

A FIXED-POINT ICA ALGORITHM FOR CONVOLUTED SPEECH SIGNAL SEPARATION

Rajkishore Prasad, Hiroshi Saruwatari, Akinobu Lee, Kiyohiro Shikano
Graduate School of Information Science, Nara Institute of Science and Technology
8916-5 Takayama-cho, Ikoma-shi, Nara, 630-0101, JAPAN
E-mail: {kishor-p, sawatari, ri, shikano}@is.aist-nara.ac.jp

ABSTRACT

This paper describes a fixed-point independent component analysis (ICA) algorithm in combination with the null beamforming technique to sieve out speech signals from their convoluted mixture observed using a linear microphone array. The fixed-point algorithm shows fast convergence to the solution, however it is highly sensitive to the initial value from which iteration starts. A good initial value leads to faster convergence and yields better results. We propose the use of a null beamformer-based initial value for iteration and explore its effects on separation performance under different acoustic conditions by examining the noise reduction rate (NRR) and convergence speed. The result of the simulation confirms the efficacy and accuracy of the proposed algorithm.

1. INTRODUCTION

Blind source separation (BSS) addresses the problem of estimating unobserved signals or sources from several observed mixtures without observing the sources and knowing the mixing procedure. This complete lack of a priori knowledge is compensated by assuming statistical independence of each source, and typically observed signals are obtained using an array of sensors that contains spatially sampled signals of the unobserved sources. This technique has wide applicability, and in recent years, it has gained much research attention. This technique has been coined as a solution to the cocktail party problem in spoken communication, which is related to a human's ability to focus auditory attention on a particular speech in the presence of multiple speech sources. One technical characterization of this ability is to separate a particular signal from mixtures of signals. This is what BSS does for speech signals and hence it has applicability in developing a noise-robust speech recognition system and hands-free communication systems.

The independent component analysis (ICA)-based BSS technique is performed in both the time domain and the frequency domain [1,2,3]. In the frequency-domain ICA (FDICA), a complex-valued separation matrix is calculated in the frequency domain, and in time-domain ICA, the inverse of the mixing FIR filter matrix is calculated in the time domain. Several algorithms have been developed for blind separation in an artificially controlled acoustic environment and in a real acoustic environment, however, the separation performance is still poor [4]. Real-world application requires faster methods to

perform on-line separation. To date, the algorithms developed are not sufficiently fast to satisfy real-time requirements. Frequency-domain approaches are relatively fast due to the power of FFT, yet the gradient-based FDICA techniques require a larger number of iterations to estimate the separating matrix. However, other algorithms like fastICA methods [5] based on Newton's iterative methods are fast but sensitive to the initial value from which iterative learning starts. In order to enhance the separation performance and to solve the problem of permutation and scaling arising in FDICA, the beamforming technique has been combined with the gradient-based ICA algorithm [6,7] and relatively good separation performance has been reported. In this paper, we focused on the new combination of the null beamforming (NBF) technique and the fixed-point ICA algorithm. Using the NBF technique, we solved the permutation and scaling problem and also generated a good initial value for the fixed-point ICA methods. In this way, the use of beamforming achieves three aims.

2. SIGNAL MIXING AND UNMIXING MODEL

In the real recording environment, signals reaching each microphone are not only direct-path signals but also delayed and attenuated versions of the source signal. Therefore, the speech signal picked up by a microphone array is modeled as a linear convolutive mixture of impinging signals, in which the n -dimensional signal vector $x(t)=[x_1(t), x_2(t), \dots, x_n(t)]^T$ picked up by an n -element linear microphone array is given by

$$x(t) = \sum_{-\infty}^{\infty} h(k)s(t-k), \quad (1)$$

where $s(t)=[s_1(t), s_2(t), \dots, s_m(t)]^T$ is an unobservable source signal; $h_k = m \times n$ mixing matrix in which h_{ij} represents the impulse response between the i th source and the j th microphone.

