International Journal of Asia Digital Art& Design

Received December 1, 2016; Accepted March 11, 2017

Markov Chain Based Procedural Music Generator with User Chosen Mood Compatibility

Adhika Sigit Ramanto
Institut Teknologi Bandung
Jl. Ganesha No. 10, Bandung
13512060@std.stei.itb.ac.id

Dr. Nur Ulfa Maulidevi, S.T., M.Sc
Institut Teknologi Bandung
Jl. Ganesha No. 10, Bandung
ulfa@stei.itb.ac.id

Abstract
Music is a phenomenon common in most human cultures. In a lot of cases, music is played as an accompaniment to other forms of art and activities, such as movies, video games, theatre, or as simple as background music for restaurants and museums. The music in these cases serve to set the mood the artist intends to make the consumers feel. According to previous studies, there is indeed a link between human emotions and music. One of the case that makes people feel different emotions is through the composition of the music itself. Procedural content generation is a field in computer science which creates a random content or art algorithmically within a set constraint. The goal of this study is to create a system that could randomly generate music that fits the mood from a manual user input. Markov chain is a stochastic model used in modeling the components of music composition. For the procedurally generated to fulfill the mood set by the user, different parameter values for each composition component is allotted for each mood. These components include tempo, pitch range, note values, chord type dominance, and melody notes. The implementation of the procedural music generation system is then evaluated by survey and experiments. The evaluation yielded results which assures the capability of the music generation system to fit the mood input.

Keywords: procedural content generation, music, mood.

1. Background and Research Goal

Music is a commonly experienced phenomenon involving sounds organized in a way that produces melody and harmony which pleases the listener. One interesting aspect about music is that it often can be incorporated as an aural accompaniment to other forms of art or activity which otherwise has no aural component or undisrupted by music. The usage of music is proved to be impactful in setting moods and conveying emotions. For example, melancholy in movie scenes could be amplified by using a slow piano ballad, gamers can feel the tension build up very clearly as the music becomes more layered and sped up, and customers can feel more comfortable and relaxed in a restaurant as smooth jazz plays in the background. There has been studies which show that the moods or emotions affected by music is based on the components of the music composition itself. With recorded music, creators who want to incorporate music into their own work would have a number of limitations. Firstly, a polished and pleasing production of music often incurs a hefty cost. Creating one's own music from composition to mastering requires a lot of tools that can be expensive, by then making a good recording also requires additional skills such as sound engineering and instrument playing. Alternatively, one can copyright the works of other musicians to use in one's own creation, this too can be expensive considering said musician would also need the tools and skills mentioned. In addition, recorded music would have no inherent variety once embedded or incorporated into another work. Movies are sent from studios along with their soundtrack reel, video games are limited to use the set of sound and songs included in their development library when they are released. The solution to this might be the advancements in computer science and technology, which gave way to creations featuring procedurally generated contents. Examples include No Man's Sky, a video game with a procedurally generated universe, fractal digital paintings, and abundant-music.com, an online music generator. With those points in mind, this study aims to create a procedural music generation system which suits the mood chosen by the user and produce a pleasing music that is adequate to a conventionally composed music.

2. Analysis and Solution

There are three main points that needs to be analysed to achieve the goals stated. The first is to understand moods and emotions and its classification, the second is to be able to compose music according to a theory or a widely acceptable convention and how the composition would relate to moods. Lastly, to devise a way to procedurally compose the music while still adhering to said theory

2.1 Mood Classification

Mood and emotion is subject to ongoing research in the field of psychology. While some studies may differ the definition of mood and emotion, the general definition that we are seeking in this paper is more akin to mood. In order to dissect human mood into an adequate input for our procedural music generation system, a scientific classification system is needed. One of the most prominent classification system is from Paul Ekman (2003), a discrete division of emotions based on human facial features when experiencing said emotions. It gained prominence as the animated movie *Inside Out* uses most of the emotions explained in the paper. The study says that emotion can discretely be divided into Joy, Sadness, Anger, Disgust, Fear, and Surprise, the last one excluded in the Disney-Pixar movie. But using this classification would prove troublesome as there is no clear connection between one emotion and the others, therefore parameter tuning in the music composition would be difficult.

