

## ABSTRACT

Title of Thesis: ON THE INCORPORATION OF  
PSYCHOLOGICALLY-DRIVEN ‘MUSIC’  
PREFERENCE MODELS FOR MUSIC  
RECOMMENDATION

Monique Kathryn Dalton, Ethan J. Ferraro,  
Meg Galuardi, Michael L. Robinson,  
Abigail M. Stauffer, Mackenzie Thomas Walls

Thesis Directed By: Professor Ramani Duraiswami,  
Department of Computer Science

There are hundreds of millions of songs available to the public, necessitating the use of music recommendation systems to discover new music. Currently, such systems account for only the quantitative musical elements of songs, failing to consider aspects of human perception of music and alienating the listener’s individual preferences from recommendations. Our research investigated the relationships between perceptual elements of music, represented by the MUSIC model, with computational musical features generated through *The Echo Nest*, to determine how a psychological representation of music preference can be incorporated into recommendation systems to embody an individual’s music preferences. Our resultant model facilitates computation of MUSIC factors using *The Echo Nest* features, and can potentially be integrated into recommendation systems for improved performance.

ON THE INCORPORATION OF PSYCHOLOGICALLY-DRIVEN 'MUSIC'  
PREFERENCE MODELS FOR MUSIC RECOMMENDATION

by

Monique Kathryn Dalton  
Ethan J. Ferraro  
Meg Galuardi  
Michael L. Robinson  
Abigail M. Stauffer  
Mackenzie Thomas Walls

Thesis submitted in partial fulfillment of the requirements of the Gemstone Program  
University of Maryland, College Park  
2016

Advisory Committee:  
Professor Ramani Duraiswami, Chair  
Dr. Daniel J. Levitin  
Dr. L. Robert Slevc  
Dr. Shihab Shamma  
Dr. Carol Espy-Wilson

© Copyright by

Monique Kathryn Dalton, Ethan J. Ferraro, Meg Galuardi, Michael L. Robinson,  
Abigail M. Stauffer, Mackenzie Thomas Walls  
2016

## Preface

This thesis is the culmination of three years of research by Team MUSIC of the Gemstone Program in the University of Maryland Honors College. We are a diverse group of undergraduate students, ranging from majors of Electrical Engineering to Neurobiology and Physiology. All of our members enjoy listening to and discovering new music, as was the inspiration for joining this project for some members, and most of us play some type of musical instrument. The past few years have been both a rewarding and challenging experience, as we originally started with twelve members, and over the years found ourselves as a small, tightly knit group of six. We have had the wonderful opportunity to present our research at the Summer 2015 Society for Music Perception and Cognition Conference, and the pleasure of meeting and communicating with a few experts in the field of music information retrieval along the way.

## Acknowledgements

Team M.U.S.I.C. would like to thank everyone who made our unique undergraduate research experience possible, most notably the outstanding Gemstone staff, including Dr. Kristan Skendall, Dr. Frank Coale, and Vickie Hill. We would also like to recognize those who have provided invaluable help to us along this intensive research process: our mentor, Professor Ramani Duraiswami; our librarian, Steve Henry; as well as researchers Dr. David Meichle, Dr. Bob Slevc, Dr. Mehmet Vurkac, Dr. Jason Rentfrow, and Dr. Daniel Levitin. Thank you to those who funded our team to attend and present at the 2015 Meeting of the Society for Music Perception and Cognition. Finally, we would like to thank all of the discussants that constitute our advisory board.

## Table of Contents

Preface.....	ii
Acknowledgements.....	iii
List of Tables .....	vi
List of Figures.....	vii
Chapter 1: Introduction.....	1
1.1 Overview of Research.....	1
1.2 Overview of Current Music Recommendation Systems.....	2
1.3 Early Stages of Research .....	3
1.4 General Research Questions .....	5
1.5 Purpose and Rationale of the Study .....	6
1.6 Method Framework.....	7
1.6.1 Study 1 – Focus Groups on Music Recommendation.....	7
1.6.2 Study 2 – Online Survey for Music Preference .....	7
1.6.3 Study 3 – Modeling the MUSIC Factors with Echo Nest Features .....	8
1.6.4 Study 4 – Model Evaluation .....	8
Chapter 2: Literature Review.....	9
2.1 Automated Music Recommendation.....	9
2.1.1 <i>Expert Analysis</i> : Pandora Internet Radio .....	9
2.1.2 <i>User Trend Approach</i> : iTunes Genius .....	11
2.1.3 Quantitative Approach: Spotify .....	12
2.2 The MUSIC Model .....	13
2.2.1 MUSIC Model Overview.....	13
2.2.2 MUSIC Model History .....	13
2.2.3 The Five Factors of the MUSIC Model .....	18
2.2.4 Model Development.....	19
2.2.5 Model Versatility .....	21
2.2.6 Application of the MUSIC Model to Music Recommendation .....	22
2.3 The Echo Nest.....	24
Chapter 3: Study 1: Focus Groups on Music Recommendation.....	26
3.1 Methodology .....	26
3.1.1 Motivation.....	26
3.1.2 Participant Enrollment .....	27
3.1.3 Focus Group Operation.....	28
3.1.4 Focus Group Structure .....	29
3.1.5 Data Analysis .....	32
3.2 Results.....	33
3.2.1 General/Demographic Information.....	33
3.2.2 Attitudes towards Current Recommender Systems .....	34
3.2.3 Prominent Features Noted in a First Listen .....	36
3.3 Discussion .....	37
Chapter 4: Study 2: Survey on Music Preference.....	41
4.1 Methodology .....	41
4.1.1 Motivation.....	41
4.1.2 Survey Structure.....	41

4.1.3	Participant Enrollment .....	42
4.1.4	Song Selection .....	43
4.1.5	Data Analysis .....	45
4.2	Results.....	45
4.3	Discussion.....	52
Chapter 5:	Studies 3 and 4: Model Development and Evaluation .....	55
5.1	Methodology .....	55
5.1.1	Motivation.....	55
5.1.2	Data Sets .....	55
5.1.3	Weka Machine Learning Environment.....	59
5.1.4	Approach to Machine Learning Analysis .....	60
5.2	Results.....	62
5.2.1	Model Development Approach.....	62
5.2.2	Classifier Determination .....	63
5.2.3	Study 4: Model Evaluation .....	68
5.3	Discussion.....	70
Chapter 6:	Conclusion.....	74
6.1	Future Research .....	76
References	.....	78
Appendices	.....	83
Appendix A:	Study 1 Supplemental Information .....	83
A1:	Focus Group Flyer Posting Locations.....	83
A2:	Study 1 Consent Form .....	84
A3:	Study 1 Demographic Information .....	86
Appendix B:	Study 2 Supplemental Information.....	87
B1:	Survey Script.....	87
B2:	Study 2 Genre Preferences.....	92
Appendix C:	Studies 3 and 4 Supplemental Information .....	93
C1:	Song List Used for Machine Learning Training .....	93
C2:	Additional Classifier Predictions .....	97

## List of Tables

Table 1. Summary of current music recommendation system methods .....	2
Table 2. Description of the factors present in the MUSIC model .....	19
Table 3. Songs featured in survey.....	43
Table 4. Description of The Echo Nest features.....	58
Table 5. M factor prediction errors.....	64
Table 6. U factor prediction errors.....	64
Table 7. S factor prediction errors .....	65
Table 8. I factor prediction errors .....	65
Table 9. C factor prediction errors.....	66
Table 10. MUSIC factor Isotonic Regression predictions .....	67
Table 11. Correlations of MUSIC factors to The Echo Nest features .....	72
Table 12. Evaluation of reproduction of MUSIC model values .....	73
Table A1. Study 1 Demographics: Gender.....	86
Table A2. Study 1 Demographics: Ethnicity .....	86
Table A3. Study 1 Demographics: Genre Preference.....	86
Table C1. Songs used in machine learning model development .....	93
Table C2. MUSIC Factor Gaussian Processes Predictions.....	97



## List of Figures

Figure 1: Simplified methodology used for the development of the MUSIC model..	20
Figure 2: Simplified methodology for data extraction from <i>The Echo Nest</i> .....	25
Figure 3: Prevalence and opinions on sources of new music .....	35
Figure 4: Quotes from focus groups regarding recommendation systems .....	36
Figure 5. Distribution of model performances (coefficient of determination) .....	50
Figure 6. Evaluation of direct participant models (p-values) .....	51
Figure 7. Coefficient p-values for the MUSIC variables.....	52
Figure 8. MUSIC factor linear regression models .....	68
Figure 9: Comparison of percentage of songs classified correctly .....	69
Figure A1: Focus group flyer posting locations .....	83
Figure B1. Genre preferences specified by Study 2 participants.....	92

# Chapter 1: Introduction

## *1.1 Overview of Research*

Music recommendation services are continually developing in response to the overwhelming library of music available, as it has become increasingly challenging for people to discover relevant music with an average of over 75,000 albums released every year in the U.S. (Peoples, 2011). While innovative ways to classify and organize the rapidly expanding body of songs have been applied by recommendation systems such as Spotify, Pandora, and iTunes, these systems fail to consistently predict songs that ideally represent user preference (Barrington, Oda, & Lanckriet 2009). It has been suggested that current recommendation systems are missing a vital component – one which may ultimately produce more accurate recommendations: a connection between the intrinsic qualities of the music, and how listeners perceive their music (Laplante, 2014). For the purposes of our research, we refer to the intrinsic qualities of music as any musical attribute, such as tempo, key signature, and instrumentation, which characterizes an individual song. Spotify and Pandora have generated characteristic analytical approaches to modeling songs, while iTunes integrated user-trend analysis in an attempt to predict user preference. Yet, as far as our review of literature has uncovered, there has been little to no unification of these bodies of data – preferential and analytical – to produce recommendations that are not only based in the inherent qualities of the music, but additionally based upon how the user perceives each song.

## 1.2 Overview of Current Music Recommendation Systems

In general, current recommendation systems rely upon data mining of homogenous data sets of user trends or quantitative music models. By relying primarily on only one type of data, these systems fail to account for the methods in which listeners perceive the different combinations of features prominent in a song. While the amount of active users of these systems is substantial, reaching tens of millions in the case of Pandora and Spotify (Adegoke, 2014; “Spotify vs. Pandora: Number of active users,” 2014), the recommendation accuracy can still be improved upon in various aspects. The overall performance of these systems may be inherently limited by their “one-sided” data set, as illustrated below in Table 1, and discussed in further detail in the literature review section (Section 2). Our focus group study results (Section 3) also provide examples of user attitudes toward the performance of current music recommendation systems.

**Table 1.** Summary of current music recommendation system methods

<i>Recommendation System</i>	<i>Approach</i>	<i>Recommender Details</i>	<i>Deficiencies</i>
Pandora Internet Radio	Expert Analysis	Songs modeled by 450 attributes evaluated by professional musicologists	<ul style="list-style-type: none"> <li>• Inefficient for large collections of songs</li> <li>• Human bias</li> <li>• Cultural bias</li> </ul>
Spotify	Quantitative Analysis	Signal processing method to automatically generate attribute vector for each song	<ul style="list-style-type: none"> <li>• Relies heavily on intrinsic song similarity</li> <li>• No known interaction with music perception</li> </ul>
iTunes Genius	User Trends	Songs are related using general trend of overall user preferences	<ul style="list-style-type: none"> <li>• Neglects intrinsic song information</li> <li>• Generalizes user preferences</li> </ul>

### ***1.3 Early Stages of Research***

In the initial stages of this research, we desired to create a music recommendation system or model that would be based on musical features. We originally proposed the development of a model in which the feature metrics were generated through signal processing and analysis of the songs. Through signal processing, we would generate musically intrinsic features as a basis for more objective classification of songs (as compared to human designation of values). The resultant feature analysis would be representative of the quantitative audio signal of each song. Additionally, we proposed that an automated signal processing system would be more effective, as song recommendations could be readily produced without requiring human evaluation to assign values to each feature, as done in Pandora's Music Genome Project ("About the Music Genome Project," 2014).

In regards to extracting music features, we had planned to use signal processing toolkits in MATLAB, such as the MIR (Music Information Retrieval) Toolbox (Lartillot & Toivainen, 2007), to gather information regarding song features. Of the preprocessing methods we researched, we discovered that a common processing technique involved the transformation of the audio signal in its frequency domain using the Mel Scale. This scale is a logarithmic transformation of an audio signal to account for the timbral ranges perceived by the human ear (Seyerlehner, 2010). We attempted to acquire this transformation by computing Mel Frequency Cepstral Coefficient (MFCC) features in MATLAB (Tzanetakis & Cook 2002; Seyerlehner, Widmer, & Pohle 2008). These features were successful in some cases, but ultimately they proved insufficiently discriminative to be useful for our purposes.

In addition to these features, we investigated the use of features based on Block Level and Frame Level processing of audio signals and their spectra. Frame Level analysis takes millisecond fragments of the audio signal and pulls out characteristic information that describes musical features of those small sections, while Block level analysis groups a collection of frames together to produce a vector for different sections of the song, each describing a collection of features (Seyerlehner, 2010). We attempted this methodology through use of the sample LabRosa MATLAB functions related to MFCC generation and visual representation of speech processing provided by Dan Ellis on his MATLAB Audio Processing Examples webpage (<http://labrosa.ee.columbia.edu/matlab>). Using these examples as a learning basis, we planned to develop a global feature vector for each song, which would be representative of the song's signature. In order to generate song recommendations, we would compare feature vectors between different songs through a machine learning program.

After a period in which our attempts to perform these signal processing methods were ineffective, we discovered that a large portion of musical signal processing had already been performed by *The Echo Nest* ("Spotify Echo Nest API"), whose features were obtainable via an open source API (application programming interface). The extensive database of this company had a large number of features readily available, such as valence, loudness, and danceability – and many of these features have already been produced for an immense number of songs ("Acoustic Attributes Overview"). Various programs such as Spotify use *The Echo Nest* feature data to power their recommendation engines. After discovering this useful tool, we

concluded that the previous methodology was too broad for us to feasibly accomplish within the time we had left to complete our research. We ultimately decided to focus on incorporating predetermined features from *The Echo Nest* into our approach that would use information regarding music perception, and analyzing how they could be used more effectively to produce quality music recommendations.

#### **1.4 General Research Questions**

Our shift in research direction and scope led us to new research questions that were more suitable for our team size and timeframe. Our new research questions became the following:

- 1. How can music perception be incorporated into an automated recommendation system?*
- 2. How will such a system effectively and accurately embody an individual's music preferences?*

Currently, music recommendation methods have a tendency to account only for the quantitative musical elements of songs, metadata, and overarching listener trends. This is ultimately causing music recommendation to become an indistinctive, overgeneralized procedure, with a noticeable absence of analysis toward individual perception of music. Many state-of-the-art research studies in the field of music information retrieval (Laplante, 2014; Rentfrow, Goldberg, & Levitin, 2011; Soleymani, Aljanaki, Wiering, & Veltkamp, 2015) have demonstrated that music perception is an integral aspect for formation of music preferences. We decided to conduct our research such that we could investigate how to fill this gap in music recommendation.

Based on our new research direction, we formulated the hypothesis that an integration of an effectual model of music perception for preference into existing systems will lead to an improved music recommendation system. As we looked for models that could represent perception and cognition of music in humans, we discovered a well-regarded, highly cited model: the MUSIC Model, presented by Rentfrow, Goldberg, & Levitin (2011). This model describes a five-factor layout of the structure of individual musical preference, based on the more psychological qualities of music perception. In addition to analyzing the applicability of this model, we aimed to investigate how this model could be used to develop an automated, simplified five-factor model to be implemented in song recommendation for a more individualized experience.

### ***1.5 Purpose and Rationale of the Study***

Our research presents the combination of musically intrinsic features of songs with the five-factor MUSIC model for music preference proposed by Rentfrow et al. (2011). Similar models have been developed through multiple international studies to show that music preferences can be depicted through simplified systems of only a few factors (George, Stickle, Faith, & Wopnford, 2007; Desling, Bogt, Engels, & Meeus, 2008; Schafer & Sedlmeier, 2009; Brown, 2012; Langmeyer, Guglhör-Rudan, & Tarnai, 2012). While there have been discrepancies in the opinions of researchers on the exact number of factors to include to develop an appropriate model, they tend to gravitate towards models of four to five-factors, like that of the MUSIC model. This model has only been derived from principal component analysis of participant data on song preference. While the MUSIC model may be an effective way to predict user

preference, it is not currently developed in a way that is efficient enough to be applied to millions of song recordings. The music features developed by *The Echo Nest* have been shown to accurately represent the acoustic attributes of a song (Bertin-Mahieux, Ellis, Whitman, & Lamere, 2011; Schindler & Raubner, 2014) in a manner that is efficient and objective. We therefore propose a means of producing MUSIC model factor values from *The Echo Nest* features, to produce a model that is simple, efficient, informed by research on human music preferences, and hence widely-applicable.

## **1.6 Method Framework**

### 1.6.1 Study 1 – Focus Groups on Music Recommendation

Five focus group sessions were conducted with students from the University of Maryland at College Park to gain insight on the level of satisfaction with current recommendation systems. We additionally asked our participants to listen to and analyze thirty-second song segments chosen from ten different songs. They identified song characteristics that they found prominent when listening to the selections, and also described how they felt about each piece. This provided insight into what potential features users might find important when listening to a piece of music and evaluating the performance of a music recommendation system.

### 1.6.2 Study 2 – Online Survey for Music Preference

To evaluate the MUSIC model's ability to predict song preference, 100 participants from across the United States completed an online survey powered by the Qualtrics Online Survey Software (Qualtrics Survey Tool). Those surveyed were



tasked primarily with listening to one-minute segments from twenty songs previously analyzed by the MUSIC model (Rentfrow et al., 2011). A strong measure of correlation between the previously reported MUSIC factors and the preference ratings over a majority of the users would support the application of this model to a personalized music recommendation system.

### 1.6.3 Study 3 – Modeling the MUSIC Factors with Echo Nest Features

The Waikato Environment for Knowledge Analysis (Weka) (Hall et al., 2009) machine learning software suite was used to develop evaluate different classifiers' modeling capabilities for determining a song's MUSIC model from its *The Echo Nest* features. Sequential analysis of classifiers provided by Weka, such as Gaussian Processes and Isotonic Regression, were applied to unify *The Echo Nest* features and MUSIC factor values for all songs from Rentfrow et al. (2011, 2012) found in *The Echo Nest* library, excluding those used in Study 2 for testing purposes.

### 1.6.4 Study 4 – Model Evaluation

Once an appropriate model was determined, it was applied to fifteen of the twenty songs (as not all songs were found in *The Echo Nest* database) from Study 2 to experimentally determine their MUSIC factor loadings. These new values were applied to the preference data from the online survey via the same correlation metrics to evaluate the applicability of the MUSIC values determined from our model to personalized music recommendation.

## Chapter 2: Literature Review

### *2.1 Automated Music Recommendation*

Our research targets an area of music recommendation that has been largely neglected in commercial use: individual music perception. The systems in popular use are usually adequate for forming basic recommendations, and are capable of sorting songs into broad categories. However, a majority of users believe that there are areas for improvement and refinement of the recommendation process, as evident from an analysis of qualitative studies and commercial behavior of such programs. By understanding how these recommendation systems function, it is easier to recognize their limitations and determine why a more rigorous recommendation method – such as the one we propose in this thesis – is necessary. In this Section (2.1), we will review the current state of commercialized recommendation systems Pandora, Spotify, and iTunes Genius.

#### 2.1.1 Expert Analysis: Pandora Internet Radio

Pandora Internet Radio (Pandora) is one of the most prolific and enduring examples of a commercialized music recommendation system (Celma, 2010), boasting a user base of 81.5 million active users ("Spotify vs. Pandora: Number of active users," 2014). Pandora has a thorough, ambitious approach to song recommendation, known as the “Music Genome Project”. The Music Genome Project was an attempt to model songs by a defined vector of approximately 450 attributes, each of which are assigned by human “professional musicologists” (Barrington et al.,

2009). These experts are responsible for defining the descriptors for each song in Pandora's database by assigning values to these attributes in an effort to envelope the song's acoustic content. Attributes range from major and minor tonality and the degree of syncopation within a song, to the gender of the vocalist (Barrington et al., 2009).

While the impressive number of features within Pandora's model suggests some level of versatility and reliability, the Music Genome Project-powered method faces inherent problems. The amount of time and effort required for analysts to rate a song's attributes is fundamentally too intensive to sustain a constantly expanding song library, thus rendering the method infeasible in regards to automated recommendation within the scope of hundreds of millions of songs. One comparative music recommendation system research study (Magno & Sable, 2008) noted the "slow rate of content edition to the Pandora database."

Furthermore, this method introduces an element of human bias through the input of musicologists, as people tend to perceive music uniquely, especially groups from different cultures. Music enculturation has been documented by experts in the field of music perception as a phenomenon that affects scale pitches and note patterns (Morrison & Demorest, 2009). Music from a different culture has been shown to evoke a response different to music from a listener's personal culture; this has specifically been cited in cases with Western listeners (Morrison & Demorest, 2009). By relying on musical experts to rate songs, Pandora introduces a level of subjectivity which may include cultural and human bias, thus reducing the reliability of its recommendations to an individual user.

