

An Audio-based Real-time Beat Tracking System and Its Applications

Masataka Goto[†] and Yoichi Muraoka

School of Science and Engineering, Waseda University
3-4-1 Ohkubo Shinjuku-ku, Tokyo 169-8555 JAPAN

goto@etl.go.jp, muraoka@muraoka.info.waseda.ac.jp

ABSTRACT: This paper describes a real-time beat tracking system that recognizes a rhythmic structure in real-world audio signals sampled from popular-music compact discs. Most previous beat-tracking systems dealt with MIDI signals and had difficulty in processing, in real time, audio signals containing sounds of various instruments and in tracking beats above the quarter-note level. Our system can process music without drums and music with drums and can — on the basis of three kinds of musical knowledge (of onset times, of chord changes, and of drum patterns) — recognize the hierarchical beat structure comprising the quarter-note, half-note, and measure levels. This paper also introduces several beat-tracking applications, such as beat-driven real-time computer graphics and lighting control.

1 Introduction

Our goal is to build a real-time system that can track musical beats in real-world audio signals like those sampled from compact discs. We think that an important initial step in the computational modeling of music understanding is to build such a system which, even in its preliminary implementation, can work in real-world environments. This is important because, as known from the scaling-up problem [Kitano, 1993] in the domain of artificial intelligence, it is hard to scale-up a system whose preliminary implementation works only in laboratory (toy-world) environments. Moreover, this real-world oriented approach facilitates the implementation of various practical applications in which music synchronization is necessary.

Most previous beat-tracking related systems, however, had difficulty working in real-world acoustic environments. Most of those systems [Dannenberg and Mont-Reynaud, 1987; Desain and Honing, 1989; 1994; Allen and Dannenberg, 1990; Driesse, 1991; Rosenthal, 1992a; 1992b; Rowe, 1993; Large, 1995] used as their input MIDI-like representations. Since it is quite difficult to obtain complete MIDI representations from real-world audio signals, the applications of MIDI-based systems are limited. Some systems [Schloss, 1985; Katayose *et al.*, 1989; Vercoe, 1994; Todd, 1994; Todd and Lee, 1994; Scheirer, 1998] dealt with audio signals, but they either did not consider the higher-level beat structure above the quarter-note level or did not process, in real time, popular music sampled from compact discs. Although we developed two beat-tracking systems for real-world audio signals, one for music with drums [Goto and Muraoka, 1994; 1995; 1998] and the other for music without drums [Goto and Muraoka, 1996; 1997b], they were separate systems and were not able to recognize the measure level.

In this paper we describe a beat-tracking system that can deal with the audio signals of popular-music compact discs in real time and can recognize the *hierarchical beat structure* comprising the *quarter-note level* (almost regularly spaced beat times), the *half-note level*, and the *measure level* (bar-lines). This structure is shown in Figure 1. In addition, the system can process music without drums and music with drums. The system assumes that the time-signature of an input song is 4/4 and that its tempo is roughly constant and is either between 61 M.M. and 185 M.M. (for music with drums) or between 61 M.M. and 120 M.M. (for music without drums).

The main issues in recognizing the beat structure in real-world musical acoustic signals can be summarized as (1) detecting beat-tracking cues in audio signals, (2) interpreting the cues to infer the beat structure, and (3) dealing with ambiguity of interpretation. First, it is necessary to develop methods for detecting effective musical cues in audio signals. Note that most of the cues available in MIDI signals are hard to detect in complex audio signals. Second, higher-level processing using musical knowledge is indispensable for inferring each level of the hierarchical beat structure from the detected cues. Third, it must be taken into account that multiple interpretations of beats are possible at any given point. Because there is not necessarily a single specific sound that directly indicates the beat position, there are various ambiguous situations, such as ones where several detected cues may correspond to a beat and where different *inter-beat intervals* (the temporal difference between two successive beats) seem plausible.

In the following sections we introduce our new approach to the beat-tracking problem and propose a beat-tracking model that addresses the issues mentioned above. We then show experimental results of our system implemented on the basis of the model. Finally, we present several beat-tracking applications we have already developed.

