European Journal of Applied Sciences 8 (2): 76-86, 2016
ISSN 2079-2077

DOI: 10.5829/idosi.ejas.2016.8.2.22872

SImplementation of Frequency Domain Approach Using Instantaneous Mixing Auto Recursive for Separation of Speech Signals

[1]*C. Anna Palagan and* [2]*K. Parimala Geetha*

[1]ECE, Rajas Engineering College, Vadakkangulam, Tirunelveli, Tamilnadu, India
[2]ECE, Ponjesly College of Engineering, Nagercoil, Tamilnadu, India

Abstract: In the present work a novel algorithmic rule by taking the speech from two different microphones and separate these speeches by prediction of separating speech mixtures that is predicated on separation matrices is planned. In multi- talker applications so as to boost individual speech sources from their mixtures is done by Blind source Separation (BSS) ways. From the previous published works of separation of speech signals, the main disadvantage is that the incidence of distortion present within the signal that affects separated signal with loud musical noise. The idea for speech separation in standard BSS ways is simply one sound source in a single room. The proposed methodology uses as a network that has the parameters of the IMAR model for the separation matrices over the complete frequency vary. An attempt has been made to estimate the best values of the IMAR model parameters, Φ_W and Φ_G by suggests that of the maximum-likelihood estimation methodology. Based on the values of these parameters, the source spectral part vectors are estimated. The entire set of TIMIT corpus is employed for speech materials in evolution results. The Signal to Interference magnitude Relation (SIR) improves by a median of 6dB sound unit over a frequency domain BSS approach.

Key words: Blind Source Separation (BSS) · Separation Matrices · Instantaneous Mixing plus Auto Regressive (IMAR) Model · Maximum Likelihood Estimation

INTRODUCTION

The audio and speech signal processing applications, the separation of speech signals is very important done by using Blind Source Separation (BSS) technique. The BSS has been used in multi talker applications and acoustics signal processing. Several adaptive step size methods for BSS for Robot audition systems have been proposed. The parameters are not adjusted manually and there is no need of additional preprocessing. For the moving sources [1-19] the positions and velocities of the source is obtained from the 3-D tracker based on Markov Chain Monte Carlo particle filter which results in high sampling frequency. In this method it provides separation of the sources without any prior knowledge of moving sources and it is used to perform real time speech enhancement. In frequency domain, the permutation problems [13] a new technique which partitioned the full frequency bands into small regions by using correlation of separate signal powers. The region wise permutation alignment is performed by region growing manner. Here the permutation alignment is based on inter frequency dependence of separated power signal.

For multi channel acoustic echo cancelation [8] the ICA is jointly perform source separation and multichannel acoustic cancellation through semi BSS without double track detection. To reduce the effect of non uniqueness the matrix constraint is used. For detecting a time varying mixing matrix the short time Fourier transform [20] is used. In frequency domain [13, 6] for reducing the computational complexity and increase the speed frequency domain BSS is used. In this algorithm the higher order frequency dependencies which employing real conference room recordings. For extracting independent from array signals the Bi iterative algorithm [3] is used. For solving the non unitary joint diagonalization problem in BSS a simultaneous Bi-iterative algorithm is introduced. In the multiple sources case, in order to find the accurate estimation of propagation time delays [7]. Generalized State Coherence

Corresponding Author: C. Anna Palagan, ECE, Rajas Engineering College, Vadakkangulam, Tirunelveli, Tamilnadu, India.

Transform (GSCT) which is a non-linear transform of the space represented by the whole demixing matrices. In convex divergence ICA algorithm, [11] the source signals of the blind sources is derived by the characteristics of Parzen window based distribution. Based on the experimental results this algorithm is the fast algorithm for the blind sources which involves speech and music signals. For cancelling the echo's in the blind sources during continuous double track [10] in order to estimate the blind source signals the maximum likelihood approach is used.

Blind source separation (BSS) methods aim to achiseve this goal, based on some prior knowledge of the source signal properties. Following the physics of sound mixing, let us consider N sources $s_m(t)$, m = 1, •••, N, to be convolutively mixed. At M sensors, the recorded mixture signals $a_i(t)$, i = 1, •••, M, is denoted by

$$a_i(t) = \sum_{m=1}^{N}\sum_{d=1}^{L} g_{is}(d)\, s_m(t-d) \qquad (1)$$

where L is that the delay length on the order of $10^3 - 10^4$ faucets (each faucet last $1/F_s$ second wherever F_s is that the sampling frequency) in an exceedingly commonplace area $g_{is}(d)$ is that the separate Green's perform of the area, conjointly called the area impulse response (RIR). The (severely ill-posed) mathematical drawback is to recover each $G_{is}(d)$ and $s_m(t)$ from $A_i(t)$. A serious branch of BSS is that the thus referred to as freelance element analysis (ICA) that assumes that the supply signals square measure orthogonal to (or freelance of) one another [2]. ICA is a more general methodology than recovering sound signals. The time domain ICA [1, 5] attempts to estimate the g_{is}'s directly in order to deal with a high dimensional nonconvex optimization problem [2]. Frequency domain ICA [15] & [16] solves an instantaneous (L = 0) version of (1) in each frequency bin after applying the discrete Fourier transform (DFT) to (2) frame by frame:

$$A_i(f,\tau) \approx \sum_{m=1}^{N} G_{is}(f)\, S_m(f,\tau) \qquad (2)$$

where (A_i, G_{is} & S_m) square measure the T -point DFT of (a_i, g_{is} & s_m) severally and τ is that the frame variety. The larger T/l square measure, the higher the approximation. Because of the absence of the regularity in d of g_{is} and s_m, DFT doesn't rework convolution to native product precisely. The frequency domain approach is limited to use a long DFT. In addition to computations to sort out scaling and permutation ambiguities when synthesizing multi-frequency estimation of $S_m(f,\tau)$ back to a time domain output [2, 14]. Imperfections and errors in scaling and permutation in the frequency domain may lead to artifacts in the time domain signals at the final output.

