A RULE-BASED GENERATIVE MUSIC SYSTEM CONTROLLED BY DESIRED VALENCE AND AROUSAL

Isaac Wallis, Todd Ingalls, Ellen Campana, and Janel Goodman
School of Arts, Media, and Engineering
Arizona State University
Tempe, AZ, USA 85281
{iwallis | testcase | ecampana | jgoodman}@asu.edu

ABSTRACT

This paper details an emotional music synthesis (EMS) system which is designed around music theory parameters and previous research on music and emotion. This system uses a rule-based algorithm to generate the music from scratch. Results of a user study on this system show that listener ratings of emotional valence and arousal correlate with intended production of musical valence and arousal.

1. EMOTIONAL MUSIC SYNTHESIS

Historically, music has been used in ceremonial, religious, and artistic settings for the purpose of affecting listeners emotionally and setting a mood. Emotional music synthesis is the computerized generation of music which has a recognizable emotion, for the purpose of setting a mood. Music can influence the emotions of listeners in multiple ways [1], but most EMS systems use one of two psychological mechanisms: (1) they manipulate musical expectations, or (2) they manipulate structural or performance features of the music in order to set up a perceptible mood.

In the approach manipulating musical expectations, the melody, harmony, and rhythm of the music are designed so that listeners will begin to expect specific musical events. Sometimes these expectations stem from standard musical theory: examples of these include harmonic cadences, in which specific chord patterns (e.g. the V-I resolution) are common. If these expectations are satisfied, listeners will experience a sense of resolution. If these expectations are not satisfied, listeners may experience emotions such as surprise or of being "left hanging." [2]

In the structural mood-based approach, features of the music are changed in order to create the perception that the music itself has a mood. Over the last century, music psychologists have identified the emotional effects of many musical features such as tempo or harmonic mode; a large list of these features and their mappings can be found in Gabriellson and Lindström [3]. However, most of these studies were performed using western forms of music, such as classical, and enlisted participants who were familiar with those forms of music.

Musical features, and their emotional mappings, can be thought of using a framework of emotional mirroring: when we perceive emotion in an agent, whether that agent be a person, a pet, a movie, or a piece of music, what we are really doing is performing a structural analysis on that agent. Some component of the structural analysis answers the question: "What emotions would drive me, personally, to exhibit these structural features?" If we hear someone singing sweetly we assign a pleasant emotion to that music—the same emotion we would feel if singing that way ourselves. When we hear fast-paced music we assign it the same high-energy emotions we might feel if we were to move our bodies at a fast pace, or watch a movie with a fast-paced editing style, or even drive a car at high speed. The structural feature moving rapidly is applicable to many situations, and results in similar emotional connotations.

There are limits to the emotional resonance of EMS. For example, musical emotions will not always be internalized by listeners. This makes sense in the context of day-to-day life—if the musical mood were always internalized, no one would choose to listen to sad or angry music. It is possible that expectation-based EMS, since it deals with internal expectations of listeners, could be more successful in affecting felt emotions. Another limit is that complex emotions such as shame or jealousy are difficult to convey by music alone, though some evidence suggests that this is not impossible [4]. Narrative in the form of words or visuals must usually be included in order to supply the necessary emotional focus; without context, there is not much to feel jealous of [5]. Also, although evidence exists to suggest that listeners may internally feel multiple conflicting emotions in response to music [6], there is none to suggest that music is capable of having, in itself, simultaneous conflicting moods.

EMS is a fairly new field of interactive music, but examples of working EMS systems do exist. Friberg's pDM system takes, as input, pre-composed musical note-lists and outputs emotionalized versions of them by manipulating features of their performance [7]. Winter created a system expanding on pDM which also manipulates harmonic features of the music in real-time [8]. Livingstone et al. [9] took an approach similar to Friberg and Winter. Oliveiro and Cardoso [10] are in the process of developing an EMS system in which the musical features to be manipulated were derived through analysis of a corpus of emotion-tagged music files. They intend to use these features to transform the emotion in other pieces of music.

Transformative EMS systems like these take pre-existing musical compositions as input, which are then transformed emotionally and output. The mood of the resulting music is different, but the music itself remains recognizable. This document presents an EMS system which generates music from scratch using rule-based algorithmic composition.

2. SYSTEM DESIGN

This paper builds on previous work by evaluating an EMS system which is described in depth in [11]. This section provides a brief description of the parameters and design of this EMS system.