Pandora's Music Genome Project is an example of a content-based filtering algorithm (Bogdanov 2011), in which a song is described using self-contained, intrinsic characteristics, rather than metadata about the song (such as artist, album, or genre). In contrast to a content-based filtering system, Apple's iTunes Genius makes use of collaborative filtering, wherein its recommendations are based upon the related song purchases of other consumers (Cremonesi, Garzotto, Negro, Papadopoulos, & Turrin, 2011).

### 2.1.2 User Trend Approach: iTunes Genius

For over 10 years, Apple's iTunes has been available to the public as a media player and library with a digital media store, the iTunes Store. This provides Apple with a massive database of buyer trends from years of users downloading millions of songs and albums, in addition to other media. This metadata provides the backbone for iTunes' recommendation service, iTunes Genius (Barrington et al., 2009). iTunes uses Gracenote's MusicID to recognize songs, artists, and albums from a user's music library. This information combined with the user's song ratings are compared with the metadata Genius has collected to generate playlists from the provided library and recommend new song purchases (Barrington et al., 2009). Genius therefore relies on little to no content-based information, and instead relies on the mindset that general buying trends should coincide with each user's preferences. This type of collaborative filtering method has a tendency to keep lesser-known artists in obscurity, as the recommendations are based on the more popular artists and albums that are bought by users (Magno & Sable, 2008; Barrington, Oda, & Lanckriet, 2009).

### 2.1.3 Quantitative Approach: Spotify

Spotify expanded its function in the music industry from primarily a music streaming site to offering user-specific recommendations upon acquiring *The Echo Nest* in 2014 (“Spotify acquires *The Echo Nest*,” 2014). *The Echo Nest* is a platform that utilizes signal processing methods to generate an understanding about songs as characterized by a set of multiple features that can be utilized by companies or programs through an application program interface (API). This API is marketed for use in music personalization, display of dynamic information about musical artists, audio fingerprinting for recognition purposes, and for applications in interactive or remixed music (“We Know Music...”). As of March 2016, *The Echo Nest* has analyzed over 37 million songs characterized by upwards of 1.2 trillion data points (“We Know Music...”). Spotify accesses from *The Echo Nest* song data including standard attributes such as tempo, key, time signature; self-developed characteristics such as energy and loudness; and artist data such as biographies, news stories, and similar artists (“Spotify Echo Nest API”). Spotify combines this quantitative data in an undisclosed manner to generate album recommendations and personalized playlists for its users. It is believed that Spotify relies too heavily on intrinsic song similarity to determine recommendations without incorporating the manner in which each song is comprehended by its users (“Spotify Echo Nest API,” 2015).

In sum, the largest flaws with the current systems using one type of data set are inefficiency, inaccuracy, reliance on a significant amount of accumulated information from the user population, and failure to effectively utilize musical information inherent within the songs themselves.

## ***2.2 The MUSIC Model***

### 2.2.1 MUSIC Model Overview

A large portion of our model for determining music preference and perception is based on a five-factor model developed by Rentfrow et al. (2011). Their work is based on the psychological effects that music has on listeners, ranging from dopamine release to mental stimulation to the use of musical listening habits as self-identifying personality features. The researchers wanted to examine the different aspects of a musical composition and examine what specifically impacted the preference of a listener toward one specific type of music over another. Many other studies have researched similar ideas and have come to inconsistent conclusions as to exactly how many factors are taken into account by a listener when determining if they like a piece of music, these numbers typically falling between three and seven factors. Rentfrow and his colleagues performed several studies to determine if underlying psychological factors are the basis for musical preference, and in the first study they established a rudimentary version of the five-factor MUSIC model (Rentfrow et al., 2011).

### 2.2.2 MUSIC Model History

Dr. Jason Rentfrow and his colleagues first proposed a version of the MUSIC model in 2003 in the paper “The Do Re Mi’s of Everyday Life.” In this paper, they lay the groundwork for their future studies, as well as establish a framework for examining musical preferences as they relate to the personality of the listener. They examined four different questions relating to this overarching idea: how important is

music to the listener, what are the dimensions of music preference, how can these dimensions be characterized, and how do these dimensions correspond to extant personality dimensions. They examined these questions through a series of six studies, the first of which examined the importance of music in everyday life, the second, third, and fourth examined structure of musical preference, the fifth examined psychological attributes of music, and the sixth examined music preferences, personality, cognitive ability, and self-perception (Rentfrow and Gosling, 2003).

#### 2.2.2.1 – MUSIC Model First Study Review

The first study utilized a questionnaire packet to examine the attitudes of the participants regarding their beliefs and attitudes regarding the importance of music in everyday life, whether or not subjects believe music preference has indication of their personality, and in what contexts people generally listen to music. This study was performed on 74 undergraduate psychology students at the University of Texas Austin. The results of this study demonstrated that the subjects considered music to be about as important as hobbies are to their everyday life, and that music preference is significantly meaningful. Additionally, the subjects believed that musical preferences revealed a large amount about their personality, with their importance being on a similar level to hobbies and bedrooms ( $M = 69.4$ ,  $M = 76.5$ , and  $M = 63.4$  respectively, where  $M$  is importance). This study revealed the importance of music in everyday life, and served as a justification for the development of the MUSIC model (Rentfrow, 2003).

#### 2.2.2.2 – MUSIC Model Second Study Review

The second study utilized a test developed by Rentfrow and Gosling (2003), the Short Test Of Music Preferences (STOMP). This test was used to examine whether factor dimensions could be established for music preference. The study sample was comprised of 1,704 students at the University of Texas Austin. The subjects first completed the STOMP and a number of personality tests, then completed the STOMP three weeks later. The results of this study revealed four different rudimentary factors: Reflective and Complex, Intense and Rebellious, Upbeat and Conventional, and Energetic and Rhythmic. Each of these factors included a number of genres with inherent qualities that corresponded to each factor, for example, jazz (factor loading of 0.83), metal (factor loading of 0.75), pop (factor loading of 0.59), and electronica (factor loading of 0.60) would fall under the factors in the same order they are listed. Since the study took place over a number of weeks, they were able to examine the stability of musical preference over time, and confirmed that the musical tastes of the subjects tended to stay relatively stable over time. These results suggested to Rentfrow and Gosling that there existed an underlying structure to musical preference.

#### 2.2.2.3 – MUSIC Model Third Study Review

The third study utilized the same methods from the second study, but attempted to confirm the generalizability of the factors and proposed structure across sample populations. As such, the sample was comprised of 1,383 University of Texas Austin undergraduate students, none of which had participated in the second study. Upon analysis of the results, Rentfrow and Gosling found that the factors and



structure did indeed remain consistent across populations, but there was little significant data regarding crossovers between factors. This ultimately provided more support for the four factor model they were proposing.

#### 2.2.2.4 – MUSIC Model Fourth Study Review

The fourth study tested the four factor model and musical preference structure to see if it remained true across geographical locations. Using audiogalaxy.com, Rentfrow and Gosling categorized the music libraries of 500 users from 50 states, totaling 10 randomly selected users per state, into genres, and then 20 songs were randomly selected per user from their respective pools. They then employed judges to attempt classification and coding of each song into a genre covered by the STOMP, and upon completion the user preference for a specific factor was determined by the number of songs within that category. The results of the study correlated with the results from the previous two studies, and each factor loading was strong and properly oriented. Combined with the previous studies, these results show consistency across time, geographic region, and sample population. Based on these results, we can say that even this rudimentary model is a very strong predictor of individual music preference and can be generalized to any population in the U.S.

#### 2.2.2.5 – MUSIC Model Fifth Study Review

The fifth study examined the attributes of music that allowed songs to be categorized into the four factor groups, and ultimately determined what qualities were descriptive of the four factors. To do this, Rentfrow and Gosling chose 25 songs for each of the 14 genres contained in the STOMP, and had judges evaluate each song on a five point scale to determine where each song would fall categorically. The results

showed that within the four factors, there were distinct musical attributes such as positive affect, negative affect, rhythm and complexity that defined the contained songs.

#### 2.2.2.6 – MUSIC Model Sixth Study Review

The sixth study served to examine this established factor model and determine if there were correlations between the factors and various personality traits of listeners. To determine this, Rentfrow and Gosling administered a series of personality tests on the subjects from studies one, two, and three. The tests they used were: Big Five Inventory, the Personality Research Form – Dominance, the Social Dominance Orientation Scale, the Brief Loquaciousness and Interpersonal Response Test, the Rosenberg Self-Esteem Scale, the Beck Depression Inventory, the Wonderlic IQ Test, and a self-view test designed by Rentfrow and Gosling. The results of this experiment showed that different personality traits correlated to the genres of music that one would expect. For example, openness and athleticism correlated to Intense and Rebellious, extraversion and liberalism correlated to Energetic and Rhythmic, and conscientiousness and agreeableness correlated to Upbeat and Conventional. Additionally, the factors had overall correlations with the tests of 0.977 for Reflective and Complex, 0.863 for Intense and Rebellious, 0.923 for Upbeat and Conventional, and 0.851 with Energetic and Rhythmic. This study also showed that there was no strong correlation between emotional stability and depression with musical preference factors.

The sheer amount of testing and reproducibility of these rudimentary factors show that this model, even in its early stages, was and is very reliable in terms of determining the musical preferences of users, and as such, it can be used extensively for user analysis, and from that a recommendation system tailored to individuals can be established. The five-factor MUSIC model discussed in Section 2.2.3 below is very closely based on the original four factor model that Rentfrow and Gosling had generated in their initial research.

### 2.2.3 The Five Factors of the MUSIC Model

The factors in this model - Mellow, Unpretentious, Sophisticated, Intense, and Contemporary - attempt to model the psychological basis of how people perceive music. Mellow describes the relaxedness, slowness, sadness, quietness, and other non-aggressive aspects of a piece. Unpretentious describes the lack of complexity, unaggressive, softness and acoustic nature. Sophisticated describes the complexity, intelligence, and dynamic nature of a piece. Intensity describes the distortion, tenseness and aggression of a piece. Contemporary describes the percussive nature, rhythmic nature, the current, and the danceability of a piece (Rentfrow et al., 2012). Table 2 provides descriptions of the MUSIC factors and typical genres that commonly match the key characteristics of each factor.

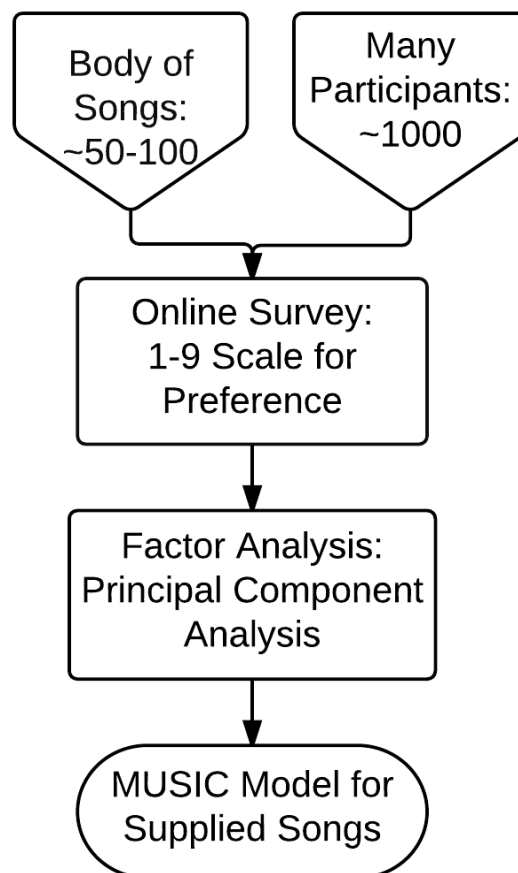
**Table 2.** Description of the factors present in the MUSIC model, as described by Rentfrow et al. (2012).

<i>Model Factor</i>	<i>Abbreviation</i>	<i>Description</i>	<i>Typical Genres</i>
Mellow	M	Romantic, Relaxing, Unaggressive, Sad, Slow, Quiet	Soft Rock, R &B, Adult Contemporary
Unpretentious	U	Uncomplicated, Relaxing, Unaggressive, Soft, Acoustic	Country, Folk, Singer/Songwriter
Sophisticated	S	Inspiring, Intelligent, Complex, Dynamic, Cultured	Classical, Operatic, Avant-Garde, World Beat, Traditional Jazz
Intense	I	Distorted, Loud, Aggressive, Tense, Not Relaxing, Not Romantic, Not Inspiring	Classic Rock, Punk, Heavy Metal, Power Pop
Contemporary	C	Percussive, Electric, Rhythmic, Danceable, Not Sad	Rap, Electronica, Latin, Acid Jazz, Euro Pop

#### 2.2.4 Model Development

Each of these factors has weighted values that additionally describe how important a given factor is to a person when they listen to and make preference evaluations for music. Rentfrow et al. utilized Principal-Component Analysis (PCA) with Varimax rotation to determine that the first feature accounted for 27% of the variance in the test population, and that a following parallel analysis showed that five eigenvalues of the first factor were the most indicative of a trend. To develop the ratings for the factors, the Rentfrow team utilized the same technique on the music ratings that the survey participants gave to the sample songs. To determine factor loading, Rentfrow and his team performed a PCA with Varimax rotation on the data from an additional study where users ranked preference of song excerpts and genre

lists. This provided a basis for the final weighting, where five hierarchical regression analyses were performed by overlaying the loadings with the genres, resulting in correlations showing values from .74 to .96, with a p value less than .05 for each factor, qualifying them as very significant (Rentfrow 2012). These factors were tested on multiple occasions by Rentfrow and his colleagues, who determined them to be significantly accurate descriptors for musical perception (Rentfrow, 2012). A simplified depiction of the general approach sequentially used by the researchers to develop the MUSIC model is shown in Figure 1.



**Figure 1.** Simplified methodology used to develop the MUSIC model through a series of survey studies by Rentfrow et al. (2011, 2012)

### 2.2.5 Model Versatility

Additionally, a group at the University of Geneva was able to replicate the results from the Rentfrow experiments and apply them to previously uncharacterized songs, and obtained a confidence interval of about 70%, indicating that the model is well-made, and will be accurate even for songs that were not initially used in the original dataset (Soleymani et al., 2015). There is some debate on exactly how many factors are needed to comprise music preference and cognition, with numbers generally varying from four to seven, but we chose this five-factor model because it has been more thoroughly researched. The MUSIC Model has been proven to be accurate across cultures by a number of research groups in locations such as Japan (Brown, 2012), the Netherlands (Delsing et al., 2008), and Malaysia (Chamorro-Premuzic, 2009) using culturally and geographically appropriate musical excerpts. The fact that this model works outside of the demographic it was originally designed around is a strong indicator of its versatility and overall strength as a model of musical preference.

Another study was performed at the University of Montreal that examined what improvements could be made on musical recommendation systems, and named the incorporation of individualized musical taste as one of the more important factors (Laplante, 2014). They examined a large body of literature surrounding the influence of individual characteristics such as political orientation, race, ethnicity, religious beliefs, age, and several others on musical taste and preferences. They noted that many of these characteristics corresponded to between four and five features, which could be characterized fairly well by the MUSIC model (Laplante, 2014). Combined

with the cross-cultural effectiveness of the model, as well as its ability to account for a large number of individual characteristics, as a result the MUSIC Model has been shown to be an excellent indicator of individual musical preference and taste, which led us to select it as our primary factor model for our research.

#### 2.2.6 Application of the MUSIC Model to Music Recommendation

The research of Soleymani et al. was useful in understanding the methodologies that might be useful for our research question, since it is the only example of applying the MUSIC model to music recommendation. It showed how the MUSIC model of Rentfrow (2011) could be used to overcome certain issues with current content-based music recommendation methods. Soleymani et al. identified several major flaws that content-based recommendation systems of the time were unable to overcome, even with the additional help of metadata. The largest problem this research set out to address was the cold start problem, where a computer model cannot adequately draw inferences for new users or items because it has no created points of reference for them. Because of this, the recommendations for new users or for new items are inaccurate for a significant period of time. There is also a problem from the significant bias created towards already popular items, also identified by Soleymani et al.

Soleymani et al. used the samples of music from all of Rentfrow et al.'s five-factor studies in 2011 ("The Structure of Musical Preferences: A Five-Factor Model") and 2012 ("The Song Remains the Same: A Replication and Extension of the MUSIC Model"), for a total of 249 songs spread across 26 genres. They considered it important to use not only the studies on Jazz and Rock samples but also the ones that

used diverse genres of music, because one of their goals was to separate the five-factors and the user preferences they modeled from qualitative genres, which are problematic in content-based music recommender systems. The five-factor approach would also allow a music-recommender system to bypass a similar flaw for genre preferences that Soleymani et al. observed, wherein “people tend to associate musical preferences with social stereotypes. And though the music might not appeal to a user, the stereotype does, which might influence his choices, especially in adolescence” (2015). Soleymani et al. believes that these “pitfalls” can be avoided by having users give preferences on sample clips of songs, as was done in Rentfrow et al.’s research.

To test the validity of the five-factor approach to solving the aforementioned problems, Soleymani et al. used modulation analysis to extract timbral features from the music clips, using methodology from Sturm and Noorzad’s paper, “On automatic music genre recognition by sparse representation classification using auditory temporal modulations.” They claim to have made this decision because prior research has shown that auditory temporal modulation features “work well for genre recognition”. A series of transformations on this data were used to estimate the MUSIC factors. Three methods of training the model were tested: multilinear regression, support vector regression, and regression with sparse representation. Across the board, RSS had the best results. The results of the baseline methods (user’s average ratings, genre-based, similarity based, and artist based) and the attribute-based system were compared using root-mean-square error (RSME) values, and attribute-based recommendations were shown to have the lowest error with an RMSE value of  $0.251 \pm 0.039$ . This leads Soleymani et al. to conclude that the five-



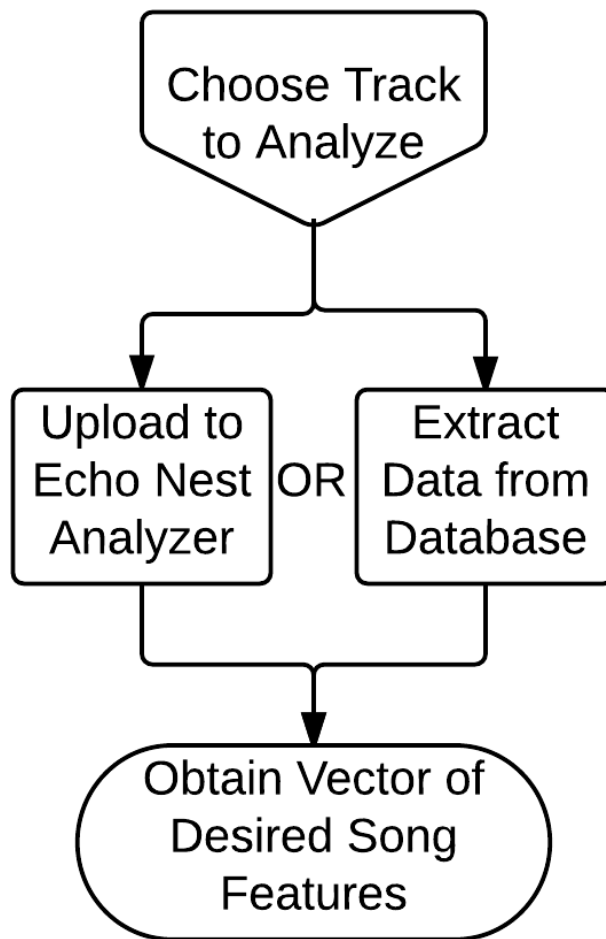
factor attribute-based model is able to address the shortcomings of current music-recommender system methodologies including genre ambiguity, a bias towards popular songs and the cold start problem.

### **2.3 *The Echo Nest***

Given the propensity of current music recommendation systems to utilize *The Echo Nest* for metadata purposes, there exists a large amount of information available from, and about *The Echo Nest* body of musical features. *The Echo Nest* is a very broad body of data that takes into account the more concretely defined features of music, such as tempo and key of a piece of music. *The Echo Nest* is used by some currently available music recommendation engines, Spotify being the most noteworthy one (“Spotify Acquires *The Echo Nest*,” 2014). Ellis and his colleagues have performed studies on *The Echo Nest* to determine its accuracy given a 10 million song dataset. From this work, they were able to extrapolate numerically similar songs given the similarities between the fingerprints of songs (Ellis, Whitman, Jehan, & Lamere, 2010). A song’s fingerprint is its collection of features from *The Echo Nest* values that are assigned to the song. *The Echo Nest* library can then be applied to the Million Song Dataset, which was developed by Ellis and Bertin-Mahieux solely to be a large amount of data for analysis using things like fingerprinting or metadata analysis (Bertin-Mahieux et al., 2011)

This has provided us with a large body of data to test the crossover between the psychological model of preference and quantitative makeup of the songs. Since there is a distinct lack of literature surrounding the use of the MUSIC model in conjunction with *The Echo Nest* body of data, we have identified a gap in research

that we can fill. Given the accuracy and use of both, such a pairing could allow for the development an accurate and useful model that combines the cognitive factors with the computational features of *The Echo Nest*. A simplified outline of the way in which *The Echo Nest* could be used to obtain numerical feature values is shown in Figure 2.



