AN ERROR CORRECTION FRAMEWORK BASED ON DRUM PATTERN PERIODICITY FOR IMPROVING DRUM SOUND DETECTION

Kazuyoshi Yoshii† *Masataka Goto*‡ *Kazunori Komatani*† *Tetsuya Ogata*† *Hiroshi G. Okuno*†

†Department of Intelligence Science and Technology
Graduate School of Informatics, Kyoto University
Sakyo-ku, Kyoto 606-8501, Japan
{yoshii,komatani,ogata,okuno}@kuis.kyoto-u.ac.jp

‡National Institute of Advanced
Industrial Science and Technology (AIST)
Tsukuba, Ibaraki 305-8568, Japan
m.goto@aist.go.jp

ABSTRACT

This paper presents a framework for correcting errors of automatic drum sound detection focusing on the periodicity of drum patterns. We define drum patterns as periodic structures found in onset sequences of bass and snare drum sounds. Our framework extracts periodic drum patterns from imperfect onset sequences of detected drum sounds (bottom-up processing) and corrects errors using the periodicity of the drum patterns (top-down processing). We implemented this framework on our drum-sound detection system. We first obtained onset sequences of the drum sounds with our system and extracted drum patterns. On the basis of our observation that the same drum patterns tend to be repeated, we detected time points which deviate from the periodicity as error candidates. Finally, we verified each error candidate to judge whether it is an actual onset or not. Experiments of drum sound detection for polyphonic audio signals of popular CD recordings showed that our correction framework improved the average detection accuracy from 77.4% to 80.7%.

1. INTRODUCTION

The concept of music information retrieval (MIR) has attracted a lot of attention. MIR enables us to acquire musical pieces by executing a query about music contents such as rhythms and melodies. To create an MIR system, we are working on an automatic rhythm description. Because drums are closely related to the rhythm, many drum-sound detection systems [1, 2] have been proposed. We developed a system, called *AdaMast* [3, 4], based on adaptation and matching of drum-sound spectrogram templates. However, bottom-up methods are required to describe higher-level content (e.g., tempo). Although AdaMast can automatically detect drum-sound onsets in polyphonic audio signals of CD recordings, detection errors often occur because the task for those signals is very difficult with low-level processing only. One of solutions for reducing the errors is to use higher-level content as top-down constraints for drum-sound onsets.

We therefore focus on *drum patterns* which are higher-level content of the rhythm. There are a few studies on drum sound detection that focus on the periodicity of drum patterns. In these studies, bar-line times and bar-line intervals are considered as start times and lengths of drum patterns. For example, to detect drum sounds in audio signals of drum tracks, Paulus *et al.* [5] used periodic N-grams for modeling the transition of drum patterns. Gillet *et al.* [6] proposed a post-processing method that corrects detected onsets on the basis of the periodicity. Although their method was applied to detect drum sounds in polyphonic audio signals, the performance was not improved. A main limitation of these methods is that they depend on the accuracy of bar-line time estimation. They were only tested in ideal conditions; correct bar-lines were manually given.

As mentioned above, start times and lengths of drum patterns are defined as bar-line times and bar-line intervals in general. To obtain the drum patterns, it is necessary to *split* onset sequences of drum sounds by estimating bar-line times. However, some drum patterns are not always useful for the periodicity-based error correction when they are not periodic. In addition, estimation errors of bar-line times often give fatal damages to the onset detection and correction.

We define drum patterns as *periodic structures* found in onset sequences of drum sounds in popular songs. We try to *extract* periodic structures which are useful for the error correction from onset sequences of drum sounds using a bottom-up method. The start time and length of an extracted pattern are often (i.e., not always) equal to the bar-line time and bar-line interval.

In this paper, we propose a new error correction framework that evaluates the reliability of each detected onset on the basis of the periodicity of drum patterns and that verifies unreliable onsets. The drum patterns are used as top-down constraints, which are obtained by results of the bottom-up processing to onset sequences of drum sounds. Our framework does not use any prior information of bar lines and it is robust to the bar-line estimation errors. We found that the framework worked well for polyphonic audio signals of popular CD recordings including various instrument sounds.

The rest of this paper is organized as follows. Section 2 and Section 3 describe our error correction framework and our drum-sound detection system with the error correction function. Section 4 explains the actual implementation of this framework. Section 5 shows evaluation results. Finally, Section 6 summarizes the paper.

2. ERROR CORRECTION FRAMEWORK

An error correction framework is described first. It can be applied to drum-sound onsets obtained by any drum-sound detection systems.

