A Novel Algorithm for Source Localization Based on NonnegativeMatrix Factorization using $\alpha\beta$ -Divergence in Cochleagram

M. E. Abd El Aziz, Wael khedr
Department of Mathematics, Faculty of Science, Zagazig University, Zagazig 44519, Egypt
abd_el_aziz_m@yahoo.com

Abstract: In this paper, a localization framework based on a nonnegative matrix factorization using a family of αβ-divergence and cochleagram representation is introduced. This method provides accurate localization performance under very adverse acoustic conditions. The system consists of a three-stage analysis,the first stage: the source separation using NMF based on αβ-divergence where the decomposition performed in cochleagram domain. In the second stage the estimated mixing matrix used to estimate the Time Difference of arrival (TDOA). Finally the Time Difference of Arrival estimates can be exploited to localize the sources individually using the Scaled Conjugate Gradient algorithm (SCG) ,where SCG has advanced compared to other conjugated gradient algorithms. Experiments in real and simulated environments with different microphone setups in 2-D plane and 3-D space, are discussed, showing the validity of the proposed approach and comparing its performance with other state-of-the-art methods.

Keywords:Blind source separation (BSS), Nonnegative Matrix Factorization (NMF), αβ divergence, sound source localization (SSL).

1. Introduction

The problem of the source separation and localization using a microphone array is an important problem in multichannel speech signal processing with many applications, for instance, in teleconference systems , seismic, biomedicine, sonar, radar and communications [1,2,3]. There are several algorithms proposed for solving this problem and these mainly differ in two aspects: the type of signal features used and the way to find the clusters of these features. The normalized time-frequency samples were used as features in [1], their amplitudes and phases were used in [3,4], and Hermitian angle in [5]. A combination of using time-frequency masking and ICA was proposed in [6,7] for underdetermined blind source separation. This algorithm first aims to convert the situation to determine by removing a number of sources from the mixed signals, which is performed by employing the clustering-based Direction of Arrival (DOA) estimation and time-frequency masking, and then applies the ICA to separate the remaining sources.

F.Nesta [8], presents a method of frequency-domain blind source separation (FD-BSS) by recursively regularizing ICA (RR-ICA) over the frequencies based on the assumption of a priori knowledge: the demixing matrix and the time-activity of the sources is expected to vary continuously across frequencies.

The RR-ICA algorithm has some drawbacks such as it require independence assumption and data length to be not small. Recently, non-negative matrix factorization (NMF) has been studied in BSS [9]. Which does not require the independence assumption, and not restricted to data length to be not small, and yields an equally important basis function. A major difference between NMF and ICA is , the basis function are ranked by the non-gaussianities in ICA, while in NMF, the basis functions are not ranked, but represent intrinsic properties of the data set. From a view point of data analysis, NMF is attractive because it takes into account spatial and temporal correlations between variables more accurately than ICA, and it provides usually sparse common factors or hidden (latent) components with physiological meaning and interpretation. RR-ICA also use STFT that will produce errors especially when complicated transient phenomena such as the mixing of speech and music occur in the analyzed signal.

The aim of this work is to remedy the drawbacks of RR-ICA through two stages ,the first stage: for source separation we formulate a multichannel NMF model that accounts for convolutive mixing based on αβ–divergence [10], in which this algorithm is the extension of previous work for multichannelseparation[1] [11] by using cochleagram. Unlike the spectrogram which deals only with uniform resolution, the gammatonefilterbank produces nonuniform TF domain (termed as the cochleagram)whereby each TF unit has different resolution. We prove that themixed signal is significantly more separable in the cochleagram than the spectrogram and the log-frequency spectrogram (constant-Q transform). The proposed family of αβ-multiplicative NMF algorithm is shown to improve robustness of separation with respect to noise and outliers in multichannel convolution, also the problem produced from STFT is solved by performing the decomposition in cochleagram domain. The second stage: for sound source localization (SSL) the TDOA is estimated from unmixing matrix, where TDOA can be used to localize the sources through using Scaled Conjugate Gradient (SCG) algorithm. In which it has the advantage of requiring virtually no parameter tuning. Second-order information (the Hessian) is approximated using the gradient only [12].

The remaining of this paper is organized as follows. Mixing model is introduced in section 2. In section 3 the αβ-NMF in Cochleagram domain. Section 4 presents the source localization. Section 5 presents the results of our algorithm of source separation in various settings. Conclusions are drawn in section 6.

2. The Mixing model

Given a sampled signal$x_i(t)$ generated as Junknown convolutive noisy mixture of point source signals $s_j(t)$ such that

$$x_i(t) = \sum_{j=1}^{J} a_{ij} s_j(t - \tau_{ij}) + b_j(t), i = 1,2,\dots,I \quad (1)$$

where$b_j(t)$ is additive noise, $a_{ij} \in R$ and $\tau_{ij} \in R^+$ are the attenuation factor and the propagation delay (in seconds), respectively, between the jth source and the ith sensor, The time-domain mixing given by (1) can be approximated in the short-time Fourier transform (STFT) domain as:

$$x_{ifn} \cong \sum_{j=1}^{J} a_{ij} s_{jfn} + b_{ifn} \quad (2)$$

where x_{ifn} and s_{jfn} are the complex-valued STFTs of the corresponding time signals,$j = 1, \dots, J$ is the source index and b_{ifn} is complex-valued STFTs of noise , $f = 1, \dots, F$ is a frequency bin index, $n = 1, \dots, N$ is a time frame index. The time-frequency model (2) can be rewritten as

$$X(f) = H(f)S(f) + B(f), \quad (3)$$

Where$X(f) = [x_{1fn}, \dots, x_{Ifn}]$, $\mathbf{S}(f) = [s_{1fn}, \dots, s_{Jfn}]$, $[H(f)]_{ij} = a_{ij} e^{\frac{-2\pi j}{F}(f-1)D_{ij}}$, $D_{ij} = F_s \tau_{ij}$, is Time of Arrival (TOA) , for any sensor-source pair (i, j) , the vector h_{ij}:

$$\mathbf{h}_{ij} = [\mathbf{H}]_{ij,1:F} = \\ = \left[a_{ij,} a_{ij} e^{\frac{-2\pi j}{F}(f-1)D_{ij}}, \dots, a_{ij} e^{\frac{-2\pi j}{F}(F-1)D_{ij}}\right]^T \quad (4)$$

