Proceedings of The National Conference
on Undergraduate Research (NCUR) 2011
Ithaca College, New York
March 31 – April 2, 2011

Discovering GEMS in Music: Armonique Digs for Music You Like

Amber Anderson
Miriam Weaverdyck
Psychology of Music
Bethel College
North Newton, Kansas 67117 USA

Faculty Advisor: Dwight Krehbiel

Abstract

There are many methods to measure which emotions are elicited by music. The recently developed Geneva Emotional Music Scale (GEMS) includes 33 emotion terms, reduced by factor analysis to 9 dimensions. In this study, participants rated music using these 9 dimensions on a scale from "not at all" to "very much." Armonique is a music search engine that suggests songs based on structural similarities revealed by power-law metrics. Previous studies using a small set of emotion dimensions have demonstrated similar emotional responses to songs judged to be similar by Armonique. Using these resources, we sought to answer the questions: (1) Which musically elicited emotions are predictive of liking? (2) Does Armonique's structural analysis translate into specific emotional responses given by the GEMS dimensions? Thirty participants listened to 6 excerpts (3 similar, 3 dissimilar) chosen by Armonique in relation to an original piece, to which they also listened, and rated their own emotional responses. In previous experiments, only three dimensions were used to measure emotion while this study combined these three measures with the 9 GEMS ratings. The data were then compiled and analyzed to look for correlations between emotions. Liking ratings were positively correlated ($.6 \leq r \leq .7$) with ratings of pleasantness, joyful activation, and inspiration, while there were lower correlations ($.3 \leq r \leq .5$) with tenderness, nostalgia, peacefulness, and power. By contrast, liking was negatively correlated with tension. Similarity to the original piece was greater for similar than for dissimilar songs in ratings of liking, pleasantness, tension, joyful activation, nostalgia, and wonder. Through the use of GEMS, this research suggests several emotional factors that may influence listeners' like or dislike for a particular piece of music and provides a more detailed validation of the performance of the Armonique search engine.

Keywords: Music, Emotion, Liking

1. Introduction

Music appears everywhere in our lives. We simply cannot escape it. Every culture has defined their own style of music, using it in situations from the celebration of birth to the mourning of death and every event in between. In every situation, however, there is a commonality: they all involve forms of emotional stimulation[1]**.** The nature of emotions that music elicits has been the subject of many studies, as has the question of what aspects of a piece cause a person to like that music[1,2,3,4]**.** Previous research has shown that a variety of factors may be predictive of liking but that they may not be necessary for a piece to be preferred[5]**.** Familiarity with a composition or the genre is among the factors that have been shown to predict liking[6,7,8]**.** This finding suggests that liking may develop partly through enculturation. Infants have exhibited preferences for their own culture's style of music by the age of two[9]**.** As children age, they begin to form preferences for certain types of music around adolescence, solidifying their favorite types by around twenty[9]**.** These preferences are largely based on what they are exposed to during that time and even become a part of their identity in that they can be a defining factor of social circles of which they are a member[9]**.**

Another predictor of preference is complexity of the music[10]**.** If the selection is very simple, we find it too predictable and become bored, whereas if the selection is too complex, we find it too unpredictable and unpleasant because we cannot easily discern any patterns within its structure[1]. Pleasantness is strongly correlated with liking[8]. Sadness, on the other hand, does not seem to be correlated to liking at all, positively or negatively[10]. While intuition might say sadness is unpleasant and therefore should correspond negatively to liking, studies have found that this is not the case[1].

For many years, the study of musically-elicited emotions was hampered by the use of ratings that only measured the basic, "coarse" emotions[11]. When further research was done to gather more accurate ratings about the nuanced emotions that music is able to arouse, nine dimensions of emotions were developed[11].

It is widely known that music elicits emotion. The problem lies in understanding which emotions those are, since many people are not accustomed to putting these emotions into words, and expressing how music makes one feel is very different than explaining how the music sounds. A recent study introduced a new device to measure emotions elicited by music, called the Geneva Emotional Music Scale[11]. The GEMS scale includes nine factors and forty highly correlated terms, allowing researchers to go beyond the traditional terms used to describe emotions and into the next level of emotion description.