2 Beat-Tracking Problem (Inverse Problem)

In our formulation the beat-tracking problem is defined as a process that organizes musical audio signals into the hierarchical beat structure. This problem can be considered the *inverse problem* of the following three forward processes

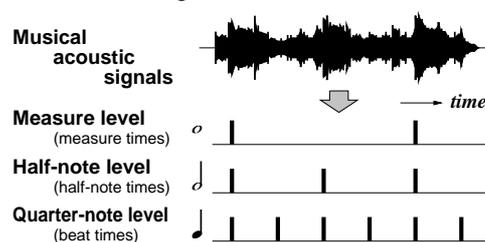


Figure 1: Beat-tracking problem.

[†] Currently at Machine Understanding Division, Electrotechnical Laboratory. 1-1-4 Umezono, Tsukuba, Ibaraki 305-8568 JAPAN.

by music performers: the process of indicating or implying the beat structure in musical elements when performing music, the process of producing musical sounds (playing musical instruments), and the process of acoustic transmission of those sounds. Although music is temporally organized along the hierarchical beat structure in the brains of performers, the structure is not explicitly expressed in music; it is implied in the relations among various musical elements, which are not fixed and are dependent on musical genres or pieces. All the musical elements constituting music are then transformed into audio signals through the process of musical sound production and the process of acoustic transmission.

The principal reason that beat tracking is intrinsically difficult is that it is the problem of inferring the original beat structure that is not explicitly expressed in music. The degree of beat-tracking difficulty is therefore not determined simply by the number of musical instruments performing a musical piece; it depends on how explicitly the beat structure is expressed in the piece. The main reason that different musical genres and instruments have different tendencies with regard to beat-tracking difficulty is that they have different customary tendencies with regard to the explicitness with which their beat structure is indicated.

In audio-based beat tracking, however, it is also difficult to detect musical elements as beat-tracking cues in the input audio signal. In that case, the more musical instruments played simultaneously and the more complex the audio signal, the more difficult it is to detect those beat-tracking cues.

3 Beat-Tracking Model (Inverse Model)

To solve this inverse problem, we propose a beat-tracking model that consists of two component models: the *inverse model* of the process of indicating the beat structure, and a model of extracting musical elements (Figure 2).¹ The three issues raised in the Introduction are addressed in our beat-tracking model as described in the following three sections.

3.1 Detecting Beat-Tracking Cues in Audio Signals

In the model of extracting musical elements, three kinds of musical elements — onset times, chord changes, and drum patterns — are detected as the beat-tracking cues. The onset times can be detected by a frequency analysis process that takes into account such factors as the rapidity of an increase in power and the power present in nearby time-frequency regions [Goto and Muraoka, 1997b]. It is difficult, however, to detect chord changes and drum patterns by this kind of bottom-up frequency analysis. We therefore proposed a method for detecting them by making use of provisional beat times, which are inferred on the basis of onset times, as top-down information [Goto and Muraoka, 1995; 1996; 1997b; 1998].

Each musical element is represented in our system as follows. Onset times are represented as an *onset-time vector* that consists of the onset times of all the frequency ranges. A chord change is represented as a *chord change possibility* that indicates how much dominant frequency components included in chord tones and their harmonic overtones change in a frequency spectrum. A drum pattern is represented as a temporal pattern of playing a bass drum and a snare drum.

3.2 Interpreting Beat-Tracking Cues to Infer the Hierarchical Beat Structure

Each level of the beat structure is inferred on the basis of the inverse model of the process of indicating the beat structure. The inverse model is represented by the following three kinds of musical knowledge (heuristics) corresponding to the three kinds of musical elements.

- musical knowledge of onset times
To infer the quarter-note level, the system uses the following knowledge: “onset times tend to coincide with beat times (i.e., sounds are likely to occur on beats)” and “a frequent inter-onset interval is likely to be the inter-beat interval.” By using autocorrelation and cross-correlation of the onset-time vectors, the system determines the inter-beat interval and predicts the next beat time as described in [Goto and Muraoka, 1996; 1997b].
- musical knowledge of chord changes
To infer each level of the structure, the system uses the following knowledge: “chords are more likely to change on beat times than on other positions between two successive correct beats,” “chords are more likely to change on half-note times than on other positions of beat times,” and “chords are more likely to change at the beginnings of measures than at other positions of half-note times.” Figure 3 shows a sketch of how to determine the half-note and measure times based on the chord change possibilities [Goto and Muraoka, 1997b].
- musical knowledge of drum patterns
When an input drum pattern that is currently detected in the audio signal is well matched with one of the prestored drum patterns that represent how drum-sounds are utilized in a large class of popular music, the system uses the