Is a Vandermonde vector, this specific structure will be enforced on its estimate $\hat{h}_{ij}$ at each step of the iterative algorithm proposed in Section 3. The equation (3) can be solved by many algorithms that depend on ICA such as RR-ICA algorithm [8] and ICA-based DOA estimation [6]. These algorithms use STFT and this will lead to some drawbacks such that the classical spectrogram is computed by the STFT has an equal-spaced bandwidth across all frequency channels. Since speech signals are characterized as highly non-stationary and non-periodic whereas music changes continuously; therefore, application of the Fourier transform will produce errors especially when complicated transient phenomena such as the mixing of speech and music occur in the analysed signal. Unlike the spectrogram, the log spectrogram possesses non-uniform TF resolution. However, it does not exactly match to the nonlinear resolution of the cochlear since their center frequencies are distributed logarithmically along the frequency axis and all filters have constant-Q factor [13]. On the other hand, these drawbacks solved by using gammatone filters, in which gammatone filters are approximated logarithmically spaced with constant-Q forfrequencies from$f_s/10$ to $f_s/2$and approximated linearly spaced for frequenciesbelow$f_s/10$.Hence, this characteristic results in selective non-uniform resolution in the TF representation of the analyzed audio signal [14], as in Fig 1, That shows an example of the frequency response for different types transform.

3. $\boldsymbol{\alpha\beta}$-NMF in Cochleagram

[1] Its a novel algorithm for using NMF in source localization.

This section describes a NMF algorithm for minimization of the likelihood objective function(5). The algorithm is similar in spirit to the algorithm in [11], except that here we consider a multichannel in Cochleagram :

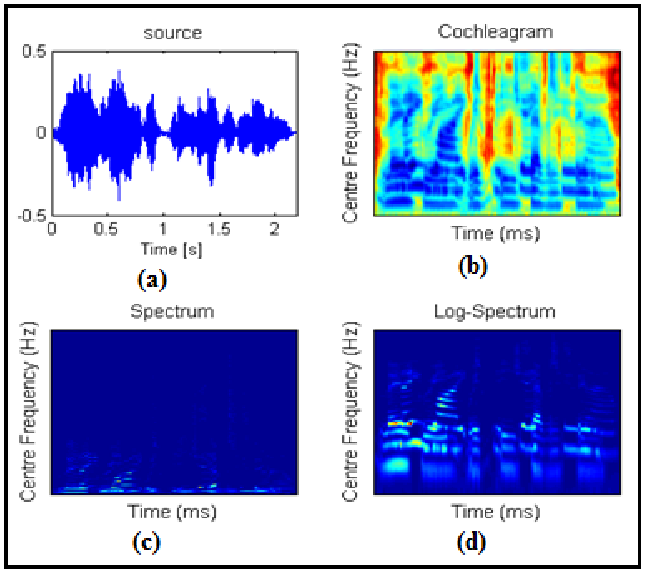

Fig1: different types transform (a) Original source (b) Cochleagram (c) Spectrum (d) Log-Spectrum.

$$\mathrm{O}(\theta) = \sum_{\mathrm{ifn}} \mathrm{d}_{\alpha\beta}(|\mathrm{x}_{\mathrm{ifn}}| \; || \; \hat{\mathrm{x}}_{\mathrm{ifn}}) \qquad (5)$$

whereθ is a scalar parameter of the set $\{\mathrm{S}, \mathrm{H}\}$, $\hat{\mathrm{x}}_{\mathrm{ifn}}$ is the structure defined by (3) and $\mathrm{d}_{\alpha\beta}$ is the $\alpha\beta$-divergence defined as [10]:

$$\mathrm{d}_{\alpha\beta} = \frac{-1}{\alpha\beta} \sum_{\mathrm{ifn}} \mathrm{x}_{\mathrm{ifn}}^{\alpha} \hat{\mathrm{x}}_{\mathrm{ifn}}^{\beta} - \frac{\alpha}{\alpha+\beta} \mathrm{x}_{\mathrm{ifn}}^{\alpha+\beta} - \frac{\beta}{\alpha+\beta} \mathrm{x}_{\mathrm{ifn}}^{\alpha+\beta} \;, (6)$$

whereα and β are parameters and $\beta + \alpha \neq 0$ (for more details about optimum choice of α and β see [10]), The $\alpha\beta$-divergence hasthe property of scale invariant, i.e.,$\mathrm{d}_{\alpha\beta}(\kappa a|\kappa b) = \mathrm{d}_{\alpha\beta}(a|b)$for any κ. This implies that any low energy components(a, b)will bear the same relative importance as the high energy ones$(\kappa a|\kappa b)$. This is particularly important to situations where$|\mathbf{X}|$is characterized by large dynamic range such as the audio shorttermspectra.

The derivative of $\mathrm{O}(\theta)$w.r.t θ:

$$\nabla_{\theta}\mathrm{O}(\mathbf{X}||\hat{\mathbf{X}}) = \sum (\nabla_{\theta}\hat{x}_{ifn}) \; \mathrm{d}'_{\alpha\beta}\left(|x_{ifn}| \; || \; \hat{x}_{ifn}\right) \quad (7)$$

where$\mathrm{d}'_{\alpha\beta}(|\mathrm{x}_{\mathrm{ifn}}| \, || \, \hat{\mathrm{x}}_{\mathrm{ifn}})$ is the derivative of $\mathrm{d}_{\alpha\beta}(|\mathrm{x}_{\mathrm{ifn}}| || \, \hat{\mathrm{x}}_{\mathrm{ifn}})$, w.r.t. $\hat{\mathrm{x}}_{\mathrm{ifn}}$,given by

$$\mathrm{d}'_{\alpha\beta}(|\mathrm{x}_{\mathrm{ifn}}| || \, \hat{\mathrm{x}}_{\mathrm{ifn}}) = \frac{1}{\alpha}\left(\mathrm{x}_{\mathrm{ifn}}^{\alpha+\beta-1} - \mathrm{x}_{\mathrm{ifn}}^{\alpha}\hat{\mathrm{x}}_{\mathrm{ifn}}^{\beta-1}\right) \quad (8)$$

