COMPUTATIONAL RHYTHM AND BEAT ANALYSIS

Nicholas Berkner

University of Rochester

ABSTRACT

One of the most important applications in the field of music information processing is beat finding. Humans have the ability to almost immediately determine the pulse of a piece of music as well as larger hierarchical structures of rhythm such as meter. Previous work in this topic has focused on either developing computational models which operate on a symbolic input and simulate the cognitive processes of the human brain, or using a variety Digital Signal Processing techniques to directly extract rhythmic information from an audio file. Because these methods operate using different kinds of inputs, it is often difficult to compare the two. Thus, this report focuses on the necessary signal processing required to convert an audio input into a quantized note onset file, which is the most basic form of symbolic input used in computational models.

1. INTRODUCTION

The final goal of this paper is to develop an automated process in which a real acoustical input, such as an audio wave file, can be converted into a quantized note onset vector, commonly used in computational meter finding. Inspiration for the audio processing portion comes from past work by Eric D. Scheirer. In "Tempo and Beat Analysis of Acoustic Musical Signals," Scheirer describes a method for finding note onsets from an audio signal by using the derivative of its envelope. Figure 1) shows a diagram of his system. A similar method is used in this study, but where Scheirer uses a mostly theoretical approach to determining the rhythmic qualities, this paper focuses on a practical approach, which will give the correct result while using the least amount of processing possible. While note onset vectors are typically derived by hand from simple melodies for the sake of testing the effectiveness of computational meter finding models, the note onset vectors in this experiment will be derived from an actual audio file.

2. METHOD

The overall process can be summarized by the flow chart below. This represents the optimized final solution, which is the result of testing many variations on the following methods which will be described below. The first stage involves converting the audio input into a note onset vector (NOV), and for the most part resembles Scheirer's model. The second stage is used to determine the tempo and meter by analyzing the frequency spectrum of the onset vector. The third and final stage uses the tempo and meter information to quantize the note onset vector so that the durations are normalized to common metrical note lengths.

5.1. Finding Note Onsets

Figure 2) is a flow chart of the process used to calculate the NOV. Because the subsequent stages use the NOV as an input, accuracy is very important. Though it is difficult to achieve a perfect NOV, especially with a polyphonic input, care must be taken to avoid misfires, where an onset is identified where it should not be, since all onsets hold the same weight in later stages. Thus, it is better to miss on onset rather than identify an onset when there is none, since the missed note will likely be on a weak beat anyway.

5.1.1. Envelope Detection

Rather than finding the envelope for the original input signal, it is better for accuracy to calculate the envelopes for individual frequency bands independently, as seen in Figure 3a) - Figure 3d). There are a variety of methods to finding the envelope, but since the end results are mostly identical, the most important factor becomes processing speed. The method used involves calculating the Spectrogram of the signal, seen in Figure 4), and summing the energies for each frequency band over time. This is more efficient than using multiple band pass filters and results in a sharper cutoff frequency for each band. Because we are now only interested in frequencies in the "natural range," i.e. from 0 to 20 Hz, a LPF with a cut off at 10 Hz is used to smooth the envelope.

5.1.1. Note Onsets

The next task is to differentiate the envelope, which can be equated to finding the attack rates of the onsets. To simplify the later stages, all note onsets are given the same weight. Therefore a threshold must be set for the derivative signal over which the sample will qualify as an onset. This proves to be a very important value, since setting the threshold to low will result in many misfires in the NOV, while setting

the threshold too high results in an empty NOV. The derivative is normalized so that the maximum value is 1, so that the same threshold can be used for multiple audio files with different volumes and instrumentations. The code is also modified so that a note onset is only triggered once when the threshold is crossed, so that there are not multiple onsets per peak. The threshold is also made dependent on the frequency band, since lower frequency sounds, such as those made by percussive instruments, have a much faster attack than melodic instruments. Below are the plots of the NOV's for the frequency bands. Note that 1st and 3rd band, corresponding to low and high frequencies are quite accurate, while the middle and coloration band tend to misfire often. This is found to be a characteristic in most audio files, which suggests that the accuracy of the NOV can be improved by combining only the onsets in these bands while omitting the others. The accuracy of the final note onset file can be qualitatively measured either visually in Figure 5) or aurally, by adding a beep signal corresponding to the NOV to the original input signal and listening.

