IJCSC International Journal of Computer Science & Communication Vol. 1, No. 1, January-June 2010, pp. 219-222

# Separation of Tabla from Singing Voice using Percussive Feature Detection

Parveen Lehana[1] Neeraj Dubey[2] & Maitreyee Dutta[3]

[1]Dept of Physics and Electronics, University of Jammu
[2]Dept of CSE, GCET, Jammu
[3]Dept of CSE, NITTTR, Chandigarh
[2]singneeraj@rediffmail.com

ABSTRACT

In many signal processing applications, different sources of sound are to be separated for further modifications. Here, a method for the separation of tabla sound from a mixer of vocal and tabla is presented. For this, the short-time Fourier transform (STFT) of the mixed signal is taken. The log difference of each frequency component between consecutive frames in the magnitude spectra is obtained. If the log difference of the magnitude of frequency components exceeds a user specified threshold value (Th) the bin corresponding to that position is incremented by '1'. If the threshold condition is met, it is deemed to belong to a percussive onset. The final value of this counter, once each frequency bin has been analyzed, is then taken to be a measure of percussivity of the current frame. Once all frames have been processed, we have a temporal profile which describes the percussion characteristics of the signal. This profile is then used to modulate the spectrogram before resynthesis. The magnitude of all the frequency components in the consecutive frames for which the log difference exceeds threshold value is then added with the original phase. The inverse FFT is then computed of the resultant signal to generate a signal which is none other than separated tabla signal. In addition to producing high quality separation results, the method we describe is also a useful pre-process for tabla transcription in the mixer of tabla and vocal. Although the separated tabla sound does not contain any residual of vocal sound, the quality of the sound needs to be further enhanced.

## 1. INTRODUCTION

Only a few systems directly address the separation of music instrument from singing voice. A system proposed by meron and hirose [1] aims to separate piano accompaniment from singing voice. In recent years, some focus has shifted from pitched instrument transcription to drum transcription [2]; and likewise in the field of sound source separation, some particular attention has been given to drum separation in the presence of pitched instruments [3]. Algorithms such as ADRess [4] and those described in [5] are capable of drum separation in stereo signals if certain constraints are met. Other algorithms such as [6] [7] have attempted drum separation from single polyphonic mixture signals with varying results. In this paper we present a fast and efficient way to decompose a spectrogram using a simple technique which involves percussive feature detection which results in the extraction of the tabla parts from a polyphonic mixture of singing voice. The algorithm is applicable for the separation of almost any audio features which exhibit rapid broadband fluctuations such as tabla in music or plosives, fricatives and transients in speech. Automatic tabla separation and transcription is in itself can be a useful tool in applications such as processing of old song records.

In the present work we investigate audio features suitable for use in a threshold based detector to detect tabla segments from a mixture of tabla and singing voice.

## 2. METHOD OVERVIEW

Tabla used in popular music can be characterized by a rapid broadband rise in energy followed by a fast decay. The tabla consists of a pair of drums, one large base drum, the bayan, and a smaller treble drum, the dayan. Tabla percussion consists of a variety of strokes, often played in rapid succession, each labeled with a mnemonic. Two broad classes of strokes, in terms of acoustic characteristics, are: 1. tonal strokes that decay slowly and have a near-harmonic spectral structure and 2. impulsive strokes that decay rapidly and have a noisy spectral structure. The pitch percept by tonal tabla strokes falls within the pitch range of the human singing voice. It was found that while all the impulsive strokes had similar acoustic characteristics, there was a large variability in those of the different tonal strokes. On the other hand the pitch dynamics (the evolution of pitch in time) of singing voice tends to be piece-wise constant with abrupt pitch changes in between. A percussive temporal profile is derived by computing STFT of the signal per frame and assigning a percussive measure to it. The frame is then

scaled according to this measure. It should be seen that regions of the spectrogram with low percussive measures will be scaled down significantly. Upon resynthesis, only the percussive regions remain.

