



## Investigation of the Quality of Decomposed Signals from a Mixture of Percussive and Non-Percussive Instrumental Sounds

**Alpana Gupta**

Student, M.Tech.  
PTU, India

**Jaswinder Singh**

Department of Electronics  
BCET Gurdaspur, India

**Parveen Lehana\***

Department of Physics & Electronics  
University of Jammu, India

---

**Abstract**—Separation of musical signals from a mixture of percussive and non-percussive instrumental sounds has been a hot area of research over the past several years but this has still remained a complex and challenging job, especially when quality of the separated sound becomes an important criterion. In this paper we investigate the quality of separated percussive sound from a mixture of sounds of percussive and non-percussive instruments using flexible audio source separation toolbox (FASST) as a function of mixture composition. The investigations were carried out using spectrograms, time domain waveforms, and source to distortion ratio (SDR). The analysis of the results showed that the quality of the separated Tabla sound is satisfactory.

**Keywords**— Separation of sound, percussive instruments, non-percussive instruments, flexible audio source separation toolbox

---

### I. INTRODUCTION

Percussive instruments such as Tabla and Drum are those musical instruments which produce sound when they are struck or stroked by an object. Essentially it's a collision of the instrument with another body which generates sound in rhythm. Non-percussive instruments such as Banjo and Flute on other hand create melody rather than rhythm and they do not require a hitting on the instrument to produce sound [1]. These musical instruments have been in use for a long time and numerous source separation methods have been used to separate individual signals from the mixture of recorded musical sources. Obtaining the individuals sounds from a mixture has been motivation for developing algorithms whose target is not only to separate sounds but also to have satisfactory quality of the separated sound [2].

Separation of sound from a mixture of percussive and non-percussive instruments has been a challenge even though various methods have been invented and have been used. Percussive feature detection [3] has been used to identify drum sound from a polyphonic sound. The frequency domain techniques [4] provide estimation of short-time magnitude spectrum, which is scaled according to estimated time-amplitude. In [3], separation of drum sources from a mixture of polyphonic sound was investigated. Using some of these methods, separation of sound has been carried out, but desired quality of the separated sound could not be accomplished. In [5], using percussive feature detection method, separation of Tabla sound from singing voice was processed and it was reported that separated Tabla sound did not contain any residual vocal sound, but the quality of sound was not satisfactory.

Human beings have the ability to focus their listening attention to a single person in a conversation taking place in a group of people where there is a lot of noise and mixture of sounds - this is commonly known as the cocktail party effect [6] [7]. Many factors such as spatial distances between the sources, differences in pitch, and sound quality have been investigated to segregate mixtures of different sources. One such method is blind source separation (BSS) that provides an estimation of the component of the mixture of sound by using a minimum of prior information [8]. Independent component analysis (ICA) is a special type of approach to BSS, which assumes that individual source signals of a mixture are statistically independent. In ICA, it is assumed that the sources are not co-related and are essentially non-Gaussian.

In [9], information-theoretic approach and a projection pursuit approach have been used for linear ICA, exploring fixed-point algorithm's properties. The decomposition of sound from a mixture is not easy and the complexity increases when the recoding is carried out from large music groups; in such case mixing process becomes difficult since sound of each instrument bleeds to the other microphones as the recorded instruments are not perfectly isolated. To overcome this, Alex, Aaron and Garrett used ICA method [10] to separate signals from different mixes, but the results obtained were not satisfactory. Their method of separation relied upon the volume decay and propagation delay present in the recording of multiple instruments. Using various approaches for ICA is a cumbersome task; it takes too much of effort and it is also time consuming as well; importantly desired quality level of separated sound is difficult to achieve.

The problem of separation of percussive sound from a mixture has remained a difficult task. There is high amount of effort and time required to separate the sound. The decomposed signal often lacks in quality. There are other problems as well such as problem of convergence and dimensionality [11].

