several third party libraries are also used in the system, including PyGame for MP3 decoding and audio playback, eyeD3 for ID3 tag manipulation, NumPy for numerical processing and arrays, and BioPython for k-NN classification.

The feature extraction framework includes The Echo Nest's ENCLIANalyzer and CLAM's ChordExtractor. ENCLIANalyzer is an API available to a small group of Media Laboratory researchers and developers that provides access to The Echo Nest's proprietary audio analysis and feature extraction tools. ENCLIANalyzer extracts tempo, rhythm, and loudness features from an audio file. The API can be run separately from the mood classification system by simply dragging an MP3 file from iTunes onto the API application. An XML file is then output to the directory of the MP3 file. This process may also be executed internally using Python's 'os' library:

```
os.system("./EN/ENCLIANalyzer.app/Contents/MacOS/ENCLIANalyzer " + fileName)
```

As with the drag n' drop interface, the command line version will perform the analysis of the audio file and extract the features to an XML file in the directory of the original MP3 file.

CLAM is used in this system to extract pitch material from an audio file, including the root and quality of the chords in the song. Though CLAM is open source, the ChordExtractor

utility is a pre-compiled C++ binary. Similar to the ENCLIANalyzer, the ChordExtractor can be run from the command line, thus allowing it to be run from within the mood classification system using the following Python command:

```
os.system("./CLAM/ChordExtractor -s CLAM/Chords.sc -f .chords " + fileName)
```

The ChordExtractor uses the 'Chords.sc' schema to analyze the input MP3 file, 'fileName'. An XML file is output using a '.chords' extension.

The output of both the ENCLIANalyzer and ChordExtractor is in XML format. This requires an XML parser to extract the relevant fields from the XML file. Python provides a variety of parsing tools, including Expat, which is used here. XML is used primarily to allow various systems to share data in a common format.

Several fundamental issues exist with both the playback and feature extraction of MP3 files. For this reason, a LAME decoder and OGG encoder have been included in the system to ensure that every audio file will be analyzable and playable. In addition to MP3 input, the CLAM ChordExtractor will also accept WAV and OGG input provided by the LAME decoder and OGG encoder. Similarly, the PyGame library will decode OGG audio files for playback. Furthermore, if PyGame is unable to play an MP3 file, it will convert the file to the OGG format.

ID3 tags provide convenient and cross-platform compatible data representation and storage for information such as a song's mode, harmony, tempo, rhythm, and loudness. These features can be stored in user-defined text frames, which are included in the ID3 metadata of an MP3 audio file. The tags are processed using the eyeD3 Python module. This module supports both ID3v1 and ID3v2 formats. ID3v2.3 is used in this system.

In addition to issues with MP3 playback, there is a similar problem with song lyrics. Only recently have song lyrics been included with MP3 files in the form of an ID3 tag. For this reason, many songs do not include lyrical data. This fact is compensated for through the use of LyricWiki, an online open source lyric database. The database is accessible through a



SOAP web service API, which follows the WSDL standard. LyricWiki's PHP server accepts incoming requests for song lyrics, which can be searched for using the name and artist of the song. The lyrics returned by the SOAP web service are stored locally using the MP3 file name with the addition of a '.lyrics' extension. The current state of lyric storage in ID3 tags is not well supported and can be both unpredictable and unreliable, therefore the lyrics will not be stored in the MP3 file's ID3 tag. If no lyrics are returned from the server then the song is assumed to contain no lyrics and is classified strictly as an instrumental song.

After retrieving the song lyrics from LyricWiki, Montylingua and ConceptNet are used to analyze the song lyrics. These open source natural language processing and commonsense reasoning toolkits are also used to process the user's textual queries. Both Montylingua and ConceptNet are written in Python, which allows for a streamlined integration of the tools with the mood classification system. ConceptNet is run as an XML-RPC server where query results are returned to the mood classification client.

The mood space visualization is incorporated into the PyGTK interface in the form of a graphical drawing area where songs are positioned in using a polar coordinate system. Each mood corresponds to a specific coordinate location around the circular space. The user may also select a location in the mood space to retrieve a subset of songs that are closely related to the desired mood.



## Chapter 4

# Evaluation

The following section begins with observations regarding the performance and accuracy of the mood classification system. Next, the results of the mood classification system are thoroughly tested against popular music metadata systems. A set of songs are examined with respect to their emotional and contextual attributes and evaluated using data obtained from social tagging and classification websites. Following this analysis, results from a user evaluation are presented.

### 4.1 System Performance and Accuracy

Overall, the mood classification system described here performed well and showed promising results as a tool for music classification and listening. The playlist generation and mood browsing features were very easy to use and returned relevant results in the form of playlists that suited situations like studying/reading, in which case the song *A Case Of You* by Joni Mitchell was used to seed the generation of a list of 20 similar songs. The resulting list contained *Orinoco Flow* by Enya and *All I Have To Do Is Dream* by The Everly Brothers, both of which are quiet, relaxing songs.

#### 4.1.1 Classification of Music Database

The process of initially classifying a song by mood involves a great deal of processing power, and consequently the initial startup time of the application is somewhat slow. However, the data relating to a song's features are stored locally in meta files once a song has been analyzed, which means that consequent classification of this song will be instantaneous.

A database of 372 songs was analyzed. The analysis of a song began with the extraction of the MP3 file's ID3 metadata, such as Title, Artist, Album, Genre, and Duration. Next, the CLAM ChordExtractor was instantiated to obtain the chordal information, which was followed by the ENCLIAalyzer which extracted the tempo, rhythm, and loudness features of the song. The lyrics for the song were then fetched from LyricWiki and analyzed using ConceptNet. The values from the five features and the lyric analysis were then input to both the decision tree and k-NN algorithms.

The results of the feature extraction and analysis are shown in Figure 4-1. The five features in the left-hand column correspond to a song from each of the eight mood classes (see Table 4.1), which are distributed along the x-axis. The y-axis relates to the polarity of the feature, where 0 is negative and 100 is positive (0 = minor mode, 100 = major mode, for example). The right-hand column displays a histogram of the entire database for each of the five features. Here the x-axis represents a feature's polarity and the y-axis is the percentage of songs.

