Computational Modeling of Emotional Content in Music

Kristine Monteith (kristinemonteith@gmail.com)

Tony Martinez (martinez@cs.byu.edu)

Dan Ventura (ventura@cs.byu.edu)

Department of Computer Science
Brigham Young University
Provo, UT 84602

Abstract

We present a system designed to model characteristics which contribute to the emotional content of music. It creates *n*-gram models, Hidden Markov Models, and entropy-based models from corpora of musical selections representing various emotions. These models can be used both to identify emotional content and generate pieces representative of a target emotion. According to survey results, generated selections were able to communicate a desired emotion as effectively as human-generated compositions.

Keywords: Music cognition; computational modeling; learning; music composition.

Introduction

Music and emotion are intrinsically linked; music is able to express emotions that cannot adequately be expressed by words alone. Often, there is strong consensus among listeners as to what type of emotion is being expressed in a particular piece (Gabrielsson & Lindstrom, 2001; Juslin, 2001). There is even some evidence to suggest that some perceptions of emotion in music may be innate. For example, selections sharing some acoustical properties of fear vocalizations, such as sudden onset, high pitch, and strong energy in the high frequency range, often provoke physiological defense responses (Ohman, 1988). Researchers have demonstrated similar low-level detection mechanisms for both pleasantness and novelty. (Scherer, 1984, 1988). There also appears to be some inborn preference for consonance over dissonance. In studies with infants, researchers found that their subjects looked significantly longer at the source of sound and were less likely to squirm and fret when presented with consonant as opposed to dissonant versions of a melody (Zentner & Kagan, 1996).

There are a variety of theories as to what aspects of music are most responsible for eliciting emotional responses. Meyer theorizes that meaning in music comes from following or deviating from an expected structure (Meyer, 1956). Sloboda emphasizes the importance of associations in the perception of emotion in music and gives particular emphasis to association with lyrics as a source for emotional meaning (Sloboda, 1985). Kivy argues for the importance of cultural factors in understanding emotion and music, proposing that the "emotive life" of a culture plays a major role in the emotions that members of that culture will detect in their music (Kivy, 1980). Tolbert proposes that children learn to associate emotion with music in much the same way that they learn to associate emotions with various facial expressions (Tolbert, 2001). Scherer presents a framework for formally describing the emotional effects of music and then outlines factors that contribute to these emotions, including structural, performance, listener, and contextual features (Scherer, 2001).

In this paper, we focus on some of the structural aspects of music and the manner in which they contribute to emotions in music. We present a cognitive model of characteristics of music responsible for human perception of emotional content. Our model is both discriminative and generative; it is capable of detecting a variety of emotions in musical selections, and also of producing music targeted to a specific emotion.

Related Work

A number of researchers have addressed the task of modeling musical structure for the purposes of building a generative musical system. Conklin summarizes a number of statistical models which can be used for music generation, including random walk, Hidden Markov Models, stochastic sampling, and pattern-based sampling (Conklin, 2003). These approaches can be seen in a number of different studies. For example, Hidden Markov Models have been used to harmonize melodies, considering melodic notes as observed events and a chord progression as a series of hidden states (Allan & Williams, 2005). Similarly, Markov chains have been used to harmonize given melody lines, focusing on harmonization in a given style in addition to finding highly probable chords (Chuan & Chew, 2007).

Wiggins, Pearce, and Mullensiefen present a system designed to model factors such as pitch expectancy and melodic segmentation. They also demonstrate that their system can successfully generate music in a given style (Wiggins, Pearce, & Mullensiefen, 2009). Systems have also been developed to produce compositions with targeted emotional content. Delgado, Fajardo, and Molina-Solana use a rule-based system to generate compositions according to a specified mood (Delgado, Fajardo, & Molina-Solana, 2009). Rutherford and Wiggins analyze the features that contribute to the emotion of fear in a musical selection and present a system that allows for an input parameter that determines the level of "scariness" in the piece (Rutherford & Wiggins, 2003). Oliveira and Cardoso describe a wide array of features that contribute to emotional content in music and present a system that uses this information to select and transform chunks of music in accordance with a target emotion (Oliveira & Cardoso, 2007). The authors have also developed a system that addresses the task of composing music with a specified emotional content (Monteith, Martinez, & Ventura, 2010). In this paper, we illustrate how our system can be interpreted as a cognitive model of human perception of emotional content in music.