However we are dealing with frequency-domain ICA, so the same signal model in the frequency domain is expressed in terms of simple multiplication as follows:

$$X(f) = A(f)S(f), \quad (2)$$

where $X(f)=[X_1(f), X_2(f), \dots, X_n(f)]^T$ is the observed signal vector; $S(f)=[S_1(f), S_2(f), \dots, S_m(f)]^T$ is the original signal source, and

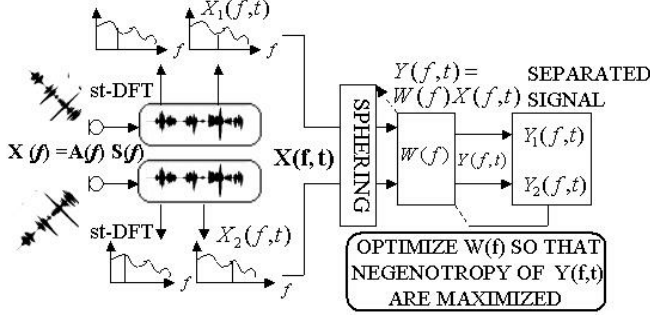


Fig. 1. ICA in frequency domain.

$$A(f) = \begin{bmatrix} A_{11}(f) & \dots & A_{1n}(f) \\ \vdots & \ddots & \vdots \\ A_{m1}(f) & \dots & A_{mn}(f) \end{bmatrix} \quad (3)$$

is the complex-valued mixing matrix.

In this study, we have considered the case of a two-microphone two-source, i.e., $n=m=2$, BSS problem for which mixing and demixing processes are shown in Figure 1.

Samaragdis [2] exploited the transformation of convolutive mixing into simple multiplicative operation and proposed the application of short-time Fourier transform (STFT) to the mixed signal $x(t)$ and then to separate independent components in every frequency bin. Thus, in the frequency domain, the entire process of convolved signal separation is transformed into the computation of the separation matrix $W(f)$ in each frequency bin for each source.

3. FIXED-POINT ICA

Fixed-point ICA was first developed and proposed in [5] for the separation of the instantaneous mixture. The key feature of this algorithm is that it converges faster than other algorithms, like natural gradient-based algorithms, with almost the same separation quality. In [8], fast algorithm has been extended to complex-valued signals, however this algorithm has no strategy for solving the problem of permutation and scaling arising in speech signal separation. In [9], Mitianoudis et al. have proposed the application of the fixed-point algorithm for speech signal separation with a time-frequency-model-based likelihood ratio jump scheme as a solution for permutation. The nonlinearity used in [9] to measure negentropy is also different from that used in [8]. There are also other methods used to solve the permutation problem. In order to combine array signal processing techniques with fixed-point ICA, we propose to use a directivity-pattern-based technique to solve the permutation and scaling problem. Fixed-point-iteration-based ICA is very sensitive to the initial value from which iteration starts. A good initial value results in early convergence and better separation quality. Hence we have used a null-beamformer-based initial value for the separation matrix.

The fixed-point ICA algorithm is based on two-step approaches, namely, prewhitening or sphering and rotation of

the observation vector. Sphering is half of the ICA task and results in spatially decorrelated signals; the remaining involves rotating the whitened signal vector by the separation matrix. In the deflation-type algorithm, the rotation step consists of one-unit ICA which is used to estimate one separation vector w (one row of the separation matrix) such that the separated component $Y(f)=w^T X(f)$ equals one independent component. Here we place the derivation of such a one-unit fixed-point algorithm using negentropy as the measurement of nongaussianity of the separated signal. The negentropy $J(y)$ of the separated signal y is given by

$$J(y) = H(y_{gauss}) - H(y), \quad (4)$$

where y_{gauss} is the gaussian random vector with the same variance as that of the estimated independent component y , and $H(\cdot)$ denotes the differential entropy.

