Automatic Generation of Music for Inducing Emotive Response

Kristine Monteith, Tony Martinez, and Dan Ventura

Computer Science Department
Brigham Young University
kristine.perry@gmail.com, martinez@cs.byu.edu,ventura@cs.byu.edu

Abstract. We present a system that generates original music designed to match a target emotion. It creates n-gram models, Hidden Markov Models, and other statistical distributions based on musical selections from a corpus representing a given emotion and uses these models to probabilistically generate new musical selections with similar emotional content. This system produces unique and often remarkably musical selections that tend to match a target emotion, performing this task at a level that approaches human competency for the same task.

1 Introduction

Music is a significant creative achievement. Every culture in history has incorporated music into life in some manner. As Wiggins explains, "musical behavior is a uniquely human trait...further, it is also ubiquitously human: there is no known human society which does not exhibit musical behaviour in some form" [1]. Perhaps one of the reasons musical behavior is tied so closely to humanity is its ability to profoundly affect human physiology and emotion. One study found that, when subjects were asked to select music that they found to be particularly pleasurable, listening to this type of music activated the same areas of the brain activated by other euphoric stimuli such as food, sex, or illegal drugs. The authors highlight the significance of the fact that music has an effect on the brain similar to that of "biologically relevant, survival-related stimuli" [2].

Computing that possesses some emotional component, termed affective computing, has received increased attention in recent years. Picard emphasizes the fact that "emotions play a necessary role not only in human creativity and intelligence, but also in rational human thinking and decision-making. Computers that will interact naturally and intelligently with humans need the ability to at least recognize and express affect" [3]. From a theoretical standpoint, it seems reasonable to incorporate emotional awareness into systems designed to mimic (or produce) human-like creativity and intelligence, since emotions are such a basic part of being human. On a more practical level, affective displays on the part of a computerized agent can improve function and usability. Research has shown that incorporating emotional expression into the design of interactive agents can improve user engagement, satisfaction, and task performance [4][5]. Users may

also regard an agent more positively [6] and consider it to be more believable [7] when it demonstrates appropriate emotional awareness.

Given music's ability to alter or heighten emotional states and affect physiological responses, the ability to create music specifically targeted to a particular emotion could have considerable benefits. Calming music can aid individuals in dealing with anxiety disorders or high-anxiety situations. Joyful and energizing music can be a strong motivating force for activities such as exercise and physical therapy. Music therapists use music with varied emotional content in a wide array of musical interventions. The ability to create emotionally-targeted music could also be valuable in creating soundtracks for stories and films.

This paper presents a system that takes emotions into account when creating musical compositions. It produces original music with a desired emotional content using statistical models created from a corpus of songs that evoke the target emotion. Corpora of musical data representing a variety of emotions are collected for use by the system. Melodies are then constructed using n-gram models representing pitch intervals commonly found in the training corpus for a desired emotion. Hidden Markov Models are used to produce harmonies similar to those found in the appropriate corpus. The system also selects the accompaniment pattern and instrumentation for the generated piece based on the likelihood of various accompaniments and instruments appearing in the target corpus. Since it relies entirely on statistics gathered from these training corpora, in one sense the system is learning to imitate emotional musical behavior of other composers when producing its creative works. Survey data indicates that the system composes selections that are as novel and almost as musical as human-composed songs. Without creating any rules for emotional music production, it manages to compose songs that convey a target emotion with surprising accuracy relative to human performance of the same task.

Multiple research agendas bear some relation to our approach. 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 [8]. 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 [9]. 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 [10].

Genetic algorithms have also been used in music composition tasks. De la Puente and associates use genetic algorithms to learn melodies, employing a fitness function that considers differences in pitch and duration in consecutive notes [11]. Horner and Goldberg attempt to create more cohesive musical selections using a fitness function that evaluates generated phrases according agreement with a thematic phrase [12]. Tokui and Iba focus their attention on using genetic algorithms to learn polyphonic rhythmic patterns, evaluating patterns with a neural network that learns to predict which patterns the user would most likely rate highly [13].

