A Real-time Beat Tracking System for Audio Signals

Masataka Goto† and Yoichi Muraoka
School of Science and Engineering, Waseda University
3-4-1 Ohkubo Shinjuku-ku, Tokyo 169, JAPAN.
+81-3-3209-5198 / {goto, muraoka}@muraoka.info.waseda.ac.jp

ABSTRACT: This paper presents a beat tracking system that processes musical acoustic signals and recognizes temporal positions of beats in real time. Most previous systems were not able to deal with audio signals that contained sounds of various instruments, especially drums. Our system deals with popular music in which drums maintain the beat. To examine multiple hypotheses of beats in parallel, our system manages multiple agents that predict beats by using autocorrelation and cross-correlation according to different strategies. In our experiment with eight pre-registered drum patterns, the system implemented on a parallel computer correctly tracked beats in 42 out of 44 commercially distributed popular songs.

1 Introduction

Beat tracking is an important initial step in computer emulation of human music understanding, since beats are fundamental to the perception of Western music. Even if a person cannot completely segregate and identify every sound component, he can nevertheless track musical beats and keep time to music by hand-clapping or foot-tapping. We therefore first build a computational model of beat perception and then extend the model, just as a person recognizes higher-level musical events on the basis of beats.

Most previous beat tracking systems [Dannenberg and Mont-Reynaud, 1987; Allen and Dannenberg, 1990; Rosenthal, 1992; Desain and Honing, 1994] have dealt with MIDI as their input. Their reliance on MIDI, however, limited the input source to electronic instruments. Those systems generally dealt with classical works, in particular piano solo. Although some systems [Katayose *et al.*, 1989; Vercoe, 1994] dealt with audio signals, they were not able to process music played on ensembles of a variety of instruments, especially drums.

Our beat tracking system, called *BTS*, processes monaural acoustic signals that contain sounds of various instruments in real time. BTS deals with popular music in which drums maintain the beat. Not only does BTS predict the temporal position of the next beat (*beat time*); it also determines whether the beat is strong or weak (*beat type*)[1]. In other words, BTS can track beats at the half-note level.

Our previous system (*PreBTS*) [Goto and Muraoka, 1994] based on multiple-agent architecture had the following problems: (1) PreBTS assumed only the case where a bass drum and a snare drum usually sounded on the strong and weak beats, respectively. (2) PreBTS sometimes failed to infer correct beat type when characteristic frequencies of drum-sounds were not acquired correctly. (3) PreBTS implemented a mechanism for correcting a typical beat-tracking error, which consisted of having agent-pairs track alternative hypotheses that differed only in phase. Though effective, this system was nevertheless not sufficiently flexible for certain situations. (4) PreBTS occasionally made double-tempo or half-tempo errors.

Our solutions to these problems are outlined as follows:[2] (1) BTS leverages musical knowledge represented as pre-registered drum patterns of the bass drum and the snare drum. These patterns represent how drum-sounds are used in a large class of popular music. The results of matching these patterns with the currently detected drum pattern make it possible to determine the beat type and which note-value a beat corresponds to. (2) To improve the method of detecting the snare drum, BTS finds noise components that are widely distributed along the frequency axis, since such a frequency profile is typical of the snare drum. (3) Multiple agents that track beats according to different strategies utilize auto- and cross-correlation of detected onset times to predict the next beat. Each agent first calculates an inter-beat interval, and then determines the appropriate beat position by evaluating every beat-position possibility that the obtained inter-beat interval could support. (4) Agents are grouped into pairs; in each pair one agent tries to track beats at a relatively higher tempo and the other tracks them at a lower tempo. These two agents then inhibit each other. This enables one agent to track the correct beats even if the other agent tracks beats with double or half the correct tempo.

To perform this computationally-intensive task in real time, BTS has been implemented on a parallel computer, the Fujitsu AP1000. In our experiment with eight pre-registered drum patterns, BTS correctly tracked beats in 42 out of 44 songs sampled from compact discs. Moreover, we have developed an application with BTS that displays a computer graphics dancer whose motion changes with musical beats in real time.

---

† A Research Fellow of the Japan Society for the Promotion of Science.

[1] In this paper, a *strong beat* is either the first or third quarter note in a measure; a *weak beat* is the second or fourth.

[2] The general issues of tracking beats in acoustic signals and our solutions are described in our previous papers [Goto and Muraoka, 1994; Rosenthal and Goto *et al.*, 1994].

2 System Description

Figure 1 shows an overview of our beat tracking system. BTS assumes that the time-signature of an input song is 4/4, and that its tempo is constrained to be between 65 M.M. and 185 M.M. and almost constant; these assumptions fit a large class of popular music. The emphasis in our system is on finding the temporal positions of quarter notes in audio signals rather than on tracking tempo changes; in the repertoire with which we are concerned, tempo variation is not a major factor. BTS reports *beat information* (*BI*) that consists of the *beat time*, its *beat type*, and the current tempo, in time to the input music.