The idea of using negentropy as a measure of nongaussianity is grounded in the central limit theorem which states that a mixture of two or more nongaussian variables is more gaussian than the individuals. Thinking in reverse, it advocates that nongaussianity of individuals ensures statistical independence. Thus maximizing nongaussianity leads to separate components. Hyvarinen et al. [10] proposed the approximation of negentropy using the nonquadratic nonlinear function G as follows:

$$J(y) = \sigma \{E\{G(y) - E\{G(y_{gauss})\}\}, \quad (5)$$

where σ is a positive constant.

The performance of the fixed-point algorithm depends on the nonquadratic nonlinear function G used. Some of the nonquadratic functions used for complex-valued signal separation are

$$G_1(y) = \sqrt{a_1 + y} ; a_1 = .01 \quad (6a)$$

$$G_2(y) = \log(a_2 + y) ; a_2 = .01 \quad (6b)$$

$$G_3(y) = \frac{y}{|y|} ; \forall y \neq 0. \quad (6c)$$

Based on negentropic measurement of nongaussianity, we derive here a one-unit fixed-point ICA learning equation using the Lagrangian multipliers method of constraint optimization. The speech signal is also modeled as a spherically symmetric variable, and as pointed out in [10], for a spherically symmetric variable, modulus-based contrast function can be used to measure nongaussianity. Accordingly, we use the same contrast function

$$J(y) = E\{G(|w^H x|^2)\}, \quad (7)$$

where w is an n -dimensional complex vector such that

$$E\{|w^H x|^2\} = 1 \Rightarrow |w| = 1. \quad (8)$$

This contrast function may have m local or global optimum solutions w_i ($i=1,2, \dots, m$); the corresponding y_i will give an estimation of m separated sources. Thus the source separation problem calls for the following constraint optimization:

$$\text{Maximize } J(y) = E\{G(|w^H x|^2)\},$$

under constraint $E\{|w^H x|^2\} = 1 \Rightarrow |w| = 1$.

The maxima of $J(y)$ can be found by solving the Lagrangian function $L(w, w^H, \lambda)$ of the above, given as

$$L(w, w^H, \lambda) = E\{G(|w^H x|^2)\} \pm \lambda[E\{w^H x\} - 1] \quad (9)$$

For the maxima of the contrast function, the following simultaneous equations must be solved.

$$\frac{\partial L}{\partial w} = 0; \quad \frac{\partial L}{\partial w^H} = 0; \quad \frac{\partial L}{\partial \lambda} = 0 \quad (10)$$

These give

$$\frac{\partial L}{\partial w} = E\{g(|w^H x|^2)w^H\} + \lambda w^H = 0 \quad (11)$$

$$\frac{\partial L}{\partial w^H} = E\{g(|w^H x|^2)(x^H w)x\} + \lambda w = 0 \quad (12)$$

$$\frac{\partial L}{\partial \lambda} = |w|^2 - 1 = 0 \quad (13)$$

From here, we proceed further in light of the following two theorems [11]:

THEOREM 1: *If function $f(z, z^*)$ is analytic with respect to z and z^* , all stationary points can be found by setting the derivative with respect to either z or z^**

THEOREM 2: *If $f(z, z^*)$ is a function of the complex-valued variable z and its conjugate, then by treating z and z^* independently, the quantity directing the maximum rate of change of $f(z, z^*)$ is $\nabla z^* f(z)$*

Accordingly, the final solution using Newton's iterative method is given by

$$w_{new} = w - \left[\frac{\partial L}{\partial w^H} \right] \left[\frac{\partial}{\partial w} \left(\frac{\partial L}{\partial w^H} \right) \right]^{-1} \quad (14)$$

After each iteration, updated w is normalized as

$$w_{new} = w_{new} / \text{norm}(w_{new}) \quad (15)$$

This leads to a simplified learning equation as follows.

$$w_{new} = w(E\{g(|w^H x|^2) + (|w^H x|^2)g'(|w^H x|^2)\} - E\{g(|w^H x|^2)(x^H w)x\}) \quad (16)$$

This is the update equation for w .