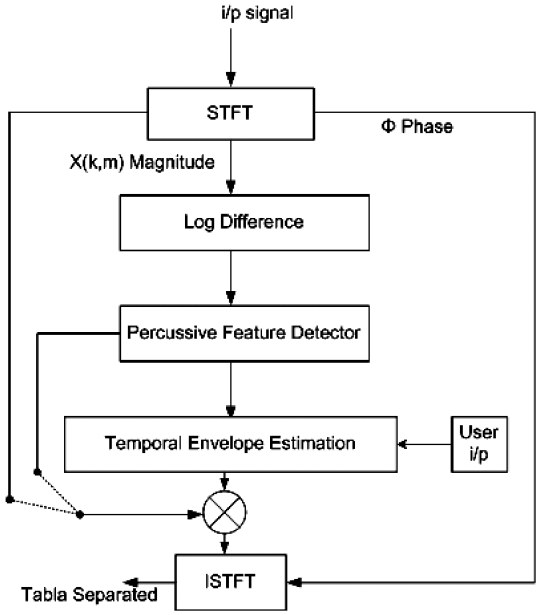


**Fig 1: System Overview**

The above Fig.1: demonstrates the general operation of the algorithm. The magnitude STFT of the mixture (Tabla+Vocal) is taken, while the phase ϕ is retained for resynthesis purposes at the later stage. The log difference of each frequency component between consecutive frames in the magnitude spectra is obtained. This measure effectively tells us how rapidly the spectrogram is fluctuating. If the log difference of the magnitude of frequency components exceeds a user specified threshold value ($T_h$) the bin corresponding to that position is incremented by '1'. If the threshold condition is met, it is deemed to belong to a percussive onset and a counter is incremented. The final value of this counter, once each frequency bin has been analyzed, is then taken to be a measure of percussivity of the current frame. Once all frames have been processed, we have a temporal profile which describes the percussion characteristics of the signal. This profile is then used to modulate the spectrogram before resynthesis. The magnitude of all the frequency components in the consecutive frames for which the log difference exceeds threshold value is then added with the original phase. The inverse FFT is then computed of the resultant signal to generate a signal which is none other than separated tabla signal.

## 3. SYSTEM IMPLEMENTATION

The given i/p signal comprises of a mixture of vocal & tabla signal. The short-time Fourier transform (STFT) of the mixed signal is taken. The magnitude $X(k, m)$ STFT (Short Time Fourier Transform) of the given input signal is taken and the phase (Φ) is retained for resynthesis purposes later on

$$X(k,m) = abs[\sum_{n=0}^{N-1} w(n)x(n+mH)e^{-j2\pi nk/N} \quad (1)$$

where $X(k, m)$ is the absolute value of the complex STFT given in equation 1 and where $m$ is the time frame index, $k$ is the frequency bin index, $H$ is the hop size between frames and $N$ is the FFT window size and where $w(n)$ is a suitable window of length $N$ also.

The log difference of each frequency component between consecutive frames in the magnitude spectra is obtained.

$$X'(k,m) = 20\log_{10}\frac{X(k,m-1)}{X(k,m)} \quad (2)$$

For all $m$ and $1 \le k \le K$

If the log difference of the magnitude of frequency components exceeds a user specified threshold value ($T_h$) the bin corresponding to that position is incremented by '1'. If the threshold condition is met, it is deemed to belong to a percussive onset and a counter is incremented.

$$Pe(m) = \sum_{k=1}^{K} \begin{bmatrix} P(k,m)=1 & if \quad X'(k,m)>T_h \\ P(k,m)=0 & otherwise \end{bmatrix} \quad (3)$$

The Final value of this counter once each frequency bin has been analyzed is then taken to be a measure of percussivity of the current frame. Once all frames have been processed, we have a temporal profile which describes the percussion characteristics of the signal. This profile is then used to modulate the spectrogram before resynthesis.

$$Y(k,m) = Pe(m)^{\Psi} X(k,m) \quad (4)$$

For all $m$ and $1 \le k \le K$

The magnitude of all the frequency components in the consecutive frames for which the log difference exceeds threshold value is then added with the original phase (Φ).

$$Y(k,m) = Pe(m)^{\Psi} X(k,m)P(k,m) \quad (5)$$

The inverse FFT is then computed of the resultant signal to generate a signal which is none other than separated Tabla signal.

$$Y(n+mH) = w(n)\left[\frac{1}{K}\sum_{k=1}^{K} Y(k,m).e^{j\angle x_{\omega}(k,m)}\right]^{norm} \quad (6)$$

## 4. RESULTS

The investigations were carried out using different proportions of tabla and vocal sounds. The analysis of the results showed that it is possible to extract the tabla sound from a mixture of tabla and vocal sounds by adjusting the parameters ($T_h$ and Ψ) in the modification and resynthesis

equations. The value of the threshold ($T_h$) was fixed at 15. It was observed that the value of Ψ was dependent upon the proportion of tabla and vocal sounds. For main tabla and supporting vocal, Ψ was to be adjusted as 0.45. Similarly, for main vocal and supporting tabla, the value of Ψ was to be adjusted as 0.1 for satisfactory quality of the output. It was also found that for increasing proportion of tabla sound in the mixer, the value of Ψ had to be increased.