Mood	Song	Artist
1. Pleasure	Love Shack	The B-52's
2. Excitement	One More Time	Daft Punk
3. Arousal	Like A Virgin	Madonna
4. Distress	Black Star	Grateful Dead
5. Displeasure	Shine On You Crazy Diamond	Pink Floyd
6. Depression	Glory Box	Portishead
7. Sleepiness	Freddie's Dead	Curtis Mayfield
8. Relaxation	Orinoco Flow	Enya

Table 4.1: Eight songs classified by the system

Of the eight mood classes, all except Sleepiness returned satisfactory results. Though the classification algorithms were trained on Hevner's findings, it appears that Sleepiness

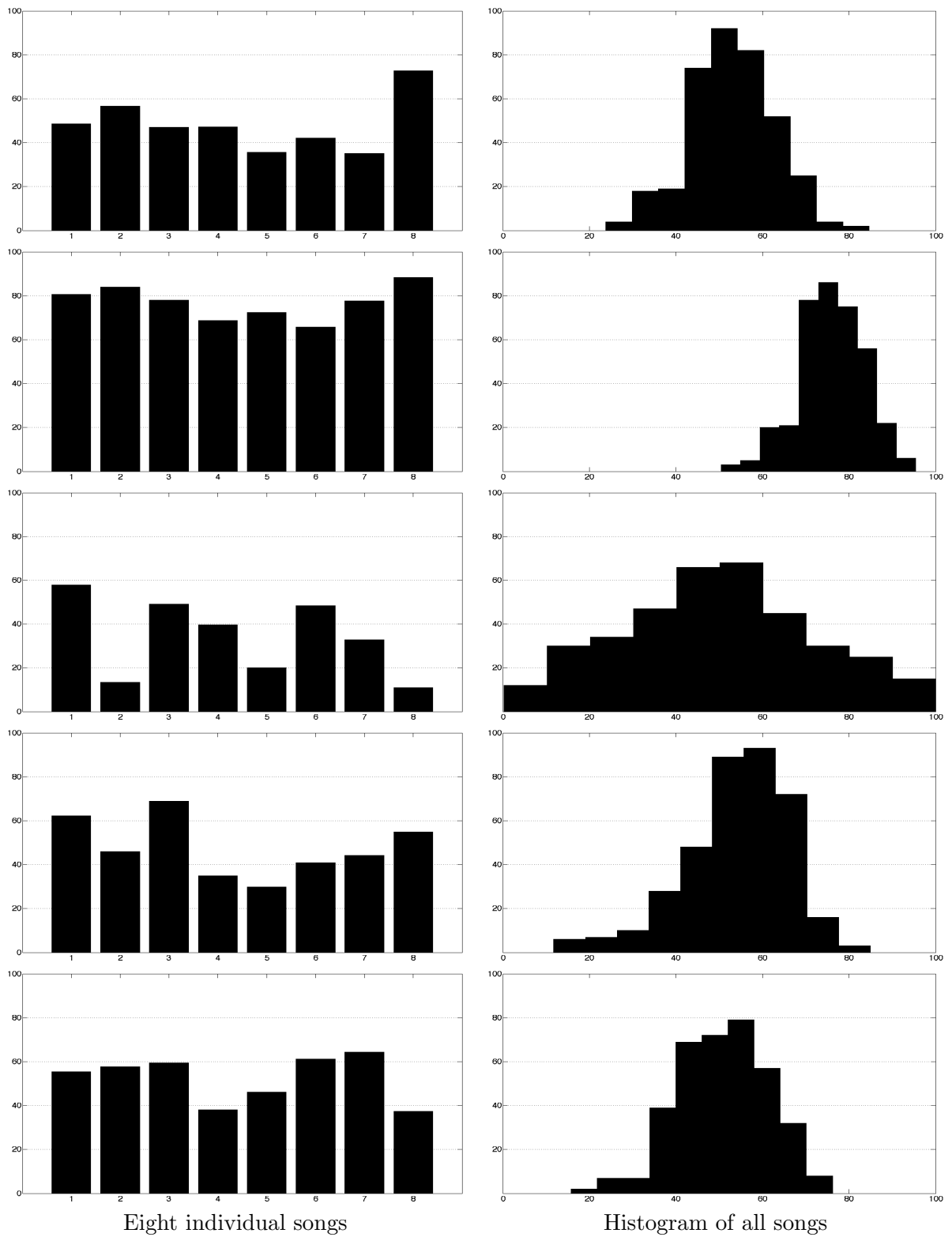


Figure 4-1: Results for the five features: Mode, Harmony, Tempo, Rhythm, Loudness

does not correlate precisely with Hevner’s affective group, Dreamy (see Table 3.1). This is particularly apparent with the song, *Freddie’s Dead*, which is a funk song that laments the death of Fat Freddie, a character in the film *Superfly* (1972) who is run over by a car. Though this song is by no means upbeat, it is definitely not characterized by Sleepiness. From the data in Figure 4-1 it appears that the Tempo and Rhythm are slightly higher than they should be, were this to be a ‘sleepy’ song.

The histograms produced from the analysis of the 372 songs found in Figure 4-1 show a uniform Gaussian distribution for each of the five features. However, the distribution for Harmony is not completely symmetric. This error can easily be corrected by re-evaluating the Harmony schema, which classifies Major, Minor, and Fifth chords as having simple harmony. Apparently, these three chords constitute the majority of the chords extracted from the songs in this database. A simple transposition of this data would result in a more even distribution.

#### 4.1.2 Classification of Lyrics

The analysis of the lyrical content of the eight songs listed above (Table 4.1) is displayed in Table 4.2. This table shows a correlation between keywords in the lyrics and the top three resulting moods from the analysis returned by ConceptNet. The values below each mood indicate the weighting, or relevance, of that mood in the song lyrics. For example, *Freddie’s Dead* is classified as Displeasure because of the keywords ‘dead’, ‘abused’, and ‘misery’, to name a few. The moods returned by ConceptNet were generally accurate, however ConceptNet lacked any sense of the song-level semantic content. ConceptNet dealt directly with phrase-level semantics, and as a result the lyrics of several songs were misclassified.