Methodology

The proposed system constructs statistical and entropic models for various emotions based on corpora of human-labeled musical data. Analysis of these models provides insights as to why certain music evokes certain emotions. The models supply localized information about intervals and chords that are more common to music conveying a specific emotion. They also supply information about what overall melodic characteristics contribute to emotional content. To validate our findings, we generate a number of musical selections and ask research subjects to label the emotional content of the generated music. Similar experiments are conducted with human-generated music commissioned for the project. We then observe the correlations between subject responses and our predictions of emotional content.

Initial experiments focus on the six basic emotions outlined by Parrott (Parrott, 2001)—love, joy, surprise, anger, sadness, and fear—creating a data set representative of each. A separate set of musical selections is compiled for each of the emotions studied. Selections for the training corpora are taken from movie soundtracks due to the wide emotional range present in this genre of music. MIDI files used in the experiments can be found at the Free MIDI File Database.[1] These MIDI files were rated by a group of research subjects. Each selection was rated by at least six subjects, and selections rated by over 80% of subjects as representative of a given emotion were then selected for use in the training corpora. Selections used for these experiments are shown in Figure 1.

Next, the system analyzes the selections to create statistical models of the data in the six corpora. Selections are first transposed into the same key. Melodies are then analyzed and *n*-gram models are generated representing what notes are most likely to follow a given series of notes in a given corpus. Statistics describing the probability of a melody note given a chord, and the probability of a chord given the previous chord, are collected for each of the six corpora. Information is also gathered about the rhythms, the accompaniment patterns, and the instrumentation present in the songs.

The system also makes use of decision trees constructed to model the characteristics that contribute to emotional content. These trees are constructed using the C4.5 algorithm (Quinlan, 1993), an extension of the ID3 algorithm (Quinlan, 1986) that allows for real-valued attributes. The decision tree classifiers classifiers allow for a more global analysis of generated melodies. Inputs to these classifiers are the default features extracted by the "Phrase Analysis" component of the freely available jMusic software.[2] This component returns a vector of twenty-one statistics describing a given melody, including factors such as number of consecutive identical pitches, number of distinct rhythmic values, tonal deviation, and key-centeredness. These statistics are calculated for both the major and minor scales.

Love:
Advance to the Rear
Bridges of Madison County
Casablanca
Dr. Zhivago
Legends of the Fall
Out of Africa

Surprise:
Addams Family
Austin Powers
Batman
Dueling Banjos
George of the Jungle
Nightmare Before Christmas
Pink Panther
The Entertainer
Toy Story
Willie Wonka

Fear:
Axel's Theme
Beetlejuice
Edward Scissorhands
Jaws
Mission Impossible
Phantom of the Opera
Psycho
Star Wars: Duel of fhe Fates
X-Files: The Movie

Joy:
1941
633 Squadron
Baby Elephant Walk
Chariots of Fire
Flashdance
Footloose
Jurassic Park
Mrs. Robinson
That Thing You Do
You're the One that I Want

Anger:
Gonna Fly Now
James Bond
Mission Impossible
Phantom of the Opera
Shaft

Sadness:
Forrest Gump
Good Bad Ugly
Rainman
Romeo and Juliet
Schindler's List

Figure 1: Selections used in training corpora for the six different emotions considered.

A separate set of classifiers is developed to evaluate both generated rhythms and generated pitches. The first classifier in each set is trained using analyzed selections in the target corpus as positive training instances and analyzed selections from the other corpora as negative instances. This is intended to help the system distinguish selections containing the desired emotion. The second classifier in each set is trained with melodies from all corpora versus melodies previously generated by the algorithm, allowing the system to learn melodic characteristics of selections which have already been

[1] http://themes.mididb.com/movies/

[2] http://jmusic.ci.qut.edu.au/

accepted by human audiences.

For the generative portion of the model, the system employs four different components: a Rhythm Generator, a Pitch Generator, a Chord Generator, and an Accompaniment and Instrumentation Planner. The functions of these components are explained in more detail in the following sections.

Rhythm Generator

The rhythm for the selection with a desired emotional content is generated by selecting a phrase from a randomly chosen selection in the corresponding data set. The rhythmic phrase is then altered by selecting and modifying a random number of measures. The musical forms of all the selections in the corpus are analyzed, and a form for the new selection is drawn from a distribution representing these forms. For example, a very simple AAAA form, where each of four successive phrases contains notes with the same rhythm values, tends to be very common. Each new rhythmic phrase is analyzed by jMusic and then provided as input to the rhythm evaluators. Generated phrases are only accepted if they are classified positively by both classifiers.