**Figure 2.** Simplified methodology for the extraction of quantitative music features from *The Echo Nest*

## Chapter 3: Study 1: Focus Groups on Music Recommendation

### **3.1 Methodology**

#### 3.1.1 Motivation

In our initial collection of data, we wanted to investigate the attitudes of music-listeners in regards to the performance of current popular music recommendation platforms, such as Spotify and Pandora. Collecting this information would allow us to identify which aspects of these systems users found satisfactory, and which aspects needed improvement for an ideal song recommendation experience. In our model, we would be able to improve upon the shortcomings of the current platforms.

We furthermore set out to discover and familiarize ourselves with musical features, such as timbre and instrumentation, which listeners identified in songs. Certain features that engaged listeners as they developed their appreciation or dislike for a song would be considered more important in the application of music recommendation, and could potentially be integrated into our recommendation system. We determined that we would collect data on the frequency of musical features and implement these features into our recommendation system.

When planning this primary data collection, we considered a variety of approaches, such as surveys and focus groups. We determined that type of the data we wanted to collect would be most appropriately gathered in the form of focus groups, for several reasons. We intended for our focus groups as a whole to provide a

more general understanding of the subject and a deeper level of communication with participants. Conducting focus group studies allowed us to experience face-to-face interaction and open-ended questioning with our research's primary audience: people who regularly use music recommendation systems as a tool to expand their song-listening base. We therefore conducted focus group sessions, in which we would facilitate discussion with participants regarding the general attitudes toward popular music recommendation systems, and musical features that are most impactful toward perceiving songs.

### 3.1.2 Participant Enrollment

Institutional Review Board (IRB) approval was obtained for up to five focus groups, jointly incorporating no more than 50 participants for \$10 compensation each. Participation was limited to the students enrolled in the University of Maryland, College Park, mainly for the ease of locational and population availability. Information about our focus groups was circulated by posting administration-approved flyers in highly populated areas around campus. The postings called for people that were "interested in talking about music" for an hour-long period, and advertised the compensation of \$10 for participation in the focus group. Posting locations included North Campus Diner, the Clarice Smith Performing Arts Center, the Engineering and Chemistry buildings, the Computer Science building, the Art and Sociology building, the Susquehanna (English) building, and the Biology-Psychology building. A map of the University of Maryland, College Park campus that indicates the posting locations can be found in Appendix A1. The same information was additionally advertised through email listservs that cater to the Honors College and

the School of Music. This spread of advertisement was intended to create a diverse sample of students.

### 3.1.3 Focus Group Operation

Our five focus groups were held in a standard-sized classroom in the Hornbake Library on the University of Maryland, College Park campus (See Appendix A1). Before entering the focus group room, participants were presented with a consent form that outlined the research that they would be partaking in, the potential risks and benefits, a confidentiality statement and researcher contact information. All participants were required to read this information and sign the form to indicate their agreement with the statements before entering the room. Participants were additionally allowed to write any name they choose to be identified by for the session on a nametag that that would be used for the remainder of the focus group. We explained to each participant that we would not release or share any identifiable or personal information in our research, and that they would receive monetary compensation of \$10 after the focus group had finished. While waiting for other participants to arrive, everyone filled out a demographic information sheet with information including their gender, ethnicity, and preferred genre of music. The full consent form can be found in Appendix A2.

Participants sat in a circle facing each other, alongside two members of the team who acted as moderators to ask questions and facilitate discussion between participants. Each session was video recorded after obtaining written consent from all participants, and created transcripts of the recordings for later analysis. Only single copies of the recordings were retained under password protection, and stored on a

secure drive to avoid compromising the participants' identities. Over a period of five sessions, each approximately an hour-long in duration, we were able to collect information from a total of thirty-seven student participants.

#### 3.1.4 Focus Group Structure

##### 3.1.4.1 Current Music Recommendation Methods

The focus groups were organized into two main stages: 1) Recommendation System Discussion, and 2) Song Feature Analyses. In the first stage, we asked participants about their typical method for discovering new songs, and whether they used any recommendation platforms to assist them in finding music. This discussion of discovering new music involved the following questions:

1. *What is your process for discovering new songs?*
2. *Do you use a music recommendation system to help you find new music?  
If so, which one(s)?*
3. *How do you feel about the results provided by the recommendation system?*

This portion of the session was designed to develop an understanding of the expectations that users had for music recommendation systems, and categorize their attitudes toward music recommendation through general positive or negative inflection.

##### 3.1.4.2 Song Feature Analysis

The second portion of our focus groups involved a procedure in which we had participants listen to song clips and then describe their reaction to hearing the music.

We chose the following songs for this process:

“Bergamasque” by *Debussy* (1:00)

“The Robots” by *Kraftwerk* (0:55)

“Mi Raza” by *Inkari* (2:10)

“All the Things You Are” by *Charlie Parker* (0:20)

“From the Sun” by *Unknown Mortal Orchestra* (0:00)

“Generation” by *Liturg* (2:10)

“Luv” by *Nujabes* (0:11)

“Flaws” by *Bastille* (2:07)

These songs were chosen as a whole to be representative of multiple genres of music that would provide contrast to each other. Each song was chosen to represent their respective genre because of their use of highly characteristic musical elements.

- “Bergamasque” by Debussy was chosen for classical for the individual piano instrumentation, its melodic quality, open soundscape, and a lack of familiarity.
- “The Robots” by Kraftwerk was chosen for techno/electronic for the heavy use of synthesizers, the distinct bassline, and repetitive robotic and futuristic sounding vocals and aesthetic.
- “Mi Raza” by Inkari was chosen for its unique instrumentation and timbre, which are typical of Latin and South American music.
- “All the Things You Are” by Charlie Parker was chosen for jazz for the strong presence of a lead saxophone, the quiet accompanying string and percussion, and the freeform flow of the music.

- “From the Sun” by Unknown Mortal Orchestra was chosen for rock for the repeating riffs, the low fidelity quality of the instruments, and the aural sensation of someone playing the instrument.
- “Generation” by Liturgy was chosen for heavy metal because of the distortion-heavy, driving, and repetitive guitar riffs.
- “Luv(sic) Part 1” by Nujabes & Shing02 was chosen for hip-hop because of the consistent bass and snare backbeat, the “scratching” sounds characteristic of hip-hop DJs, and the looped instrumental clips.
- “Flaws” by Bastille was chosen for pop for the very upbeat, light instrumentals, the standard 4:4 time signature, and the incorporation of catchy synthesizer accompaniment.

Within the scope of our focus groups, it was determined that the specific musical content of these eight songs should not be the primary concern of the experiment; rather, the observation of the participants’ responses to the types of features intrinsic to the music was the main goal. Thus, it may have been plausible to have presented differing set of songs to the participants and still have maintained the same integrity and overall purpose of the focus groups.

Segments of approximately thirty seconds within each song were chosen that feature the prominent feature of the song, as discussed above. These segments were selected to be mostly non-lyrical, as we wanted to ensure that the lyrical content of the song itself was not influential in our participants’ reactions. The starting time of the selected segment is indicated above. To ensure that focus groups would be completed in a timely manner and to minimize participant fatigue, each group only



listened to five of the eight selected songs. The songs were selected by following the order given above. This sequence was used to separate songs with similar characteristics and to allow each song to sound fresh to the participant. While this approach could be seen to reduce the amount of data points we were able to collect, we felt that this allowed for better conversation of the songs among participants without surpassing the one-hour limit. Possible bias arising from the songs each group listened to was avoided in the analysis of the transcripts as solely on the types of features that participants discussed were extracted, not the exact details of the feature at the pertains to the song.

Participants were asked questions following this format:

- 1. How would you describe the music?*
- 2. What descriptors would you use to describe the song?*
- 3. What did you focus on the most when listening to the song?*

After listening to and discussing the five songs specified for the given session, the moderators facilitated a discussion regarding listening to music in general, asking participants to talk about characteristics of music that made them react in a certain favorable or unfavorable way. Many of the participants discussed how the song clips they had heard sounded similar to other songs they have heard previously.

### 3.1.5 Data Analysis

We created transcripts from the video recordings to be used for analysis of the focus groups. For the first part of each focus group session, occurrences were tabulated corresponding to each time a music recommendation service was named or discussed in a novel manner. Therefore cases of agreement were not counted as novel

occurrences. Additionally, each time a service was discussed it was categorized as positive, negative, or neutral, based on the experience being described or the word choice used by the participant. All cases that were considered ambiguous were classified as neutral. For the latter portion of the focus group sessions, comments on music were classified with respect to general music features. All transcripts were analyzed by at least two researchers to minimize any imposition of bias in this process.

## **3.2 Results**

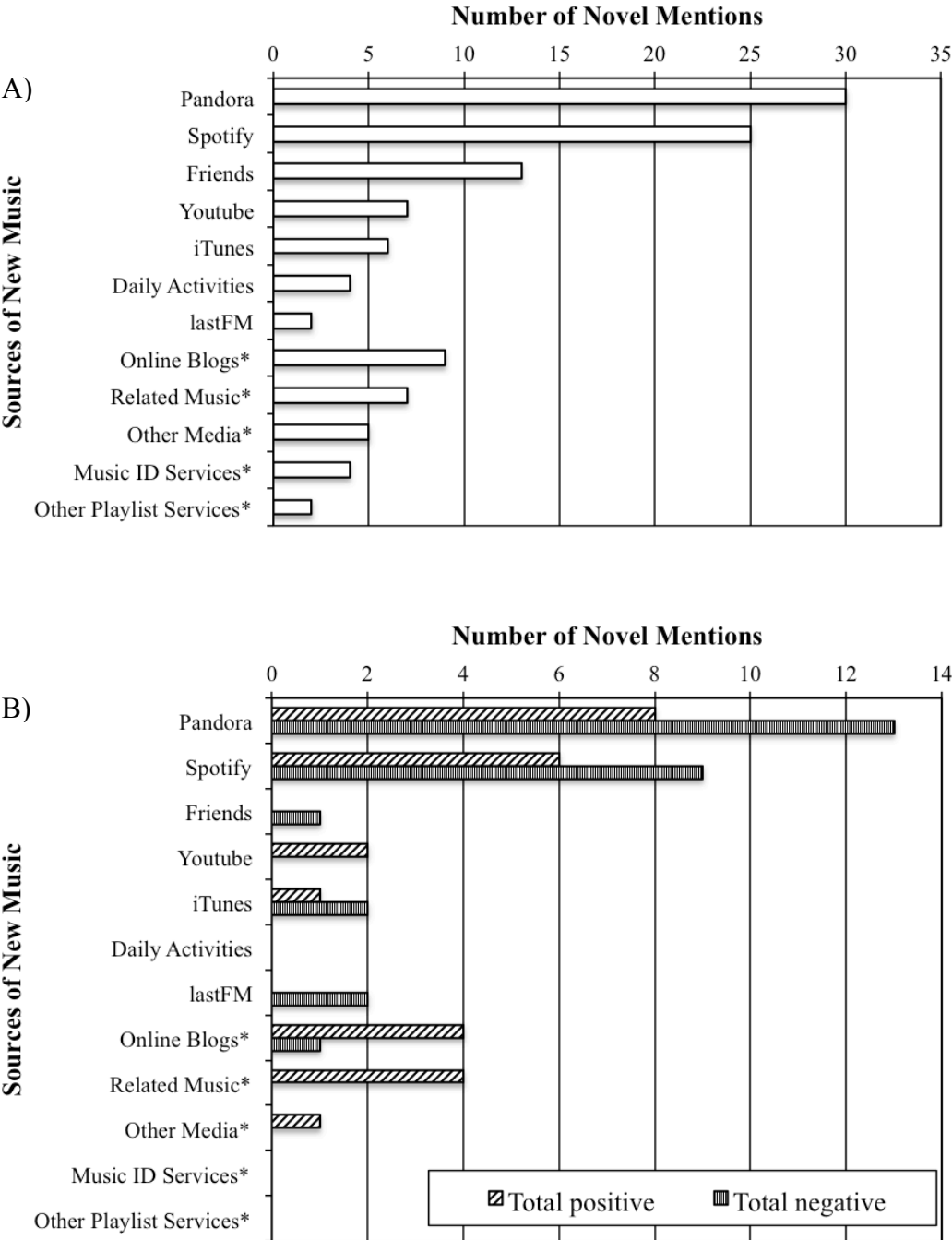
### **3.2.1 General/Demographic Information**

We initially conducted several focus group sessions for the purpose of finding out several pieces of information regarding music listening trends, frequently used services, as well as what sort of musical qualities were important in developing song preference. The participants were comprised of college age students, ranging from 18 to 22 years of age. The participants comprised a variety of majors, from scientific and non-scientific backgrounds. We ran five sessions with a total of 37 participants. Of these, 20 identified as female, and 17 as male. The ethnicities were as follows: 11% Asian/Pacific Islander, 19% Black, 8% Hispanic, 49% White, and 11% were combinations of the aforementioned ethnicities. Participant listed music genre preferences were fairly diverse, with larger tendencies toward Alternative Rock/Pop, Classic Rock, Metal, Pop, and India Pop/Folk. A complete breakdown of the focus group participant demographics can be found in Appendix A3.

### 3.2.2 Attitudes towards Current Recommender Systems

From these focus group sections, we were able to determine that a majority of the participants were constantly looking for superior music recommendation systems to the ones they were using at the time. The services that were most frequently mentioned were Pandora and Spotify, which were mentioned twenty two and sixteen times respectively as the primary method of finding and being recommended new music. All of the mentioned methods are shown in Figure 3A. Of those instances, a majority of the participants were dissatisfied with the results they were seeing from the services, with eight participants stating they were satisfied to fourteen stating dissatisfaction with Pandora, and six participants stating they were satisfied to ten stating dissatisfaction with Spotify's recommendation system. The total population of participants stated that they were generally dissatisfied with extant recommendation systems, stating that they felt the systems were too generalized for mass amounts of users, and did not have much focus on what aspects of music the users actually liked. These statistics for all of the mentioned methods of finding new music are displayed in Figure 3B, and several quotes from participants can be seen in Figure 4.

**Figure 3.** Prevalence and options on sources of new music. A) The number of time each method of finding music was mentioned over all sessions is indicated. B) Expressions of positive or negative feelings toward music recommendation methods are shown. In each figure, sources with an asterisk (\*) indicate a conglomerate source. “Online Blogs” is defined as mentions of Tumblr, Billboard, various blogs, Myspace, music review sites, Google, and Soundcloud. “Related music” is defined as mentions of related artist lists and CDs. “Other Media” is defined as mentions of television and radio. Music ID Services is defined as mentions of Shazam and Soundhound. “Other Playlist Services” is defined as mentions of Songza and 8tracks.



**Figure 4:** Quotes from focus groups regarding recommendation systems.

Spotify	“...see the bands that sound like it, that are recommended, and they’re all just bands that I’ve heard of, guys that have like 40, or like no followers or something.”	“If I start an artist radio based on a song I like, it’ll be like, ‘eh, it’s alright,’ and just keep skipping past stuff until I find stuff that I like.”	“The songs [Spotify] recommend, I already know.”
Pandora	“I wouldn’t go to Pandora to find different types of music.”	“It’ll like put you in a box. If you listen to one thing it’ll only suggest stuff that is related or sounds exactly like it.”	“You get pigeonholed and [Pandora] starts repeating songs.”
Genius	“[Genius] was terrible, I didn’t like it. It never gave me good songs.”		“They never have any of the songs that I have in my library.”
Soundcloud	“I’m not gonna hear about the new classical rock EP or something just by going through my Soundcloud feed.”		

### 3.2.3 Prominent Features Noted in a First Listen

From the transcripts of the focus groups, we analyzed the number of mentions for specific musical features or qualities in an attempt to determine what the participants focused on predominantly when deciding whether or not they liked a piece of music. Over all of the sessions, we found that four specific features stood out, each being mentioned at least 30 times: instrumentation, perceived genre, rhythm, and song structure. The majority of other features were not mentioned nearly as much as these, so we can designate them as the several of the most important features for music categorization.

### 3.3 *Discussion*

From the trends observed from our focus groups, we were able to draw several conclusions regarding the current level of satisfaction within usage of current popular recommendation systems, and how they could be improved if necessary. Within the focus groups' discussions, the general dissatisfaction of the performance of popular recommendation systems suggested the need for a refined, different approach to music recommendation, providing validity to the development of our system. While some participants found these systems to be adequate in terms of musical discovery, others felt that they could be improved in the way of playlist generation and suggestion of new songs based on the user's individual tastes. These results had demonstrated how there is room for improvement in the current state of music recommendation. Over half of the focus group participants voiced dissatisfaction with the tendency of systems, such as Spotify and Pandora, to neglect specific user profiles and personalities in favor of catering to a more generalized audience. While it may be logistically more efficient to design a more generalized recommendation system to process information for millions of users, these systems still exhibit deficiencies in terms of user specificity and how listeners respond to different song features. After conduction of the focus groups, we examined methods of how this user-system disconnect could be corrected by implementation of additional information regarding the user. We came to recognize that the deficit in current systems, as observed in our focus groups, could potentially be corrected through the integration of listener perception models, especially those formed through the expanding field of extant listener-psychology research (Rentfrow & Gosling, 2003; George et al., 2007;

Desling et al., 2008; Schafer & Sedlmeier, 2009; Rentfrow et al., 2011; Brown, 2012; Langmeyer et al., 2012; Rentfrow et al. 2012). As iterated in numerous psychological studies, music perception plays a significant role in the formation of musical preferences. Furthermore, many of our participants suggested that music recommendation systems should be able to take into account their mood - which provided additional reason for us to implement the music perception model into recommendation.

In the song-listening portion of our focus groups, we had asked our participants to discuss the features of the music that they felt were most prominent and engaging in the listening process. This line of questioning provided insight into the relationship between personal song preference, and identification and perception of musical features. The most prominent features presented by our participants were instrumentation, perceived genre, rhythm, and song structure. We suspect that these features were represented with such frequency due to being relatively easy to understand for the average listener. Many listeners identify with a preferred genre or appreciate a song for its melody, which is broken down into key components, such as rhythm and song structure. Instrumentation has several associated features - particularly, the timbre and acoustic properties of the instruments used in the song. The song structure (e.g., melody) and timbre are often the most easily recognized and most defining attributes of a song (Schellenberg & Habashi, 2015), so it logically follows that our participants had also identified these features as important song properties for their song preferences. The results of the feature identification portion of the focus groups are further supported by previous studies (Istok, Brattico,

Jacobsen, Ritter, & Tervaniemi, 2013) that have confirmed the importance of music genre preferences in cognitive responses to music.

Following the focus groups, we looked for ways of incorporating these primary listening features into our novel music recommendation method, while still allowing the recommendation to be automated to account for a large volume of songs. Within *The Echo Nest* API, we were able to obtain *tempo*, *danceability*, *energy*, *speechiness*, *liveness*, *acousticness*, and *valence* as features derived from the intrinsic properties (e.g., time signature, beats per minute) of the songs. On the basis of instrumentation acting as an important feature for song listeners in developing preference, we theorized that acousticness would be one of the more influential *The Echo Nest* factors of our recommendation method. In addition, we could integrate inferred information regarding the genre and structure of the song through the *energy* and *valence* features provided by *The Echo Nest*.

There are some possible limitations to our focus group design. For instance, the number of participants was highly variant between sessions - some sessions had over ten participants, while others had seven or under. This variability was mainly due in part to the prospective participants' decision to attend the focus group session, as we had scheduled each session to have close to the same amount of participants. Furthermore, some may consider our selection of the songs to be non-objective. We believe that the actual musical content of these songs was relatively arbitrary, since we were aiming to measure the reactivity of participants to principle song features rather than specific song components. Finally, our focus groups demographics and age group were limited to the student body of the University of Maryland - if we had



had more time and more resources to collect the data from a more general population,  
we would have attempted to do so.

## Chapter 4: Study 2: Survey on Music Preference

### 4.1 *Methodology*

#### 4.1.1 Motivation

The study and results obtained from using the MUSIC model (Rentfrow et al., 2011) provided data that offered insights into the psychological aspects of music perception that was not directly accounted for in the feature values for songs from *The Echo Nest* API. Before attempting to bridge the gap between the mathematically precise *The Echo Nest* computations and the music perception data, we first tested the validity of the MUSIC model ourselves. We anticipated using the results from an online survey to see if trends in music preferences followed the trends in MUSIC factor values across the songs. This would justify using the MUSIC factor values along with *The Echo Nest* feature values in a machine learning model.