An EMS algorithm was designed to emulate piano accompaniment techniques. Models of emotion and musical emotion were explored for insight into algorithm interfaces. The most rigorously developed and accurate such models tend to be multidimensional models such as the Geneva Emotion Music Scale (GEMS) [6]. GEMS has nine dimensions: wonder, transcendence, tenderness, nostalgia, peacefulness, power, joyful activation, tension, and sadness. These nine dimensions can be factored down to three: sublimity, vitality, and unease. This model was designed for classification of induced musical emotions only, and no attempt was made to form relationships between musical features and emotions.

Although models like GEMS point to an interesting area of interface research, at this stage of our research we decided to use a simpler model as our interface. One of the simplest scientifically validated emotional models is the Circumplex Model [12], a two-dimensional model where the Y-axis corresponds to emotional arousal or intensity and the X-axis corresponds to emotional pleasure or valence. Common emotion words, if rated along these axes by enough people, will fall in a ring around the origin with words like anger in the upper left, joy in the upper right, serenity in the lower right, and sadness in the lower left. Since it is a two-dimensional model, it allows the implementation of a simple canvas interface—depicted in Figure 1—allowing users to select valence and arousal with a single click. This style of interface has been used in other EMS systems, such as pDM [7].

The circumplex model is intuitive and easy to use in interfaces, but it is not entirely emotionally accurate. For example, anger and fear are close to one another on the circumplex, but most people perceive these as very different emotions; far more different than other closely-spaced emotions such as sadness and depression. The only way to overcome this drawback is adding more dimensions to the model, such as a dominance axis separating anger and fear [13]; the logical result of this is a complex model such as GEMS.

Ten musical parameters—pitch register, loudness, rhythmic roughness, tempo, articulation, harmonic mode, and upper extensions—were developed using Gabrielsson and Lindström's work [3] as a guide, then fine-tuned by ear. As most of the musical features in Gabrielsson and Lindström's review are standard concepts from music theory, Persichetti [14] was also a useful resource in parameter de-

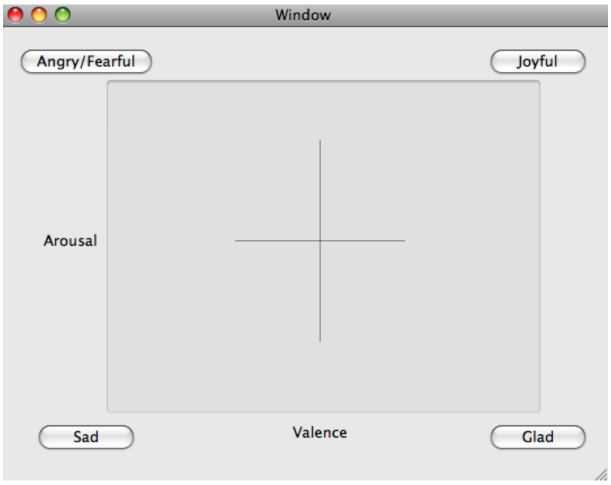

Figure 1. Clickable valence/arousal interface.

velopment. These musical parameters were attached to the valence and arousal of the interface. This algorithm needs direct control in real-time of information regarding rhythm and timing, harmony, and note selection (it does not have melodic parameters because it is designed for chordal piano accompaniment). The following subsections describe these ten musical parameters in greater detail.

2.1 Rhythmic/Timing Parameters

The rhythmic module has three lower-level parameters attached to it: rhythmic roughness, tempo, and articulation. Rhythmic roughness is a musical feature first studied by Gundlach [15]. This parameter determines the variation in note lengths over a measure of music: if all notes are of equal duration, roughness is low, and if notes are of varying length, roughness is high. In this system, at the lowest roughness each measure is populated with sixteen equal-length notes. As roughness increases, pairs of notes are selected at random and joined into single notes (Figure 2). This means that lower roughness correlates with higher note density over time, which has profound effects on the perceived rapidity of the music.

