Issues in Evaluating Beat Tracking Systems

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

Abstract

This paper discusses the main issues in evaluating beat tracking systems and proposes a method of evaluating the accuracy of these systems' output. Although there have been a few attempts to evaluate beat tracking systems, they have not sufficiently addressed quantitative evaluation of tracking accuracy, in particular for audio signals. Our method compares subjective hand-labeled beat positions with computer-parsed beat positions and enables us to quantitatively analyze the deviation error of the tracked rhythm (beats) not only at the quarter-note level but also at the half-note and measure levels. It also considers several typical errors such as half-tempo and double-tempo errors as well as correct parsing. In testing our beat tracking system for audio signals without drum-sounds on 40 popular songs sampled from compact discs, we used the proposed method to evaluate overall recognition rates and tracking quickness and accuracy.

1 Introduction

A great deal of research related to beat tracking has been undertaken in recent years [Schloss, 1985; Dannenberg and Mont-Reynaud, 1987; Desain and Honing, 1989; 1994; 1995; Desain, 1992; Katayose *et al.*, 1989; Allen and Dannenberg, 1990; Driesse, 1991; Rosenthal, 1992a; 1992b; Rowe, 1993; Vercoe, 1994; Todd, 1994; Todd and Lee, 1994; Large, 1995; Smith, 1996; Scheirer, 1996; Goto and Abe, 1996; Goto and Muraoka, 1994; 1995a; 1995b; 1996; 1997] because beat tracking is an important initial step in the computational modeling of music understanding. It is quite difficult to understand Western music without perceiving beats, since the beat is a fundamental unit of the temporal structure of music. Furthermore, beat tracking — also called *beat induction*, *foot-tapping*, or *rhythm-tracking* — has applications in such various fields as human-computer improvisation, video/audio editing, stage lighting control, and the synchronization of computer graphics with music.

Since various studies have addressed various issues, purposes, and applications, it is generally difficult to comparatively evaluate previous beat tracking models. There have, however, been a few discussions of evaluation issues. Desain and Honing [1995] described the evaluation of three different rule-based beat-induction models from various points of view. Goto and Abe [1996] also examined several cognitive models of rhythmic interpretation from the psychological viewpoint. Rosenthal [1992a] pointed out that evaluating the ideas behind many of previous systems is not a straightforward task and that different researchers have worked on different aspects of the problem. Because most studies have dealt with MIDI signals or used onset times as their input, the evaluation of audio-based beat tracking has not been discussed enough.

To take the first step toward evaluating beat tracking systems, we propose a method of measuring the accuracy of beat tracking results that enables those results to be evaluated quantitatively. This method compares subjective hand-labeled beat positions with computer-parsed beat positions while considering different rhythmic levels and several typical errors as well as the correct parsing. For example, it evaluates not only the quarter-note level but also the half-note and measure (bar) levels and also analyzes half-tempo and double-tempo errors.

In the following sections we mainly describe issues in the evaluation of beat tracking systems and introduce a measure of beat-tracking accuracy. Although the proposed evaluation method has been applied only to the output of our audio-based beat-tracking system, we think that it could also be applied to other systems for MIDI signals or onset times.

2 Evaluation Issues

Since beat tracking systems can have a variety of inputs, purposes, levels of rhythmic structure, and so on, various issues should be taken into consideration when evaluating these systems. This section deals with some of the important issues.

2.1 Inputs

The input sources of beat tracking systems can be classified into three categories: onset times (a sequence of inter-onset intervals), MIDI signals (a standard MIDI file or a real-time MIDI input), and audio signals. Within each category there are the score-like uninflected inputs, the real performances with roughly constant tempo, and the expressive performances with tempo changes. In terms of musical genres, previous systems have dealt with classical works, folksongs, national anthems, popular music, rock music, and so on.

This variety causes researchers to address various aspects of the beat-tracking problem. That research all has its own value; however it is generally difficult to compare the results of different systems and to integrate the ideas behind them. A first step toward this comparison would be a quantitative evaluation of the results of each system.

2.2 Purposes

There are various purposes for tracking beats: understanding human rhythm perception, building a computational model understanding rhythm, following live performances, beat-quantization for automatic transcription, and synchronization of computer graphics with music. Beat-tracking researchers have accordingly specified their problems in different ways: finding beat positions in onset times, following tempo changes of MIDI signals, parsing the rhythmic structure of MIDI signals, and understanding the rhythmic structure of audio signals.