#### 4.1.2 Survey Structure

Over a series of three articles, Rentfrow and his colleagues (Rentfrow & Gosling 2003; Rentfrow et al., 2011; Rentfrow et al., 2012) presented work regarding music perception and intrinsic human psychological traits. We crafted an online survey to include songs from the first article from the researchers that presents a fully developed five-factor model of music perception (Rentfrow et al., 2011) and tasked participants with reporting their preference on a numerical scale for the selected songs. Additional questions about song preference, genre preference, ethnicity, age, opinions on current music recommendation systems, comfort with listening to

different songs, and self-reported MUSIC factor values were also included. The survey required 32 questions to be answered in whole. See Appendix B1 for the full list of online survey questions.

#### 4.1.3 Participant Enrollment

Institutional Review Board (IRB) approval was obtained to conduct the survey and use our results in publication. Using the online survey service, Qualtrics, we were able to obtain survey data from 100 participants within the United States with a demographic spread comparable to that of the country as a whole. Due to the additional restrictions that would come from working with minors, we limited our survey participants to those 18 years of age and older. We also limited our participant to those less than 50 years of age because they are outside the target audience for music recommendation systems (Sikora 2015). 28% of our participants were 18 to 24 years old, 28% were 25 to 34, 30% were 35 to 44, and 14% were 45 to 50 years old. The average age of our participants was 32 years old. Our participants reported a diverse range of ethnicities, with 4% identifying as American Indian or Alaskan Native, 6% as Asian or Pacific Islander, 20% as Black or African, 17% as Hispanic or Latino, and 61% as White or Caucasian. These percentages are roughly similar to the racial breakdown of the country, so we do not believe our data was unduly influenced by a certain ethnic background. There was more disparity in the gender of our respondents; 77% of our respondents were female and only 23% were male. We have no explanation for why the participants provided to use were so disproportionately female, but we do not believe that it impacted the validity of our results.

#### 4.1.4 Song Selection

Given that we wanted to keep the length of the survey at about 30 minutes, due to the desire to maintain participant interest and ensure that the cost of the survey was manageable, we decided to incorporate twenty songs into the main body of the survey. The songs presented in the survey can be found in Table 3 and the audio clips can be shared upon request. We imposed several criteria for choosing the songs from “Table 3: Five Varimax-Rotated Principal Components Derived from Music Preference Ratings in Study 1” from “The structure of music preferences: a five-factor model” (Rentfrow et al., 2011). The selection process revolved around

**Table 3.** Songs from Rentfrow et al. (2011) featured in survey with one-minute song segment starting at specified time. MUSIC Model factor loadings for selected songs additionally provided with permission. Bolded values correspond to significant loadings. Songs with asterisk (\*) were not used in Study 4 (See Chapter 5).

<i>Artist</i>	<i>Song</i>	<i>Starting Time</i>	<i>M</i>	<i>U</i>	<i>S</i>	<i>I</i>	<i>C</i>
Karla Bonoff	Just Walk Away	1:35	<b>0.65</b>	0.27	0.26	0.15	-0.02
Ace of Base	Unspeakable	1:16	<b>0.63</b>	0.21	0.13	0.18	-0.01
Billy Paul	Brown Baby	1:51	<b>0.46</b>	0.35	0.26	0.17	0.16
The Mavericks	If You Only Knew	1:05	0.22	<b>0.73</b>	0.07	0.12	0.00
Uncle Tupelo	Slate	1:20	0.14	<b>0.72</b>	0.12	0.20	0.02
Flaming Groovies	Gonna Rock Tonight*	1:58	0.21	<b>0.46</b>	0.12	0.26	-0.03
William Boyce	Symphony No. 1 in B Flat Major	2:50	0.15	-0.03	<b>0.78</b>	0.05	-0.16
Oscar Peterson	The Way You Look Tonight	1:00	0.20	0.02	<b>0.74</b>	0.00	0.09
Ornette Coleman	Rock the Clock	1:18	-0.21	0.22	<b>0.49</b>	0.13	0.35
Iron Maiden	Where Eagles Dare	2:30	0.05	0.15	-0.02	<b>0.71</b>	0.03
Owsley	Oh No the Radio	1:58	0.07	0.06	0.04	<b>0.69</b>	0.09
BBM	City of Gold	1:35	0.16	0.29	0.07	<b>0.59</b>	0.04
D-Nice	They Call Me D-Nice*	1:32	0.02	0.08	0.10	0.17	<b>0.76</b>
Ludacris	Intro	0:00	0.03	-0.06	0.02	0.25	<b>0.72</b>
Age	Lichtspruch*	5:11	0.1	0.07	0.3	0.11	<b>0.45</b>
Frankie Yankovic	My Favorite Polka	1:00	0.00	<b>0.41</b>	<b>0.59</b>	0.09	0.1
Booney James	Backbone*	1:45	<b>0.40</b>	0.07	<b>0.55</b>	-0.03	0.23
Eilen Ivers	Darlin Corey*	1:41	-0.02	<b>0.40</b>	<b>0.45</b>	0.21	0.1
Doc Watson	Interstate Rag	1:00	-0.06	<b>0.57</b>	<b>0.44</b>	0.06	0.11
Adrian Belew	Big Blue Sun	1:30	0.28	0.12	0.12	0.35	0.02

choosing samples with a representative range of statistically significant factor loadings for each of the five MUSIC factors, such that no artists and genres were repeated. Statistically significant loadings were shown by the authors as those with values whose absolute value was greater than or equal to 0.40. All songs chosen as being representative of a MUSIC factor had a statistically significant loading on that factor. We chose two examples of high loading values for each factor. This was determined by looking at the highest factor loading listed, while also considering the genre and artist to avoid repetitiveness. We wanted the songs to be from a variety of artists and genres to minimize the bias towards any one music style. The genres represented included: Classical, Traditional Jazz, Acid Jazz, Mainstream Country, Country Rock, Rock-n-roll, Heavy Metal, Power Pop, Classic Rock, Soft Rock, Europop, R&B/soul, Rap, Electronica, Polka, Quiet Storm, Celtic, and Bluegrass, as identified by the authors of the article. We additionally chose one mid-level loading example for each of the factors. This was selected as a song whose factor loading for the respective factor was close to 0.40 while still being significant. This resulted in a total of three songs selected to represent each factor, or 15 songs in total. When possible, these representative songs were chosen so as not include songs with the highest or lowest significant loading in its representative factor. The testing set was rounded out with five additional songs that contained additional interesting characteristics. These included songs with multiple or zero statistically significant high loadings to see if they provided unique results about the connection between MUSIC factor loadings.

We chose one-minute segments of those songs so that we could present the listeners with a relatively high number of songs without losing a valid representation of the song itself. We decided that samples near the middle section of a piece would show the development of the introduction into the main themes of the song as well as the climax. We believe that this range of characteristics would provide the most representative one-minute segment of the song as a whole. To eliminate bias song to song on what this qualification meant, we developed a routine. We calculated the time when 40% of the song was completed, plus or minus five seconds, as the beginning of our one-minute segment. The five seconds was built in to remove awkwardness in the cutoff, such as beginning or ending in the middle of a word. The resulting starting times for the selected songs are given in Table 3.

#### 4.1.5 Data Analysis

Analysis of the data collected from this survey was completed mainly through regression analysis between the MUSIC factor loadings and each individual participant's provided preference values. The ability of the MUSIC factors' ability to model each participant's preference was analyzed. Finally, the results of this survey were additionally incorporated into an evaluation of our proposed method for determining MUSIC values for songs via an automated method (discussed and presented in Chapter 5).

## **4.2 Results**

Our survey participants were able to select up to 5 preferred genres of music, and reported diverse preferences. The most preferred genres were Pop/Rock, R&B,

Country, and Rap with 74, 63, 42, and 40 respondents preferring them respectively. This was an unsurprising result, as these are four of the most prominent genres in contemporary American music. See Appendix B2 for a graph of the genre preferences.

To assess the participants overall responses to the songs, we compared the minimum value, 1<sup>st</sup> quartile, median, 3<sup>rd</sup> quartile, and maximum value of the preferences provided for each song. No song in the survey was universally loved or hated by the participants: every song in the survey had at least one participant rank it as a 1 and one rank it as a 6. The median and quartile values instead give a better idea of overall participant response to each song. The least popular songs were “Rock the Clock” by Ornette Coleman and “My Favorite Polka” by Frankie Yankovic. It is possible that these songs were generally disliked because their genres, acid jazz and polka respectively, are not likely familiar to our participants. The most popular songs included Intro by Ludacris, and My Name is D-Nice by D-Nice. Both of these songs are rap songs, and a large number of participants identified as preferring the rap genre, so this result is unsurprising. However, these songs had a fair number of detractors, suggesting that rap is a polarizing genre. This is consistent with the anecdotal evidence from our focus group studies. The other songs with high preference rankings included pop songs like Unspeakable by Ace of Base and Oh No the Radio by Owsley. This shows that among our participants pop music is generally considered inoffensive even if not preferred. A few participants were noticeably negative in their preferences, with an average rating of less than 2 for all of the songs.

One participant was very positive, with an average song ranking above 5.75. Most of the participants, however, ranked the songs an average score between 3 and 5.

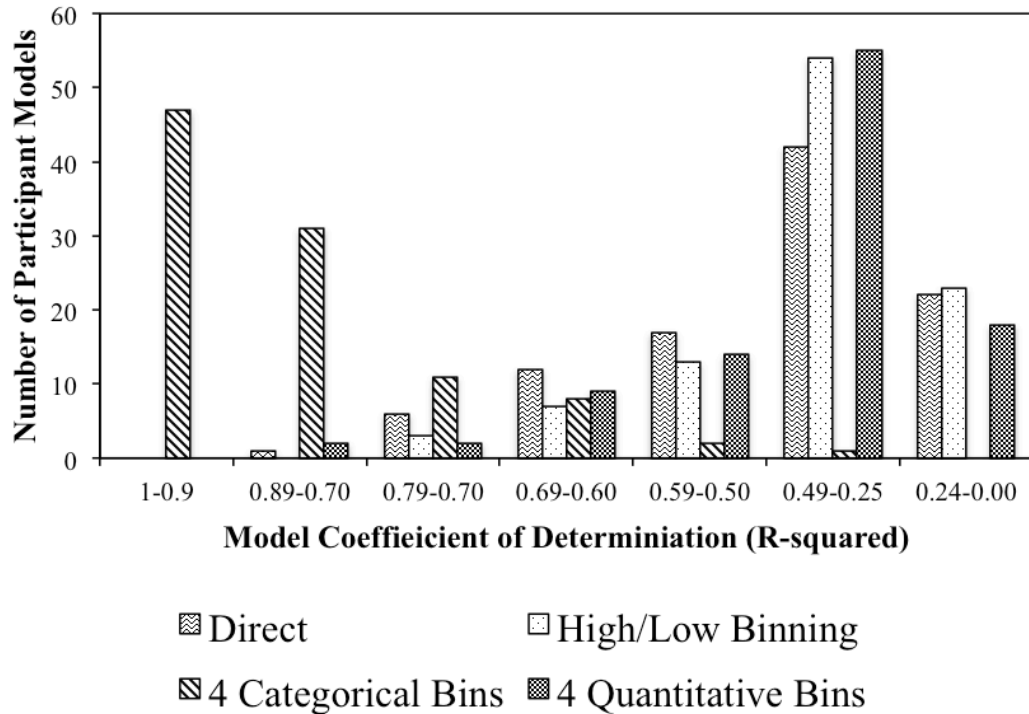
In addition to ranking their preferences of songs, we asked the participants questions to create self-reported MUSIC factor values (Q34-Q38, Appendix B1), and questions about their music listening habits (Q9-Q11, Appendix B1). To self-report MUSIC factor values, participants were asked to indicate their preference on a slider bar from 1-6 of how much they prefer songs that could be described by adjectives used to define the MUSIC factors. These self-reported values were then transformed to a scale of -0.25 to 0.75. After reviewing the values, we decided that other methods of analysis would yield better insight into modeling music preferences. When we asked participants to describe how comfortable they are when listening to music that is of a different genre, 84% of participants responded that they were somewhat comfortable or comfortable. No participants responded that they were completely uncomfortable listening to music outside of their usual genres. There was a lot of variety in how often the participants listen to music outside of their preferred genres, but 4% reported that they never do, 15% reported that they do daily, and on average our participants listen to music outside their preferred genres several times a month. We asked our participants how they would rate the ability of current recommendation systems to recommend music that matches their preferences, and 60% reported that they find current recommendation systems to be less than satisfactory. This is consistent with the previous literature on the subject and the results of our focus groups, suggesting that a significant portion of users do not find that the current recommendation systems on the market meet their needs.



In order to investigate the MUSIC model's ability to model user preference, we applied multiple linear regression. Minitab Statistical Software was used to generate and analyze each regression model. A participant model was generated by applying multiple linear regression to the MUSIC values from the Rentfrow et al. studies (2011, 2012) and the song preference data of each participant in turn. A few different values were recorded for each model, including coefficient of determination  $R^2$  (as shown in Figure 5), but these parameters were not analyzed in depth for this study.

The main method we analyzed our model was using an ANOVA test. ANOVA tables were generated for each participant model, which we were able to test for two types of hypotheses. The first hypothesis test was that all slope parameters were equal to zero. This gives an indication of the overall significance of our model. The second hypothesis was that each slope parameter on each MUSIC factor is zero. This yields a more specific analysis of the significance of each MUSIC factor in modeling music preference. Residuals were analyzed for major deviations from the inherent assumptions of this approach. The residuals checked for are independence, normally distribution, and for equal variance. This process was repeated for the preferences of each of the 100 survey participants for the four models we tested. These models were Direct Data, High/Low, Four-Point Categorical, and Four-Point Scale, or Quantitative (see Figure 5). The Direct Data model used the MUSIC factor values for the songs as given in Rentfrow et al.'s (2011) article. While this method ultimately revealed the most relevant results, we decided to explore the sensitivity of the MUSIC factor values by running a more simple models where the factor loadings

where binned into different groups. The simplest of which included just two bins: high and low. This constituted the High/Low Model. The values were considered high if greater than or equal to 0.4 (indication of a significant factor loading). The values were considered low if less than 0.4. A slightly more complicated, yet still simplified model, divided the factor values into 4 bins. This constituted the Four-Point models. The high and low bins were each split into two bins. The first bin included values - 0.25 to 0.09, the second included values 0.10 to 0.34 the third included values 0.35 to 0.54, and the fourth bin was values above 0.54. The four-point scale was analyzed in two ways: one that considered the bins to be consecutive numerical values (Four-Point Quantitative) and another that considered each as a nominal descriptor (Four-Point Categorical). We took this approach of binning the MUSIC factors because we believed that there could be a decent amount of variation in what the exact factor loading as they were determined via surveys, not through any sort of qualitative analysis of the song. The purpose of evaluating these different models was to investigate how simple the model can be made while still holding meaning. From the linear regression, we used Minitab to calculate several parameters including S, R-squared, R-squared(adj), PRESS, and R-squared(pred), however the only values used in our study were the R-squared values. These were calculated for each participant model for all four approaches, as shown in Figure 5. While we did not examine the binning methods in depth, we did notice some potential issues with them. We believe that the High/Low method is an oversimplification of the factors that lessens their fidelity. While the R-squared values for the Four-Point Categorical Model looked promising, the residuals for this model were strongly heavily tailed, and this



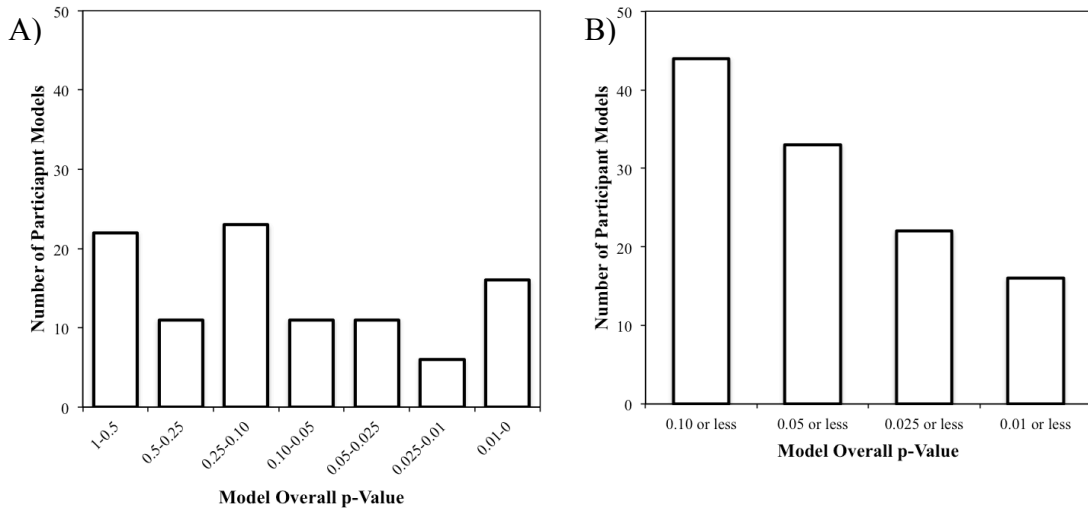
**Figure 5.** Distribution of model performances as indicated by coefficient of determination for different model methodologies. Note that despite the notable performance of the 4 Categorical Bins Method, the regression of these models had largely non-normal residuals.

abnormality with the residuals could possibly invalidate this data analysis approach. The Four-Point Scale data set seems to retain more significance than the High/Low approach while still maintaining a more condensed representation of the MUSIC factor loadings. As the main purpose of our study was to investigate the MUSIC factor model's ability to predict user preferences and not to optimize this method, we chose not to look further into incorporating transformations to allow for more proper residuals in modeling. Direct Data seemed to present the most accurate results than any of the optimization methods attempted, and so it was analyzed in the most depth from this point forward.

After our brief experimentation with different methods of binning and optimization, we returned to a more in depth look at the analysis that ANOVA

provided for the Direct Data participant models. We analyzed the p-values for the MUSIC model overall to test our first hypothesis (that all slope parameters are zero), and for each of the MUSIC factors individually to test for our second hypothesis (that each slope parameter on each MUSIC factor is zero). The p-value corresponds to the probability that the null hypothesis is observed. A significant p-value indicates that at least one of the slope parameters is not zero. If we consider a p-value of 0.05 to be significant, 33 participant models meet the criteria. If we are considering 0.01 to be significant, 16 participant models are meeting it. While there is a general downward trend as shown in Figure 6B, upon closer examination (Figure 6A) there is a spike of p-values between 0.01 and 0, which correspond to extremely significant models, and suggests that there is legitimacy to our approach in applying the MUSIC model to represent music preference.

In addition to analyzing the model overall, we ran an ANOVA on each of the MUSIC factors individually. This was to determine if any of the factors was more or

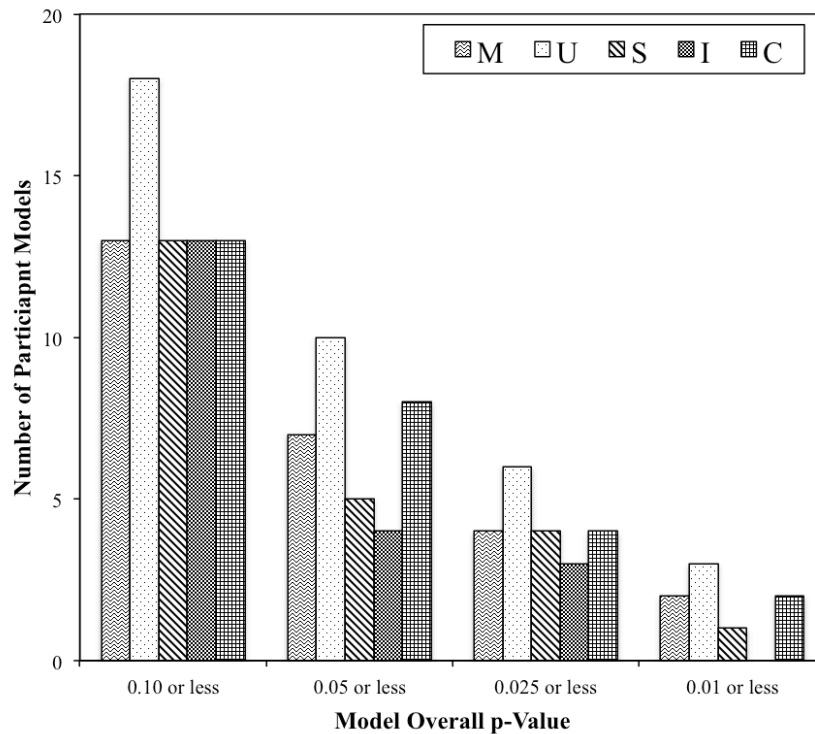


**Figure 6.** Evaluation of Direct participant models as indicated by the p-value for the multiple regression model obtained for each participant, n=100. A) Histogram for all participant models with respect to model p-value. B) Total number of participant models whose p-values satisfy the specified significance level.

less significant in modeling user preferences according to our method. We considered the idea that if certain factors proved to be significantly better predictors of taste than others, then a model could be built on those factors alone. As evident in Figure 7, overall this was not the case and the factors are generally consistent. U does appear to be a slightly more significant for lower p-values, and I is less significant. The C value remains reasonably significant across p-values relative to the other factors. None of the MUSIC factor values have such low statistical significance that any of them are noticeably more or less relevant in terms of prediction modeling.