Tempo and articulation are well-known musical timing concepts. In this system, tempo is a global speed value (in beats-per-minute) affecting how quickly the notes of each measure is played—there are four beats per measure. A chord change also takes place upon each measure. Articulation refers to the actual duration each note is played in comparison with its allotted time. Notes with long articulation (i.e. legato) take up all the time available between

Figure 2. (a) Smoothest rhythm, consisting of sixteen equal-length rhythmic events. (b) Rougher rhythm, in which four of the sixteen events are joined at random. (c) Roughest rhythm, in which five additional events (nine total) are joined.

the previous note and the upcoming one, while notes with short articulation (i.e. staccato) take up just a fraction. In this system, when the articulations are set to the longest setting, notes will overlap by extending beyond the allotted times.

2.2 Harmonic Parameters

Two musical parameters were harmonic in nature: upper extensions and harmonic mode. These are each well-known musical terms, although upper extensions is a concept more commonly applied in jazz than other forms of music.

2.2.1 Harmonic Mode

The harmonic mode parameter is discrete, switching between six of the natural harmonic modes: Lydian, Ionian, Mixolydian, Dorian, Aeolian, and Phrygian. Locrian is left out; it is less commonly used in classical or jazz modal composition because it has a dissonant tonic chord. When the parameter changes and a new mode is selected, all chords in the progression are replaced by chords from the new mode which serve the same chord functions—tonic, subdominant, or dominant—as the original progression. In this way, the new progressions sound similar to the old progression, but the flavor of the new mode is introduced. This EMS algorithm is distinctive in its use of six harmonic modes; to our knowledge, other EMS algorithms incorporate only the major and minor modes. This parameter is mapped to valence so that the darkest mode [1] is at the lowest valence setting and the brightest mode is at the highest valence.

2.2.2 Upper Extensions

This system uses chord progressions made of triads. Triads are chords consisting of a scale degree plus the notes at the intervals of a third and a fifth above that root scale degree. On any triad within a mode, there are somewhere between two and four notes—forming seventh, ninth, eleventh or thirteenth intervals with the root—which can be played without forming discordant intervals with any of the three original triadic notes. These extra notes are called upper extensions, and when the upper extensions parameter increases, more of these notes are allowed into the chords played by the EMS system.

2.3 Note Generation

The system tempo controls the rate of an underlying chord progression which is partially determined by the harmonic parameters. When the system needs a new note, any note from the underlying chord could be generated, subject to a few note generation parameters. The first two parameters deal with the volume and thickness of each musical event. The loudness parameter specifies average note velocity, and the voicing size parameter determines how many notes are used simultaneously upon each event; or, put another way, how many fingers the virtual pianist uses on each chord. Voicing size is perceptually confounded with loudness, because playing more simultaneous notes will lead to a louder overall sound.

In addition, there are three parameters related to the rule-based note generation in the system: voice spacing, voice leading, and pitch register. Whenever a note is needed, the system selects a chord-tone from the underlying chord in a probabilistic way, with the probabilities being weighted by the three rules. Since the rules conflict with one another, the result is a music generator which is unpredictable, yet remains within certain musical constraints.

2.3.1 Rule 1: Pitch Register

The new note should be selected from a range which is determined by the pitch register parameter, and should tend to be more frequently selected from the center of this range. This keeps the music from going too high or too low.

2.3.2 Rule 2: Voice Spacing

As this is a piano emulation system, new notes should not play on top of notes which are already playing. In addition, new notes should tend to avoid playing near existing notes to a degree specified by the voice spacing parameter. As this parameter increases, the area of avoidance around existing pitches widens.

2.3.3 Rule 3: Voice Leading

Pianists often follow a principle called voice-leading where new chords are voiced to be as similar as possible, in terms of the intervals and placement on the keyboard, as previous chords; this minimizes the required arm and finger movement. We capture this phenomenon by specifying that new notes will tend to be generated where other notes have been recently released. Notice that this rule conflicts with the previous rule: existing notes repulse new notes, but once those existing notes are released, new notes are attracted to the same area. This conflict would lead to a directionality of note generation, where new notes are generated at an increasingly high or low pitch, but the first rule eliminates this possibility.

2.4 Mappings

Table 1 and Figure 3 show how parameters are manipulated in the EMS system. Each parameter is linearly mapped to valence or arousal in the most continuous possible way; for example, harmonic mode thresholds the valence selection space into six equal regions, while loudness increments between low and high arousal settings. The system is designed so that musical parameters can be easily re-scaled or re-mapped, so the values in Table 1 represent the system in only one configuration. In practice, often a scaled-down subset of parameters was used, in which the voice spacing and voice leading parameters were fixed at a central value across all valence settings. In preliminary study, these simplifications seemed to have little effect on the emotional connotations of the system. Tempo was also fixed at 80 bpm across all arousal settings; the rhythmic roughness parameter alone was sufficient to affect the perceived rapidity and arousal of the music.