Musical selections can also be generated through a series of musical grammar rules. These rules can either be specified by an expert or determined by statistical models. For example, Ponsford, Wiggins, and Mellish use n-gram statistical methods for learning musical grammars [14]. Phon-Amnuaisuk and Wiggins compare genetic algorithms to a rule-based approach for the task of four-part harmonization [15].

Delgado, Fajardo, and Molina-Solana use a rule-based system to generate compositions according to a specified mood [16]. 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 [17]. 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 [18].

Like these previously mentioned systems, our system is concerned with producing music with a desired emotional content. It employs a number of statistical methods discussed in the previously mentioned papers. Rather than developing rule sets for different emotions, it composes original music based on statistical information in training corpora.

2 Methodology

In order to produce selections with specific emotional content, a separate set of musical selections is compiled for each desired emotion. Initial experiments focus on the six basic emotions outlined by Parrot [19]—love, joy, surprise, anger, sadness, and fear—creating a data set representative of each. Selections for the training corpora are taken from movie soundtracks due to the wide emotional range present in this genre of music. MIDIs used in the experiments can be found at the Free MIDI File Database.[1] These MIDIs 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.

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.

Since not every melody produced is likely to be particularly remarkable, the system also makes use of multilayer perceptrons with a single hidden layer to evaluate the generated selections. Inputs to these neural networks are the de-

---

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

fault 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.

A separate set of two networks is developed to evaluate both generated rhythms and generated pitches. The first network 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 network in each set is trained with melodies from all corpora versus melodies previously generated by the algorithm. In this way, the system learns to emulate melodies which have already been accepted by human audiences.

Once the training corpora are set and analyzed, 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.

2.1 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 neural network rhythm evaluators. Generated phrases are only accepted if they are classified positively by both neural networks.

2.2 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 the melody lines of the corpus with the desired emotion. A melody is initialized with a series of random notes, selected from a distribution that model which notes are 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 what note is most likely to follow the given series of n notes. 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 neural network pitch evaluators. Generated melodies are only selected if they are classified positively by both neural networks.

---

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

2.3 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. 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. 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.

2.4 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, chords sounding on repeated rhythmic patterns, and a low base note followed by chords on non-downbeats. (A few of the accompaniment patterns such as "Star Wars: Duel of the Fates" and "Addams Family" had to be rejected or simplified; they were so characteristic of the training selections that they were too recognizable in the generated song.) Instruments for the melody and harmonic accompaniment are also probabilistically selected based on the frequency of various melody and harmony instruments in the corpus.

3 Results

Colton [20] suggests that, for a computational system to be considered creative, it must be perceived as possessing skill, appreciation, and imagination. The system could be considered "skillful" if it demonstrates knowledge of traditional music behavior. This is accomplished by taking advantage of statistical knowledge to train the system to behave according to traditional musical conventions. The system may be considered "appreciative" if it can produce something of value and adjust its work according the preferences of itself or others. This is addressed through the neural networks evaluators. The "imaginative" criterion can be met if the system can create new material independent of both its creators and other composers. Since all of the generated songs can be distinguished from the songs in the training corpora, this criterion is met at least on a basic level. However, to further evaluate all of these aspects, the generated songs were subjected to human evaluation. Twelve selections were generated for testing purposes.[3] Each selection was then played for thirteen individuals, who were asked to answer the following questions:

---

[3] These selections are available at http://axon.cs.byu.edu/emotiveMusicGeneration

1. What emotions are present in this selection (circle all that apply)?
2. On a scale of one to ten, how much does this sound like real music?
3. On a scale of one to ten, how unique does this selection sound?

The first two questions target the aspects of skill and appreciation, ascertaining whether the system is skillful enough to produce something both musical and representative of a given emotion. The third question evaluates the imagination of the system, determining whether or not the generated music is perceived as novel by human audiences.