### 4.3 Discussion

The intent of the online survey we conducted was to demonstrate that the



**Figure 7.** Coefficient p-values for the MUSIC variables in each participant multiple linear regression model. This data representation matches that of Fig. 6B, as it indicates the number of participant models whose p-values satisfy the specified significance level, but no specified for each factor. (M=Mellow; U=Unpretentious; S=Sophisticated; I=Intense; C=Contemporary)

MUSIC model constructed by Rentfrow et al. could predict song preferences with some degree of accuracy, through analyzing the correlations between each song's designated MUSIC factors and the participants' corresponding preference of the song. Lack of correlation between the elements would have suggested that the two are not related, and that the MUSIC model is not an appropriate predictor of a listener's song preferences. However, we found that for 32 out of 100 survey participants (Figure 6B), the MUSIC model had displayed a standard significant correlation to their preference for a song (for a critical value of 0.05). While this percentage indicates a minority number of participants, we believe it is still non-trivial, as it demonstrates that the MUSIC model shows a non-negligible degree of correlation to the participants' preference ratings. Upon closer examination of the results for participants with a p-value less than 0.05, Figure 6A shows that half of these participants (16 of 32) displayed a p-value of less than 0.01, indicating an even higher level of significance of correlation between their preferences and the MUSIC model. This may additionally be an overly harsh analysis, as in implementation, a user will not likely care whether or not a song arose from a *statistically significant* recommendation model, but more so whether or not they like or dislike the recommended song. With this in mind, additional analysis of the implementation of this data with a machine learning system to predict whether or not a user will like a song is considered in Study 4 (see Chapter 5).

This analysis is not enough to justify using the MUSIC model as a stand-alone model on which to generate song recommendations. Based on the results of our survey analysis, we believe that the MUSIC model is best used as a high-level

categorization feature that may serve to improve the performance of current recommendation systems. The MUSIC model has the potential to embody the music preferences of a listener according to an emotional orientation, which is an element often neglected in favor of automation of song recommendation.

One area of our survey that we would like to have changed is the number of songs analyzed by each participant. Using twenty songs to generate each participant model may have been too small, and cause a misrepresentation of the MUSIC model's capability to model preferences for a wide range of individuals.

We also believe that there was a possibility of optimizing the model by manipulating the MUSIC factor values of the songs. While the untransformed data provided substantial results, we believe that there is more to be explored with this model by testing the sensitivity of the MUSIC factor values in modeling overall perception. We also believe that a larger sample size of songs would illustrate a more conclusive trend on the influence that the individual MUSIC factors have on preference. As shown in Figure 7, Unpretentious (U) was slightly more significant and Intense (I) was slightly less significant, although all five factors were fairly closely correlated. Data from a greater sample size of songs could either increase or decrease the uniformity observed across factors.

## Chapter 5: Studies 3 and 4: Model Development and Evaluation

### 5.1 *Methodology*

#### 5.1.1 Motivation

In order to develop a novel method of music recommendation that combined elements of automatic data processing and psychological music factors, we needed a program that would calculate concrete, mathematical values for songs that could then be translated into a corresponding psychological model. To combine both music perception data from the MUSIC Model (Rentfrow et al., 2011), a widely reviewed and accepted psychological model with calculated data from the content of the song, we needed to collect song feature data.

#### 5.1.2 Data Sets

At the start of our project, we planned to calculate musical feature values ourselves using signal processing in MATLAB. However, we later realized that this step of the research could be streamlined entirely through the use of *The Echo Nest*. *The Echo Nest* is a music intelligence company that provides customers an Application Programming Interface (API) which allows people to query their database through use of an API key. We registered for an account to receive an API key, which provided us with the means to obtain numerical values for different features of songs in their database.

Features in *The Echo Nest* API dataset include metadata, such as artist name and song name, along with data about energy, valence, tempo, and many other



features. Many of the features allowed for public use are: acousticness, danceability, duration, key, liveness, loudness, mode, speechiness, time signature (“Acoustic Attributes Overview”), artist id, artist name, song id, title, timbre, and pitch (Jehan & DesRoches, 2014). To build the balance between psychological and numerical music information we searched for expert opinions on music perception with respect to music recommendations. This led us to discover the MUSIC model, as presented by Rentfrow et al. in 2012.

The article provided information about a study which provided MUSIC – where M: Mellow, U: Unpretentious, S: Sophisticated, I: Intense, and C: Contemporary – factor loadings for a variety of songs related to psychological traits. We then manually inserted the song MUSIC factor values from Tables 1, 3, and 5 of Rentfrow et al., 2012, as well as those from an earlier article on the MUSIC model (Rentfrow et al., 2011), into an Excel file. The tables contained five Varimax-rotated principal components with the first table using song excerpts from a variety of genres, the third table using only jazz excerpts, and the fifth table using only rock excerpts. The table with various genres included: Avant-garde classical, Latin, Polka, Celtic, World beat, and Acid jazz. Once the information was compiled, we wrote a Python script to import this list of song data (song title and artist name) into *The Echo Nest* API to retrieve feature values (Figure 2 summarizes the methodology employed by the python script).

In order to gather data from The Echo Nest API about songs we created a python file, which along with the appropriate input file would output the song features we needed. We also retrieved an API key for The Echo Nest API giving us

permission to use their database. The input text file we used contained a list of songs and their artists. We used the Pyen Python library for *The Echo Nest*, which allowed us to call The Echo Nest API. We specifically searched for the “audio\_summary” information for each song present in the input file and from that we saved the song title, song ID (a tag for that specific version of the song in The Echo Nest API), tempo, danceability, energy, speechiness, liveness, acousticness, and valence to an output file. In this manner we collected all of the feature information we needed for the songs in the Rentfrow et al. article (2011, 2012) to create our quantitative data set.

The features searched for included: tempo, danceability, energy, speechiness, liveness, acousticness, and valence. . We chose these specific features for two main reasons. First, they are the features most correlated with those brought up during our Focus Groups (described in detail in Chapter 3). We wanted the features we chose to be those identified as most important in determining music preference and since they were mentioned in the Focus Groups we made sure to incorporate them in our model. Second, we chose the features that were most visible when working with the API and since these were located in the audio\_summary category they were more accessible. All of these features, except for tempo, take on values from 0.0 to 1.0 (*The Echo Nest Acoustic Attributes*). These Echo Nest features are described in detail in Table 4.

**Table 4.** Description of *The Echo Nest* features employed to predict the values of the MUSIC model. Descriptions come directly from *The Echo Nest* developers (Acoustic Attributes Overview)

<i>Echo Nest Feature</i>	<i>Description</i>
Tempo	<ul style="list-style-type: none"> <li>• Represents the average speed of the song</li> </ul>
Danceability	<ul style="list-style-type: none"> <li>• Describes the suitability of a track for dancing</li> <li>• Considers numerous elements including tempo, rhythm stability, beat strength, and regularity</li> </ul>
Energy	<ul style="list-style-type: none"> <li>• Represents intensity throughout the track</li> <li>• Contributing features include dynamic range, perceived loudness, timbre, and entropy</li> </ul>
Speechiness	<ul style="list-style-type: none"> <li>• Indicates the presence of spoken word in the track</li> </ul>
Liveness	<ul style="list-style-type: none"> <li>• Indicates the presence of an audience in the track</li> </ul>
Acousticness	<ul style="list-style-type: none"> <li>• Indicates the likelihood that the track was created by solely acoustic means</li> </ul>
Valence	<ul style="list-style-type: none"> <li>• Provides a description of the level of positivity portrayed by the musical features of the track</li> </ul>

The outputted values were contained in a spreadsheet next to the MUSIC factor loadings, for those that could be found on *The Echo Nest* API. We found numerous results from the same artist on certain song searches; however, overcame this by observing that the MUSIC factor values were similar and choosing the first result on the assumption that it was closest to the original. We were unable to find a portion of the songs in the articles within *The Echo Nest* API and, therefore, could not retrieve *The Echo Nest* feature values for those songs so these songs were excluded from further use in our study. There is a feature in *The Echo Nest* API where we could have uploaded song files for analysis that were not in their database; however, we opted against this since we did not have the original song files that were not found.

After this process, we realized that approximately 100 data points was not sufficient to build a model. In order to build a larger dataset we collected more *The Echo Nest* features using songs from Tables 1, 2, and 3 in Rentfrow et al.'s 2011 article. This brought the total number of songs in our dataset to 161. A complete list of the songs found from the Rentfrow et al. articles that was used in our analysis (2011, 2012) can be found in in the Appendix C1. The tables described five Varimax-rotated principal components from studies with songs from various genres. We were cautious to compare the MUSIC factor loadings from this article to that of the 2012 article because the definitions of the MUSIC factors had changed. In the 2011 article, the categories are I: Mellow, II: Urban, III: Sophisticated, IV: Intense, and V: Campestral. In terms of the 2012 article, I corresponds to S, II corresponds to U, III corresponds to I, IV corresponds to C, and V corresponds to M (Rentfrow et al., 2012). This collected data lent itself to be input into a machine learning tool in order to generate a computationally-defined cognitive model of music preference.

### 5.1.3 Weka Machine Learning Environment

We learned that it would not be feasible for us to create a model integrating the MUSIC factors and *The Echo Nest* values without the use of a machine learning tool. Machine learning is a process in which a computer program studies the patterns present in an existing body of data and is then able to make predictions about new data. We eventually discovered Weka, an open source machine learning program. We studied this machine learning tool (starting in the Spring of 2015) by watching online

tutorials by Professor Ian Witten from the University of Waikato, New Zealand (Witten, 2013).

As we observed the videos and the example machine learning techniques employed in them, we came to realize that Weka is a great resource for creating algorithms to make predictions. We used Weka by providing it with *The Echo Nest* feature values for songs, and having the program predict the corresponding MUSIC values.

We created a CSV file that contained all of the MUSIC values and *The Echo Nest* values from over 100 songs in the articles “The Song Remains the Same” and “The Structure of Musical Preferences: A Five-Factor Model” (Rentfrow et al., 2011; Rentfrow et al., 2012). We excluded the twenty songs that were used in our online survey from this file in order to have a data set that the final model could be tested on for accuracy. Our plan was to see how well our constructed Echo Nest - MUSIC model would be able to predict the MUSIC values of the songs we excluded from the input data set. In addition, we planned to observe how the machine learning generated values aligned with the music preference data collected for 100 participants in the online survey.

#### 5.1.4 Approach to Machine Learning Analysis

Initially, using various modeling tools in Weka, such as the J48 and ZeroR classifiers, we collected baseline data giving predictions for mapping MUSIC factors based on *The Echo Nest* features. The J48 and ZeroR classifiers are meant for classification problems (categorizing nominal input groups into nominal output groups), and what we needed instead was a regression classifier, which takes in

numerical values, and outputs predicted numerical values. Thus, we tried out several different regression classifiers to predict the MUSIC values for the songs.

We tested the Multilayer Perceptron, RBF Network, SMOreg, Linear Regression, and LeastMedSq classifiers in Weka. Several of these classifiers, namely Multilayer Perceptron and SMOreg were recommended to us by machine learning professor Dr. Vurkac at the Society for Music Perception and Cognition (SMPC) conference where we presented our research poster in the Summer of 2015. Other machine learning tests proposed by various interested parties at the conference involved SVM, PNN, Genetic Algorithms, and ID3 - however, these options were not available as classifiers in Weka.

Our procedure in the initial testing of the classifiers was as follows:

1. Upload the 161 song dataset, containing *The Echo Nest* values and MUSIC values into Weka.
2. Select the MUSIC attribute of interest, and remove all others. For example, keep “M” and remove “U”, “S”, “I”, and “C”. (This was necessary as Weka is able to predict only one attribute at a time.)
3. Select the classifier of interest.
4. Run the regression classifier.

We used a 10-fold cross validation as our test option in order for Weka to train the model on a different parts of the data set as it constructed the predictive algorithm for a MUSIC value. With each run of the program, Weka gave us a summary of the performance of the classifier giving a variety of error measures. This showed how

well the algorithm it produced was able predict the MUSIC factor values using the given *The Echo Nest* features.

## **5.2 Results**

### 5.2.1 Model Development Approach

We tested our input consisting of 161 songs, with attached MUSIC factors and *The Echo Nest* features in Weka as described in the above Methodology section (see Appendix C1 for complete list of songs). We systematically tested many of the regression classifiers available in the Weka Explorer application in order to see which classifiers would produce the lowest error values in response to the test set given. The classifiers we chose were Multilayer Perceptron, SMOreg, Linear Regression, Pace Regression, Gaussian Processes, and Isotonic Regression. In order to evaluate these classifiers, we measured the error by looking at the mean squared error values and correlation coefficient values for each model that Weka generated. These error values are standard for determining the degree of accuracy of a model (Armstrong & Callopy, 1992). In general, low mean squared error values (high correlation coefficient values) indicate that the values the classifier is able to predict are close to the actual values from the dataset upon which the program constructs its model. We noticed, as expected, the values with the lowest mean squared error values also, for the most part, had the largest correlation coefficient values. From the error values provided by Weka, we were able to compare the results of models using different classifiers within Weka.

### 5.2.2 Classifier Determination

Our output of the error values for different classifiers and different combinations of *The Echo Nest* features in are shown in Tables 5-9 below, where ‘Baseline’ represents using all of our desired *The Echo Nest* features - tempo, danceability, energy, speechiness, liveness, acousticness, and valence - in the prediction. Whereas ‘No Tempo’ or ‘No ...’ represents running the test on all of the features except the one listed after ‘No’. We took out different *The Echo Nest* features to observe the how the error changed with varying combinations of the features to check that all features were necessary as we conjectured from the Focus Group sessions in Chapter 3. The classifiers with the smallest error values were for Gaussian Processes (M, S, I, C) and Isotonic Regression (U), as seen in Tables 5-9, where the numbers in these tables take values between 0 and 1 representing the percent error in terms of correlation coefficient and root mean squared error. Overall, we decided that using all of *The Echo Nest* features (‘Baseline’) and allowing the model to reduce weighting of a certain feature for a certain MUSIC value would be the best approach since the error values in the last column of Tables 5-9 did not change significantly from that of the ‘Baseline’ column.



**Table 5.** M factor prediction errors as indicated by correlation coefficient (CC) and root mean squared error (RMSE). The highest correlation coefficient was 0.6286 and lowest root mean squared error was 0.1981, both using the Gaussian Processes Classifier.

	<i>Baseline</i>	<i>No Tempo</i>	<i>No Danceability</i>	<i>No Energy</i>	<i>No Speechiness</i>	<i>No Liveness</i>	<i>No Acousticness</i>	<i>No Valence</i>	<i>Min MSE/Max Correlation Coefficient</i>
<b>Multilayer Perceptron</b>									
CC	0.5157	0.5256	0.5157	0.4413	0.4881	0.5758	0.5306	0.5137	0.5758
RMSE	0.2539	0.2326	0.2372	0.2496	0.2534	0.23	0.2526	0.2516	0.23
<b>Linear Regression</b>									
CC	0.6017	0.5916	0.597	0.4986	0.5281	0.6029	0.5908	0.552	0.6029
RMSE	0.2033	0.205	0.204	0.2209	0.2162	0.203	0.2052	0.2124	0.203
<b>Gaussian Processes</b>									
CC	0.6148	0.5946	0.6095	0.5498	0.5632	0.6286	0.6009	0.5876	0.6286
RMSE	0.2008	0.2045	0.2018	0.2126	0.2101	0.1981	0.2035	0.2062	0.1981
<b>Isotonic Regression</b>									
CC	0.509	0.509	0.509	0.509	0.4349	0.509	0.509	0.509	0.509
RMSE	0.2194	0.2194	0.2194	0.2194	0.2301	0.2194	0.2194	0.2194	0.2194

**Table 6.** U factor prediction errors as indicated by correlation coefficient (CC) and root mean squared error (RMSE). The highest correlation coefficient was 0.3811 and lowest root mean squared error was 0.2349, both using the Isotonic Regression Classifier.

	<i>Baseline</i>	<i>No Tempo</i>	<i>No Danceability</i>	<i>No Energy</i>	<i>No Speechiness</i>	<i>No Liveness</i>	<i>No Acousticness</i>	<i>No Valence</i>	<i>Min MSE/Max Correlation Coefficient</i>
<b>Multilayer Perceptron</b>									
CC	0.1614	0.2787	0.1856	0.195	0.189	0.164	0.2223	0.2242	0.2787
RMSE	0.3291	0.2849	0.3027	0.3042	0.2894	0.3132	0.2985	0.2917	0.2849
<b>Linear Regression</b>									
CC	0.3041	0.3197	0.2565	0.2973	0.2534	0.2608	0.3098	0.2437	0.3197
RMSE	0.2425	0.2408	0.2464	0.2429	0.2461	0.2463	0.2418	0.2471	0.2408
<b>Gaussian Processes</b>									
CC	0.3223	0.3515	0.3122	0.3144	0.2679	0.3185	0.3207	0.3129	0.3515
RMSE	0.2402	0.2372	0.241	0.2408	0.2449	0.2405	0.2401	0.2407	0.2372
<b>Isotonic Regression</b>									
CC	0.3811	0.3811	0.3811	0.3811	0.3811	0.3811	0.0681	0.3811	0.3811
RMSE	0.2349	0.2349	0.2349	0.2349	0.2349	0.2349	0.2669	0.2349	0.2349

**Table 7.** S factor prediction errors as indicated by correlation coefficient (CC) and root mean squared error (RMSE). The highest correlation coefficient was 0.5755 and the lowest root mean squared error was 0.2176 both using the Gaussian Processes Classifier.

	<i>Baseline</i>	<i>No Tempo</i>	<i>No Danceability</i>	<i>No Energy</i>	<i>No Speechiness</i>	<i>No Liveness</i>	<i>No Acousticness</i>	<i>No Valence</i>	<i>Min MSE/Max Correlation Coefficient</i>
<b>Multilayer Perceptron</b>									
CC	0.3662	0.4542	0.3343	0.3779	0.4839	0.4325	0.3963	0.3988	0.4839
RMSE	0.2867	0.2568	0.2887	0.2752	0.2438	0.2558	0.2687	0.27	0.2438
<b>Linear Regression</b>									
CC	0.5638	0.5137	0.5266	0.5599	0.5638	0.5638	0.537	0.5116	0.5638
RMSE	0.2185	0.2275	0.2252	0.2192	0.2185	0.2185	0.2233	0.2274	0.2185
<b>Gaussian Processes</b>									
CC	0.5673	0.5553	0.5238	0.5488	0.566	0.5658	0.561	0.5419	0.5673
RMSE	0.2176	0.2197	0.2253	0.2209	0.2179	0.2178	0.2186	0.222	0.2176
<b>Isotonic Regression</b>									
CC	0.4819	0.4819	0.4819	0.4819	0.4819	0.4819	0.4157	0.4819	0.4819
RMSE	0.2337	0.2337	0.2337	0.2337	0.2337	0.2337	0.2424	0.2337	0.2337

**Table 8.** I factor prediction errors as indicated by correlation coefficient (CC) and root mean squared error (RMSE). The highest correlation coefficient was 0.8228 and lowest root mean squared error was 0.1681, both using the Gaussian Processes Classifier.

	<i>Baseline</i>	<i>No Tempo</i>	<i>No Danceability</i>	<i>No Energy</i>	<i>No Speechiness</i>	<i>No Liveness</i>	<i>No Acousticness</i>	<i>No Valence</i>	<i>Min MSE/Max Correlation Coefficient</i>
<b>Multilayer Perceptron</b>									
CC	0.6607	0.695	0.6702	0.5024	0.6727	0.6298	0.6987	0.6454	0.6987
RMSE	0.2586	0.2342	0.2622	0.3226	0.2425	0.2575	0.2278	0.2539	0.2278
<b>Linear Regression</b>									
CC	0.7501	0.7461	0.7071	0.6961	0.7398	0.7375	0.7551	0.7384	0.7551
RMSE	0.1942	0.1955	0.2075	0.2107	0.1975	0.1983	0.1924	0.198	0.1924
<b>Gaussian Processes</b>									
CC	0.8143	0.8228	0.7915	0.7639	0.8025	0.8127	0.8034	0.8014	0.8228
RMSE	0.1713	0.1681	0.1803	0.1907	0.1759	0.1717	0.1759	0.1763	0.1681
<b>Isotonic Regression</b>									
CC	0.7774	0.7774	0.7774	0.7144	0.7774	0.7774	0.7774	0.7774	0.7774
RMSE	0.1849	0.1849	0.1849	0.2063	0.1849	0.1849	0.1849	0.1849	0.1849

**Table 9.** C factor prediction errors as indicated by correlation coefficient (CC) and root mean squared error (RMSE). The highest correlation coefficient was 0.5342 and lowest root mean squared error was 0.1650, both using the Gaussian Processes Classifier.