Another approach to classifying mood is dimensional classification, this approach argues that for each human emotion or mood there is a connection explaining the proximity of one mood with the other. Fear might be closer to surprise than it is to happiness, but surprise might also be close to happiness, but as far from sadness as happiness is far from it. The dimensional classification was first proposed by Schlosberg and since then it has underwent refinements from Russel (1980), and Watson et al. (1999). The iteration most useful to our purposes is called the Tellegen-Watson-Clark Circumplex. It has been used by Trohidis (2008) and Thompson (2010) to identify moods in music. It identifies moods as eight poles produced by four axes, each axis representing an aspect of mood. The aspects are Engagement, Pleasantness, Positive Affect, and Negative Affect as shown in Figure 1.

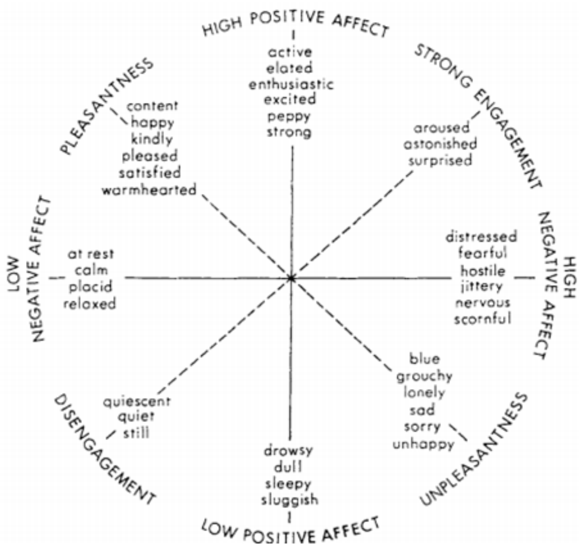

Figure 1 Tellegen-Watson-Clark Circumplex

2.2 Music Composition and Relation to Mood

The general definition of music can be very broad, with genres including noise and experimental defined as music. However, there is a convention well-established in the history of western music that is now accepted worldwide as music theory. The procedural music generation system will follow said music theory in composing music.

There are some key elements to composing music. The key elements we will be using in composing procedural music is tempo, pitch, notes, note values, and chords. Staves and bars are the canvas upon which music is composed. On that canvas the composer write notes. Notes itself has some properties that differentiate one to another.. The vertical position indicates the pitch the note is in, while horizontal position indicates the time it is played as notes are played from left to right of the staves. Each note has its value of the relative duration the note will be played, represented by the shape of the written note. In music, there are fundamental frequencies that makes up the pitch of the notes, these set of pitches are called scales. Scales vary greatly in terms of providing the "correct" notes available to have in the composition, the unifying name for a scale is called a key. However, there is a way to simplify these scales using what is called a solfege. First, we simplify the numerous scales in music by using only keys that have heptatonic scales meaning the scale has seven pitches. Then, music is composed in this solfege which has the note representation Do, Re, Mi, Fa, Sol, La, Si and back to Do. The music composed will hen be able to be transpositioned into any key which has the heptatonic scale. Chords are several notes that are played together as a "mode" sustained usually through a whole bar. The most common type of chord is the tonal triad in which the third ad fifth note in the scale after the first or "tonal" note is played together. In every scale we are using there are three types of tonal triad: the major chord which happens with the 1^{st}, 4^{th}, 5^{th} note in a scale as the tonal note (1maj,4maj,5maj), the minor chord which happens with the 2^{nd}, 3^{rd}, and 6^{th} note in the scale as the tonal note (2min, 3min, 6min), and the diminished chord which happens with the 7^{th} note in the scale as the tonal note (7dim).