Source separation problems require detailed knowledge of the problem. This makes the process complicated and since there are no generalized steps used, the separation methods cannot be made general and one has to create these steps every time when one executes a new source separation, this makes reuse of the method difficult and cumbersome, which

ultimately leads to spending of extra efforts and time. As shown in Figure 1, for each source separation of sound, the process starts by model design wherein problem formulation is carried out. Then the algorithm is designed and implemented. The mixture is applied to the system and output is obtained. If the output is satisfactory, the steps are repeated. Flexible audio source separation toolbox (FASST) [11] solves the problem of a source separation by generalizing existing methods. This framework improves time-consuming steps and provides flexibility to incorporate prior knowledge of a particular problem. The block diagram of the FASST is shown in Figure 2.

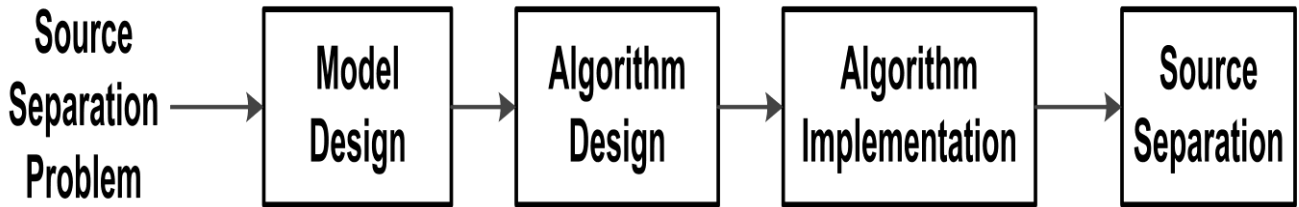


Figure 1: Current Approach for separating sound [11]

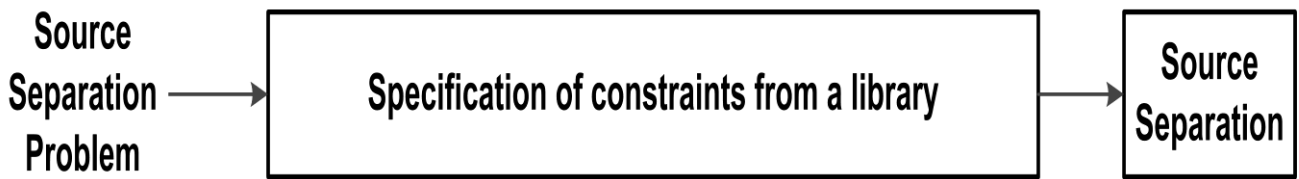


Figure 2: FASST Framework [11]



Figure 3: Tabla



Figure 4: Banjo

The objective of this paper is to explore the application of FASST for the separation of the musical sounds of Tabla (Figure 3) from the mixture of Tabla and Banjo (Figure 4) as a function of mixture proportion. Tabla is a popular percussive instrument which consists of two drums, one large drum called Bayan and small treble drum called Dayan. Variety of strokes often played in rapid succession generates Tabla percussion and these results into rise in energy followed by a fast decay. Banjo is a string based non-percussive instrument, popular in various parts of world. The methodology of the investigation is given in the following section. The results are presented in Section III and the conclusions drawn from the investigations are given in Section IV.

## II. METHODOLOGY

The musical sounds of Tabla and Banjo were recorded at a sampling frequency of 16 kHz and quantisation of 16 bits in an acoustically treated room using a data acquisition system. The musical signals recorded were mixed in different proportions varying from 0.1 to 0.9. The mixtures were applied to FASST system and separated sounds were stored in the computer. This process is shown in Figure 5. The mixture when passed to FASST framework [12] which was based on local Gaussian model [13] [14] gets decomposed into its individual components.