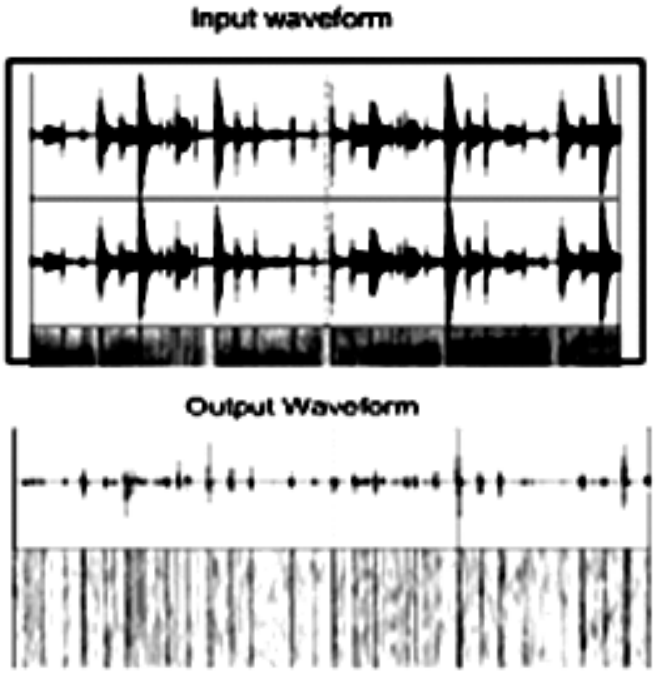


**Fig 2: The Plot Shows the Original i/p file "Main Tabla and Supporting Vocal" and the Tabla Separation Which Resulted.**