Existing Methods and New Idea: BSS refers to the problem of recovering signals from several observed linear mixtures. Up to now, solving the BSS problem in an underdetermined case has mainly consisted in assuming that the speech signals were sufficiently sparse [22, 21], However, due to unexpected discontinuous zero-padding, such separated signals have considerable distortion and therefore a loud musical noise is heard. In [4], an estimated mixing matrix was used for solving the determined BSS problem. Our suggestion for eliminating the distortion matter is to combine sparseness with mixing matrix estimation. Indeed we can obtain more information about the signals to be separated and to reduce the zero-padding effect, from which the musical noise originates. Whereas Vielva *et al*., Rickard and Yilmaz worked on an undetermined instantaneous case employing sparseness [22, 21, 4] and Deville on a determined instantaneous case utilizing a mixing matrix estimation [4], here, we are dealing with undetermined BSS in a convolutively case.

Blind Source Separation of Speech and Music Signals: Blind source separation (BSS) could be a technology for separating mixtures of multiple speech signals. This technology has been studied extensively and vital progresses are remodeled the last decade. However, typical BSS ways performs terribly poor once the reverberation time is massive. Many researchers have addressed this downside, however it's still associate open question. Our approach to overcoming this limitation is to unify BSS and performing BSS will realize BSS even under highly reverberant environments.

Figure 1 is an example for speech separation of an indoor auditorium problem. For these applications, the instant combination model could also be applicable as a result of the propagation delays is negligible. However, in real environments substantial time-delays could occur Associate in a design and algorithmic rule is required to account for the blending of time-delayed sources and convolved sources. It focuses on the implementation of the learning algorithm and on issues that arise when separating speakers in room recordings. It used associate infomax approach during a feed forward network enforced

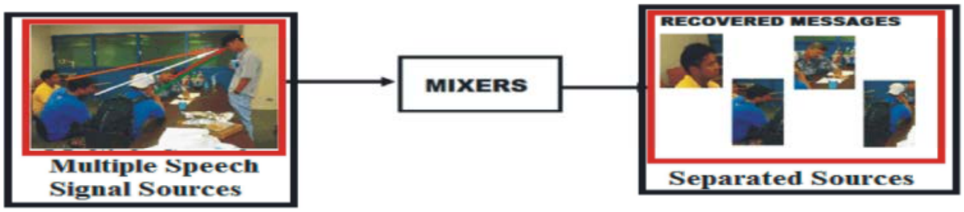


Fig. 1: Indoor Auditorium Problem

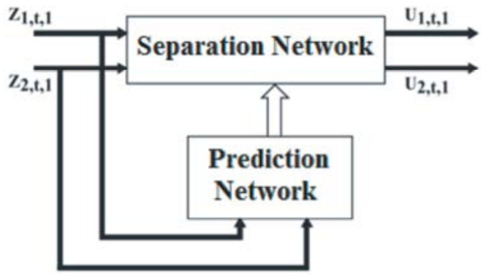


Fig. 2: Diagram of Blind Source Separation method

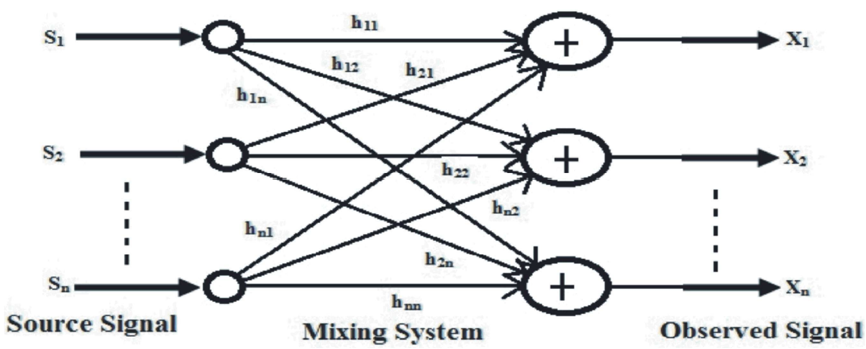


Fig. 3: Structure for Mixing of Speech signals

within the frequency domain mistreatment the polynomial filter algebra technique. Beneath minimum-phase mix conditions this preprocessing step was adequate for the separation of signals. These strategies with success separated a recorded voice with music within the background (indoor auditorium problem).

Proposed Method

Imar Model to Generate Microphone Signals: In the proposed method the BSS is recovered of source signals by LTI filter and permutation. The time domain Blind Source Separation approach is used here. In this proposed method the estimation of blind source signal is in the form of source signal vector B_s (n) by applying an I_M input signals and I_S output separation filter value to observed signal vector O (n). In the time domain BSS approach for separating sound mixtures the order of the separation filter is set a value that exceeds the room reverberation time. The order of the separation filter becomes very large for the reverberation time is long. So the convergence rate is poor and the cost for computation is very high. The estimation of source spectral component vector in the frequency domain BSS approach is done by applying a separation matrix to the observed spectral component vector.