Variation of harmonization in music is endless; therefore, we need a minimalistic convention to show that mood can be affected by the elements of composition before venturing through the limitless possibilities of harmonization. Hindemith (1970) proposes a method called Two-Voice Framework in which there are two parts/voices, the lower bass voice which forms a skeleton and foundation for an upper melody voice. In this case we will use chords as the bass voice.

In the research done by Trohidis (2008), a number of these elements has psychological effects on human mood. Tempo affects mood on the Engagement axis of the circumplex, while pitch affects mood on the Pleasantness axis. A faster tempo leads to moods of Strong Engagement while slow tempo leads to moods of Disengagement. Likewise, higher frequency pitches leads to moods of pleasantness while lower frequency pitches leads to moods of Unpleasantness. Another study argues that the famous notion of major chords sounding happy and minor chords sounding sad is true to an extent (Jones et al, 2010). The notion might be true because popular music does use that convention of chords to convey said emotion but there is no direct psychological effects unlike tempo or pitch.

2.3 Procedural Music Generation

In procedural content generation, the content generated must still be comprehensive enough in achieving the goal of creating the same kind of content in the first place. For music, that goal of the procedural composition would be to produce an enjoyable or at least acceptably composed musical piece.

This can be done by adhering to a rule and setting up constraints for generating said content. In the former chapter we analysed music theory and to summarize, there are composition elements that fit better as constraints set in the beginning in neer changed throughout the song such as tempo, and composition elements that can be modeled as variables constantly changing throughout the musical piece such as the current note's pitch and value.

To generate music that fits the user-input mood, the system will use the aforementioned connection of compositional elements to the Tellegen-Watson-Clark circumplex, while the poles of the circumplex itself will be our users' input. This means that here will be eight possibilities of input by users in accordance to the number of poles in the mood circumplex.

Setting up the constants in the generated musical piece would be simple as we only need to allocate a range of possible values for each mood and choose a specific value in that range for composing the mood-fitting music. Compositional elements that will be constants are : tempo, key, pitch range, and chord type dominance. However, we will need to translate the composition variables to something that we can still fine tune the probabilities of the variable having certain values which fits the musical theory and sound good. To achieve this we model these compositional variables into markov chains.

Markov chain is a stochastic state machine in which the probabilities of future events is not dependent on the previous happenings in the state machine itself, in other words, the state that will be chosen as the next state is only dependent on the current state. Compositional elements that will be modeled as markov chains are : solfege (Do,Re, etc.), relative tonal triad chords (1maj,2min,3min), octave movement, and note value. The possible values of these variable compositional elements would be the states of the markov chain. The states of each markov chain and the chosen value of the constants will shape the music composition. The final product of composition will follow Hindemith's Two-Voice Framework. This forms the architecture of the procedural music generation system as shown in Figure 2.