To provide a baseline, two members of the campus songwriting club were asked to perform the same task as the computer: compose a musical selection representative of one of six given emotions. Each composer provided three songs. These selections were also played and subjects were asked to evaluate them according to the same three questions. Song order was randomized, and while subjects were told that some selections were written by a computer and some by a human, they were not told which selections belonged to which categories.

Table 1 reports on how survey participants responded to the first question. It gives the percentage of respondents who identified a given emotion in computer-generated selections in each of the six categories. Table 2 provides a baseline for comparison by reporting the same data for the human-generated pieces. Tables 3 and 4 address the second two survey questions. They provide the average score for musicality and novelty (on a scale from one to ten) received by the various selections.

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. In four cases, it was ranked highest. Respondents tended to think that the love songs sounded a little more like joy than love, and that the songs portraying fear sounded a little sadder than fearful. But surprisingly, the computer-generated songs appear to be slightly better at communicating an intended emotion than the human-generated songs. Averaging over all categories, 54% of respondents correctly identified the target emotion in computer-generated songs, while only 43% of respondents did so for human-generated songs.

Human-generated selections did tend to sound more musical, averaging a 7.81 score for musicality on a scale of one to ten as opposed to the 6.73 scored by computer-generated songs. However, the fact that a number of the computer-generated songs were rated as more musical than the human-produced songs is somewhat impressive. Computer-generated songs were also rated on roughly the same novelty level as the human-generated songs, receiving a 4.86 score as opposed to the human score of 4.67. As an additional consideration, the computer-generated songs were produced in a more efficient and timely manner than the human-generated ones. Only one piece in each category was submitted for survey purposes due to the difficulty of finding human composers with the time to provide music for this project.

Table 1. Emotional Content of Computer-Generated Music. Percentage of survey respondents who identified a given emotion for songs generated in each of the six categories. The first column provides the categories of emotions for which songs were generated. Column headers describe the emotions identified by survey respondents.

	Love	Joy	Surprise	Anger	Sadness	Fear
Love	**0.62**	0.92	0.08	0.00	0.00	0.00
Joy	0.38	**0.69**	0.15	0.00	0.08	0.08
Surprise	0.08	0.46	**0.62**	0.00	0.00	0.00
Anger	0.00	0.00	0.08	**0.46**	0.38	0.69
Sadness	0.09	0.18	0.27	0.18	**0.45**	0.36
Fear	0.15	0.08	0.00	0.23	0.62	**0.23**

Table 2. Emotional Content of Human-Generated Music. Percentage of survey respondents who identified a given emotion for songs composed in each of the six categories.

	Love	Joy	Surprise	Anger	Sadness	Fear
Love	**0.64**	0.64	0.00	0.09	0.09	0.00
Joy	0.77	**0.31**	0.15	0.00	0.31	0.00
Surprise	0.00	0.27	**0.18**	0.09	0.45	0.27
Anger	0.00	0.09	0.18	**0.27**	0.73	0.64
Sadness	0.38	0.08	0.00	0.00	**0.77**	0.08
Fear	0.09	0.00	0.00	0.27	0.55	**0.45**

Table 3. Musicality and Novelty of Computer-Generated Music. Average score (on a scale of one to ten) received by selections in the various categories in response to survey questions about musicality and novelty.

	Musicality	Novelty
Love	8.35	4.12
Joy	6.28	5.86
Surprise	6.47	4.78
Anger	5.64	4.96
Sadness	7.09	4.40
Fear	6.53	5.07
Average:	6.73	4.86

Table 4. Musicality and Novelty of Human-Generated Music. Average score (on a scale of one to ten) received by selections in the various categories in response to survey questions about musicality and novelty.