2.1. Concept

Onset detection errors are corrected through three steps, as shown in Fig. 1. Each step is briefly explained as follows:

Step 1: Drum pattern extraction

To extract drum patterns which are useful for evaluating the reliability of detected onsets on the basis of the periodicity, only periodic drum patterns are extracted. Because an extracted drum pattern is periodic, similar successive structures appear in the neighborhood of the extracted pattern even if the estimation of bar-line times fails, as shown in Fig. 2. Therefore, the reliability of each detected onset is appropriately evaluated by examining these successive structures.

Thanks to the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Grant-in-Aid for Scientific Research for funding.

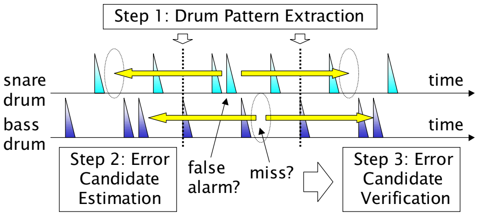
Step 1: Drum Pattern Extraction
snare drum
time
bass drum
time
Step 2: Error Candidate Estimation
false alarm?
miss?
Step 3: Error Candidate Verification

Fig. 1. Overview of error correction framework.

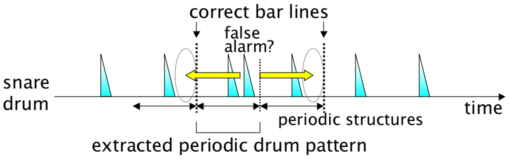

Fig. 2. Error candidate estimation based on the periodicity of successive structures in case of bar-line estimation errors.

Step 2: Error candidate estimation
Because there are two types of detection errors — false alarms and misses, candidate times of each error type need to be detected according to the drum pattern periodicity. *False-alarm candidates* indicate unreliable onsets among the detected onsets. *Miss candidates* indicate potential onsets that were not detected by a drum transcription system, although the probability that they are actual onsets is comparatively high.

Step 3: Error candidate verification
Because those candidates do not always correspond to actual errors, each candidate needs to be verified carefully with an evaluation measure. For example, we can reuse a drum transcription system with dynamically changing judgment thresholds according to the reliability of each candidate.

2.2. Approach

Our system implements this framework by making the following two assumptions on input audio signals.

1. The time signature is 4/4. If the actual time signature of a song is 2/4, the time signature of the song is assumed to be 4/4 by concatenating two successive measures.
2. The tempo is between 60 and 200 M.M. [1]

These assumptions fit into a large class of popular music.

First of all, we should prepare a reference pattern. It represents prior onset distributions of bass and snare drum sounds in a bar-line interval (e.g., onsets of bass drum sounds tend to be in the first beat). The reference pattern is obtained by averaging various drum patterns sampled from many MIDI files of popular songs.

To extract drum patterns, we perform two steps: *period length calculation* and *reference pattern matching*. First, we calculate the period length of onset sequences at each time by using a short-time Fourier transform (STFT). Next, we find time points when the correlation between the reference pattern and the onset sequences takes the local maximum while shifting and extending the reference pattern along the time axis on the basis of the period length. Finally, the drum patterns are extracted from these time points.

To estimate error candidates in a drum pattern, we examine the same time points in successive structures of both sides of the pattern (see horizontal arrows in Fig. 1 and Fig. 2). If there is an onset in a time point in the drum pattern but there are few onsets in the examined points, we judge that the onset is a false-alarm candidate. If there is no onset in a time point but there are some onsets in the examined points, we judge that the time point is a miss candidate.

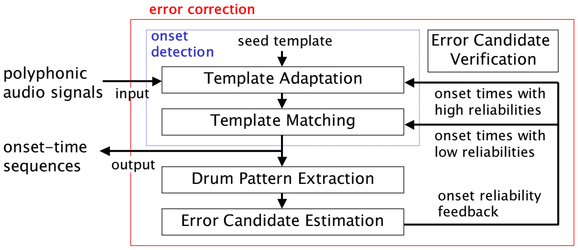

Fig. 3. Overview of feedback architecture of our system.

3. DRUM SOUND DETECTION SYSTEM WITH ERROR CORRECTION FUNCTION

Fig. 3 shows a novel architecture of our drum-sound detection system that implements the error correction framework. The architecture comprises two parts. An onset-detection part detects onset times of the bass and snare drum sounds on the basis of adaptation and matching of spectrogram templates. An error-correction part evaluates the reliability of the onset times on the basis of the drum pattern periodicity and verifies them by reusing the onset-detection part.