By substitute (8) in (7), we obtain the following derivatives for each parameter:

$$\nabla_{\mathrm{h}}\mathrm{O}(\mathrm{X}||\hat{\mathrm{X}}) = \sum_{\mathrm{n}=1}^{\mathrm{N}} \mathrm{s}_{\mathrm{jfn}} \mathrm{d}'_{\alpha\beta}(|\mathrm{x}_{\mathrm{ifn}}|||\hat{\mathrm{x}}_{\mathrm{ifn}})$$

$$= \frac{1}{\alpha} \sum_{n=1}^{\mathrm{N}} s_{jfn}(\hat{x}_{ifn}^{.[\alpha+\beta-1]} - x_{ifn}^{.[\alpha]}\hat{x}_{ifn}^{.[\beta-1]}) \qquad (9)$$

$$\nabla_{s}\mathrm{O}(\mathrm{X}||\hat{\mathrm{X}}) = \sum_{\mathrm{i}=1}^{\mathrm{I}} \mathrm{h}_{\mathrm{ij}} \mathrm{d}'_{\alpha\beta}(|\mathrm{x}_{\mathrm{ifn}}|||\hat{\mathrm{x}}_{\mathrm{ifn}})$$

$$= \frac{1}{\alpha} \sum_{i=1}^{I} h_{ij}(\hat{x}_{ifn}^{.[\alpha+\beta-1]} - x_{ifn}^{.[\alpha]}\hat{x}_{ifn}^{.[\beta-1]}) \quad (10)$$

The previous equations can be written in the following matrix form:

$$\nabla_{\mathrm{H}}\mathrm{O}(\mathrm{X}||\hat{\mathrm{X}}) = \frac{1}{\alpha} \sum (\hat{\mathbf{X}}(f)^{\alpha+\beta-1}$$
$$- \mathbf{X}(f)^{\alpha}\hat{\mathbf{X}}(f)^{\beta-1})\mathbf{S}(f)^{\mathrm{T}} \qquad (11)$$

$$\nabla_{S}\mathrm{O}(\mathrm{X}||\hat{\mathrm{X}}) = \frac{1}{\alpha} \sum \mathrm{H}(\mathrm{f})^{\mathrm{T}}(\hat{\mathrm{X}}(\mathrm{f})^{\alpha+\beta-1}$$
$$- \mathrm{X}(\mathrm{f})^{\alpha}\hat{\mathrm{X}}(\mathrm{f})^{\beta-1}) \qquad (12)$$

So the multiplication update rule [2] for both H and S in matrix form are

$$\mathrm{S}\;(\mathrm{f}) = \mathbf{S}(f) \circledast \left(\frac{\left(\mathbf{H}(f)^{\mathrm{T}}\left(\mathbf{X}(f)^{.[\alpha]}\hat{\mathbf{X}}(f)^{.[\beta-1]}\right)\right)}{\left(\mathbf{H}(f)^{\mathrm{T}}\hat{\mathbf{X}}(f)^{.[\alpha+\beta-1]}\right)} \right) \qquad (13)$$

$$\mathrm{H}(\mathrm{f}) = \mathrm{H}(\mathrm{f}) \circledast \left(\frac{\left(\left(\mathrm{X}(\mathrm{f})^{.[\alpha]}\hat{\mathrm{X}}(\mathrm{f})^{.[\beta-1]}\right)\mathrm{S}(\mathrm{f})^{\mathrm{T}}\right)}{\left(\hat{\mathrm{X}}(\mathrm{f})^{.[\alpha+\beta-1]}\mathrm{S}(\mathrm{f})^{\mathrm{T}}\right)} \right) \qquad (14)$$

Where $\circledast$ is the Hadamard (components-wise) product, after the estimation of (f) , the ratios between the elements of two generic rows k and l of the matrix $\hat{\mathrm{H}}(\mathrm{f})$ are scaling invariant [8], i.e., for the jth column.

$$\mathrm{r}_{\mathrm{j}}^{(\mathrm{k,l})}(\mathrm{f}) = \frac{\mathrm{h}_{\mathrm{kj}}(\mathrm{f})}{\mathrm{h}_{\mathrm{lj}}(\mathrm{f})} \qquad (15)$$

Assuming the permutation problem to be solved, each ratio represents the acoustic propagation of the jth source with respect to the microphone pair (k, l) at the fth frequency bin. Indeed, (15) can rewrite as

$$\mathrm{r}_{\mathrm{j}}^{(\mathrm{k,l})}(\mathrm{f}) = \left|\mathrm{r}_{\mathrm{j}}^{(\mathrm{k,l})}(\mathrm{f})\right| \mathrm{e}^{-2\pi \mathrm{j} \mathrm{f} \Delta \mathrm{t}_{\mathrm{j}}^{(\mathrm{k,l})}(\mathrm{f})} \qquad (16)$$

where$\Delta\mathrm{t}_{\mathrm{j}}^{(\mathrm{k,l})}(\mathrm{f})$ is the TDOA for the jth source at the chosen microphone pair (k, l) observed at the fth frequency bin. In anechoic conditions, where the reverberation is absent, the magnitude $\left|\mathrm{r}_{\mathrm{j}}^{(\mathrm{k,l})}(\mathrm{f})\right|$and the TDOA$\Delta t_j^{(k,l)}(f)$ are expected to be invariant with respect to the frequency, and consequently the phase of (16) must vary linearly.Hence, as long as the acoustic waves related to the propagation along the direct paths dominate over the secondary reflections, each ratio can be considered as a state observation of the free-field propagation model of the jth source [8].