Pitch Generator

Once the rhythm is determined, pitches are selected for the melodic line. These pitches are drawn according to the *n*-gram model constructed from melody lines of the corpus with the desired emotion. A melody is initialized with a series of random notes, selected from a distribution that models notes most likely to begin musical selections in the given corpus. Additional notes in the melodic sequence are randomly selected based on a probability distribution of note mosts likely to follow the given series of *n* notes.

For example, with the "joy" corpus, the note sequence (C4, D4, E4) has a 0.667 probability of being followed by an F4, a 0.167 probability of being followed by a D4, and a 0.167 probability of being followed by a C4. If these three notes were to appear in succession in a generated selection, the system would have a 0.167 probability of selecting a C4 as the next note.

The system generates several hundred possible series of pitches for each rhythmic phrase. As with the rhythmic component, features are then extracted from these melodies using jMusic and provided as inputs to the pitch evaluators. Generated melodies are only selected if they are classified positively by both classifiers.

Chord Generator

The underlying harmony is determined using a Hidden Markov Model, with pitches considered as observed events and the chord progression as the underlying state sequence (Rabiner, 1989). The Hidden Markov Model requires two conditional probability distributions: the probability of a melody note given a chord and the probability of a chord given the previous chord. The statistics for these probability distributions are gathered from the corpus of music representing the desired emotion.

For example, C4 is most likely to be accompanied by a C major chord, and F4 is most likely to be accompanied by a G7 chord in selections from the "love" corpus (probabilies of 0.099 and 0.061, respectively). In the "sadness" corpus, C4 is most likely to be accompanied by a C minor chord (probability of 0.060). As examples from the second set of distributions, the G7 chord is most likely to be followed by the G7 or the C major chord in selections from the "love" corpus (both have a probability of 0.105). In selections from the "sadness" corpus, the G7 chord is most likely to be followed by the G7 or the C minor chord (probabilities of 0.274 and 0.094 respectively).

The system then calculates which set of chords is most likely given the melody notes and the two conditional probability distributions. Since many of the songs in the training corpora had only one chord present per measure, initial attempts at harmonization also make this assumption, considering only downbeats as observed events in the model.

Accompaniment and Instrumentation Planner

The accompaniment patterns for each of the selections in the various corpora are categorized, and the accompaniment pattern for a generated selection is probabilistically selected from the patterns of the target corpus. Common accompaniment patterns included arpeggios, block chords sounding on repeated rhythmic patterns, and a low base note followed by chords on non-downbeats.

For example, arpeggios are a common accompaniment pattern in the corpus of selections expressing the emotion of "love." Two of the selections in the corpus feature simple, arpeggiated chords as the predominant theme in their accompaniments, and two more selections have an accompaniment pattern that feature arpeggiated chords played by one instrument and block chords played by a different instrument. The remaining two selections in the corpus feature an accompaniment pattern of a low base note followed by chords on non-downbeats. When a new selection is generated by the system, one of these three patterns is selected with equal likelihood to be the accompaniment for the new selection.

Instruments for the melody and harmonic accompaniment are also probabilistically selected based on the frequency of various melody and harmony instruments in the corpus. For example, melody instruments for selections in the "surprise" corpus include acoustic grand piano, electric piano, and piccolo. Harmony instruments include trumpet, trombone, acoustic grand piano, and acoustic bass.

Evaluation

In order to verify that our system was accurately modeling characteristics contributing to emotional content, we presented our generated selections to research subjects and asked them to identify the emotions present. Forty-eight subjects, ages 18 to 55, participated in this study. Six selections were generated in each category, and each selection was played for four subjects. Subjects were given the list of emotions and asked to circle all emotions that were represented in each song. Each selection was also played for four subjects who had not seen the list of emotions. These subjects were asked to write down any emotions they thought were present in the music without any suggestions of emotional categories on the part of the researchers. Reported results represent percentages of the twenty-four responses in each category. To provide a baseline, two members of the campus songwriting club were also asked to perform the same task: compose a musical selection representative of one of six given emotions. Each composer provided selections for three of the emotional categories. These selections were evaluated in the same manner as the computer-generated selections, with four subjects listening to each selection for each type of requested response. Reported results represent percentages of the four responses in each category.