In this architecture, bottom-up processing (i.e., drum pattern extraction from onset sequences) and top-down processing (i.e., error correction based on onset reliability) are linked through the evaluation feedback of the onset reliability based on the drum pattern periodicity. In other words, this architecture has a self-refining function based on self-evaluation results to yield more reasonable outputs.

3.1. Onset Detection based on Adaptation and Matching of Spectrogram Templates of Drum Sounds

This part takes polyphonic audio signals as inputs and outputs onsets of the bass and snare drum sounds. Our system [3, 4], which comprises the following successive stages, achieves this purpose.

Template Adaptation Its purpose is to obtain actual power spectrograms of the bass and snare drum sounds in an input audio signal. First, two initial spectrograms are prepared. These are called *seed templates* in a previous study [3]. Note that they are different from the actual spectrograms of the drum sounds in the input audio signal. This stage updates each seed template iteratively by setting multiple drum-sound spectrograms found in the input audio signal for the adaptation targets.

Template Matching Its purpose is to detect all the onset times of the bass and snare drum sounds in the polyphonic audio signal, even if other instrument sounds overlap them. This stage judges whether each onset candidate is an actual onset or not; the onset candidate times are obtained in advance by peak-picking the song spectrogram. To enable this, we designed a distance measure that is robust for polyphonic mixtures. In this measure, the distance is calculated between the adapted template and a song-spectrogram segment at each onset candidate. The distance threshold is automatically determined.

3.2. Error Correction based on Evaluation Feedback of Onset Reliability

In the error candidate estimation step, we can also detect *reliable onsets* which accord with the periodicity of drum patterns well. Be-

[1] Mälzel's Metronome: the number of quarter notes per minute.

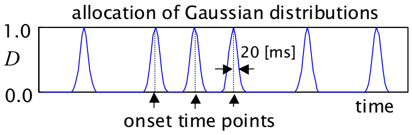

Fig. 4. Calculation of detected onset distribution.

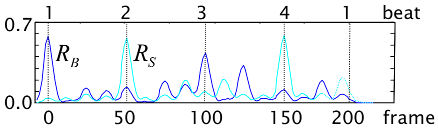

Fig. 5. Calculation of reference onset distributions.

cause it is highly likely that spectrograms at the reliable onsets include the target drum-sound spectrograms, the reliable onsets are useful clues for the accurate adaptation. The error candidate verification step is implemented by reusing the onset detection part.

Template Re-adaptation The template updating is performed again by setting multiple spectrograms extracted from reliable onsets for the adaptation targets.

Template Re-matching The template matching is performed again for false-alarm and miss candidates by dynamically changing the threshold according to the reliability of each onset.

4. SYSTEM IMPLEMENTATION

To use a STFT for calculating the period length, discrete onset sequences are transformed to quasi-continuous functions by allocating a Gaussian distribution to each onset time, as shown in Fig. 4. The standard deviation of the Gaussian is 20 [ms]. The time resolution of the functions is 10 [ms], which corresponds to 1 [frame]. We call them *detected onset distributions*, which are represented as D_B and D_S for the bass and snare drums, respectively. Henceforth, we often omit the subscripts $_B$ and $_S$ for convenience.

4.1. Drum Pattern Extraction

The following two steps are performed to determine the lengths and start times of drum patterns.

Period length calculation A STFT with a Hanning window is applied to detected onset distributions D_B and D_S. The window length is 2048 [frames], and the shifting interval is 1 [frame]. These two amplitude spectrograms are summed, and then a total amplitude spectrogram is obtained. The period length $L(t)$ [frames] is obtained at each frame t by calculating the peak interval of the spectral autocorrelation.

Reference pattern matching First, a reference pattern is obtained by averaging measure-length drum patterns in MIDI files of popular music database RWC-MDB-P-2001 [7] after these files are normalized at 120 M.M. Let R_B and R_S be *reference onset distributions*, which are obtained by convoluting the Gaussian with the reference pattern, as in calculating D. Fig. 5 shows R_B and R_S. Their length is 200 [frames].

