# Blind Source Separation in Farsi Language by Using Hermitian Angle in Convolutive Enviroment

**Atefe Soltani(1) - Hamid Mahmoodian(2) - Ali Hashemi(3)**
(1) MsC – Department of Electrical Engineering, Najafabad Branch, Islamic Azad University
(2) Assistant Professor – Department of Electrical Engineering, Najafabad Branch, Islamic Azad University
(3) Assistant Professor – Department of Electrical Engineerin, Majlesi Branch, Islamic Azad University

Receive: Summer 2012 Accept: Spring 2013

**Abstract:** This paper presents a T-F masking method for convolutive blind source separation based on hermitian angle concept. The hermitian angle is calculated between T-F domain mixture vector and reference vector. Two different reference vectors are assumed for calculating two different hermitian angles, and then these angles are clustered with k-means or FCM method to estimate unmixing masks. The well-known permutation problem is solved based on k-means clustering of estimated masks which are partitioned to small groups. The experimental results show an improvement in performance when using two different reference vectors compared to only one.

**Index Terms:** Bind source separation (BSS); sparsity, w-disjoint orthogonality, hermitianangle;time-frequency masking.

**List of Symbols and Abbreviations**

| | |
|---|---|
| BSS | Blind Source Separation |
| COR (x, y) | Correlation between x and y |
| DFT | Discrete Fourier Transform |
| DOA | Direction of Arrival |
| DUET | Degenerate Un-mixing Estimation Technique |
| FCM | Fuzzy c-means |
| ICA | Independent Component Analysis |
| $R_{CM}$ | Correlation matrix calculated between C and M |
| SAR | Signal to Artifact Ratio |
| SCA | Sparse Component Analysis |
| SDR | Signal to Distortion Ratio |
| SIR | Signal to Interference Ratio |
| STFT | Short Time Fourier Transform |
| TF | Time-Frequency |

## 1. Introduction

This paper considers using hermitian angle in separating linear convolutive mixtures of audio signals on acoustic mixtures. The Blind Source Separation (BSS) or cocktail-party problem has been defined to characterize the task of retrieval speech in a room of simultaneous and independent speakers. Convolutive blind source separation (BSS) has often been used to solve the problem as it carries the promise to recover the sources exactly. The essential idea behind BSS is that on linear noise free acoustic model with multiple inputs (sources) and multiple outputs (sensors), original signals can be extracted under some reasonable assumptions with appropriately chosen multi-dimensional filters. Actually, this problem can be divided into time domain [1], [2] and frequency domain approaches [3], [4]. Many algorithms have been developed in different cases such as delay less model [5], [6], delayed source, noise free or with noise. Frequency domain solutions have been proposed based on many algorithms such as linear/circular convolution [7], [8], subband filtering [9] and or permutation ambiguity [10-14]. In this paper, T_F masking method based on hermitian vectors has been developed to make mask functions which is used in signal separation by clustering.

Suppose that source signals and microphone outputs aretermed as $s_1, s_2, \ldots, s_q$ and $x_1, x_2, \ldots, x_p$ respectively, convolutive BSS can be expressed as

$$X_p(n) = \sum_{q=1}^{Q} \sum_{l=0}^{L-1} h_{pq}(l) S_q(n-l) \quad (1)$$

where P is the number of microphones, Q is the source number ($q = 1, \ldots, Q$, $p = 1, \ldots, P$) and L is the mixing filter length.The output signal vectorsand $p^{th}$ microphone output samplesare shown as:

$$x = [x_1, x_2, \ldots, x_p]^T \ , x_p = [x_p(0), \ldots, x_p(N-1)]^T$$

where in this relation'T'is transpose operator and N is the number of total sample scolumn vector of sources. The $q^{th}$ source samples are defined as:

$$s = [s_1, s_2, \ldots, s_p]^T, \ s_q = [s_q(0), \ldots, s_q(N-1)]^T$$

The impulse response from $q^{th}$source to$p^{th}$microphone is supposed to be $h_{pq}(l), \quad l = 0, \ldots L-1$.