Music search engines are used by many listeners to find music they like by entering in a song that is already a favorite and allowing the search engine to suggest similar pieces. Last.fm is an online music recommendation system that suggests music based on factors such as: music a person is currently listening to, which artists or songs they play most often, how much a person has played specific music over a certain amount of time, etc. Last.fm uses "The Scrobbler," which the listener must download, to deliver personal recommendations to the listener from their growing song database currently listed at 43 billion songs. The more popular iTunes Genius recommends songs based on the listener's iTunes playlist and then compares it to another listener's playlist that includes some of the same songs. It then applies latent-factor algorithms to predict songs the listener would like. Pandora radio, however, is quite different in its method of suggesting music in that musicologists rate individual songs based on melody, harmony, rhythm, instrumentation, lyrics, and vocal harmony. It is also different in that the listener is more interactive with this system since Pandora allows for listeners to tell them what they think of the music they are recommending by responding with, "I like this song," "I'm tired of this song," or "I don't like this song." Based on their response, Pandora will adjust its recommendations. Thus, Armonique[12] may not seem unique in its ability to produce similar songs. However, Armonique uses an entirely different technique to achieve the same purpose. Armonique is preferable to the aforementioned search engines because it requires no human ratings, and it aids in discovering music that listeners are unlikely to find otherwise. Armonique uses Zipf's Law and over 250 metrics based on power laws to search its music database for songs that are aesthetically similar or dissimilar to a piece chosen by the listener.

Objects, sounds, etc. that show a 1/f, or harmonic, proportion have repeatedly shown to be aesthetically pleasing to humans. This proportion is related to the golden ratio by means of logarithmic principles. Manaris has integrated this golden ratio into his Armonique search engine to develop metrics to measure individual pieces of music, allowing assessment of their similarity to other pieces[12].

Armonique's intricate system raises the question of applicability. Does Armonique's structural analysis translate into specific emotional responses of listeners as measured along the GEMS dimensions? If so, then which musically elicited emotions are predictive of liking?

In light of previous findings on Armonique[12], it was hypothesized that responses along at least some of the GEMS dimensions would be similar for pieces judged to be similar by Armonique. In addition, since previous experiments had demonstrated that pleasantness is a predictor of liking, the present study was designed to explore which positive emotions from the GEMS dimensions might predict liking.

2. Methodology

2.1 song excerpts

The song excerpts used in this experiment were chosen because prior research in this laboratory had already demonstrated some aspects of affective and physiological responses to these pieces. They had been chosen as follows by Armonique: an original piece (henceforth designated O) was selected from the database and then three songs considered very similar (most similar = MS, second most similar = MS2, third most similar = MS3) and three songs considered very dissimilar (likewise labeled MD, MD2, MD3) to the original were also selected, making a total of seven songs. The songs were then shortened so that only the first minute of each song was used in the experiment. It should be noted that the original and similar songs were in the "classical" genre.

2.2 participants

Participants in this study were students recruited from the General Psychology classes on campus and received extra course credit. Of the 31 participants, 21 were female and 10 male. Participants' preference rankings of classical music among 8 total genres varied from 1 (favorite genre) to 8 (least favorite).

2.3 instruments

The instruments used in this study consisted of two sets of ratings displayed on successive front panels of a LabVIEW virtual instrument[13] (screenshots shown below), and iTunes software, which played the songs in a different random order for each participant**.**