In this proposed method the assessment is a multiple sound source case, where $I_M = 2$. In this consideration the frequency domain BSS approach as shown in Figure. 2 is by using WPE method as a preprocessor which illustrate the case of $I_S = I_M = 2$. Because in the first step we use prediction error with first microphone for BSS process. Figure 3 shows the structure for Mixing of Speech signals.

For any microphone ~~as~~ the prediction target as

$$P_{n,u,v} = O_{n,u,v} - \sum_{s=L_i}^{L_i+M_i-1} h^{G}_{n,s,v} O_{n,u,v}; v \le n \le I_M \quad (3)$$

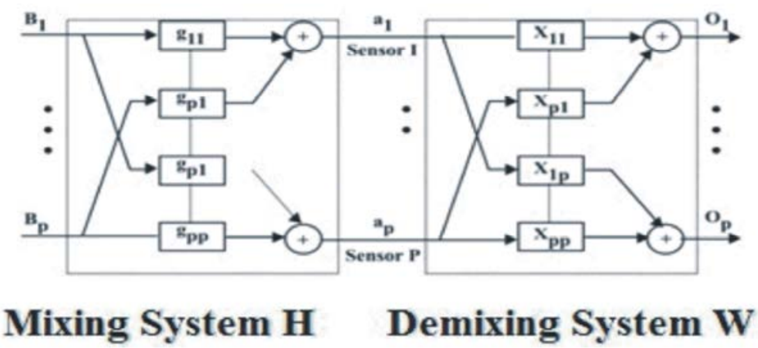


Fig. 4: Filter representation of Mixing and Demixing of speech signals

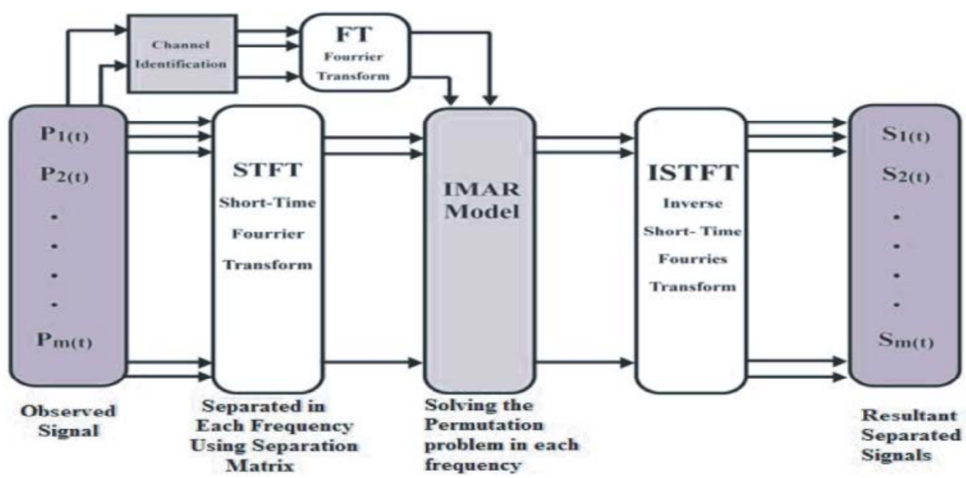


Fig. 5: Representation of Instantaneous Mixing plus Auto Regressive model

where $\{h_{n,s,v}\}$ L_i= L_i+M_i-1 denoted the prediction filter for the I^{th} microphone spectral component and $P_{n,u,v}$ is the corresponding prediction error. The different spectral component outputs $P_{1,u,1}$…..$P_{Im,u,1}$ can be obtained. The instantaneous mixtures of the source spectra components were considered for these components. Based on the theory of Multichannel linear prediction the values of $P_{1,u,1}$,……$P_{Im,u,1}$ become nearly instantaneous mixtures by using appropriate prediction filters although such prediction filters may not be able to obtained with the WPE method. For the m^{th} microphone the prediction filter values are $D_{n,s,v}$.

It is assumed that the bin indices is 1 for all frequencies from the set of exists values of X_1 and $\{D_{n,s,v}\}$ L_l=L_l+M_l-1that is equalize the output spectral component vector $B_{u,1}$ based on these assumptions we identified $Z_{u,1}$ with $B_{u,1}$ is

$$Q_{u,v} = \sum_{S=L_l}^{L_i+M_l^{-1}} H_{S,v}^{G} O_{u-s,v} + P_{u,v} \tag{4}$$

The assumption taken in the equation (4) will not completely hold in real time so further experimental detail is desired. So in IMAR model it performs high separation of speech signal based on the possible assumption is at least partially show the practical validity of this assumption. Figure 4 shows the filter representation of mixing demixing of speech signals.

The set of equations (3) and (4) represents the generation of IMAR model for the observed spectral component vector $O_{u,1}$. In need of this mode may be interpreted as follows. The individual sound source signals and the spectral component of the sound source was given in equation (4) are instantaneously mixed together with mixing X_1^{-1} to form $P_{u,1}$. In the equation (3) the mixing of the remaining elements present in $P_{u,1}$ with the multichannel AR system with regression or prediction matrices $\{H_{s,1}\}$= L_l=s=L_l+s_l-1 to generate the observed spectra component vector $O_{u,1}$.