The relation between the lyric data and the audio data can be difficult to accurately predict. In this case, the results from the lyric analysis are incorporated into the audio analysis results if there is an adequately high presence of specific mood in the lyrics. Madonna’s *Like A Virgin* contains elements of Depression, Excitement, and Distress; each comprising of roughly 30% the lyrical content. As such, these three moods contribute equally to the

<b>Song</b>	<b>Mood 1</b>	<b>Mood 2</b>	<b>Mood 3</b>	<b>Keywords</b>
Love Shack	Pleasure 0.371	Arousal 0.269	Depression 0.213	glitter, love, huggin', kissin', dancin', funky, groovin', jukebox money
One More Time	Pleasure 0.176	Arousal 0.164	Excitement 0.086	celebrate, dancing, feeling, free, tonight
Like A Virgin	Depression 0.383	Excitement 0.367	Distress 0.307	sad, blue, shiny, heart, love, fear, scared, cold
Dark Star	Arousal 0.246	Displeasure 0.222	Depression 0.222	dark, crashes, searchlight casting, shatters, flowers, velvet
Shine On You Crazy Diamond	Depression 0.308	Displeasure 0.218	Pleasure 0.140	black, night, crazy, cried, threatened, shadows
Glory Box	Depression 0.261	Arousal 0.219	Pleasure 0.165	tired, temptress, playing, love, flowers, bloom
Freddie's Dead	Displeasure 0.369	Depression 0.282	Distress 0.137	dead, terrible, junkie, misused, abused, misery
Orinoco Flow	Excitement 0.170	Pleasure 0.106	Arousal 0.000	sail away, waves, flow, beach, beyond, power, crash

Table 4.2: Results from the lyric analysis of eight songs

overall mood of the piece, but they are not influential enough to assign a new mood label to the song.

## 4.2 Vs. Music Classification Experts

Current music classification and recommendation tools were used as a reference point from which to gage the results obtained in this thesis work. All Music Guide (AMG) provides a large database of music classified by genre, style, mood, and theme. Each artist, album, or song is assigned several moods, which, when averaged, convey a general mood for the given item. Likewise, the AMG themes provide relevant contextual information for a song. Another source of contextual music data is Pandora Internet Radio. Their website contains a subset of features associated with a particular song. The Music Genome Project catalogued the selected features for this song, along with several hundred other attributes related to the song's content.

AMG's classification of Radiohead's Kid A album is shown in Table 4.3. Unfortunately, the

AMG classification values are freely available only for entire albums. Data pertaining to specific songs is not accessible from their public website, though it is possible to browse the top songs associated with a particular mood. Regardless, the AMG moods provide useful insight into the emotional qualities of an artist’s music. The moods presented in Table 4.3 are generally located in the emotional space of negative valence and negative arousal, which suggests that the Kid A album translates to Depression in Russell’s circumplex model of emotion. Accordingly, songs from the Kid A album classified using our system contained the moods of Displeasure, Depression, and Distress, as seen in Table 4.4.

The AMG themes can be used in conjunction with Pandora’s expert feature labeling to gain contextual information about a song. Several emotionally relevant features pertaining to Radiohead’s Kid A album contain keywords like abstract, heartbreaking, meandering, mellow, and ambient<sup>1</sup> (Table 4.4). These features and moods relate to the AMG themes, and together suggest that this album would work well as ambient background music for a quiet, solitary activity, such as reading or writing.

We now return to the eight songs presented in Table 4.1. A mood classification of these songs performed by our system is presented in Table 4.5. The results are compared with AMG’s mood classification and several emotionally relevant features from Pandora. Several of these features correlate with the mood label assigned to the songs. For example, Daft Punk’s *One More Time* is classified by both AMG and Pandora as ‘party’ music, which translates accordingly to Excitement. Similarly, Enya’s *Orinoco Flow*, classified by our system as relaxing, is a very soothing and meditative song with elements of ambient synths and breathy, unintelligible vocals.

A comparison of the mode extracted from the audio file and Pandora’s mode label in Table 4.6 reveals for the most part a similarity between the two. The only discrepancy that exists is with respect to *Love Shack* and *Dark Star*, whose modes lie near the boundary of major and minor tonality. *Love Shack*’s mode is, in this case, misclassified by our system, but manages to maintain its classification as a pleasurable song. As seen in Table 3.2, the emotion space of Pleasure is defined by a moderately major mode, which correlates with

---

<sup>1</sup><http://www.pandora.com/music/album/241c44e16b7af970>



Genre <sup>a</sup>	Styles	Moods	Themes
Rock	Indie Electronic Alternative Pop/ Rock Experimental Rock	Cold Suffocating Austere Atmospheric Bleak Nocturnal Cerebral Gloomy Complex Somber Ethereal Wintry Detached Clinical Light Insular Tense/Anxious Hypnotic Brooding Angst-Ridden	Late Night Solitude Background Music Feeling Blue Reflection The Creative Side Introspection

Table 4.3: All Music Guide’s classification of Radiohead’s album Kid A

<sup>a</sup><http://www.allmusic.com/cg/amg.dll?p=amg&sql=10:0ifyxq90ldae~T00>

Song	Mood (Audio)	Mood (Lyrics)	Pandora Features
Kid A	Displeasure	Displeasure	abstract lyrics unusual vocal sounds
How To Disappear Completely	Pleasure	Pleasure	mellow rock instrumentation
Treefingers	Distress	N/A	new age aesthetics use of ambient synth
Optimistic	Depression	Distress	meandering melodic phrasing mild rhythmic syncopation
Motion Picture Soundtrack	Distress	Arousal	heartbreaking lyrics

Table 4.4: Audio and lyric mood classification of songs from Radiohead’s Kid A album and their respective features from Pandora

Pandora’s ‘major key tonality’ classification.