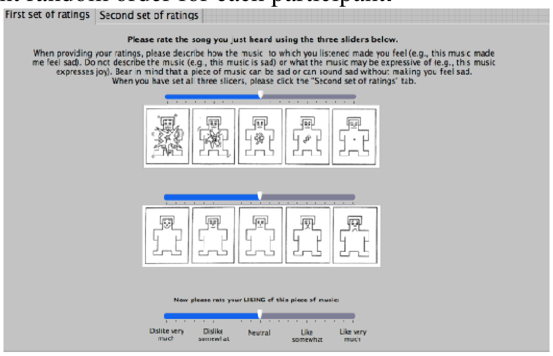

Figure 2.3(a): Screenshot of the first set of ratings

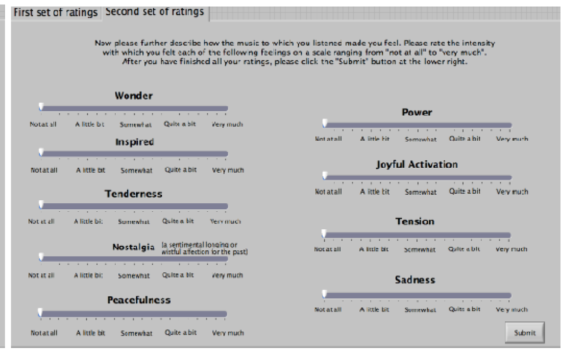

Figure 2.3(b): Screenshot of the second set of ratings

Figure 2.3(a) shows the set of ratings commonly used in psychology of music experiments in this laboratory. The instructions at the top read,

> When providing your ratings, please describe how the music to which you listen made you feel (e.g., this music made me feel sad). Do not describe the music (e.g., this music is sad) or what the music may be expressive of (e.g., this music expresses joy). Bear in mind that a piece of music can be sad or can sound sad without making you feel sad. When you have set all three sliders, please click the "Second set of ratings" tab.

The first picture rating represented "activation" rated on a scale of 0-10 where 10 is to the far left and 0 is the far right. The second picture rating was represented as "pleasantness" with the same scale as the first set. These images are taken from the Self-Assessment Manikin of Bradley and Lang[14]. The third is simply a rating of "liking" from 0-10, where 0 is "Dislike very much" and 10 is "Like very much."

Figure 2.3(b) is the "Second set of ratings" tab displaying the nine categories derived by GEMS researchers. Each emotion is rated on a scale of "Not at all" (a rating of 0 in our data) to "Very much" (a rating of 10). The nine factors are "wonder," "inspired," "tenderness," "nostalgia," "peacefulness," "power," "joyful activation," "tension," and "sadness." The instructions for this tab read,

> Now please further describe how the music to which you listened made you feel. Please rate the intensity with which you feel each of the following feelings on a scale ranging from "not at all" to "very much." After you have finished all your ratings, please click the "Submit" button at the lower right.

It should be noted that a definition for "nostalgia" (but no other emotion) was provided for participants. This was added after several subjects inquired about the meaning of "nostalgia" in the pilot study.

2.4 procedure

Participants chosen from a Bethel College psychology class were scheduled by appointment to listen to seven excerpts of music. The instructions were read to them before they made the original emotion ratings (e.g. liking, activation, pleasantness), followed by the nine GEMS emotion ratings (e.g. wonder, inspired, tenderness, nostalgia, peacefulness, power, joyful activation, tension, and sadness). The participants were shown that the original emotion ratings were on a separate page than the nine GEMS emotion ratings. They were allowed to move from the first page to the next page, as well as from song to song, at their own discretion. The participants were asked if they understood the instructions before they began. After all songs had been rated, the participants filled out a debriefing form.

2.5 data analysis

LabVIEW was used to record all twelve emotion ratings. These were transferred to an Excel spreadsheet to be analyzed in R Commander[15]**.** The debriefing forms provided data about participants, which were combined with the ratings data in hierarchical linear modeling analyses using HLM software[16].

The GEMS dimensions were transformed by means of Box-Cox transformations, using the exponents in Table 1 to normalize the data[17,18]**.**

Table 1: transformation exponents for the nine GEMS dimensions

	Joyful Activation	**Inspiration**	**Tenderness**	**Nostalgia**	**Peacefulness**	**Power**	**Tension**	**Wonder**	**Sadness**
Transformation Exponent	0	0	0	-1	0	-.5	-1	0	-1