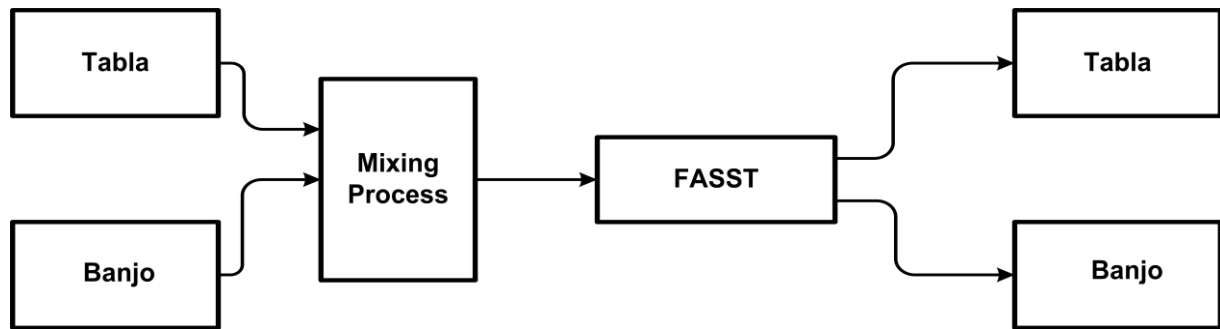


Figure 5: Separation of individual sounds using FASST

FASST solves problem of M-step by optimizing different parameter subsets alternatively; in addition to this it takes into account the corresponding constraints and makes use of nonnegative matrix factorization (NMF) methodology when using multiplicative update rules [11]. All of these attribute to the closed form update equations as proposed by Gaussianity. The framework is able to solve M-step problem with the help of a generalized expectation-maximization (GEM) [15] algorithm; on which this framework is implemented. FASST contains library of components by making use of audio-specific structures for source spectral power and it makes an assumption that in a given region of space and time, source power is constant. For NMF-like separation of sources, it proposes a new structure where nonnegative linear combination of time-localized temporal patterns represents each spectral pattern. It allows constraining of spectral patterns as harmonic, in-harmonic or noise-like. For quadratic time-frequency representations in Gaussian modelling, it provides a proper probabilistic formulation.

The investigation of the quality of separated Tabla sound was carried out using spectrograms, time domain waveforms, informal listening, and source to distortion ratio (SDR). SDR evaluates performance of a separation algorithm by considering gain or filtering of indeterminacies of BSS algorithms and in addition to this, it also takes into account the amount of residual crosstalk in each estimated source [16]. For estimating SDR, the estimated source  $\hat{S}_j$  is written as

$$\hat{S}_j = S_t + e_{interf} + e_{noise} + e_{artif}$$

where  $S_j$  is the true source,  $S_t = f(s_j)$  is a version of  $S_j$  modified by an allowed distortion  $f \in F$  where  $F$  is a family of distortions,  $e_{interf}$  the interferences,  $e_{noise}$  the noise, and  $e_{artif}$  the artifacts error. Giving SDR by [16]

$$SDR = 10 \log_{10} \frac{\|S_t\|^2}{\|e_{interf} + e_{noise} + e_{artif}\|^2}$$

Table 1: SDR for different values of the weightages of Tabla and Banjo

S.N.	Weightage of		SDR
	Tabla	Banjo	
1	0.1	0.2	32.28
2	0.2	0.1	32.07
3	0.2	0.3	27.05
4	0.3	0.2	27.48
5	0.4	0.2	25.67
6	0.5	0.2	26.75
7	0.6	0.2	24.92
8	0.7	0.2	24.63
9	0.8	0.3	23.17
10	0.9	0.3	22.41