[1] According to Persichetti, the modes in order from darkest to lightest are: Locrian, Phrygian, Aeolian, Dorian, Mixolydian, Ionian, and Lydian [14].

	Valence Min	Valence Max
Mode	Phrygian	Lydian
Extensions	2-4 Added Pitches	No Added Pitches
Pitch Register	Centered on C4	Centered on C6
Voice Spacing	Avoid 2nd Int.	Avoid 6th Int.
Voice Leading	w/in 3rd interval	w/in 5th interval

	Arousal Min	Arousal Max
Roughness	9 Joinings	0 Joinings
Tempo	60 BPM	100 BPM
Articulations	Overlapping Notes	Half-Length Notes
Loudness	Velocity 50	Velocity 70
Voicing Size	2 Notes	8 Notes

Table 1. Shows how musical features map to valence and arousal in this EMS system.

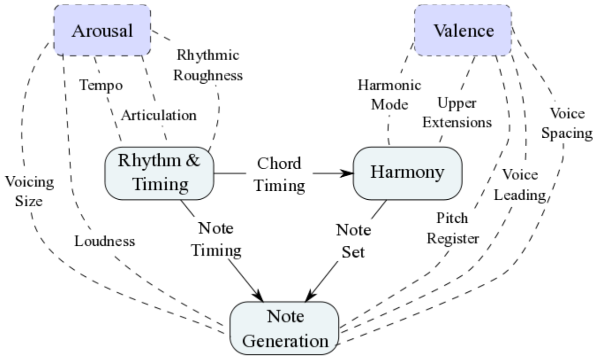

Figure 3. Flow of algorithm operation. Solid edges indicate control or information flow, dashed edges indicate an "implemented by" relationship.

3. EXPERIMENT

A user study was performed in order to determine how well the intended valence and arousal system parameters corresponded with actual listener evaluations of valence and arousal in the music. For this experiment, the scaled-down subset of musical parameters discussed in Section 2.4 was used.

3.1 Participants

Eleven participants—five female, six male—recruited from a student pool at ASU consisting of psychology students, whose ages ranged from eighteen to twenty-one years of age. As per the requirements of recruiting from this student pool, each participant was given class credit in return for participation.

3.2 Procedure

Each participant was stationed at a computer terminal consisting of monitor, keyboard and mouse, and speakers. Several practice trials were completed, and when the participant felt ready the experiment began. Each trial proceeded as follows: First, the EMS algorithm generated music at a specific valence and arousal setting—randomly selected from thirty-six valence/arousal configurations evenly dividing the parametric space into a six-by-six grid—for fifteen seconds. Once the music stopped, a clickable grid appeared on the graphical user interface and the participant clicked at the valence/arousal point which, in his or her opinion, most closely matched the music. After the click, the grid disappeared and another trial began. Four blocks of thirty-six trials were completed by each participant.

3.3 Results

Each trial's X and Y mouse-click location was recorded along with the valence and arousal of the generated music. Correlations and t-tests were performed on this data as depicted in Table 2. Also, graphs were generated showing the mean clicked valence or arousal for every intended valence or arousal setting, and the standard deviations over these means (Figures 4 and 5). Data was not normalized before these operations.

	Intended Valence	Intended Arousal
Perceived Valence	$r = 0.475 (p < .001)$ $t = 36.82 (p < .001)$	$r = 0.364 (p < .001)$ $t = 18.09 (p < .001)$
Perceived Arousal	$r = 0.008 (p < .001)$ $t = 0.44 (p < .66)$	$r = 0.679 (p < .001)$ $t = 23.61 (p < .001)$

Table 2. Shows correlations between perceived and intended arousal and valence.

4. DISCUSSION

This user study shows that our EMS algorithm is fairly well designed since changes to the settings of the valence or arousal parameters resulted in participants hearing corresponding changes in the emotion of the music. One caveat exists, which is the fact that there is some crossover between intended arousal and perceived valence (Figure 5b). Ideally, changes to the arousal parameter would result in no change at all in the perceived valence, but this was not the case. The fact that the crossover is not symmetrical—perceived valence correlates with intended arousal, but perceived arousal does not significantly correlate with intended valence—is interesting. This indicates that the crossover is not a result of mapping a musical parameter to the wrong emotional axis. Therefore, the crossover is probably a result of a non-orthogonal relationship between arousal and valence with regard to one or more of the features. It could also be the result of uncontrolled factors such as cultural effects. In any case, this crossover does not greatly affect the desired operation of the EMS system, because the within-dimensions results behave as expected.