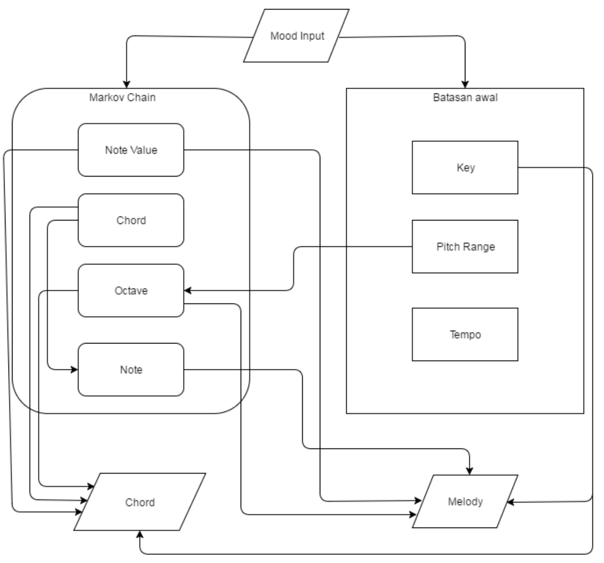

Figure 2 Procedural Music Generation System

As there are eight different mood inputs, we need to allocate a range of possible values for the constants, and the transition probabilities for the markov chain-modeled variables. For tempo, using Harnum's commonly used tempo ranges we have a range of tempo that we can use to conventionally compose music. Tempo is one of the most important factors in affecting someone's mood. Trohidis, et al. (2008) states that tempo directly relates to the Tellegen-Watson-Clark circumplex of emotion. A faster tempo or higher beats per minute (bpm) would bring a movement to the Strong Engagement pole of the circumplex, this would mean that the slower tempo and lower bpm would bring a movement to the opposite side, which is Disengagement. Combining Harnum's tempo convention and Trohidis' research we can allocate a tempo on the Engagement poles as well as the moods between the poles by using the proximity provided by the circumplex model of emotions. Figure 3 is a perspective-shifted circumplex for the convenience of allocating tempo to each mood.

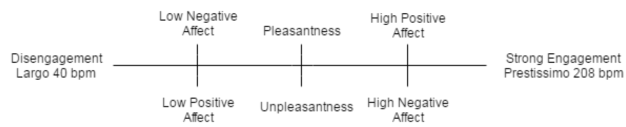

Figure 3 Tempo allocation for circumplex moods

As for pitch, the system will be using piano sounds through MIDI. In this case we will be using the pitch range of an 88-key piano as the universal constraint for pitch. From Trohidis, et al. (2008) it is found that higher pitches causes movement of the circumplex into the area of Pleasantness whereas the lowering of pitch would cause the movement of the circumplex into the area of Unpleasantness. Combining the research and the 88-key piano constraint would give as a range of usable pitch and the emotion that is tied into the range of pitches, using the proximity of different moods provided by the Tellegen-Watson-Clark circumplex, we can allocate a range of pitch for each mood. Figure 4 is a perspective-shifted circumplex for pitch allocation.

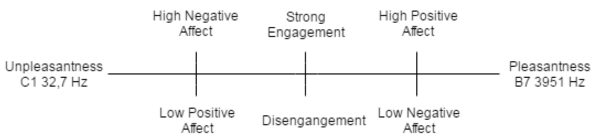

Figure 4 Pitch Allocation for Circumplex Moods

Using supporting adjectives given by Watson et. al. on each arm of the circumplex, then dividing them into positive and negative adjective words, we find that Low Negative Affect, Pleasantness, High Positive Affect, and Strong Engagement more suited to music with dominant major chords while High Negative Affect, Unpleasantness, Low Positive Affect, and Disengagement are more suited suited to music with minor chord dominance.

Table 1 Chord Dominance for Circumplex Moods

Major chord dominance	**Minor chord dominance**
Low Negative Affect	High Negative Affect
Pleasantness	Unpleasantness
High Positive Affect	Low Positive Affect
Strong Engagement	Disengagement

To fill the matrix which houses the transition probability of each state to all the other state including itself we use a likelihood table then translate into the 0 to 1 probability.

Figure 5 is an example of a markov chain that represents note values probabilites for notes generated by the system. Table 2 is its transition likelihood

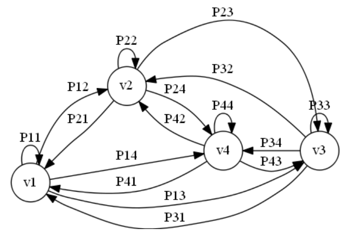

Figure 5 Markov Chain Example for Note Value

Table 2 Transition Likelihood for Markov Chain

Value	Quaver	Crotchet	Minim	Semibreve	percentage
Quaver	1	2	3	3	0.111
Crotchet	1	2	3	3	0.111
Minim	1	2	3	3	0.111
Semibreve	1	2	3	3	0.111

From the likelihood itself we calculate the base value of each likelihood unit before multiplying it with the likelihood value itself into the 0 to 1 range of probability values we have in Table 3