Overdetermined and underdetermined BSS are related to the number of sources and sensors (microphones) comparing to each other. Many algorithms have been proposed to solve BSS problems. In the overdetermined case, in many algorithms, the sources are assumed to be independent or at least decorrelated. The separation criteria can be divided in to the methods based on higher order statistics (HOS) and second order statistics (SOS)[15]-[19]. In the underdetermined BSS (P< Q),SCA is the most popular method. There are several methods [20]-[23] that work based on the sparseness of the source signals. If the signals are sufficiently sparse, it could be assumed that the sources rarely exist simultaneously. Therefore, each source could be estimated by gathering the observation samples that belong to one of the sources. In fact the original signals are extracted by applying appropriate masks to the mixtures in the TF domain. For speech signals, w-disjoint orthogonality property is the most used assumption in the TF domain. It means at most of the time only one source is dominant at any time-frequency slot.

DUET, (Degenerate Unmixing Estimation Technique), is one of the most popular methods in this field [24]. In this method an arbitrary number of sources could be separated from just two anechoic mixtures provided that the time-frequency transformation of sources do not overlap too much, which is true for speech signals.The ratio of the mixtures in T-Fdomain can be used to find two-dimensional weighted histogram. This histogram is used to cluster the mixing parameters, sofrequency variable and delay could be extracted and finally original sources could be extracted.

In [25], the direction of arrival (DOA) information is used to estimate the binary masks. $\theta_{DOA}(k,t)$ is estimated at each T-F point, then histogram of this angles is plotted for any point (k,t) and the peak of this histogram is taken as the DOAs of the sources. By using these angles, the binary masks are estimated for extracting sources.

Many algorithms have been developed based on the ratio of mixtures or their phase difference between observations in T-F domain and make some features where estimation of sources have done by clustering these features[26,27].These methods fail in a reverberant environment or when sources are very close or collinear. Also some algorithms need additional geometrical information which is compensated in the proposed method in this study. In addition, in our method the well-known scaling problem of the frequency domain algorithms is eliminated.

In this paper, hermitian angle method which has been proposed by [28], has been developed to separate the BSS between sources. Two hermitian angles are selected as the reference vectors to make masks used in separation. This paper organized as follows. In the second section, w-disjoint orthogonality property is explained. In the third section,the proposed algorithm for estimation of masks and solving permutation problem is described which is followed by experimental results in the fourth section and finally the conclusion is described in the last section.

## 2- Definition of W-Disjoinnnt Orthogon Ality

Let's w(t) be a given window function, if the short Fourier transform of two time –signals do not have overlap, then two time-signals are defined w-disjoint orthogonal. Since STFT is defined as:

$$\hat{s}_j(\tau, w) = F^w[s_j](\tau, w) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} w(t-\tau) S_j(t) e^{-iwt} dt$$

$$F(W) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} f(t) e^{-jwt} \, dt \quad (2)$$

w-disjoint orthogonality could be summarize as:

$$\hat{s}_j(\tau, w)\,\hat{s}_k(\tau, w) = 0 \ \forall \tau, w, \ \forall j \neq k, \hat{s}_j(\tau, w) \ \hat{s}_j(\tau, w) \neq 0 \quad (3)$$

This means that it is likely that each T-F point in mixture with significant energy is dominated by the contribution of only one source. If $S_j(w)$ becomes Fourier transform of $s_j(t)$ ,w-disjoint orthogonality is defined as:

$$\hat{s}_j(w)\,\hat{s}_k(w) = 0 \ \ \forall j \neq k \ \forall w \quad (4)$$

Therefore, using this property and finding binary masks, mixture could be separated into its component sources. For extracting any source its binary mask has been defined as follows:

$$M_j(\tau, w) := \begin{cases} 1 & \hat{s}_j(\tau, w) \neq 0 \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