Figure 5 represents the Instantaneous Mixing plus Auto Regressive (IMAR) model. The latent spectral component vector is unobservable by $P_{u,v}$. The separation

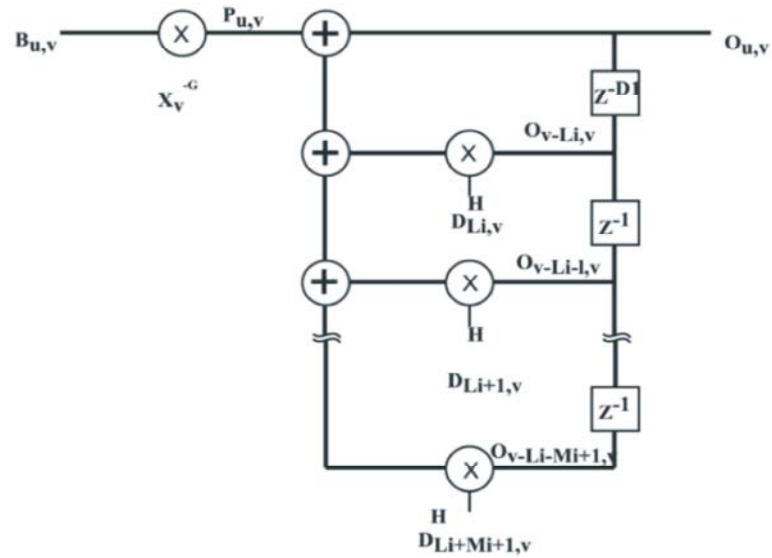


Fig. 6: Analysis filter for Instantaneous Mixing plus Auto Regressive model

matrices and prediction matrices parameters of IMAR model over the entire frequency range is denoted by Φ_W and Φ_G. Figure 6 shows the analysis filter for instantaneous mixing plus auto regressive model. The following expressions represent the set of separation matrices and prediction matrices as

$$\Phi_W\{W_v\};\quad 0 = v = K\text{-}1 \tag{5}$$

$$\Phi_G = \{\{D_{s,v}\}L_l \le s \le L_v + S_v^{-1};\ 0 \le v \le K\text{-}1 \tag{6}$$

By using the maximum likelihood estimation the ideal values of IMAR model parameters Φ_W and Φ_G. The parameter values of the source spectral component vectors are estimated from equations (3) and (4). The matrix form is represented for the IMAR model in the MIMO filter which gives the relationship between the IMAR model and the frequency domain. The expression for matrix form of $X_{s,v}$ as

$$X_{s,v} = \begin{cases} X_v & ;\ if \quad s = 0 \\ 0 & ;\ if \quad 1 \le s \le L_v \\ D_{s,v} X_v & ;\ else \end{cases} \tag{7}$$

where 0 is $z_{u,v} = \sum_{s=0}^{l_v+s_v-1} X^G_{s,v} O_{u-s,v}$ a zero matrix. Then (3) and (4) may be summarized in one equation as

Parameters Estimation Based on Maximum-likelihood Function: The IMAR mode parameters Φ_W and Φ_G was optimized by using the maximum log likelihood function. For finding the source spectrum IMAR model first find out the time varying all pole model. By using the time varying all pole and IMAR model we derive the log likelihood function. The set of Linear Predictive Coding and PRPs over the whole observation period we use Φ_s and it is represented as

$$\Phi_k = \{a_{m,t}, \sigma^2_{m,t}\}\ 1 \le m \le M,\ 0 = t = N_F \text{-}1 \tag{9}$$

The Linear Predictive Coding and PRPs are collectively called the all pole parameters. It is assumed that, the speech signal is taken as short time based signal which has a short time frame is 18ms. So that in the present works the short time analysis of frame length 18ms. In added with this the smallest frame length less than 18ms is not considered because its violates the assumption.

The source spectrum parameters of both LPCs and PRPs can be observed by means of probability distribution function is $H_S = \{h_{u,v}\}$ $0 = u = M_G\text{-}1$, $0 = l = D\text{-}1$. The probability density function for the source spectral component vector $k_{u,v}$ by the assumptions of h (2) and h (3) can be written as

$$p(k_{u,v};\Phi_k) = N_C\{\ k_{u,v};0;kA_{u,v}\} \tag{10}$$

where o is a zero vector $kA_{u,v}$ is the diagonal matrix defined as

$$kA_{u,v} = \begin{bmatrix} k\lambda_{1ut,v} & 0 \\ 0 & k\lambda_{N,u,v} \end{bmatrix} \tag{11}$$

The $P_{u,v}$ is the probability distribution function of the latent spectral component vector which is taken by the equations (5) and (11).

$$p(P_{u,v;}\ \Phi_S, \Phi_W) = N_C\ \{P_{u,v};0,pA_{u,v}\} \tag{12}$$

where the covariance matrix $pA_{u,v}$ is given by

$$pA_{u,v} = (X_{MN}A_{u,v}{}^{-1}X_I^G)^{-1} \tag{13}$$

The past sequence of the probability function for observed spectral component vector is taken from the equations (5) and (13) and it is represented by

$$p(O_{u.v} \mid O_{u-1.v}\cdots,O_{0,v};\Phi_S,\Phi_W,\Phi_G = $$

$$N_C\left\{O_{u,v};\sum_{s=L_V}^{L_V+S_V-1} H_{s,v}^G O_{u-s.v}, PA_{u,v}\right\} \tag{14}$$