Next, the following processing is performed at each frame t. Fig. 6 shows an overview of this processing. To reduce bar-line estimation errors (double-tempo errors), the length of R is extended to $L(t)$ and $2L(t)$ [frames] between 120 [frames] (200 M.M.) and 400 [frames] (60 M.M.). The correlation between D starting from frame t and each extended R is calculated; it is normalized with the length of R. Then, the correlations of both the drum types are summed for each length of R. The total correlation $C(t)$ is determined as the maximum summed correlation. Let $L'(t)$ be the length of R ($L(t)$ or $2L(t)$) that yields the maximum correlation.

Finally, the start times of the drum patterns are obtained by picking frame t at which $C(t)$ is larger than a threshold. The lengths of the drum patterns are determined as $L'(t)$ at picked frame t. If a drum pattern overlaps with another one, the pattern that has the larger correlation is extracted.

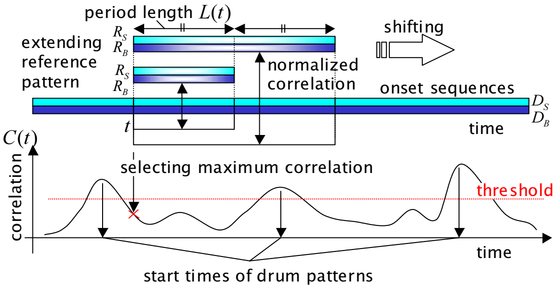

Fig. 6. Drum pattern extraction using reference pattern.

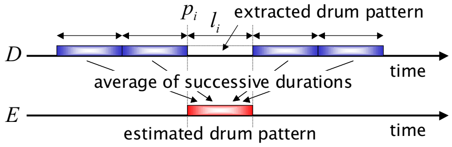

Fig. 7. Calculation of estimated onset distribution.

4.2. Error Candidate Estimation

To estimate actual onset times in each drum pattern, the average of detected onset distributions is calculated in four successive structures of both sides of the drum pattern, as shown in Fig. 7. Let E be an *estimated onset distribution*, obtained by

$$E(p_i+\delta_i) = \frac{1}{4}\sum_{m=\{-2,-1,1,2\}} D(p_i+\delta_i+m\,l_i) \qquad (1)$$

where p_i and l_i $(i = 1, \cdots, N)$ are the start time and length of the i-th extracted pattern. N is the number of extracted drum patterns. δ_i is an offset time from start time p_i $(0 \le \delta_i \le l_i)$.

Each detected onset is grouped into one of three classes (i.e., a class of reliable onsets and two classes of false-alarm candidates) from the viewpoint of reliability by comparing the detected onset distribution D with the estimated onset distribution E. In addition, a class of miss candidates is considered. As shown in Fig. 8, there are a total of four classes, which are obtained by

Reliable onsets: $\{t|D(t) = 1.0,\ E(t) \ge 0.8\}$,

False-alarm candidates 1: $\{t|D(t) = 1.0,\ 0.8 > E(t) \ge 0.05\}$,

False-alarm candidates 2: $\{t|D(t) = 1.0,\ 0.05 > E(t)\}$,

Miss candidates: $\{t|D(t) = 0.0,\ E(t) \ge 0.4\}$.

4.3. Error Candidate Verification

As described in Section 3.2, reliable onsets are used for re-adaptation. In addition, re-matching is performed while decreasing the distance threshold gradually as the reliability decreases for false-alarm candidates or while increasing the threshold for miss candidates.

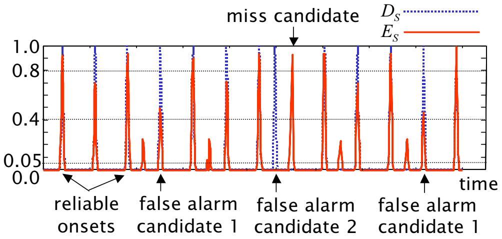

Fig. 8. Example of onset reliability evaluation.

Table 1. Musical pieces used for experiments.

piece number (No.) in RWC-MDB-P-2001
1,5,6,7,8,10,11,12,13,14,18,20,21,22,23,25,26,30, 33,35,36,37,40,41,43,44,46,47,48,50,51,52,53,54, 58,59,61,62,63,66,70,83,84,85,87,88,89,90,92,98

Table 2. Notation of tested methods.

TM	Template Matching
TA	Template Adaptation
ECM	Error Correction based on Re-Matching
ECA	Error Correction based on Re-Adaptation

5. EXPERIMENTAL EVALUATION

We performed comparative experiments of detecting bass and snare drum sounds for polyphonic audio signals including various sounds. The four methods described in Section 3 were enabled one by one to evaluate the different performance improvements.