Table 3 Transition Probability for Markov Chain

Value	Quaver	Crotchet	Minim	Semibreve
Quaver	0.111	0.222	0.333	0.333
Crotchet	0.111	0.222	0.333	0.333
Minim	0.111	0.222	0.333	0.333
Semibreve	0.111	0.222	0.333	0.333

As we can see from Figure 2, the states of the markov chains of the left column will decide what properties each note has. Most markov chains change states each beat, with the exception of the chord change markov chain which changes states every bar. The bar used in the procedural music generation system is 4/4 . All songs are composed in its solfege in the C major scale before transpositioned into a random key.

The design for the markov chains themselves will vary according to the needs of the composition. In this study we are aiming for a balance of variability inside constraints that differentiate the moods.

The first component that will be modeled using the markov chain is the note or solfege movement that acts as the main melodic component in the music composition. There are seven solfeges (Do-Re-Mi-Fa-Sol-La-Si) that can be translated as seven states, and these seven states can transition into all the other states including the state the markov chain is in. The transition matrix will be defined by the current chord the composition is in. The transition likelihood of the notes in the tonal triad of the chord will be bigger than the other notes, while the chance to go to the same note will be less than transitioning to different notes.

The second component that will be modeled using the markov chain is the chord movement that acts as the notes for the bass hand in the Two-Voice Framework and to determine the value for transition matrix in the note movement markov chain. There are seven possible conventional chords in a key (1maj, 2min, 3min, 4maj, 5maj, 6min, 7dim), which translates to seven possible states with similar properties as the note movement markov chain. The transition matrix values of the chord movement markov chain itself is determined by the chord dominance for each mood, the dominant chords will have twice as much likelihood than its counterparts. Figure 6 is the shape for the markov chain with seven states in it.

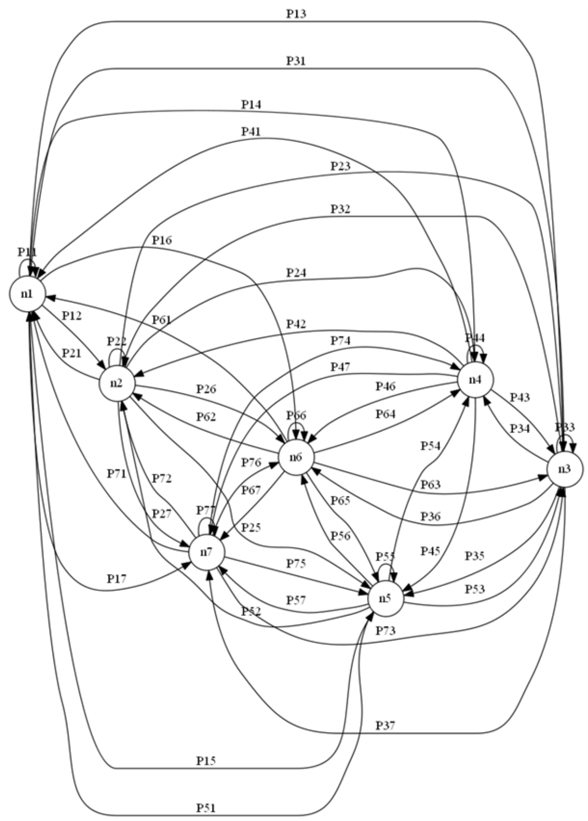

Figure 6 Markov Chain Topology for Note and Chord Movement

The third component that is modeled into the markov chain is the octave movement. By using three octave ranges (Low-Default-High), it enables the composition to be more varied while still being in the pitch range for the mood. The markov chain for octave movement has three states representing the default octave, the lower octave, and higher octave. The likelihood to transition to the default octave will be twice as much than the lower and higher octave. Figure 7 is the markov chain model for octave movement