The probability density function for observed data is given by

$$P(O_F;\Phi_S;\Phi_W;\Phi_G) = \sum_{v=0}^{D-1}\sum_{u=0}^{N_F-1} p(O_{u,v} \mid O_{u-1.v},\cdots O_{0,v};\Phi_S\Phi_W\Phi_G) \tag{15}$$

The maximum log likelihood function was obtained by taking the log value of equation (15).

$$L(\Phi_S\Phi_W\Phi_G\Phi_F) = \sum_{v=0}^{D-1}\sum_{u=0}^{N_F-1}\left\{\log\left|pA_{u,v}\right|^{-1} - \right.$$

$$\left.\left(O_{u,v} - \sum_{s=L_1}^{L_1+S_1-1} H_{u,v}^G O_{u-s,v}\right)^G\right\} \tag{16}$$

From the equations [1-16] it is identified that the maximum log likelihood function not depends on the separation matrices Φ_W and prediction matrices Φ_G but also depend on the pole parameter Φ_S. So the maximum log likelihood function is calculated which depends on Φ_S, Φ_W and Φ_G is used as the estimates of the ideal values of Φ_W and Φ_G respectively.

RESULTS AND DISCUSSIONS

In the proposed method two sources and two microphones are used for testing the speech signals. The complete test used a set of TIMIT corpus which includes 38 male speakers, 8 female speakers and 145 utterances. The sampled frequency used for the testing of acoustic signals of these utterances is 14 KHz and the bandwidth is limited between 70 Hz to 4 Khz frequency range. The data's from 26 male speakers are taken and it can be used to from male – male utterances pairs and the data's from 5 female speakers are taken and it can be used for female – female utterances pairs. The remaining data's from 12 male speakers and 3 female speakers are taken and it can be paired to form the remaining utterances. So in total 55 male – male, 20 female – female and 70 male – female utterances pairs were generated. We take each utterances pair and it can be mixed with the acoustic signals of the two utterances with the room impulse responses measured in a varechoic chamber to simulate signals which can be observed from the microphones

Figure 7 shows an experimental setup developed for the present work. The Signal to Interference Ratio (SIR) and the Direct to Reverberation Ratio (DRR) should be evaluated from each trial taken by the experimental setup. The I_M^{th} source from the I_M^{th} microphones are described by the following equations

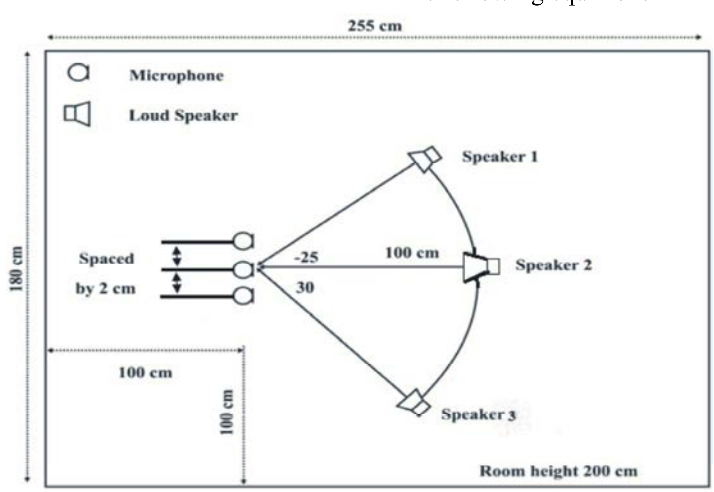


Fig. 7: Experimental setup

$$SIR_{I_S,I_M} = 10 \ \log 10 \frac{\sum_{n=0}^{N_T-1} O_{I_M}^{I_S}(n)^2}{\sum_{n=0}^{N_T-1}\left(\sum_{I \neq I_S} O_{I_M}^{I}(n)\right)^2} \tag{17}$$

The component of I_M^{th} microphone signal is given by $O_{Im}^{Is}(n)$ which is originating from the I_S^{th} source. Then the value of $O_{Im}^{Is}(n)$ is given by the following equations as

$$O_{I_M}^{I_S}(n) = \sum_{s \geq 0} b_{I_S,I_M}(s) S_{I_S}(n-s) \tag{18}$$

The room impulse response is given by $\{b_{Is,Im}(s)\}$ with $s \geq 0$ from the I_s^{th} source speech signal to the I_m^{th} microphone. The index value of the microphone is obtained from I_s where the source speech signal I_s appears most prominently as

$$\text{mic}(I_S) = \text{argmax}\ \{SIR_{Is,Im}\} \tag{19}$$

By using the source signal I_s^{th} the input SIR and DRR value is computed as

$$SIR_{Is} = SIR_{Is,mic(Is)} \tag{20}$$

$$DRR_{I_S} = 10\log_{10}\frac{\sum_{n=0}^{N_T-1} O_{mic(I_S)}^{I_S,D}(n)^2}{\sum_{n=0}^{N_T-1} O_{mic(I_S)}^{I_S,R}(n)^2} \tag{21}$$