Following the findings in [10], we use nonquadratic function (6b) whose first- and 2nd-order derivatives are given by

$$g = \frac{1}{(a_2 + (|w^H x|^2)^2)} \quad (17a)$$

$$\text{and } g' = -\frac{0.5}{(a_2 + (|w^H x|^2)^2)^2} \quad (17b)$$

4. ESTIMATION OF SEPARATION MATRIX

The above update equation is used to estimate separation vector w for each source one by one. First the observation signals are windowed using overlapping Hanning windows. Then P point STFT is carried out in each frame of each mixed signal. Next the time series $X(f, t) = [X_1(f, t), X_2(f, t), \dots, X_n(f, t)]^T$ is obtained by stacking signal samples of same frequency bins of every frame of every microphone signal. Then signal separation is performed on the time series data $X(f, t)$ of every frequency bin. As stated in the previous section, before performing ICA, the time series data in every frequency bin is centered and whitened using the Mahalanobis transform [12]. The whitened signal in the p th bin is obtained as

$$X_w(f_p, t) = Q(f_p)X(f_p, t) \quad (18)$$

with

$$Q(f_p) = \Lambda_x^{-0.5} V_x, \quad (19)$$

where

$$\Lambda_x = \text{diag}\left\{\frac{1}{\sqrt{\lambda_1}}, \dots, \frac{1}{\sqrt{\lambda_n}}\right\} \quad (20)$$

is the diagonal matrix with positive eigenvalues

$\lambda_1 > \lambda_2 > \dots > \lambda_n$ of the covariance matrix of the observed

vector in the p th frequency bin f_p , and

$$V_x = \text{Orthogonal matrix.} \quad (21)$$

After whitening, whitened data is used in the ICA stage where weight vector w is computed using iterative learning equation (16). After each iterative update, w is normalized since updating changes the norm and it is essential to keep the norm unity. As this is a deflationary algorithm, independent sources are extracted one by one in the order of decreasing negentropy from the mixed signal. Thus, after each iteration, it is also essential to decorrelate w to prevent its convergence to the previously converged point. For this, Gram-Schmidt sequential orthogonalization has been used, in which components of all previously obtained separation vectors falling in the direction of the current vector are subtracted as

$$w_i = w_i - \sum_{j=1}^{i-1} (w_i^T w_j) w_j \quad (22)$$

After finishing ICA in every frequency bin for m sources (less than or equal to the number of microphones), a separation matrix $W(f)$ is obtained in every frequency bin for every source:

$$W(f) = \begin{bmatrix} w_{11}(f) & \dots & w_{1n}(f) \\ \vdots & \ddots & \vdots \\ w_{m1}(f) & \dots & w_{mn}(f) \end{bmatrix} \quad (23)$$

Each row of this separation matrix corresponds to a separation vector for each source. Because this separation matrix has been obtained using whitened signals, its premultiplication with whitened signals in the frequency domain gives time series $Y(f, t) = [Y_1(f, t), Y_2(f, t), \dots, Y_m(f, t)]^T$ of the separated signal, i.e.,

$$Y(f, t) = W(f)X_w(f, t). \quad (24)$$

However, in order to use $W(f)$ of Equation (23) in the time domain to form an FIR filter, it is essential to dewhiten the separation filter as follows:

$$W(f) = W(f)Q(f)^{-1}. \quad (25)$$

Then using dewhitened W , an FIR filter of length P can be formulated to separate the signal as

$$y(t) = \sum_{r=0}^p w(r)x(t-r). \quad (26)$$

5. PERMUTATION AND SCALING PROBLEM

As mentioned earlier, each row of the separation matrix corresponds to a separation vector for different sources. However, the row order of W is arbitrary in every frequency bin, but for source separation, it is also essential to ensure the same order and arrangement of separation vector for each source in every frequency bin. For this, the directivity-pattern-based method is used, which requires the direction of arrival (DOA) of each source to be known. In the totally blind setup, this cannot be known so it is estimated from the directivity pattern (DP) of the separation matrix. The directivity pattern $F_k(f, \theta)$ of the microphone array is given by [13]

$$F_l(f, \theta) = \sum_{k=1}^m W_{lk}^{(ICA)}(f) \exp[j2\pi d_k \sin \theta / c], \quad (27)$$

where $W_{lk}^{(ICA)}(f)$ is an element of the separation matrix.