The fact that this EMS system generates music algorithmically instead of transforming pre-composed music has benefits and drawbacks. One benefit is that the generated music is completely new to listeners, therefore previous exposure to the music could not bias user evaluations in this study. A potential drawback is that, although the music in this system is randomly-generated and non-repeating, it still has stylistic similarities with itself. Whereas other EMS algorithms are capable of more variety because they transform pre-existing pieces, this system sounds like a single composition. This leads to the possibility that this

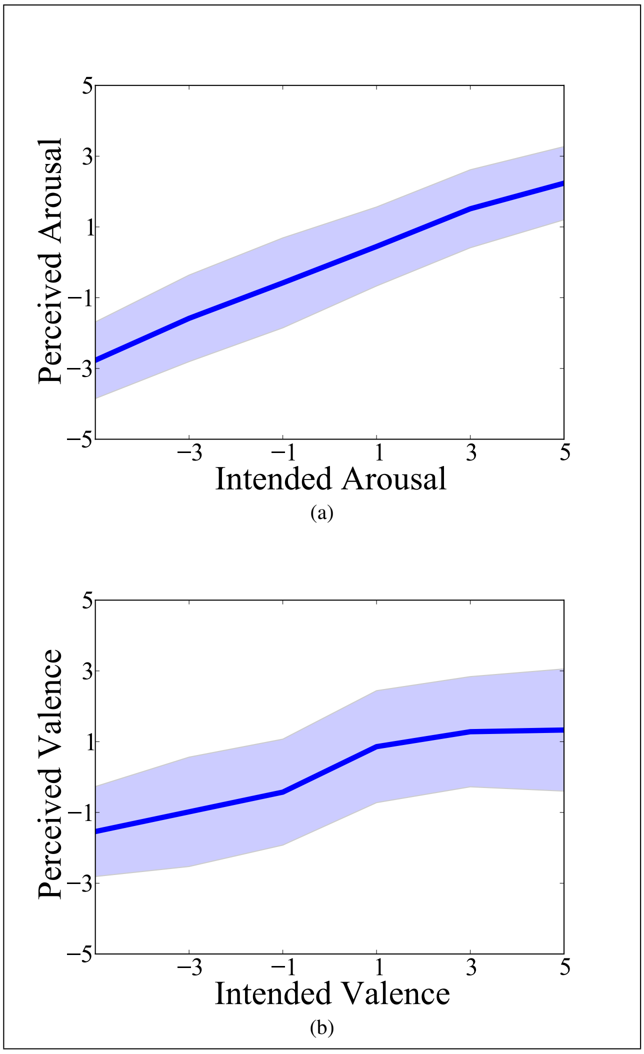

Figure 4. Within-dimensions results of study. These graphs show the mean clicked valence or arousal (and standard deviation above and below the mean), at each generated valence or arousal setting.