The criteria used to evaluate beat-tracking accuracy depend on the purpose. For example, when the main research purpose is to analyze tempo changes and expressiveness in human performances, the criteria should be strict and the time-resolution should be fine, since the timing of beats is important. On the other hand, when the main application is music-synchronized computer graphics, the time-resolution of beat times is resampled at the frame-rate resolution (usually 30 frames per second) in the output images. In such a case, the stability of tracking is more important than the precision of the beat times. In the case of beat tracking for audio/video editing systems, even beat-tracking results that include typical errors such as half-tempo and double-tempo errors might be useful since human correction can be expected. An evaluation method whose measurement results can be considered with regard to a variety of criteria is hence preferable.

Although real-timeness is not directly related to the evaluation of beat-tracking accuracy, it is an important issue for several applications. Even if the preliminary implementations of a system do not work in real time, it is often preferable to process the information sequentially because such processing will facilitate future real-time implementations.

2.3 Rhythmic Structure

In this paper the rhythmic structure is defined as a hierarchical musical structure with several levels of rhythm: *the quarter-note level*, *the half-note level*, *the measure level*, and so on (Figure 1). The beginnings of a quarter note, a half note, and a measure are respectively called a *beat time*, a *half-note time*, and a *measure time*. For the sake of brevity, we consider only the 4/4 time-signature in this paper.

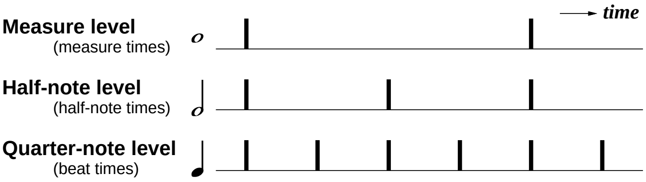


Figure 1: Rhythmic structure.

Most previous systems dealt with only a single level — usually the quarter-note level — which is the level of regular isochronal pulses that a person feels in music. One system [Goto and Muraoka, 1994; 1995a; 1995b; 1996] also tracked beats at the half-note level, and other systems [Rosenthal, 1992a; 1992b; Goto and Muraoka, 1997] tracked beats at the measure level as well as the lower levels. In evaluating those systems, it is necessary to take these rhythmic levels into account.

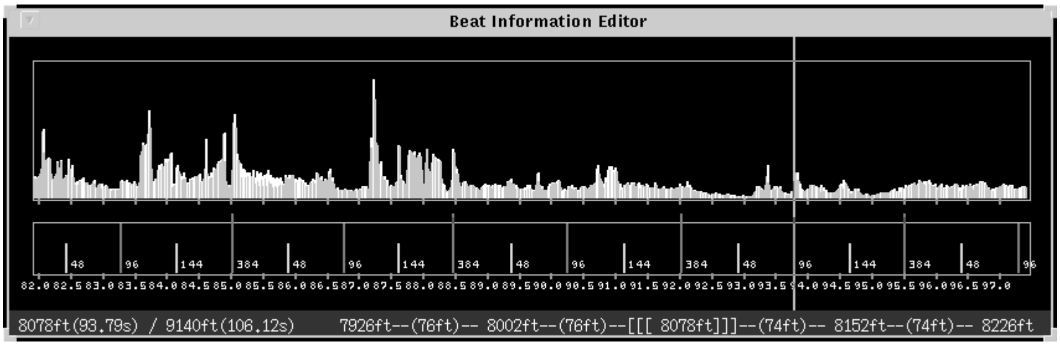
3 Quantitative Measure of Accuracy

As an initial step toward the complete evaluation of beat tracking systems, we propose a measure that can be utilized to analyze how accurate the tracked beat times are. Although this is not a perfect measure considering all evaluation issues, it is useful for analyzing the accuracy of beat-tracking results at several rhythmic levels and identifying typical errors. Because we have been working on a real-time beat-tracking system for audio signals [Goto and Muraoka, 1994; 1995a; 1995b; 1996; 1997], in this section we concentrate on describing the measure for evaluating audio-based beat-tracking systems.

3.1 Beat Correctness

Since the beat is a perceptual concept that a person feels in music, it is generally difficult to define the correct beat in an objective way. We think that one of goals of beat-tracking research is to establish a precise definition of the “beat” concept and to verify this definition using an actual system.

In designing our measure for evaluation, we considered subjective hand-labeled beat positions to be the correct



```
8078ft(93.79s) / 9140ft(106.12s)      7926ft--(76ft)-- 8002ft--(76ft)--[[[ 8078ft]]]--(74ft)-- 8152ft--(74ft)-- 8226ft
```

Figure 2: Beat-position editor program.

beat times. To label these positions, we developed a beat-position editor program that enables a user to mark the beat positions in a digitized audio signal while listening to the audio and watching its waveform (Figure 2). The positions can be finely adjusted by playing back the audio with click tones at beat times, and the user also defines a hierarchical rhythmic structure — the quarter-note, half-note, and measure levels — corresponding to the audio signal.