The DP of the separation matrix contains nulls in each source direction only. However, the position of nulls varies in each frequency bin for the same source directions. Hence by calculating the null direction in each frequency bin, the DOA of the source can be estimated as

$$\hat{\theta}_l = \frac{2}{N} \sum_{p=1}^{P/2} \theta_l(f_p), \quad (28)$$

where $\theta_l(f_p)$ denotes the null direction in the f_p th frequency bin for the l th source.

As an alternative, DOA can also be estimated by plotting a histogram of null directions occurring in each frequency bin. The center of the histogram corresponding to the maximum score of the null directions gives the DOA of the sources. However, this method is sensitive to permutation of the separation matrix. Therefore, before plotting the histogram, the separation matrix is depermuted using techniques based on similarity in the separation matrix for different successive bins. In each frequency bin (say the p th bin), separation matrix $W(f_p)$ is permuted by interchanging rows to generate separation matrix $W^{Per}(f_p)$ and then it is compared with separation matrix $W(f_{p-1})$ of the immediately prior frequency bin as follows:

$$D_1 = \sum_{i,j} |w_{ij}(f_p) - w_{ij}(f_{p-1})|, \quad (29)$$

$$D_2 = \sum_{i,j} |w_{ij}^{per}(f_p) - w_{ij}(f_{p-1})|. \quad (30)$$

The separation matrix is selected as

$$W(f_p) = \begin{cases} W(f_p) & \text{if } D_1 < D_2 \\ W^{per}(f_p) & \text{if } D_1 > D_2 \end{cases}. \quad (31)$$

Then the histogram method is used to compute DOA of the sources.

After estimating DOA, the gain value in each frequency bin is normalized in each source direction. Gain in the m th source direction in the p th frequency bin is given by

$$\alpha_m = \frac{1}{F_m(f_p, \hat{\theta}_m)}, \quad (32)$$

where $\hat{\theta}_m$ is the estimated direction of the m th source. Thus, a scaled separation matrix is obtained as

$$W(f) = \begin{bmatrix} \alpha_1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & \alpha_m \end{bmatrix} \begin{bmatrix} w_{11}(f_p) & \dots & w_{1n}(f_p) \\ \vdots & \dots & \vdots \\ w_{m1}(f_p) & \dots & w_{mn}(f_p) \end{bmatrix}. \quad (33)$$

This scaled and depermuted matrix is used, with Equation (22), to separate the signal in each frequency bin. Then by using the overlap-add technique, the signal is reconstructed. It is important to note that the estimated separation matrix have been estimated from whitened data so the separated signal is dewhitened in every frequency bin. If the signal is to be separated in the time domain, the separation matrix is dewhitened and then the P point FIR filter is used to separate signals, with Equation (26).

As the Newton's iterative formula is sensitive to the initial value of separation matrix, we use here NBF-based initial value for separation vector. The NBF-based initial value of $W(f)$ is generated as calculated in [13]. We give here the same result for the f_p th frequency bin in Equations (34(a,b,c,d)) which are based on forming the null in the source direction to be rejected, and gain is set to be unity in the source direction to be separated.