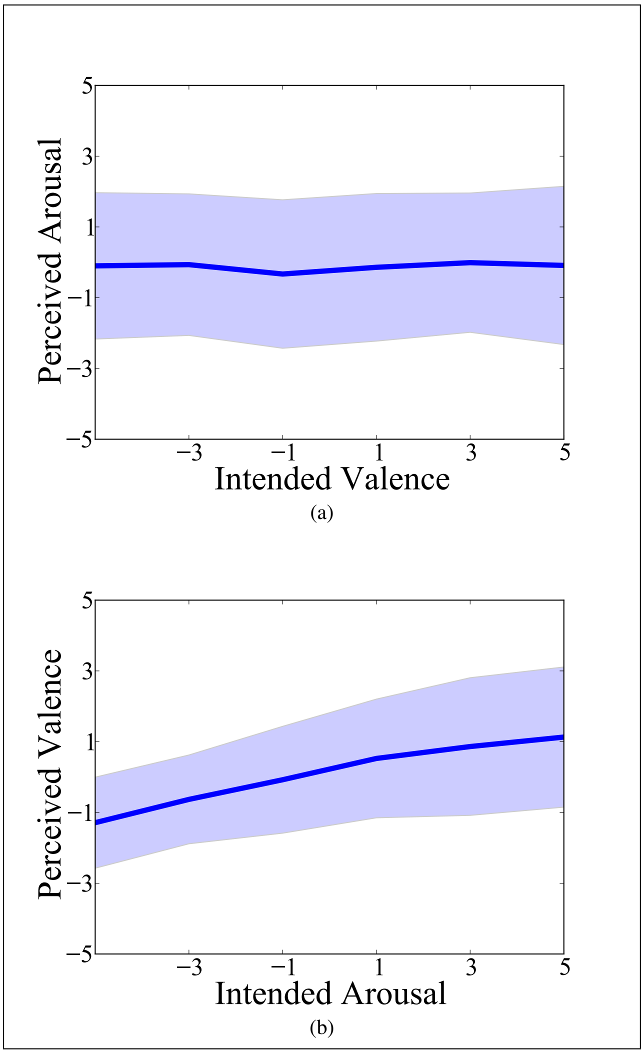

Figure 5. Between-dimensions results of study. These graphs show the mean clicked valence or arousal (and standard deviation above and below the mean) at each generated valence or arousal setting.

music could be emotionally biased; that this entire system, in comparison with other music in the world, may have an emotional value in the same way many people feel certain genres of music are inherently angry or sad (e.g. "hard rock sounds angry").

4.1 Topics for Further Study

Although the musical parameters manipulated by this EMS system affected listener perceptions of musical emotion as expected, the complex relationships between parameters mean that separating the parameters or using them in different configurations will have unpredictable results. For this reason, this user study is presented only as an evaluation of this EMS algorithm; we make no claims about musical emotion in the overall sense. However, during the course of this project, certain insights emerged about EMS design. These need further study as they have not been empirically proven. However, they seem to be plausible hypotheses. These hypotheses are described in the following sub-sections.

4.1.1 Timing and Volume Parameters Should be Mapped to Arousal, While Tonality and Timbre Parameters Should be Mapped to Valence

All system mappings between the valence/arousal space and the musical features were first based on mappings in Gabriellson and Lindström's compilation of emotional music study results, and then were subject to much experimentation before a final design was finalized. In the end, all timing-based parameters (rhythmic roughness, tempo, articulation) and all loudness-based parameters (voicing size, loudness) were mapped to arousal. Similarly, all parameters related to tonality (mode, register, voice spacing) were mapped to valence. Although this division of parameters is simplistic, it seems to suffice for the purposes of generative EMS. The arousal link with timing is in keeping with existing theory on proprioceptive emotional contagion [16]. Also, some evidence suggests that valence and tonal parameters are cognitively linked with language and prosody [17], suggesting that many timbral features of music, which were not directly manipulated in this system, would also map to valence. Studies of timbre with regard to emotion indicate that this is the case, although some aspects of timbre do correlate with arousal [18]. Exploring the salience of this mapping hypothesis—that timing and volume parameters should be mapped to arousal while tonal and timbral parameters should be mapped to valence—is an interesting and necessary direction for future study.

4.1.2 Density is Most Important Parameter For Arousal

Of all musical features, event density seems most important for determining arousal in EMS systems. In our system, event density was dictated by a combination of rhythmic roughness and tempo, but either would have been sufficient by itself to control arousal. Had we mapped only one of these features to arousal with no other features, it seems likely that listeners would have perceived the intended arousal. On the other hand, had we mapped all other arousal features (loudness, voicing size, articulation) to arousal without including either tempo or rhythmic roughness, listeners would not have strongly perceived the intended arousal.

4.1.3 Mode is Most Important Parameter For Valence

In EMS algorithms based on western music, mode will likely be the most important musical feature for determining valence. In Figure 4b, the steepest part of the graph is the center, which divides the Dorian mode from the Mixolydian mode. These modes are very similar, differing only in that the Dorian mode flattens its third scale degree. However, modes with a flattened third degree (such as Phrygian, Aeolian, and Dorian, which are on the left side of the graph) are considered to be minor modes, and modes without a flat third pitch such as Mixolydian, Ionian, and Lydian, which are on the right) are considered to be major modes. Had we mapped only the harmonic mode parameter to intended valence and no other musical parameters, most listeners would still have perceived the intended valence. However, had we mapped all the other parameters except harmonic mode (upper extensions, pitch register, etc.) the intended valence would not have been perceived as strongly.