	Musicality	Novelty
Love	7.73	4.45
Joy	9.15	4.08
Surprise	7.09	5.36
Anger	8.18	4.60
Sadness	9.23	4.08
Fear	5.45	5.45
Average:	7.81	4.67

4 Discussion and Future Work

Pearce, Meredith, and Wiggins [21] suggest that music generation systems concerned with the computational modeling of music cognition be evaluated both by the music they produce and by their behavior during the composition process. The system discussed here can be considered creative both in the fact that it can produce fairly high-quality music, and that it does so in a creative manner. In *Creativity: Flow and the Psychology of Discovery and Invention* (Chapter 2), Csikszentmihalyi includes several quotes by the inventor Rabinow outlining three components necessary for being a creative, original thinker [22]. The system described in this work meets all three criteria for creativity.

As Rabinow explains, "First, you have to have a tremendous amount of information...If you're a musician, you should know a lot about music..." Computers have a unique ability to store and process large quantities of data. They have the potential even to have some advantage over humans in this particular aspect of the creative process if the knowledge can be collected, stored, and utilized effectively. The system discussed in this paper addresses this aspect of the creative process by gathering statistics from the various corpora of musical selections and using this information to inform choices about rhythm, pitch, and harmony.

The next step is generation based on the domain information. Rabinow continues: "Then you have to be willing to pull the ideas...come up with something strange and different." The system described in this work can create a practically unlimited number of unique melodies based on random selections from probability distributions. Again, computers have some advantage in this area. They can generate original music quickly and tirelessly. Some humans have been able to produce astonishing numbers of compositions; Bach's work alone fills sixty volumes. But while computers are not yet producing original work of Bach's creativity and caliber, they could easily outdistance him in sheer output.

The final step is evaluation of these generated melodies, Rabinow's third suggestion: "And then you must have the ability to get rid of the trash which you think of. You cannot think only of good ideas, or write only beautiful music..." Our system addresses this aspect through the neural network evaluators. It learns to select pieces with features similar to musical selections that have already been accepted by human audiences and ones most like selections humans have labeled as expressing a desired emotion. It even has the potential to improve over time by producing more negative examples and learning to distinguish these from positive ones. But finding good features for use in the evaluating classifiers poses a significant challenge. First attempts at improving the system will involve modifications in this area.

As previously mentioned, research has been done to isolate specific features that are likely responsible for the emotional content of a song [17, 18]. Incorporating such features into the neural network evaluators could provide these evaluators with significantly more power in selecting the melodies most representative of a desired emotion. Despite the possible improvements, it is quite encouraging to note that even naïve evaluation functions are able to produce fairly musical and emotionally targeted selections.

Additional improvements will involve drawing from a larger corpus of data for song generation. Currently, the base seems to be sufficiently wide to produce songs that were considered to be as original as human-composed songs. However, many of the generated pieces tend to sound somewhat similar to each other. On the other hand, sparseness of training data actually provides some advantages. For example, in some cases, the presence of fewer examples in the training corpus resulted in similar musical motifs in the generated songs. Phrases would often begin with the same few notes before diverging, particularly in corpora where songs tended to start on the same pitch of the scale. Larger corpora will allow for the generation of more varied songs, but to maintain musicality, the evaluation mechanism might be extended to encourage the development of melodic motifs among the various phrases.

The type and magnitude of emotions can often be indicated by concurrent physiological responses. The format of these experiments lends itself to the additional goal of generating music targeted to elicit a desired physiological response. Future work will involve measuring responses such as heart rate, muscle tension, and skin conductance and how these are affected by different musical selections. This information could then be used to create training corpora of songs likely to produce desired physiological responses. These could then be used to generate songs with similar properties. The format also allows for the generation of songs that can switch emotions at a desired point in time simply by switching to using statistical data from a different corpus.

The system described here is arguably creative by reasonable standards. It follows a creative process as suggested by Rabinow and others, producing and evaluating reasonably skillful, novel, and emotionally targeted compositions. However, our system will really only be useful to society if it produces music that not only affects emotions, but that people will listen to long enough for that effect to take place. This is difficult to demonstrate in a short-term evaluation study, but we do appear to be on the right track. A few of the generated pieces received musicality ratings similar to those of the human-produced pieces. Many of those surveyed were surprised that the selections were written by a computer. Another survey respondent announced that the program had "succeeded" because one of the computer-generated melodies had gotten stuck in his head. These results show promise for the possibility of producing a system that is truly creative.