The direct reverberation components from the experimental values of O_{Im}^{IsR} (n) are find out from the values of O_{Im}^{IsD} and O_{Im}^{IsR} respectively. So from the above experimental values the direct to dereverberent components are described as

$$O_{I_M}^{I_S,D}(n) = \sum_{s=0}^{\Delta} b_{I_S,I_M}(s) S_{I_S}(n-s) \tag{22}$$

$$O_{I_M}^{I_S,R}(n) = \sum_{s=\Delta+1}^{\infty} b_{I_S,I_M}(s) S_{I_S}(n-s) \tag{23}$$

The value of early reflection delay Δ was set as 40 ms. The signal to interference ratio and the direct to reverberation ratio output also defined in the same way. From the experimental output each output signal is decomposed according to the sources in order to find the values of output signal to interference ratio. Let us consider $W_{Io}(n)$ denote the output signal I_O^{th} and $W_{Io}^{Is}(n)$ denote the I_s^{th} source component. For these calculations of processing microphone signals $O_1^{Is}(n)$,….., $O_{Im}^{Is}(n)$ with the estimated prediction and separation matrices. For calculating the direct to reverberation ratio we need the impulse responses from the sources to the outputs. For estimating the impulse responses it uses the least mean squares matching. The impulse response is estimated from the source speech signal I_S to the output speech signal I_O. The matching errors are taken by the assumptions of -20 dB. This experimental result indicates the IMAR model is effective for BSS.

Figure 8 shows the input samples of speech signals from two sources. The corresponding short time Fourier transform was illustrated in Figure. 9. The instantaneous mixing of the samples based on the IMAR model was shown in Figure. 10. the separated speech signals of the blind sources are illustrated in different forms for various separable speech signals are given in Figure. 11.

Figure. 12 and Figure. 13 illustrated the average variations SIRs and DRRs based on different reverberation. From that figure we can clearly mentioned the IMAR model gives better outperformed in the frequency domain blind source separation for both reverberation conditions in terms of average SIR. Figure 14 shows the separated speech with echo canceller by using IMAR for μ=0,025 and 0.04. Figure 15 represented the separated speech with noisy environment.

Figure 16 shows the average SIRs for each reverberation condition. We infer that the methodological difference between the ICA, CSD and IMAR models leads to the difference in SIR improvement.