3.2 Designing Measure

The basic concept of our measure of beat-tracking accuracy is to compare the hand-labeled beat times (*the correct times*) with the beat times of the system output (*the examined times*)[1]. In other words, we compare the system's parsing with our own subjective parsing.

Basic Measure

We first define the basic measure for analyzing the beat-tracking accuracies at the quarter-note, half-note, and measure levels. These accuracies are each represented as a measurement set {Q, H, M}$[\tau, \mu, \sigma, M]$, where Q[], H[], and M[] respectively represent the measures at the quarter-note, half-note, and measure levels. The term τ is the correctly-tracked period (the period in which the rhythm is tracked correctly), and the terms μ, σ, and M are respectively the mean, standard deviation, and maximum of difference between the correct time and the examined time.

The measurement set at each level is calculated as follows:

1. *Pairing the correct times with the examined times*

   Let C_n be the n-th correct time and B_m be the m-th examined time. We first find the B_m which fulfills Condition (1) and is nearest to C_n in order to make a pair $[C_n, B_m]$:

   $$C_n - \frac{1}{2}\,I_{n-1} \leq B_m < C_n + \frac{1}{2}\,I_n \qquad (1)$$

   $$I_n = C_{n+1} - C_n \qquad (2)$$

   where I_n is the n-th inter-beat interval (the temporal difference between two successive beats). If there is no B_m which fulfills Condition (1), C_n is labeled as *unpaired*. An examined time B_m that is not paired with a C_n is also labeled as *unpaired*.

2. *Evaluating the error in each pair*

   The normalized difference (deviation error) P_n between C_n and B_m in each pair (Figure 3) is given by

   $$P_n = \begin{cases} \frac{|B_m - C_n|}{i_n} & (C_n \text{ is paired}) \\ 1 & (C_n \text{ is unpaired}) \end{cases} \qquad (3)$$

   $$i_n = \begin{cases} I_n\,/\,2 & (B_m \geq C_n) \\ I_{n-1}\,/\,2 & (B_m < C_n) \end{cases} \qquad (4)$$

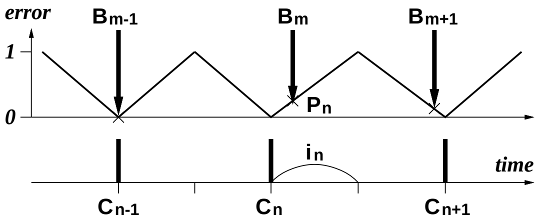


Figure 3: Evaluating the error.

3. *Finding the longest correctly tracked period*

   The longest correctly tracked period $\tau = [C_{Ns}, C_{Ne}]$ is then found in C_n according to P_n, where Ns and Ne are respectively the beginning and end suffix of C_n during the period. The correctly tracked period should fulfill two conditions: every P_n during the period must be $P_n < \Phi$ where Φ (equal to 0.35 in our evaluation) is a constant threshold, and there must be no unpaired B_m during the period. In particular, $[C_{Ns}, -]$ means that a beat tracking system keeps on tracking the correct rhythm if once it starts to track the correct one at C_{Ns}, and [!, !] means that a system cannot track the correct rhythm at all.

---

[1] In evaluating the quarter-note, half-note, and measure levels, we use the beat, half-note, and measure times respectively.

4. *Calculating the undecided terms of the measurement set*
   Finally, the mean μ, standard deviation σ, and maximum M of P_n during the period τ are calculated.

Advanced Measure

We upgrade the basic measure to an advanced measure that considers the typical errors: a half-tempo or double-tempo error and a π*-phase error*. The half-tempo and double-tempo errors are respectively those in which the tracked inter-beat interval is twice or half the correct one. The π-phase error is the half-period displacement error in which the difference between the correct times and the examined times is half of the interval between two successive correct times. These kinds of errors should to some extent be considered to be appropriate because such rhythm-parsings are better than a random parsing.

The advanced measure of each rhythmic level is represented as a measurement set $\{\mathrm{Q, H, M}\}[\tau, \mu, \sigma, M, f_T, f_p]$ based on the basic measure, where f_T is a flag indicating the tempo error and f_p is a flag indicating the phase error. The flag f_T takes a value of $-$ (the correct tempo), *hlf* (the half-tempo error), or *dbl* (the double-tempo error), and the flag f_p takes a value of either 0 (the correct phase) or π (the π-phase error). The f_T of all the measurement sets $\{\mathrm{Q, H, M}\}[]$ takes the same value.

