IDENTIFICATION OF DRUM OVERHEAD-MICROPHONE TRACKS IN MULTI-TRACK RECORDINGS

Keita Arimoto

YAMAHA Corporation
keita.arimoto@music.yamaha.com

ABSTRACT

This paper proposes a system that automatically identifies drum overhead-microphone tracks out of multi-track recordings. The system is designed to perform a binary classification between drum overhead-microphone tracks and the rest based on the following characteristics that are commonly observed in drum overhead-microphone tracks: A) more onsets than the other drum tracks B) catch more percussive sounds than non-percussive C) high onset coincidence with the other percussive tracks. The system is evaluated against 21 multi-track recordings that involve harmonic tracks (e.g. Vocal, Guitar, Piano, …) as well as drums. The system achieved a recall of 95.23% on drum overhead-microphone tracks (42 tracks) and a recall of 92.14% on the rest of the tracks (458 tracks).

1. INTRODUCTION

Musical instrument labels are one of the most basic and important pieces of meta-data for multi-track recordings. There are several studies for automatic musical instrument identification (e.g. [1]) that are expected to be applicable for automatic labeling. However, the concept is not directly applicable for drum overhead-microphone tracks as they are intended to catch the sound coming from the entire drum set instead of a specific target.

A possible option for the identification of drum overhead-microphone tracks is extending the concept of drum transcription for polyphonic audio (e.g. [2]). However, it is hard to build a general system that works for any input source as the transcription-based approach usually relies on models or knowledge of sounds lying in a polyphonic signal. On the other hand, Ronan *et al.* [3] tried automatic subgrouping of multi-track audio using audio features and clustering. The idea is more or less relevant even though it does not directly address the identification of individual tracks.

The most relevant work is done by Scott and Kim [4]. They tried an automatic classification of drum tracks, including overhead-microphone, by means of global feature for an entire track and a multi-class SVM classifier.

Following the concept, this paper proposes a method for an automatic identification of drum overhead-microphone tracks based on the feature extraction for each onset instead of global feature for an entire track. There are advantages in the onset-based approach as it allows focusing on the following three characteristics that are commonly observed in drum overhead-microphone tracks:

A) more onsets than the other drum tracks.
   This is quite reasonable as they are intended to catch the entire drum set.

B) catch more percussive sounds than non-percussive.
   The overhead-microphones are expected to catch more percussive drum sounds even if there are interferences coming from non-percussive sources.

C) high onset/offset coincidence with the other drum tracks.
   If a drum is played, the sound is caught by a dedicated microphone located close to the drum, as well as the overhead microphones. This results in a high onset/offset coincidence between the overhead-microphone tracks and the other drum tracks.

Another advantage of the proposed system is being designed to be general and robust against any input source as it only requires a general classifier between percussive sound and non-percussive sound.

2. SYSTEM DESCRIPTION

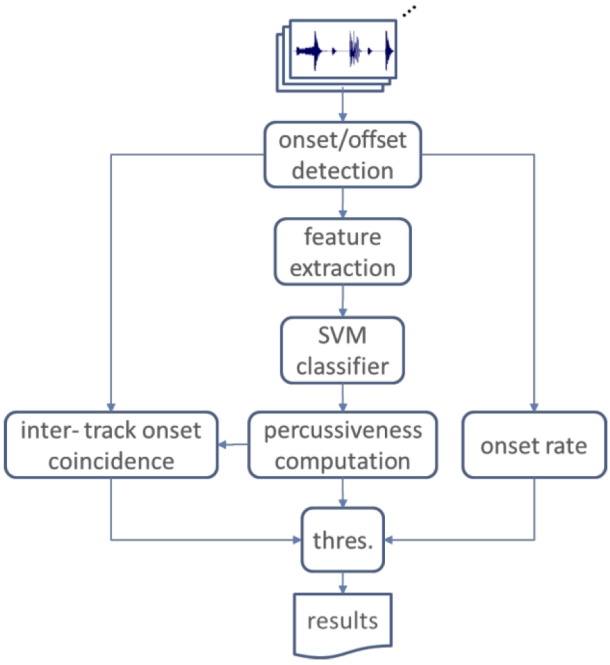

Figure 1. System overview.

Fig.1 illustrates the overview of the proposed system. The audio signals have onset/offset detection applied as a first step. The onset detection is done by a simple method based on temporal envelope. Onset is identified as a minimum of the envelope function just prior to a peak in the envelope function. Small peaks in the envelop function are ignored by a threshold in order to cancel the contribution of noise and small interferences coming from other sources. The corresponding offset is identified at a time location which first satisfies either of i) start of the next onset, or, ii) envelope becomes less than the 60% of a previous peak.