$$w_{11}^{BF} = -\exp[-q_1 \sin \theta_2] \times \{-\exp[q_2 (\sin \theta_1 - \sin \theta_2)] + \exp[q_2 (\sin \theta_1 - \sin \theta_2)]\}^{-1}, \quad (34a)$$

$$w_{12}^{BF} = -\exp[-q_2 \sin \theta_2] \times \{-\exp[q_1 (\sin \theta_1 - \sin \theta_2)] + \exp[q_2 (\sin \theta_1 - \sin \theta_2)]\}^{-1}, \quad (34b)$$

$$w_{21}^{BF} = -\exp[-q_1 \sin \theta_1] \times \{-\exp[q_1 (\sin \theta_2 - \sin \theta_1)] - \exp[q_2 (\sin \theta_2 - \sin \theta_1)]\}^{-1}, \quad (34c)$$

$$w_{22}^{BF} = -\exp[-q_2 \sin \theta_1] \times \{-\exp[q_1 (\sin \theta_2 - \sin \theta_1)] - \exp[q_2 (\sin \theta_2 - \sin \theta_1)]\}^{-1}, \quad (34d)$$

where

$$q_1 = j2\pi d_1 f_p \sin \theta_1 / c, \quad (34e)$$

Table 1. Signal analysis conditions

Frame length	20ms
Step size	10ms
Window	Hanning
FFT length	512 points

$$q_2 = j2\pi d_2 f_p \sin \theta_2 / c ,$$

and θ_1 and θ_2 are estimated DOAs of source 1 and source 2.

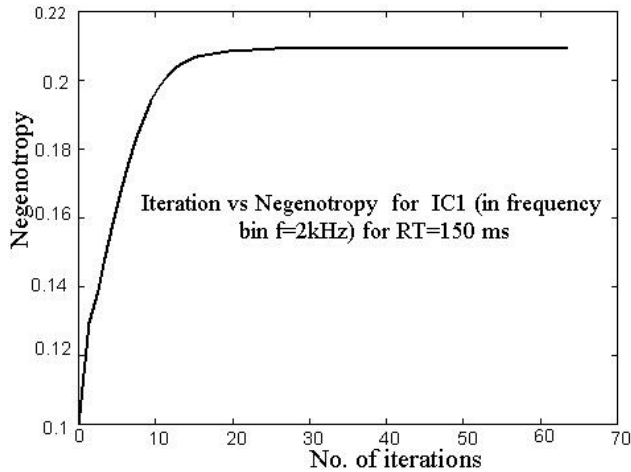


Fig. 2. Convergence curve for source 1 in 2 kHz.

6. EXPERIMENTS AND RESULTS

In the experiment, we used simulated data for a two-element linear microphone array with interelement spacing of 4 cm. Voices of two speakers (male and female) at a distance of 1.15 meters in the direction of -30° and 40° are used to generate mixed signals. Mixed signals at each microphone are obtained by convolving the seed speech with room impulse response, recorded under different acoustic conditions, characterized by different reverberation times (RTs), e.g., RT=0 ms, RT=150 ms and RT=300 ms. The unconvolved speech reaching each microphone are used as reference signals. The signal analysis conditions are shown in Table 1.

In the first phase of the experiment, we set random initial values of the separation matrix in each frequency bin and then used the proposed algorithm to compute the separation matrix for RT=0 ms, RT=150 ms and RT=300 ms data. The convergence curve for RT=150 ms is shown in Figure 2. The algorithm begins to converge after 20 iterations and stops when the stopping criterion is satisfied. As the stopping criterion, we measure the difference between separation vectors before and after the update and if the difference goes below 0.001, the iteration is stopped. Using directivity-pattern-based methods, DOAs of the sources are estimated. The directions of the 1st source S1 and 2nd source S2 estimated using Equation (26) and the histogram of the directivity pattern are presented in Table 2.

Table 2. DOA results

RT	RT= 0 ms		RT=150 ms		RT=300 ms	
	S1	S2	S1	S2	S1	S2
Averaging	-31.1	40.0	-32.2	39.0	-28.1	42.1
Histogram	-30.6	40.8	-30.9	37.8	-27.0	39.0

The histogram of the null directions for RT=150 ms is shown in Figure 3 and Figure 4. Using the estimated source direction, the separation matrix is scaled using Equation (29). As in the random-initial-value-based method, there is no information on DOA. Therefore, before calculating DOA by the histogram method, the separation matrix $W(f)$ is depermuted using Equation (30).