As described in Chapter 2, researchers have had a difficult time defining rhythm. Rhythm is not a straightforward feature to evaluate as it contains a variety of subjective elements. Pandora’s rhythmic classification seen in Table 4.6 contains terms like ‘mild rhythmic syn-copation’ and ‘a light swing groove’. The corresponding rhythmic values extracted by our mood classification system show a connection with several of Pandora’s classifications. For example, *One More Time* has a low rhythmic value, suggestion irregularity, which matches Pandora’s ‘busy beats’ label. As well, *Like A Virgin* contains both the ‘prevalent use of groove’ and a highly regular rhythmic value.

### 4.3 Vs. Social Tagging Services

Last.FM, Qloud, and MyStrands provide services where users can add descriptive tags to artists, albums, and songs. With the rise of online social networks, user-generated web-content is a useful, though sometimes noisy, reference point. These social tags are representative of how someone feels when they hear a particular song.

Table 4.7 presents a comparison of a popular tagging vocabulary associated with Radio-head’s song, Idioteque. The tags were gathered from a number of social music services, including Last.FM, Qloud, and MyStrands. As Qloud and MyStrands are relatively new services, they contain fewer tags than their more established counterpart, Last.FM. Many of the tags found on the web relate to genres, which is a result of genre classification’s long-standing dominance over how music is presented to its audience. However, a vast amount of contextual and mood information is being generated with the rise of social music tagging services. This can be seen by several of the more descriptive tags from Last.FM, such as ‘Angry’, ‘Dark’, and ‘Atmospheric’, and from Qloud, including the very popular ‘Depressive’ tag (see Figures 4-2 and 4-3).

In regards to the eight songs listed above (Table 4.1), the tags related to the artists of these songs show a strong correlation with the songs’ respective moods (see Table 4.8).

<b>Song</b>	<b>Mood</b>	<b>Mood (AMG)</b>	<b>Pandora Features</b>
Love Shack	Pleasure	Fun	danceable grooves, humorous lyrics
One More Time	Excitement	Party	upbeat lyrics, lyrics about partying
Like A Virgin	Arousal	Provocative	house roots, new wave & disco influences, radio friendly stylings
Dark Star	Distress	Spacey	abstract lyrics, mellow rock instrumentation, folk influences, a dynamic male vocalist
Shine On You Crazy Diamond	Displeasure	Epic	string section beds, a prominent saxophone part, prominent organ
Glory Box	Depression	Gloomy	a laid back swing feel
Freddie's Dead	Sleepiness	Street-Smart	flat out funky grooves, funk roots
Orinoco Flow	Relaxation	Soothing	an overall meditative sound, new age influences, a breathy female lead vocalist, use of ambient synths, an unintelligible vocal delivery

Table 4.5: Our mood classification of eight songs versus AMG and Pandora

<b>Song</b>	<b>Mode</b>	<b>Pandora Mode</b>	<b>Rhythm</b>	<b>Pandora Rhythm</b>
Love Shack	48.5	major key tonality	62.3	mild rhythmic syncopation
One More Time	56.6	n/a	45.9	a bumpin' kick sound busy beats
Like A Virgin	47	n/a	68.9	prevalent use of groove
Dark Star	47.3	minor key tonality	35	mild rhythmic syncopation hard swingin' rhythm
Shine On You Crazy Diamond	35.6	minor key tonality	29.9	a twelve-eight time signature triple note feel
Glory Box	42.1	minor key tonality	40.8	a light swing groove prominent synth drums
Freddie's Dead	35.2	minor key tonality	44.3	mild rhythmic syncopation
Orinoco Flow	72.8	major key tonality	54.9	n/a

Table 4.6: Comparison of musical qualities of eight songs

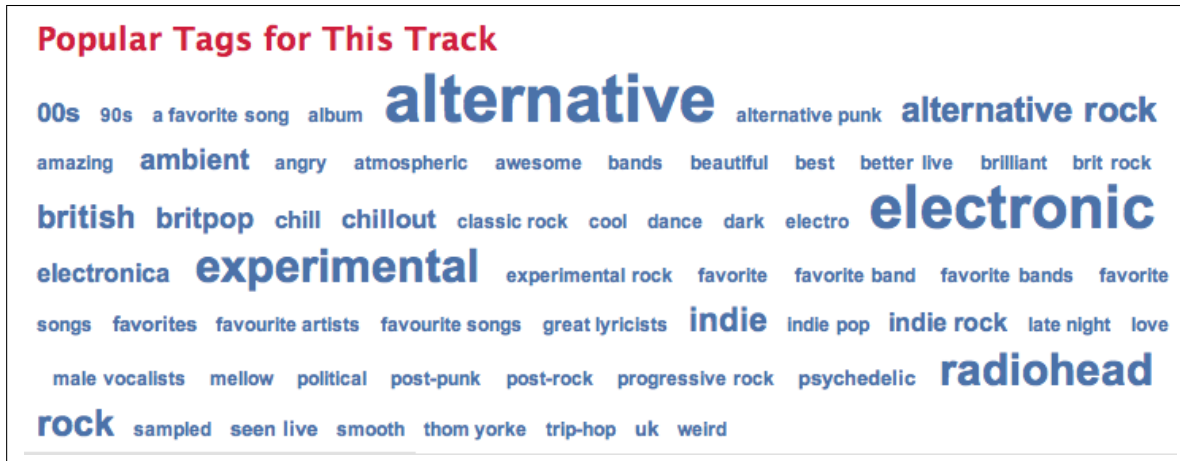


Figure 4-2: Popular Tags from Last.fm for Radiohead's Idioteque



Figure 4-3: Popular Tags from Qloud for Radiohead's Idioteque

<b>Last.FM<sup>a</sup></b> (Top 10)	<b>Last.FM</b> (Moods)	<b>Qloud<sup>b</sup></b>	<b>MyStrands<sup>c</sup></b>
Alternative	Angry	Psychedelic	Brit Pop
Electronic	Atmospheric	Depressive	Rock & Pop
Experimental	Beautiful	High Quality	
Radiohead	Brilliant	Live	
Rock	Chill	Alternative Rock	
Alternative Rock	Dark	Beat	
British	Mellow	British	
Britpop	Smooth	Epic	
Indie			
Ambient			

