Beat Tracking and Emotion Within Music Signals to Control Lighting

Jedrik D. Eliasen

September 10, 2007

Preface

This thesis was written by Jedrik D. Eliasen as a requirement for the Master of Science in Computer Animation at Bournemouth University.

This thesis was written solely by the author and is his true work. All code and animations included with this thesis are also solely his work.

Acknowledgements

I would first like to thank Jesus Christ for giving me the opportunity to study at Bournemouth University.

I would also like to thank my parents for their love and support, both emotionally and financially.

Finally, thanks for Mya Yee Win and James Whitworth for their expert knowledge in Houdini and Maya respectively.

Abstract

Music and light are two of the stimuli that evoke some of the strongest emotional responses within humans. Due to their shared vernacular and wave natures it is reasonable to think of them together. This document aims to show the author's attempt at creating music-driven, 3D lighting. It will present the implementation of beat tracking in both Maya and Houdini, as well as compare and contrast the performance and usability in both pieces of software. This document will also discuss and present work on emotion detection of audio signals and propose a viable model for the implementation of such detection within a 3D software package.

Contents

Figures

Chapter 1: Introduction

Music and lighting are the two man-made stimuli that are the most emotionally evocative. They are consistently used in cinema, theater, and concerts to provide audiences with a memorable entertainment experience. Many times music is described in terms linked with light, for example, one can talk about the “color of chords” or describe a piece's mood as being “dark” or “light.” In the same way, one can talk about light in musical terms, as in the “harmony of a color scheme.” It is reasonable to deduce from this language, as well as observable emotional responses, that one should be able to use music to control lighting. This can described as a sort of bleeding of the emotionality and movement induced by the auditory sense to musical stimuli into the visual one. Music driven lighting is more feasible and more intuitive than light driven music. This is because music generates a greater emotional response than lighting and because the full range of signal processing techniques can be used in such a system.

Chapter 2 will provide background information and on emotion. Section 1 will talk about what emotion actually is, different theories of emotion, and how this relates to both the brain and the body. Section 2 will delve a bit into musical theory and also describe the role of music as an emotion stimuli. Section 3 will talk about lighting and it's uses within a concert experience.

Chapter 3 discusses beat induction. Section 3.1 is an introduction into beat induction. It explains the various roles and types of rhythm in music and how these differences are important in beat induction. Section 3.2 is a review of work already done on beat induction. Section 3.3 presents the Frederic Patin's algorithm and my implementation of that algorithm in both Maya and Houdini 3D software packages. In Section 3.4 I will present the results of the implementation. Section 3.5 will then compare Houdini and Maya, and the advantages of using one software over another for this kind of analysis. Section 3.6 will talk about the problems and Section 3.7 will conclude the chapter.

Chapter 4 will present and discuss a potential emotion detection system. Section 4.1 review current work on emotion detection. Section 4.2 explains my thoughts on the use of emotion detection with 3D software as well as its use with lighting.

Chapter 5 summarizes the paper and provides recommendations for future work.

Chapter 2: Emotion

2.1 Emotion, the Body, and the Brain

Most every human experiences emotions. We get angry when insulted, sad when a loved one dies, happy when something good happens, embarrassed when we act foolishly, and overwhelmed by beauty in nature. We are constantly stimulated by the world outside of ourselves and adjust our internal states accordingly. Humans regularly take this regulation for granted. Since music derives much of it's importance from emotion, an overview of the role of the brain and the body on emotion is provided.

Currently there are many theories on what emotions are and how they are created. Most of these theories are philosophical in nature, but are not argued "in isolation from the approaches of other disciplines, particularly psychology, neurology, evolutionary biology, and even economics."[1] The two classic theories of emotion are the James-Lange and the Cannon-Bard Theories of Emotion. The James-Lange theory states that emotions arise from the physiological changes that occur as the body responds to environmental stimuli. "We lose our fortune, are sorry and weep; we meet a bear, are frightened and run; we are insulted by a rival, and angry and strike."[2] The Cannon-Bard theory is the James-Lange's antithesis. It states that people first feel an emotion which produces a physiological effect.

Most of today's current theories can be grouped into different types of theories: feeling theories (of which the James-Lange and Canon-Bard are a part), appraisal, cognitivist, perceptual, computational, and dynamical systems' theories.[3] Appraisal and computational theories are especially helpful in the current study.

> Appraisal theories can be described as taking a functional approach to emotion...This approach suggests that the space of emotions can be conceptualized as multidimensional...however, so-called dimensional theories simplify the problem of representation by reducing these to just two or three. Typically these include 'arousal' and 'valence'.[4]

The use of arousal and valence is useful when considering how a musical stimuli effects emotions. This is because arousal and valence are easily mapped on to a coordinate system with valence as one axis and arousal as the other. Such a mapping lends itself well to any algorithmic solutions developed involving the use of dynamically generated emotional labels.

1 de Sousa, Ronald. "Emotion." *Stanford Encyclopedia of Philosophy.* Introduction: Emotion
2 Glynn, Ian. *An Anatomy of Thought*. pg. 334.
3 de Sousa, Ronald. "Emotion." *Stanford Encyclopedia of Philosophy.*
4 de Sousa, Ronald. "Emotion." Stanford Encyclopedia of Philosophy. Section: 4 Psychological and Evolutionary Approaches

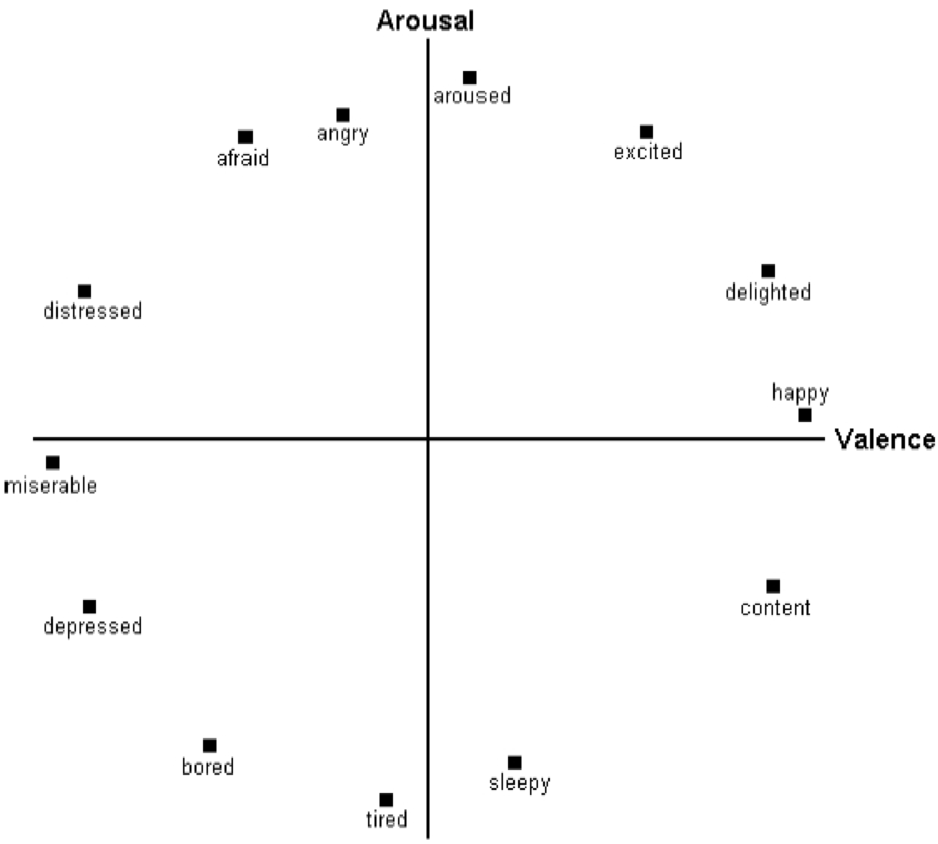

Figure 2.1 – the coordinate system created by arousal and valence and the location of emotions on the graph.[5]

Using design architecture normally used in computer systems, Aaron Sloman et al state "The study of designs satisfying requirements for autonomous agency can provide new deep theoretical insights at the information processing level of description of mental mechanisms. Designs for working system...can systematically explicate old explanatory concepts and generate new concepts that allow new and richer interpretations of human phenomena."[6] If such a model is completed later on, then the implementation of an emotional simulation will be able to be created, thus allowing a comprehensive simulation of human emotional response from not just musical or light stimuli, but from many addition stimuli as well. Thus, stimuli would act as parameters in a program to create emotional reactions.

Regardless of the theory or the theorist, it is recognized that emotion is inexorably tied to physiology and the brain. The Online Medical Dictionary defines emotion as "a strong feeling, aroused mental state, or intense state of drive or unrest directed toward a definite object and evidenced in both behaviour and in psychologic changes, with accompanying autonomic nervous system manifestations."[7]

The processing of emotions in the brain is a very complex process. The limbic system,

5 Korhonen, Mark David. *Modeling Continuous Emotional Appraisals of Music Using System Identification.* pg 9.
6 Wright, Ian and Aaron Sloman, Luc Beaudoin. "Towards a Design-Based Analysis of Emotional Episodes." Abstract
7 http://cancerweb.ncl.ac.uk/cgi-bin/omd?action=Search+OMD&query=emotion

which consists of the amygdala, hippocampus, medial thalamus, nucleus accumbens, and basal forebrain, is responsible for emotion, motivation, behavior, and various autonomic functions. Of these areas of the brain, the amygdala is one of the most important.