	<i>Baseline</i>	<i>No Tempo</i>	<i>No Danceability</i>	<i>No Energy</i>	<i>No Speechiness</i>	<i>No Liveness</i>	<i>No Acousticness</i>	<i>No Valence</i>	<i>Min MSE/Max Correlation Coefficient</i>
<b>Multilayer Perceptron</b>									
CC	0.3377	0.2305	0.3514	0.3091	0.4158	0.465	0.4298	0.3014	0.465
RMSE	0.223	0.2266	0.2101	0.2386	0.2073	0.1862	0.1888	0.2028	0.1862
<b>Linear Regression</b>									
CC	0.498	0.4519	0.3026	0.5101	0.4934	0.498	0.5149	0.498	0.5149
RMSE	0.1694	0.1742	0.1864	0.1678	0.1697	0.1694	0.1672	0.1694	0.1672
<b>Gaussian Processes</b>									
CC	0.5283	0.5069	0.4312	0.521	0.5047	0.5307	0.5228	0.5342	0.5342
RMSE	0.1657	0.1683	0.176	0.1666	0.1683	0.1654	0.1664	0.165	0.165
<b>Isotonic Regression</b>									
CC	0.3985	0.3985	0.1906	0.3985	0.3985	0.3985	0.3985	0.3985	0.3985
RMSE	0.1813	0.1813	0.1965	0.1813	0.1813	0.1813	0.1813	0.1813	0.1813

Since a few of the classifiers had similar error values, we decided to experiment with different mixes of classifiers to find an overall combination that would produce optimal results for predicting MUSIC values. First we collected the models for them using the classifiers that resulted in the lowest mean squared error and highest correlation coefficient for each of the MUSIC factors, resulting in six models (due to the overlap of some lowest mean squared error occurrences with high correlation coefficient occurrences). After collecting this model information, we decided to choose the next best classifier that had a linear type model equation so that we could implement the models ourselves outside of Weka. The classifiers for these results included: Pace Regression, SMOreg, and Linear Regression.

After using the equations from the linear-like models to create predictions, we observed that the Gaussian Processes Classifier and the Isotonic Regression Classifier would better suit the needs of our predictor. The Gaussian Processes Classifier,

having the least error in general over all of the MUSIC factors, was then used to model new MUSIC factors. The Isotonic Regression Classifier was also chosen due to its ability to tie a MUSIC factor to its most important and telling *The Echo Nest* feature. The Isotonic Regression ties between MUSIC factors and *The Echo Nest* features were that: M was based on Speechiness, U was based on Acousticness, S was based on Acousticness, I was based on Energy, and C was based on Danceability (see Table 10).

We also included the linear regression model, shown in Figure 8, for each of the factors to get an alternate visualization of the weights of the various *The Echo Nest* features. Here the most positively correlated *The Echo Nest* features associated with MUSIC factors were: Danceability with M, Danceability with U, Acousticness with S, Speechiness with I, and Danceability with C.

**Table 10.** MUSIC factor Isotonic Regression predictions

<i>Song Title</i>	<i>Predicted M</i>	<i>Predicted U</i>	<i>Predicted S</i>	<i>Predicted I</i>	<i>Predicted C</i>
	Based on <b>Speechiness</b>	Based on <b>Acousticness</b>	Based on <b>Acousticness</b>	Based on <b>Energy</b>	Based on <b>Danceability</b>
Just Walk Away	0.422	0.277	0.353	0.015	0.08
Unspeakable	0.247	0.095	0.065	0.117	0.263
Brown Baby	0.108	0.277	0.248	0.086	0.117
If You Only Knew	0.422	0.277	0.201	0.117	0.117
Slate	0.422	0.277	0.201	0.066	0.117
Symphony No. 1 in B Flat Major	0.247	0.285	0.5	0.014	0.117
The Way You Look Tonight	0.108	0.285	0.435	0.014	0.21
Rock the Clock	-0.041	0.285	0.435	0.384	0.095
Where Eagles Dare	0.108	0.065	0.046	0.667	0.04
Oh No the Radio	0.247	0.046	0.046	0.19	0.117
City of Gold	0.108	0.277	0.201	0.384	0.163
Intro	-0.041	0.277	0.248	0.117	0.263
My Favorite Polka	0.247	0.277	0.201	0.074	0.263
Interstate Rag	0.108	0.285	0.435	0.117	0.095
Big Blue Sun	0.247	0.277	0.248	0.066	0.117

**Figure 8.** MUSIC factor Linear Regression models.

$$\mathbf{M} = 0.0013 * \mathbf{Tempo} + 0.2493 * \mathbf{Danceability} + -0.5502 * \mathbf{Energy} + -1.4067 * \mathbf{Speechiness} + -0.2048 * \mathbf{Acousticness} + -0.2357 * \mathbf{Valence} + 0.5347$$

$$\mathbf{U} = 0.3513 * \mathbf{Danceability} + -0.2089 * \mathbf{Energy} + -0.7 * \mathbf{Speechiness} + 0.2058 * \mathbf{Liveness} + 0.2313 * \mathbf{Valence} + 0.0695$$

$$\mathbf{S} = -0.0021 * \mathbf{Tempo} + -0.3738 * \mathbf{Danceability} + 0.3173 * \mathbf{Acousticness} + 0.2053 * \mathbf{Valence} + 0.4834$$

$$\mathbf{I} = 0.001 * \mathbf{Tempo} + -0.5668 * \mathbf{Danceability} + 0.6586 * \mathbf{Energy} + 1.044 * \mathbf{Speechiness} + -0.1953 * \mathbf{Liveness} + -0.1764 * \mathbf{Valence} + 0.0459$$

$$\mathbf{C} = -0.0013 * \mathbf{Tempo} + 0.5846 * \mathbf{Danceability} + -0.1049 * \mathbf{Acousticness} + 0.065$$

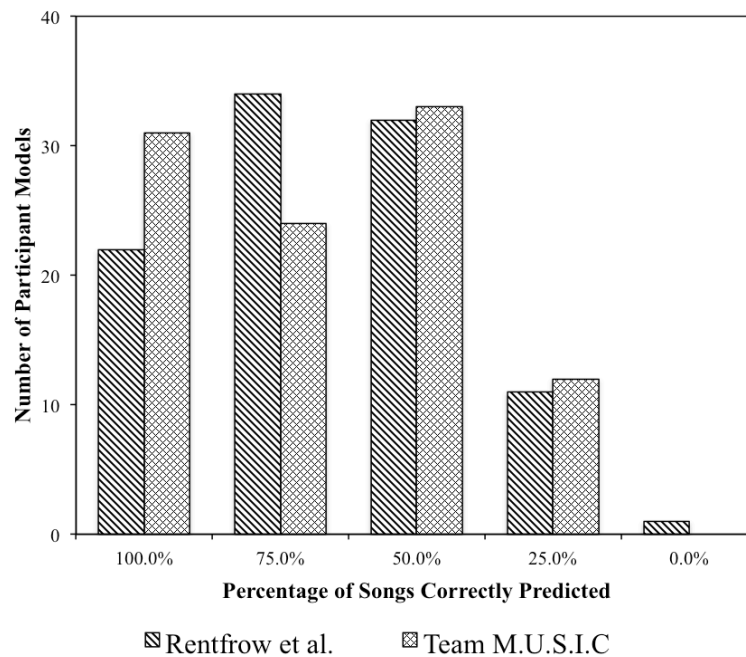
### 5.2.3 Study 4: Model Evaluation

Once we had tested the ability of these models to generate MUSIC factors we used them to create a new set of MUSIC factors for the survey songs described in Chapter 4 selected from Rentfrow et al., 2011 which were absent in the training set made to create the models. The predictions based on these classifiers are in Appendix C2.

Then, we used the predictions in conjunction with the preferences in the survey to attempt a prediction of preference from our predicted MUSIC factors. We used the Multilayer Perceptron Classifier on eleven of fifteen survey songs from Chapter 4 to build a model for music preference prediction. Only fifteen of the twenty survey songs were used because the rest could not be found in *The Echo Nest* database (see Table 3). Furthermore, only eleven of those fifteen were chosen so that we could specify a test set to judge the accuracy of the preference predictions. The preferences were represented by converting the numerical likings from the survey into a binary ‘yes’ and ‘no’ representation.

After we gathered data on how the Rentfrow et al. (2011) MUSIC model handled predicting preferences on its own as well as how our predicted MUSIC model using *The Echo Nest* data handled preference prediction we compared the

results in the Figure 9. Of the four songs we tested on, we were able to successfully predict all four of them with our model for approximately 31 of the participants, whereas the MUSIC model had that success rate for about 22 of the participants. We also had no case of incorrectly predicting all four preference likings, whereas the MUSIC model had approximately one case of this occurrence. The disparity in results of the MUSIC model and our model could have to do with the MUSIC model being better suited for a larger dataset, and therefore potentially more generalizable than our model based on fewer recordings. In general, our model had results similar to the MUSIC model.



**Figure 9.** Comparison of percentage of songs classified correctly using the Rentfrow et al. model and our model.

### 5.3 Discussion

The Isotonic Regression predictions based on *The Echo Nest* features showed the connection between each MUSIC factor and a specific *The Echo Nest* feature with which it was most correlated. The M factor was based on Speechiness, the U factor was based on Acousticness, the S factor was based on Acousticness, the I factor was based on Energy, and the C factor was based on Danceability. The Mellow factor was defined as “romantic, relaxing, and slow” (Rentfrow et al., 2012), whereas Speechiness was defined as having a certain percent of spoken words ("*The Echo Nest* Acoustic Attributes"). The connection between these descriptors could mean that songs considered relaxing have more speech than those that are not; however, this connection was the most tenuous of the group. The Unpretentious factor described “uncomplicated, unaggressive, soft sounding, and acoustic” songs (Rentfrow et al., 2012) and the Acousticness feature that it was matched up with represents the degree to which a sound was created by using acoustic instruments or voice ("*The Echo Nest* Acoustic Attributes"). Acoustic sound has been described as soft (and the definition of the U factor included acoustic as a descriptor), and therefore this connection seems like a natural choice. The Sophisticated factor was used for “intelligent, complex, and cultured” music (Rentfrow et al., 2012), and its match was also Acousticness. Although the U and S factors seem to differ in terms of complexity, Acousticness does also seem to fit for S because orchestral instruments have high values for Acousticness and since the S factor songs typically include classical type music this seems appropriate. The Intense factor was used for “loud, tense, and aggressive” music (Rentfrow et al., 2012) and it was tied to the Energy feature which is described

as measuring the intensity of a piece tied to examples of “fast, loud, and noisy” tracks (*The Echo Nest Acoustic Attributes*). The Contemporary factor described music in terms of it being “current, rhythmic, and danceable” (Rentfrow et al., 2012) and it was matched with Danceability, which is supposed to tell how musical elements such as “rhythm stability, beat strength, and overall regularity” intertwine to make a song that people could find danceable (*The Echo Nest Acoustic Attributes*). It is also interesting to note that all of these Isotonic Regression connected features are considered acoustic attributes by *The Echo Nest* which model subjectivity using values from 0.0 to 1.0 (*The Echo Nest Acoustic Attributes*).

The Linear Regression models for prediction the MUSIC factors from *The Echo Nest* features gave more insight into *The Echo Nest* features that were most positively correlated and most negatively correlated to a given factor (see Figure 8 for the full equations). It is important to note that although these equations give relations between the factors and features, this classifier was not the best performing from Tables 5-9, so these are not meant to judge the other classifiers but rather bolster them when it agrees. Agreements with the Isotonic Regression classifier are shown in the positive correlations for the S and C factors, as shown in Table 11.

The error in our predictions of MUSIC factors using Gaussian Processes with *The Echo Nest* features were shown when we ran the percentage and paired t-tests seen in Table 12. The error values, more specifically in the paired t-test values in the bottom row were not statistically significant giving us confidence that our predictions are comparable to the results of Rentfrow et al. Overall, the percentage error also



**Table 11.** Most positive and most negative correlations of MUSIC factors to *The Echo Nest* features with bolded showing agreement with Isotonic Regression Classifier.

Factor	Positive Correlation	Negative Correlation
M	Danceability (+0.2493)	Speechiness (-1.4067)
U	Danceability (+0.3513)	Speechiness (-0.7000)
S	<b>Acousticness</b> (+0.3173)	Danceability (-0.3738)
I	Speechiness (+1.044)	Danceability (-0.5668)
C	<b>Danceability</b> (+0.5846)	Acousticness (-0.1049)

stayed within about 24% where the error in I seemed to be the smallest. This could have to do with the Intense factor having such a solid relation to *The Echo Nest* features used.

Our models were limited by the classifiers we chose: Gaussian Processes, Isotonic Regression, Multilayer Perceptron, and Linear Regression. Due to the limited timeframe of our project we chose classifiers that were more accessible in terms of creation of models and interpretation of results. However, given more time we could have explored other machine learning classifiers better suited for our dataset size.

Although we performed a thorough modeling of our data, our results are limited by the size of our initial dataset of 161 songs. It is preferable to have a larger data set in order to make better predictions. We attempted to combat this by using cross-fold validation and making sure to keep a small subset of our data unseen from our model to test on later. Our results would be stronger and more generalizable had we used a larger dataset.

**Table 12.** Evaluation of our ability to reproduce MUSIC model values. Percent change is calculated by the difference in predicted value and original loading divided by 2, the total possible range for the factor loadings.

<i>Song</i>	<i>Percent Change in M</i>	<i>Percent Change in U</i>	<i>Percent Change in S</i>	<i>Percent Change in I</i>	<i>Percent Change in C</i>
Just Walk Away	-11.55%	-4.00%	6.85%	-6.85%	4.10%
Unspeakable	-20.80%	6.20%	1.70%	-2.35%	15.25%
Brown Baby	-12.85%	-5.05%	3.20%	-1.85%	2.90%
If You Only Knew	-0.75%	-16.55%	6.95%	2.40%	5.35%
Slate	18.75%	-23.40%	5.55%	-8.00%	6.80%
Symphony No. 1 in B Flat Major	0.65%	18.25%	-14.30%	-0.55%	13.20%
The Way You Look Tonight	5.50%	15.00%	-18.85%	1.30%	1.65%
Rock the Clock	15.30%	2.10%	-13.55%	6.50%	-10.70%
Where Eagles Dare	1.25%	-4.00%	6.95%	-2.30%	0.60%
Oh No the Radio	3.45%	9.00%	3.55%	-11.60%	-1.70%
City of Gold	-3.40%	-6.30%	3.45%	-9.65%	10.70%
Intro	0.00%	8.65%	7.90%	-0.65%	-17.75%
My Favorite Polka	14.15%	-0.45%	-18.80%	-4.35%	7.30%
Interstate Rag	2.60%	-14.85%	6.20%	5.40%	0.65%
Big Blue Sun	8.35%	7.05%	6.85%	-16.70%	8.80%
<i>p-value</i>	0.628	0.857	0.874	0.066	0.180

## Chapter 6: Conclusion

The MUSIC model demonstrates that by using features selected from *The Echo Nest*, a dataset of acoustic information can be used to describe an individual's perception of music, which suggests that musical taste is linked to musicality. The overall results obtained from the study indicate that the MUSIC model is a viable tool in prediction of music preference. The MUSIC model's perception-based, psychological approach allowed for statistically significant correlations to be drawn between its five-factors and an individual's tastes. With access to a sufficiently large dataset, these correlations could then be used for generating music recommendations.

Our focus groups at the initial stages of research were used as a measure for determining what aspects of songs had significant meaning to the users of music recommendation services, as well as how they felt about the recommendation services that they had previously used. The noteworthy levels of dissatisfaction with extant recommendation systems demonstrated that there is certainly room for improvement. The participants indicated that they mostly valued broad, holistic traits of songs – such as instrumentation and perceived genre (see Chapter 3 Focus Groups section for relevant data). An analysis of the data collected from the online survey allowed us to draw connections between an individual's indicated musical preferences and the psychological factors of the MUSIC model. The mathematical values that are calculated by *The Echo Nest* are then used to predict the psychological factors of the MUSIC model. In this fashion it is possible to incorporate a psychological

perspective into the recommendation system – all while keeping free of ambiguity in calculations – through the machine learning capabilities of Weka.

This demonstrates a confirmation of our original hypothesis, that the integration of an effect model of music perception would allow for improved music recommendations when combined with current systems. It was proven in our study that the MUSIC factors showed significant correlations to an individual's taste. These results help to fill the existing gap between automated recommendations of songs and users' perception of music, which went previously unfulfilled by popular recommendation systems.

This psychological preference model cannot simply generate music recommendations at its current state. The MUSIC model, as a representative type of data, is limited because it is only derived from acoustic information – but combined with other types of data, it can present musical information in a contrasting and complementary way. The strength of the correlations between the MUSIC factors and survey participants' preferences are non-negligible, but not reliable enough, indicating the model's potential as an auxiliary recommendation tool. The five-factors from the MUSIC model are applicable as a high-level categorization feature, and be used to guide more popular metadata models in a direction that is more grounded in musical psychology and emotional response. The most effective way to use the MUSIC model for song recommendation is by employing it in conjunction with other recommendation systems or techniques. In this way, the model serves as an augmentation to the extant methods of recommendation.

## **6.1 Future Research**

We believe the MUSIC model, if integrated into an existing recommendation system, would allow for recommendation results that are more closely based on facets of musical perception. By using this method, the recommendations are not overly simplified, based solely on the musical content (i.e., tempo or key signature) – which may be the case in Spotify’s recommendation system – and are also still viable for automated processing of millions of songs, thereby forgoing the processing limitations observed by Pandora’s recommendation method.

To incorporate the MUSIC model into an existing system, one would need to calculate the MUSIC values of songs by inputting those songs’ *Echo Nest* values into Weka. Based on these newly generated MUSIC values, the system can then compare and organize songs using the MUSIC factors as an additional basis. These MUSIC factors can assist the user by acting as a functional categorization of songs. For example, the user might want to listen to songs from a certain genre, such as rock, and also wants a song with a high “Intense” aspect and low “Contemporary” aspect from the MUSIC model – the recommendation program, assuming it had information regarding genre, would first examine the body of songs that fall under the category of “rock”, and then further fine-tune its selection by seeing which rock songs have those loadings of MUSIC values. This specificity would likely lead to results that are more representative of the user’s interest, as it allows the user to pick a type of music that fits the general sense of what they want to listen to.

In essence, the MUSIC model becomes a valuable tool for any song search engine, since it allows the user to discover songs that best represent how they perceive music. The model may not be fully optimized to act as a standalone recommendation system, but it is certainly capable of providing contextual information for use by music recommendation systems. We encourage future research to examine the use of the MUSIC model as a supplemental tool for recommendation systems.

## References

- About the Music Genome Project*. (n.d.). Retrieved March 31, 2016, from <http://www.pandora.com/corporate/mgp.shtml>
- Acoustic Attributes Overview*. (n.d.). Retrieved December 2014, from <http://developer.echonest.com/acoustic-attributes.html>
- Adegoke, Y. (2014, May 21). Spotify Now Has 10 Million Paid Subscribers, 3 Million In U.S. (Exclusive). Retrieved March 31, 2016, from <http://www.billboard.com/biz/articles/news/digital-and-mobile/6092226/spotify-now-has-10-million-paid-subscribers-3-million>
- Armstrong, J., & Collopy, F. (1992). "Error Measures For Generalizing About Forecasting Methods: Empirical Comparisons" (PDF). *International Journal of Forecasting* 8 (1): 69–80.
- Barrington, L., Oda, R., & Lanckriet, G. (2009). SMARTER THAN GENIUS? HUMAN EVALUATION OF MUSIC RECOMMENDER SYSTEMS. *ISMIR*, 9, 357-362. Retrieved March 10, 2016.
- Bertin-Mahieux, T., Ellis, D. P., Whitman, B., & Lamere, P. (2011). The Million Song Dataset. *Columbia University Academic Commons*.
- Bogdanov, D. & Herrera, P. (2011). How much metadata do we need in music recommendation? A subjective evaluation using preference sets. Retrieved from <http://ismir2011.ismir.net/papers/PS1-10.pdf>
- Brown, R. A. (2012). Music preferences and personality among Japanese university students. *International Journal of Psychology*, 47(4), 259-268.

- Celma, O. (2010). *Music recommendation and discovery: The long tail, long fail, and long play in the digital music space*. Heidelberg: Springer.
- Cremonesi, P., Garzotto, F., Negro, S., Papadopoulos, A. V., & Turrin, R. (2011). Looking for “Good” Recommendations: A Comparative Evaluation of Recommender Systems. *Human-Computer Interaction – INTERACT 2011 Lecture Notes in Computer Science*, 152-168.
- Delsing, M. J., Bogt, T. F., Engels, R. C., & Meeus, W. H. (2008). Adolescents' music preferences and personality characteristics. *Eur. J. Pers. European Journal of Personality*, 22(2), 109-130.
- Ellis, D. (2006, November 21). Matlab Audio Processing Examples. Retrieved March 31, 2016, from <http://labrosa.ee.columbia.edu/matlab/>
- Ellis, D. P., Whitman, B., Jehan, T., & Lamere, P. (2010). *The Echo Nest musical fingerprint. ISMIR 2010 Utrecht: 11th International Society for Music Information Retrieval Conference, August 9th-13th, 2010*. International Society for Music Information Retrieval.
- George, D., Stickle, K., Rachid, F., & Wopnford, A. (2007). The association between types of music enjoyed and cognitive, behavioral, and personality factors of those who listen. *Psychomusicology: A Journal of Research in Music Cognition*, 19(2), 32-56.
- Getting the tempo, key signature, and other audio attributes of a song*. (n.d.). Retrieved December 2014, from [http://developer.echonest.com/raw\\_tutorials/faqs/faq\\_04.html](http://developer.echonest.com/raw_tutorials/faqs/faq_04.html).
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: An Update. *SIGKDD Explor. Newsl. ACM SIGKDD Explorations Newsletter*, 11(1), 10.