The audio signal between onset and offset is sliced into frames with hop size of 128 samples and FFT size of 1024 samples at a sampling rate of 12 kHz. 13-dimensional MFCC vectors are computed for each frame and they are finally summarized into a set of 65-dimensional features:

- MFCC average over frames (13 dimensions),
- MFCC variance over frames (13 dimensions),
- Initial 3 DCT coefficients obtained by applying the DCT against the initial 8 frames of MFCC value for each MFCC dimension. (3 × 13 = 39 dimensions).

The last aims to capture the temporal behavior of timbre in the vicinity of attack [5].

The 65-dimensional feature vector is fed to a SVM classifier that is to give a binary prediction if a sound belongs to percussive sound or non-percussive sound. The training is done beforehand using 3626 instances (1813 instances for each class). All the training data comes from clean sound without any interference. They are the combination of the RWC musical instrument sound database [6], home recordings and commercial libraries. Train and test with 10-fold cross validation achieved 96.73% average accuracy in F-measure.

The outputs of the SVM classifier are summarized into "percussiveness", a ratio for the number of percussive onsets and non-percussive onsets in a track.

The "percussiveness" is utilized to emphasize the contribution of percussive tracks at the computation of "inter-track onset coincidence". This is a summary of the onset-offset overlap durations against all the other tracks.

The "onset rate" is the number of onsets in a track normalized by that of the track having most onsets.

A track is finally identified to be drum overhead-microphone if it clears all the thresholds for the three criteria corresponding to A-C described in the previous section.

3. EXPERIMENT

An experiment was performed against 21 multi-track recordings in the format of sampling rate = 12 kHz, bit depth = 8 bit. They are all those recordings that include at least a drum overhead-microphone track out of multi-track recordings available from [7] in wav format. The tracks named "room", "MonoDr" and "beat" were eliminated beforehand in order to simplify the experiment. Finally 42 drum overhead-microphone tracks and 458 "the rest" tracks were selected out of the 21 recordings.

After parameter tuning on the threshold for A-C, the system finally achieved a recall of 40 / 42 = 95.23% on the overhead-microphone tracks and a recall of 422 / 458 = 92.14% on the rest of the tracks. Note that neither precision nor F-measure makes sense in this case as the definition implicitly assumes that positive and negative instances are equally balanced in the test data. However, it is not the case in this experiment as the ratio is 42:458.

The 458 – 422 = 36 of the errors on "the rest" come from the Snare (18), Hi-Hat (12), Percussion (3), Guitar (2), Bass (1) tracks, respectively. The reason for the error in the Snare track is because the signals observed at overhead-microphones sometimes get close to that of the Snare track depending on the phrase and microphone arrangement. On the other hand, the reason for Hi-Hat is because it catches both Hi-Hat and Snare due to the microphone that is usually located close to Snare. Therefore, the observation in a Hi-Hat track becomes like a mixture of both Hi-Hat and Snare that is likely to get close to the observation at an overhead-microphone.

4. CONCLUSION

A system for automatic identification of drum overhead-microphone track is proposed. The system finally achieved a recall of 95.23% on the overhead-microphone tracks and a recall of 92.14% on the rest of the tracks.

5. REFERENCES

[1] P. Herrera, A. Dehamel and F. Gouyon, "Automatic labeling of unpitched percussion sounds," 114th Convention of the Audio Engineering Society, Amsterdam, The Netherlands, pp. 1 – 14, 2003.

[2] C. Dittmar and D. Gartner, "Real-Time Transcription and Separation of Drum Recordings Based on NMF Decomposition," Proc. of the 17th International Conference on Digital Audio Effects, pp. 187–194, 2014.

[3] D. Ronan, H. Gunes, D. Moffat and J. D. Reiss, "Automatic Subgrouping of Multitrack Audio," Proc. of the 18th International Conference on Digital Audio Effects, pp. 187–194, 2015.

[4] J. Scott and Y. E. Kim, "Instrument Identification Informed Multi-track Mixing," Proc. of the 14th International Society for Music Information Retrieval, pp. 305-310, 2013.

[5] R. Marxer and H. Purwins, "Unsupervised Incremental Online Learning and Prediction of Musical Audio Signals," IEEE/ACM Transactions on Audio, Speech, and Language Processing (Volume: 24, Issue: 5), pp. 863 – 874, 2016.

[6] M. Goto, H. Hashiguchi, T. Nishimura, and R. Oka, "RWC music database: Music genre database and musical instrument sound database," Proc. of the 4th International Conference on Music Information Retrieval, pp. 229 – 230, 2003.

[7] B. De Man, M. Mora-Mcginity, G. Fazekas and J. D. Reiss, "The Open Multitrack Testbed," 137th Convention of the Audio Engineering Society, Los Angeles, USA, 2014.