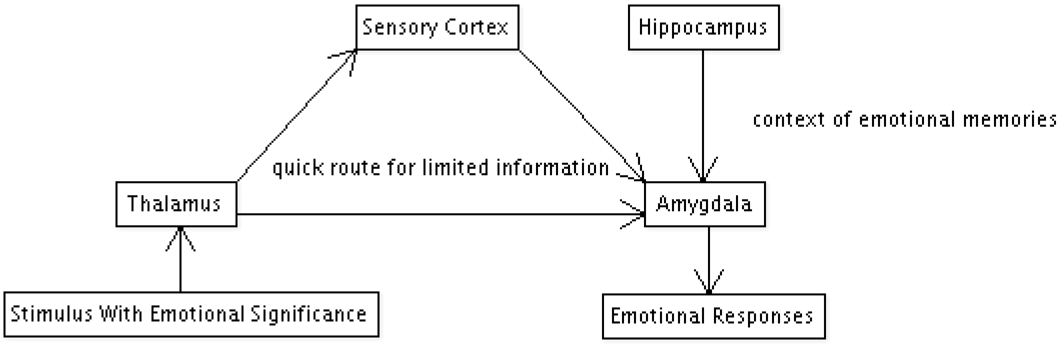

Figure 2.2 – The method by which emotion responses are created by the amygdala depends from the information comes.[8]

The amygdala can receive emotional input from three different locations in the brain: the thalamus, sensory cortex, and hippocampus (see Figure 2.2). Experiments have shown that the thalamus initially receives emotional stimuli. From there the information can be passed either immediately to the amygdala or to the sensory cortex.[9] Information passed to the sensory cortex allows for further processing of incoming stimuli; thus allowing a more reasoned, emotional response. The hippocampus is important in the formation of memories. When memories are formed, emotional associations get attached to them. This means that when people experience stimuli they may feel the emotion attached to the stimuli from a previous experience.

Emotions are felt through physiological responses because:

> The subcortical emotion-processing parts of the brain affect the rest of the body through two basic mechanisms: the release of chemical molecules into the blood that act on various part of the body; and the spread of neural activation to various brain centres and muscles.[10]

Thus certain physiological responses are associated with certain emotions. For example, fear is usually experienced with such responses as cold sweat, increased heartbeat, and reactional hyper-responsiveness.

2.2 Music Theory and Emotion

2.2.1 Music – An Emotional Stimulus

Emotional stimuli not only come from our environment and other organisms. Music, the art of sound that unfolds through time, does not exist in nature, yet it too, causes emotion. It

8 Glynn, Ian. *An Anatomy of Thought.* pg. 339.
9 Glynn, Ian. *An Anatomy of Thought.* pg. 338-339.
10 Trainor, L.J. and L.A. Schmidt. “Processing Emotions Induced By Music.” pg. 312.

has been apart of human life in every culture around the world, in distinct forms, since before history started to be recorded. Flutes thousands of years old have been found at various location around the world.[11,12] There is no doubt that music touches each us differently than any other art or sound. We swoon over a beautifully woven melody and cringe over thundering dissonance. But is this a matter of biology, culture, or choice?

Robert Jourdain, in his book *Music, The Brain and Ecstasy,* uses the discrepancy theory of emotion to demonstrate music's ability to rouse emotional responses. Actions anticipate results that will resolve themselves in either fulfilling or unfulfilling ways. The emotionally neutral state, which occurs before the action, will either become a positive state after the action if anticipations are met or a negative state if anticipations are not met. He says that "[m]usic sets up anticipations and then satisfies them. It can withhold its resolutions, and heighten anticipation by doing so, then to satisfy the anticipation in a great gush of resolution."[13]

Many different musical features contribute to it's emotionality. This can be seen research done on emotion detection of audio signals. Different numbers of musical variables are used in differing models to map emotions to music. Mark David Korhonen used 16 features;[14] Tao Li and Mitsunori Ogihara used 30 features[15]; and the online radio service Musicovery uses 40 parameters, each of which is capable of taking 10 values each.[16] The features normally include, but are not limited to, dynamics, mean pitch, variation in pitch, timbre, harmony, articulation, tempo, texture, vibrato, note onset, melodic contour, and rhythm. This will be discussed more in Section 4.1.

2.2.2 Consonance and Dissonance

Consonance and dissonance also play an important role in creating musical expectations, though they can be confusing concepts to understand. This is because they are constantly given multiple definitions and meanings. In it's simplest, and perhaps truest, definition consonance is a stable sound or one that is pleasant. Dissonance, then, would be a sound that is unstable and hence unpleasant. The most consonant chords in the tonal system are the octave, perfect fourth, and the perfect fifth; less consonant chords would be the third and the sixth. However, it must be noted that consonance is mutable across time and culture. "Although there are important physical and neurological facts important to understanding the idea of dissonance, the precise definition of dissonance is culturally conditioned."[17] What was once considered dissonant in Western culture is now considered consonant.

11 http://today.msnbc.msn.com/id/3077403/
12 http://www.bnl.gov/bnlweb/pubaf/pr/1999/bnlpr092299.html
13 Jourdain, Robert. M*usic, The Brain and Ecstasy*. pg. 311-312.
14 Korhonen, Mark David. *Modeling Continuous Emotional Appraisals of Music Using System Identification.* pg. 31-33
15 Li, Tao and Mitusonit Ogihara. "Detecting Emotion in Music."
16 Personal communication: see Appendix A.
17 http://en.wikipedia.org/wiki/Consonance_and_dissonance

Dissonant chords have a need to resolve, or transition, into consonant ones. This is because the music wants to find a more stable chord on which to stand. Our ears are wont to this transition, from unstable to stable, so the longer the dissonant chord is held the more tension it will create. This is one of the most important ways that consonance / dissonance elicits emotion in a piece of music.

Musical consonance can be further broken down into sensory consonance and harmony. Terhardt explains sensory consonance as “the auditory phenomenon that sounds of any kind in general differ with respect to how 'pleasant', or 'unannoying' they are to a listener.”[18] is directly effected by roughness, sharpness, and tonalness, the most important of which is roughness. Harmony refers to the use and study of relationships between different simultaneous sounds. It normally is thought of as multiple pitches producing a sound pleasing to the ear. It also aims to provide rules for producing pleasing pitch relations.

2.2.3 Biological Issues

In terms of music's biological innateness, it is a known fact that music induces autonomic action; however, whether it can be deduced which emotion is felt from a musically created physiological response is still debated. According to various studies, the autonomic signatures associated with specific emotions differed when stimuli came from musical sources as opposed to other sources.[19] Cerebrally, it has been demonstrated that many of the same cortical areas that process emotion from musical stimuli are the same as those that process emotion from other stimuli.[20] However it is not merely from these emotional centers that we derive pleasure from music.

> The brain must reassemble the sound components detected by auditory cortex. It does this by modeling relations between the various features -- frequency, intensity, location, rates of change -- and then relations among those relations, and so on... it is the act of modeling deep relations among sound components that constitutes our final comprehension.[21]

It is these relations that keep the listener's attention and help give meaning to the music, and it is within this meanings that much emotional content is transmitted. However, the research that shows how the brain produces emotion from music is limited. It is still unclear as to why or how music causes emotional processing in the brain, we simply know that it does.

2.3 Lighting

Lighting is an art form that specializes in the conveyance of mood, the creation of ambience, and the direction of attention. However, it also performs distinct roles depending on

18 http://www.mmk.ei.tum.de/persons/ter/top/senscons.html
19 Trainor, L.J. and L.A. Schmidt. “Processing Emotions Induced By Music.” pg. 313.
20 Trainor, L.J. and L.A. Schmidt. “Processing Emotions Induced By Music.” pg. 314, 316.
21 Jourdain, Robert. *Music, the Brain, and Ecstasy.* pg. 55.

the time, place, and purpose of an event. For example, in cinema, lighting is one of the key methods by which emotion and mood are visually created and communicated to the audience. Theater lighting is just as encompassing, acting as the source of illumination, divulging spatial and temporal information, as well as helping establish composition. Since each form of lighting has it's own needs, and because I aim to focus on the creation of music-driven lighting, I shall focus my attention on concert lighting.

Concert lighting is used to enhance the concert going experience. Lighting designer John Schlick elaborates this point well:

> Most people go to SEE a concert. Yes, there are people who will say they want to go hear a band, but the active word, most of the time is SEE. Most people want for the show that evening to be an experience, they want it to be a memory...Philosophically, what it means when a band gets onstage is that they are prepared to offer the audience an evening of spectacular, memorable entertainment...this performance philosophy implies is that artists ought to strive to get the most out of the music visually as well as acoustically.[22]

Just as lighting in theater or cinema can will direct attention to a specific event spatially, the the lighting will highlight certain events temporally for the concert. “Lights that change offbeat usually are more distracting than enhancing to a show.”[23] The light also acts to establish the mood or feeling of a specific song and also reflects the persona of the band. The lighting for a band like AC/DC is going to be completely different to the lighting for Julio Iglesias. Pink Floyd concerts offer a great example of the marriage lighting and music can attain. These concerts are not merely musical events, but music and lighting events. Both music and lighting are impressive spectacles on their own, but when combined the sum is greater than the parts.