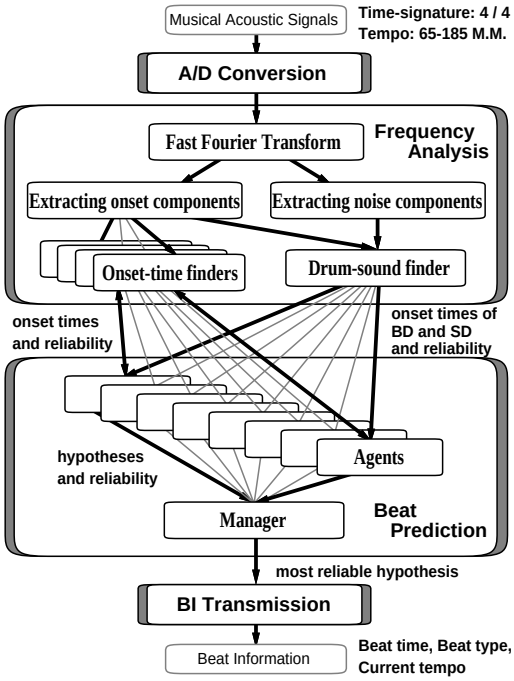

Figure 1: Overview of BTS

The two main stages of processing are *Frequency Analysis*, in which a variety of cues are detected, and *Beat Prediction*, in which multiple hypotheses of beat positions are constructed and evaluated (Figure 1). In the *Frequency Analysis* stage, BTS employs multiple onset-time finders that detect various cues such as onset times in several different frequency ranges, and onset times of two different kinds of drum-sounds: a bass drum (*BD*) and a snare drum (*SD*). In the *Beat Prediction* stage, BTS manages multiple agents that make parallel hypotheses based on detected onset times according to different strategies. Each agent first calculates an inter-beat interval (*IBI*) using autocorrelation[3] and predicts the next beat time using cross-correlation; it then infers the beat type, and evaluates the reliability of its own hypothesis. The manager then determines the position of the next beat on the basis of the most reliable hypothesis. Finally, in the *BI Transmission* stage, BTS transmits BI to other application programs via a computer network.

The following sections describe the main stages of Frequency Analysis and Beat Prediction.

2.1 Frequency Analysis

Onset components and noise components are first extracted from the frequency spectrum calculated by the Fast Fourier Transform. Onset-time finders then detect onset times in different frequency ranges and with different sensitivity levels. At the same time, another process, a drum-sound finder, detects BD and SD.

2.1.1 Extracting onset components / Extracting noise components

Frequency components whose power has been rapidly increasing are extracted as onset components. The onset components and their degree of onset (rapidity of increase in power) are obtained by a process that takes into account the power present in nearby time-frequency regions.

BTS extracts noise components as a preliminary step to detecting SD. Because non-noise sounds typically have harmonic structures and peak components along the frequency axis, frequency components whose power is roughly uniform locally are extracted and considered to be potential SD sounds.

2.1.2 Onset-time finders

Fourteen onset-time finders use different sets of frequency-analysis parameters. Each finder sends its onset information to a particular *agent-pair*. Each onset time is given by the peak time found by peak-picking in $D(t)$ along the time axis, where $D(t) = \sum_f d(t, f)$, and $d(t, f)$ is the degree of onset of frequency f at time t. The sum $D(t)$ is linearly smoothed with a convolution kernel before its peak time is calculated.

2.1.3 Drum-sound finder

A drum-sound finder detects BD from the onset components and SD from the noise components. Note that BTS cannot simply use the detected drums to track beats, because the results of this detection include many mistakes. The detected drums are used only to label a beat time with its beat type.

[Detecting onset times of BD]

Because the sound of BD is not known in advance, BTS learns the characteristic frequency of BD corresponding to a particular song. The finder finds peaks in the onset components along the frequency axis and histograms them (Figure 2). The finder then judges that BD has sounded at times when an onset's peak frequency

[3]The paper [Vercoe, 1994] also proposed using a variant of autocorrelation for rhythmic analysis.

coincides with the characteristic frequency that is given by the lowest-frequency peak of the histogram.

[Detecting onset times of SD]

BTS detects noise components widely distributed along the frequency axis as SD. First, the noise components are quantized (Figure 2). Second, the finder calculates how widely noise components are distributed along the frequency axis in the quantized noise components (*degree of wide distribution* $c(t)$). Finally, the onset time of SD is obtained by peak-picking of $c(t)$ in the same way as in the onset-time finder.