Acknowledgments

References

1. Wiggins, G.: A preliminary framework for description, analysis and comparison of creative systems. Journal of Knowledge Based Systems **19**(7) (2006) 449–458

2. Blood, A.J., Zatorre, R.J.: Intensely pleasurable responses to music correlate with activity in brain regions implicated in reward and emotion. Proceedings of the National Academy of Sciences **98**(20) (2001) 11818–11823
3. Picard, R.W.: Affective computing. MIT Technical Report No. 321 (1995)
4. Partala, T., Surakka, V.: The effects of affective interventions in human-computer interaction. Proceedings of the Conference on Human Factors in Computing Systems **16** (2004) 295–309
5. Klein, J., Moon, Y., Picard, R.: This computer responds to user frustration: Theory, design, results, and implications. Interacting with Computers **14** (2002) 119–140
6. Ochs, M., Catherine, P., Sadek, D.: An empathic virtual dialog agent to improve human-machine interaction. Autonomous Agent and Multi-Agent Systems (2008) 89–96
7. Bates, J.: The role of emotion in believable agents. Communications of the ACM **37**(7) (1994) 122–125
8. Conklin, D.: Music generation from statistical models. Proceedings AISB Symposium on Artificial Intelligence and Creativity in the Arts and Sciences (2003) 30–35
9. Allan, M., Williams, C.K.I.: Harmonising chorales by probabilistic inference. Advances in Neural Information Processing Systems **17** (2005) 25–32
10. Chuan, C., E.Chew: A hybrid system for automatic generation of style-specific accompaniment. Proceedings International Joint Workshop on Computational Creativity (2007) 57–64
11. de la Puente, A.O., Alfonso, R.S., Moreno, M.A.: Automatic composition of music by means of grammatical evolution. Proceedings of the 2002 conference on APL (2002)
12. Horner, A., Goldberg, D.: Genetic algorithms and computer-assisted music composition. Proceedings of the International Conference on Genetic Algorithms (1991)
13. Tokui, N., Iba, H.: Music composition with interactive evolutionary computation. Proceedings of the Third International Conference on Generative Art (2000)
14. Ponsford, D., Wiggins, G., Mellish, C.: Statistical learning of harmonic movement. Journal of New Music Research **28**(2) (1998) 150–177
15. Phon-Amnuaisuk, S., Wiggins, G.: The four-part harmonization problem: A comparison between genetic algorithms and a rule-based system. Proceedings of the AISB'99 Symposium on Musical Creativity (1998)
16. Delgado, M., Fajardo, W., Molina-Solana, M.: Inmamusys: Intelligent multi-agent music system. Expert Systems with Applications **36**(3-1) (2009) 4574–4580
17. Rutherford, J., Wiggins, G.: An experiment in the automatic creation of music which has specific emotional content. Proceedings of MOSART, Workshop on current research directions in Computer Music (2003) 35–40
18. Oliveira, A., Cardoso, A.: Towards affective-psychophysiological foundations for music production. Affective Computing and Intelligent Interaction (2007) 511522
19. Parrott, W.G.: Emotions in Social Psychology. Psychology Press, Philadelphia (2001)
20. Colton, S.: Creativity versus the perception of creativity in computational systems. Creative Intelligent Systems: Papers from the AAAI Spring Symposium (2008) 14–20
21. Pearce, M.T., Meredith, D., Wiggins, G.A.: Motivations and methodologies for automation of the compositional process. Musicae Scientiae **6**(2) (2002)
22. Csikszentmihalyi, M.: Creativity: Flow and the Psychology of Discovery and Invention. Harper Perennial, New York (1996)