where $M_j$ is a mask which separate $S_j$ from its mixture by

$$\hat{s}_j(\tau, w) = M_j(\tau, w)\hat{x}_1(\tau, w), \ \forall w \quad (6)$$

## 3. Proposed Algorithm

### 3.1. Transformation to Time-Frequency Domain

Fig.1 shows the flow of proposed method. First, time-domain mixture signals which are sampled at frequency $f_s$ are transformed in to time-frequency domain using STFT analysis. Time-Frequency transformation of equation (1) is:

$$X(k,t) = H(k)S(k,t) = \sum_{q=1}^{Q} H_q(k) S_q(k,t) \quad (7)$$

where X(k,t) is T-F transformation of microphone output vectors and S(k,t) is the STFT of source signals .i.e.:
$X(k,t) = [X_1(k,t), \ldots, X_p(k,t)]^T$,
$S(k,t) = [S_1(k,t), \ldots, S_p(k,t)]^T$
The impulse response H(k) and $q^{th}$source column vector of impulse response in the $k^{th}$frequency bin are:
$H(k) = [H_1(k), \ldots, H_q(k)]$,
$H_q(k) = [H_{1q}(k), \ldots, H_{pq}(k)]^T$
where $H_{pq}(k)$is the impulse response from $q^{th}$ source to $p^{th}$ microphone at the $k^{th}$ frequency bin. In our experiment Hanning window is used for STFT of time domain signals.

### 3.2. Feature Extraction

Suppose that the original sources are disjoint orthogonal, then the separation can be realized with gathering T-F points which belongs to only one source. To estimate such T-Fpoints one or more features should be calculated by using the time-frequency domain mixtures. In this study, hermitian angle in complex vector space between STFT of microphone outputs and two reference vectors have been considered as the features in the procedure. In this method the feature space has been expanded in two dimensions comparing with proposed method in [28], which one vector has been considered as the reference.
In complex space, the cosine of the complex-valued angle between two complex vectors is defined as[23]:

$$\cos(\theta_C) = \frac{u_1^H u_2}{\|u_1\|\|u_2\|} \quad , \ \|u\| = \sqrt{u^H u} \tag{8}$$

where$u_1$and $u_2$are two complex vectorsand H is the complex conjugate transpose operation.$\cos(\theta_C)$in (8) is defined as

$$\cos(\theta_C) = \rho e^{j\emptyset} \qquad , \rho \le 1 \tag{9}$$

where,

$$\rho = \cos(\theta_H) = |\cos(\theta_C)| \tag{10}$$

Using (9) and(10), two different angles aredefined as

$$0 \le \theta_H \le \frac{\pi}{2} \quad , \quad -\pi \le \phi \le \pi$$

where $\theta_H$ and $\emptyset$ are called hermitian and pseudo angle respectively between two complex vectors $u_1$ and $u_2$. It is proven that [28], if two complex vectors are multiplied by any complex scalars, the hermitian angle between them will not change, but pseudo angle willbe affected.In other words, multiplication of a complex vector by any complex scalars does not change the hermitian angle between that vector and another vector (reference vector). This concept can be extended toBSS problem,and design T-F masks. The hermitian angle between$H_q(k)$, impulse response vector and r, reference vector will not change if it multiplied by complex scalar(source signal $S_q(k,t)$). The same rule applies tohermitian angle between mixture vector, $X_p$ and reference vector, where reference vector consists of P random elements with all the elements equal to 1+j1.
For P mixtures and Q sources, the hermitian angle between each of the mixture vectors in the $k^{th}$ frequency bin and reference vector with p element is explained as:

$$\cos(\theta_C(k1,t1)) = \frac{X(k_1,t_1)H_r}{\|X(k_1,t_1)\|\|H_r\|} \tag{11}$$

$$\theta_H^{(k_1)}(t_1) = \cos^{-1}(|\cos(\theta_C(k_1,t_1))|) \tag{12}$$

The original source signals $S_q, q = 1,\ldots,Q$ are assumed sparse in T-F plane i.e. at any T-F point it is at most contribution of one source. The hermitian angle between mixture vector and reference vector is equal tothe hermitian angle between impulse response$H_q(k)$ and r reference vector, where $S_q(k,t)$is corresponding to the present source at that point (k,t). In our method, the hermitian angle is calculated by using two different reference vectors and therefore, two different features are generated. The expanded space of features should be used to recognize the sources of the samples.

$$\theta_{H1} = \text{hermitian} \ (X(k,t), r_1)$$

$$\theta_{H2} = \text{hermitian} \ (X(k,t), r_2)$$

Generally, for P microphones, P different reference vectors are taken for extracting P different features. In other words, in this way a P-dimensional space is made in such a way that any P dimensional vector could be estimated more accurately than using only one reference vector.