4. SignalSource Localization

The purpose of this stage is to localize the J sources from the TDOA estimates w.r.t. , the reference sensor l. Let $p_j = [x_j, y_j, z_j]^T$,denote the unknown vector of Cartesian coordinates of the jth source in a 3D propagation medium $p_i^* = [x_i, y_i, z_i]^T$,the vector of known coordinates of the ith sensor. Choose the reference sensor l as the origin of the new system of coordinates. Let us compute the relative range difference estimates $\widehat{rd}_{ij} = \Delta t_{ij} v$[2], where v denotes the wave velocity in the propagation medium. The position of sources found by minimizing the following cost function:

$$J_{TDOA}(p_j) = \sum_{i=2}^{I} \left(\widehat{rd}_{ij} - u_{ji} + u_{j1}\right)^2 \qquad (17)$$

where$u_{ji} =$

$$\sqrt{\left(x_j - x_i\right)^2 + \left(y_j - y_i\right)^2 + \left(z_j - z_i\right)^2} \qquad (18)$$

$$u_{j1} = \sqrt{(x_j - x_1)^2 + (y_j - y_1)^2 + (z_j - z_1)^2} \quad (19)$$

$J_{TDOA}(p_j)$is optimized by using the SCG algorithm [12].Compared to other gradient descent algorithms, the SCG has the advantage of requiring virtually no parameter tuning. Second-order information (the Hessian) is approximated using the gradient only. This is particularly friendly in the localization case, where the dimensionality is quite large. The first derivatives of (17),w.r.t. x_j, y_j, and z_j are:

$$\frac{\partial J_{TDOA}(p_j)}{\partial x_j} = 2\sum_{i=2}^{I} \left(\widehat{rd}_{ij} - u_{ji} + u_{j1}\right) \times \left(\frac{-(x_j - x_i)}{u_{ji}} + \frac{(x_j - x_1)}{u_{j1}}\right) \quad (20)$$

$$\frac{\partial J_{TDOA}(p_j)}{\partial y_j} = 2\sum_{i=2}^{I} \left(\widehat{rd}_{ij} - u_{ji} + u_{j1}\right) \times \left(\frac{-(y_j - y_i)}{u_{ji}} + \frac{(y_j - y_1)}{u_{j1}}\right) (21)$$

$$\frac{\partial J_{TDOA}(p_j)}{\partial z_j} = 2\sum_{i=2}^{I} \left(\widehat{rd}_{ij} - u_{ji} + u_{j1}\right) \times \left(\frac{-(z_j - z_i)}{u_{ji}} + \frac{(z_j - z_1)}{u_{j1}}\right) \ (22)$$

In finally we conclude our algorithm (called αβ-NMFC) as in algorithm 1. (The algorithm SCG mentions in the Appendix B).

```
Algorithm 1
Input :X input data (mixture),$p^*$ sensor position
Output: S ,H and $\hat{p}_j$
Begin
Initialization for $S$and H
X =cochleagram (X)/* cochleagramdomain*/
Repeat                    /* update S and H */
    Update$H(f)$ using (14)
    Calculate $\Delta t_{ij}$ from (16)
    Update$S(f)$ using (13)
Until  a stopping criterion is met
$\hat{X}$ =inv_cochleagram($\hat{X}$)
$for\ j = 1{:}J$
    $\hat{p}_j = scg\left(J_{TDOA}(p_j), \frac{\partial J_{TDOA}(p_j)}{\partial p_j}\right)$
$end$
 End
```

5. Results and Analysis

A. Experimental setup

This experiment was performed using the mixture generation, from a group of speakers'(male and female) were selected from TIMIT speech database [15]. All mixtures are sampled at 16 kHz, in which there are a fixed number of sensors and sources that are randomly placed in a 3D room of size $12\text{m} \times 8\text{m} \times 3\text{m}$. The proposed algorithm has been tested in Matlab.

B. Criteria

The performance of separation is evaluated with the BSS EVAL toolbox, which is based on the criteria proposed in [16], using time-invariant filters of 1024 taps to represent the family ofallowed distortions.The source-to-interferences ratio (SIR), the source-to-distortion ratio (SDR) and the sources to artifacts

[2] $\Delta t_{ij} = \Delta t_j^{(i,l)}$,since$l$ is the origin and we set $l = 1$.

ratio (SAR) are evaluated using the whole separated signals.The performance of source localization can be evaluated by using the Mean Square Error (MSE) ρ:

$$\rho = \frac{1}{J}\sum_{j=1}^{J}\left\|p_j - \hat{p}_j\right\|^2 \qquad (23)$$

whereJ is the number of source, and p_j and $\hat{p}_j$,are the estimated and ground truth positions of the jth source, respectively.

C. The efficiency ofαβ-NMFC in Source Separation

In this example we compare our algorithm αβ-NMFC with RR-ICA algorithm for source separation, in which the TF representation for RR-ICA is computed by normalizing the time-domain signal to unit power and computing the STFT using 1024 points Hamming window FFT with 50% overlap. For αβ-NMFC the cochleagram based on Gammatonefilterbank of 128 channels (filter order of 4) and the output is divided into 20-ms time frame with 50% overlap between consecutive frames. In all cases, the sources are mixed with equal average power over the duration of the signals. Fig.3 shows the time domain of the original source (male, female and music) and the cochleagram of three sources; Fig.4 shows the mixtures of sources and its Cochleagram. Fig.5 shows the final recovered time-domain sources.