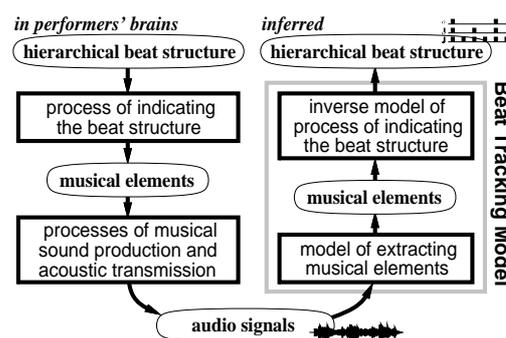


Figure 2: Beat-tracking model.

¹The basic concept of this model can be applied to more general music understanding models by replacing the words like “beat structure.”

following knowledge to infer the quarter-note and half-note levels: “the input drum pattern has the appropriate inter-beat interval” and “the beginning of the input drum pattern indicates a half-note time.” Figure 3 also shows a sketch of how to determine the half-note times based on the best-matched drum pattern [Goto and Muraoka, 1995; 1998].

In inferring the quarter-note and half-note levels, the two kinds of musical knowledge of chord changes and drum patterns are selectively applied according to the presence or absence of drum-sounds. This is done in order to give priority to the knowledge of drum patterns for music with drum-sounds. We therefore propose a method for judging whether drums are currently being played in the input audio signal. Since the detection of the drum-sounds is noisy, the system cannot simply use the detected results for this judgement. According to the fact that in popular music a snare drum is usually played on the second and fourth quarter notes in a measure, the method judges that the input audio signal includes drum-sounds only when autocorrelation of the snare drum’s onset times is high enough.

3.3 Dealing with Ambiguity of Interpretation

To handle ambiguous situations in interpreting the beat-tracking cues, we introduced a multiple-agent model in which multiple agents examine various hypotheses of the beat structure in parallel [Goto and Muraoka, 1996]. Each agent makes a hypothesis by using the inverse model described in 3.2 according to different strategy. An agent interacts with another agent to track beats cooperatively and adapts to the current situation by adjusting its strategy. It then evaluates the reliability of its own hypothesis according to how well the inverse model can be applied. The final beat-tracking result is determined on the basis of the most reliable hypothesis.

3.4 System Overview

Figure 4 shows the processing model of the system. In the frequency-analysis stage, the system detects the onset-time vectors, detects onset times of bass drum and snare drum sounds, and judges the presence of the drum-sounds. In the beat-prediction stage, each agent infers the quarter-note level by using autocorrelation and cross-correlation of the onset-time vectors. A higher-level checker corresponding to each agent then detects chord changes and drum patterns by using the quarter-note level as the top-down information. Using those detected results, each agent infers the higher levels and evaluates the reliability of its own hypothesis. Finally, the beat-tracking result is transmitted to other application programs via a computer network by using the network protocol *RMCP* [Goto et al., 1997], which is designed for sharing musical information.

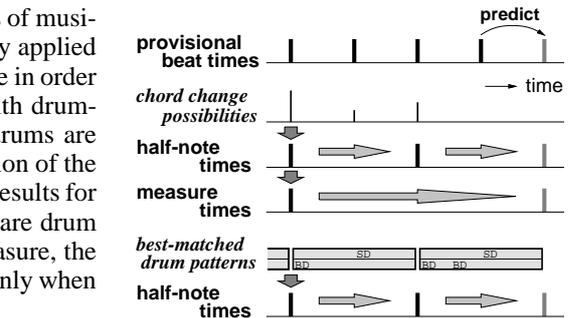


Figure 3: Knowledge-based inferring.

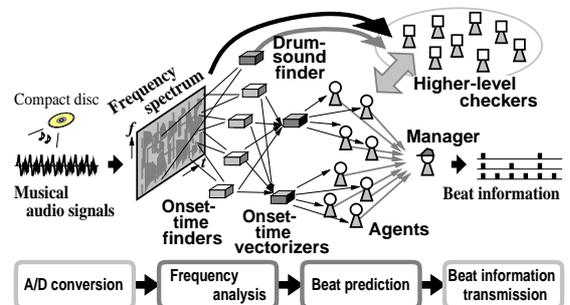


Figure 4: Overview of our beat-tracking system.