Results

Figure 2 outlines the characteristics identified by the decision trees as being responsible for emotional content. For example, if a piece had a Dissonance measure over 0.107 and a Repeated Pitch Density measure over 0.188, it was classified in the "anger" category. Informally, angry selections tend to be dissonant and have many repeated notes. Similar information was collected for each of the different emotions. Selections expressing "love" tend to have lower repeated pitch density and fewer repeated patterns of three, indicating these selections tend to be more "flowing." Joyful selections have some stepwise movement in a major scale and tend to have a strong climax at the end. The category of "surprise" appears to be the least cohesive; it requires the most complex set of rules for determining membership in the category. However, repeated pitch patterns of four are present in all the surprising selections, as is a lack of stepwise movement in the major scale. Not surprisingly, selections expressing "sadness" adhere to a minor scale and tend to have a downward trend in pitch. Fearful selections deviate from the major scale, do not always compensate for leaps, and have an upward pitch direction. Downward melodic trends do not deviate as much from the major scale. Our model appears to be learning to detect the melodic minor scale; melodies moving downward in this scale will have a raised sixth and seventh tone, so they differ in only one tone from a major scale.

Tables 1 and 2 report results for the constrained response surveys. Row labels indicate the corpus used to generate a given selection, and column labels indicate the emotion identified by survey respondents. Based on the results in Table 1, our system is successful at modeling and generating music with targeted emotional content. For all of the emotional categories but "surprise," a majority of people identified the emotion when presented with a list of six emotions. In all cases, the target emotion ranked highest or second highest in terms of the percentage of survey respondents identifying that emotion as present in the computer-generated songs. As a general rule, people were more likely to select the categories of "joy" or "sadness" than some of the other emotions, perhaps

Love:
RepeatedPitchDensity $<=$ 0.146
- RepeatedPitchPatternsOfThree $<=$ 0.433: Yes
- RepeatedPitchPatternsOfThree $>$ 0.433: No
RepeatedPitchDensity $>$ 0.146: No

Joy:
PitchMovementByTonalStep $<=$ 0.287: No
PitchMovementByTonalStep $>$ 0.287
- ClimaxPosition $<=$ 0.968
- - ClimaxTonality $<=$ 0: No
- - ClimaxTonality $>$ 0
- - - PitchMovementByTonalStep(Minor) $<=$ 0.535: No
- - - PitchMovementByTonalStep(Minor) $>$ 0.535: Yes
- ClimaxPosition $>$ 0.968: Yes

Surprise:
RepeatedPitchPatternsOfFour $<=$ 0.376: No
RepeatedPitchPatternsOfFour $>$ 0.376
- PitchMovementByTonalStep (Minor) $<=$ 0.550
- - ClimaxPosition $<=$ 0.836: Yes
- - ClimaxPosition $>$ 0.836
- - - LeapCompensation $<=$ 0.704: No
- - - LeapCompensation $>$ 0.704
- - - - KeyCenteredness $<=$ 0.366: No
- - - - KeyCenteredness $>$ 0.366: Yes
- PitchMovementByTonalStep(Minor) $>$ 0.550: No

Anger:
Dissonance $<=$ 0.107: No
Dissonance $>$ 0.107
- RepeatedPitchDensity $<=$ 0.188: No
- RepeatedPitchDensity $>$ 0.188: Yes

Sadness:
TonalDeviation(Minor) $<=$ 0.100
- OverallPitchDirection $<=$ 0.500: Yes
- OverallPitchDirection $>$ 0.500: No
TonalDeviation (Minor) $>$ 0.100: No

Fear:
TonalDeviation $<=$ 0.232: No
TonalDeviation $>$ 0.232
- LeapCompensation $<=$ 0.835
- - OverallPitchDirection $<=$ 0.506
- - - TonalDeviation $<=$ 0.290: Yes
- - - TonalDeviation $>$ 0.290: No
- - OverallPitchDirection $>$ Yes
- LeapCompensation $>$ 0.835: No

Figure 2: Decision tree models of characteristics contributing to emotional content in music.

because music in western culture is traditionally divided up into categories of major and minor. A higher percentage of people identified "joy" in songs designed to express "love" or "surprise" than identified the target emotion. "Fear" was also a commonly selected category. More people identified angry songs as fearful, perhaps due to the sheer amount of scary-movie soundtracks in existence. Themes from "Jaws," "Twilight Zone," or "Beethoven's Fifth Symphony" readily come to mind as appropriate music to accompany frightening situations; thinking of an iconic song in the "anger" category is more of a challenging task. Averaging over all categories, 57.67% of respondents correctly identified the target emotion in computer-generated songs, while only 33.33% of respondents did so for the human-generated songs.