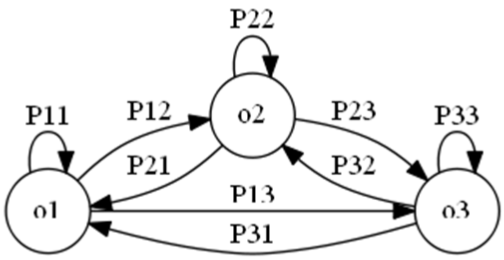

Figure 7 Markov Chain Topology for Octave Movement

The fourth and final component to be modeled using the markov chain is the note value, which has been explained in the example for the markov chain usage above.
The implementation used for testing is built with Java and the Jmusic library as it allows for fine tuning the MIDI according to musical elements where needed without worrying about the low-level parameters required to activate MIDI.

3. Implementation and Evaluation

To test the capability of the designed system we need to implement it into a working application, we then evaluate the application by survey and fine tune the parameters according to the results of the survey

3.1. Implementation

The implementation used for testing is built with Java and the Jmusic library as it allows for fine tuning the MIDI according to musical elements where needed without worrying about the low-level parameters required to activate MIDI.
The classes in the Java implementation are MarkovChain.java as the implementation of the markov chain, EmotionHandler.java which houses the parameter values appropriate for each mood in the Tellegen-Watson-Clark circumplex, MidiHandler.java to handle the initiation of the Jmusic library, and MusicGenerator.java as the main class to handle the input, logic, and output of the application.
After inputting the mood, an iteration begins. Each iteration represents one beat, notes are only manipulated when the note value counter is depleted, chord markov chains are changed every four beats. Songs are always first composed in the solfege of C, before in the end tranpositioned randomly in one of twelve possible keys.

3.2. Evaluation

Evaluation is done through survey. The subject criteria for the survey is male and female from the age of 17 upwards and is in a music organization or has been to a concert at least once. The survey attempts to answer these questions

1. Does the generated music fit the mood chosen?
2. If not, what mood fits the music better?
3. Can you enjoy the music? In other words, is it adequately composed?
4. Can you (the user) tell me what is lacking from the generated music?

The evaluation is done through three waves of survey. The first two waves are followed by experimentation and tuning determined by the results of the survey. Each respondent chooses all eight moods for the system to generate music for which means music is generated 80 times in each of the first two surveys.
The first two waves have each 10 respondents, while the last wave of survey has 25 respondents including some of the people involved in the first two waves.
In the first wave of survey, 62,5% of answers approves that the music generated fits the mood chosen as the input while 76,4% approves that the music is adequately composed. With further inspection to each answer, we find that the lowest approvement rate for mood and music fit are the emotions High Positive Affect with only 20% saying the music fits the mood while Low Negative Affect and Disengagement has the approval rate of 50%. High Positive Affect was fine tuned so it would generate major chords more often than before, while Low Negative Affect is slowed down. Disengagement was fine tuned so the note values generated are often longer.

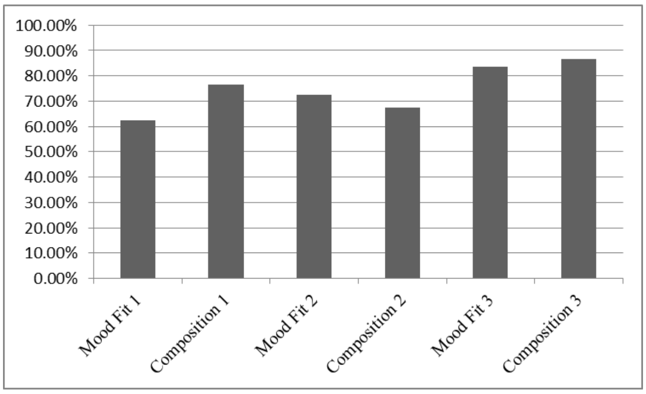

Figure 8 Comparison of Survey Results

In the second wave of survey, 72,5% of answers approves that the music generated fits the mood chosen as the input while 67,5% approves that the music is adequately composed, which is a degradation from the previous evaluation. With further inspection into the fourth question, we find that often respondents hear dissonance in the music, resulting in parts that are not pleasant to listen to and ruins the mood. The problem is solved by explicitly handling dissonant notes, which are immediate notes adjacent to the note being played.
In the third wave of survey, 83,5% of answers approves that the music generated fits the mood chosen as the input while 86,5% approves that the music is adequately composed. These are satisfactory numbers of approval for both goals of the procedural music generation system.