To further analyses the performance of all the above methods in separating the mixed signal and capturing the TF patterns of the sources, the time domain of the each recovered source has been plotted in Fig 5. In Fig 5, panels (a) and (b) denote the recovered of the sources by using the αβ-NMFC and RR-ICA algorithms, respectively. In particular, it is noted that both RR-ICA and αβ-NMFC algorithms exhibit good reconstruction. However, the RR-ICA algorithm fails to identify several missing components as indicated in the red box marked area of panel (b). Hence, less accuracy is obtained in the estimation of the source as compared with the αβ-NMFC algorithm which has successfully estimated sources with high accuracy.The major reason for the large discrepancy between them is the resulting spectrogram fails to infer the dominating source. This leads to a high degree of ambiguity in TF domain and causes lack of uniqueness in extracting the spectral-temporal features of the sources. The cochleagram enables the mixed signal to be more separable and thereby reduces the mixing ambiguity between $|S_1|^2$and $|S_2|^2$. This explains the performance of separating mixture music and female utterance is highest among all the mixtures because both sources have very distinguishable TF patterns in the cochleagram. In summary, all the results in Table 1 and Figs 5 unanimously show the importance of using the αβ-NMFC factorization algorithm in order to correctly estimate the spectral and temporal features of each source.

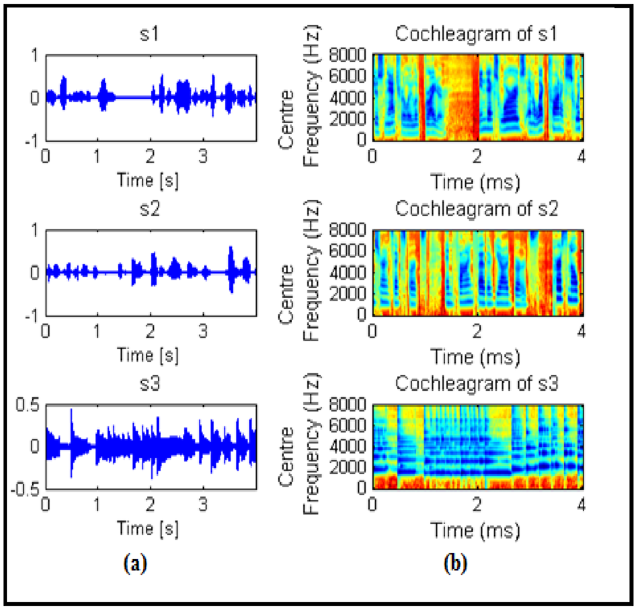

Fig2 : (a) Original source . (b) Cochleagramof the original.

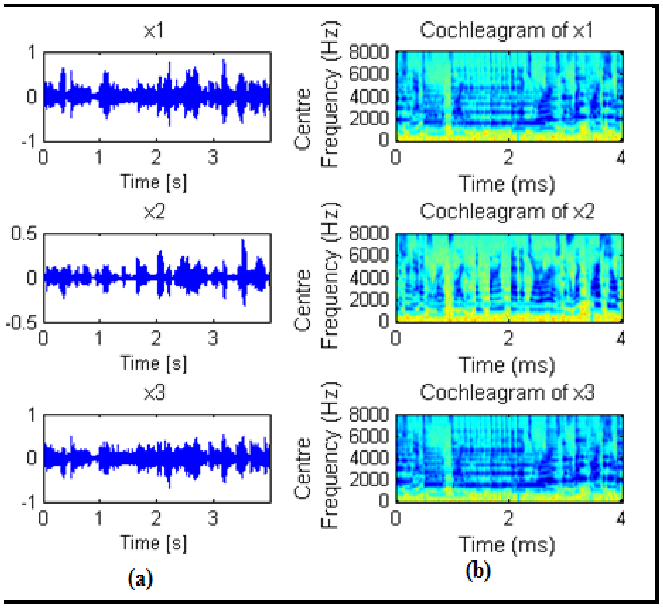

Fig3: :(a) Mixture and (b) Cochleagram of mixture.

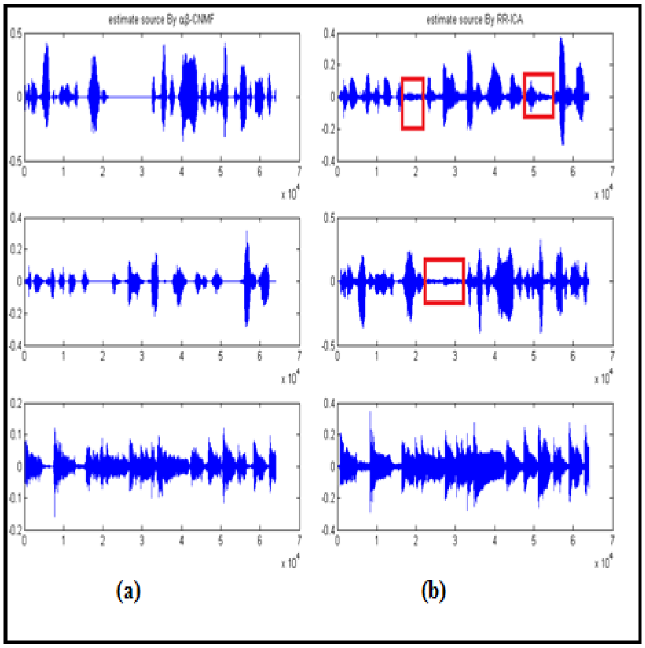

Fig4 :Time separation of source (a) **αβ**-NMFC ,(b) RR-ICA algorithm.

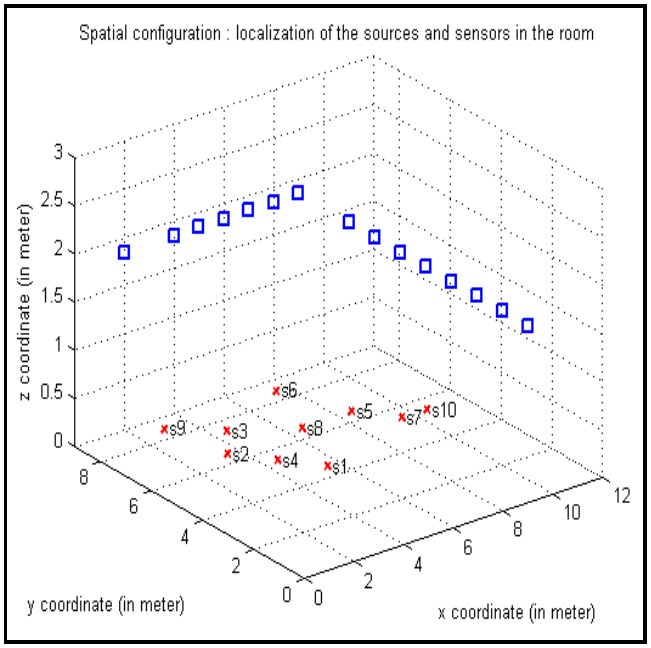

Fig5:Original sources with red (x) and microphone with blue square.