Lighting is directly associated with color; since I aim to utilize music in the dynamic production and movement of light, it would make sense to have the color of the lighting match the color of the music. This will be explained in more detail in Section 4.3.

22 http://www.exotic-lighting.com/WhyHireLD.html

23 http://www.exotic-lighting.com/HowLDRunShow.html

Figure 2.3 – Pink Floyd performing during Live 8 2005.[24]

24 http://www.nrk.no/musikk/1.1881311

Chapter 3: Beat Tracking

3.1 Lighting and the Beat

Finding the beat in a piece of music is useful for lighting. During a concert (or consequently at a club) the lights on a light rig will sway back and forth in time to the music. A viable method for accomplishing this movement 3D could be done using a sine wave who periodicity matches the beat of a given piece of music. When lights change colors they also will do so on the beat, often cycling through 2-3 colors on a larger period that is a multiple of beat-period. It is also helpful to find out when the “colorful” or “flavorful” accents occur so that movement and color switching do not become tired or repetitious.

3.2 Beat Tracking vs. Beat Induction

Music is an art form that takes place through time. Rhythm is music's grammar: it syntactically organizes music into coherent and comprehensible patterns, guides the ear and brain by highlighting the music's most important features, and modulates the listener's attention accordingly.[25] The most basic unit of rhythm is the beat or pulse. This is the constant ticking that regularly continues throughout a piece of music. In Western music beats are grouped within a larger rhythmic structure called meter. Meter is a hierarchical grouping of beats into a regular pattern consisting of strong and weak beats. In a musical score the meter is represented by the time signature, id est 4/4, 3/4, 6/8, etc. Eventually, the movement of the meter becomes inculcated into the listener's mind such that the music need not adhere to it's pattern and the listener will still perceive the same meter. When this happens beats can occur that go against the meter, producing what is called a syncopation. It is important to note that this break down of rhythm is a cultural phenomena. For instance, raga based music in India is rhythmically divided into talas, which consist of patterns that usually are considerably longer than those found in Western music. “Moreover, its subdivisions consist of units of unequal length that combine to form a freely flowing musical continuum within the tala.”[26]

Discussing the extraction or detection of rhythmic elements from a piece of music is difficult. That is because the literature constantly confuses and ascribes multiple uses and definitions to the same set of terms. Authors and researchers on beat induction and beat

25 rhythm, Music and the Brain. 6 section 1.3
26 http://encarta.msn.com/encyclopedia_761553014_2/Musical_Rhythm.html

tracking use the idea of tapping ones foots along with the music as a valid example of the process they are researching, which adds to the confusion. Goto and Muraoka[27] write that beat tracking systems “recognize temporal positions or quarter notes, just as people keep time to music by...foot-tapping.” Ellis[28] begins with “Beat tracking – i.e. deriving from a music audio signal a sequence of beat instants that might correspond to when a human listener would tap his foot.” Meudic[29] says the aim of beat tracking is to extract the beat from a music file. Gouyon et. al[30] describe pulse (beat) induction as the highlighting the intrinsic periodicities of musical features, while beat tracking deals with “strategies to cope with deviations from constant tempo.” Davies[31] describes beat tracking as the process which tries to recover “a sequence of beat times from a musical input consistent with human foot taps.” In another paper, Dixon[32] writes that beat induction is the ability to “subconsciously tap a foot time with the music.” Desain and Honing[33] define beat induction as “the process in which a regular isochronous pattern (the beat) is activated while listening to music.”

Many of these same authors, nevertheless, relate a difference between beat induction and beat tracking. The most helpful definition comes from Dixon and Cambouropoulos:[34] “The action of find an appropriate beat rate (tempo) will be referred to as beat induction, whereas the task of following the beat, that is finding the locations of beats, will be referred to as beat tracking.” According to this definition, beat / pulse induction would be better described as meter induction. I will be using this definition throughout and

3.2.1 Literature Review

Gouyon and Herrera[35] present a method of detecting a beat using an musical audio signal. They first define the beat as “'the sequence of equally spaced phenomenal impulses which define a tempo for the music.'” This beat is characterized by a phase and period. The model they developed takes features from several others researchers: it makes use of the autocorrelation method advocated by Brown, but extends J. Foote and S. Uchihashi's use of representing signals by low-level descriptors computer frame-by-frame. However, they make no use of onset detection.

The beat period is perceived as a multiple of the tick period. The tick indexes of the signal are fed to the algorithm, as well as the upper and lower bounds for the tempo. They then go through six phases: 1) computation of the low-level, energy features; 2) computation of tick features; 3) autocorrelation of features; 4) periodicity seeking in autocorrelation functions;

27 Goto, Masataka and Yoichi Muraoska. “ A Beat Tracking System for Acoustic Signals of Music.”
28 Ellis, Daniel P.W. “Beat Tracking with Dynamic Programming.”
29 Meudic, Benoit. “A Causal algorithm for beat-tracking.”
30 Ellis, Daniel P.W. “Beat Tracking with Dynamic Programming.”
31 Davies, Matthew E.P. “Context-Dependent Beat tracking of Musical Audio.”
32 Dixon, Simon. “Beat Induction and Rhythm recognition.”
33 Desain, Peter and Henkjan Honing. “Computational Models of Beat Induction: The Rule-based Approaches.”
34 Dixon, Simon and Emilios Cambouropoulos. “Beat tracking with Musical Knowledge.”
35 Gouyan, Fabien and P. Herrera. “A Beat Induction Method For Musical Audio Signals.”

5) selection of the beat, period, and phase for each feature with the highest confidence; and 6) the building of a weighted histogram through integration of the results.

According to the results of their experiments, their algorithm detects beats fairly accurately. They recommend that the periodicities sought from 4) should be constructed from periodicities in frequency subbands rather than the whole frequency range and further comparison with other models and the creation of an annotated database.

3.3 Implementation

3.3.1 Patin's Algorithm

In order to accomplish this goal of creating a beat induction program, I first created a stand-alone C++ program, implementing Frederic Patin's algorithm as presented on GameDev.net.[36] I then took the core of my program and converted it into a Maya Node and then as a Houdini Chop Node. I used FMOD Ex libraries to handle to music analysis for both plug-ins. For Maya this was self-evident since it does not have any built in music analysis capabilities, however, Houdini does. I chose to stick with FMOD Ex in Houdini because the Houdini Development Kit (HDK) is not very well documented, especially for CHOPs; the few example nodes provided with the HDK are good for learning how to setup a CHOP node, but contain little to no explanation of the Houdini types and functions themselves. Because I started this phase of the project late in the term, my main priority was to produce a working CHOP node so that I could compare it with Maya.

Patin's algorithm starts with an assumption: a sound will be registered by the brain as a beat if it's energy is "largely superior" to the sound's energy history. This can be accomplished by comparing the "instant" energy with the average energy of the last second. He chooses length of a second somewhat arbitrary, any length, so long as it results in a "nearby" time length, will do. This time length must be "nearby" because of changes in volume, tempo, rhythm, time signature, etc.; a song may have many different types of beat throughout the piece. If the instant energy, using the current N_i samples, is greater than the average energy, the previous N_{ae} samples, we have found a beat.

It can be daunting to keep a buffer N_{ae} samples since N_{ae} may be as large a number as 44100 or more. It is better to add the average of the N_i samples to the energy buffer and not the samples themselves. This will result in a buffer with N_{ae} / N_i values as opposed to N_{ae}. In Patin's terms, the algorithm is also "colorblind" at this point because the algorithm cannot accurately detect beats for "noisy" music like heavy metal. He explains using the example of a guitar and flute alternating between making notes of constant amplitude:

> Each time the first finishes the other starts. The note made by the guitar and the note made by the flute have the same energy but the ear detects a certain rhythm because

36 Patin, Federic. "Beat Detection Algorithms." GameDev.net.

the notes of the instruments are at different pitch. For our algorithm...it is just an amplitude constant noise with no energy peaks.[37]

This situation is remedied by using a Fast Fourier Transform on the N_i samples to move from a time domain to a frequency. The frequency can then be divided into X number of subbands, on each of which the average energy can be computed. The instantaneous energy for a certain frequency can then be compared with the average energy for a particular subband.

The algorithm can be summarized by the following steps:

1. Sample the left and right channels of the signal using N_i samples. Find their average.

$$LE_i[N_i] = \text{sample}(\text{ Left}, N_i)$$
$$RE_i[N_i] = \text{sample}(\text{ Right}, N_i)$$
$$E_i[N_i] = \frac{LE_i[N_i] + RE_i[N_i]}{2}$$

Figure 3.1 – Step 1

2. Run a Fast Fourier Transform on the average samples. Divide the samples into X subbands.

$$E_s[j] = \frac{X}{N_i} * \sum_{k = j * (Ni / X)}^{(j+1)(Ni / X)} E_i[k] \quad , \quad 0 < j < X - 1$$

Figure 3.2 – Step 2

3. Compute the average sound energy (E_a) of the previous N_{ae} samples of the signal.

$$E_a = \frac{N_i}{N_{ae}} * \sum_{k=0}^{(Ni / Nae) - 1} E_s[k]$$

Figure 3.3 – Step 3

4. Compare the instant energy to the average energy multiplied by the sensitivity constant. If $E_i > C * E_a$ then a beat has been detected.
5. Remove the oldest N_i samples from the history buffer and add the current N_i to it.