Table 4.7: Tagging of Radiohead’s Idioteque

<sup>a</sup>[http://www.last.fm/music/Radiohead/\\_/Idioteque/+tags](http://www.last.fm/music/Radiohead/_/Idioteque/+tags)

<sup>b</sup><http://www.qloud.com/track/Idioteque%20by%20Radiohead/1817056>

<sup>c</sup><http://www.mystrands.com/track/2135374/ref/12>

It appears in some cases that users have tagged songs using existing taxonomies, such as that of AMG, which can be seen in such tags as ‘street-smart’ for Curtis Mayfield and ‘atmospheric’ for Enya. For the most part these tags are relevant to both the mood and context of an artist’s music, though there is the odd tag that is unfitting (see Grateful Dead and ‘tasty’). Contextual information, such as Madonna’s ‘power walk’, ‘running’, and ‘workout’, indicate that her music is generally upbeat, energetic, and exciting. Likewise, the mood descriptors ‘ethereal’, ‘ambient’, and ‘meditative’ associated with Enya and her song, *Orinoco Flow*, match the mood label assigned by our system - Relaxation.

## 4.4 User Evaluation

A user study was conducted to evaluate the proposed system and to gain feedback on the usefulness of the classification tools. Twelve male and female students, between the ages of 22 and 33, took part in this survey. Their musical experience ranged from trained musician (8) to avid listener (9) to background listener (6). They were first asked to navigate the mood space and respond to specific questions and tasks, such as finding a particular song based on its mood. Afterwards, a questionnaire was provided to gather their comments on the overall experience.

Artist	Last.FM	Qloud	MyStrands
The B-52's	cheerful, energetic, fun, gleeful, happy, humorous, lively, mischievous, quirky, silly, whimsical, witty	dance, freaky, fun, party, party music	party
Daft Punk	ambient, chillout, dance, fun	dance, night, party, work out	booty shakin, dance, fun, party
Madonna	chillout, dance, energetic, ethereal, fun, melancholy, sexy, stylish	dance, feel good, power walk, running, workout	driving, erotica, groovy, happy, hypnotic, party, romance, sensual, touching, upbeat
Grateful Dead	chillout, psychedelic	folksy, glee, happy, mellow, party, road trip	psychedelic, tasty
Pink Floyd	ambient, chillout, dance, dreamy, psychedelic	chill, epic, pain, sad, sexy, sleep, soft, sorrow, travel, trippy	angry, confident, cry, eternal, gritty, inspiring, relaxing, rousing, searching, swaggering, theatrical, thinking
Portishead	ambient, chill, ethereal, mellow, smooth	groove, melancholic, sad	deep, eclectic, hypnotic, mellow, slow
Curtis Mayfield	aggressive, confrontational, empowering, exuberant, fiery, funky, menacing, plaintive, provocative, storm, raucous, street-smart	work out	blaxploitation, schmoozy, beatcake
Enya	atmospheric, calm, chill, circular, dreamy, ethereal, mellow, reserved, sexy, smooth	ambient, choral, new age, slow	beautiful, contemplative, driving, energy, heaven, mediative, rain, relaxing

Table 4.8: Social tagging of eight artists

All twelve participants agreed that they would like to be able to choose their music based on mood, and only two had experience with other mood-related music environments, notably Rhapsody and AMG. The majority currently use a combination of online radio stations, recommendations of friends, and mood to choose what music they would like to listen to. Others simply choose their music randomly or by musical feature, such as tempo. Eleven of the twelve participants prefer mood classification over genre and artist classification. The twelfth participant was unsure of the advantages of mood classification. There was a general consensus that mood classification is very subjective, but that it can be very useful for finding music that suits one's current state of mind.

With respect to the Mood Player interface, all twelve participants found that browsing by mood was useful, and most thought that the layout was intuitive. Suggestions for a more intuitive interface included a more graphical, exploratory interface. One user noted that it would be interesting to hide the artist, album, and song information because these names are already very emotionally charged and consequently can detract from the song's mood classification. The moods and the emotional model used by the Mood Player made sense to eight of the participants, while the remaining four had mixed feelings regarding the clarity of the model and gave suggestions for other moods that they would prefer, including sweet, unsettling, nervous, in love, cerebral vs. trance-inducing, wintry, complacent/satisfied, impatient, and distracted. Some found the mood labels misleading, and postulated that the system could be improved by changing some of the terms to less commonplace synonyms or by labeling the moods as 'exciting' instead of 'excitement', for example.

Nine people said they would like to use this mood classification system as a plug-in to their current digital music player, while the others found the novelty of the interface to be reason enough to use it as a standalone player. The participants suggested that this type of interface would be useful for such contexts as parties and background music while programming or working, and that it would also be useful as a tool to find new music.

The results of this user study reveal that there is a definite desire and need for mood classification and exploration tools that is not being addressed by current digital music players. The majority of music listeners use mood as their primary means of music selection, and presently must rely solely on personal knowledge of their music. However, with the proliferation of music compression algorithms, such as MP3, listeners are amassing copious amounts of music at a rate that prevents them from developing an intimate knowledge of their music collection. Therefore, tools such as the mood-based music classification and exploration system evaluated here will organize and present the listener's music using an intuitive and innovative approach, which will allow them to have a more enjoyable listening experience regardless of their understanding of the music.





## Chapter 5

# Conclusion

The research presented in this thesis is an attempt to design, implement, and evaluate a mood classification system for a music database. The goal of this work was to provide music listeners with an interface with which they could navigate their digital music collection in a less constrained manner. Most media players supply the user with a simple list of songs in the user's music library and the option to sort or search this list, and it is becoming increasingly difficult to quickly and easily find a set of songs that match a criteria set forth by the user, such as music to accompany a dinner party, a road trip, or an exercise schedule. The mood-based classification system is meant to enhance the music listening experience by removing the hassle of searching through a massive database for music that suits a specific context or mood. The motivation behind this work came from the lack of mood-based music navigation systems as well as the apparent shortcomings of today's music classification and recommendation systems described in Chapter 2.