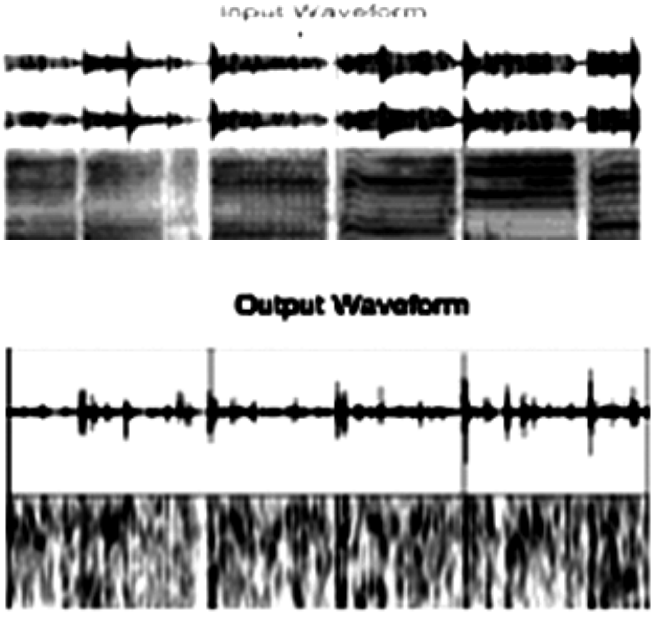


**Fig 3: The Plot Shows the Original i/p file "Equal Tabla Equal Vocal" and the Tabla Separation Which Resulted.**

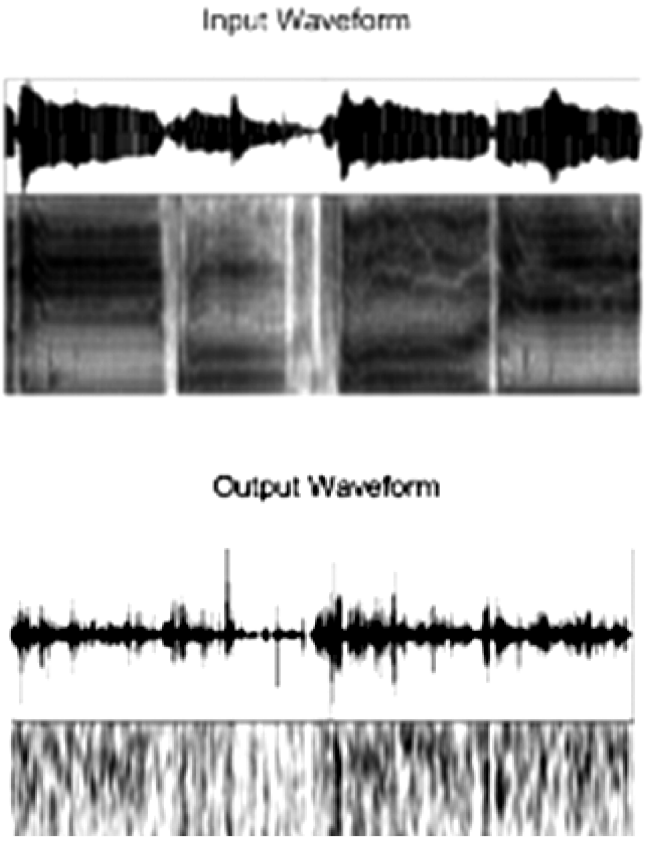


**Fig 4: The Plot Shows the Original i/p file "Main Vocal and Supporting Tabla" and the Tabla Separation Which Resulted.**

## 5. CONCLUSIONS

In addition to producing high quality separation results, the method we describe is also a useful pre-process for tabla transcription in the mixer of tabla and vocal. Although the separated tabla sound does not contain any residual of vocal sound, the quality of the sound needs to be further enhanced. The enhancement of quality and investigations of the effect of random noise on the tabla separation is on our future plan.

## REFERENCES

[1] Y Meron and K. Hirose, "Separation of Singing and Piano Sounds", *in Proc. of the 5th International Conference on Spoken Language Processing*, 1998.

[2] D. FitzGerald, E. Coyle, and B. Lawlor, "Sub-band Independent Subspace Analysis for Drum Transcription", *in Proc. Digital Audio Effects Conference*, Hamburg, pp. 65-69, 2002.

[3] D. FitzGerald, E. Coyle, and B. Lawlor, "Drum Transcription in the Presence of Pitched Instruments Using Prior Subspace Analysis", *in Proc. Irish Signals and Systems Conference 2003*, Limerick. July 1-2 2003.

[4] D. Barry, R. Lawlor, and E. Coyle, "Real-time Sound Source Separation Using Azimuth Discrimination and Resynthesis", *in Proc. Audio Engineering Society Convention*, October 28-31, San Francisco, CA, USA, 2004.

[5] C. Avendano, "Frequency Domain Source Identification and Manipulation in Stereo Mixes for Enhancement, Suppression and Re-panning Applications", *in Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, pp. 55-58 New Paltz, NY, October 19-22, 2003.

[6] A. Zils, F. Pachet, O. Delerue, and F. Gouyon, "Automatic Extraction of Drum Tracks from Polyphonic Music Signals", *in Proc. of the 2nd International Conference on Web Delivering of Music*, Darmstadt, Germany, Dec. 9-11, 2002.

[7] C. Uhle, C. Dittmar, and T. Sporer, "Extraction of Drum Tracks from Polyphonic Music Using Independent Subspace Analysis", *in Proc. of 4th International Symposium on Independent Component Analysis and Blind Signal Separation*, April 2003, Nara, Japan.

[8] P. Masri, and A. Bateman, 1996. "Improved Modeling of Attack Transients in Music Analysis Resynthesis", *in Proc. International Computer Music Conference*, pp. 100-103, 1996.

[9] D. Barry, R. Lawlor, and E. Coyle, "Comparison of Signal Reconstruction Methods for the Azimuth Discrimination and Resynthesis Algorithm", *in Proc. 118th Audio Engineering Society Convention*, May 28-31, Barcelona, Spain, 2005.

[10] D. W. Griffin, J.S. Lim, "Signal Estimation from Modified Short-time Fourier Transform", *IEEE Transactions on Acoustics*, Speech, and Signal Processing, vol. ASSP-32, no. 2, April 1984.

[11] D. Barry, R. Lawlor, and E. Coyle, "Drum Source Separation using Percussive Feature Detection and Spectral Modulation", in ISSC 2005,Dublin, September 1-2 2005.