The core implementation of the algorithm is the same for both Maya and Houdini when using FMOD Ex. It is the manner in which the information is used and the amount of information available that changes with each program.

37 Patin, Federic. "Beat Detection Algorithms." GameDev.net.

3.3.2 Pseudocode

1) Initialize the FMOD Sound system. Create and initialize a system object.

```
result = FMOD::System_Create(&system);
ERRCHECK(result, "FMOD::System_Create");
result = system->init();
ERRCHECK(result, "system->init");
```

2) Load the music file into memory

```
LoadFileIntoMemory(songName, &musicBuffer, &bufferLength);
```

3) Create and initialize a sound object to the music file in memory.

```
result = system->createSound(musicBuffer, use_hardware,..., &sound);
ERRCHECK(result, "createSound");

// Free the memory from the buffer
free(musicBuffer);

result = sound->getLength(&songLength, milliseconds);
ERRCHECK(result, "getLength");
```

4) Play the song (we can only use the FMOD getSpectrum function as the song is playing)

```
result = system->playSound(free_channel, sound);
ERRCHECK(result, "playSound");
```

5) While the song is playing:

```
do
{
```

5a. use getSpectrum to get the frequency domain for the left channel (if there is only one channel then skip to setp D) store this into an array

```
result = system->getSpectrum(leftchanneldata[], numSamples, LEFT,
        use_dsp_fft_window_rect);
ERRCHECK(result, "getSpectrum");
```

```
if(song is stereo)
{
```

5b. use getSpectrum to get the frequency domain for the right channel store this into an array

```
result = system->getSpectrum(rightchanneldata[], numSamples, RIGHT,
            use_dsp_fft_window_rect);
ERRCHECK(result, "getSpectrum");
```

5c. average the two spectrums arrays together.

```
for(int count = 0; count < numSamples; count++)
    averagechanneldata[count] = (leftchanneldata[count] + rightstereodata[count]) * .5;
} // end if
```

5d. Loop through the spectrum array for a certain number of frequencies and sum the amplitudes of the frequencies.

```
for(int i = 0; i < numSubbands; i++)
{
```

5e. Set the sum equal to the energy converted average.

```
        for(int j = i*numFreqPerSubband; j < (i+1)*numFreqPerSubband; j++)
                sum += averagestereodata[j];

        energy[i] = calculateEnergy(sum);
```

5f. Push this onto an array of deque (each deque represents a subband) located at position i

```
        Subband[i].push_back(energy[i]);
} //end for
```

5g. For each subband

```
for(int count = 0; count < SUBBANDS; count++)
{
```

$5g_1$.Average each value in the deque together to get the average energy for the subband

```
for(int i = 0; i < SAMPLEGROUPSPERSEC; i++)
{
    // Sum the energy of freq count.
    ave += the32of43[count].at(i);
}
// Now we have an array with the average energy for each freq subband
aveeng[count] = (1.0 / SAMPLEGROUPSPERSEC) * ave;
ave = 0.0;
```

$5g_2$. Remove the oldest energy value from current subband

```
        the32of43[count].pop_front();
    } // end for
```

5h. Loop through each subband and compare average energy for that subband with the instantaneous energy. If the instantaneous energy is greater, we have a beat.

```
    for(int count = 0; count < SUBBANDS; count++)
    {
        if(energy[count] > (BEATSENS(context.myTime) * aveeng[count]))
        {
            if(count == cf)
            {
                btae.timeMS = ms;
                btae.energy = energy[count];
                if(!isVectorMember(beatinfo,btae))
                    beatinfo.push_back(btae);
            }
        }
        else
            break;
    }
} while(ms < songlength);
```

6) Release the FMOD System

```
result = sound->release();
ERRCHECK(result, "sound->release");
result = system->release();
ERRCHECK(result, "system->release");
```

3.3.3 Houdini Pseudocode

In Houdini the beat induction information is written to a channel, so this must occur directly after the beats are detected.

7) Create the tracks for the clip

```
destroyClip();
samplerate = RATE(context.myTime);
if(isZero(samplerate))
{
    addError(CHOP_ERROR_ZERO_SAMPLE_RATE);
    return error();
}
myClip->setSampleRate(samplerate);
```

8) Based on the user parameters, draw the beat as a channel to the CHOP viewport.

```
for(int frame = 0; frame < (endFrame-startFrame+1); ++frame)
{
    for(int h = 0; h < beatinfo.size(); ++h)
    {

        timeF = msTOframes(beatinfo[h].timeMS);
        if(fabs(timeF - frame) < 1)
        {
            chopViewportData[frame] = beatinfo.at(h).energy;
            break;
        }
        chopViewportData[frame] = 0;
    }
}
```

3.3.4 Maya

In order to create beat tracking in Maya I decided to create a node plugin as opposed to a command plugin. This would allow a user more flexibility as well as the ability to link the resultant data energy of a given beat from the node to any animatable attribute. The time in milliseconds the beat occurred as well as the energy of the beat were paired together in a struct and then stored in a C++ vector. The node only has to cook when one of the user defined parameters is changes, not when the time changes. When the animation is played after cooking, the time in frames is passed to the node, converted to milliseconds and compared with the times at which the beats occurred. If the difference between the nodes is less than 20 (because each frame is 41.667 ms when played at 24 frames per second) the frame is sufficiently close to the actual time the beat occurs, and so a beat in the music has occurred at that frame. Even if two beats are both sufficiently close to a certain frame, there is no problem since the human body can detect at most 300 beats per minute before

3.4 Results

The implementation of Patin's algorithm was successful for both programs (see Figure 3.4 and 3.6). Tests were run manually using 10 different musical pieces; 2 had single, percussive instrumentations; 4 were rock songs; and 4 were rap songs. The lengths of the test songs were 7 and 25 seconds for the percussive pieces and 30 seconds for all the other pieces. The amplitude of the occurring beat was connected to the radius of a sphere so that there could be a direct visual validation of the beat's aural occurrence. The animations were then played (a better word might be simulated or scrubbed) within the software. A test was deemed successful if the change of the radius occurred when the beat occurred. A better test would have been to render out the animated sphere and combine with the song sample. This is because the search time for the beat in the vector is linearly dependent on the number of beats; lag occurs when scrubbing the greater the number of beats detected. This lag time, however, is very minuscule for most songs. Only simulations in which a very low beat intensity, such as 1.05, and common freq, anywhere between 200-800, especially around 440, would cause a severe lag.