A framework that is built on a model of human emotion was used to classify music by mood based on both its audio and lyrical content. James Russell's dimensional model of emotion was chosen for its congruence with music psychology and music classification, and Kate Hevner's studies in music and emotion were especially useful for correlating the musical features of a song with specific emotions. Our research also looked to current studies on music and emotion, including the work of Emery Schubert, Alf Gabrielsson, and Erik

Lindström, and we were able to incorporate their ideas to create a system that successfully classified music by mood.

The combination of audio and lyrical content is a crucial factor in our mood classification system. Many recent attempts at music mood classification have relied solely on audio content, as mentioned in Chapter 2; however, the inclusion of lyric analysis data was found to provide a great deal of contextual information that is integral to the classification of a song. It was possible to ascertain the particularities of a song through the lyric analysis, including meaning and context, which subsequently enabled the system to assign a more accurate mood label to a song.

Music is a complex subject to analyze as it contains a multitude of independent and dependent parameters. The five musical features employed by this system; mode, harmony, tempo, rhythm, and loudness, were chosen both because they convey emotional meaning in music (Cariani 2007) and for their relative ease of extraction from an audio file. However, a certain amount of information was lost as a result of this decision. Features such as pitch, melody, and timbre, which weren't included in this system, also provide information about a song's emotional characteristics. Though the mood classification system performed relatively well, the addition of these features would potentially improve the performance of the system.

One of the weaknesses of the classification system lies in its audio feature extraction stage. A misclassified musical feature, such as tempo or mode, will have a large impact on the mood classification of a song. A common issue that exists with tempo estimation of an audio file is the doubling or halving of the tempo value. This error is prevalent because it has not yet been determined how to extract the song's *feel* or *groove*. Consequently, if a song's tempo is classified as 160 bpm, while its true value is 80 bpm, then the song might be misclassified as Arousing instead of Pleasing. One solution to this problem is to define tempo ranges for each style or genre of music, where the upper range is less than the lower range multiplied by two.

The choice to use decision trees and k-nearest neighbor classification algorithms stemmed from the need to classify several different musical features into a set of eight emotion classes.

Supervised learning techniques were used because they enabled the classification system to create a model containing the eight classes after being trained on a set of songs. With a sufficiently trained model this system was able to accept and assign a class label (Pleasure, Excitement, etc.) to a new song that was added to the database. These two methods performed reasonably well and were easy to implement into the system for the dynamic classification of any song.

In conclusion, the resulting product is an innovative mood-based music classification and exploration system. The ultimate goal of this system was to provide the user with useful tools with which to browse and query their music library, and as such the interface provides a variety of novel playlist generation and search options. Playlists may be generated using any combination of natural language input, the presence of certain musical features (mode, harmony, etc.), or the song's overall mood. In addition, the clear and coherent presentation of mood-related information and the visualization of the individual music features of mode, harmony, tempo, rhythm, and loudness create a listening environment that is both interactive and informative. By incorporating this type of mood-based digital music player into one's everyday listening patterns the avid music listener is afforded the ability to discover new music, retrieve lost or forgotten music, and above all enhance and stimulate their overall listening experience.

## **5.1 Future Work**

### **5.1.1 Improvements**

As audio feature extraction techniques improve, the musical features extracted from the audio file will become more accurate. Melody extraction, for example, has seen several successful attempts resulting from MIREX's 2006 Audio Melody Extraction contest<sup>1</sup>. Progress in this area would improve the feature extraction portion of our current system, thus allowing for a wider range of musical parameters available to both the classification algorithms

---

<sup>1</sup>[http://www.music-ir.org/mirex2006/index.php/Audio\\_Melody\\_Extraction\\_Results](http://www.music-ir.org/mirex2006/index.php/Audio_Melody_Extraction_Results)

and to the user. The possibility of having more high-level features that can be extracted from an audio file will vastly improve classification in a system such as the one presented here.

Regression analysis with subjectively rated moods could be incorporated into the system enabling it to be more statistically objective. This would produce an unbiased calculation of Hevner's weightings for each musical feature. The use of a rating scale of emotion in place of the current eight mood categories could better suit the context of this work and would remove the circular constraint of the current emotional model. However, there needs to exist a compromise between an intuitive user interface and a comprehensive emotional model.

The decision tree and k-nearest neighbor classification algorithms used in this system provided moderate success. This aspect of the system could be improved either through further testing and modification of the current algorithms or through the use of alternative algorithms, such as Support Vector Machines or Neural Networks. A variety of different classification algorithms have been implemented in systems similar to this one and these systems could provide insightful information in choosing the best algorithm for this particular context.

Improvements to the user interface and tighter coupling of the external libraries and frameworks will help to make the system more robust and efficient. The system is currently implemented in Python using PyGTK graphics. A more elegant solution would be to port this interface to a native Mac OS X or Windows environment; however, this would diminish the system's cross-platform compatibility.

Currently, this system is designed for popular music classification. As well, the natural language processing tools that analyze the song lyrics are applicable only to the English language. The addition of cross-culturally relevant classification criteria and multi-lingual support would be a definite improvement for this system.

### 5.1.2 Applications

Potential applications of this work include the communication of personal emotions through greeting cards. Music classified by mood can be selected by the donor to match a desired sentiment to accompany a traditional or digital greeting card. When the recipient receives the greeting card they will be able to listen to music that conveys the mood conceived by the donor. Our mood classification system would be an ideal tool for one to classify their music in this context.

Alternatively, a possible evolution of this system would be to expand the system's capabilities beyond popular music to the areas of orchestral music, environmental sounds, speech, and video. Emotion detection in video is a particularly relevant progression of this work. Digital signal processing techniques are similar regardless of whether an audio signal or a video signal is used. Similarly, the analysis of a song's lyric metadata can easily be translated to the dialog or transcript of a video. Therefore, one could easily develop this work further for applications related to video.

Lastly, the information provided by this system would pair well with data from sites such as Last.FM, where the listening patterns of a user are recorded and analyzed. The addition of mood information to these patterns would allow one to visualize the emotional effects and evolution of their music and listening habits. This could likewise be applied to artists and bands as a way of visualizing the emotional content of their albums and songs.