4 Experiments and Results

We tested the system implemented on a distributed-memory parallel computer, the Fujitsu AP1000. The input monaural audio signals were sampled from commercial compact discs of popular music. Eighty-five songs, each at least one minute long, 45 with drum-sounds (32 artists, tempo range: 67-185 M.M.) and 40 without drum-sounds (28 artists, tempo range: 62-116 M.M.), were used as the inputs. Each song had the 4/4 time-signature and a roughly constant tempo.

In our experiment using the quantitative evaluation measures for analyzing the beat-tracking accuracy that were proposed in [Goto and Muraoka, 1997a], the recognition rates at each level of the beat structure were more than 86.7 percent (Table 1). This result shows that the system is robust enough to deal in real time with real-world audio signals containing sounds of various instruments.²

5 Applications

Since beat tracking can be utilized to automate the time-consuming tasks that must be completed in order to synchronize events with music, it is useful in various applications, such as video editing, audio editing, and human-computer improvisation. In the following, we introduce three applications based on our beat-tracking system.

- Beat-driven real-time computer graphics

Our system makes it easy to create real-time computer graphics synchronized with music. In fact, we have developed a system that displays virtual dancers and several graphic objects whose motions and positions change in time to beats (Figure 5). This system has several dance sequences, each for a different mood of dance motions.

²Further evaluations described in [Goto, 1998] showed the validity of the three kinds of musical knowledge used in the inverse model.

Table 1: Recognition rates.

Beat structure	Without drums	With drums
Measure level	32 / 34 songs (94.1 %)	34 / 39 songs (87.2 %)
Half-note level	34 / 35 songs (97.1 %)	39 / 39 songs (100 %)
Quarter-note level	35 / 40 songs (87.5 %)	39 / 45 songs (86.7 %)

While a user manually selects a dance sequence, the timing of each motion in the selected sequence is automatically determined on the basis of the beat-tracking results. Such a computer graphics system is suitable for live stage, TV program, and *Karaoke* uses.

- Stage-lighting control

Beat tracking facilitates the automatic synchronization of computer-controlled stage lighting with beats of a musical performance. Various properties of lighting, such as color, brightness, and direction, can be changed in time to music. At the moment this application is simulated on a computer graphics display with the virtual dancers.

- Intelligent drum-machine

We have implemented a preliminary system that can play drum patterns in time to input musical audio signals without drum-sounds. This application has the potential for automatic MIDI-audio synchronization and intelligent computer accompaniment.

We have also developed a beat-structure editor program that enables a user to correct the system output for practical uses. While listening to an audio signal and watching its waveform, a user can simply adjust the beat structure or can make the whole hierarchical beat structure corresponding to the audio signal from scratch. The positions can be finely adjusted by playing back the audio with click tones at beat times.

6 Conclusion

We have described the beat-tracking problem in dealing with real-world audio signals, our beat-tracking model, and three applications based on our real-time beat-tracking system. Since our system employs a higher-level processing using three kinds of musical knowledge and supported by sophisticated frequency analysis processes based on top-down information, it can make musical decisions about complex audio signals and infer the hierarchical beat structure. Experimental results show that our system can recognize the quarter-note, half-note, and measure levels in compact discs' audio signals. We have also confirmed that the system is effective in practical applications.

We plan to upgrade the system to follow tempo changes and to generalize to other musical genres. Future work will include its application to various interactive-media systems for which beat tracking is useful.

References

(Our papers are available on "<http://www.etl.go.jp/~goto/>".)

