Automatic Rhythm Transcription from Multiphonic MIDI Signals

Haruto Takeda Takuya Nishimoto Shigeki Sagayama
Graduate School of Information Science and Technology, The University of Tokyo
Hongo, Bunkyo-ku, Tokyo 113-8656 Japan / {takeda,nishi,sagayama}@hil.t.u-tokyo.ac.jp

Abstract

For automatically transcribing human-performed polyphonic music recorded in the MIDI format, rhythm and tempo are decomposed through probabilistic modeling using Viterbi search in HMM for recognizing the rhythm and EM Algorithm for estimating the tempo. Experimental evaluation are also presented.

1 Introduction

We are investigating automatic transcription of MIDI (Musical Instrument Digital Interface) signals. As the MIDI format already includes the pitch information, the problem here is to recognize the note values, i.e., intended nominal lengths of notes as shown in Fig. 1, which we refer to "rhythm recognition."

Conventionally, it has been done by "quantization" of IOIs (Inter-Onset Intervals) of played notes. We used HMM (Hidden Markov Model) to solve this problem (Saito et al. 1999) by modeling both fluctuating note lengths and probabilistic constraints of note sequences. In this work, we also included multiple tempos in the HMM to find the best-matching tempo. In other works, tempo was included as hidden variables of probabilistic models (Cemgil et al, 2000; Raphael, 2001), or determined by clustering IOIs (Dixon, 2001), and rhythm was estimated based on the tempo.

In this paper, we treat rhythm recognition as a problem of probabilistically decomposing the observed IOIs into rhythm and tempo components.

2 Rhythm and Tempo

The observed duration (IOI) x [sec] of note in the performed MIDI data is related both to intended note value q [beats] in the score and to tempo variable τ [sec/beat] (average time per beat) by:

$$x\text{[sec]} = \tau\text{[sec/beat]} \times q\text{[beats]} \tag{1}$$

Rhythm recognition can be defined as a decomposition of IOIs $X = \{x_1, \cdots, x_N\}$ into tempo $T = \{\tau_1,\cdots,\tau_N\}$ and rhythm $Q=\{q_1,\cdots,q_N\}$. This is a kind of ill-posed problem since $\hat{Q}$ and $\hat{T}$ are not determined uniquely. In principle, any rhythm can be expressed in various ways, e.g. twice note values and half tempo gives the same note duration in Eq. 1. Furthermore, fluctuation of tempo and rhythm can not be completely separated. Decomposition is possible only in a probabilistic sense

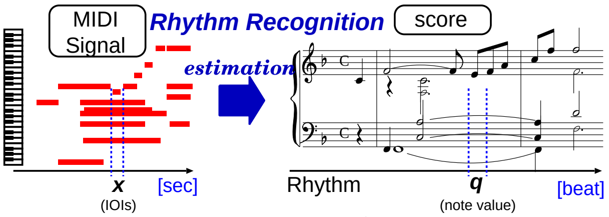

Figure 1: Rhythm recognition for automatic transcription

assuming that T is constant or slowly changing (at least within phrases), and that Q often fit common rhythm patterns. Human often can recognize rhythm from the musical performance because they have *a priori* knowledge of rhythm, e.g., what type of rhythm patterns are likely to appear. In our approach, "most likely rhythm patterns" for the given MIDI data are estimated by search in the proposed probabilistic models, whose parameters are optimized by stochastic training with existing scores and performances.

Our goal is to separate rhythm and tempo by iterating the estimation of the two. First, we estimate rhythm from the IOIs of the given MIDI using tempo-invariant feature parameters. Then, using the estimated rhythm and the given IOIs, the tempo is estimated. Rhythm and tempo are alternately re-estimated using the estimated counterpart. In the next sections, we discuss first two steps.

3 HMM Using Rhythm Vectors

This section describes the method to estimate rhythm from observed IOIs (Saito, 1999; Otsuki, 2001; Takeda, 2002).