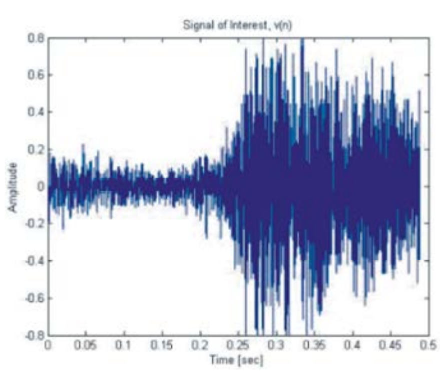


Fig. 8(a): Input Signal

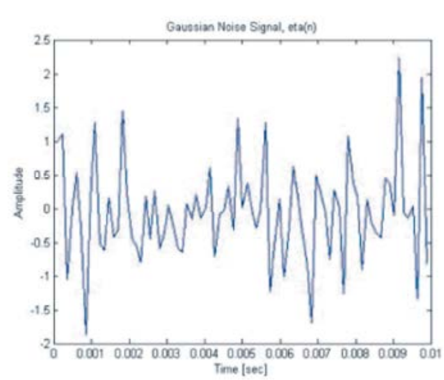


Fig. 8(b): Noise Signals

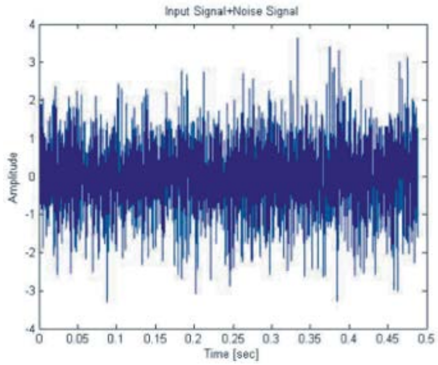


Fig. 9(a): Input Signal + Noise Signal

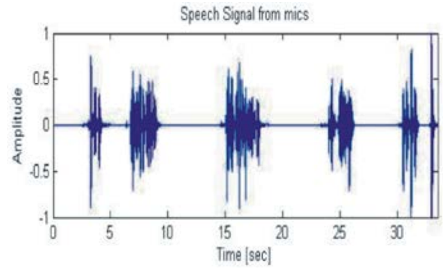


Fig. 9(b): Speech Signals from Mics

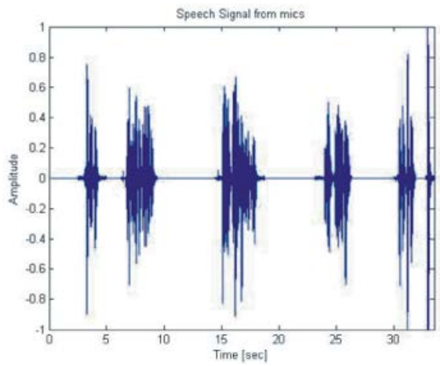


Fig. 10(a): Speech Signals from mic

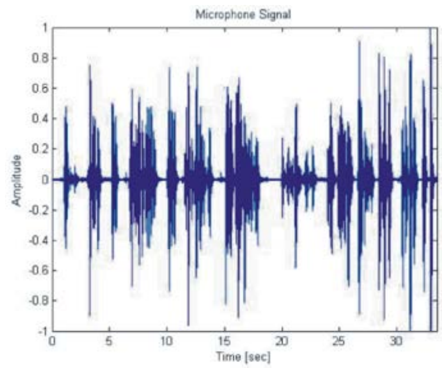


Fig. 10(b): Speech Signals from mic

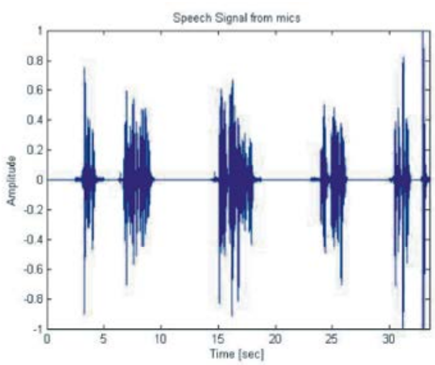


Fig. 11(a): Speech Signals from mics

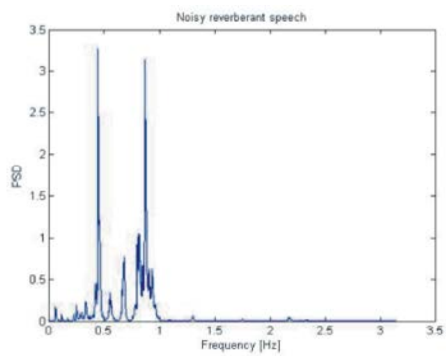


Fig. 11(b): Noisy Reverberant Signal

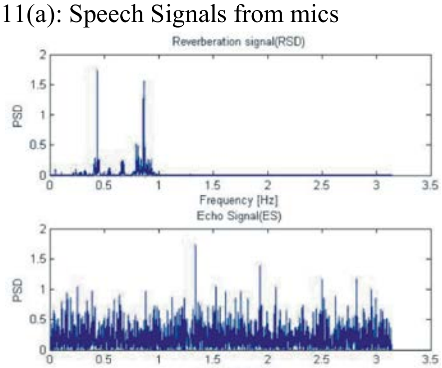


Fig. 12(a). Reverberation and Echo Signal

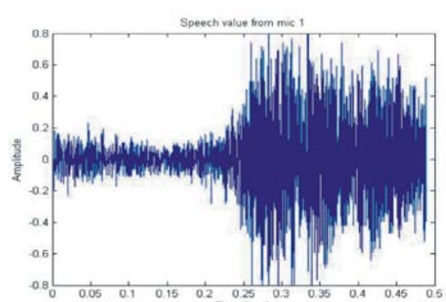
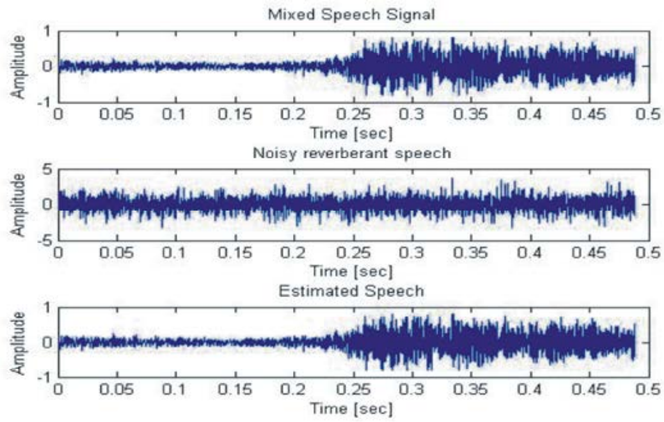


