DRUM TRANSCRIPTION IN POLYPHONIC MUSIC USING SEMI-SUPERVISED NON-NEGATIVE MATRIX FACTORIZATION

First author
Affiliation1
author1@ismir.edu

Second author
Retain these fake authors in submission to preserve the formatting

Third author
Affiliation3
author3@ismir.edu

ABSTRACT

In this paper, a drum transcription algorithm using semi-supervised non-negative matrix factorization has been presented. This method allows users to separate percussive activities from harmonic activities with pre-trained drum templates, and detect drum events from the extracted percussive activity matrix. A polyphonic subset from ENST drum dataset has been used for training and testing of the algorithm. The system has been tested using different rank settings, and a cross-performer validation process has been performed to evaluate the reliability of the system. The results show that the system can achieve 61 to 77% recognition rate on multiple drums in polyphonic music. In the future, more efforts will be put on differentiating different playing styles and more drum parts, leading toward a complete drum transcription system.

1. INTRODUCTION

Transcribing music content into sheet music or any form of score is an essential skill to musicians for analysis and composition purposes. However, It is often considered a time-consuming and non-trivial task, for it requires repetitive listening and integral knowledge of music and instruments. With the advance of computing power and various machine learning techniques, a system that has the abilities to automatically recognize music content has become a plausible idea and interests many researchers in the field of Music Information Retrieval (MIR) [1]. In general, Automatic Music Transcription (AMT) systems could not only serve as a tool to record the music content, identify notes from improvisations, but also lead to the realization of a music intelligence system [2]. To build a complete AMT system, many subtasks and challenges, such as multi-pitch detection, onset detection, instrument recognition, and rhythm extraction, have been attempted [1]. Comparing with pitched instruments, transcribing unpitched music events such as percussive sounds seem to be less addressed, and a robust algorithm to detect drum sounds in polyphonic music is still an open question in this field.

Drum track, especially in pop music, often plays an important role in determining music structure, style, rhythm and tempo. In many cases, it shares the same importance as the harmonic part in the music. That said, a good drum transcription system could provide essential information of the music content, and could potentially facilitate the realization of many interesting applications. For example, a drum transcription and drum source separation task could mutually benefit from each other and provide the possibility to manipulate drum sounds within an existing drum track [3], [4]. Also, music education could be an application of a drum transcription system, for it could help students identify drum event within a polyphonic music, and provide instructions by analyzing and comparing the differences between the input and reference performances. Furthermore, such system could also be integrated as a part of machine listening, improving the existing systems of robotic musicians[5].

Therefore, this study aims to explore alternative solutions to drum transcription in polyphonic music. The final goal of this paper is to find a potential direction to push the limit of this task, leading toward further musical applications.

2. RELATED WORKS

The early attempts to transcribe percussive sounds main focused on classifying monophonic signals [6][7]. With standard approaches such as feature extraction and classification, fairly high accuracy were reported in the previous studies. However, in the real use case, a drum transcription system is expected to work in polyphonic signal instead of monophonic. Therefore, solving this problem in the context of polyphonic music had become another criterion, and different methods could be found in previous research [8][15]. According to [15], recent studies on the drum transcription in polyphonic music could be categorized into three types: segment and classify, separate and detect, match and adapt. For the first type of approaches, the common procedure starts by applying onset detection to the audio signal in order to segment the music event. Once the event has been detected, various features from time or spectral domain of the signal will be extracted, and a classifier will be trained to classify the event based on the extracted features. This type of approaches seem to perform well when the data and features are well chosen [13], [15]. However, to get good results, sufficient amount of data, careful preprocessing and training steps are required in this type of systems. Additionally, to handle the situation of simultaneous sounds, more classes need to be trained.

The second type of approaches is based on the assumption that music signal is a superposition of different sound sources. By decomposing the signal and into different source templates and corresponding activities, the music content could be transcribed by detecting onsets of these activities. Different methods such as Independent Subspace Analysis (ISA) [16], Prior Subspace Analysis (PSA) [8], and Nonnegative Matrix Factorization (NMF) [14], [17], [18] are the examples in this category. This type of approaches is usually easier to interpret, since most of the decompositions have been done on the spectrogram of the signal. Also, the separate nature allows simultaneous events to be handled easily in this case. However, to be able to transcribe different kinds of music, a large template might be required prior to the decomposition. Moreover, the number of rank during the decomposition process could be difficult to determine in some cases.

The third type of approaches uses pre-trained templates to detect drum events [19][20]. An iterative process has been taken to search for the closest matches to these templates and adapt them. The proposed system has been evaluated and performed well in MIREX 2005 drum detection competition. However, to coverage a wider range of sounds, multiple seed templates need to be prepared prior to the process.

In this paper, an approach that falls into the second category has been presented. Based on the popular NMF based

3. METHOD

about 1 page

3.1 Algorithm Description

This part I will first mention the co-factorization paper, and then introduce my modification of the algorithm, the cost function and the updating methods...