- Istok, E., Brattico, E., Jacobsen, T., Ritter, A., & Tervaniemi, M. (2013). 'I love Rock 'n' Roll' - Music genre preference modulates brain responses to music. *Biological Psychology*, 92(2), 142-151.
- Jehan, T., & DesRoches, D. (2014, January 7). Analyzer Documentation. Retrieved December 2014, from [http://developer.echonest.com/docs/v4/\\_static/AnalyzeDocumentation.pdf](http://developer.echonest.com/docs/v4/_static/AnalyzeDocumentation.pdf).
- Langmeyer, A., Guglhör-Rudan, A., & Tarnai, C. (2012). What Do Music Preferences Reveal About Personality? *Journal of Individual Differences*, 33(2), 119-130.
- Laplante, A. (2014). Improving music recommender systems: What can we learn from research on music tastes? *ISMIR*, 451-456. Retrieved March 10, 2016.
- Lartillot, O., & Toiviainen, P. (2007). A Matlab Toolbox for Musical Feature Extraction From Audio. In *International Conference on Digital Audio Effects*. Bordeaux.
- Magno, T., & Sable, C. (2008). A Comparison of Signal Based Music Recommendation to Genre Labels, Collaborative Filtering, Musicological Analysis, Human Recommendation and Random Baseline. *ISMIR 2008, Session 2a – Music Recommendation and Organization*.
- Morrison, S. J., & Demorest, S. M. (2009). *Cultural Constraints on Music Perception and Cognition* (Vol. 178) (J. Y. Chiao, Ed.). Amsterdam, The Netherlands: Elsevier.
- Peoples, G. (2011, February 18). Business Matters: 75,000 Albums Released In U.S. In 2010 -- Down 22% From 2009. Retrieved March 2016, from <http://www.billboard.com/biz/articles/news/1179201/business-matters-75000-albums-released-in-us-in-2010-down-22-from-2009>

- Qualtrics Survey Tool - University of Maryland - MIM Central. (n.d.). Retrieved March 31, 2016, from <http://mim.umd.edu/qualtrics-survey-tool>
- Rentfrow, P. J., Goldberg, L. R., & Levitin, D. J. (2011). The structure of musical preferences: A five-factor model. *Journal of Personality and Social Psychology*, *100*(6), 1139-1157.
- Rentfrow, P. J., Goldberg, L. R., Stillwell, D. J., Kosinski, M., Gosling, S. D., & Levitin, D. J. (2012). The Song Remains the Same: A Replication and Extension of the MUSIC Model. *Music Perception: An Interdisciplinary Journal*, *30*(2), 161-185.
- Rentfrow, P. J., & Gosling, S. D. (2003). The do re mi's of everyday life: The structure and personality correlates of music preferences. *Journal of Personality and Social Psychology*, *84*(6), 1236-1256.
- Schafer, T., & Sedlmeier, P. (2009). From the functions of music to music preference. *Psychology of Music*, *37*(3), 279-300.
- Schellenberg, E. G., & Habashi, P. (2015). Remembering the melody and timbre, forgetting the key and tempo. *Mem Cogn Memory & Cognition*, *43*(7), 1021-1031.
- Schindler, A., & Rauber, A. (2014). Capturing the Temporal Domain in Echonest Features for Improved Classification Effectiveness. *Adaptive Multimedia Retrieval: Semantics, Context, and Adaptation Lecture Notes in Computer Science*, 214-227.
- Seyerlehner, K. (2010). *Ontent-based music recommender systems: Beyond simple frame-level audio similarity* (Unpublished doctoral dissertation). Johannes Kepler University Linz.
- Seyerlehner, K., Widmer, G., & Knees, P. (2008). Frame level audio similarity-a codebook approach. *Proc. of the 11th Int. Conf. on Digital Audio Effects (DAFx-08)*.

Soleymani, M., Aljanaki, A., Wiering, F., & Veltkamp, R. C. (2015). Content-based music recommendation using underlying music preference structure. *2015 IEEE International Conference on Multimedia and Expo (ICME)*. Retrieved March 11, 2016.

*Spotify Acquires The Echo Nest*. (2014). Retrieved November 2015, from <https://press.spotify.com/us/2014/03/06/spotify-acquires-the-echo-nest/>.

*Spotify Echo Nest API*. (n.d.). Retrieved October 2015, from <https://developer.spotify.com/spotify-echo-nest-api/>.

*Spotify vs. Pandora: Number of active users 2014*. (2014). Retrieved March 31, 2016, from <http://www.statista.com/statistics/293749/spotify-pandora-number-active-users/>

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

*We Know Music...* (n.d.). Retrieved March 2016, from <http://the.echonest.com/>.

Witten, I. (2013). Weka Predictive Analytics Tutorial. Retrieved July, 2016, from <https://www.youtube.com/playlist?list=PLzVF1nAqI9VmC96TbvOPMkXTToSmBMHJn7>

# Appendices

## Appendix A: Study 1 Supplemental Information

### A1: Focus Group Flyer Posting Locations



**Figure A1:** Focus group flyer posting locations. Flyers for our focus groups were posted throughout the University of Maryland-College Park campus in the general locations marked by the grey page icons. The focus groups were held in a classroom in Hornbake Library, a central location of the campus, as shown by the yellow sun icon. Map courtesy of Conferences and Visitors Services.

A2: Study 1 Consent Form

<b>Project Title</b>	Gemstone Team MUSIC: Focus Group
<b>Purpose of the Study</b>	This research is being conducted by Prof. Ramani Duraiswami and Gemstone Team MUSIC at the University of Maryland, College Park. We are inviting you to participate in this research project because you have shown an interest in music and discovering ways to find new music. The purpose of this research project is to develop a method of automated extraction of musical features from an audio file by using signal processing and computational techniques, with the end goal of enhancing song recommendation capabilities. This music feature-extraction and recommendation program paves the way for detailed trend analysis and unique music recommendations to be applied commercially and academically. Focus groups are being used to determine which musical traits are sought after when finding new music. Furthermore, we intend to find out the other individuals' opinions on currently available music recommendation services and what individuals believe could be improved upon.
<b>Procedures</b>	Involvement in this study includes participation in a single focus group, which will entail discussing past experiences with music recommendation services as well as the prominent musical features in a few clips of audio. A video camera will be placed in the room to record the session. The session will take place in the Gemstone Suite in Ellicott Hall at the University of Maryland and last no longer than one hour.
<b>Potential Risks and Discomforts</b>	A possible risk of embarrassment can arise, but is extremely unlikely during the course of the focus group. While the questions are designed to avoid sensitive topics, remember that you do not have to answer any question that makes you uncomfortable. Discomfort may also arise due to the presence of the video camera in the room recording the session. Furthermore, there is a potential for the loss/breach of confidentiality given the nature of the focus group and our inability to control other participants outside of the focus group. You may choose to go by an alias to decrease this risk.
<b>Potential Benefits</b>	This research is not designed to help you personally, but the results may help the investigators learn more about how others view currently available music recommendation system. We furthermore hope to learn more about the different musical features that individuals tend to focus on in songs they find more appealing.
<b>Confidentiality</b>	Any potential loss of confidentiality will be minimized by password protecting all recordings from the focus groups. These recordings will be deleted upon conclusion of the study. If we write a report or article about this research project, your identity will be protected to the maximum extent possible. Your information may be shared with representatives of the University of Maryland, College Park or governmental authorities if you or someone else is in danger or if we are required to do so by law.
<b>Compensation</b>	You will receive \$10. Check here if you expect to earn \$600 or more as a research participant in UMCP studies in this calendar year. You must provide your name, address and SSN to receive compensation. Check here if you do not expect to earn \$600 or more as a research participant in UMCP studies in this calendar year. Your name, address, and SSN will not be collected to receive compensation.
<b>Right to Withdraw and Questions</b>	Your participation in this research is completely voluntary. You may choose not to take part at all. If you decide to participate in this research, you may stop participating at any time. If you decide not to participate in this study or if you stop participating at any time, you will not be penalized or lose any benefits to which you otherwise qualify. If you are an employee or student, your employment status or academic standing at UMD will not be affected by your participation or non-participation in this study. If you decide to stop taking part in the study, or if you have questions, concerns, or complaints, please contact the principle investigator and/or the two co-investigators whose contact information is displayed below:

	<p align="center"><b>Principal Investigator: Ramani Duraiswami</b>  Email: *Redacted*  Telephone: *Redacted*</p> <p align="center">Co-Investigator: Mackenzie Walls  Email: *Redacted*  Telephone: *Redacted*</p> <p align="center">Co-Investigator: Mary Galuardi  Email: *Redacted*  Telephone: *Redacted*</p>	
<b>Participant Rights</b>	<p align="center">If you have questions about your rights as a research participant or wish to report a research-related injury, please contact:</p> <p align="center"><b>University of Maryland College Park  Institutional Review Board Office  1204 Marie Mount Hall  College Park, Maryland, 20742  E-mail: <a href="mailto:irb@umd.edu">irb@umd.edu</a>  Telephone: 301-405-0678</b></p> <p>This research has been reviewed according to the University of Maryland, College Park IRB procedures for research involving human subjects.</p>	
<b>Statement of Consent</b>	<p>Your signature indicates that you are at least 18 years of age; you have read this consent form or have had it read to you; your questions have been answered to your satisfaction and you voluntarily agree to participate in this research study. You will receive a copy of this signed consent form.</p> <p>If you agree to participate, please sign your name below.</p>	
<b>Signature and Date</b>	<b>NAME OF PARTICIPANT</b> <b>[Please Print]</b>	
	<b>SIGNATURE OF PARTICIPANT</b>	
	<b>DATE</b>	

Note: In “Right to Withdraw and Questions” section, investigator phone numbers and emails redacted for solely this document, but were faithfully provided on the actual form.

### A3: Study 1 Demographic Information

**Table A1:** Study 1 Demographics: Gender.  
n=37

<i>Gender</i>	<i>Occurrences</i>
Female	20
Male	17
Other	0
Prefer not to specify	0

**Table A2:** Study 1 Demographics: Ethnicity.  
n=37

<i>Ethnicity</i>	<i>Occurrences</i>
Asian/Pacific Islander	4
Black	7
Hispanic	3
White	18
Black/Asian	1
Black/Hispanic	1
White/Asian	1
White/Hispanic	1
Other	0
Prefer not to specify	1

**Table A3:** Study 1 Demographics: Genre Preference. n=37. Participants could list multiple genres

<i>Genre</i>	<i>Occurrences</i>
Alternative	11
Rock/Pop	
Rock/Metal	6
Classic Rock	6
Pop	5
Indie	5
Pop/Folk	
Jazz	3
Hip Hop	3
Rap	2
Classical (Piano)	2
Country	2
R&B	1
Electronic	1
Reggae	1
Other	2

## ***Appendix B: Study 2 Supplemental Information***

### B1: Survey Script

Q1: The purpose of this survey is to assess the effectiveness of a novel music recommendation method. This method incorporates a model of how we perceive music and how it connects to our intrinsic human characteristics.

This survey will contribute to the research efforts of Team MUSIC to develop novel ways to recommend music. Your contribution to this survey will remain anonymous and unidentifiable, and you may withhold information if you desire. To complete the survey, you must answer all the questions. Thank you for agreeing to participate in our study.

Q2. Read the below consent form in its entirety and confirm your statement of consent by selecting yes, as appropriate.

<b>Project Title</b>	Gemstone Team MUSIC: Online Survey
<b>Purpose of the Study</b>	This research is being conducted by Team MUSIC of the Gemstone Honors Program at the University of Maryland, College Park, under the supervision of Prof. Ramani Duraiswami. The purpose of this research project is to develop a novel music recommendation method by using a music perception model to organize music in an application programming interface, <i>The Echo Nest</i> . This survey will examine the viability of the music perception model for song recommendation.
<b>Procedures</b>	Participants in this study will answer a series of questions in an online survey. You will be asked to listen to 20 song segments, and indicate your individual preference for each song. This survey is estimated to take 30 minutes to complete.
<b>Potential Risks and Discomforts</b>	Some songs in this survey may contain explicit or distressing content, and may be considered offensive by certain audiences. There are no other known risks that result from participating in this survey.
<b>Potential Benefits</b>	There are no direct benefits from participating in this research. However, the results of this research survey may help us learn more about how the cognitive music modeled presented can be incorporated into <i>The Echo Nest</i> platform for the most effective music recommendation method possible.
<b>Confidentiality</b>	All data collected through this online survey will be confidential, and will not be associated with an identifiable information. Only the investigators will be allowed to access to the data from this study. If we write a report or article about this research project, your identity will be protected to the maximum extent possible. Your information may be shared with representatives of the University of Maryland, College Park or governmental authorities if you or someone else is in danger or if we are required to do so by law.



<b>Compensation</b>	You will be compensated by Qualtrics.
<b>Right to Withdraw and Questions</b>	<p>Your participation in this research is completely voluntary. You may choose not to take part at all. If you decide to participate in this research, you may stop participating at any time. If you decide not to participate in this study or if you stop participating at any time, you will not be penalized or lose any benefits to which you otherwise qualify. If you decide to stop taking part in the study, if you have questions, concerns, or complaints, please contact the principal investigator:</p> <p style="text-align: center;"><b>Principal Investigator: Ramani Duraiswami</b>  Email: *Redacted*  Telephone: *Redacted*</p>
<b>Participant Rights</b>	<p>If you have questions about your rights as a research participant or wish to report a research-related injury, please contact</p> <p style="text-align: center;"><b>University of Maryland College Park  Institutional Review Board Office  1204 Marie Mount Hall  College Park, Maryland, 20742  Email: irb@umd.edu  Telephone: 301-405-0678</b></p> <p>This research has been reviewed according to the University of Maryland, College Park IRB procedures for research involving human subjects.</p>
<b>Statement of Consent</b>	<p>By clicking “Yes” below, you are confirming the following:  You are at least 18 years of age.  You have read this consent form or have had it read to you.  You voluntarily agree to participate in this research study.  You may request a copy of this consent form by emailing the principal investigator.</p>

Q3. By selecting the “Yes” choice below, you indicate that you are at least 18 years of age; you have read the above consent information; and you voluntarily agree to participate in this research study. If you would not like to give consent, please select “No”

Q4. Please answer the following questions honestly and to the best of your ability. Your answers will remain confidential, and will never be publicly available.

Q5. Please specify your gender.  
(Choices: Male, Female, Prefer not to answer)

Q6. Please specify your ethnicity. (Please select all that apply.)  
(Choices: American Indian or Alaskan Native, Asian or Pacific Islander, Black or African American, Hispanic or Latino, White/Caucasian, Other, Prefer not to answer)

Q7. What is your age?  
(Choices: 18-24, 25-34, 35-44, 45-50, 50 or older, Prefer not to answer)

Q8. Which type(s) of music do you prefer to listen to or identify with? (Please select a maximum of 5)

(Choices: Avant-Garde, Blues, Classical, Comedy-Spoken, Country, Electronic, Folk, Holiday, International, Jazz, Latin, New Age, Pop/Rock, R&B, Rap, Reggae, Religious, Vocal, Prefer not to answer)

Q9. How would you describe your experience when listening to music that is from a different genre or style than the music you usually listen to?

(Choices: Uncomfortable, Somewhat Uncomfortable, Somewhat Comfortable, Comfortable, Not Applicable)

Q10. How often do you listen to songs that are not associated with the genre or style of music that you usually listen to?

(Choices: Never, Less than Once a Month, Once a Month, 2-3 Times a Month, Once a Week, 2-3 Times a Week, Daily, Not Applicable)

Q11. Of the current music recommendation systems that you have used, how would you rate its ability to recommend songs that match your music preferences?

(Choices: Unsatisfactory, Somewhat unsatisfactory, Neutral, Somewhat Satisfactory, Satisfactory, Not Applicable – have not used any such system)

Q12. On the following pages, you be presented with 1-minute segments from 20 songs in a randomized order. Please listen to each segment in its entirety; you will be unable to move onto the next page until you have done so. Please listen to the songs using headphones or speakers in a quiet, non-distracting environment.

After each segment, you will be asked to rate your preference for the song using a slider bar.

A rating of “6” corresponds to “**High Preference**,” meaning that you **enjoyed** listening to the segment and would **prefer to listen to it again** in the future.

A rating of “1” corresponds to “**Low Preference**,” meaning that you **did not enjoy** listening to the segment and would **not prefer to listen to it again** in the future.

We ask you to have an open mind when making your indication. In making your selection, consider whether the piece has features that you may find interesting or cause you to want to listen to it again in the future even if it is not a style of music you may listen to.

If you recognize the song, try to answer to question based on your initial impression of the song when you first heard it.

We encourage you to answer the question based on listening to the segment only once, indicating your initial preference after listening to the entire segment. You may listen to the segment again if necessary to answer the question.

\*Q13-32 Comprised of presenting a randomized song clip and asking the participant to respond in the following manner:

Please listen to the following clip in its entirety. We ask you to have an open mind to different music genres while considering your preference for this song.

Q#. Please use the slider bar below to indicate your preference for this piece. A rating of “6” corresponds to “**High Preference**,” meaning that you **enjoyed** listening to the segment and would **prefer to listen to it again** in the future. A rating of “1” corresponds to “**Low Preference**,” meaning that you **did not enjoy** listening to the segment and would **not prefer to listen to it again** in the future.

\*Slider bar was from 1 to 6, and this question was repeated for each of the 20 song segments\*

Q33. This is the final portion of the survey. We will ask you to rate your preference of songs that have certain descriptors. We ask that you consider these questions to be independent from the questions you have already answered. Please answer these following questions as honestly as possible.

Q34. How much do you prefer songs that are **romantic, relaxing, unaggressive, sad, slow, and/or quiet**?

A rating of “6” corresponds to “**High Preference**,” meaning that you would enjoy listening to a song that can be described by at least one of the attributes above. A rating of “1” corresponds to “**Low Preference**,” meaning that you **would not** enjoy listening to a song that can be described by at least one of the attributes above. (Slider bar was from 1 to 6)

Q35. How much do you prefer songs that are **uncomplicated, relaxing, unaggressive, soft, and/or acoustic**?

A rating of “6” corresponds to “**High Preference**,” meaning that you would enjoy listening to a song that can be described by at least one of the attributes above. A rating of “1” corresponds to “**Low Preference**,” meaning that you **would not** enjoy listening to a song that can be described by at least one of the attributes above. (Slider bar was from 1 to 6)

Q36. How much do you prefer songs that are **inspiring, intelligent, complex, and/or dynamic**?

A rating of “6” corresponds to “**High Preference**,” meaning that you would enjoy listening to a song that can be described by at least one of the attributes above. A rating of “1” corresponds to “**Low Preference**,” meaning that you **would not** enjoy listening to a song that can be described by at least one of the attributes above. (Slider bar was from 1 to 6)

Q37. How much do you prefer songs that are **distorted, loud, and/or aggressive**? These songs may also be described as having qualities that are **not** relaxing, romantic, or inspiring.

A rating of “6” corresponds to “**High Preference**,” meaning that you would enjoy listening to a song that can be described by at least one of the attributes above.

A rating of “1” corresponds to “**Low Preference**,” meaning that you **would not** enjoy listening to a song that can be described by at least one of the attributes above.

(Slider bar was from 1 to 6)

Q38. How much do you prefer songs that are **percussive, electric, rhythmic, and/or danceable**? These songs may also be described as having qualities that are **not** sad.

A rating of “6” corresponds to “**High Preference**,” meaning that you would enjoy listening to a song that can be described by at least one of the attributes above.

A rating of “1” corresponds to “**Low Preference**,” meaning that you **would not** enjoy listening to a song that can be described by at least one of the attributes above.

(Slider bar was from 1 to 6)

Q39. Thank you for completing our survey. By doing so, you have helped us gain more insight on how human perception of music can be incorporated into music recommendation. Your results will not be published, but conclusions drawn from the results of all of the participants will be incorporated in our research and future publications.

If you have any questions or concerns, please feel free to contact us at [music.gemstone@gmail.com](mailto:music.gemstone@gmail.com).

Your involvement in our research is greatly appreciated.

This survey is based on content from “The Structure of Musical Preferences: A Five-Factor Model” by Rentfrow et al. (2011), and “The Song Remains the Same: A Replication of the MUSIC Model” by Rentfrow et al. (2012).