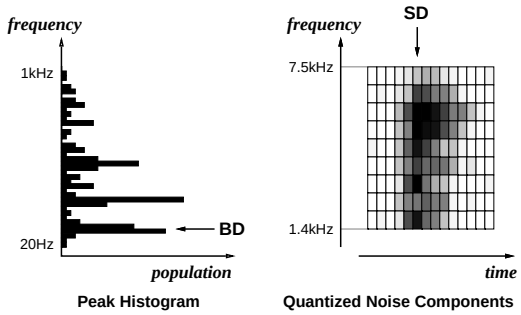

Figure 2: Detecting BD and SD

2.2 Beat Prediction

Twenty-eight agents maintain their own hypotheses, each of which consists of a predicted next-beat time, its beat type, and the current IBI. These hypotheses are gathered by the manager (Figure 1), and the most reliable one is selected as the output. The twenty-eight agents are grouped into fourteen agent-pairs. Each agent-pair is different in that it receives onset information from a different onset-time finder. As mentioned in the first section of this paper, two agents in a pair try to track beats with different ranges of tempi. In other words, the two agents have the different assigned ranges of IBI.

The following sections describe the formation and management of hypotheses. First, each agent predicts the next beat time using auto- and cross-correlation, and then evaluates its own reliability (*Predicting next beat*). Second, the agent infers its beat type and modifies its reliability (*Inferring beat type*). Finally, the manager selects the most reliable hypothesis from the hypotheses of all agents (*Managing hypotheses*).

2.2.1 Predicting next beat

Beats are characterized by two properties: IBI (period) and phase. The phase of a beat is the relative beat position to the most recent onset time. We measure phase in radians; for a quarter-note beat, for example, an eighth-note displacement corresponds to a phase-shift of π radians (Figure 3).

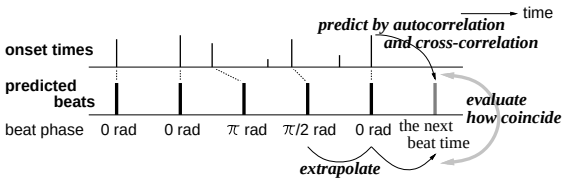

Figure 3: Predicting next beat

Each agent first calculates the current IBI (period). The IBI is given by the maximum value within the assigned IBI range in the autocorrelation function of the received onset times. To determine the beat phase, the agent then calculates cross-correlation between the onset times and a set of equally-spaced pulse sequences whose temporal interval is the IBI. The maximum value in the cross-correlation function provides the plausible beat phase. This calculation corresponds to evaluating all possibilities of the beat phase under the current IBI. The next beat time is thus predicted on the basis of the IBI and the current beat phase.

Each agent evaluates the reliability of its own hypothesis. This is determined on the basis of how the next beat time predicted by the auto- and cross-correlation coincides with the time extrapolated from the past two beat times (Figure 3). If they coincide, the reliability is increased; otherwise, the reliability is decreased.

2.2.2 Inferring beat type

Each agent determines the beat type by matching the pre-registered drum patterns of BD and SD with the currently detected drum pattern. Figure 4 shows two examples of the pre-registered patterns. These patterns represent how BD and SD are typically played in rock and pop music. The beginning of a pattern should be a strong beat, and the length of the pattern is restricted to a half note or a measure.

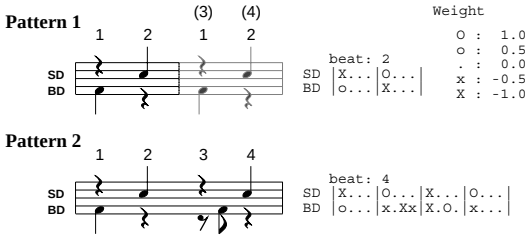

Figure 4: Examples of pre-registered drum patterns

The beat type and its reliability are obtained as follows: (1) The onset times of drums are quantized to the currently detected pattern, with one sixteenth-note resolution that is obtained by interpolating between successive beat times (Figure 5). (2) The *matching score* of each pre-registered pattern is calculated by matching the pattern with the currently detected pattern: The score is weighted by the product of the weight in the pre-registered pattern and the reliability of the detected onset. (3) The beat type is inferred from the fact that the beginning of the best-matched pattern indicates the position of the strong beat (Figure 6). The reliability of the beat type is given by the highest matching score.

If the reliability of the beat type is high, the IBI in the hypothesis can be considered to correspond to a quarter note. In that case, the reliability of the hypothesis is increased so that a hypothesis with an IBI corresponding to a quarter note is likely to be selected.