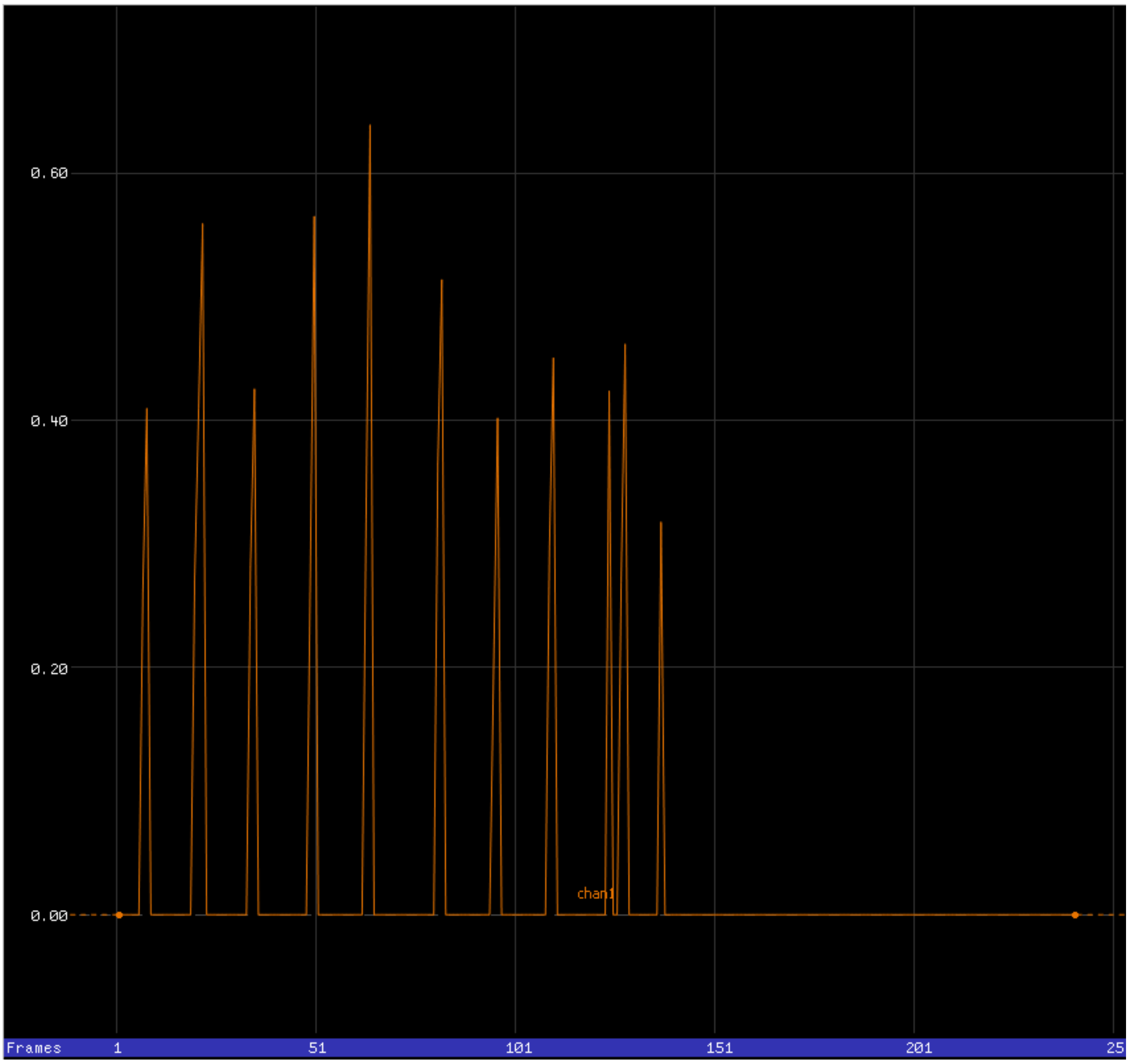

Figure 3.4 - The signal produced via beat induction using the 7 second song.

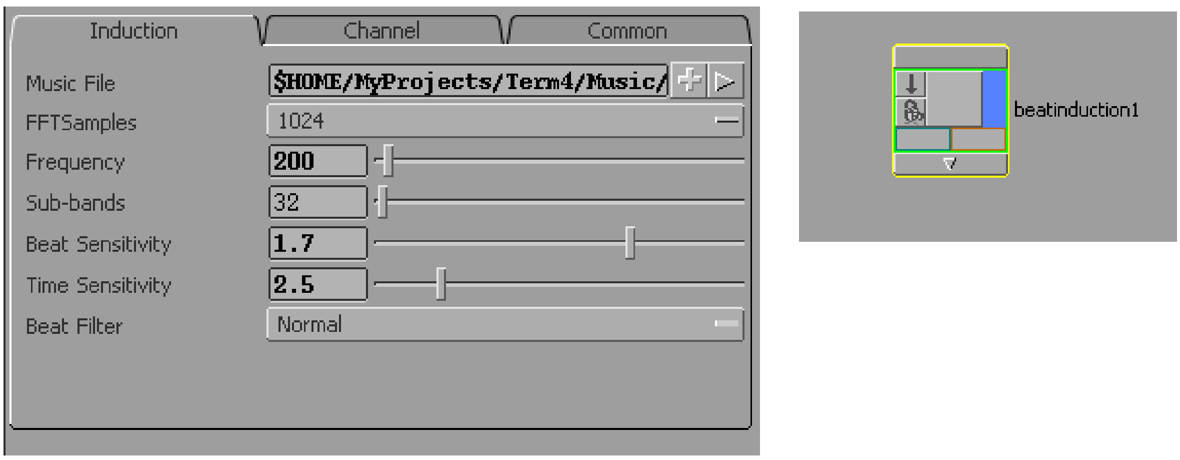

Figure 3.5 – Left: The parameters for node that produce the signal in Figures 3.4 and Fig. 3.6. Right: The network for the signal shown in Figure 3.4.

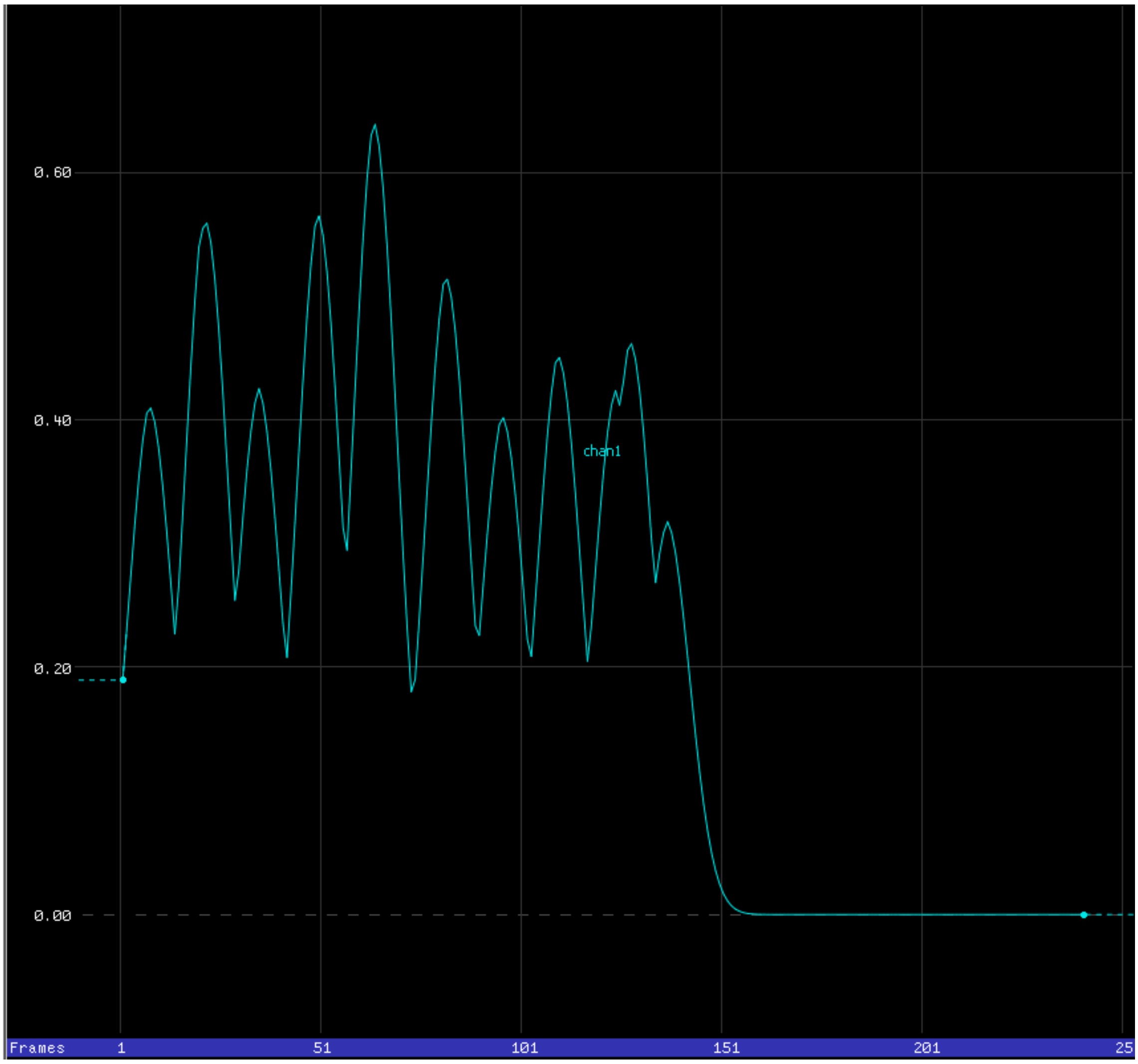

Figure 3.6 - The signal from Figure 3.4 with an envelop filter applied.

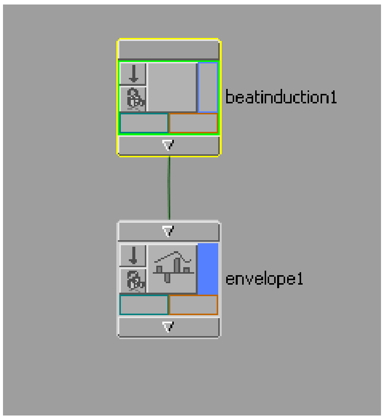

Figure 3.7 – The network for the signal show in Figure 3.6.

3.5 Problems

Even though the beat induction plugins were successful, certain problems were still encountered. Cooking the nodes takes a long time because of step 4) in the pseudo-code. When Houdini loads, each chop has to cook, so if there are multiple nodes, it could several minutes for the scene to load. Also, for some reason I have failed to discover, whenever the blue flag is displayed it will cook the node, but the channel information will not appear. It is only after the flag has been set and a parameter changes that the beat information will appear in the viewport. So one should change all the necessary parameters except for one, turn on the blue flag, let the CHOP cook, then change the last parameter, and the CHOP cook again.

In addition, there may be what are known as false positives and negatives. A false positive occurs when the a beat is detected when one does not exist. A solution to this is to increase the beat sensitivity. A false negative occurs when a beat exists and is not detected. Lowering the the beat sensitivity may solve this.

3.6 Comparison

Overall producing the beat induction node for Houdini proved far easier and more intuitive than for Maya. This is because Houdini creates a basic template by which all chops (and subsequently all Houdini nodes) must be created. All one needs to do is declare a new node operator, define the parameters, and then write the cook method, as opposed to Maya, which involves writing the uninitializePlugin() and initializePlugin(), compute(), and initialize() functions. The process of creating the parameters in Houdini consists of the PRM_Template types, which since it's a template, lends itself to greater facileness in creation. In Maya, one has to create a standard function and explicitly declare how the parameter is to be used, which can be cumbersome when one doesn't know about all the MAttribute types.

