A NOVEL MUSIC BEAT DETECTION ALGORITHM BASED ON PERIODICITY OF ENERGY FLUX

Pradeep J.

Undergraduate Student
Department of Electronics
and Communication, NITK,
Surathkal – 575 025.

pradeepjraman@yahoo.com

Ajit V. Rao

Multimedia Codecs Group,
Texas Instruments,
Bangalore – 560 017.

ajitr@ti.com

Sumam David S.

Professor
Department of Electronics
and Communication, NITK,
Surathkal – 575 025.

sumam@ieee.org

ABSTRACT

Tapping our feet to the rhythm of a song comes naturally and appears easy. However, designing an algorithm to automatically detect beats in an audio signal is a challenging problem whose solution can potentially enable a world of exciting applications. In this paper, we propose a novel algorithm to detect musical beats using an important characteristic of beats – the presence of regularly spaced time instants characterized by a strong energy flux. Preliminary experimental results indicate that the beat locations detected automatically by using the proposed algorithm strongly correlate with locations discovered manually in music containing a strong bass component.

Keywords: *beat detection, energy flux*

1. INTRODUCTION

Human beings find it natural to detect the beats in a song and they rhythmically tap their feet or dance to its rhythm. Can the process of detecting beats in music be reliably automated? If this is indeed possible, it could potentially open up an exciting array of applications.

For instance, an automatic beat detection algorithm can help in efficient content-based retrieval of audio and in the automated indexing of a music collection by genres. Applied to a disco setting, the algorithm can enable auto-DJ mixing, auto play list generation and synchronized lighting controls. Automatic beat detection also makes possible, the synchronization of computer graphics with music leading to exciting possibilities in the design of high-end audio players. In addition, one can also envision applications in the fields of music information retrieval, speech/ music discrimination, as well as in designing better error concealment schemes for audio codecs.

2. BACKGROUND

Before proceeding further, it is useful to define the beats of a musical piece. Beats correspond to a periodic sequence of impulses that define a tempo for the music. In the beat detection problem, we are interested in determining the locations of the beats and hence the tempo of the music.

Typically (but not always) beat locations are characterized by a sharp variation in the energy profile of the audio signal – referred to as energy flux. Capturing the nature of this energy flux and using it to locate beats in the music is central to solving the beat detection problem.

A significant amount of research work has already been carried out in the beat detection field. Patin's excellent primer [1] provides an overview of the commonly used algorithms and orders them on a complexity scale. We observe in the literature that many researchers (example Laroche [2]) have proposed the use of the energy flux parameter for optimal results. After the energy flux has been calculated from the audio signal, a post-processing module must then be employed to infer the locations of the pulses. To achieve this, Kirovski and Attias [3] have proposed an interesting three-stage process based on statistical modeling. The mean period of the beat is first estimated, the mean onset of the beat is then calculated, and finally the actual onset is established. Another post-processing algorithm based on dynamic programming has been proposed in [2].

The energy flux parameter is most appropriate when a strong percussion-based beat component is present in the music. While this is generally the case, a whole subset of interesting algorithms are designed to detect beats without making any assumptions about the presence of strong energy flux. For example, in [4], the authors propose the use of knowledge of chord changes, note onsets and drum patterns to detect beats. While such an algorithm may be extremely complex, it is designed to work reliably even in the absence of drumbeats. Another good example of this class of algorithms is the work by Foote et al [5] where the concept of 'beat spectrum' is used. Auto-correlation based methods combined with sub-band filtering have also been applied for beat-detection in [6] and [7].

3. PROPOSED ALGORITHM

3.1. Overview

In this paper, we propose a novel beat detection algorithm based on a frequency-weighted energy flux parameter. Following the calculation of the energy flux, we employ a post-processing step based on a "comb" approach. The comb approach ensures that we simultaneously determine the optimal locations of all the beat cycles in the audio over a short time interval.

Two aspects of percussion-based beats are well-exploited in our algorithm: regularity of the beats and a strong change in the local energy profile at the beat locations. The energy change is captured in the flux parameter while the regularity is detected via the comb approach.

3.2. Framing and Transformation

The input audio data is represented by a single-channel audio signal {s(n)}. In case of multi-channel input audio signals, down-mixing to a mono signal is performed to generate s(n). The signal s(n) is subsequently divided into non-overlapping "frames" (consisting of N samples where N depends on the sampling rate of the input audio). It is important to note that the frame length also corresponds to the maximum uncertainty in the location of a beat.

The N audio samples in each frame i,

$$s_i(n) \qquad n = 0..N-1$$

are windowed and transformed into the frequency domain via a discrete Fourier Transform. The frequency domain samples are represented by

$$S_i(k) \qquad k = 0..N-1$$

In the flux calculation, we are only interested in the frequency-dependent energy value,

$$E_i(k) = \left|S_i(k)\right|^2$$

3.3.Sub-band weighting

The frequency domain is divided into three sub-bands: $[B_0=0, B_1)$, $[B_1, B_2)$, $[B_2, B_3=N/2)$

The energy of the audio signal in each band, j, is calculated as follows:

$$Q_i(j) = \sum_{k=B_j}^{B_{j+1}-1} E_i(k) \qquad j = 1, 2,3$$

3.4. Energy Flux

The energy flux between consecutive frames is calculated in each sub-band via a half-rectified difference of the sub-band energy (sub-band energy flux).

$$D_i(j) = \begin{cases} Q_i(j) - Q_{i-1}(j) & Q_i(j) > Q_{i-1}(j) \\ 0 & Otherwise \end{cases}$$

The half-wave rectification step ensures that we can isolate the positive energy flux points, which are the locations of interest.