### 3.3. Mask Designing

It can be seen from (11) and(12), that the hermitian angle for any frequency bin, is a vector with ‘t’ different values, so thehermitian angle at $(k_1, t_1)$,point is different from hermitian angle at $(k_2, t_2)$.‘t’ is the number of time slots in T-F plane of mixture. Using Clustering algorithm to cluster the vector of hermitian angle in the $k^{th}$ frequency bin,Q different clusters are produced such that, the samples which belong to one cluster are components of one source.
Like other T-F masking methods, the membership function obtained from clustering is directly used as T-F mask. In our experiment, K-means and fuzzy c-means (FCM) are used for clustering of hermitian angles. Using k-means method, the result mask produced with k-means algorithm, has binary values (0,1) so it introduces reconstructed signals. FCM method is usedfor clustering the angles to produce softer and smoother masks with respect to the masks produced by k-means. After clustering the angles,Index vector of the k-means clustering ({0,1}) and membership function of the fuzzy c-means clustering ([0,1])have been used as masking values to extract the original sources from microphone outputs. It should be noted that permutation problem should be considered after this stage.

### 3.4. Permutation Problem

Actually, BSS of signals infrequency domain involves two important problems, the scaling and the permutation. Because of applying the generated T-F masks directly to the mixtures, scaling problem does not exist in our experiment. To address the permutation issue, many algorithms have been reported in the frequency domain[29],[30].
For speech signals in T-F domain, increasing the distance between two frequency bins decreases the correlation between them. Therefore, frequency bins are portioned to small groups that signal for one frequency group is more correlated which in turn solves the permutation problem[28].The groups are composed of few adjacent

frequency bins with overlap which in our experiment, 16 bins with 75% overlap are assumed. To solve the permutation problem for the groups, the masks are clustered by k-means method to minimize the following equation:

$$D = \sum_{q=1}^{Q} \sum_{\substack{M_i^k \in C_q \\ i=1,\dots,Q \\ k=k_{st},\dots,k_{end}}} \left(1 - r_{M_i^{(k)} C_q}\right) \quad (13)$$

where, $M_i(k)$ is the $i^{th}$ mask in the $k^{th}$ bin frequency, $C_q$ is the center of $q^{th}$ cluster, and $r_{M_i(K)C_q}$ is the correlation between them. The parameters, $k_{st}$ and $k_{end}$ are starting and ending frequency bins for each frequency group.The k-means categorizes the masks in each group into Q clusters based ondistance metric correlation to have high correlated masks in each. Ideally, for any frequency bin, each cluster has exactly one mask, but itdoes not occur in practical situation.For instance, at some frequency bins, some clusters do not have any mask and some clusters have more than one mask. In such bins the k-means is failed to cluster them correctly. In such cases, correlation between the centers and the masks has been calculated to maximize the summation of correlation values. The permutation matrix $\pi_k$ in the $k^{th}$ bin for failed clusters is defined as:

$$\pi_k = \arg_\pi \max \sum_i^F \sum_j^F (\pi . R_{CM})_{ij} \quad (14)$$

where '•' is the element wise product between two matrixes, F is the number of failed clusters, $\pi$ is a selected permutation matrix with elements 0 and 1, such that there is one '1' in each row or column. In this case permutation matrix could be selected as $\begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$ or $\begin{bmatrix}0 & 1\\1 & 0\end{bmatrix}$.