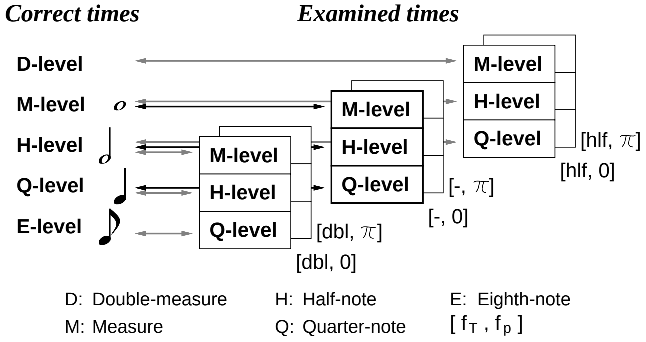


Figure 4: Advanced measure.

To determine f_T, we first evaluate six sets of the basic measure Q[] between the examined times at the quarter-note level and the six kinds of correct times at the three rhythmic levels with two phase variations: the eighth-note level with the 0- or π-phase (0-phase means the correct phase), the quarter-note level with the 0- or π-phase, and the half-note level with the 0- or π-phase (Figure 4). We then find the best-matched measure Q[] with the longest correctly tracked period τ. The flag f_T is thus determined on the basis of the level with its best measure. The higher-level measurement sets {H, M}[] are calculated while considering the level of the determined Q[] to be the lowest (Figure 4).

To determine f_p of each measurement set, we then calculate two sets of the basic measure at each level between the examined times and the two kinds of correct times at the corresponding level with the two phase variations. The flag f_p is finally determined according to the phase with the longer correctly tracked period τ. For example, if the basic measure Q[] between the examined times at the quarter-note level and the correct times at the eighth-note level with the π-phase has the longest τ, we determine $\mathrm{Q}[\tau, \mu, \sigma, M, \mathit{dbl}, \pi]$.

This advanced measure is useful for analyzing and identifying the different types of typical errors. This measure also considers that one of disputable points of human rhythm parsing is the half-tempo or double-tempo error.

4 Experimental Results

With the advanced measure we evaluated the output of our beat tracking system for monaural musical audio signals without drum-sounds (a bass drum or a snare drum). In this section we first introduce an overview of the target system and then describe recognition rates of songs, how quickly the system started to track the correct rhythm, and how accurately the system obtained the beat, half-note, and measure times.

4.1 Target System for Evaluation

As mentioned above, the target system was our real-time beat-tracking system for audio signals without drum-sounds[2] presented in [Goto and Muraoka, 1997]. This system is an upgraded version of the system presented in [Goto and Muraoka, 1996]: this system can track beats at the quarter-note, half-note, and measure levels. The system assumes that the time signature of an input song is 4/4 and that its tempo is constrained to be between 61 M.M. (Mälzel's Metronome: the number of quarter notes per minute) and 120 M.M., and is roughly constant. Since the system digitizes the input at 22050 Hz and the FFT window of the input is shifted by 256 samples, the time resolution of our measurement is 11.61 ms (= 256 / 22050 Hz).

The system is based on multiple-agent architecture in which multiple beat-position hypotheses are maintained by programmatic agents using different strategies for beat tracking. This architecture enables the system to cope with difficult beat-tracking situations: even if some agents lose track of beats, the system will track beats correctly as long as other agents maintain the correct hypothesis. Each agent predicts beat times using autocorrelation and cross-correlation of the onset times in seven different frequency ranges. The agent then recognizes the higher-level rhythmic structure and evaluates the reliability of its own hypothesis. The final output is

---

[2] We have also developed a real-time beat-tracking system for audio signals that include drum-sounds [Goto and Muraoka, 1994; 1995a; 1995b].

determined on the basis of the most reliable hypothesis out of all these agent-generated hypotheses.

To make context-dependent decisions such as those needed for recognizing the higher-level rhythmic structure and for determining which is the best hypothesis in an ambiguous situation, higher-level processing using musical knowledge is indispensable. The system therefore detects chord changes without chord name symbolization and utilizes heuristic musical knowledge of chord changes. For example, by using its knowledge that chords are more likely to change at the beginning of measures than at other positions, the system can determine the higher-level structure.

4.2 Overall Result

We tested the target system on 40 songs, each at least about one minute long, performed by 28 artists. The input audio signals were sampled from commercial compact discs of popular music and contained the sounds of various instruments (not including drums). The time-signature was 4/4 and the tempi ranged from 62 M.M. to 116 M.M. and were roughly constant.

We judged that a song was tracked correctly at a certain rhythmic level if the corresponding advanced measurement set of the song fulfilled {Q, H, M}[[$C_{Ns} < 45.0$ sec, $C_{Ne} = -$],[3] $\mu < 0.2$, $\sigma < 0.2$, $M < 0.35$, $f_T = -$, $f_p = 0$]. When the beat-tracking result for a song was evaluated as Q[[10.84 sec, −], 0.072, 0.059, 0.243, −, 0], H[[12.65 sec, −], 0.037, 0.029, 0.120, −, 0], and M[[25.05 sec, −], 0.019, 0.015, 0.046, −, 0], for example, we could judge that this song was tracked correctly at the three rhythmic levels, that the deviation error was quite small on the average, and that the maximum error was the thirty-second-note period at most.