3.2 Processing Steps

Here I will put my system flowchart, and explain my processing procedure step by step.

3.2.1 Template Extraction

1)extract templates

3.2.2 Activity Detection

2)thresholding on activity matrix

4. EVALUATION

about 2 pages

8pt

	Original	HPF	HPF+LPF	HPF+BPF+LPF
P.	0.654	0.704	0.704	0.702
R.	0.776	0.786	0.780	0.746
F.	0.617	0.618	0.569	0.587

Table 1. Results from templates of all drummers.

Drummer	Original	HPF	HPF+LPF	HPF+BPF+LPF
P.	0.654	0.704	0.704	0.702
R.	0.776	0.786	0.780	0.746
F.	0.617	0.618	0.569	0.587

Table 2. Results of cross-performer validation.

4.1 Dataset Description

1)dataset description

In this project, all of the experiments have been conducted on the minus one subset from the ENST public drum dataset [22]. This dataset consist of recordings from three different drummers performing on their own drum kits. The recording contains single hits on different drums, short phrases of drum beats, drum solos, and short play through with the accompaniments. The minus one public subset has 64 tracks of polyphonic music; each track has a length of approximately 70 seconds. The music style varies from each other. Particularly, this subset contains many drum playing techniques such as ghost notes, flam, and dragetc, which could be considered difficult for many of the existing drum transcription algorithms. The accompaniments are mixed with the wet mix in the dataset without any modification of the magnitudes.

The onset counts of each class within different drummers recordings are shown in Table 1. More details on the creation process of this dataset could be found in their corresponding paper.

4.2 Evaluation Procedure

To keep classifiers from seeing the data during the training phase, a cross-performer validation process has been performed. As shown in Fig 2, when a drummers recordings is selected as the training data, the classifier will be tested using the other drummers recordings. The process applies to every drummers recordings, and the performance of the training data is the average of the test results from the other two drummers.

The accuracy of the classification has been evaluated by calculating the precision, recall and F-scores.

4.3 Evaluation Results

3)basic results + improved results (filtering) 4)system properties (rank change, iteration and convergence, cross training and testing)

String value	Numeric value
Hello ISMIR	2014

Table 3. Table captions should be placed below the table.

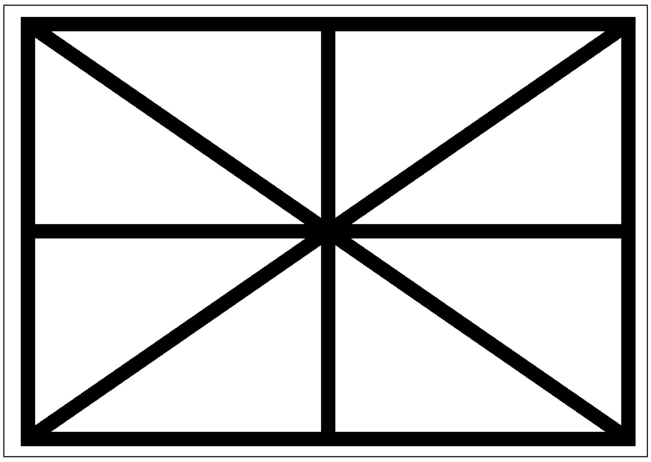

Figure 1. Figure captions should be placed below the figure.

4.4 Discussions

5)discussion

although the results seem to be comparable, NMF based method still has some advantages over instance based methods: 1. simultaneous sounds will be detected separately 2. require less labeled data during the training process 3.

5. CONCLUSION

about 0.5 page

6. REFERENCE

6.1 Figures, Tables and Captions

All artwork must be centered, neat, clean, and legible. All lines should be very dark for purposes of reproduction and art work should not be hand-drawn. The proceedings are not in color, and therefore all figures must make sense in black-and-white form. Figure and table numbers and captions always appear below the figure. Leave 1 line space between the figure or table and the caption. Each figure or table is numbered consecutively. Captions should be Times 10pt. Place tables/figures in text as close to the reference as possible. References to tables and figures should be capitalized, for example: see Figure 1 and Table 3. Figures and tables may extend across both columns to a maximum width of 17.2cm.

7. EQUATIONS

Equations should be placed on separated lines and numbered. The number should be on the right side, in parentheses.

$$E = mc^2 \tag{1}$$

8. CITATIONS

All bibliographical references should be listed at the end, inside a section named "REFERENCES," numbered and in alphabetical order. Also, all references listed should be cited in the text. When referring to a document, type the numbering square brackets [1] or [1–3].

9. REFERENCES

[1] E. Author: "The Title of the Conference Paper," *Proceedings of the International Symposium on Music Information Retrieval*, pp. 000–111, 2000.

[2] A. Someone, B. Someone, and C. Someone: "The Title of the Journal Paper," *Journal of New Music Research*, Vol. A, No. B, pp. 111–222, 2010.

[3] X. Someone and Y. Someone: *Title of the Book*, Editorial Acme, Porto, 2012.