3. Data

3.1 patterns of ratings across songs

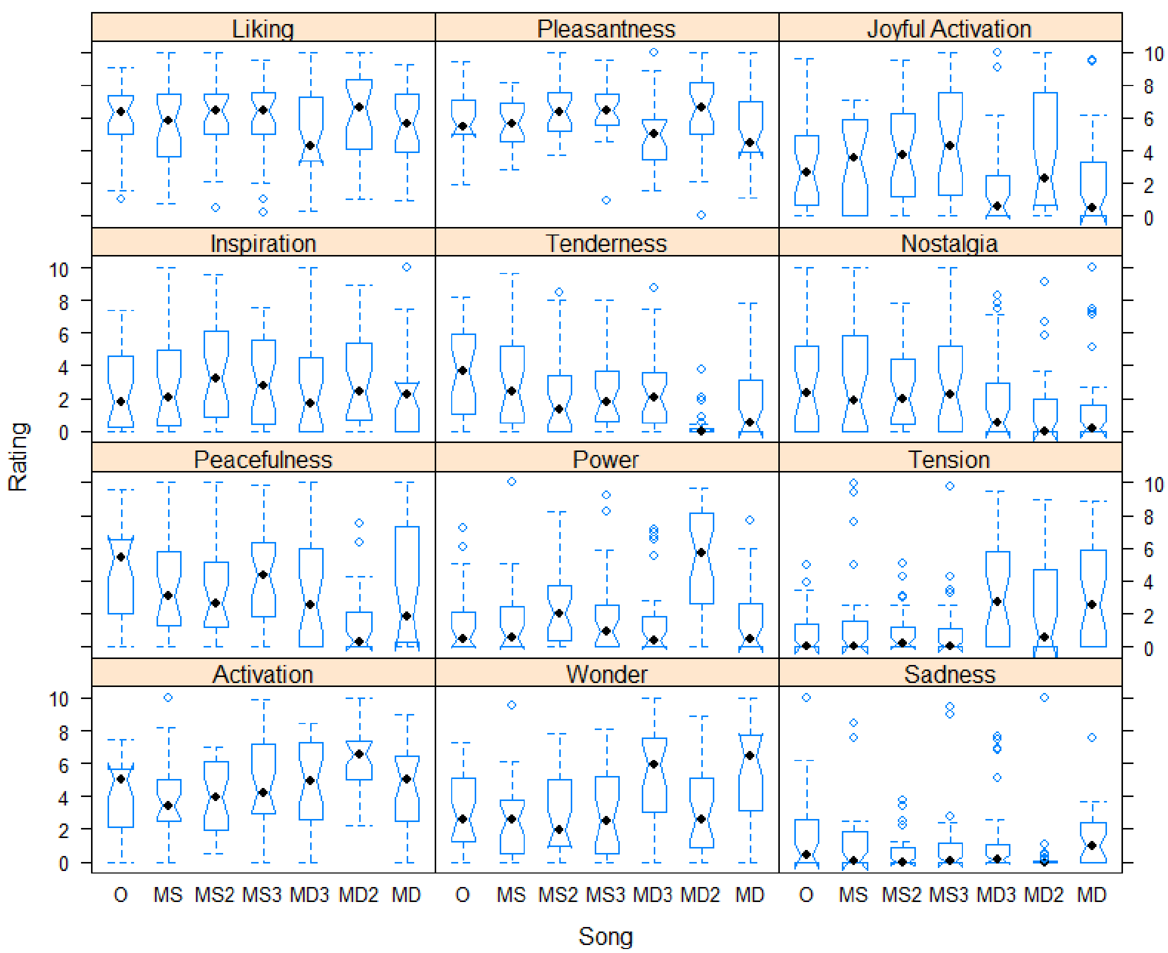

Figure 3.1: Boxplots of responses to the seven songs for each of the twelve dimensions – nine GEMS dimensions plus “Liking” and two dimensions measured with the Self-Assessment Manikin, here referred to as “Pleasantness,”

The ratings for all dimensions are displayed in Figure 3.1. This figure shows that median ratings for liking had a pattern across the seven songs that is similar to those for pleasantness, activation, and inspiration, while wonder and tension showed an inverse pattern. Preliminary analysis also showed the strongest correlations to the liking rating for pleasantness, joyful activation, and inspiration (see Table 2).

From these relationships and through further inspection of Figure 3.1, it can be seen that songs many individuals liked best were also perceived to have the least tension. Focusing only on the liking box, participants rated the O, MS, MS2, MS3, and MD2 songs similarly. However, MD2 is quite different from the original and three similar songs on a number of other dimensions, such as tenderness, nostalgia, peacefulness, and power. This finding suggests that participants' liking may be based on different emotions for some songs than for others. This figure also illustrates individual differences, as can be seen by the wide range of ratings and numerous outliers.