The use of the node is also more straightforward in Houdini. Because Houdini emphasizes the use of nodes and the network editor, more than Maya does with the hypergraph, to connect and use the nodes output is simpler. The use of the expression editor allows for easy information retrieval from a channel to any animatable node within Houdini. With Maya, however, one has to use mel script. While this may seem like a trivial matter – after all if one knows how to write smalls scripts for the expression editor then writing mel scripts should not acquire addition effort – it is the fact that Houdini allows for the specific attribute to have a script, allows for easier information access. In order for the node to be used efficaciously in Maya, it was best for a gui interface to be created, and it was in this interface that all of a scene's objects and animatable attributes could be accessed. The user interface is generated by Houdini automatically.

Houdini already has audio and signal processing nodes. This means that once the beats have been transferred into a signal, we can use those nodes to mold the information how we see fit. Hence, it helps to create a more plenary audio and channel framework, permitting the user flexibility in using the beat information. Figures 3.4 and 3.6 demonstrate this fact. Figure 3.4 shows an original beat signal while Figure 3.6 shows the signal with an envelope. Because Maya does not have an existing audio or signal processing framework built into the software, the user cannot further process the beat signal to fit his needs. This is a severe limitation when compared to Houdini.

When Maya creates a new node, only one instance of the node is created. It will then use a table to store the attributes of each "instance" of a node. For example, if we create 5 spheres, only one sphere node will get created, and Maya will create a table that has the radius, translation, rotation, etc. information for each of the 5 spheres. It will then pass this table's values to the node as attributes whenever the node needs to become cooked. Any variable stored in the class that is not on the attribute of the plug is only stored once, that value is shared between the various "instances" of the node. As was mentioned before, the beat information is stored internally as a c++ vector. Maya will only store one copy of beat information and use it for all of the instances of the beat induction node because this variable is not an attribute. That means that Maya can only store the beat information from one song at a time. Houdini circumvents this problem automatically because each CHOP node stores it's clip information locally. Since we apply the beat information to this clip, each CHOP node contains it's own beat information. We can thus have however many nodes with distinct information we want.

One advantage that because we have to define the output attributes of the node as well as the input attributes, we can pass any sort of additional information we want out of the node. For example, if we wanted to store the spectrum data in an a different array while calculating the beat, then we could create another node output attribute and connect the spectrum array to this output. With CHOPs, only the floating point numbers of the channel data can be output.

3.7 Conclusions

Patin's algorithm to be a proved a good choice for the implementation of a beat induction algorithm. Using it, I was able to successfully implement a working beat induction plugin for both Maya and Houdini. However, Houdini provides certain advantages over Maya, most specifically it's audio and signal processing nodes, which warrant use of such plugin in Houdini as opposed to Maya. FMOD Ex also provided a good choice in terms of an audio processing api as it is well documented and easy to use.

Chapter 4: Emotion Detection

4.1 Initial Approach

The first type of emotion detection that my research brought me into contact with was the emotional reaction created when consonance and dissonance are used to create tension within a piece of music and pleasing/unsatisfying harmonies. Doing research into consonance, I discovered difference between musical and sensory consonance and the use of harmony in musical consonance. At first I looked for models in which musical consonance was analyzed from musical signals, but I was not able to find any such model. My next attempt was to use find a model for sensory consonance and then a model for harmony and somehow combine the results of the two together to create musical consonance. I discovered Humdrum, a software, music analysis framework. Using Humdrum to process sensory consonance, I was able to create a spectrum file for the toolkit's dissonance command. I created the spectrum file while doing the calculations for the beat, then used a system command from Maya (I was not using Houdini at this point) to call Humdrum to convert the file into a dissonance file and then to read it back into Maya. However the minimum number of samples that FMOD Ex could provide from the spectrum was 64, while the greatest number of samples humdrum could process was three. I contacted Firelight Technologies asking for a copy of the source code so that I could change the minimum number of samples, but they said that I needed to buy a licence in order to get the source code. I also was not able to find any models on harmony, so I abandoned that approach.

I also contacted Musicovery.com (see Appendix A for the email), an online radio service that used the concepts of valence and arousal in order to create a tree structure for listeners to use to peruse the music library. They were not helpful and their response left me with the idea in my head that emotion detection would be well beyond the scope of this project. However, after doing more research I was able to find a several papers about emotion detection; but by this time I had already invested and seen results in implementing beat induction within Houdini. I decided that the best approach would be to break this project into two phases: an implemented first phase, id est implement beat induction in two 3D software packages and to compare and contrast these implementations; and then a proposed 2nd phase, an emotion detection and it's use in a lighting system.

Detecting emotion algorithmically from a music signal is a colossal undertaking. It pulls together several fields including music theory, physiology, and computer science. Even though studies into the relation between music and emotion have been going on for some time, there have been few studies about deriving emotion from music simply by analyzing the music itself. Before delving into my attempt to use musical analysis to effect lighting, it is important to see what other work on the subject has been done.

4.2 Literature Review

4.2.1 Detecting Emotion from Music Signals

Tao Li and Mitsunori Ogihara hypothesized that emotion detection could be developed by analyzing music signals.[38] The process of emotion detection they developed was divided into two steps: feature extraction and multi-label classification.[39] In order to obtain usable data, they extracted 30 second samples from 499 songs across four genres: ambient, classical, fusions, and jazz. The sound files were then labeled by a 39 year old male into 13 adjective/emotional groups and these were then grouped into 6 emotional supergroups.

In the first step, they extracted information from the music signal using MARSYAS.[40] They insist that the information should accurately represent the music and require little storage and extraction time. These features were classified as timbral or having rhythmic or pitch content. The second step builds an algorithm/mathematical model for properly labeling music. This was done by breaking down the multi-label classification problem into multiple binary classification problems, using LIBSVM to train Support Vector Machines (SVM) using the features extracted from the sounds in step one.

They then ran experiments using their system and the collected data. The "multi-label classification method was tested for classification in the thirteen adjective group and classification into the six supergroups." Overall, their algorithm resulted in low performance. They attributed this to numerous cases in which emotional labeling was difficult, resulting in borderline cases. They also note that labels were not used with the same frequency across the music types. They conclude their paper stressing the difficulty of implementing an emotion detection system and recommend expanding data sets, multiple round label collection, different adjective sets, different types of features, and the use of style and genre information to help resolve performance issues.

38 Li, Tao and Mitusoni Ogihara. "Detecting Emotion in Music."

39 en.wikipedia.org/wiki/Multi-label_classification, Multi-label classification is a concept in mathematics and machine learning. Traditional single-label classification is concerned with learning from a set of examples that are associated with a single label l from a set of disjoint labels L, $|L| > 1$. In multi-label classification, the examples are associated with a set of labels . However, they allow for multiple labels to be applied to one piece of music.

40 According to http://marsyas.sness.net/ , MARSYAS is "is an open source software framework for audio processing with specific emphasis on Music Information Retrieval applications."

4.2.2 System Identification

In his thesis to the University of Waterloo, Ontario; Mark David Korhonen presents a method for emotion detection using system Identification. This system should be able to generalize emotional appraisals for any song within a genre of music. System identification is used because it is a signal processing technique used to create "mathematical models of a system given examples of its inputs and output signals."[41]

Korhonen lists 5 reasons why system identification should be used to model time-varying emotional appraisals:

1. It is possible to model a listener's perception of emotion as a system by treating the m musical features as an input signal and the emotional appraisals as an output signal...
2. Performing time series analysis to examine the relationship between the inputs and the outputs is limited...
3. Splitting the emotional appraisal into a deterministic function of the inputs and a stochastic component is intuitive...
4. The goal of a system identification is to predict the output of a system given the inputs to learn how they affect the outputs...
5. System identification literature addresses particular challenges that occur while creating models.[42]

In order for music to be capable of being used as an input signal to a model, the music must be accurately represented by it's musical features (similar to Li and Obihara). Korhonen creates an m-dimensional vector where m is the number of musical features used to represent the selection. He uses 18 features total, which he extracts using a combination of PsySound, MARSYAS, and manual efforts. Musical features selected include such features as timbre, harmony, texture, note onset, pitch range, and rhythm.

In order to create the emotional appraisals to which the inputs will match, he uses EmotionSpace Lab, "which quantifies emotion using the dimensions valence and arousal." He used 35 volunteers to appraise 6 musical pieces. After the volunteer has has emotionally labeled the music, Korhonen estimates the emotion of a song by comparing it to the emotion most labeled to the song by the volunteer population. To do this he preprocesses the appraisals by first dealing with the outliers and missing data and then applies filters to the input and output signals to reduce noise.

He next selects the appropriate model to use. He uses simulation models, which predict "the output based entirely on the input signal and delayed versions of the input." After selecting the model's structure the parameters need to be estimated so the model agrees with the input and output data. Next, he validates the models. This involves assessing how they relate to observed data, prior knowledge and their usage. He validates them by comparing the simulated emotion appraisals with the true ones. This can be done by using the model's bias and variance.