4.1.4 Perceived Musical Emotion is Relative

It seems possible that emotion perception is relative, meaning that judgments on the emotion of music depend on comparing it to other music. This could be a comparison within a genre (e.g. "this music is angry compared to most jazz"), a comparison within a single composition (e.g. "this part is angry compared to the rest of the song"), or simply a comparison with the music one has been recently listening to. As stated earlier, the fact that this system has limited musical variation means that this study was vulnerable to emotional bias in the music. However, in Figures 4 and 5 it is obvious that the curves in all four graphs are centered on, and nearly symmetrical about, the mid-line. If the music were emotionally biased, and listener perception was not relative, some of these would be offset vertically instead of centered on the mid-line.

Figure 6. Interface to an emotional slideshow generator using EMS background music.

4.2 An Artistic Use of the EMS System

The EMS system in this paper was combined with Flickr [19] to create an emotional slideshow generator, the interface of which can be seen in Figure 6. The interface provides a drop-down list of emotions to select from, each of which has a valence and arousal value which has been determined by Russell [12]. When an emotion is selected, the EMS system begins to play at the correct valence and arousal while any Flickr images tagged with the emotion word are downloaded and displayed one after another. When the selected emotion has high arousal, the images will cycle through more rapidly.

Since image selection is dependent on tagging by Flickr users, there is nothing to ensure the emotional accuracy of the slides. With that said, the generated slideshow is, in the overall sense, remarkably apt. Sometimes unexpected connections are made which serve to highlight the use of social media; for example, setting the interface to serene might result in a slideshow consisting of mostly sleeping babies, sleeping pets, and landscapes—but once in a while, a bottle of pills may appear. In our view, these unexpected images are desirable, as this application was created for artistic/aesthetic purposes, not for any clinical use. To repurpose this application for scientific use, Flickr would need to be replaced with an image database whose images have been rigorously rated in terms of affect, such as the International Affective Picture System (IAPS) [20].

4.3 The Importance of Emotional Music Synthesis

The uses of compelling EMS are numerous and obvious; almost anywhere background music is currently used, EMS could be used. This makes it useful for the purposes of marketing and entertainment. Since EMS can be automated, it could be put to effective use in video games. Since EMS deals with emotions, it might be used in therapeutic systems for emotional imbalances and autism.

There exists a theory of emotional contagion [6, 16, 21] which is proprioceptive in nature, positing that certain structural features observed in outside agents such as music seem emotional because they correspond with movement features. These movement features may be emotional for evolutionary reasons; at some point in evolutionary history, odds of survival increased if high-intensity emotions such as anger or terror correlated with rapid rates of movement. Therefore, if we observe some agent moving rapidly, even if it only moves rapidly in a conceptual sense as in the case of high-tempo music, it can seem as though that agent has a high-intensity emotion such as anger or terror. If this theory is accurate, then affecting a person's emotions will have repercussions in that person's movement. For this reason, EMS might be an especially useful form of interactive music for systems designed to rehabilitate Parkinson's Disease or stroke-induced impairments, such as the stroke rehabilitation system described in [22].

Most interactive media is designed to engage participants. In existing narrative multimedia, such as cinema, background music seems to greatly increase the chances that the audience will become immersed in the story. Therefore, one question of importance to interactive media is the following: How can we best leverage musical emotion in order to make interactive media appealing? Hopefully, in the future, scientists will be able to deploy EMS algorithms such as this one in order to better understand the answer to this question.

5. CONCLUSIONS

In this paper, we discussed the computerized generation of music which communicates with listeners on an emotional level. We presented a system which algorithmically generates emotional music; this system differs from many current EMS systems because it composes the music from scratch rather than transforming pre-existing music. Evaluation of this system shows that listeners perceive the emotions which the EMS system is designed to produce. This project resulted in the generation of some insights on EMS design and new hypotheses for future study. Lastly, this paper discussed the importance of EMS in the context of entertainment, therapy and rehabilitation, and digital media in the context of future uses.

Acknowledgments

This material is partially supported by the National Science Foundation CISE Infrastructure Grant No. 0403428 and IGERT Grant No. 0504647. We would like to thank Laura Gonzales for her helpful comments.