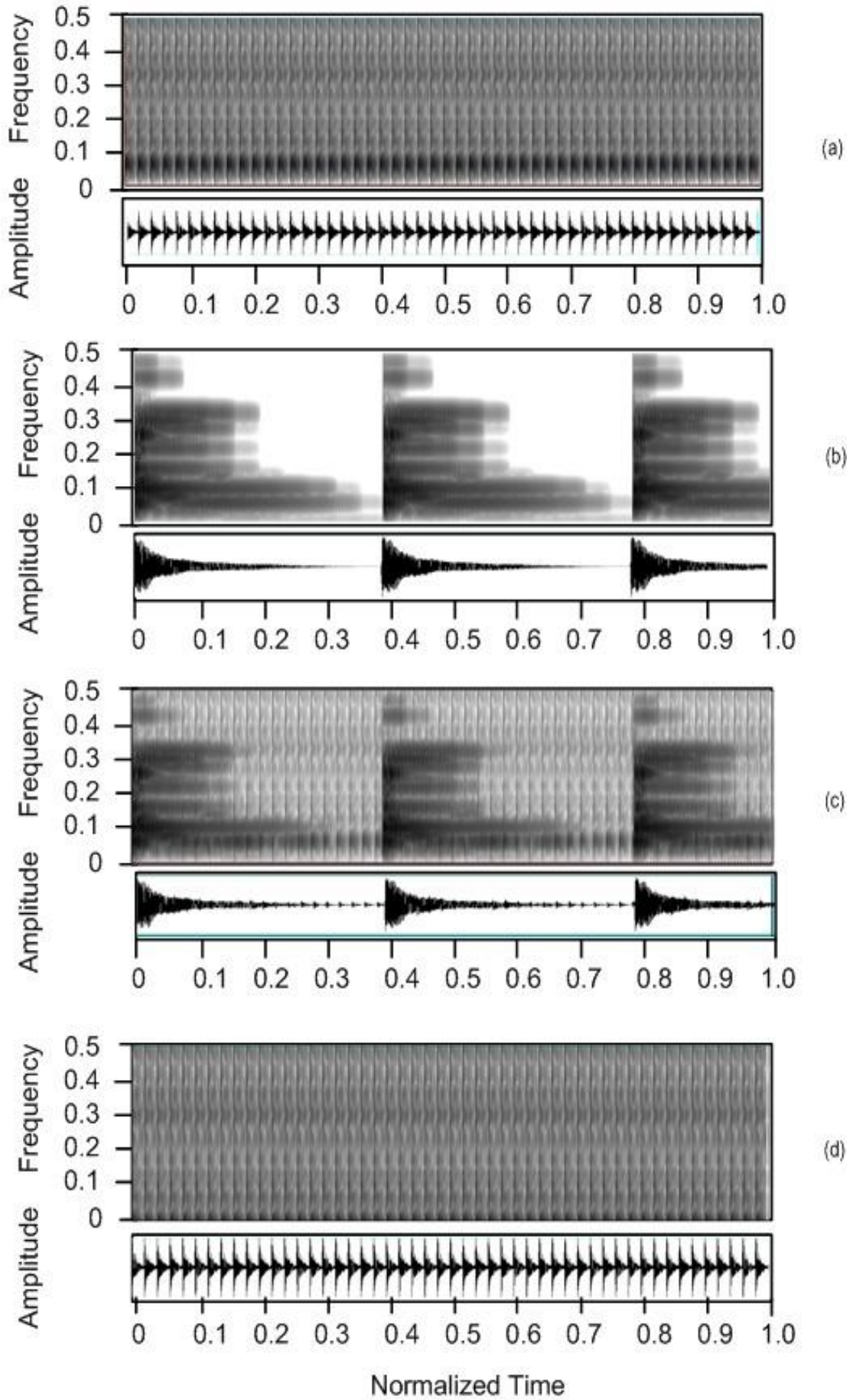


Figure 6: Spectrograms and waveforms of the recorded and processed sounds. a) Recorded Tabla, b) Recorded Banjo, c) Mixed Tabla and Banjo with weightage (0.1, 0.2), d) Decomposed Tabla.