2.2.3 Managing hypotheses

The manager classifies all agent-generated hypotheses into groups, according to beat time and IBI. Each group has an overall reliability, given by the sum of the reliability of the group's hypotheses. The most reliable hypothesis in the most reliable group is selected as the output and sent to the BI Transmission stage.

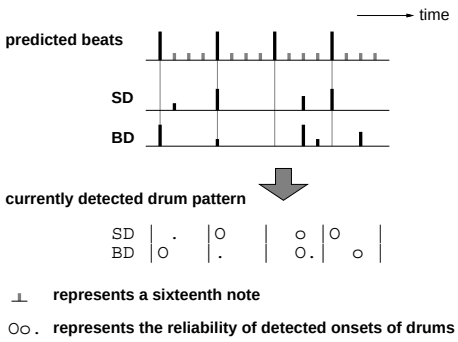

Figure 5: A drum pattern detected from an input

3 Experiments and Results

We tested BTS on 44 popular songs performed by 31 artists in the rock and pop music genres. Their tempi ranged from 67 M.M. to 185 M.M. and were almost constant. In our experiment with eight pre-registered drum patterns, BTS correctly tracked beats in 42 out of 44 songs in real time. After the BD and SD had sounded stably for a few measures, the beat type was obtained correctly.

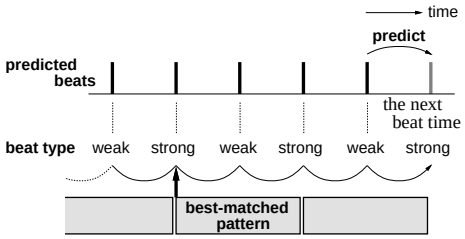

Figure 6: Inferring beat type

We discuss the reason why BTS made mistakes in two of the songs. In one song, BTS made a half-tempo error, in other words, the output IBI was double the correct IBI. Since onset times on strong beats were often not detected, agents that had the double IBI inhibited agents that had the correct one. In the other song, the beat type was wrong for most of the song. Because SD sounded on every quarter note, the detected drum pattern was not well matched with the pre-registered ones.

4 Conclusion

We have described the configuration and implementation of a real-time beat tracking system (*BTS*) that can deal with audio signals played on ensembles of a variety of instruments. BTS manages multiple agents that track beats according to different strategies in order to examine multiple hypotheses in parallel. This enables BTS to follow beats without losing track of them, even if some hypotheses become wrong. The use of drum patterns pre-registered as musical knowledge makes it possible to determine whether a beat is strong or weak and which note-value a beat corresponds to. The experimental results show that our beat-tracking model based on multiple-agent architecture is robust enough to handle real-world audio signals.

We plan to upgrade our beat-tracking model to understand music at a higher level and to deal with other musical genres. Future work will include a study on appropriate musical knowledge for dealing with musical audio signals, and application to various multimedia systems for which beat tracking is useful, such as video/audio editing, live ensemble, stage lighting control, and synchronization of real-time CG with music.

References

(Our papers can be obtained from ''http://www.info.waseda.ac.jp/muraoka/members/goto/''.)

[Allen and Dannenberg, 1990] Paul E. Allen and Roger B. Dannenberg. Tracking musical beats in real time. In *Proceedings of the 1990 International Computer Music Conference*, pages 140−143, 1990.

[Dannenberg and Mont-Reynaud, 1987] Roger B. Dannenberg and Bernard Mont-Reynaud. Following an improvisation in real time. In *Proceedings of the 1987 International Computer Music Conference*, pages 241−248, 1987.

[Desain and Honing, 1994] Peter Desain and Henkjan Honing. Advanced issues in beat induction modeling: syncopation, tempo and timing. In *Proceedings of the 1994 International Computer Music Conference*, pages 92−94, 1994.

[Goto and Muraoka, 1994] Masataka Goto and Yoichi Muraoka. A beat tracking system for acoustic signals of music. In *Proceedings of the Second ACM International Conference on Multimedia*, pages 365−372, 1994.

[Katayose *et al.*, 1989] H. Katayose, H. Kato, M. Imai, and S. Inokuchi. An approach to an artificial music expert. In *Proceedings of the 1989 International Computer Music Conference*, pages 139−146, 1989.

[Rosenthal and Goto *et al.*, 1994] David Rosenthal, Masataka Goto, and Yoichi Muraoka. Rhythm tracking using multiple hypotheses. In *Proceedings of the 1994 International Computer Music Conference*, pages 85−87, 1994.

[Rosenthal, 1992] David Rosenthal. *Machine Rhythm: Computer Emulation of Human Rhythm Perception*. PhD thesis, Massachusetts Institute of Technology, 1992.

[Vercoe, 1994] Barry Vercoe. Perceptually-based music pattern recognition and response. In *Proceedings of the Third international conference for the perception and cognition of music*, pages 59−60, 1994.