1.1-3

On-Line Musical Beat Tracking with Phase-Locked-Loop (PLL) Technique

Yu Shiu and C.-C. Jay Kuo

Department of Electrical Engineering and Integrated Media Systems Center
University of Southern California, Los Angeles, CA 90089, USA
E-mails: yshiu@usc.edu, cckuo@sipi.usc.edu

Abstract- **An on-line beat tracking algorithm based on the digital phase-locked-loop (PLL) technique is proposed in this work, which estimates both the music tempo and the position of beat pulses in real time. In addition to a detailed description of the proposed algorithm, we show that the PLL-technique is effective in its tracking performance and simple to implement.**

I. INTRODUCTION

When listening to the pop music, many people have the capability to follow the tempo by foot-tapping with the music. In other words, one could estimate the music tempo based on past observations to predict the location of the next beat. The real-time musical beat tracking system finds a wide range of applications in consumer electronics for musical signal processing. For example, in an automatic music accompanying system, where the system estimates the musical tempo of a singer and then provides the accompanying music to match the singer's tempo.

Musical beat analysis consists of two parts: tempo estimation and determination of the position of each individual beat pulse, where the musical tempo refers to the number of beats in a given time and it is often in the unit of "beats per minute" (BPM). For some music genres such as pop and rock, the tempo is likely to stay the same throughout the whole song. For some other music genres, we may see a gradual or rapid tempo change from one to the other. In classical music, gradual tempo change is used as a means to deliver expressive performance.

A song can be viewed as a quasi-periodic function whose frequency is about equal to the music tempo while the position of the individual pulse can be viewed as the phase information of a period. Thus, to identify the exact pulse location can be treated as a phase trakcing problem. The phase-locked loop (PLL) technique [1][2] has been developed and used in communication and control systems for decades. Simply speaking, it is a time-synchronizing mechanism that attempts to align an output signal with an input signal in terms of both frequency and phase. The PLL technique is applied to on-line beat tracking in this work. It not only tracks the beat (*i.e.*, estimating the time instance of beats continuously) but also adapts to the tempo change of music signals. In addition to a detailed description of the proposed algorithm, we show that the PLL technique is effective in its tracking performance and simple to implement. Consequently, it is an ideal solution to be adopted by consumer electronics for musical signal analysis and processing.

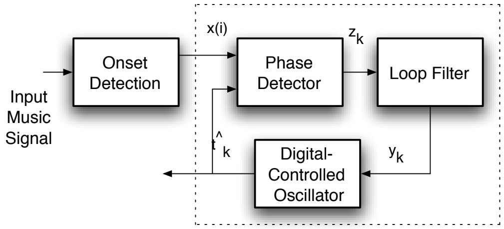

Fig. 1. The block diagram of the proposed PLL-based beat tracking system.

II. PLL-BASED BEAT TRACKING ALGORITHM

The block diagram of the proposed PLL-based beat tracking system is shown in Fig. 1. It consists of two main modules: the onset detection module and the PLL module, which is enclosed by a dotted box. The function of each individual module is described below.

Onset Detection Module

The input to this module is the audio waveform of a musical input signal, and its output is the onset signal $x(t)$ that indicates a significant change in the lowpass-filtered audio waveform. The onset signal usually consists of a sequence of pulses of varying spacing and magnitute. There have been several on-set detection algorithms proposed in the literature, for example, [3]. Here, we aim to extract music onsets for two kinds of music content changes: instantaneous noise-like pulses caused by percussion instruments and changes of music pitches/harmonies due to the arrival of new notes. They constitute beats to build a sense of constant or varying tempo. Our method to extract the music onset is to compute the change of the spectral content between two adjacent shifting windows with 50% overlap of 20-msec long. Mathematically, we adopt the following mel-scale cepstral distance measure[4]:

$$d_c^2 = \sum_{m=1}^{L} (c_m - c'_m)^2, \qquad (1)$$

where c_m and c'_m are MFCC (mel-scale frequency cepstral coefficient) for the two neiboring frames, respectively. The output of the onset detection module is the onset signal denoted by $x(i)$, which serves as the input to the next module.

Phase-Locked-Loop (PLL) Module

Without loss of generality, we consider an interval that has a constant music tempo. Under this stationary assumption, we may assume a reference clock of the PLL system of the following form

$$t_k = kT_0 + t_0, \qquad (2)$$

1-4244-0763-X/07/$20.00 ©2007 IEEE

where t_k indicates the time instance of the k-th beat pulse and T_0 is the period of the beats. The reference clock has a uniform spacing in the time domain. The actual beat pattern is aligned well with the reference clock yet with the possibility of some missing pulses. The output of the PLL module is denoted by $\hat{t}_k$ that provides the estimated position of the k-th beat pulse for the input music onset $x(i)$, $i \leq \hat{t}_k$. Note that $x(i)$ may consist of actual beat pulses as well as non-beat pulses. Even though t_k is not observable, we would like to make $\hat{t}_k$ track t_k as closely as possible under the observation of $x(i)$, $i \leq \hat{t}_k$.