Table 1: Comparison between RR-ICA and **αβ**-NMFC.

	SNR	Algorithm	SDR			SAR			SIR		
			S1	S2	S3	S1	S2	S3	S1	S2	S3
Mixture of 3 sources	10	αβ-NMFC	10.088	9.467	9.188	10.8510	9.930	9.827	23.361	19.838	18.832
		RR-ICA	-2.441	6.723	-2.918	-1.8484	12.180	1.728	10.534	8.433	1.136
	15	αβ-NMFC	8.949	8.515	8.956	10.3333	10.288	11.907	17.87	18.307	17.826
		RR-ICA	-3.016	-0.366	-0.178	-2.7027	-0.238	-0.004	13.121	18.139	16.914
	20	αβ-NMFC	5.390	5.398	5.606	6.651	6.977	6.255	12.225	13.769	13.571
		RR-ICA	-2.310	0.929	-0.211	-2.0820	1.050	0.040	14.768	19.068	15.273

D. The efficiency of$\alpha\beta$-NMFC in source localization in 3D

In this section we used the same data used in the previous example where , the configuration of the sources and microphones is shown in Fig5 in which consists of 10 sources and 15 sensors[3]. Figs 6-7 show a qualitative evaluation of the estimated 3D positions of 10 sources against the ground truth for RR-ICA[4]and αβ-NMFC respectively. The results for each source are obtained as the average over the source position errors from the 10 experimental runs for a given microphone combination and overall the two adjacent microphone combinations.To further analyses the performance of all two algorithms, Table 2 shows the comparison of the proposed algorithm αβ-NMFC with RR-ICA with different

[3] The total number of sensors and sources are 16 and 10 respectivly

[4] We used the relation between TDOA and DOA to used RR-ICA in this paper.

SNR(5,10,15,20). It is noted that αβ-NMFC is better than RR-ICA in term of MSE ρ:

E. The efficiency of SCG

This section shows the performance of SCG where , we compare between SCG , the gradient descent (GD) and Newton method[5]used to the minimize (17) after estimate the mixing matrix by αβ-NMFC and then determine the TDOAs for each source. We run 100 random trials for each configuration (three source with five sensors and four sources with 6 sensors) in 2D as in Fig8 (a) . Fig8(b) and Table 3 show that ,the GD method completely fails to find the true estimation of source position (the estimated position out of range) . However the bothmethods Newton and SCG are good but the SCGisfaster and robustness than Newton method as in Fig8(c),(d).

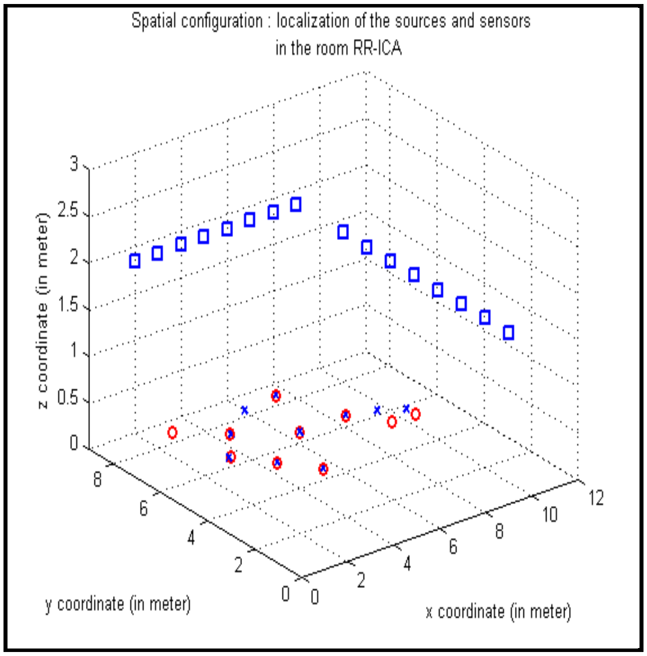

Fig6: Estimation the position of sources by RR-ICA

Table 2: comparison between RR-ICA and **αβ** NMFC**.**

SNR	RR-ICA	αβ – NMFC	
5	3.3434	0.3774	8 sources + 12 sensors
10	3.4288	0.4801	
15	4.0116	0.6175	
20	4.1397	3.5912	
5	3.9275	0.9179	10 sources + 15 sensors
10	4.3088	1.6610	
15	4.5842	1.3675	
20	4.9566	2.3607	