3.2 correlations

Table 2: correlation coefficients of each non-transformed rating to liking across all seven songs

RATING	COEFFICIENT	**RATING**	COEFFICIENT	**RATING**	COEFFICIENT
Pleasantness	.694	**Nostalgia**	.326	**Activation**	.007
Joyful Activation	.666	**Peacefulness**	.473	**Wonder**	.250
Inspiration	.597	**Power**	.340	**Sadness**	-.109
Tenderness	.306	**Tension**	-.358		

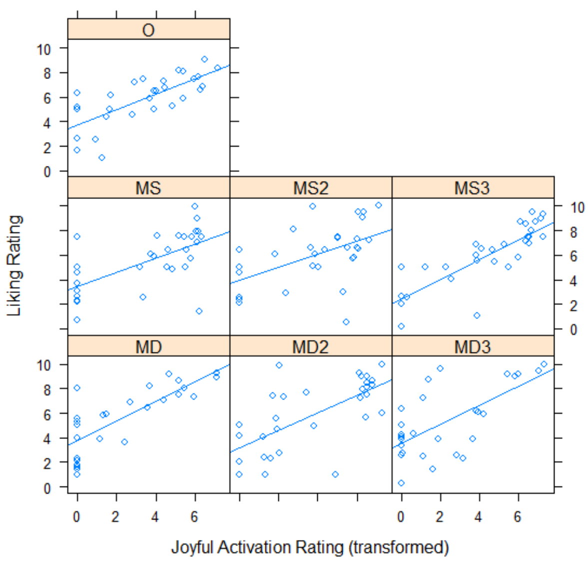

Figure 3.2(a): Scatterplots of liking ratings vs. transformed joyful activation ratings for each of the seven songs.

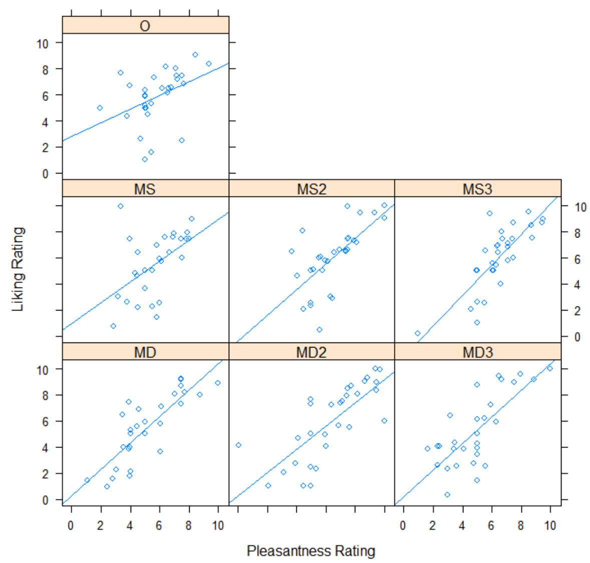

Figure 3.2(b): Scatterplots of liking ratings vs. pleasantness ratings for each of the seven songs.

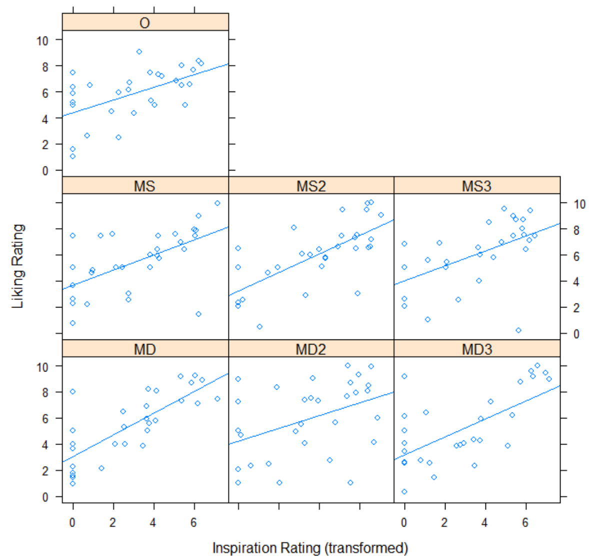

Figure 3.2(c): Scatterplots of liking ratings vs. transformed inspiration ratings for each of the seven songs.