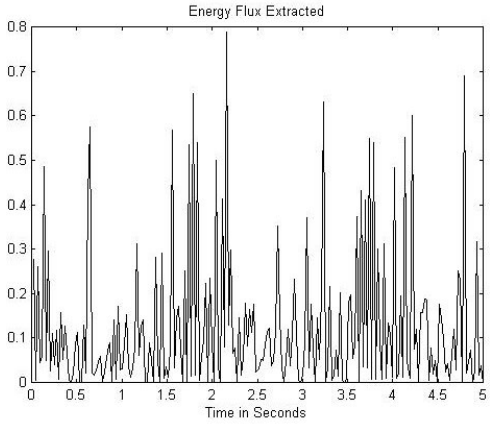

Figure 1: Energy flux parameter as a function of time for a segment of Michael Jackson' s song "Thriller"

The final energy flux parameter is calculated via a weighted sum of the sub-band flux values:

$$F_i = \sum_{j=0}^{2} W_j D_i(j)$$

Fig 1 shows an example of the variation in the energy flux parameter over time, for a five second segment of Michael Jackson's song, "Thriller".

3.5.Comb-based post-processing

Having computed the flux parameter, for every frame i we propose a post-processing step based on the "comb approach" to determine the optimal locations of the beats.

The comb approach attempts to identify the optimal locations of the periodically placed beats given a history of flux values F_i, for recent frames, i = 0,-1, -2, -3,..

The comb approach exploits the regular nature of the potential beat locations. We represent a comb by a periodic impulse train, defined completely by a frequency ("pitch") and phase. Generally, the beat frequency which is equivalent to the pitch is measured in beats per minute (bpm) rather than the mathematically traditional Hertz measure. Given our knowledge of the nature of typical musical beats, we restrict the pitch, r as follows:

$$L \le r \le H$$

Typical percussion-based music is characterized by a beat frequency between 30 to 180 bpm.

For a beat frequency of 'r' beats per minute, and phase p the comb may be defined as a train of impulses separated by z(r)-1 zeros as follows:

$$C_{r,p}(k) = \begin{cases} 1 & k = -(p + mz(r)) \quad m = 0,1..M \\ 0 & Otherwise \end{cases}$$

where

$$z(r) = \lfloor 60F_s / (Nr) \rfloor$$

The phase parameter p may take on the values

$$p = 0...z(r) - 1$$

Note that k=0 corresponds to the current frame and we are only interested in the comb $C_{r,p}(k)$ for k = 0,-1,-2, ... The locations of the non-zero values in the comb are referred to as "teeth". Note that each comb has a fixed number of teeth, M.

The output of the comb operation is defined as

$$O_{r,p} = \sum_{k=0}^{-\infty} C_{r,p}(k) F(k)$$

Having thus computed the comb outputs for different combinations of pitch values r and phase values p, we define the optima as the pitch and phase that result in the maxima of this function.

$$\{r_{k,opt}, p_{k,opt}\} = \underset{\{r,p\}}{\arg\max} \; O_{r,p}$$

Figure 2 shows a sample plot of the maximum output of the comb ($\max_{\{p\}} O_{r,p}$) versus the beat frequency, r.

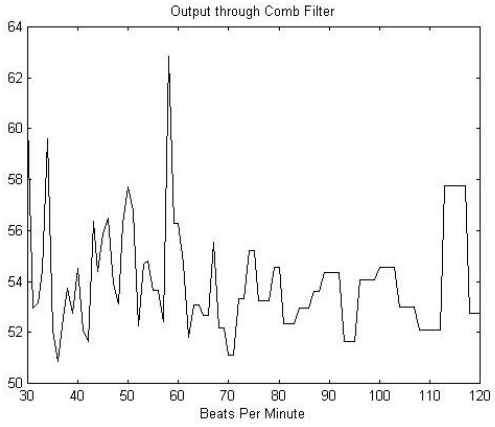

Figure 2: Plot displaying the maximum output of the comb (maximized over phase values) as a function of the beat frequency (bpm) for a five second segment of Michael Jackson's song "Thriller"

3.6.Practical Considerations

We performed extensive experiments to tune the basic approach suggested above and achieve good results in a practical scenario.

We implemented the proposed algorithm in the C programming language. For our practical experiments, we considered popular audio tracks belonging to various Western music genres – specifically rock, pop and dance. We intentionally selected songs with a strong beat component to help validate our approach. All audio tracks were sampled at the standard sampling frequency of F_s = 44100 Hz. The frame size N = 1024.