In our experiment the system correctly tracked beats at the quarter-note level in 35 of the 40 songs (87.5 %) and correctly recognized the half-note level in 34 of the 35 songs (97.1 %) in which the correct beat times were obtained. Moreover, it correctly recognized the measure level in 32 of the 34 songs (94.1 %) in which the correct half-note times were obtained.

We analyzed tracking errors at each of the three rhythmic levels using the advanced measurement sets. The system did not track the correct beats at the quarter-note level in five songs. In one of the five songs, the system made the π-phase error (f_p of Q[] $= \pi$) during most of the song (τ = [11.45 sec, −]). In the other four songs, although the system temporarily tracked the correct beats, it often tracked wrong phases or determined wrong inter-beat intervals. In the case of one song which was incorrectly tracked at the half-note level, the system had made the π-phase error (f_p of H[] $= \pi$) just after starting to track the correct beat times. There were two songs in which the system was unable to track only the measure level. In those songs, the system also made the π-phase error (f_p of M[] $= \pi$).

Figure 5 shows the histogram of the correct tempo of all the input songs. It also indicates the five songs which were incorrectly tracked at the quarter-note level and the three songs which were incorrectly tracked at the half-note or measure level.

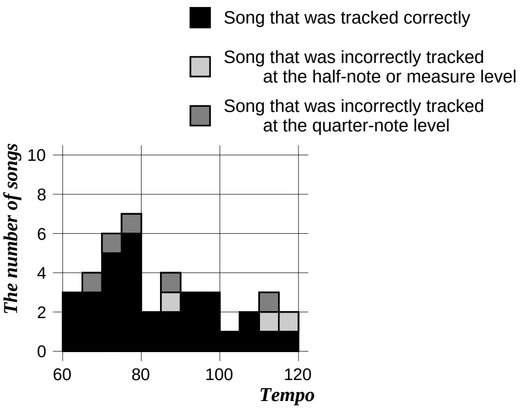


Figure 5: Histogram of the tempo of all the songs.

4.3 Tracking Quickness

We can use the C_{Ns} of each measurement set to evaluate tracking quickness — how quickly the system started to track the correct rhythm — at each level. Figure 6 shows the C_{Ns} of the correctly-tracked songs at the quarter-note, half-note, and measure levels. The horizontal axis represents the song numbers (#) arranged in order of

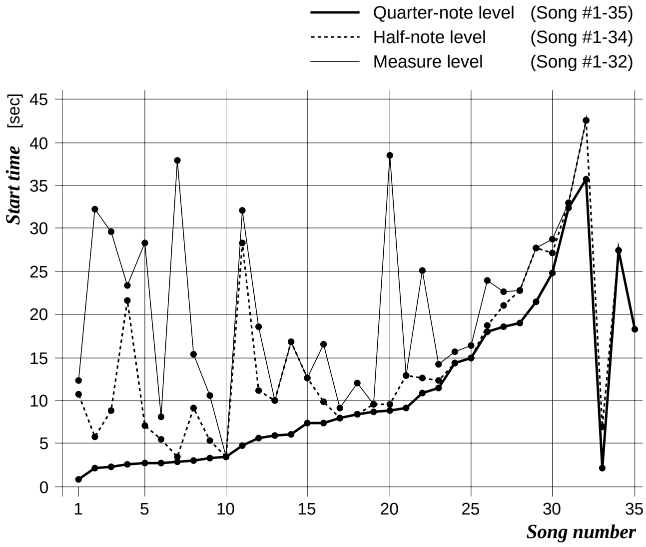


Figure 6: Start time (C_{Ns}) of tracking the correct rhythm at the quarter-note, half-note, and measure levels.

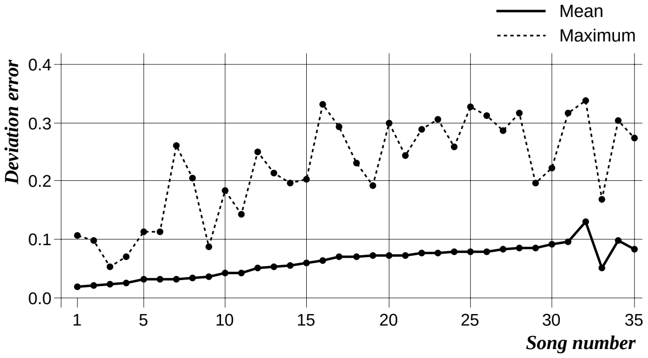
---

[3] The first correct time C_1 is considered 0 sec.