- [Allen and Dannenberg, 1990] Allen, P. E. and Dannenberg, R. B. Tracking musical beats in real time. In *ICMC 1990*, pp. 140–143, 1990.
- [Dannenberg and Mont-Reynaud, 1987] Dannenberg, R. B. and Mont-Reynaud, B. Following an improvisation in real time. In *ICMC 1987*, pp. 241–248, 1987.
- [Desain and Honing, 1989] Desain, P. and Honing, H. The quantization of musical time: A connectionist approach. *CMJ*, 13(3):56–66, 1989.
- [Desain and Honing, 1994] Desain, P. and Honing, H. Advanced issues in beat induction modeling: syncopation, tempo and timing. In *ICMC 1994*, pp. 92–94, 1994.
- [Driesse, 1991] Driesse, A. Real-time tempo tracking using rules to analyze rhythmic qualities. In *ICMC 1991*, pp. 578–581, 1991.
- [Goto and Muraoka, 1994] Goto, M. and Muraoka, Y. A beat tracking system for acoustic signals of music. In *ACM Multimedia 94*, pp. 365–372, 1994.
- [Goto and Muraoka, 1995] Goto, M. and Muraoka, Y. A real-time beat tracking system for audio signals. In *ICMC 1995*, pp. 171–174, 1995.
- [Goto and Muraoka, 1996] Goto, M. and Muraoka, Y. Beat tracking based on multiple-agent architecture — a real-time beat tracking system for audio signals —. In *ICMAS-96*, pp. 103–110, 1996.
- [Goto and Muraoka, 1997a] Goto, M. and Muraoka, Y. Issues in evaluating beat tracking systems. In *IJCAI-97 Workshop on Issues in AI and Music*, pp. 9–16, 1997.
- [Goto and Muraoka, 1997b] Goto, M. and Muraoka, Y. Real-time rhythm tracking for drumless audio signals — chord change detection for musical decisions —. In *IJCAI-97 Workshop on Computational Auditory Scene Analysis*, pp. 135–144, 1997.
- [Goto and Muraoka, 1998] Goto, M. and Muraoka, Y. Music understanding at the beat level — real-time beat tracking for audio signals —. In Rosenthal, D. F. and Okuno, H. G., editors, *Computational Auditory Scene Analysis*, pp. 157–176. Lawrence Erlbaum Associates, 1998.
- [Goto et al., 1997] Goto, M., Neyama, R., and Muraoka, Y. RMCP: Remote music control protocol — design and applications —. In *ICMC 1997*, pp. 446–449, 1997.
- [Goto, 1998] Goto, M. *A Study of Real-time Beat Tracking for Musical Audio Signals* (in Japanese). PhD thesis, Waseda Univ., 1998.
- [Katayose et al., 1989] Katayose, H. et al. An approach to an artificial music expert. In *ICMC 1989*, pp. 139–146, 1989.
- [Kitano, 1993] Kitano, H. Challenges of massive parallelism. In *IJCAI-93*, pp. 813–834, 1993.
- [Large, 1995] Large, E. W. Beat tracking with a nonlinear oscillator. In *IJCAI-95 Workshop on AI and Music*, pp. 24–31, 1995.
- [Rosenthal, 1992a] Rosenthal, D. Emulation of human rhythm perception. *CMJ*, 16(1):64–76, 1992.
- [Rosenthal, 1992b] Rosenthal, D. *Machine Rhythm: Computer Emulation of Human Rhythm Perception*. PhD thesis, MIT, 1992.
- [Rowe, 1993] Rowe, R. *Interactive Music Systems*. The MIT Press, 1993.
- [Scheirer, 1998] Scheirer, E. D. Tempo and beat analysis of acoustic musical signals. *JASA*, 103(1):588–601, 1998.
- [Schloss, 1985] Schloss, W. A. *On The Automatic Transcription of Percussive Music — From Acoustic Signal to High-Level Analysis*. PhD thesis, CCRMA, Stanford Univ., 1985.
- [Todd and Lee, 1994] Todd, N. and Lee, C. An auditory-motor model of beat induction. In *ICMC 1994*, pp. 88–89, 1994.
- [Todd, 1994] Todd, N. P. M. The auditory “primal sketch”: A multiscale model of rhythmic grouping. *JNMR*, 23(1):25–70, 1994.
- [Vercoe, 1994] Vercoe, B. Perceptually-based music pattern recognition and response. In *ICMPC 94*, pp. 59–60, 1994.

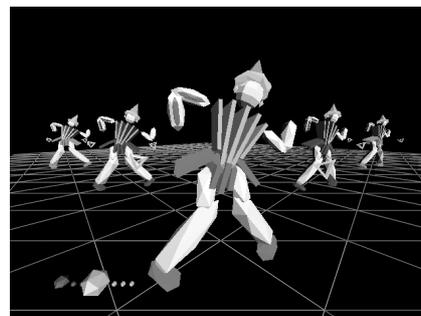


Figure 5: Virtual dancer “Cindy.”