3.1 Stochastic Modeling of Rhythm Patterns

We assume that a sequence of note values appear in music with certain probability, which can be approximated by an n-gram probability, i.e., a conditional probability $P(q_t|q_{t-1},\cdots,q_{t-n+1})$ dependent on the history of previous $n-1$ note values. Similar to the n-gram language model often used in speech recognition, the probability of rhythm $Q = \{q_1,\cdots,q_N\}$ is approximated by

$$P(Q) \approx P(q_1,\cdots,q_{n-1}) \prod_{t=n}^{N} P(q_t|q_{t-1},\cdots,q_{t-n+1}) \tag{2}$$

Conditional probabilities can be obtained through statistical training using already composed music scores.

3.2 Rhythm Vector: a Tempo-Invariant Feature

We introduce a tempo-invariant feature named "rhythm vector," since the tempo of the input data is not given in advance and it may vary throughout the data. From our assumption that the tempo is constant or changes slowly, the proportion of consecutive note lengths x is nearly independent from tempo τ according to Eq.1). Therefore, we introduce *rhythm vector* as follows:

$$\boldsymbol{r}_t = (r_t^1,\cdots,r_t^m) \;\text{ where }\; r_t^i = \frac{x_{t+i}}{x_t + \cdots + x_{t+m-1}} \tag{3}$$

Table 1: Formulating rhythm recognition with HMM

Hidden States	State Transition	Output Signal
n-tuples of note	$(n+1)$-gram	rhythm vectors

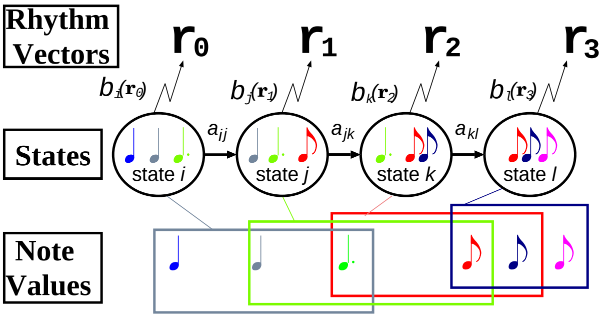

Figure 2: Rhythm vector as the output of HMM

Here, probabilistic distribution $p(\boldsymbol{r})$ is assumed to follow the normal distribution $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, whose parameters, mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$, can be obtained through statistical training using human-performed MIDI data.

3.3 Rhythm Estimation: Search Problem in HMM

The two probabilistic models of rhythm vector and rhythm pattern can be combined in the HMM framework as shown in Table 1 and Fig. 2. This HMM gives the probability $P(X|Q)P(Q)$, where $P(Q)$ is the probabilistic model of rhythm score and $P(X|Q)$ is that of rhythm vectors and tempo fluctuation.

Rhythm estimation is to find the time sequence of states in the state transition network, Q, that gives the maximum *a posteriori* probability, $P(Q|X)$, given a sequence of observed note lengths series, X. Maximizing $P(Q|X)$ is equivalent to maximizing $P(X|Q)P(Q)$ according to Bayes theorem. The optimal sequence of states in HMMs is efficiently found through the well-known Viterbi algorithm. The sequence of intended notes $\hat{Q}$ is estimated in the maximum likelihood sense.

3.4 Multiphonic Case

The above stated method can also be applied to the multiphonic case. Projecting the onset timings of all notes in a multiphonic music score onto a one-dimensional time axis, we obtain a “rhythm score” from which the n-gram “grammar” can be defined in the same way. After preprocessing for grouping nearly simultaneous onsets, the rhythm score is recognized from the input sequence of IOIs across multiple voices in the observed MIDI signal, followed by postprocessing for assigning note length to each of recognized note onsets.