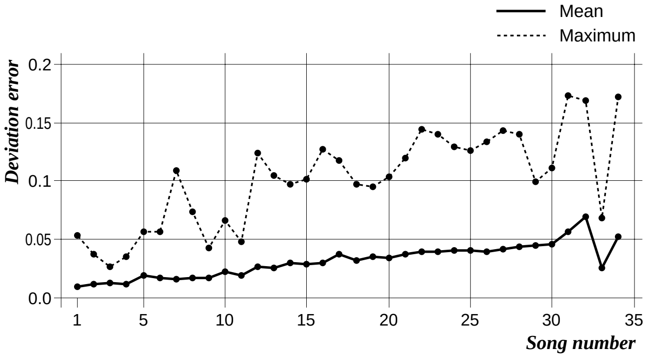
C_{Ns} of the quarter-note level up to song #32. In each song where the rhythmic structure was eventually determined correctly, the system initially had trouble determining the higher rhythmic level, even though the lower level was correct. The mean, minimum, and maximum of the C_{Ns} of all the correctly-tracked songs are shown in Table 1.

Table 1: Mean, minimum, and maximum of C_{Ns} at the quarter-note, half-note, and measure levels.

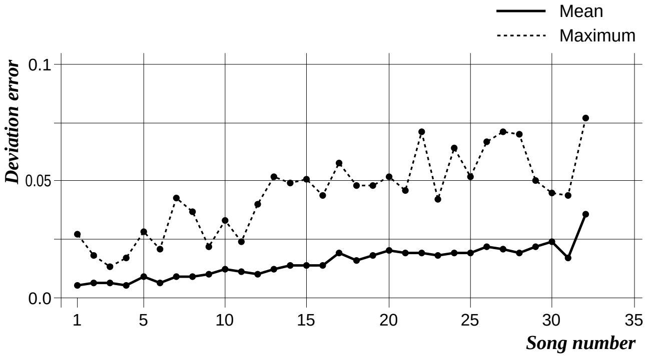
rhythmic level	mean	min	max
Quarter-note level	10.71 sec	0.79 sec	35.77 sec
Half-note level	14.70 sec	3.42 sec	42.56 sec
Measure level	20.70 sec	3.42 sec	42.56 sec



(a) Quarter-note level (Song #1-35)



(b) Half-note level (Song #1-34)



(c) Measure level (Song #1-32)

Figure 7: Mean (μ) and maximum (M) of the deviation error at the quarter-note, half-note, and measure levels.

4.4 Tracking Accuracy

We can use the μ, σ, and M of each measurement set to evaluate tracking accuracy — how accurate the examined times (the system output) were — at each level. Figure 7 shows the μ and M of the correctly-tracked songs at the three rhythmic levels. The horizontal axis represents the song numbers rearranged in order of μ of the quarter-note level up to song #32. The mean, minimum, and maximum of μ and M of all the correctly-tracked songs are shown in Table 2 and Table 3, respectively. The maximum deviation error was 0.339 and the mean of the error was relatively small on the average.

Table 2: Mean, minimum, and maximum of μ at the quarter-note, half-note, and measure levels.

rhythmic level	mean	min	max
Quarter-note level	0.062	0.019	0.130
Half-note level	0.031	0.009	0.069
Measure level	0.015	0.005	0.036

Table 3: Mean, minimum, and maximum of M at the quarter-note, half-note, and measure levels.

rhythmic level	mean	min	max
Quarter-note level	0.223	0.053	0.339
Half-note level	0.101	0.026	0.174
Measure level	0.045	0.013	0.077

5 Measuring Rhythmic Difficulty

Although the proposed measuring method enables us to quantitatively analyze the output of beat tracking systems, there are lots of other evaluation issues that we have not discussed so far. It is, for example, important but very difficult to measure the rhythmic difficulty of a song from the viewpoint of the input of beat tracking. For the input we prefer realistic songs such as ones sampled from commercial compact discs rather than artificial songs generated just for experiments. On the other hand, one reasonable solution might be to carefully compose songs with different levels of difficulty.

To take the first step toward measuring the rhythmic difficulty of realistic songs, we tried to evaluate only the power transition of the input audio signals, although it is desirable to consider various aspects of songs such as tempo changes, rhythmic patterns (complexity), time signatures, musical instruments, and musical genres. In terms of the power transition of the audio signals, it is more difficult to track beats of a song in which the power on beats is often less than on other positions between adjacent beats. In other words, the larger the number of syncopations, the greater the difficulty of tracking beats.