# Appendix A

## User Evaluation

### Music Mood Classification Survey

Name: \_\_\_\_\_

Date: \_\_\_\_\_

Age: \_\_\_\_\_

Gender: \_\_\_\_\_

\_\_\_\_\_

How do you normally choose the music you listen to?

Would you like to be able to choose your music based on mood?

Have you used mood-related music software before? If yes, which software? What was your experience with this software?

Do you think that mood classification has an advantage over other types of music classification, such as genre, for example? If yes, which classifications and why?

Does the Mood Player perform as expected? Why or why not?

Is browsing by mood intuitive or useful?

Do the moods used in the Mood Player make sense?

Are there any other moods that you would use in place of these ones?

Would you use the Mood Player in place of your current digital music player (iTunes, Songbird, Windows Media Player, etc.)?

What contexts or situations would you use the Mood Player for?

Is there anything that you would change in the Mood Player?

What is your musical experience (trained musician, avid listener, background listener, etc.)?

Other comments?



# Bibliography

- [1] X. Amatriain. *An Object-Oriented Metamodel for Digital Signal Processing*. PhD thesis, Universitat Pompeu Fabra, 2004.
- [2] X. Amatriain, J. Massaguer, D. Garcia, and I. Mosquera. The clam annotaor: A cross-platform audio descriptors editing tool. In *Proceedings of the 6th International Conference on Music Information Retrieval*, pages 426–429, London, UK, 2005.
- [3] C. Anderson. The long tail. *Wired Magazine*, 12(10):170–177, 2004.
- [4] A. Andric and G. Haus. Automatic playlist generation based on tracking user’s listening habits. *Multimedia Tools and Applications*, V29(2):127–151, 2006.
- [5] S. Baumann and A. Klüter. Super-convenience for non-musicians: Querying mp3 and the semantic web. In *Proceedings of the 3rd International Conference on Music Information Retrieval*, Paris, France, 2002.
- [6] P. Cariani. Lecture notes for hst.725 music perception and cognition, Spring 2007.
- [7] N. Corthaut, S. Govaerts, and E. Duval. Moody tunes: The rockanango project. In *Proceedings of the 7th International Conference on Music Information Retrieval*, 2006.
- [8] S. J. Cunningham, D. Bainbridge, and A. Falconer. ‘more of an art than a science’: Supporting the creation of playlists and mixes. In *Proceedings of the 7th International Conference on Music Information Retrieval*, Victoria, Canada, 2006.
- [9] S. J. Cunningham, M. Jones, and S. Jones. Organizing digital music for use: An examination of personal music collections. In *Proceedings of the 5th International Conference on Music Information Retrieval*, pages 447–454, Barcelona, Spain, 2004.
- [10] S. Dixon. A lightweight multi-agent musical beat tracking system. In *Proceedings of the Pacific Rim International Conference on Artificial Intelligence*, pages 778–788, Melbourne, Australia, 2000.
- [11] P. Ekman. An argument for basic emotions. *Cognition & Emotion*, 6(3/4):169–200, 1992.
- [12] P. R. Farnsworth. A study of the hevner adjective list. *The Journal of Aesthetics and Art Criticism*, 13(1):97–103, 1954.

- [13] P. R. Farnsworth. *The social psychology of music*. The Dryden Press, 1958.
- [14] P. R. Farnsworth. *The social psychology of music*. Iowa State University Press, Ames, Iowa, 2nd edition, 1969.
- [15] B. Fehr and J. A. Russell. Concept of emotion viewed from a prototype perspective. *Journal of Experimental Psychology: General*, 113(3):464–486, 1984.
- [16] Y. Feng, Y. Zhuang, and Y. Pan. Music information retrieval by detecting mood via computational media aesthetics. In *Proceedings of the IEEE/WIC International Conference on Web Intelligence*, page 235, Washington, USA, 2003.
- [17] A. Gabrielsson and E. Lindström. *Music and Emotion: Theory and Research*, chapter The Influence of Musical Structure on Emotional Expression, pages 223–248. Oxford University Press, 2001.
- [18] R. H. Gundlach. Factors determining the characterization of musical phrases. *The American Journal of Psychology*, 47(4):624–643, 1935.
- [19] K. Hevner. The affective character of the major and minor modes in music. *The American Journal of Psychology*, 47(1):103–118, 1935.
- [20] K. Hevner. Expression in music: A discussion of experimental studies and theories. *Psychological Review*, 42:186–204, 1935.
- [21] K. Hevner. Experimental studies of the elements of expression in music. *The American Journal of Psychology*, 48(2):246–268, 1936.
- [22] K. Hevner. The affective value of pitch and tempo in music. *The American Journal of Psychology*, 49(4):621–630, 1937.
- [23] T. Jehan. *Creating Music by Listening*. PhD thesis, Massachusetts Institute of Technology, 2005.
- [24] P. N. Juslin. Cue utilization in communication of emotion in music performance: relating performance to perception. *Journal of Experimental Psychology: Human Perception and Performance*, 26(6):1797–1813, 2000.
- [25] P. R. Kleinginna and A. M. Kleinginna. A categorized list of emotion definitions, with a suggestion for a consensual definition. *Motivation and Emotion*, 5(4):345–79, 1981.
- [26] M. Leman, V. Vermeulen, L. D. Voogdt, and D. Moelants. Using audio features to model the affective response to music. In *Proceedings of the International Symposium on Musical Acoustics*, Nara, Japan, 2004.
- [27] T. Li and M. Ogihara. Detecting emotion in music. In *Proceedings of the 4th International Conference on Music Information Retrieval*, Baltimore, USA, 2003.
- [28] T. Li and M. Ogihara. Content-based music similarity search and emotion detection. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 5, pages 705–708, 2004.