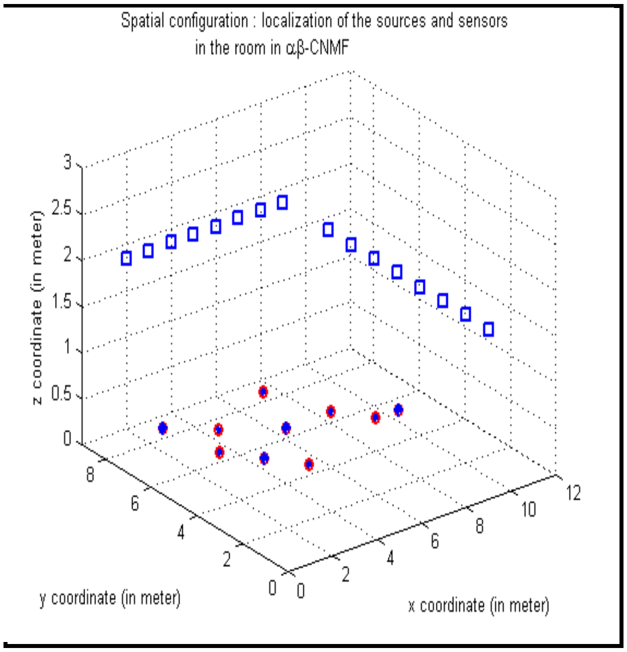

Fig7 : Estimation the position of sources byαβ-NMFC

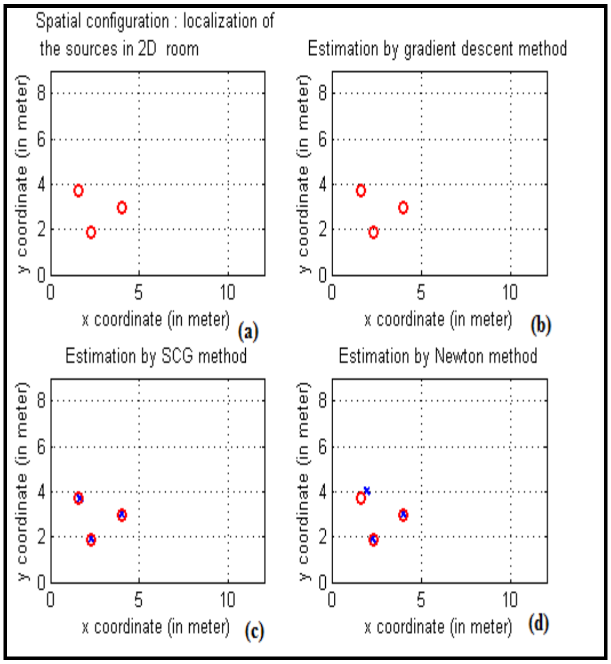

Fig8:Estimation the position of sources by (a) original position ,(b) GD ,(c) SCG , and(d) Newton.

[5] The second derivatives of (17) are calculated in Appendix B

Table 3: Comparison between SCG , Newton and Gradient descent.

SNR	SCG		Newton		Gradient descent		
	MSE ρ	Time(s)	MSEρ	Time(s)	MSE ρ	Time(s)	
10	2.0548e-005	0.0063	0.0042	0.0367	1.3606e+004	0.0237	3source
15	3.0548e-005	0.0069	0.1798	0.0583	5.3752e+004	0.0483	
20	0.0072	0.0257	0.2627	0.0633	1.3606e+004	0.0233	
10	0.0022	0.0094	0.0045	0. 0671	2.3928e+004	0.0371	4source
15	0.0468	0.0296	0.1351	0.0507	6.8489e+004	0.0517	
20	0.0744	0.0180	0.1597	0.0685	5.8337e+004	0.0459	

6. Conclusion

In this paper we proposed a novel system for source separation and localization framework using the gammatonefilterbank, where the gammatonefilterbank produces a non-uniform TF domain termed as the cochleagram whereby each TF unit has different resolution unlike the classical spectrogram which deals only with uniform resolution. Towards this end, it is shown that the mixed signal is significantly more separable in the cochleagram than the classic spectrogram and the log-frequency spectrogram (constant-Q transform). Also a family of αβ-divergence based novel nonnegative matrix factorization algorithms has been developed to extract the spectral and temporal features of the sources.

This paper also described the problem of locating the acoustic source in 2-D plane and 3-D space using microphones. Our approach of determining the time-delays using the mixing matrix that estimated from the proposed αβ-NMFC algorithm produces results that are comparable to the conventional method RR-ICA. The source positioned in a plane and space have been successfully located by αβ-NMFC. Experimental validation tests have proved that this methodology is suitable for locating arbitrary sound sources in a nonideal environment.

Appendix A

The second derivatives of (17) that used in hessian matrix (Newton method) are :

$$\frac{\partial^2 J_{TDOA}}{\partial x^2} = 2\sum_{i=2}^{I}\left(Wx^2 + \left[\left(-\frac{\left(z_j - z_i\right)^2 + \left(y_j - y_i\right)^2}{\left(u_{ji}\right)^3} + \frac{\left(z_j - z_1\right)^2 + \left(y_j - y_1\right)^2}{\left(u_{j1}\right)^3}\right)\sqrt{J_{TDOA}}\right]\right) \quad (24)$$

$$\frac{\partial^2 J_{TDOA}}{\partial x \partial y} = 2\sum_{i=2}^{I}\left((WxWy) + \left[\left(\frac{\left(x_j - x_i\right)\left(y_j - y_i\right)}{\left(u_{ji}\right)^3} - \frac{\left(x_j - x_1\right)\left(y_j - y_1\right)}{\left(u_{j1}\right)^3}\right)\sqrt{J_{TDOA}}\right]\right) \quad (25)$$

$$\frac{\partial^2 J_{TDOA}}{\partial x \partial z} = 2\sum_{i=2}^{I}\left((WxWz) + \left[\left(\frac{\left(x_j - x_i\right)\left(z_j - z_i\right)}{\left(u_{ji}\right)^3} - \frac{\left(x_j - x_1\right)\left(z_j - z_1\right)}{\left(u_{j1}\right)^3}\right)\sqrt{J_{TDOA}}\right]\right) \quad (26)$$

where

$Wx = \frac{-(x_j - x_i)}{u_{ji}} + \frac{(x_j - x_1)}{u_{j1}}$,$Wy = \frac{-(y_j - y_i)}{u_{ji}} + \frac{(y_j - y_1)}{u_{j1}}$and $Wz = \frac{-(z_j - z_i)}{u_{ji}} + \frac{(z_j - z_1)}{u_{j1}}$. The derivatives of $\frac{\partial^2 J_{TDOA}}{\partial y\, \partial z}$, $\frac{\partial^2 J_{TDOA}}{\partial y^2}and\frac{\partial^2 J_{TDOA}}{\partial z^2}$aresimilar.