$R_{CM} \in R^{F \times F}$ is correlation matrix which it's elements $(R_{CM})_{ij}$ are Pearson correlation coefficient between $i^{th}$row of C and $j^{th}$row of M and q is a number which points to the index of failed clusters. C and M are respectivelythe center matrix and the mask matrix in the failed clusters in the $k^{th}$ frequency bin, which are defined as:

$C= [\dots, C_q^T,..]^T$ , $C_q \in R^{1*T}$ , $M= [\dots, M_q^T,..]^T$ , $M_q \in R^{1*T}$

Finally the masks matrix after solving permutation in the $k^{th}$ frequency bin will be $p_k M$.

### 3.5. Signal Reconstruction

Output signal reconstruction has been done after Time-Frequency mask generation which makes the separated signals as:

$$Y_q(k,t) = M_q(k,t) X_P(k,t), \forall t, q = 1, \dots, Q \quad (15)$$

where $X_p(k,t)$ is one of the mixtures(p=1 or 2). Finally, inverse STFT (overlap and add method reported in [31]) has been used to produce the separated signals and come back it to time domain.The generated signal is the estimation of original signals.

## 4. Experimental Results

### 4.1. Data set

For performance evaluation of the proposed method some criterions are needed. Room simulator software version[1] is used to simulate a room conditions and control the microphones and sources positions. Two sets of speech signals have been used to evaluate the proposed method. Five speech signals in the first set is randomly selected from TIMIT database and five speech signals in the second set have been produced in Farsi language in an acoustic room (16 bit with 48 KHz sampling rate) to analyze the model with Farsi language speakers.

Some needed parameters for room simulation are as follows: two sources and two microphones are considered, reverberation time is 120ms, order of reflection is 28 and two sources and two microphones are assumed where the distance between two microphones is assumed 28cm and 34cm.Hanning window has been used to segment the signals and transform it into frequency domain by 2048 DFT length.

### 4.2. Performance criterions

The separation quality is measured by method suggested in [32], [33]. In this method, the estimated signals are written as

$$y_q = y_{qtarget} + e_{interf} + e_{qartif} \quad (16)$$

where, $y_{qtarget}$ is target signal with allowed deformation, $e_{interf}$ is interference due to unwanted sources and $e_{qartif}$ is artifacts introduced by the separation algorithm. So the performance criterions are defined as

$$SDR = 10 \log_{10} \frac{\|y_{qtarget}\|^2}{\|e_{qinterf} + e_{qartif}\|^2} \quad (17)$$

$$SIR = 10 \log_{10} \frac{\|y_{qtarget}\|^2}{\|e_{qinterf}\|^2} \quad (18)$$

$$SAR = 10 \log_{10} \frac{\|y_{qtarget} + e_{qinterf}\|^2}{\|e_{qartif}\|^2} \quad (19)$$

SDR is source to distortion ratio, SAR is source to artifact ratio and SIR is source to interference ratio. Since the masks are applied to $p^{th}$ microphone output, the target signal is$y_{qtarget} = h_{pq} * s_q$, where $h_{pq}$ is the impulse response from $p^{th}$ source to $q^{th}$ microphone.In our experiment the masks have been applied to the first microphone output, so the target signals are $y_{1target} = h_{11} * s_1$ and $y_{2target} = h_{12} * s_2$

In our study, five speech signals with 5 seconds duration from TIMIT database have been used and the average of the output indices (SDR,SIR,SAR)is calculated and shown in table I and table II for two different microphones distance.

### 4.3. The comparison between one and two reference vectors

The reference vector is a vector with two elements 1+j, and the results of the simulation for two clustering methods have been shown in tables (I-IV). Based on our proposed method, the hermitian angle is calculated with respect to two different reference vectors.The elements of the first and second reference vectors have been selected as 1+j and random values.The resultant angles via one reference vector are different from angles by another reference vector. So, it could be assumed that two different features for clustering exist. The new masks are different from the masks generated with only one reference vector. The experimental results for two clustering methods are shown in table I and II for TIMIT database and in tables III and IV for Farsi database.