5.1. Finding Tempo and Meter

Once the NOV finding algorithms have been tweaked to give the most accurate result for a variety of audio inputs, spectral analysis can be applied to the NOV to give some insight on the tempo and meter of the signal.

5.1.1. Fast Fourier Transform

Because the NOV can be viewed as a variation of an impulse train, its Fourier Transform will also be similar to an impulse train. As stated previously, the frequencies of interest in this application are those under 20 Hz. Since the NOV still has the same sampling rate as the input signal, its FFT will extend to the Nyquist Frequency, 22.05 kHz for a typical 44.1 kHz sampling rate. By downsampling the NOV by a factor of 44100/20=2205, the bandwidth can be limited to 10 Hz. This requires an anti-aliasing LPF, which can be achieved simply by convolving the NOV with a 2205 sample length pulse. The resulting downsampled NOV will sometimes have onsets with amplitudes or widths of 2 as a result of the combination of several very close onsets. Though this is mostly indicative of a strong beats, and therefore might be seen as useful, the amplitudes must be equalized back to 1 to achieve the best FFT result. Figure 6) shows the Spectrum and Cepstrum of some artificially created simple onset patterns. Figure 6a) corresponds to a duple meter rhythm and Figure 6b) a triple meter. Note that since the shortest metrical level for both is an eighth note, which has a frequency of 4 Hz, this is the strongest peak in both spectra and is indicative of the 'tactus' or beat. Note, however, that the separation of the other peaks is dependent on the meter. This is easily seen in the Cepstra, where the two highest peaks occur at frequencies with a ratio of 4:3 for the duple meter signal and 3:2 for the triple meter signal. The ratio of peaks in the Cepstrum can therefore be useful in determining meter for ideal inputs, but as seen in Figure 7), the Cepstrum quickly becomes too noisy for real audio input. The same functionality can be found in the frequency domain. If the highest peak corresponds to the pulse, the next peak above that frequency will correspond to the next shortest note duration. Similarly, the next peak below that frequency will correspond to the next highest note duration. In a simple duple meter, such as 4/4, the tactus can be on the eighth note, quarter note, half note etc. Assuming it is every quarter note, the next longest metrical note will either be a half note or a whole note. Though peaks can exist at dotted quarter or dotted half notes, the peak will be strongest at the true metrical level. The same reasoning can be applied to the next shortest metrical note. There is more ambiguity in triple meter. If the tactus is on the quarter note, with the strongest short note being an eighth note and the strongest long note being a dotted half note, the meter is simple triple. If the opposite is true, and the strongest note lengths have a ratio of 1:3:6, complex meter (6/8) is implied. In this fashion, the meter of a piece can be determined from the FFT of its NOV. If a ratio is greater than ½, next smallest frequency is actually the difference between adjacent peaks. (i.e. 3:4:8=>1:4:8) This can be further expanded by the following algorithm:

Find Tactus
For N metrical levels higher
    Find highest peak with frequency above that of last peak
    Calculate Ratio
For M metrical levels lower
    " " below that of last peak
    Calculate Ratio
Shortest metrical level = smallest frequency difference between peaks

Note that in the example duple meter spectrum in Figure 6a), there is no 2Hz peak. In "Pulse Detection in Synchopated Rhythms using Neural Oscillators," Ed Large calls this a "missing pulse," which is common in syncopated rhythms. In this project missing pulses can be ignored, since the next lowest note length will still indicate if the meter is duple or triple.

With the known information about meter, the tempo can be modified from the pulse frequency. Humans typically prefer tempos within a certain range, so if the pulse is outside of that range, the next closest tempo can be found by multiplying or dividing by the next higher or lower ratio value. For this project, it is assumed that the comfortable range of tempos lies between 50 and 150 BPM.

Finally, while the magnitude spectrum shows where the strongest beats are, the phase spectrum can be used to find the delay, since many audio files do not begin on the very first sample. To create a simple metronome, the phase of the tactus can simply be added to the oscillator in the metronome. The oscillator is a simple sinusoidal wave, which triggers a beep when it reaches a peak. Note that the phase does not necessarily account for pick up notes, so more advanced metronomes with meter may be off by a beat.

