Hak cipta dan penggunaan kembali:
Lisensi ini mengizinkan setiap orang untuk menggubah, memperbaiki, dan membuat ciptaan turunan bukan untuk kepentingan komersial, selama anda mencantumkan nama penulis dan melisensikan ciptaan turunan dengan syarat yang serupa dengan ciptaan asli.

Copyright and reuse:
This license lets you remix, tweak, and build upon work non-commercially, as long as you credit the origin creator and license it on your new creations under the identical terms.
Mood-based Procedural Music Generation Using Markov Chains

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 systems 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 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 undergone refinements from Russell (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](image)

**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 writes notes. Notes itself have some properties that differentiate one from 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 related to the 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 then 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 throughout a whole bar. The most common type of chord is the tonal triad in which the third and 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 1st, 4th, 5th note in a scale as the tonal note (1maj, 4maj, 5maj), the minor chord which happens with the 2nd, 3rd, and 6th note in the scale as the tonal note (2min, 3min, 6min), and the diminished chord which happens with the 7th 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 lead to moods of pleasantness while lower frequency pitches lead 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.

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 this perspective-shifted circumplex for tempo allocation.

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 probabilities for notes generated by the system. Table 5 is its transition likelihood

<table>
<thead>
<tr>
<th>Value</th>
<th>Quaver</th>
<th>Crotchet</th>
<th>Minim</th>
<th>Semibreve</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quaver</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>0.111</td>
</tr>
<tr>
<td>Crotchet</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>0.111</td>
</tr>
<tr>
<td>Minim</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>0.111</td>
</tr>
<tr>
<td>Semibreve</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>0.111</td>
</tr>
</tbody>
</table>

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>
<thead>
<tr>
<th>Value</th>
<th>Quaver</th>
<th>Crotchet</th>
<th>Minim</th>
<th>Semibreve</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quaver</td>
<td>0.111</td>
<td>0.222</td>
<td>0.333</td>
<td>0.333</td>
</tr>
<tr>
<td>Crotchet</td>
<td>0.111</td>
<td>0.222</td>
<td>0.333</td>
<td>0.333</td>
</tr>
<tr>
<td>Minim</td>
<td>0.111</td>
<td>0.222</td>
<td>0.333</td>
<td>0.333</td>
</tr>
<tr>
<td>Semibreve</td>
<td>0.111</td>
<td>0.222</td>
<td>0.333</td>
<td>0.333</td>
</tr>
</tbody>
</table>

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 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 transpositioned 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 approval 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.

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

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.

References