The relationships to liking of the three strongest correlates are displayed separately for each song in the scatterplots above. These plots indicate that for each of the seven excerpts, ratings for joyful activation, pleasantness, and inspiration are all highly positively correlated with ratings of how much they liked the piece. Note, however, that the strength of this correlation is not equal between the songs. This is particularly noticeable in the Liking-Pleasantness graph in which almost every song has a very strong correlation except the original, for which the relationship is considerably weaker. Now consider the song MD (most dissimilar to the original). In all three, it has one of the strongest correlations between liking and the emotion rating, possibly due to the wide range of responses. Since the dissimilar songs are very different from the original, they may allow for more variation in how participants react.

The three ratings that displayed the strongest correlations with liking were employed in a hierarchical linear model to determine which of the dimensions would significantly predict liking. The model showed that each of these variables was a significant predictor of liking, each adding significantly to the overall predictiveness of the model, even though the p-value for joyful activation in the overall model does not attain the traditional $\alpha = .05$ level. These results are summarized in Table 3.

Table 3: HLM model parameter estimates and significance tests

The predicted variable is Liking

Fixed Effect	**Coefficient**	**Approximate P-value**
Intercept	5.711911	0.000
Pleasantness	0.447977	0.001
Joyful Activation (transformed)	0.165921	0.078
Inspiration (transformed)	0.355829	0.001

4. Discussion

While there have been many studies of musically-elicited emotions[1,2,3,4], and some in regards to Armonique[5,6], this study was the first to take a closer look at the effects that Armonique's analysis has on the emotion associated with pieces using the more in-depth GEMS measurement scale. In accordance with past studies, the experiment suggests that Armonique is successful in finding similar songs which exhibit similar relationships between elicited emotions, particularly those of pleasantness and activation[12]. The current findings duplicate and extend results found by Manaris et al by utilizing the same song excerpts and ratings, and supplementing them with additional GEMS dimensions[12]. Furthermore and more specifically, it was found that people tend to enjoy music that is pleasant, joyful, or inspiring more than music that does not contain these qualities. When these three ratings were combined using HLM software, we found them to be strongly predictive of liking. Thus, when a particular piece of music is characterized by high levels of all three, the listener is more likely to like this piece. It is possible therefore, that listeners seek out these emotions in music.

An interesting result found in the data from this study was the negative relationship between liking and tension. While the correlation between tension and liking was negative, this pattern may be due to the abrupt, and thus unresolved, ending of the excerpts, since the music was always shortened to 1 minute lengths without regard to what was occurring in the music at that time. This seems to counteract the popular professional notion that music is only worthwhile if it contains some amount of tension. However, most compositions containing tension must resolve this tension before the piece finishes for a person to like it. When this resolution is cut off, it appears the listener reacts negatively to this unresolved stress in the music. Another explanation for the negative correlation is the purpose for which music is used. Tension was positively correlated to activation in this study showing that tension arouses physiological systems. It may be that sitting in a laboratory is not a situation in which tension is desired since no physical activities are required and participants may want to remain as relaxed as possible. Thus, it is impossible to say from this study alone what the significance of the negative correlation between liking and tension is.

Another trend of interest in the data was the unequal and asymmetric distributions in some of the ratings. As can be observed in Figure 3.1, the ratings for which the slider started in middle of the scale (those on the first tab) tended to have far less skewed results toward zero than those in which the slider began at zero. This was compensated for by performing transformations on the data[17]**.** However, even without transforming the data as seen in Figure 3.1, significant patterns among the ratings were clear.

Note that in the box and whisker graphs and scatterplots (Figure 3.1 and Figures 3.2(a), (b), and (c)), there is a large range in which the ratings fall. These wide ranges of values for any particular rating in a given song suggest that there are many individual differences in the strength of the emotion elicited in a listener.

In Figure 3.1, a pattern appears to group the similar songs together and the dissimilar songs together, except for song MD2. This excerpt consisted of a large amount of percussion and very little conventional melody; this is unlike any of the other pieces. Therefore, it is likely that MD2 was given such distinct ratings due to this dissimilarity to all the other songs, not just the original. The distinctive ratings for MD2 suggest that our findings may be limited to the classical genre.