As shown in Fig. 1, PLL consists of three components: the phase detector (PD), the loop filter (LF) and the digital-controlled oscillator (DCO) [2]. Define the beat observation selection function of $x(i)$ as

$$f_k(x(i)) = t_k + n_k \tag{3}$$

that attempts to estimate the position of the k-th beat pulse, t_k, based on observaed input $x(i)$, $i \leq \hat{t}_k$, and n_k denotes the noise term. PD calculate the time difference (or phase difference) z_k between $\hat{t}_k$ and $f_k(x(i))$; namely,

$$z_k = f_k(x(i)) - \hat{t}_k = t_k + n_k - \hat{t}_k. \tag{4}$$

LF is a lowpass filter used to remove AC components and keep primarily the DC component. Finally, DCO is adopted to accumulate the DC-only phase difference so as to update the beat pulse estimate $\hat{t}_k$. Roughly speaking, when the beat pulse calculated from $x(i)$, *i.e.* $f_k(x(i))$, leads the synthesized pulse $\hat{t}_k$ from DCO, the phase error will be positive. This positive feedback will accelerate the clock of the synthesized signal. Conversely, when the beat pulse of $f_k(x(i))$ lags behind the synthesized pulse form DCO, the phase error will be negative, which in turn slows down the clock of the synthesized signal $\hat{t}_k$. Under the stationary assumption, the synthesized signal $\hat{t}_k$ will converge to a steady-state value that get synchronized with the underlying reference clock t_k. If the tempo of the input music changes slowly, the above system can still track the changing tempo dynamically.

Beat Observation Selection Function

The simplest selection function $f_k(x(i))$ in PD is to find the location of the music onset that has the highest magnitude around estimated beat $\hat{t}_k$ within a predefined window. However, since the pulse with the largest onset value is not necessarily the beat pulse, it is vulnerable to tracking the wrong pulse and its performance is not reliable. In this work, we consider a different approach. First, the period T_0 is estimated based on a 3-sec window of previous music onsets via the autocorrelation function computation. Then, a comb filter [3] is applied to observations in the previous and the current windows. The comb filter can be written as

$$x'(i) = \alpha x'(i - T_0) + (1 - \alpha)x(i), \tag{5}$$

where $x'(i)$ is the output of the comb filter and α is a predefined parameter in $(0, 1)$. $x'(i)$ will enahnce the periodic characteristics of $x(i)$ by adding $x'(i - T_0)$, which includes part of $x(i - T_0)$, to current $x(i)$. It helps keep the periodicity of beat pulses even in the extreme case where the onset value is small or zero in a beat position. The case happens a lot in real music pieces. For example, during the transition of two musical sections, there may be no beat pulses at all. However, human can continue the beat based on previous tempo and beat pulses. Finally, among periodic pulses in comb filter's output $x'(i)$, the pulse that is closest to the estimated $\hat{t}_k$ is selected as the next beat pulse; namely, $f_k(x(i))$.

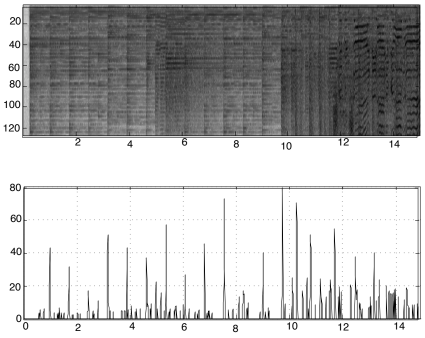

Fig. 2. The spectrogram (top) and its corresponding music onsets (bottom) for the first 15 seconds of Oasis' *Don't look back in anger*.

III. EXPERIMENTAL RESULTS AND DISCUSSION

Fig. 2 shows an example of music onset along with its spectrogram for the first 15 seconds of Oasis' *Don't look back in anger*. The beat pulse pattern is clearly observed in the first half of the interval but it looks confusing in the second half. Many onsets are caused by non-beat pulses. For example, some of them correspond to the dense drum sounds starting near 10 sec and others correspond to the vocal sound starting near 12 sec. These extra onsets make beat pulse tracking difficult. The proposed algorithm described above works well for this challenging case. The beat tracking results for the first 60 sec are shown in Fig. 3, where we plot the estimated interval between two adjacent beat pulses $T_k = t_{k+1} - t_k$ versus the beat index, k. Strictly speaking, the period of this music piece is not constant. It is time varying with a small deviation from the mean period 728 msec, which roughly corresponds to music tempo of 82.4 BPM.

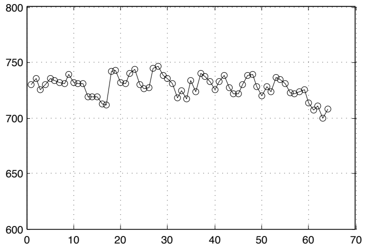

Fig. 3. The estimated time interval between two adjacent beat pulses $T_k = t_{k+1} - t_k$ as a function of beat index k.

IV. REFERENCES

[1] R. E. Best,*Phase-Locked Loops: Theory, Design and Applications*, 2nd ed., McGraw-Hill, 1993.

[2] W. C. Lindsey and C. M. Chie, "A survey of digital phase-locked loops", *Proceedings of the IEEE*, vol. 69, No. 4, pp. 410-431 1981.

[3] Eric Scheirer, "Tempo and beat analysis of acoustic musical signals", *Journal of Acoustic Society America*, vol. 103, pp. 588-601 1998.

[4] L. Rabiner and B.-H. Juang, *Fundamentals of Speech Recognition*, Prentice Hall, 1993.