In order to evaluate the performance of the algorithm with NBF, the initial value of $W(f)$ is generated for every frequency bin in accordance with Equations (32(a,b,c,d)). Using these initial values in each frequency bin, ICA is performed using Equation (16). The separation performance for both the cases of NBF-based initial value and random initial values of the separation vector have been studied in different acoustic conditions. In every frequency bin, the speed of convergence and Noise Reduction Rate (NRR; output SNR - input SNR in dB) improvement of the algorithm have been computed. Corresponding results are shown in Fig. 5 and Fig. 6, respectively. It is obvious from Fig. 6 that NRR improvements for nonreverberant cases are almost the same for both random initial value and NBF-based guess values of the separation matrix. However, for reverberant conditions, NBF-based guess value gives good result in the performance as well as the convergence speed (see Fig. 5). For the second source, the algorithm shows the convergence in very few steps because the algorithm is deflationary and test data is mixtures of two signals only. So after extracting one signal source, it converges rapidly for second source (which is also last).

7. CONCLUSION

In this study, we derived a fixed-point learning rule using the Lagrangian multipliers optimization technique. We introduced the fixed-point ICA algorithm for convoluted audio source separation in the frequency domain. The performance of the developed algorithm was evaluated with different initial guess of the separation matrix. The algorithm converged rapidly, and convergence and NRR performance were improved when a good initial value of the separation matrix was used. Also, the histogram-based method for DOA estimation deteriorated under the heavily reverberant condition. Both methods can estimate DOA in the presence of two sources only.

However the effect on performance of using other initial values based, e.g., on signal subspace or transfer function between source and microphones, is yet to be investigated. The effect of combining gradient-based FDICA with fixed-point ICA is also left for future work. The slow convergence near the convergence point of the gradient-based ICA might be improved by adopting the fixed-point algorithm.

Acknowledgement

We acknowledge the valuable discussion with Dr. H. Sawada of NTT Co, Ltd on in this study.

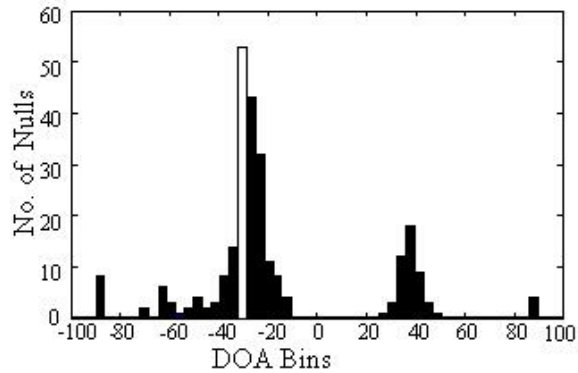


Fig. 3. Histogram of directivity pattern showing more nulls in the direction of first source.

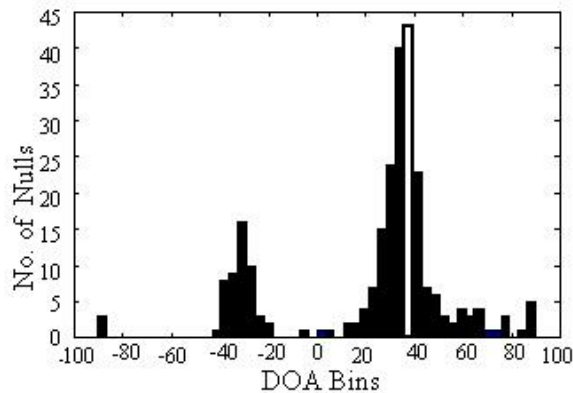


Fig. 4. Histogram of directivity pattern showing more nulls in the direction of 2nd source.