Appendix B:

w=SCG algorithm($f(\mathrm{w_k})$, f'(w1)) [12]
1. Choose a weight vector w_1 and scalars ,σ> 0 and $\bar{\lambda}_1 = 0$.

```
Set p1 = r1 = – $f$'(w1), k = 1 and success = true.
2.  If success = true then calculate second order information:
        $\sigma_k = \frac{\sigma}{|p_k|}, s_k = \frac{f'(w_k+\sigma_k p_k)-f'(w_k)}{\sigma_k|}, \sigma_k = p_k^T s_k$.
3.  Scale $s_k$ :
    $s_k$=$s_k + (\lambda_k - \bar{\lambda}_k)P_k$ ,
    $\delta_k = \delta_k + (\lambda_k - \bar{\lambda}_k){P_k}^2$.
4.  If $\delta_k \leq 0$ then make the Hessian matrix positive definite:
    $s_k$=$s_k + (\lambda_k - 2\frac{\delta_k}{{P_k}^2})P_k$,
    $\bar{\lambda}_k = 2(\lambda_k - \frac{\sigma}{{P_k}^2})$,
    $\delta_k = -\delta_k + \lambda_k {P_k}^2$ ,$\lambda_k = \bar{\lambda}_k$ .
5.  Calculate step size :
        $\mu_k = p_k^T r_k$ ,$\alpha_k = \frac{\mu_k}{\delta_k}$ .
6.  Calculate the comparison parameter :$\Delta_k = \frac{2\delta_k[f(w_k)-f(w_k+\alpha_k p_k)]}{\mu_k^2}$ .
7.  If $\Delta_k \geq 0$ then a successful reduction in error can be made :
        $w_{k+1} = w_k + \alpha_k p_k$,
    $r_{k+1} = -f'(w_{k+1})$ ,
    $\bar{\lambda}_k$=0, success=true.
        7a. If k mod N=0 then restart algorithm: $p_{k+1} = r_{k+1}$
            else create new conjugate direction:
                $\beta_k = \frac{{r_{k+1}}^2 - r_{k+1}r_k}{\mu_k}$,
                $p_{k+1} = r_{k+1} + \beta_k p_k$ .
        7b.If $\Delta_k \geq 0.75$ then reduce the scale parameter :$\lambda_k = \frac{1}{2}\lambda_k$ .
    else a reduction in error is not possible: $\bar{\lambda}_k = \lambda_k$, success=false.
8.  If $\Delta_k 0.25$ then increase the scale parameter :$\lambda_k = 4\lambda_k$
9.  If the steepest descent direction$r_k \neq 0$ then set k=k+1 and go to 2
     else terminate and return $w_{k+1}$ as the desired minimum
```

Reference

[1] S. Araki, H. Sawada, R.Mukai, and S. Makino," DOA estimation for multiple sparsesources with normalized observation vector clustering," *in Proceedings of the IEEEInternational Conference on Acoustics, Speech, and Signal Processing (ICASSP ’06),*vol. 5, pp. 33–36, Toulouse, France, 2006.

[2] S. Araki, H. Sawada, R. Mukai, and S. Makino, "Underdetermined blind sparse source separation for arbitrarily arranged multiple sensors", *Signal Processing*, vol. 87, no. 8, pp. 1833– 1847, 2007.

[3] S. Rickard and F. Dietrich,"DOA estimation of many Wdisjoint orthogonal sources from two mixtures using duet", *IEEE Signal Processing Workshop on Statistical Signal and Array Processing*, pp. 311–314, 2000.

[4] P. Bofill and M. Zibulevsky," Underdetermined blind source separation using sparse representations," *Signal Processing*, vol. 81, no. 11, pp. 2353–2362, 2001.

[5] G. Reju, S. N. Koh, and I. Y. Soon," Underdetermined convolutive blind source separationvia time-frequency masking," *IEEE Transactions on Audio, Speech and LanguageProcessing*, vol. 18, no. 1, Article ID 5061881, pp. 101–116, 2010.

[6] J. Peter, XinZou, and MunevverKokuer, "Underdetermined DOA Estimation via Independent Component Analysis and Time- Frequency Masking ,"*Journal of Electrical and Computer Engineering*,vol(2010) ,2010.

[7] S. Araki, S. Makino, A. Blin, R. Mukai, and H. Sawada, "Blind separation of more speechthan sensors with less distortion by combining sparseness and ICA," *in Proceedings of the International Workshop on Acoustic Echo and Noise Control(IWAENC '05),* pp. 271–274, Kyoto, Japan, September 2003.

[8] F. Nesta, PiergiorgioSvaizer, and Maurizio Omologo, "Convolutive BSS of ShortMixtures by ICA Recursively Regularized Across Frequencies" , *Trans. Audio Speechand Language Processing* ,vol. 19, no. 3, 2011.

[9] A. Cichocki,; R. Zdunek,; A.H. Phan,; S.Amari," Nonnegative Matrix and TensorFactorizations,"*John Wiley & Sons Ltd.: Chichester*, UK, 2009.

[10] A. Cichocki , C. Sergio and S. Amari," Generalized Alpha-Beta Divergences and Their Application to Robust Nonnegative Matrix Factorization," *Entropy*, 13,134- 170; 2011.

[11] Wael M. Khedr , M. E. Abd El-Aziz and S. M. Amer,"A Novel Algorithm forMultichannel Deconvolutive based on αβ –Divergence" . *International Journal of Computer Applications* ,vol 57, no.10, 2012.

[12] F. Møller ,"A Scaled Conjugate Gradient Algorithm for Fast Supervised Learning," , *neural network*, Vol. 6, pp. 525-533, 1993.

[13] C.Brown. Calculation of a constant Q spectral transform, J. Acoust. Soc. Am., vol. 89,no 1, pp. 425–434,1991.

[14] G. Hu and D. L. Wang, "Auditory segmentation based on onset and offset analysis, " *IEEETrans. AudioSpeech and Language Processing*, vol. 15, no. 2, pp. 396–405, Feb. 2007.

[15]www.ldc.upenn.edu/Catalog/LDC93S1.html

[16] E. Vincent, R. Gribonval, and C. Févotte," Performance measurement in blindaudiosource separation," *IEEE Trans. Audio*, *Speech,Language Process*., vol.14, no. 4, pp. 1462–1469, Jul. 2006.