The comparison between one and two references vector shown in the tables, illustrates that the latter has improvement in both SDR and SAR by FCM clustering method although k-means has more improvement in SIR. In addition, comparing the results shows that the signal separation with 34cm microphone distance is better than one with 28cm. Besides, the results with two reference vectors have been improved with respect to one reference vector.

## 5. Conclusion

T-F masking method has been used for convolutive blind source separation of speech signals same as previous work [28]. In this study, hermitian angles with respect to two reference vectors has been proposed and results show the superiority of the SDR, SIR and SAR compare to previous work where hermitian angles with one reference vector has been used.By taking two different reference vectors, two different hermitian angles are found which could be clustered.The separation performance for proposed algorithm is measured with one and two reference vectors for two clustering methods, k-means and FCM.

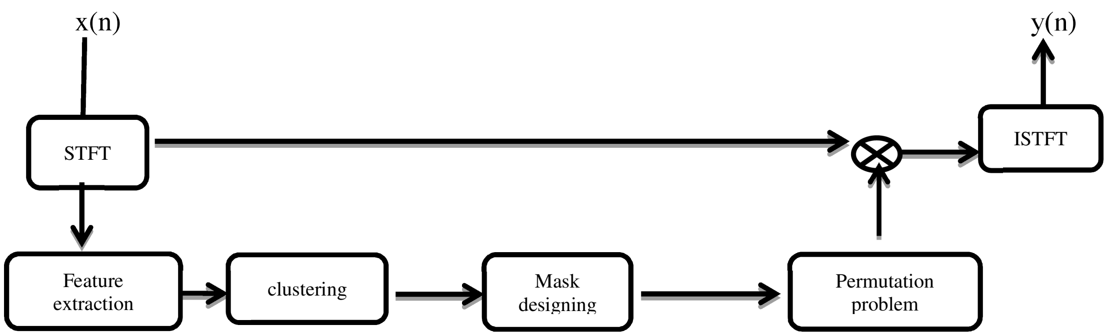


Fig. (1): System overview

Table (1): Separation performance for TIMIT (microphones distance= 28cm)

| Separation performance | Input | One reference vector | | | | Two different reference vectors | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | k-means | | FCM | | k-means | | FCM | |
| | | Output | Improvement | Output | Improvement | output | Improvement | output | Improvement |
| **SDR(db)** | 0.07 | 8.1 | 8.03 | 8.15 | 8.08 | 11.83 | 11.76 | 12.41 | **12.34** |
| **SIR(db)** | 0.07 | 12.98 | 12.92 | 11.5 | 11.08 | 17.13 | **17.06** | 16.52 | 16.45 |
| **SAR(db)** | 20 | 10.14 | -9.86 | 11.3 | -10.14 | 13.63 | -6.37 | 14.78 | **-5.22** |

Table (2): Separation performance for TIMIT (microphones distance=34cm )

| Separation performance | Input | One reference vector | | | | Two different reference vectors | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | k-means | | FCM | | k-means | | FCM | |
| | | Output | Improvement | Output | Improvement | output | Improvement | output | Improvement |
| **SDR(db)** | 0.08 | 8.73 | 8.65 | 8.75 | 8.67 | 11.87 | 11.79 | 12.53 | **12.45** |
| **SIR(db)** | 0.08 | 13.5 | 13.42 | 12.11 | 12.03 | 17.26 | **17.18** | 16.4 | 16.32 |
| **SAR(db)** | 20 | 10.75 | -9.25 | 11.52 | -8.48 | 13.63 | -6.37 | 14.84 | **-5.16** |

Table (3): Separation performance for Farsi data base(microphones distance= 28cm)

| Separation performance | Input | One reference vector | | | | Two different reference vectors | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | k-means | | FCM | | k-means | | FCM | |
| | | Output | Improvement | Output | Improvement | output | Improvement | output | Improvement |
| **SDR(db)** | 0.07 | 5.47 | 5.4 | 5.98 | 5.92 | 6.57 | 6.5 | 7.3 | **6.23** |
| **SIR(db)** | 0.07 | 9.44 | 9.37 | 9.03 | 8.96 | 10.92 | **10.85** | 10.18 | 10.11 |
| **SAR(db)** | 20 | 9.98 | -10.02 | 11.05 | -8.95 | 10.62 | -10.55 | 12.26 | **-7.74** |