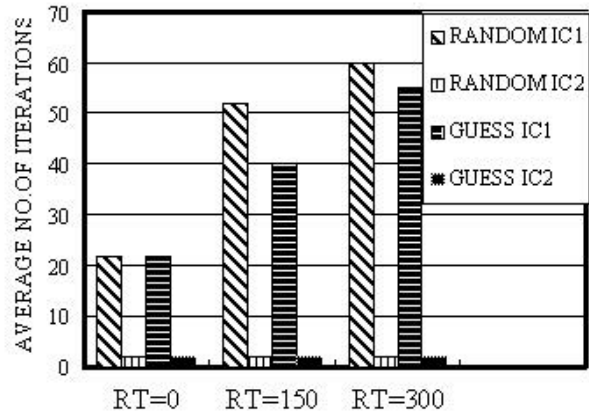


Fig. 5. Average number of iterations (Y-axis) required in different acoustic conditions characterized by different values of RT (X-axis) using random and NBF-based initial guess value.

- 8 E. Bingham et al., "A fast fixed point algorithm for independent component analysis of complex valued signal," *Int. J. of Neural System*, 10(1): 8, 2000.

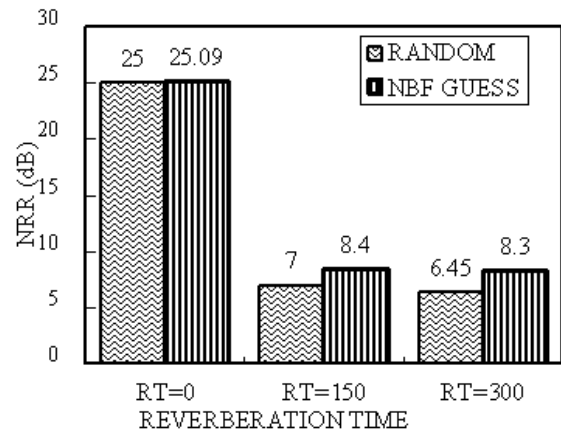


Fig. 6. NRR improvement using NBF guess and random initial value in different acoustic environment.

References

- 1 P. Common, "Independent component analysis, a new concept?," *Signal Processing*, vol.36, pp.287-314, 1994.
- 2 P. Samaragadis, "Blind separation of convolved mixture in the frequency domain," *Neurocomputing*, vol.22, pp.21-34, 1998.
- 3 T. W. Lee, "Independent component analysis," Kluwer Academic Publishers, 1998.
- 4 K. Torkkola, "Blind separation for audio signals-are we there yet?," *Proc. Workshop on ICA & BSS, France*, 1999.
- 5 A. Hyvarinen, "Survey on Independent Component Analysis," *Neural Computing Surveys*, vol.2, pp.94-128, 1999.
- 6 F. Asano et al., "Combined approach of array processing and independent component analysis for blind separation of acoustic signal," *Proc. ICASSP2001*, pp.MULT-P2.1, Salt Lake City, 2001.
- 7 H. Saruwatari et al., "Fast-convergence algorithm for blind source separation based on array signal processing," *IEICE Trans. Fundamentals*, vol.E86-A, no.3, 2003.
- 9 N. Mitianoudis, N. Davies, "New fixed point solution for convolved audio source separation," *Proc. IEEE Workshop on Application of Signal Processing on Audio and Acoustics*, New York, 2001.
- 10 A. Hyvarinen et al., "Independent component analysis," John Wiley & Sons, 2001.
- 11 H. Johnson et al., "Array Signal Processing, concepts and techniques," Prentice Hall, 1993.
- 12 A. Cichocki, S. Amari, "Adaptive blind signal and image processing, learning algorithms and application," John Wiley & Sons Ltd., pp.130-31, 2002.
- 13 H. Saruwatari, T. Kawamura, K. Shikano, "Blind source separation for speech based on fast convergence algorithm with ICA and beamforming," *Proc. EUROSPEECH2001*, pp.2603-2606, 2001.