5.1.1. Oscillator Model

The technique described above is reminiscent, but not equivalent to Ed Large's Oscillator Model. A true oscillator model is evaluative, because it tests the strength of resonance of each oscillator and then chooses the strongest one. This can be implemented either by a sweep of Comb Filters, or by physically creating the oscillator waveforms and multiplying them with the NOV. The latter method was used to create Figure 8), which have very distinguished peaks at the pulse frequency. However, as with most evaluative algorithms, the need to process the data over the entire range of possible solutions makes the systems implementing them very inefficient. The algorithm described in the previous section is similar in spirit to the Oscillator model, but can be computed in a fraction of the time, giving it a good advantage in this application. This also represents a combination of cognitive and signal processing approaches to the problem of meter finding, and suggests that the best method might involve both tactics. The human brain seems to operate in a similar fashion. First, there is a subconscious "processing" which occurs and establishes a pattern of different note lengths. Then, the blanks spots are filled in based on that information.

5.1. Quantization

Once the tempo and meter have been found, finding the quantized NOV is relatively simple. First, the note lengths from the meter finding algorithm are converted to samples according to their respective frequencies. The NOV is converted into a inter onset interval vector (IOIV), by taking the difference, in samples, between onsets. These intervals are compared and rounded to the nearest possible note lengths. The quantized IOIV can be converted back to a quantized NOV, for which every data point corresponds to the shortest existing note length. To account for very short IOI's a lower limit is set on the note length. Also, since M=N=1, note length ratios that are unspecified are assumed to be 2. Quantized NOV's are useful in many computational meter finding algorithms, such as the Povel-Essens Model and Probabilistic Models. Because the meter has already been defined, these models can be optimized and simplified to only determine pick up note status. It should be noted, however, that these models were designed with NOV's derived by hand from melodies. Because of the polyphonic nature of music, and the tendency for percussive elements to have stronger onsets, the NOV's calculated in this project will differ in several ways from those of simple melodies. First, they are not perfect. There are often notes missing and sometimes notes are added. These notes were mostly insignificant in the previous stages, because signal processing is dependent on repeated patterns and random errors have little effect. These errors may prove more significant in computational meter finding models. The second difference is that, since the NOV is derived from a polyphonic input, and there is no extra strength applied to onsets on multiple voices, the rhythms are much less diverse. Many consist of a consistent pulse beat with an occasional pickup and then some rests, so the starting point of the train of pulses is very important in determining where the beat begins. This tendency towards few distinguishing features and occasional errors will likely strain the computational models mentioned above. Luckily, some information is already known about meter, so a comparison of the two results will help to negate some of these errors.

4. SAMPLE INPUTS

In developing and testing the method described above, several audio files were used. The following pieces were selected first to test the functionality of the code, then to see how resilient it was to different types of inputs. All files were Mono with sampling rates of 44.1 kHz and were shortened to 10 or 30 second clips. The Italian Concerto by Bach was the first file used because of its simple rhythm and instrumentation. Once the code had been developed to give a satisfactory result for that input, the piece "Make the Road by Walking" by the Menahan Street Band was used, because it still had a relatively simple meter, but greatly increased the variation of sounds, causing the spectral envelope detector to be developed. With the algorithm working successfully with these duple meter pieces, the song "Living a Lie" by Sinima Beats was added to the repertoire to test the algorithm with a triple meter input. A simple 6/8 drum pattern was used to develop the compound meter algorithm. Finally, for fun and out of morbid curiosity, Dave Brubeck's "Take Five" was tested just to see what would happen. The resulting meter ratios and tempos can be seen in the table below. Attached in the .zip file are the original audio files with the note onset beeps and the metronome at the corresponding tempo added.

5. CONCLUSIONS

In conclusion, it was found that an iterative method for designing musical based algorithm could be successful. Music, after all, is an art form and thus the techniques used to analyze it must be somewhat creative at times. It is

doubtful that a meter finding algorithm will ever be developed that works for every piece of music, and if it is, someone will promptly compose a piece to baffle it. This is not to take away from the advantages of a method based purely on signal processing. Much can be learned by studying the effects of different rhythms on the domains of a signal, and much more research needs to be done to fully understand how musical properties affect a physical signal as well as our cognition.