Table (4): Separation performance for Farsi data base(microphones distance=34cm )

| Separation performance | Input | One reference vector | | | | Two different reference vectors | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | k-means | | FCM | | k-means | | FCM | |
| | | Output | Improvement | Output | Improvement | output | Improvement | output | Improvement |
| **SDR(db)** | 0.07 | 6.34 | 6.27 | 6.57 | 6.5 | 7.4 | 7.33 | 8.08 | **8.01** |
| **SIR(db)** | 0.07 | 10.7 | 10.63 | 9.69 | 9.62 | 12.5 | **12.43** | 11.23 | 11.16 |
| **SAR(db)** | 20 | 10.36 | -9.64 | 11.31 | -8.69 | 11.25 | -6.75 | 12.64 | **-7.36** |

## Appendix:

1. [Online]. Available: http://bass-db.gforge.inria.fr/BASS-dB/?show=browse & id=filters

## References


[1] D. Yellin, E. Weinstein, “Multichannel signal separation: methods and analysis, ” IEEE Transactions on Signal Processing, Vol. 44, pp. 106-118, Jan. 1996.

[2] H. Sahlin, H. Broman, “MIMO signal separation for r channels: a criterion and performance analysis, ” IEEE Transactions on Signal Processing, Vol. 48, pp. 642-649, March. 2000.

[3] J.K. Tugnait, “Adaptive blind separation of convolutive mixtures of independent linear signals,” Signal Processing, Vol. 73, pp. 139-152, 1999.

[4] J. Pesquet, B. Chen, A.P. Petropulu, , “Frequency-domain contrast functions for separation of convolutive mixtures, ” in Proc. ICASSP2001, May 2001.

[5] A. Hyv¨arinen, J. Karhunen, E. Oja, , “Independent Component Analysis”. Wiley, 2001.

[6] S. Roberts and R. Everson, “Independent Components Analysis: Principles and Practice”. Cambridge University Press, 2001

[7] V. Capdevielle, C. Servire, J.L. Lacoume, “Blind separation of wide-band sources in the frequency domain,” in ICASSP95, Vol. III, Detroit, MI, USA, May 9–12 1995, pp. 2080–2083.

[8] L. Parra, C. Spence, “Convolutive blind separation of non-stationary sources,” IEEE Trans. Speech Audio Proc., Vol. 8, No. 3, pp. 320–327, May 2000.

[9] H. Attias, “New EM algorithms for source separation and deconvolution,” in ICASSP’03, Vol. V, pp. 297–300, 2003.

[10] D.T. Pham, C. Servi`ere, H. Boumaraf, “Blind separation of convolutive audio mixtures using nonstationarity,” in ICA’03, 2003, pp. 981–986.

[11] F. Asano, S. Ikeda, M. Ogawa, H. Asoh, N. Kitawaki, “Combined approach of array processing and independent component analysis for blind separation of acoustic signals,” IEEE Trans. Speech Audio Proc., Vol. 11, No. 3, pp. 204–215, May 2003.

[12] F. Asano, S. Ikeda, M. Ogawa, H. Asoh, N. Kitawaki, “Blind source separation in reflective sound fields,” in HSC’01, Apr 9–11 2001.

[13] H. Attias, C.E. Schreiner, “Blind source separation and deconvolution: The dynamic component analysis algorithm,” Neural Computation, Vol. 11, pp. 803–852, 1998.

[14] W. Baumann, B.U. K¨ohler, D. Kolossa, R. Orglmeister, “Real time separation of convolutive mixtures,” in ICA’01, 2001, pp. 65–69.

[15] V. Capdevielle, C. Servire, J.L. Lacoume,“Blind separation of wide-band sources in the frequency domain,” in ICASSP95, Vol. III, Detroit, MI,USA, May 9–12 1995, pp. 2080–2083.