41 Korhonen, Mark David. *Modeling Continuous Emotional Appraisals of Music Using System Identification.* pg 20.
42 Korhonen, Mark David. *Modeling Continuous Emotional Appraisals of Music Using System Identification.* pg 22-23

Overall, he concludes that the first three criteria that a model should meet are met. His results also find that the fourth criteria is met for arousal appraisals but not for valence appraisals. Ergo, system identification is a valid means by which to create an emotion detection model.

4.3 Emotion Detection for Lighting

A simple method for finding the emotion of a musical piece would be to use consonance and dissonance. There are several ways of calculating consonance: frequency ratios, with ratios of lower simple numbers being more consonant than those which are higher; coincidence of harmonics, consonant sounds have a largely degree of overtones (and so the quality of the sounds is directly linked with the timbre of the sound); Period length or neural-firing coincidence; or Fusion or pattern matching: fundamentals may be perceived through pattern-matching of the separately analyzed partials to a best-fit exact-harmonic template.[43]

$$\frac{P}{P_0} = e^{\frac{-0.7R}{R_0}} e^{\frac{-1.08S}{S_0}} \left(1.24 - e^{\frac{-1.08T}{T_0}}\right) e^{-\left(\frac{0.23N}{N_0}\right)^2}$$

Figure 4.1 - P - sensory pleasantness; S – sharpness; R – roughness; T – tonality; N - loudness[44]

Using system identification seems to be a better approach for detecting emotion from audio signals; Korhonen's method, though complex and involved, appears to be a viable way to do so. This is for several reasons. Firstly, the use of MARSYAS and to extract the necessary features from a musical signal is advantageous. Because MARSYAS is a software framework, it can be implemented into Houdini using the HDK. So all musical analysis that Korhonen performs can thus be done in Houdini. PsySound poses a larger problem since it is a stand alone program for the Mac. If PsySound existed for Linux, then it would be capable to use a system command to call and run PsySound from Houdini to produce the information. However a better choice would be to find another software framework that can extract these same features. This is because we can use these frameworks in the HDK to create nodes that will perform the feature extraction. Between C/C++ audio frameworks such as Aubio, Vamp, and The Melisma Music Analyzer[45] the musical features Korhonen extracted using PsySound should be extractable, however it will take some research to figure out which framework(s) is best. The second problem would come from developing the SVMs to train the models. Again a system such LIBSVM would be advantageous here.

Once the emotion detection is implemented it would be a matter of decided how to use the emotional labels to adjust the lighting. One of the most immediate ways to do so would be

43 http://en.wikipedia.org/wiki/Consonance_and_dissonance

44 Karjalaninen, M. “Chapter 7: Other psychoacoustic concepts.”

45 Information on these framework can be found at http://aubio.piem.org/ , http://www.vamp-plugins.org/ , and http://www.link.cs.cmu.edu/music-analysis/ respectively.

to control the color of the lights. There are no shortage of theories pertaining to this. "Artists such as Kandinsky, Klee and Vantongerloo; composers such as Scriabin, Berlioz and Debussy; and scientists such as Newton, Kepler and Helmholtz have all explored links between colour and music."[46] How, it would be difficult to use this. This is because the music changes so much. Most theories either link notes with colors based on wavelengths or intervals. It would not be viable to change the color of the lights every time the melody changed, because the notes change so readily that if the lights changed as much the people wouldn't like it. They would get sick. It is more likely that the light colors would change on certain beats and cycle through 2-3 colors. So it might be more advantageous to find out the emotion of a song, and choose a color palette fitting that emotion. Then one could use the beat information to change the light color according to the colors produced from the emotion detection.

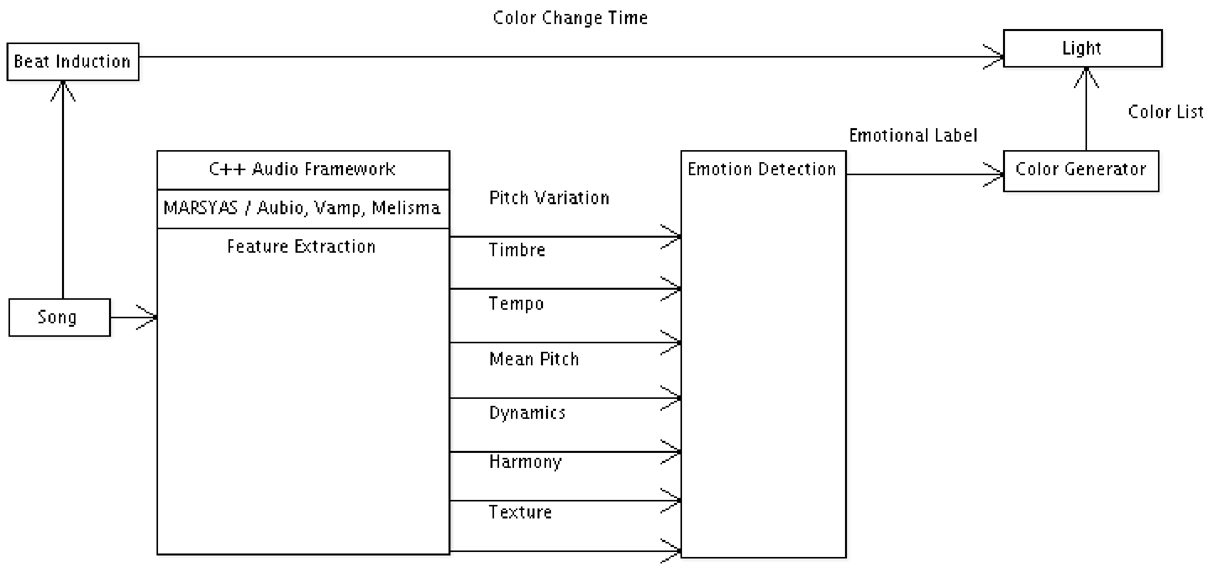

Fig. 4.2 – Emotion Detection Model for controlling the color of a light

46 http://www.katlubar.com/html/theories.html

Chapter 5: Conclusions

As was demonstrated, the relationship between emotion, music and lighting is very strong. Music is a very strong stimuli and the brain seems to be especially equipped and designed with this in mind. Emotional elicitations are further honed by the use of lighting and colors that are reflective of the emotional content of the music.

In addition, I succeeded in accomplishing both proposed phases: the implementation and comparison of beat tracking within a 3D software package, as well as coming up with a proposal for implementing emotion detection and it's use with lighting.

After having established the definition of beat tracking as finding the locations of beats, I was able to implement it in both in Maya and Houdini using Frederic Patin's energy sampling algorithm. Overall, the results in Houdini were far more promising than the results in Maya, and any future work done on beat tracking or beat induction should be done in Houdini. This is because of Houdini's audio and signal processing capabilities as well as the ease of implementing a custom node using the HDK.

The complexity of the emotion detection problem can be seen in the size and complexity of Mark David Korhonen's thesis for the University of Waterloo, Ontario Canada. The thesis is very long, complex, and gets into many issues about signal and musical analysis, yet it presents a window into which an emotion detection system can be created. It does this by establishing features to extract, describing them in ways in terms of C++, audio frameworks. This makes would make it easier to implement within a software package like Houdini.

Bibliography

1) emotion. *CancerWEB.* Posted on March 5, 2000. Accessed on August 18, 2007. <http://cancerweb.ncl.ac.uk/cgi-bin/omd?action=Search+OMD&query=emotion>

2) “Brookhaven Lab Expert Helps Date Flute Thought to be Oldest Playable Musical Instrument.” *Brookhave National Laboratory.* Posted on September 22, 1999. <http://www.bnl.gov/bnlweb/pubaf/pr/1999/bnlpr092299.html>

3) “Consonance and dissonance.” *Wikipedia.* Accessed on July 20, 2007. <http://en.wikipedia.org/wiki/Consonance_and_dissonance>

4) "Musical Rhythm," Microsoft® Encarta® Online Encyclopedia 2007 http://encarta.msn.com © 1997-2007 Microsoft Corporation. All Rights Reserved.

5) Barker, Paul. *Pink Floyd under Live 8-konserten i Hyde Park i 2005. Vil det skje igjen?.* Hyde Park, London. 2005. Accessed on September 6, 2005. <http://www.nrk.no/musikk/1.1881311>

6) Boyle, Alan. “Listening to the sounds of science.” *Today*. Posted on February 21, 2000. Accessed on August 18, 2007. <http://today.msnbc.msn.com/id/3077403/>

7) Davis, James W. “A Response to W. Garner's Observations on the Relationship Between Colour and Music.” *Leonardo.* Vol 12. 1979. pg. 218-219.