- [29] H. Liu, H. Lieberman, and T. Selker. A model of textual affect sensing using real-world knowledge. In *Proceedings of the 8th International Conference on Intelligent User Interfaces*, pages 125–132, Miami, USA, 2003. ACM Press.
- [30] H. Liu and P. Singh. Conceptnet: a practical commonsense reasoning toolkit. *BT Technology Journal*, 22(4):211–226, 2004.
- [31] B. Logan, A. Kositsky, and P. Moreno. Semantic analysis of song lyrics. In *IEEE International Conference on Multimedia and Expo*, volume 2, pages 827–830, 2004.
- [32] L. Lu, D. Liu, and H.-J. Zhang. Automatic mood detection and tracking of music audio signals. *IEEE Transactions on Audio, Speech and Language Processing*, 14(1):5–18, 2006.
- [33] M. I. Mandel, G. E. Poliner, and D. P. W. Ellis. Support vector machine active learning for music retrieval. *Multimedia Systems*, 12(1):3–13, 2006.
- [34] E. McKean, editor. *New Oxford American Dictionary*. Oxford University Press, New York, 2nd edition, 2005.
- [35] A. Mehrabian. Pleasure-arousal-dominance: A general framework for describing and measuring individual. *Current Psychology*, 14(4):261–292, 1996.
- [36] O. Meyers. Lyricator: An emotional indicator of lyrics. <http://www.media.mit.edu/~meyers/lyricator.php>, 2005.
- [37] O. Meyers. Mysoundtrack: A commonsense playlist generator. <http://www.media.mit.edu/~meyers/mysoundtrack.pdf>, 2006.
- [38] M. Minsky. *The Society Of Mind*. Simon and Schuster, New York, 1986.
- [39] G. Peeters and X. Rodet. Automatically selecting signal descriptors for sound classification. In *Proceedings of the International Computer Music Conference*, Göteborg, Sweden, 2002.
- [40] T. Pohle, E. Pampalk, and G. Widmer. Evaluation of frequently used audio features for classification of music into perceptual categories. In *Proceedings of the 4th International Workshop on Content-Based Multimedia Indexing*, Riga, Latvia, 2005.
- [41] M. G. Rigg. *What Features of a Musical Phrase Have Emotional Suggestiveness?*, volume 36 of *Bulletin of the Oklahoma Agricultural and Mechanical College*. Oklahoma Agricultural and Mechanical College, Stillwater, Oklahoma, 1939.
- [42] M. G. Rigg. The mood effects of music: A comparison of data from four investigators. *The Journal of Psychology*, 58:427–438, 1964.
- [43] J. A. Russell. Affective space is bipolar. *Journal of Personality and Social Psychology*, 37(3):345–356, 1979.
- [44] J. A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39(6):1161–1178, 1980.

- [45] E. Schubert. Update of the hevner adjective checklist. *Perceptual and Motor Skills*, 96:1117–1122, 2003.
- [46] E. Schubert. Modeling perceived emotion with continuous musical features. *Music Perception*, 21(4):561–585, 2004.
- [47] P. Singh, T. Lin, E. T. Mueller, G. Lim, T. Perkins, and W. L. Zhu. *On the Move to Meaningful Internet Systems 2002: DOA/CoopIS/ODBASE 2002*, volume 2519 of *Lecture Notes in Computer Science*, chapter Open Mind Common Sense: Knowledge Acquisition from the General Public, pages 1223–1237. Springer-Verlag, Heidelberg, 2002.
- [48] J. Skowronek, M. F. McKinney, and S. van de Par. Ground truth for automatic music mood classification. In *Proceedings of the 7th International Conference on Music Information Retrieval*, Victoria, Canada, 2006.
- [49] J. A. Sloboda and P. N. Juslin. *Music and Emotion: Theory and Research*, chapter Psychological Perspectives on Music and Emotion, pages 71–104. Oxford University Press, 2001.
- [50] A. Tellegen, D. Watson, and L. A. Clark. On the dimensional and hierarchical structure of affect. *Psychological Science*, 10(4):297–303, 1999.
- [51] R. E. Thayer. *The Biopsychology of Mood and Arousal*. Oxford University Press, Oxford, 1989.
- [52] R. E. Thayer. *The origin of everyday moods: managing energy, tension, and stress*. Oxford University Press, New York, 1996.
- [53] M. Tolos, R. Tato, and T. Kemp. Mood-based navigation through large collections of musical data. In *Consumer Communications and Networking Conference*, pages 71–75, Las Vegas, USA, 2005.
- [54] G. Tzanetakis and P. Cook. Marsyas: a framework for audio analysis. *Organised Sound*, 4(3):169–175, 1999.
- [55] G. Tzanetakis and P. Cook. Music analysis and retrieval systems for audio signals. *Journal of the American Society for Information Science and Technology*, 55(12):1077–1083, 2004.
- [56] G. S. Vercoe. Moodtrack: practical methods for assembling emotion-driven music. Master’s thesis, Massachusetts Institute of Technology, 2006.
- [57] M. Wang, N. Zhang, and H. Zhu. User-adaptive music emotion recognition. In *Proceedings of the International Conference on Signal Processing*, Istanbul, Turkey, 2004.
- [58] K. B. Watson. The nature and measurement of musical meanings. In *Psychological Monographs*, volume 54, pages 1–43. The American Psychological Association, Evanston, IL, 1942.

- [59] C. M. Whissell. *Emotion: Theory, Research, and Experience*, volume 4, chapter The Dictionary of Affect in Language, pages 113–131. Academic Press, New York, 1989.
- [60] B. A. Whitman. *Learning the Meaning of Music*. PhD thesis, Massachusetts Institute of Technology, 2005.
- [61] A. Wiczorkowska, P. Synak, R. Lewis, and Z. Ras. Extracting emotions from music data. In *Proceedings of the 15th International Symposium on Methodologies for Intelligent Systems*, Saratoga Springs, USA, 2005.
- [62] D. Yang and W. Lee. Disambiguating music emotion using software agents. In *Proceedings of the 5th International Conference on Music Information Retrieval*, Barcelona, Spain, 2004.
- [63] T. Zhang. Semi-automatic approach for music classification. In *Proceedings of the Conference on Internet Multimedia Management Systems*, Orlando, USA, 2003.