5.1. Conditions

Our methods were tested on fifty songs sampled from popular music database RWC-MDB-P-2001 developed by Goto *et al.* [7]. Table 1 shows a list of the songs. The audio signals were sampled with CD quality, and they were converted to monaural recordings. Table 2 shows a notation list of the tested methods. Each method was evaluated by performing comparative experiments with a different combination of our methods. The evaluation measures are defined as

$$\begin{aligned}\text{recall rate} &= \frac{\#(\text{correctly detected onsets})}{\#(\text{actual onsets})},\\ \text{precision rate} &= \frac{\#(\text{correctly detected onsets})}{\#(\text{detected onsets})},\\ \text{f-measure} &= \frac{2\cdot\text{recall rate}\cdot\text{precision rate}}{\text{recall rate}+\text{precision rate}}.\end{aligned}$$

The error tolerance is set to 25 [ms].

5.2. Results

Table 3 and Table 4 show the results of detecting the bass and snare drum sounds, respectively. Our methods improved the f-measures, and combining them yielded further improvements. A fully-enabled error correction method (ECA+ECM) improved the f-measure from 76.8% to 81.1% (a 18.7% error reduction) in the bass drum sound detection and from 78.0% to 80.3% (a 10.6% error reduction) in the snare drum sound detection. The precision rates of the four classes in Section 4.2 are 91%, 77%, 66%, and 23% in the bass drum sound detection. They are 92%, 37%, 51%, and 6% in the snare drum sound detection. These results proved the effectiveness of our error correction framework.

Table 3. Experimental results of bass drum sound detection.

method combination	recall	precision	f-measure
TM	70.122%	70.109%	70.115%
TM+TA (baseline)	75.838%	77.758%	76.786%
TM+TA+ECM	76.194%	78.872%	77.510%
TM+TA+ECA	79.691%	81.463%	80.567%
TM+TA+ECM+ECA	**79.835%**	**82.449%**	**81.121%**

Table 4. Experimental results of snare drum sound detection.

method combination	recall	precision	f-measure
TM	67.126%	68.891%	67.997%
TM+TA (baseline)	77.968%	78.025%	77.996%
TM+TA+ECM	78.106%	80.747%	79.404%
TM+TA+ECA	78.191%	80.523%	79.340%
TM+TA+ECM+ECA	**78.283%**	**82.464%**	**80.319%**

Table 5. Experimental results of tempo estimation.

correct	double-tempo errors	other errors
40 songs (80%)	9 songs (18%)	1 song (2%)

As a secondary effect, our methods will help with tempo estimation. The tempo is determined by converting the majority of $L'(t)$ to M.M. $\left(\frac{60{,}000[\text{ms}]}{L'(t)\cdot 10[\text{ms}]}\cdot 4\ [\text{quarter notes}]\right)$. Table 5 shows the results of the tempo estimation. They are promising.

6. CONCLUSION

We presented an error correction framework and introduced an implementation in our template-based drum-sound detection system. An onset detection part in the system is equipped with a self-refining function based on evaluation feedback of the onset detection reliability. This is an integration of bottom-up processing (i.e., drum pattern extraction from onset sequences) and top-down processing (i.e., onset correction based on drum pattern periodicity). We demonstrated proof of the concept with comparative experiments. We believe that our framework could be helpful for many studies.

7. REFERENCES

[1] C. Dittmar and C. Uhle, “Further steps towards drum transcription of polyphonic music,” in *AES, 116th Conv.*, 2004.

[2] D. FitzGerald, B. Lawlor, and E. Coyle, “Drum transcription in the presence of pitched instruments using prior subspace analysis,” in *ISSC2003*, pp. 202–206.

[3] K. Yoshii, M. Goto, and H.G. Okuno, “Automatic drum sound description for real-world music using template adaptation and matching methods,” in *ISMIR2004*, pp. 184–191.

[4] K. Yoshii, M. Goto, and H.G. Okuno, “Drum sound recognition for polyphonic audio signals by adaptation and matching of spectrogram templates with harmonic structure suppression,” *IEEE Trans. SAP (submitted)*, 2005.

[5] J. Paulus and A. Klapuri, “Conventional and periodic N-grams in the transcription of drum sequences,” in *ICME2003*, pp. 737–740.

[6] O. Gillet and G. Richard, “Drum track transcription of polyphonic music using noise subspace projection,” in *ISMIR2005*.

[7] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka, “RWC Music Database: Popular, Classical, and Jazz Music Databases,” in *ISMIR2002*, pp. 287–288.