We therefore consider differences between the power on beats and the power on other positions as a measure of the rhythmic difficulty. This measure, called the *power-difference measure*, is calculated as follows:

1. *Finding the local maximum of the power*

   Let $a(t)$ be the power of the input audio signal at time t,[4] which is the sum of the absolute value of the digitized audio data during the corresponding period. We first calculate two kinds of local maximum of the power, L_n^b and L_n^o:

$$L_n^b = \max_{C_n - \epsilon \;\le\; t \;<\; C_n + I_n/4 - \epsilon} (a(t)) \tag{5}$$

$$L_n^o = \max_{C_n + I_n/4 - \epsilon \;\le\; t \;< C_{n+1} - \epsilon} (a(t)) \tag{6}$$

   where C_n is the n-th correct beat time of the quarter-note level and I_n is the n-th inter-beat interval (Figure 8). ϵ is a margin to obtain appropriate maximum values because the power of sounds played on the beat sometimes increases just before its beat time. In our current implementation, ϵ is equal to 2 (23.22 ms). Thus the L_n^b represents the maximum power on the n-th beat and the L_n^o represents the maximum power on the other positions between the n-th beat and $(n+1)$-th beat.

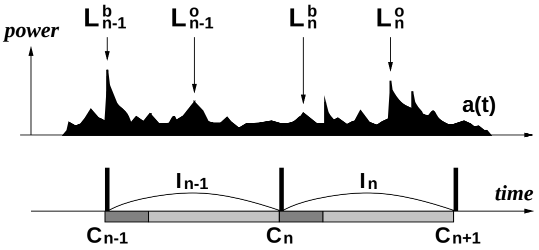


Figure 8: Finding the local maximum of the power.

2. *Calculating the normalized power difference of each beat*

   The normalized power difference d_n between L_n^o and L_n^b is then calculated:

$$d_n = 0.5 \, \frac{L_n^o - L_n^b}{\max\,(L_n^o, L_n^b)} + 0.5 \tag{7}$$

3. *Calculating the mean of the normalized power differences in a song*

   We finally calculate the power-difference measure D, which is the mean of all the d_n in the song:

$$D = \frac{1}{\mathrm{N}-1} \sum_{n=1}^{\mathrm{N}-1} d_n \tag{8}$$

   where N is the number of the correct beat times. The D takes a value between 0.0 (easiest) and 1.0 (most difficult). A regular pulse sequence with a constant interval, for example, takes 0.0 of this measure. Practically speaking, the D of a realistic song cannot take 1.0.

With the power-difference measure D, we evaluated the rhythmic difficulty of all 40 songs that we used for testing the beat tracking system described in Section 4. Figure 9 shows the histogram of D of all the songs. It also indicates the five songs which were incorrectly tracked at the quarter-note level and the three songs which were incorrectly tracked at the half-note or measure level. Although this measure is not perfect for evaluating the rhythmic difficulty, it showed a tendency to agree with our subjective ordering of the rhythmic difficulty of the songs.

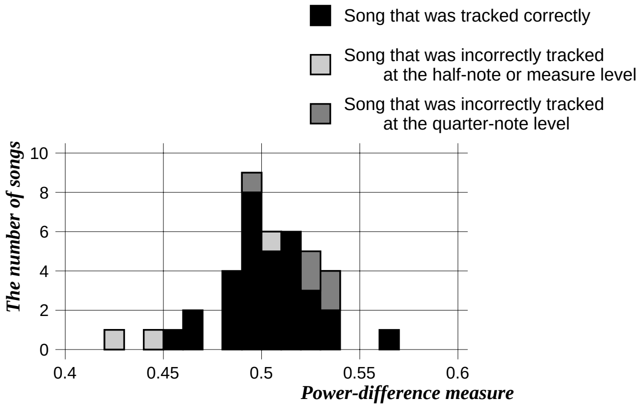


Figure 9: Histogram of the power-difference measure of all the songs.

6 Conclusion

We have discussed the main issues in evaluating beat tracking systems and have addressed several of them by proposing a quantitative measure that analyzes the accuracy of a system's rhythm-parsing. Although this is not perfect measure addressing all evaluation issues, it can analyze the accuracy of the rhythm tracking while considering a hierarchical rhythmic structure — the quarter-note, half-note, and measure levels — and several typical errors such as the half-tempo and double-tempo errors and the π-phase error. In our experiment the proposed measure was applied and found to be useful for evaluating the recognition rates and the tracking quickness and accuracy of our audio-based beat-tracking system.

---

[4] The time resolution of this measurement is 11.61 ms (= 256 / 22050 Hz) in our current implementation.

We plan to address other evaluation issues and to consider an advanced framework of beat-tracking research evaluation. Future work will also include upgrading our beat tracking system so that it can understand the rhythmic structure of music in a human-like fashion.

References

[Allen and Dannenberg, 1990] Paul E. Allen and Roger B. Dannenberg. Tracking musical beats in real time. In *Proc. of the 1990 Intl. Computer Music Conf.*, pages 140–143, 1990.