4 Tempo Estimation

Tempo T, the sequence of instantaneous tempo estimated by $\tau_t = x_t/\hat{q}_t$ from observed IOIs X and the estimated rhythm $\hat{Q}$, often contain errors as shown in Fig. 3 such as double tempo, half tempo, and errors due to confusion by triplets mainly caused by the nature of rhythm vector. We model the distribution of estimated tempo T by a Gaussian mixture distribution and apply the EM (Expectation-Maximization) algorithm to estimate the true tempo. After estimating the tempo, note values are estimated again.

5 Experimental Evaluation

The proposed method was evaluated by using 3 classical music pieces listed in Table 2 recorded in the MIDI format, which were performed 2 times by 5 players for each piece. 19 kinds of note values (whole note, quarter note, etc.) were treated and represented by 6859 ($= 19^3$) hidden states in the HMM, whose transition probabilities were trained by 13 classical pieces containing 4355 note values, and whose output probabilities were trained by 2 music pieces by 2 players containing 1288 IOIs. Rhythm recognition rate RRR-1 (first estimation of rhythm) and RRR-2 (rhythm estimation after tempo estimation) were obtained as shown in Table 3. “Prep” shows the rate of correct preprocessing (synchronizing grace notes, etc.).

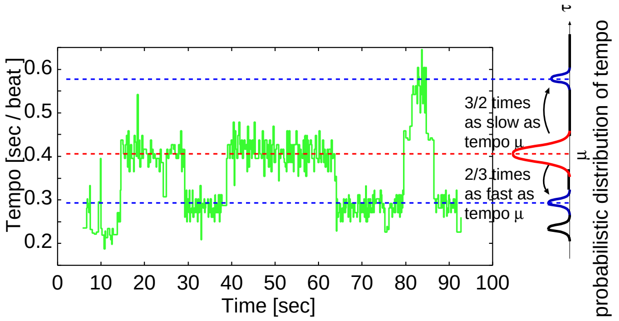

Figure 3: Tempo plot: derived from rhythm recognition results

Table 2: Testing music pieces for rhythm recognition

Fuga	J. S. Bach: Fuga in C minor, BWV847, Well-Tempered Clavier Book I.
Sonata	L. v. Beethoven: Piano Sonata No. 20, 1st Mov.
Träumerei	R. Schumann: “Träumerei” from “Kinderszenen,” Op. 15, No.7.

Table 3: Rhythm recognition results [%]

Data	Prep.	RRR-1	RRR-2
Fuga	97.5	94.1	94.3
Sonata	97.4	60.7	78.5
Träumerei	97.5	68.4	72.0

6 Conclusion

We proposed a method for rhythm recognition of MIDI signals performed by humans. In the experimental evaluation, the original rhythm was successfully estimated from MIDI piano performances.

Acknowledgment

This research was partly supported by CREST (Auditory Brain Project).

References

A. Cemgil, B. Kappen, P. Desain, H. Honing, “On tempo tracking: Tempogram Representation and Kalman filtering,” Journal of New Music Research, 2000.

S. Dixon, “Automatic Extraction of Tempo and Beat from Expressive Performances,” 2001.

T. Otsuki, N. Saito, M. Nakai, H. Shimodaira, and S. Sagayama, “Musical Rhythm Recognition Using Hidden Markov Model,” Transaction of Information Processing Society of Japan, Vol. 43, No. 2, pp. 245–255, 2002. (in Japanese)

C. Raphael, “Automated Rhythm Transcription,” Proc. of ISMIR, pp. 99–107, 2001.

N. Saito, M. Nakai, H. Shimodaira, S. Sagayama, “Hidden Markov Model for Restoration of Musical Note Sequence from the Performance,” Proceedings of the Joint Conference of Hokuriku Chapters of Institutes of Electrical Engineers, Japan, 1999, F-62, p.362, Oct 1999 (in Japanese).

H. Takeda, T. Otsuki, N. Saito, M. Nakai, H. Shimodaira, S. Sagayama, “Hidden Markov Model for Automatic Transcription of MIDI Signals,” Proc. 2002 MMSP, 2002.