For the open-ended questions, responses were evaluated by similarity to Parrott's expanded hierarchy of emotions. Each of the six emotions can be broken down into a number of secondary emotions, which can in turn be subdivided into tertiary emotions. If a word in the subject's response matched any form of one of these primary, secondary, or tertiary emotions, it was categorized as the primary emotion of the set. Results are reported in Tables 3 and 4. Again, row labels indicate the corpus used to generate a given selection, and column labels indicate the emotion identified by survey respondents.

The target emotion also ranked highest or second highest in terms of the percentage of survey respondents identifying that emotion as present in the computer-generated songs for the open-ended response surveys. Without being prompted or limited to specific categories, and with a rather conservative method of classifying subject response, listeners were still often able to detect the original intended emotion. Once again, the computer-generated songs appear to be slightly more emotionally communicative. 21.67% of respondents correctly identified the target emotion in computer-generated songs in these open-ended surveys, while only 16.67% of respondents did so for human-generated songs.

Listeners cited "fondness," "amorousness," and in one rather specific case, "unrequited love," as emotions present in selections from the "love" category. One listener said it sounded like "I just beat the game." Another mentioned "talking to Grandpa" as a situation the selection called to mind. Reported descriptions of selections in the "joy" category most closely matched Parrott's terms. These included words such as "happiness," "triumph," "excitement", and "joviality." Selections were also described as "adventurous" and "playful."

None of the songs in the category of "surprise" were described using Parrott's terms. However, this is not entirely unexpected considering the fact that Parrott lists a single secondary emotion and three tertiary emotions for this category. By comparison, the category of joy has six secondary emotions and 34 tertiary emotions. The general sentiment of "surprise" still appears to be present in the responses. One listener reported that the selection sounded like an ice cream truck. Another said it sounded like being literally drunken with happiness. "Playfulness," "childishness," and "curiosity" were also used to describe the selections.

Angry songs were often described using Parrott's terms of "annoyance" and "agitation." Other words used to describe angry songs included "uneasy," "insistent," and "grim." Descriptions for songs in the "sad" category ranged from "pensive" and "antsy" to "deep abiding sorrow." A few listeners described a possible situation instead of an emotion: "being somewhere I should not be" or "watching a dog get hit by a car." Fearful songs were described with words such as "tension," "angst," and "foreboding." "Hopelessness" and even "homesickness" were also mentioned.

Table 1: Emotional Content of Computer-Generated Music. Percentage of survey respondents who identified a given emotion for selections generated in each of the six categories. Row labels indicate the corpus used to generate a given selection, and column labels indicate the emotion identified by survey respondents.

	love	joy	surprise	anger	sadness	fear
love	58%	75%	12%	4%	21%	0%
joy	58%	88%	25%	0%	4%	0%
surprise	4%	54%	38%	0%	12%	8%
anger	4%	04%	46%	50%	17%	88%
sadness	0%	8%	25%	42%	62%	58%
fear	17%	21%	29%	12%	67%	50%

Table 2: Emotional Content of Human-Generated Music.

	love	joy	surprise	anger	sadness	fear
love	50%	0%	25%	25%	100%	0%
joy	100%	25%	0%	0%	75%	0%
surprise	0%	0%	50%	75%	50%	50%
anger	25%	25%	0%	25%	50%	50%
sadness	75%	25%	25%	25%	0%	25%
fear	50%	0%	0%	0%	100%	50%

Conclusion

Pearce, Meredith, and Wiggins (Pearce, Meredith, & Wiggins, 2002) suggest that music generation systems concerned with the computational modeling of music cognition be evaluated both by their behavior during the composition process and by the music they produce. Our system is able to successfully develop cognitive models and use these models to effectively generate music. Just as humans listen to and study the works of previous composers before creating their own compositions, our system learns from its exposure to emotion-labeled musical data. Without being given a set of preprogrammed rules, the system is able to develop internal mod-

Table 3: Emotional Content of Computer-Generated Music: Unconstrained Responses.