[16] S. Icart, R. Gautier, “Blind separation of convolutive mixtures using second and fourth order moments,” in ICASSP’96, Vol. 5, 1996, pp. 3018–3021.

[17] D. Yellin, E. Weinstein, “Criteria for multi channel signal separation,” IEEE Transactions on Signal Processing, Vol. 42, No. 8, pp. 2158–2168, Aug 1994.

[18] P. Comon, E. Moreau, L. Rota, “Blind separation of convolutive mixtures: A contrast-basedjoint diagonalization approach,” in ICA’01, 2001, pp. 686–691.

[19] J. Thomas, Y. Deville, S. Hosseini, “Time domain fast fixed-point algorithms for convolutiveICA,” IEEE Signal Processing Lett., Vol. 13, No. 4, pp. 228–231, Apr 2006.

[20] O. Yilmaz, S. Rickard, “Blind separation ofspeech mixtures via time-frequency masking,” IEEETransactions Signal Processing, Vol. 52, No. 7, pp. 1830–1847, Jul. 2004.

[21] N. Roman, “Auditory-based algorithms for sound segregation in multisource and reverberant environtments,”Ph.D. dissertation, The Ohio State University,Columbus, OH, 2005.

[22] A. Blin, S.Araki, S.Makino, “Underdetermined blind separation og convolutive mixtures of speechusing time-frequency mask and mixing matrix estimation,”IEICE Trans. Fundamentals, vol. E88-A,no. 7, pp. 1693–1700, Jul 2005.

[23] S. Xie, L.Yang, J.Yang, G.Zhou, Y. Xiang ,“Time frequency Approach To Underdetermined Blind Source Separation,” IEEE Transaction on neural networks and lerning systems, Vol. 23, No. 2, february 2012.

[24] S. Rickard, The DUET Blind Source Separation Algorithm, S. Makino et al. (eds.), Blind Speech Separation, Springer, 217–237, 2007.

[25] S. Araki, S. Makino, A. Blin, R. Mukai, H. Sawada, “Underdetermined blind separation for speech in real environments with sparseness and ICA,” in Proc. ICASSP, pp. III-881–III-884, May 2004.

[26] O¨. Yılmaz, S. Rickard, Blind separation of speech mixtures via time–frequency masking, IEEE Transactions on SP 52 (7) 1830–1847, 2004.

[27] A. Jourjine, S. Rickard, O¨. Yılmaz, Blind separation ofdisjoint orthogonal signals: demixing N sources from 2mixtures, in: Proceedings of the ICASSP 2000, Vol. 12, pp. 2985–2988, 2000.

[28] V.G. Reju, S.N. Koh, I.Y. Soon, "Underdetermined convolutive blind source separation via time-frequency masking", IEEE Trans. on Audio, Speech and Language Processing, Vol. 18, No. 1, pp. 101–116, Jan. 2010.

[29] S. Dubnov, J. Tabrikain, M. Arnon-Targan, "Amethod for directionally-disjoint source separationin convolutive environment," in ICASSP'04, Vol. V, pp. 489–492, 2004.
[30] A. Hiroe, "Solution of permutation problem in frequencydomain ICA, using multivariate probabilitydensity functions," in ICA'06, pp. 601–608, 2006.
[31] A.V. Oppenheim, R. W. Schafer, J.R. Buck, Discrete-Time SignalProcessing. Upper Saddle River, NJ: Prentice-Hall, 2003.
[32] E. Vincent, R. Gribonval, C. Fevotte, "Performance measurement in blind audio source separation," IEEE Trans. Audio, Speech, Lang. Process., Vol. 14, No. 4, pp. 1462–1469, Jul. 2006.
[33] C. Fevotte, R. Gribonval, E. Vincent, "BSS EVAL Toolbox User Guide, IRISA Tech. Report 1706", Tech. Rep. Rennes, France, 2005 [Online]. Available: http://www.irisa.fr/metiss/bss eval/