8) Davies, Matthew E.P. “Context-Dependent Beat tracking of Musical Audio.” *IEE Transaction on Audio, Speech, and language processing, Vol. 15, No.3 March 2007.* Accessed on September 10, 2007. <http://64.233.183.104/search?q=cache:1Wa-Dj5U-VIJ:www.elec.qmul.ac.uk/people/markp/2007/DaviesPlumbley07-tnn.pdf+difference+beat+tracking+beat+induction&hl=en&ct=clnk&cd=9&client=safari>

9) Desain, P. and Honing, H. “Computational Models of Beat Induction: The Rule-Based Approach. Journal of New Music Research.” 28(1). 1999. pg. 29-42. Accessed on September 3, 2007. <http://www.nici.kun.nl/mmm>

10) Dixon, Simon. “Beat Induction and Rhythm Recognition.” *Proceedings of the Australian Join Conference on Artificial Intelligence.* Accessed on September 3, 2007. <http://citeseer.ist.psu.edu/cache/papers/cs/2911/http:zSzzSzwww.cs.flinders.edu.auzSzPeoplezSzSimon_DixonzSzpubzSzai97.pdf/dixon97beat.pdf>

11) Dixon, Simon and Emilios Cambouropoulos. “Beat tracking with Musical Knowledge.” *ECAI 2000: Proceedings of the 14th European Conference on Artificial Intelligence.* 2000. pg. 626-630. Amsterdam, IOS Press. Accessed on September 10, 2007.

<http://citeseer.ist.psu.edu/cache/papers/cs/14839/http:zSzzSzwww2.ai.univie.ac.atzSz~simonzSzpubzSzecai2000.pdf/dixon00beat.pdf>

12) Ellis, Daniel P.W. “Beat Tracking with Dynamic Programming.” Accessed on September 10, 2007. <http://www.ee.columbia.edu/~dpwe/pubs/Ellis06-beattrack.pdf>

13) Glynn, Ian. *An Anatomy of Thought: The Origin and Machinery of the Mind.* Oxford university Press. 1999. pg. 315 – 350.

14) Goto, Masataka and Yoichi Muraoska. “ A Beat Tracking System for Acoustic Signals of Music.” *Proceedings of the second ACM international conference on Multimedia*. 1994. pg.365-372. Accessed on September 10, 2007.
<http://delivery.acm.org/10.1145/200000/192700/p365-goto.pdf?key1=192700&key2=0172939811&coll=GUIDE&dl=ACM&CFID=29201323&CFTOKEN=22131379>

15) Gouyon, Fabien and Gerhard Widmer, Xavier Serra, Arthur Flexar. “Acoustic Cues to Beat Induction: A Machine Learning Perspective.” August 25, 2006. Accessed on July 15, 2007. pg. 1-3. <http://www.iua.upf.edu/mtg/publications/bc6559-MP-2006.24.2.177-Gouyon.pdf>

16) Gouyan, Fabien and P. Herrera. “A Beat Induction Method For Musical Audio Signals.” Accessed on July 15, 2007. <http://www.iua.upf.es/mtg/publications/WIAMIS03-fgouyon-beat.pdf>

17) Gouyon, Fabien and Simon Dixon, Gerhard Widmer. “Evaluating low-level features for beat classification and tracking.” Accessed on September 10, 2007.
<http://www.cp.jku.at/research/papers/Gouyon_etal_ICASSP_2007.pdf>

18) Heimiller, Joseph. “Where Math meets Music.” Posted in 2002. Accessed on July 15, 2007.
<http://www.musicmasterworks.com/WhereMathMeetsMusic.html>

19) Jourdain, Robert. *Music, the Brain, and Ecstasy.* William Morrow and Company, Inc. 1997. pg. 1 – 333.

20) Karjalaninen, M. “Chapter 7: Other psychoacoustic concepts.” Online Slides. Slide 12. Accessed on September 6, 2007. <http://www.acoustics.hut.fi/teaching/S-89.3320/KA7.pdf>

21) Korhonen, Mark David. 2004. *Modeling Continuous Emotional Appraisals of Music Using System Identification.* Thesis (MASc). University of Waterloo.

22) Lubar, Katherine. “Colour Intervals: Applying Concepts of Musical Consonance and Dissonance to Colour.” Accessed on August 7, 2007. <http://www.katlubar.com/html/theories.html>

23) Li, Tao and Mitusonit Ogihara. “Detecting Emotion in Music.” 2003. Accessed on August 19, 2007. <http://ismir2003.ismir.net/papers/Li.PDF>

24) Meudic, Benoit. “A Causal algorithm for beat-tracking.” Accessed on September 10, 2007. <http://mediatheque.ircam.fr/articles/textes/Meudic02b>

25) Mitchell, Lon H. “Advanced Mathematics in Music Theory.” Section 2 Music Theories. Accessed on July 19, 2007. <http://www.math.ku.edu/~mitchell/mm/mathmusic.html>

26) Patin, Federic. “Beat Detection Algorithms.” *GameDev.net.* Accessed on July, 14 2007. <http://www.gamedev.net/reference/programming/features/beatdetection/>

27) Ratey, John. *A User's Guide to the Brain.* Little, Brown, and Company. 2001. pg. 222 – 251.

28) Schlick, John. “How does a Lighting Designer actually run a show?” Accessed on September 6, 2007. <http://www.exotic-lighting.com/HowLDRunShow.htmll>

29) Schlick, John. “Why Hire a Lighting Designer?” Accessed on September 6, 2007. <http://www.exotic-lighting.com/WhyHireLD.html>

30) Skovenborg, Esben and Søren H. Nielsen. “Measureing Sensory Consonance by Auditory Modelling.” Proc. of the 5th Int. Conference on Digital Audio Effects (DAFX-02). Hamburg, Geramny. September 26-28, 2002. Accessed on July <http://citeseer.ist.psu.edu/cache/papers/cs/27023/http:zSzzSzwww.unibw-hamburg.dezSzEWEBzSzANTzSzdafx2002zSzpaperszSzDAFX02_Skovenborg_Nielsen_consonance.pdf/skovenborg02measuring.pdf>

31) de Sousa, Ronald. “Emotion.” *Stanford Encyclopedia of Philosophy.* Posted on February 3, 2003. Substantive revision June 5, 2007. Accessed on August 17, 2007. <http://plato.stanford.edu/entries/emotion/>

32) Terhardt, Ernst. "Harmony." Posted on March 9, 2000. Accessed on September 4, 2007. <http://www.mmk.ei.tum.de/persons/ter/top/harmony.html>

33) Terhardt, Ernst. "Musical consonance." Posted on February 17, 2000. Accessed on July 20, 2007. <http://www.mmk.ei.tum.de/persons/ter/top/muscons.html>

34) Terhardt, Ernst. "Sensory consonance." Posted on February 15, 2000. Accessed on July 20, 2007. <http://www.mmk.ei.tum.de/persons/ter/top/senscons.html>

35) Trainor, L.J. and L.A. Schmidt. "Processing Emotions Induced By Music." *The Cognitive Neuroscience of Music.* Oxford University Press. 2003. pg. 312 - 324.

36) Uhle, Christian. "Tempo Induction by Investigating the Metrical Structure of Music Using a Periodicity Signal that Relates to the Tatum Period." Accessed on August 18, 2007. <http://www.music-ir.org/evaluation/mirex-results/articles/tempo/uhle.pdf>

37) Wright, Ian and Aaron Sloman, Luc Beaudoin. "Towards a Design-Based Analysis of Emotional Episodes." Accessed on August 17, 2007. <http://citeseer.ist.psu.edu/cache/papers/cs/2786/http:zSzzSzwww.cs.bham.ac.ukzSz~axszSzcog_affectzSzWright_Sloman_Beaudoin_grief.pdf/towards-a-design-based.pdf>

Appendix A: Personal Communication

From : Contact <contact@musicovery.com>

Sent : Thursday, August 16, 2007 3:25 AMTo : "Jedrik Eliasen" <jedrik>

Subject : Re: Algorithmic Question

Dear Eliasen
Matching music and emotion is very complex.
You take the view the vertical axe (energy), the easiest part, can be measured by beat detection, well it is only a small part of the story, there are at least 5 key parameters broadly not correlated and bit rate is not a measure corresponding to tempo perceived by human beings. The horizontal axe (valence, dark to positive) is much more complex to analyse.

At Musicovery we listen to every single peaces and attributes 40 parameters (each parameter can take 10 values each). We created special algorithms to make the projection a on 2D graphics.

We doubt there is any possibility in a near future to automatize the characterisation process, unless you accept to do a very approximate job.

Best regards

Vincent castaignet, Co-founder
Musicovery

----- Original Message ----- From: "Jedrik Eliasen" <jedrik>
To: <contact@musicovery.com>
Sent: Wednesday, August 15, 2007 3:34 PM
Subject: Algorithmic Question

To Whom It May Concern:

My name is Jedrik Eliasen and I am currently completing my MSc at Bournemouth University in Computer Animation. I am working on a project that will analyze a piece of music (an mp3, wav, etc) and figure out how fast and "emotional" (ie happy or sad) it is. Right now I am trying to figure out a method of measuring a piece of music's emotion. I was wondering how you were able to create the emotion graph on your page. I have already implemented beat detection, so I can find out if a song's spend, ie calm or energetic, but I have not been successful in finding anything out about calculating consonance/emotion from a song (or signal). The only research I have been able to find has been has been on sensory consonance/dissonance. Right now I am using the fmod library for my musical processing.

I was wondering if you might be able to shed some light on this issue. Any help you could give would be greatly appreciated, especially about any specific algorithms, articles, or pieces of software that you used.

Thanks,
Jedrik Eliasen
MSc. Compute Animation
Bournemouth University