To calculate the flux parameter, we defined the three frequency bands as 0-200 Hz, 200-4500 Hz 4500- 22050 Hz. In terms of the FFT coefficient indices, this choice corresponds to the values B_1=4 and B_2=104. Furthermore, weights were assigned to the sub-bands to ensure that the middle sub-band was accorded the lowest importance. This was achieved by the choice

$$W_0=W_2=1.0, \quad W_1=0.01.$$

In the post-processing algorithm, the candidate pitch values were constrained by the limits L = 30 and H = 120 bpm with the number of comb teeth M = 30.

One important observation that we made is that the sharp nature of the comb $C_{r,p}(k)$ suggested in the previous section poses a problem in two cases: (i) when the frequency of the beats corresponds only approximately to the candidate pitch and (ii) when there is a slight variations in the regularity of the beat cycle. In both cases, the sharp and regular nature of the comb filter prevents the maxima from being correctly identified. To overcome this difficulty, we chose to supplement each impulse in the comb, $C_{r,p}(k)$ by two impulses, one on either side as follows:

$$C'_{r,p}(k) = \begin{cases} 1 & C_{r,p}(k \pm 1) = 1 \\ C_{r,p}(k) & Otherwise \end{cases}$$

In addition, it was also found that it is necessary to position low amplitude (0.5) "teeth" half -way between the locations of the teeth in $C'_{r,p}(k)$ to account for "half -beats" that are typical in music. For conciseness, we do not elaborate further on this minor modification to the basic approach.

3.7. Results

The final algorithm was tested on one randomly selected 60-second segment from each of 12 songs belonging to the rock, pop and dance genres. The automatically detected beat locations were then compared against manually identified locations of the beats – an error of upto 25 ms was allowed to compensate for the frame size of the algorithm. The results obtained by automated detection agreed largely with manual detection on the songs tested. The results were better in songs with a strong bass component and strongly defined beats. Table 1 summarizes the songs tested and the results of the comparison between the automatically detected beats and the manually identified ones. In 50% of the songs tested, the automatically determined beat locations agree perfectly with the manually identified ones. In the remaining 50%, we note that the main problem is the doubling and halving of the true pitch value.

While pitch doubling/halving does not necessarily harm in some end applications (auto-indexing of music and automated disco lighting for example), we have planned a further investigation of this problem and improvements to our proposed algorithm.

We note here also, that the complexity of our proposed algorithm is quite low. The average time taken to identify the beats in a 60 second sample was less than 3 seconds on a Pentium 4 system at 1.5 GHz running on the Windows XP platform.

4. FUTURE WORK

We believe that our experimental results are somewhat preliminary. They validate the basic approach strongly. However, we plan to direct a significant amount of our future effort towards solving the pitch doubling and halving problems. In addition, we are planning additional enhancements to the algorithm to handle fast variations in the beat frequencies. We have also planned to use the proposed algorithm on some of the applications such as auto DJ mixing and auto-playlist generation.

5. REFERENCES

[1] F. Patin, "Beat Detection Algorithms," http:\\www.yov408.free.fr

[2] J. Laroche. "Efficient Tempo and Beat Tracking in Audio Recordings", Journal of the Audio Engr. Soc. vol 51, no 4, pp 226-233, 2003.

[3]D. Kirovski and H. Attias, "Audio Watermark Robustness to Desynchronization via Beat Detection", Microsoft Research.

[4]M. Goto, "An Audio-based Real-time Beat Tracking System for Music With or Without Drum Sounds", *Journal Of New Music Research*, Vol 30, no 2 , pp.159 - 171 , 2001.

[5]J. Foote and S. Uchihashi, " The Beat Spectrum: A new approach to rhythm analysis", *Int. Conf. On Multimedia & Expo,* 2001.

[6]R. Jarina, N. O'Connor, S. Marlow, and N. Murphy, "Rhythm detection for speech-music discrimination in MPEG compressed domain", Centre for Digital Video Processing.

[7]G. Tzanetakis, G. Essl, and P. Cook, " Audio Analysis using the Discrete Wavelet Transform", *Proc. Conf. In Acoustics and music theory applications,* WSES, Sept. 2001.

[8]C. Chan and K. Liu, "Beat Tracking Strobe", Cornell University.

Song	Artist	Comments	Result
Thriller	Michael Jackson	Strong bass	Success
Formula 1	D J Visage	Techno/Bass	Success
Daddy Cool	Boney M	Strong Bass	Pitch halving
Every Breath you take	Police	Strong Beats	Success
Money for nothing	Dire Straits	Strong Bass	2/3 Pitch
Bad	Michael Jackson	Strong Beats	Success
Tubthumping	Chumbawumba	Strong Beats	Pitch Halving
My heart will go on	Celiene Dion	None	Success
More than words	Extreme	No Bass	Success
Pretty Woman	Ray Orbinson	Strong Bass	Pitch halving
That thing you do	The wonders	Strong Bass	Pitch doubling
I'll be there for you	The Rembrands	Bass	Pitch halving

Table 1: Table of songs used for experiments on automatic beat detection and results obtained