4. Conclusion & Suggestions

From the analysis, design, and evaluation above. There are a few conclusions that we can infer. First, we have analysed that the Tellegen-Watson-Clark circumplex is the most appropriate model in representing mood in relation to music. Secondly, The elements of composition that needs to be manipulated to affect mood are tempo, pitch, and chord type dominance while notes, chords, note values, and octave movement can be modeled with markov chains to simulate variance in music composition. Lastly, the implemented application of the system has a 83,5% mood fit approval rate and 86,5% music composition approval rate, which answers the goals of the system.

This procedural music generator has a lot of potential to be developed further Additionally, this research can be useful if applied for other studies which explores the nature of emotion, mood, and its connection to different kinds of stimuli including but not limited to sound. The applied system can also be useful for the purposes of easily creating music composition or at least can be the base of inspiration for music that needs to fit the mood of a certain work. A few suggestions for further development are as followings:

1.There are a militude of music composition styles as opposed to the classical western style used in this research. Those composition styles can be used as complementary features on the music generation or substitute the compositions style used in this research.

2. The reference used in this research only explores the connection between mood or emotion with the intrinsic part of music which is called the composition and does not explore the connection between the same emotions and the shape of the sound itself, such as the kind of instruments being used in the music generation. For example, an electric guitar playing the same notes as a piano might evoke different emotions. Furthermore, there might be a connection between percussive instruments and emotions that can be explored.

3. This research uses a manually input mood choice by a user that consciously chooses what mood the user wants the music to be compatible with. On further development, this research can be improved by using with other studies which enables the generator to detect the mood of the user automatically. The only caveat is that the study should use the same model of emotions.

4. The values of the transition matrix in this study only uses a convention in music composition that is widely used in composing music, instead of applying different real world cases of music composition as the possibility can be endless. This research can be improved by using machine learning to study acutal compositions across different genres and artists to find the values of the transition matrices that better reflect actual compositions. This can even be done to simulate the composing style of certain artists.

References

[1]J. A. Russel, “A Circumplex Model of Affect.” Journal of Personality and Social Psychology, 1980.

[2]J. Harnum, “Basic Music Theory : How to Read, Write, and Understand Written Music.” 2001.

[3]J. X. Parreira, “Chapter 11 : Markov Chain,” 2009.

[4]N. Shaker, G. Smith, and G. N. Yannakakis, “Evaluating content generators,” 2016.

[5]K. Trohidis, G. Tsoumakas, G. Kalliris, and I. P. Vlahavas, “Multi-Label Classification of Music into Emotions.,” in ISMIR, 2008, vol. 8, pp. 325–330.

[6]W. F. Thompson, L. Quinto, and others, “Music and emotion: psychological considerations,” The Aesthetic Mind: Philosophy and Psychology, pp. 357–75, 2011.

[7]M. Riess Jones, R. R. Fay, and A. N. Popper, Eds., Music Perception, vol. 36. New York, NY: Springer New York, 2010.

[8]H. Von Helmholtz, On the Sensations of Tone as a Physiological Basis for the Theory of Music. Kessinger Publishing, 2005.

[9]T. Adam, “Procedural Music Generation and Adaptation Based on Game State,” California Polytechnic State University San Luis Obispo, 2014.

[10]P. Hindemith, The Craft of Musical Composition: Theoretical Part, vol. 1. Schott & Co Ltd, 1970.

[11]D. Watson, J. Vaidya, and A. Tellegen, “The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence.,” Journal of personality and social psychology, vol. 76, no. 5, p. 820, 1999.