12. FIGURES

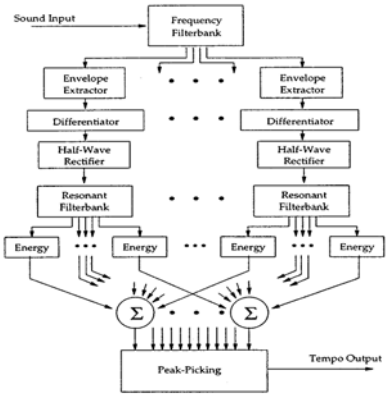


Figure 1) Scheirer's Model for meter finding

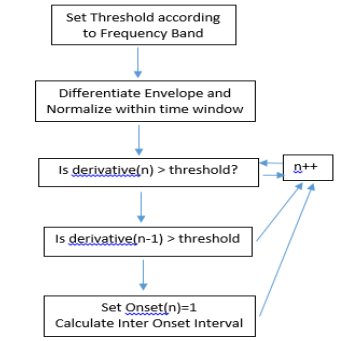
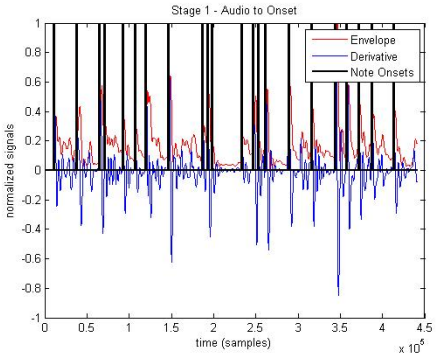
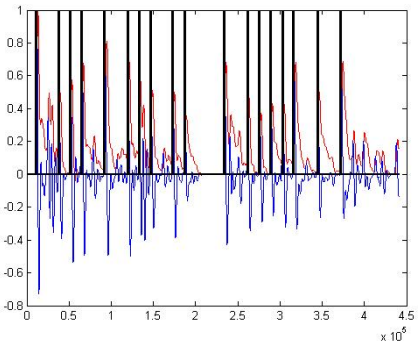


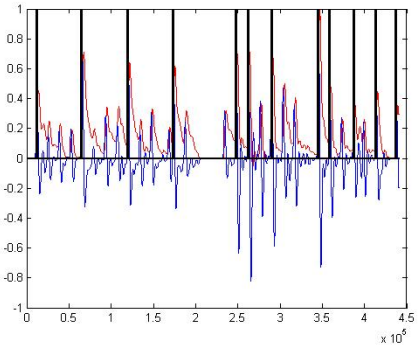
Figure 2) Flow Chart for NoteOnsets.m



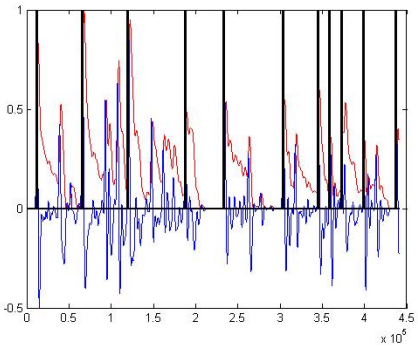
a) 8192 - 16384 Hz



b) 20148 - 8192 Hz



c) 512 2048 Hz



a) 20 512 Hz

Figure 3) Note Onset Vectors for different frequency bands.