6. REFERENCES

[1] P. N. Juslin and D. Västfjäll, "Emotional responses to music: The need to consider underlying mechanisms," *Behavioral and Brain Sciences*, vol. 31, no. 5, pp. 559–621, October 2008.

[2] L. B. Meyer, *Emotion and meaning in music*. University of Chicago Press, 1956.

[3] A. Gabrielsson and E. Lindström, "The influence of musical structure on emotional expression," in *Music and Emotion: Theory and Research, Series in Affective Science*, P. Juslin and J. Sloboda, Eds. Oxford University Press, 2001, ch. 10, pp. 223–248.

[4] R. Bresin, "What is the color of that musical performance?" in *Proc. Int'l. Computer Music Conference (ICMC)*, Barcelona, Spain, 2005, pp. 367–370.

[5] G. Collier, "Why does music only express some emotions? a test of philosophical theory," *Empirical Studies of the Arts*, vol. 20, no. 1, pp. 22–31, 2002.

[6] M. Zentner, G. Didier, and K. Scherer, "Emotions evoked by the sound of music: Characterization, classification, and measurement," *Emotion*, vol. 8, no. 4, pp. 494–521, August 2008.

[7] A. Friberg, "pdm: An expressive sequencer with real-time control of the kth music-performance rules," *Comput. Music J.*, vol. 30, pp. 37–48, March 2006.

[8] R. Winter, "Interactive music: Compositional techniques for communicating different emotional qualities," Master's thesis, KTH Royal Institute of Technology, 2006.

[9] S. R. Livingstone, R. Muhlberger, A. R. Brown, and W. F. Thompson, "Changing musical emotion: A computational rule system for modifying score and performance," *Comput. Music J.*, vol. 34, pp. 41–64, March 2010.

[10] A. Oliveiro and A. Cardoso, "Modeling affective content of music: A knowledge-base approach," in *Proc. Sound and Music Computing Conf.*, Berlin, Germany, July 2008.

[11] I. Wallis, T. Ingalls, and E. Campana, "Computer-generating emotional music: The design of an affective music algorithm," in *Proc. 11th Int'l. Conf. Digital Audio Effects (DAFx-08)*, Espoo, Finland, September 2008.

[12] J. A. Russell, "A circumplex model of affect," *J. Personality and Social Psychology*, vol. 39, no. 6, pp. 1161–1178, December 1980.

[13] J. A. Russell and A. Mehrabian, "Evidence for a three-factor theory of emotions," *J. Research in Personality*, vol. 11, no. 3, pp. 273 – 294, 1977.

[14] V. Persichetti, *Twentieth-century harmony: creative aspects and practice*. New York: W. W. Norton, 1961.

[15] R. Gundlach, "Factors determining the characterization of musical phrases," *American J. of Psychology*, vol. 47, no. 4, pp. 624–643, 1935.

[16] J. Chen, V. Penhune, and R. Zatorre, "Listening to musical rhythms recruits motor regions of the brain," *Cerebral Cortex*, vol. 18, no. 12, pp. 2844– 2854, December 2008.

[17] N. Cook, T. Fujisawa, and K. Takami, "Evaluation of the affective valence of speech using pitch substructure," *Audio, Speech, and Language Processing, IEEE Transactions on*, vol. 14, no. 1, pp. 142 – 151, January 2006.

[18] T. Eerola, V. Alluri, and R. Ferrer, "Emotional connotations of isolated instruments sounds," in *Proc. 10th Int'l. Conf. Music Perception and Cognition (ICMP 10)*, Sapporo, Japan, 2008, pp. 483–489.

[19] Flickr. [Online]. Available: http://www.flickr.com/

[20] P. J. Lang, M. Bradley, and B. N. Cuthbert, "International affective picture system (iaps): Instruction manual and affective ratings," NIMH Center for the Study of Emotion and Attention, Tech. Rep., 2005.

[21] P. Ekman, R. W. Levenson, and W. V. Friesen, "Autonomic nervous system activity distinguishes emotions," *Science*, vol. 221, pp. 1208–10, September 1983.

[22] I. Wallis, T. Ingalls, T. Rikakis, L. Olsen, Y. Chen, W. Xu, and H. Sundaram, "Real-time sonification of movement for an immersive stroke rehabilitation environment," in *Proc. 13th Int'l. Conf. Auditory Display (ICAD)*, Montréal, Canada, June 2007.