[Dannenberg and Mont-Reynaud, 1987] Roger B. Dannenberg and Bernard Mont-Reynaud. Following an improvisation in real time. In *Proc. of the 1987 Intl. Computer Music Conf.*, pages 241–248, 1987.

[Desain and Honing, 1989] Peter Desain and Henkjan Honing. The quantization of musical time: A connectionist approach. *Computer Music Journal*, 13(3):56–66, 1989.

[Desain and Honing, 1994] Peter Desain and Henkjan Honing. Advanced issues in beat induction modeling: syncopation, tempo and timing. In *Proc. of the 1994 Intl. Computer Music Conf.*, pages 92–94, 1994.

[Desain and Honing, 1995] Peter Desain and Henkjan Honing. Computational models of beat induction: the rule-based approach. In *Working Notes of the IJCAI-95 Workshop on Artificial Intelligence and Music*, pages 1–10, 1995.

[Desain, 1992] Peter Desain. Can computer music benefit from cognitive models of rhythm perception? In *Proc. of the 1992 Intl. Computer Music Conf.*, pages 42–45, 1992.

[Driesse, 1991] Anthonie Driesse. Real-time tempo tracking using rules to analyze rhythmic qualities. In *Proc. of the 1991 Intl. Computer Music Conf.*, pages 578–581, 1991.

[Goto and Abe, 1996] Yasuhiro Goto and Junichi Abe. On cognitive models of rhythmic interpretation. *International Journal of Psychology*, 31:51, 1996.

[Goto and Muraoka, 1994] Masataka Goto and Yoichi Muraoka. A beat tracking system for acoustic signals of music. In *Proc. of the Second ACM Intl. Conf. on Multimedia*, pages 365–372, 1994.

[Goto and Muraoka, 1995a] Masataka Goto and Yoichi Muraoka. Music understanding at the beat level — real-time beat tracking for audio signals —. In *Working Notes of the IJCAI-95 Workshop on Computational Auditory Scene Analysis*, pages 68–75, 1995.

[Goto and Muraoka, 1995b] Masataka Goto and Yoichi Muraoka. A real-time beat tracking system for audio signals. In *Proc. of the 1995 Intl. Computer Music Conf.*, pages 171–174, 1995.

[Goto and Muraoka, 1996] Masataka Goto and Yoichi Muraoka. Beat tracking based on multiple-agent architecture — a real-time beat tracking system for audio signals —. In *Proc. of the Second Intl. Conf. on Multiagent Systems*, pages 103–110, 1996.

[Goto and Muraoka, 1997] Masataka Goto and Yoichi Muraoka. Real-time rhythm tracking for drumless audio signals — chord change detection for musical decisions —. In *Working Notes of the IJCAI-97 Workshop on Computational Auditory Scene Analysis*, 1997 (in press).

[Katayose *et al.*, 1989] H. Katayose, H. Kato, M. Imai, and S. Inokuchi. An approach to an artificial music expert. In *Proc. of the 1989 Intl. Computer Music Conf.*, pages 139–146, 1989.

[Large, 1995] Edward W. Large. Beat tracking with a nonlinear oscillator. In *Working Notes of the IJCAI-95 Workshop on Artificial Intelligence and Music*, pages 24–31, 1995.

[Rosenthal, 1992a] David Rosenthal. Emulation of human rhythm perception. *Computer Music Journal*, 16(1):64–76, 1992.

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

[Rowe, 1993] Robert Rowe. *Interactive Music Systems.* The MIT Press, 1993.

[Scheirer, 1996] Eric D. Scheirer. Using bandpass and comb filters to beat-track digital audio. (unpublished), 1996.

[Schloss, 1985] W. Andrew Schloss. *On The Automatic Transcription of Percussive Music — From Acoustic Signal to High-Level Analysis.* PhD thesis, CCRMA, Stanford University, 1985.

[Smith, 1996] Leigh M. Smith. Modelling rhythm perception by continuous time-frequency analysis. In *Proc. of the 1996 Intl. Computer Music Conf.*, pages 392–395, 1996.

[Todd and Lee, 1994] Neil Todd and Chris Lee. An auditory-motor model of beat induction. In *Proc. of the 1994 Intl. Computer Music Conf.*, pages 88–89, 1994.

[Todd, 1994] Neil P. McAngus Todd. The auditory "primal sketch": A multiscale model of rhythmic grouping. *Journal of New Music Research*, 23(1):25–70, 1994.

[Vercoe, 1994] Barry Vercoe. Perceptually-based music pattern recognition and response. In *Proc. of the Third Intl. Conf. for the Perception and Cognition of Music*, pages 59–60, 1994.