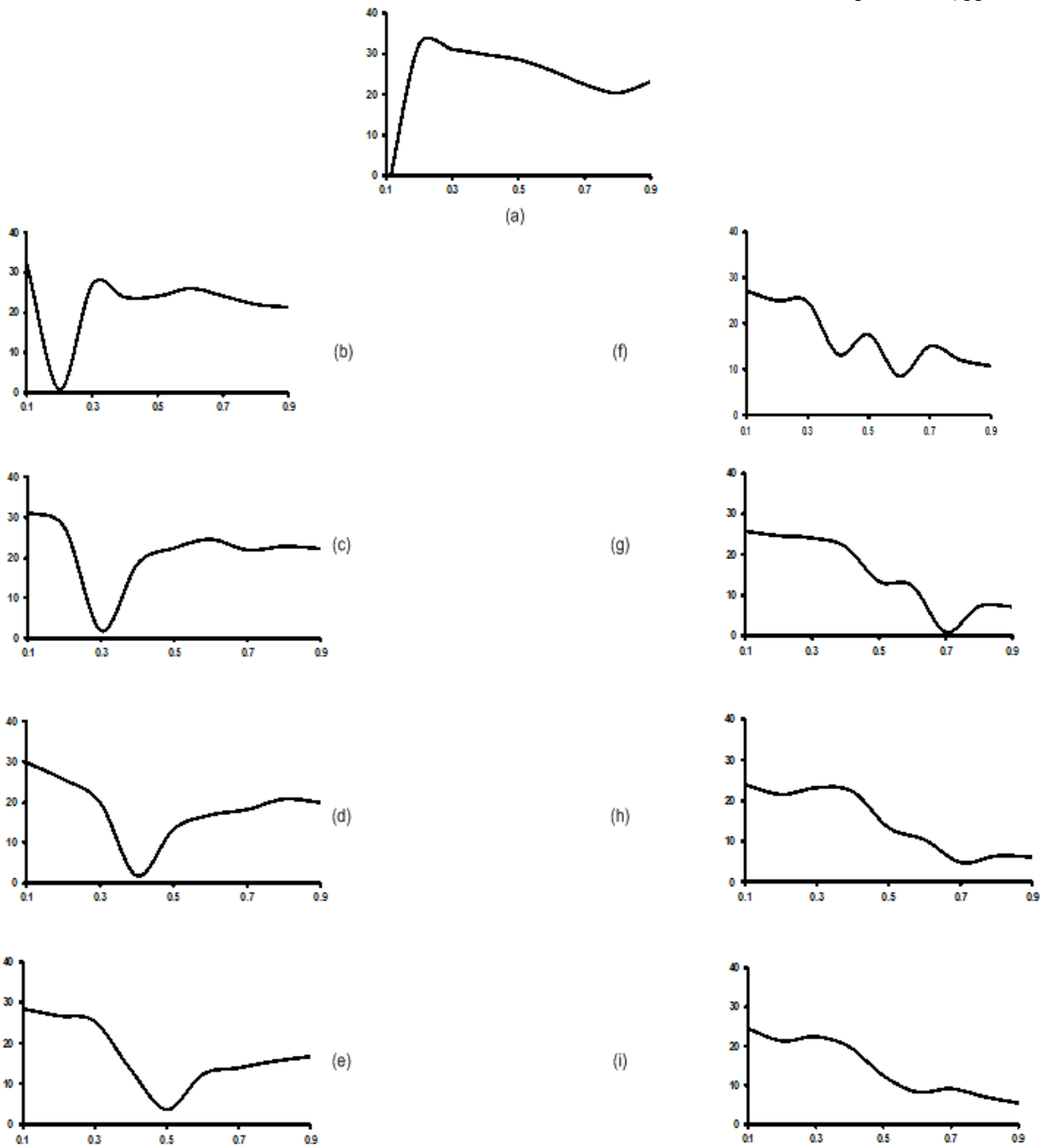


Figure 7: Variation of SDR (along y-axis) w.r.t. weightage of Banjo (along x-axis) for different proportions of Tabla. a) Tabla=0.1, b) Tabla=0.2, c) Tabla=0.3, d) Tabla=0.4, e) Tabla=0.5, f) Tabla=0.6, g) Tabla=0.7, h) Tabla=0.8, and i) Tabla=0.9.

### III. RESULTS

Spectrograms and waveforms of the recorded and processed sounds for the mixture proportion of Tabla and Banjo as (0.1,0.2) are shown in Figure 6. The analysis of the waveform and spectrogram shows that the Tabla sound has been satisfactory recovered from the mixture of Tabla and Banjo. Further, it was observed during the informal listening that quality of separated Tabla sound depends upon the weightage of the mixture proportions. Satisfactory quality was observed for the mixture with proportions as (0.1, 0.2). The estimated SDR for different propositions of mixture components is shown in Table 1 and plotted in Figure 7. It may be observed from these figures that for the mixture proposition of 0.1 of Tabla and 0.2 of Banjo, the value of SDR is maximum and for the proposition of 0.9 of Tabla and 0.3 of Banjo the value of SDR is minimum.