This study has yielded some interesting insights into emotions elicited by music and their relationship with liking. As noted earlier, songs rated as similar by Armonique do tend to arouse similar emotions in the listener. Also, two main emotions appear to be predictive of liking, joyful activation and inspiration, as well as the overall pleasantness of the piece. Further research incorporating other musical genres is necessary to determine whether other GEMS dimensions may be predictive of liking.

5. Acknowledgments

5.1 references

1. W. F. Thompson, *Music, thought, and feeling: understanding the psychology of music* (Oxford: Oxford University Press, 2008).
2. P. N. Juslin and J. A. Sloboda (Eds), *Handbook of music and emotion: theory, research, applications* (New York, NY US: Oxford University Press, 2010).
3. T. Eerola and J. K. Vuoskoski, "A comparison of the discrete and dimensional models of emotion in music," *Psychology of Music* 39(1) (2011): 18-49. doi:10.1177/0305735610362821.
4. P. G. Hunter and E. G. Schellenberg, "Music and emotion," *Music perception* (Ed. by M. R. Jones, R. R. Fay, and A. N. Popper, New York, NY US: Springer, 2010): 129-164.
5. S. Pryce and T. Burns, “You don't have to know it to like it: general familiarity as a predictor of liking for music chosen by the Armonique music search engine,” poster, Sigma Xi Student Research Conference, The Woodlands, TX US, November 14, 2009.
6. S. Barrera, “Pleasantness and familiarity predict liking of music selected by the Armonique music search engine,” poster, Sigma Xi Student Research Conference, The Woodlands, TX US, November 14, 2009.

7. S. O. Ali and Z. F. Peynircioğlu, "Intensity of emotions conveyed and elicited by familiar and unfamiliar music," *Music Perception* 27(3) (2010): 177-182. doi: 10.1525/mp.2010.27.3.177.

8. D. A. Ritossa and N. S. Rickard, "The relative utility of 'pleasantness and liking' dimensions in predicting the emotions expressed by music," *Psychology of Music* 32(1) (2004): 5-22. doi:10.1177/0305735604039281.

9. D. J. Levitin, "My favorite thing: why do we like the music we like?" *The Jossey-Bass reader on the brain and learning* (Jossey-Bass Publishers, 2007): 374-376.

10. E. G. Schellenberg, I. Peretz, and S. Vieillard, "Liking for happy- and sad-sounding music: effects of exposure," *Cognition and Emotion* 22(2) (2008): 218-237. doi: 10.1080/02699930701350753.

11. M. Zentner, D. Grandjean, and K. R. Scherer, "Emotions evoked by the sound of music: characterization, classification, and measurement," *Emotion* 8(4) (2008): 494-521. doi: 10.1037/1528-3542.8.4.494.

12. B. Manaris, P. Roos, D. Krehbiel, T. Zalonis, and J. R. Armstrong, "Zipf's law, power laws and music aesthetics," *Music data mining* (Ed. by T. Li, M. Ogihar, and G. Tzanetakis, Boca Raton, FL US: CRC Press, 2012): 169-216.

13. National Instruments, Austin, TX.

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

15. R Development Core Team, *R: a language and environment for statistical computing* (Vienna, Austria: R Foundation for Statistical Computing, 2010), accessed March 8, 2011, http://www.R-project.org.

16. S. W. Raudenbush, A. S. Bryk, Y. F. Cheong, and R. T. Congdon, Jr., *HLM 6: hierarchical linear & nonlinear modeling* (Lincolnwood, IL US: Scientific Software International, Inc., 2004).

17. R. Nishii, "Box-Cox transformation," *Encyclopaedia of mathematics* (Berlin, Germany: Springer-Verlag, 2002), accessed March 8, 2011, http://eom.springer.de/B/b110790.htm.

18. G. Snow, "R package 'TeachingDemos': demonstrations for teaching and learning," (2010), accessed March 8, 2011, http://cran.r-project.org/web/packages/TeachingDemos/TeachingDemos.pdf.