	love	joy	surprise	anger	sadness	fear
love	21%	25%	0%	0%	0%	0%
joy	0%	58%	0%	4%	0%	0%
surprise	0%	12%	0%	8%	0%	0%
anger	0%	8%	0%	17%	0%	25%
sadness	4%	0%	0%	4%	17%	17%
fear	0%	8%	0%	12%	17%	17%

Table 4: Emotional Content of Human-Generated Music: Unconstrained Responses.

	love	joy	surprise	anger	sadness	fear
love	0%	25%	0%	0%	0%	0%
joy	0%	25%	0%	0%	0%	0%
surprise	0%	0%	0%	0%	25%	0%
anger	0%	0%	0%	0%	25%	0%
sadness	0%	0%	0%	0%	25%	0%
fear	0%	0%	0%	25%	25%	50%

els of musical structure and characteristics that contribute to emotional content. These models are used both to generate musical selections and to evaluate them before they are output to the listener. The quality of these models is evidenced by the system's ability to produce songs with recognizable emotional content. Results from both constrained and unconstrained surveys demonstrate that the system can accomplish this task as effectively as human composers.

Acknowledgments

This work is partially supported by the National Science Foundation under Grant No. IIS-0856089.

References

Allan, M., & Williams, C. K. I. (2005). Harmonising chorales by probabilistic inference. *Advances in Neural Information Processing Systems*, *17*, 25-32.

Chuan, C., & Chew, E. (2007). A hybrid system for automatic generation of style-specific accompaniment. *Proceedings International Joint Workshop on Computational Creativity*, 57-64.

Conklin, D. (2003). Music generation from statistical models. *Proceedings AISB Symposium on Artificial Intelligence and Creativity in the Arts and Sciences*, 30-35.

Delgado, M., Fajardo, W., & Molina-Solana, M. (2009). Inmamusys: Intelligent multi-agent music system. *Expert Systems with Applications*, *36*(3-1), 4574-4580.

Gabrielsson, A., & Lindstrom, E. (2001). The influence of musical structure on emotional expression. *Music and Emotion: Theory and Research*, 223-248.

Juslin, P. N. (2001). Communicating emotion in music performance: A review and a theoretical framework. *Music and Emotion: Theory and Research*, 223248.

Kivy, P. (1980). *The corded shell: Reflections on musical expression*. Princeton, NJ: Princeton University Press.

Meyer, L. (1956). *Emotion and meaning in music*. Chicago: Chicago University Press.

Monteith, K., Martinez, T., & Ventura, D. (2010). Automatic generation of music for inducing emotive response. *Proceedings of the International Conference on Computational Creativity*, 140-149.

Ohman, A. (1988). Preattentive processes in the generation of emotions. *Cognitive perspectives on emotion and motivation*, 127-144.

Oliveira, A., & Cardoso, A. (2007). Towards affective-psychophysiological foundations for music production. *Affective Computing and Intelligent Interaction*, 511522.

Parrott, W. G. (2001). *Emotions in social psychology*. Philadelphia: Psychology Press.

Pearce, M. T., Meredith, D., & Wiggins, G. A. (2002). Motivations and methodologies for automation of the compositional process. *Musicae Scientiae*, *6*(2).

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, *1*(1), 81-106.

Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. San Mateo, CA: Morgan Kaufman.

Rabiner, L. R. (1989). A tutorial on hidden markov models and selected applications in speech recognition. *Proc. IEEE*, *77*(2), 257-285.

Rutherford, J., & Wiggins, G. (2003). An experiment in the automatic creation of music which has specific emotional content. *Proceedings of MOSART, Workshop on current research directions in Computer Music*, 35-40.

Scherer, K. R. (1984). On the nature and function of emotion: A component process approach. *Approaches to emotion*, 293-317.

Scherer, K. R. (1988). On the symbolic functions of vocal affect expression. *Journal of Language and Social Psychology*, *7*, 79-100.

Scherer, K. R. (2001). Emotional effects of music: Production rules. *Music and Emotion: Theory and Research*, 223248.

Sloboda, J. (1985). *The musical mind: The cognitive psychology of music*. Oxford: Oxford University Press.

Tolbert, E. (2001). Music and meaning: An evolutionary story. *Psychology of Music*, *24*, 103-130.

Wiggins, G. A., Pearce, M. T., & Mullensiefen, D. (2009). Computational modelling of music cognition and musical creativity. *Oxford Handbook of Computer Music and Digital Sound Culture*.

Zentner, M., & Kagan, J. (1996). Perception of music by infants. *Nature*, *383*(29).