### IV. CONCLUSIONS

In this paper, investigations of the quality of the separated percussive sound from a mixture of percussive and non-percussive sounds using FASST were carried out. Investigations were carried out using spectrograms, time domain waveforms, informal listening tests, and SDR. The analysis of the results showed that the FASST is able to separate the Tabla sound with satisfactory quality provided the proportion of Banjo sound is low.

#### ACKNOWLEDGEMENT

This experiment was performed in DSP lab of Department of Physics and Electronics in University of Jammu and we are very thankful to research scholars of the lab for providing the help whenever we needed it.

#### REFERENCES

- [1] S. Abburu and S. G. Babu, "Indian Music Instruments Semantic Knowledge Representation," *Int. J. of Comput. Applicat.*, vol. 71, no. 15, 2013.
- [2] Y. Meron and K. Hirose, "Separation of singing and piano sounds," in *Proc. 5th Int. Conf. Spoken Language Process.*, 1998.
- [3] D. Barry, D. Fitzgerald, E. Coyle and R. Lawlor, "Drum source separation using percussive feature detection and spectral modulation," in *Proc. of the IEE Irish Signals and Syst. Conf.*, Dublin, 2005.
- [4] C. Avendano, "Frequency-domain source identification and manipulation in stereo mixes for enhancement, suppression and re-planning applications," in *IEEE Workshop on Applicat. of Signal Process. to Audio and Acoust.*, pp. 55-58, 2003.
- [5] N. Dubey, P. Lehana, and M. Dutta, "Separation of tabla from singing voice using percussive feature detection in a polyphonic channel," *Int. J. of Latest Trends Computing*, vol. 1, no. 2, 2010.
- [6] A. Hyvarinen and E. Oja, "Independent component analysis: algorithms and applications," *Neural Networks*, vol. 13, no. 4-5, pp. 411-430, 2000.
- [7] C. Uhle, C. Dittmar, and T. Sporer, "Extraction of drum tracks from polyphonic music using independent subspace analysis," in *Proc. 4th Int. Symp. on Independent Component Analysis and Blind Signal Separation*, 2003.
- [8] J. F. Cardoso, "Blind signal separation: statistical principles," *Proc. of the IEEE*, vol. 86, no.10, pp. 2009-2025, 1998
- [9] A. Hyvarinen: "Fast and robust fixed-point algorithm for independent component analysis", *IEEE Trans. on Neural Networks*, vol. 10, no. 3, pp. 626-634, 1999.
- [10] A. Favaro, A. Lewis, and G. Schlesinger, "ICA for musical signal separation," *CS 229 Machine Learning Final Projects, 2011*. [Online] Available: <http://cs229.stanford.edu/proj2011/FavaroLewisSchlesinger-IcaForMusicalSignalSeparation.pdf>
- [11] A. Ozerov, E. Vincent, and F. Bimbot, "A general flexible framework for the handling of prior information in audio source separation," *IEEE Trans. on Audio, Speech, and Language Process.*, vol. 20, no. 4, pp. 1118 – 1133, 2012.
- [12] "Flexible Audio Source Separation Toolbox (FASST)." [Online] Available: <http://bass-db.gforge.inria.fr/fasst/>
- [13] H. Attias: "New EM algorithms for source separation and deconvolution with a microphone array," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process.*, vol. 5, pp. 297–300, 2003.
- [14] C. Fevotte and J. F. Cardoso: "Maximum likelihood approach for blind audio source separation using time-frequency Gaussian source models," in *Proc. IEEE Workshop on Applicat. of Signal Process. to Audio and Acoust.*, pp. 78–81, 2005.
- [15] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. of Royal Statistical Soc.*, vol. 39, pp. 1–38, 1977.
- [16] E. Vincent, R. Gribonval, and C. Fevotte, "Performance measurement in blind audio source separation," *IEEE Trans. on Audio, Speech, and Language Process.*, vol. 14, no. 4, pp. 1462–1469, 2006.