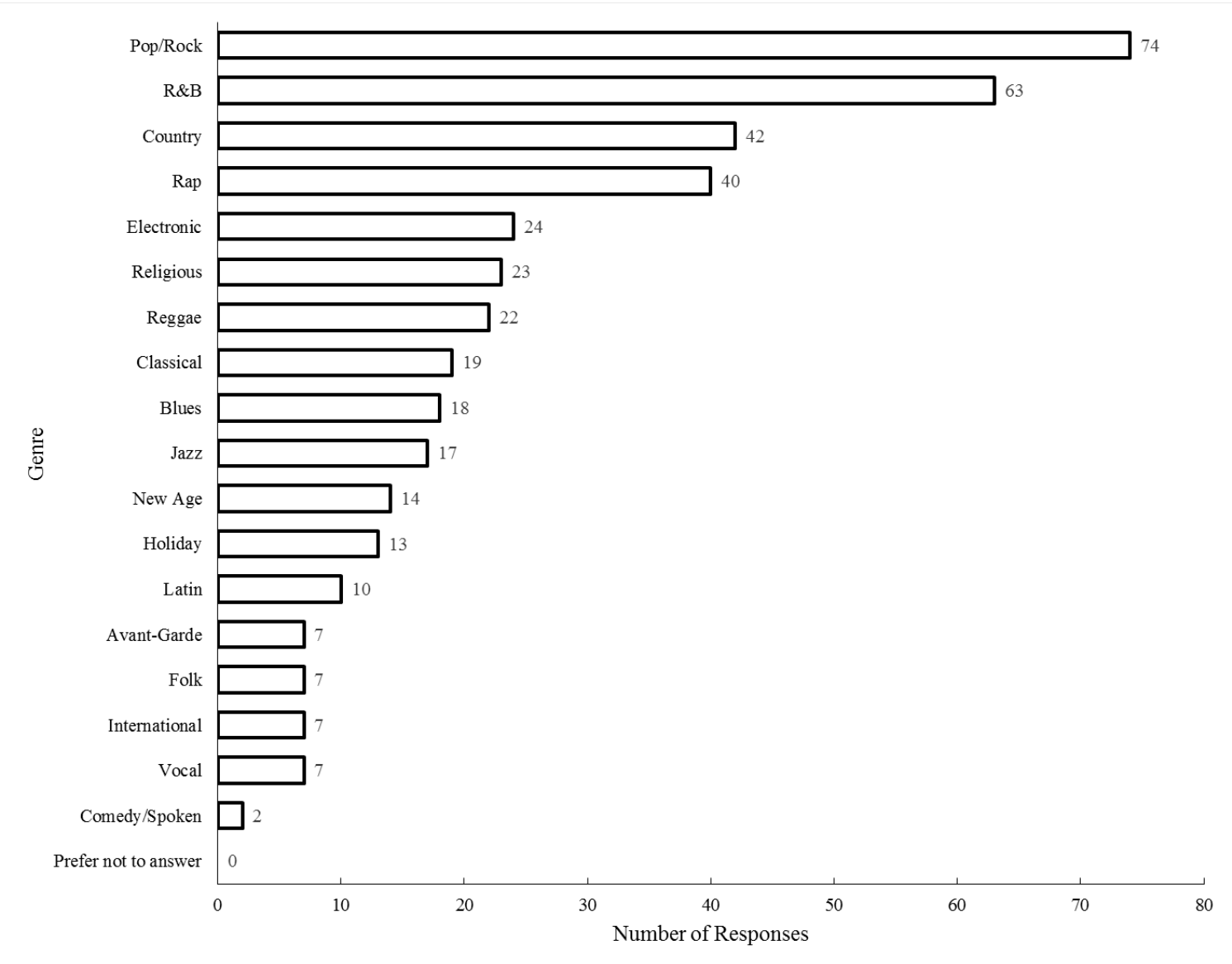
#### Notes:

- In “Right to Withdraw and Questions” section, investigator phone number and email redacted for solely this document, but were faithfully provided on the actual form

- Questions 13-32 presented in a random order to each participant. These questions additionally required the participant to stay on the page for at least 60 seconds to encourage them to listen to the entire song segment

- Questions 34-38 presented in a random order to each participant

B2: Study 2 Genre Preferences



**Figure B1.** Genre preferences specified by Study 2 participants. n=100. Participants were able to select up to five preferences.

*Appendix C: Studies 3 and 4 Supplemental Information*

C1: Song List Used for Machine Learning Training

**Table C1.** Songs from Rentfrow et al. (2011; 2012) that were used in machine learning model development

<i>Song Title</i>	<i>Artist</i>	<i>M</i>	<i>U</i>	<i>S</i>	<i>I</i>	<i>C</i>
Children Of Spring	Bruce Smith	0.65	0.14	0.38	-0.05	0.01
Sweet Scene	Ali Handal	0.52	0.38	0.31	0.03	0.01
Sweet 5	Kush	0.5	0.02	0.49	-0.06	0.29
Newsreel Paranoia	Babe Gurr	0.13	0.76	0.18	-0.04	0.03
<b>Penny Black</b>	<b>Bob Delevante</b>	0.25	0.75	0.13	0.01	0.05
<b>Lana Marie</b>	<b>Five Foot Nine</b>	0.34	0.71	0.1	-0.05	0.03
<b>Praying For Time</b>	<b>Carey Sims</b>	0.47	0.65	-0.02	0.05	0.08
That's Not Rockabilly	Hillbilly Hellcats	-0.11	0.64	0.27	0.03	-0.02
Famous Right Where I Am	Laura Hawthorne	0.46	0.62	-0.12	0.01	0.1
<b>Seltzer, Do I Drink Too Much?</b>	<b>Ljova</b>	0.04	0.11	0.82	0.06	0.07
<b>Trumpet Concerto in C major</b>	<b>Antonio Vivaldi</b>	0.21	0.05	0.75	0	-0.06
<b>Who Are You</b>	<b>Paul Serrato &amp; Co.</b>	0.07	0.1	0.68	0.02	0.25
<b>North Africa's Destiny?</b>	<b>Moh Alileche</b>	0.06	0.2	0.67	0.01	0.07
<b>Fernando Esta Feliz</b>	<b>Lisa McCormick</b>	0.06	0.27	0.63	-0.03	0.25
<b>Let's Love</b>	<b>Lisa McCormick</b>	0.3	0.29	0.51	0.02	0.16
<b>Face The Failure</b>	<b>Bankrupt</b>	-0.05	0	-0.01	0.85	-0.02
<b>Michigan</b>	<b>Squint</b>	-0.04	0.03	-0.02	0.83	-0.06
<b>Over Now</b>	<b>Straight Outta Junior High</b>	-0.06	0.11	0.05	0.82	0
<b>Death Before Dishonor</b>	<b>Five Finger Death Punch</b>	0.08	-0.12	-0.02	0.8	-0.01
<b>Johnny Fly</b>	<b>The Tomatoes</b>	-0.06	0.14	0.04	0.79	0.01
<b>Dick Dater</b>	<b>Cougars</b>	-0.15	0.21	0.08	0.76	0.08
<b>White Knuckles</b>	<b>Five Finger Death Punch</b>	-0.11	-0.12	-0.05	0.74	0.01
<b>Out Of Lies</b>	<b>Dawn Over Zero</b>	0.14	-0.1	-0.03	0.72	0.1
<b>Get The Party Started</b>	<b>Sammy Smash</b>	-0.2	0.13	-0.09	0.06	0.76
<b>Brooklyn Swagger</b>	<b>Ciph</b>	-0.1	0.15	-0.05	0.07	0.75
<b>Go Away</b>	<b>The Cruxshadows</b>	0.3	-0.21	0.25	0.08	0.5
Forever In Love - Live	Kenny G	0.84	0.07	0.09	-0.01	0.12
After hours	Mezzoforte	0.82	0.01	0.15	0.01	0.16
Long Ago and Far Away	Earl Klugh	0.81	0.11	0.09	0.08	0.18
Sister Rose	Kenny G	0.79	0.06	0.19	0.02	0.12
In All My Wildest Dreams	Joe Sample	0.72	0.12	0.12	0.02	0.3
Angela	Bob James	0.72	0.12	0.16	0.03	0.31
Come Away With Me	Norah Jones	0.54	0.35	-0.04	0.05	0.25
Smooth Operator	Sade	0.54	0.17	0.01	-0.06	0.47
The St. Louis Blues	Bessie Smith	-0.06	0.78	0.18	0	0.05
Gut Bucket Blues	Louis Armstrong	0.03	0.78	0.25	0.03	0.07

All Of Me	Billie Holiday	0.09	0.76	0.25	-0.14	0.16
Minnie & The Moocher	Cab Calloway & his Orchestra	-0.11	0.74	0.25	-0.02	0.15
Charmaine	Jacques Montagne	0.18	0.63	0.45	0.1	0.06
Hobo's Blues	Paul Simon	0.25	0.61	0.18	0.21	0.04
Origin	Pharoah Sanders	0.12	0.12	0.85	0.11	0.16
Moments Notice	Pharoah Sanders	0.23	0.18	0.82	0.05	0.08
Italian Concerto: Presto	Jacques Montagne	0.14	0.24	0.82	0.08	0.11
Ko Ko	Charlie Parker	0.11	0.29	0.78	-0.02	0.09
Directions I	Miles Davis	-0.08	0.11	0.73	0.22	0.15
John McLaughlin	Miles Davis	0.06	0.2	0.58	0.41	0.19
Cloudburst	Lambert, Hendricks & Ross	-0.18	0.31	0.42	0.1	0.39
Park's Place	Royal Crown Revue	-0.09	0.4	0.41	0.19	0.37
Blue Wind	Jeff Beck	0.01	0.1	0.15	0.8	0.12
Nothing Like the Sound of Bebop	Den Sidran	0.18	0.1	0.1	0.18	0.73
Sax-A-Go-Go	Candy Dulfer	0.25	0.08	0.26	0.25	0.63
Rose Rouge	St. Germain	0.1	0.16	0.46	0.1	0.55
Pick Up	noJazz	0.36	-0.01	0.35	0.15	0.51
No Surprises	Radiohead	0.81	0.12	0.04	0	0.08
Fake Plastic Trees	Radiohead	0.79	0.11	0.02	0.01	0.13
Reckoner	Radiohead	0.73	0.09	0.18	0.12	0.08
Nude	Radiohead	0.66	0.11	0.29	-0.05	0.11
Dream Brother	Jeff Buckley	0.65	0.02	0.13	0.2	0.19
Weird Fishes/Arpeggi	Radiohead	0.64	0.07	0.38	0	0.06
Girls	Death in Vegas	0.62	0.12	0.18	-0.06	0.17
Deep Blue	Arcade Fire	0.6	0.16	0.18	0.04	0.24
15 Step	Radiohead	0.58	0.1	0.35	0.03	0.05
Hallelujah	Jeff Buckley	0.53	0.19	-0.01	-0.08	0.29
Guess I'm Doing Fine	Beck	0.46	0.18	0.1	0.17	0.38
Eleanor Rigby	The Beatles	0.4	0.3	0.03	0.02	0.38
Honey Don't	The Beatles	0.13	0.81	0.08	-0.02	0.19
Hot Dog	Led Zeppelin	0	0.78	0.22	0.07	0.1
Act Naturally	The Beatles	0.09	0.77	0.03	-0.03	0.14
Parchman Farm Blues	Jeff Buckley	0.33	0.76	0.03	0	0.05
Bron-Yr-Aur Stomp	Led Zeppelin	0.23	0.75	0.07	0.06	0.1
Boogie With Stu	Led Zeppelin	0.11	0.63	0.18	0.11	0.3
Rich Man's Welfare	The RH Factor	0.23	0.15	0.74	0.06	-0.01
One On One	Hall & Oates	0.11	0.06	0.7	-0.18	0.35
Lover Lay Down	Dave Mathews Band	0.29	0.12	0.65	-0.17	0.24
Inca Roads	Frank Zappa	0.32	0.11	0.63	0.06	-0.03
You Enjoy Myself	Phish	0.12	0.21	0.59	0.2	0.27
Black Mountain Side	Led Zeppelin	0.39	0.28	0.43	0.12	-0.05
L.S.F.	Mark Ronson	0.31	0.31	0.41	0.21	0.2
Tension Head	Queens of the Stone Age	-0.01	-0.11	0.06	0.81	0.03
Battery Acid	Queens of the Stone Age	0.08	0.07	-0.01	0.8	-0.02

Burning Inside	Ministry	-0.1	-0.08	0	0.8	-0.06
Kick Out The Jams	Jeff Buckley	0.02	0.1	0.02	0.79	0.01
Quick And To The Pointless	Queens of the Stone Age	0	-0.04	-0.08	0.78	-0.13
Misfit Love	Queens of the Stone Age	0.07	0.01	0.08	0.78	0.02
The Invisible Man	Queen	-0.06	0.04	0.24	0.71	0.06
Higher Ground	Red Hot Chili Peppers	-0.04	0.05	0	0.71	0.26
Sabotage	Beastie Boys	0.12	-0.02	-0.19	0.63	0.09
Down on the Street	The Stooges	0.07	0.38	0.13	0.56	0.3
Electioneering	Radiohead	0.45	0.12	-0.06	0.52	-0.03
Jerry Was A Race Car Driver	Primus	0.18	0.27	0.08	0.39	0.03
Saturday Night	Ozomalti	0.04	0.03	0.12	0.26	0.19
Under Pressure	Queen	0.24	0.15	-0.07	0.12	0.64
Crazy Little Thing Called Love	Queen	0.16	0.4	-0.01	-0.03	0.64
When Doves Cry	Prince	0.15	0.05	0.11	-0.01	0.63
Listen To The Music	The Doobie Brothers	0.06	0.42	0.24	-0.07	0.59
Comfortably Numb	Pink Floyd	0.31	0.2	0.11	0.09	0.53
Symphony No. 3	Philip Glass	0.1	-0.02	0.83	0.13	-0.1
Jumping the Creek	Charles Lloyd	0.04	-0.01	0.71	0.06	0.26
Bohemian Beer Party	Walter Legawiec & His Polka Kings	-0.02	0.34	0.66	0.08	0.1
I Wish You Love	Mantovani	0.33	0.07	0.64	-0.06	-0.14
Mambo Numero Cinco	Hilton Ruiz	0.17	0.07	0.64	-0.06	0.17
You Brought Me Up	Meav	0.22	0.09	0.61	0.11	0.11
Take Me in Your Arms	Dean Martin	0.21	0.28	0.55	0.08	-0.08
Waxing Moon	Jah Wobble	-0.06	0.13	0.53	0.14	0.32
I Fell in Love	Carlene Carter	0.18	0.79	-0.11	0.08	-0.02
Texas Tornado	Tracy Lawrence	0.21	0.76	-0.13	0.08	-0.01
Let the Mystery Be	Iris Dement	0	0.65	0.27	0.06	0.11
Razzle Dazzle	Bill Haley and His Comets	0.16	0.47	0.36	0.14	0.02
Cold Feelings	Social Distortion	-0.04	0.05	0.02	0.78	0.08
Get it On	Kingdom Come	0.06	0.16	0	0.66	0.02
When our Love Passed out on the Couch	X	-0.14	0.1	0.26	0.66	0.15
Wildflower	Skylark	0.68	0.24	0.13	0.06	-0.01
I Love You	Kenny Rankin	0.59	0.26	0.26	0.14	-0.02
Laughter in the Rain	Earl Klugh	0.58	0.08	0.37	-0.19	0.19
Seltzer Do I Drink Too Much?	Ljova	0.16	0.05	0.77	0.12	0.01
The Keel Laddie	Golden Bough	-0.19	0.31	0.71	0.05	0.13
North Africa's Destiny	Moh Alileche	0.06	0.09	0.7	0.1	0.17
Sonata A Major	Bruce Smith	0.26	0.18	0.69	-0.03	-0.01
Concerto in C	Antonio Vivaldi	0.27	0.09	0.68	0.05	-0.05
Still Too Late	Ron Sunshine	0.33	0.04	0.59	0.01	0.2
And What You Hear	Twelve 20 Six	-0.07	0.14	0.59	0.28	0.26
Falling Down	Ezekiel Honig	0.07	0.12	0.54	0.11	0.3
Night of the Living Mambo	Mamborama	0.24	0.05	0.47	0.05	0.34
With the North Wind	The Tossers	0.11	0.33	0.44	0.13	0.05



Never Mind	Linn Brown	0.35	0.41	0.42	0.02	-0.02
Fernando Esta Feliz	Lisa McCormick	0.3	0.17	0.38	0.08	0.32
Who Are You?	Paul Serrato & Co.	0.16	-0.04	0.38	0	0.13
Penny Black	Bob Delevante	0.11	0.75	0.2	0.03	0.16
Newsreel Paranoia	Babe Gurr	-0.09	0.73	0.23	0.08	0.18
Lana Marie	Five Foot Nine	0.19	0.72	0.19	-0.07	0.11
Praying For Time	Carey Sims	0.29	0.72	0.04	0.07	0.15
Hard to Get Over Me	Babe Gurr	0.03	0.69	0.02	0.01	0.07
My Remembrance of You	Diana Jones	-0.02	0.58	0.44	0	0.05
Carrots and Grapes	Curtis	0.01	0.54	0.29	0.18	0.22
Passing Through	Mark Erelli	0.23	0.52	0.33	-0.04	-0.04
Once in a Lifetime	Doug Astrop	0.34	0.49	0.27	-0.03	0.29
Sweet Scene	Ali Handal	0.37	0.47	0.37	-0.03	0.05
Angel	Epic Hero	0.32	0.39	0.29	0.15	0.07
Let Me In	Travis Abercrombie	0.39	0.39	-0.18	0.33	0.07
Michigan	Squint	0.03	0.03	0.09	0.83	0.06
Johnny Fly	The Tomatoes	0.14	0.04	0.03	0.8	0.05
Death Before Dishonor	Five Finger Death Punch	0.09	-0.04	0.07	0.77	0.18
Over Now	Straight Outta Junior High	0.02	0.1	-0.02	0.76	0.01
Salvation	Five Finger Death Punch	-0.08	-0.07	0.08	0.76	0.11
Face the Failure	Bankrupt	-0.15	0.06	0.14	0.76	0.15
Dick Dater	Cougars	0.03	0.08	0.07	0.76	0.14
Out of Lies	Dawn Over Zero	-0.13	-0.14	0.18	0.75	0.15
Girlfriend	The Peasants	0.09	0.09	-0.12	0.73	-0.03
Prove It To Me	Tiff Jimber	0.31	0.2	-0.08	0.68	0.05
White Knuckles	Five Finger Death Punch	-0.19	-0.07	0.18	0.63	0.17
Brooklyn Swagger	Ciph	-0.11	0.07	0.04	0.15	0.68
Get the Party Started	Sammy Smash	-0.19	0.09	-0.07	0.15	0.65
Go Away	The Cruxshadows	0.09	0.04	0.18	0.28	0.56
Sesame Hood	Grafenberg All-Stars	-0.28	0.26	0.14	0.31	0.51
Close	The Alpha Conspiracy	0.21	0.07	0.36	0.24	0.5
Big City	Michael Davis	0.29	0.12	0.41	0.16	0.43
Children of Spring	Bruce Smith	0.5	0.39	0.4	-0.07	0.14
Let's Love	Lisa McCormick	0.46	0.34	0.37	0.08	-0.08
Seltzer, do I Drink Too Much?	Ljova	0.1	-0.03	0.82	-0.02	-0.01
Johnny Fly	The Tomatoes	0.02	0.14	0.12	0.73	-0.06
Sweet Scene	Ali Handal	0.73	0.06	0.23	-0.01	0.03

Notes:

- The bolded items represent songs that had two sets of MUSIC values from different studies from Rentfrow et al. (2011; 2012). These were kept to reduce bias in choosing one and strengthen our model when confronted with multiple versions.

## C2. Additional Classifier Predictions

**Table C2:** MUSIC factor Gaussian Processes predictions

Song Title	Predicted M	Predicted U	Predicted S	Predicted I	Predicted C
Symphony No. 1 in B Flat Major	0.163	0.335	0.494	0.039	0.104
The Way You Look Tonight	0.31	0.32	0.363	0.026	0.123
Rock the Clock	0.096	0.262	0.219	0.26	0.136
If You Only Knew	0.205	0.399	0.209	0.168	0.107
Slate	0.515	0.252	0.231	0.04	0.156
Where Eagles Dare	0.075	0.07	0.119	0.664	0.042
Oh No the Radio	0.139	0.24	0.111	0.458	0.056
City of Gold	0.092	0.164	0.139	0.397	0.254
Just Walk Away	0.419	0.19	0.397	0.013	0.062
Unspeakable	0.214	0.334	0.164	0.133	0.295
Brown Baby	0.203	0.249	0.324	0.133	0.218
Intro	0.03	0.113	0.178	0.237	0.365
My Favorite Polka	0.283	0.401	0.214	0.003	0.246
Interstate Rag	-0.008	0.273	0.564	0.168	0.123
Big Blue Sun	0.447	0.261	0.257	0.016	0.196