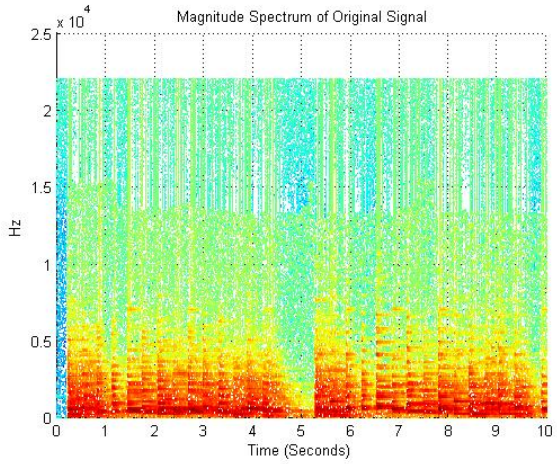


Figure 4) Spectrogram of Italian Concerto

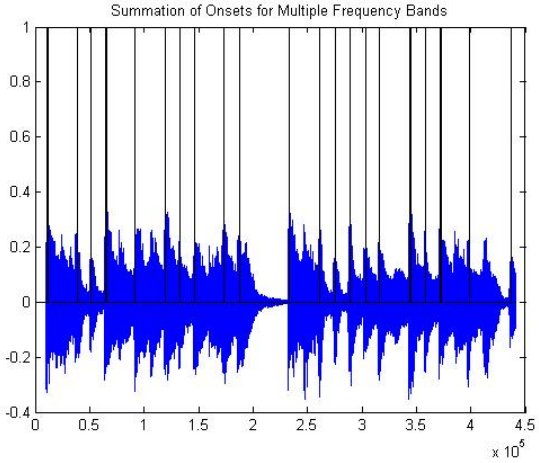
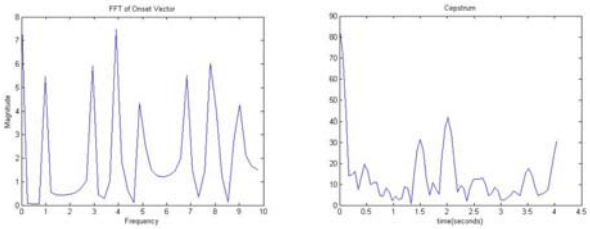
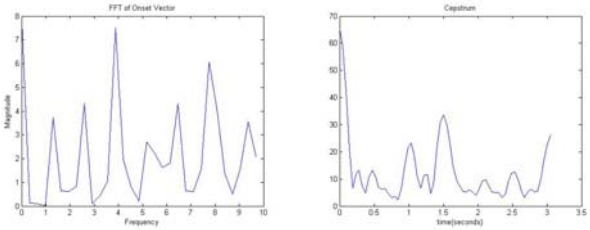


Figure 5) Note Onsets compared to original signal



a) Duple Meter: Tactus 4Hz



b) Triple Meter: Tactus 4Hz

Figure 6) FFT and Cepstrum of Duple/Triple Meter Rhythm

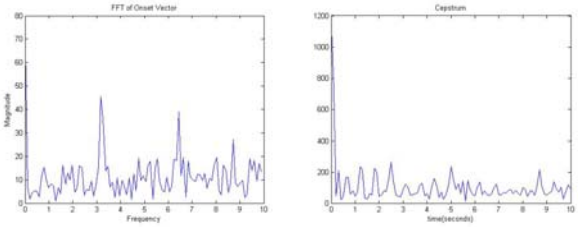
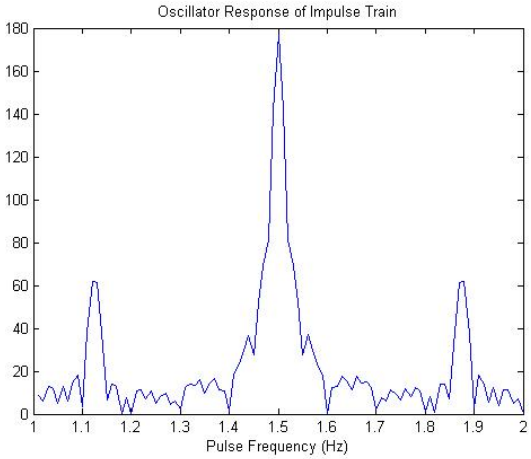
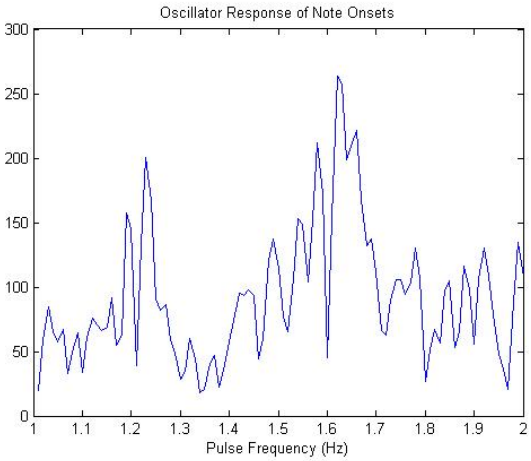


Figure 7) FFT and Cepstrum of Italian Concerto NOV



a) Pulse Train: 1.5 Hz



b) Italian Concerto: 1.62 Hz = 97.2 BPM

12. REFERENCES

[1] Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time varying events. Psychological Review, 106, 119–1 59.

[2] Scheirer, E. D. (1998). Tempo and beat analysis of acoustic musical signals. Journal of the Acoustical Society of America, 103, 588–6 01.