Fig. 12(b): Speech value from Mic



Fi.g. 13: Estimated Speech Signal

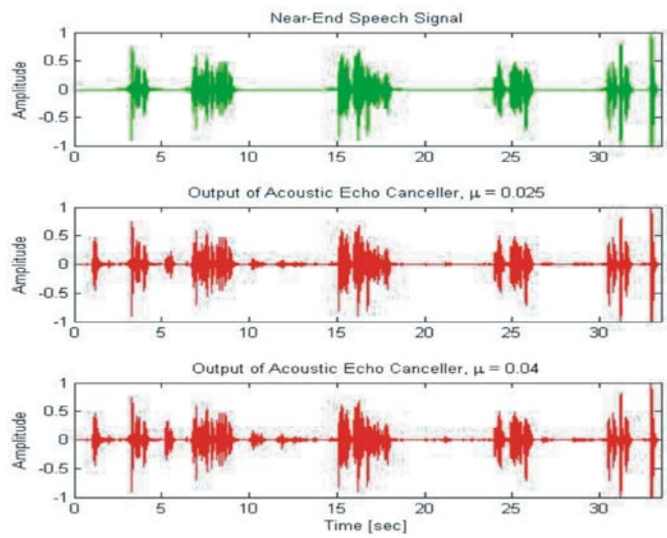


Fig. 14: Separated Speech with Echo Canceller by using IMAR for μ=0,025 and 0.04

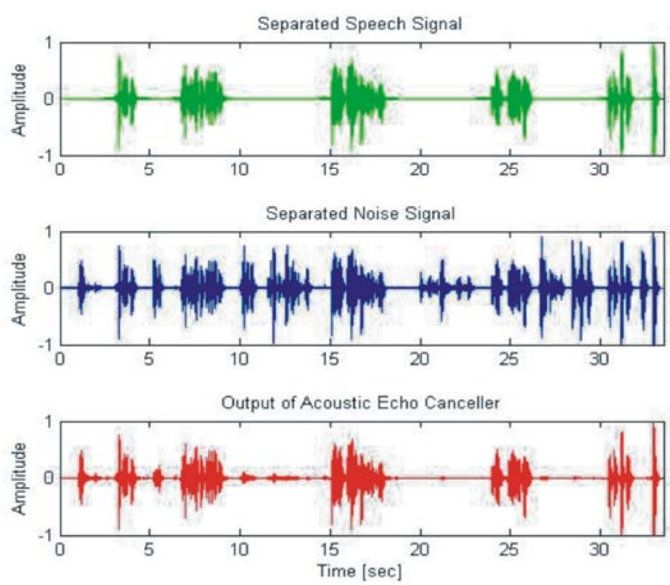


Fig. 15: Separated Speech with noisy environment

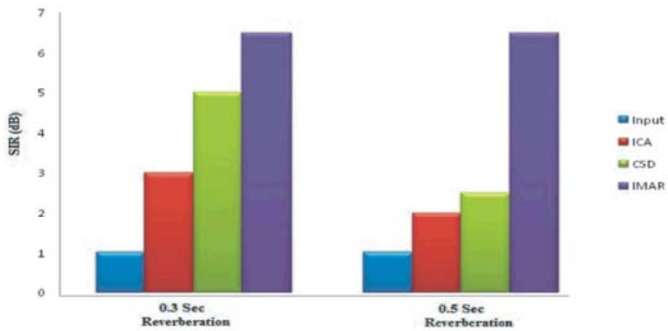


Fig. 16: Average SIRs for each reverberation condition

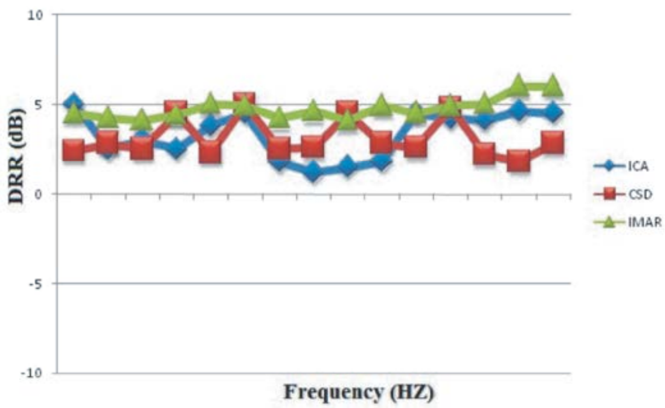


Fig. 17: Comparison of DRR between ICA, CSD by using IMAR model

Table 1: Average SIR and DRR Increases In Decibel

Gender for Speaker 1		Male	Female
Gender for Speaker 2		Female	Male
SIR	Speaker 1	5.21	4.72
	Speaker 2	5.26	4.88
	Speaker 3	5.30	4.92
DRR	Speaker 1	4.32	4.12
	Speaker 2	4.51	4.42
	Speaker 3	4.68	4.58

Table 2: Comparison of SIR Improvement Between Different Algorithms and the Proposed Method

Algorithms	SIR Improvements in dB
ICA	03
CSD	04
IMAR	6.5

For an experimental study the 0.3 sec and 0.5 sec reverberation time are considered. The effects of male voices and female voices for the separation of speech signals are estimated. Table I represent the average changes in SIR and DRR for male female and female male pair by taking speaker 1 is male then speaker 2 is female and another one is reverse of this. Table II lists the difference between the three algorithms based on the SIR improvements in those methods. Finally the graphs represent the comparison between the ICA, CSD and IMAR methods used for separation of speech signals. From the above results and comparative tables we proved that IMAR model is the best one which is used for the applications of separation of speech signals. Figure 17 shows the comparison of DRR between ICA, CSD by using IMAR model [17-22].

CONCLUSION

The present work is carried out to design the effectively separate of speech signal from the blind Source Separation by using the method of Instantaneous Mixing Auto Regressive method and the maximum likelihood function. The key features presented in the Instantaneous Mixing Auto Regressive method is that optimized separation of speech signals and thereby enabling us to perform a blind source separation process in consideration. In the present method the signal to interference rate improves over 6 dB. By using Instantaneous Mixing Auto Regressive method it attained good signal to interference ratio and direct to reverberation ratio even when a reverberation time was 0.3 s. It is concluded that the Instantaneous Mixing Auto Regressive method provides a powerful tool for microphone array signal processing in a reverberant room impulse response.

REFERENCES

1. Amari, S., S.C. Douglas, A. Chichocki and H.H. Yang, 1997. Multichannel blind deconvolution and equalization using the natural radiant, Proc. IEEE Workshop Signal Proc. Adv. Wireless Commun, pp: 101-104.
2. Choi, S., A. Cichocki, H. Park and S. Lee, 2005. Blind source separation and independent component analysis: A review. Neural Inform. Process. Lett. Rev., 6: 1-57.
3. Da-Zheng Feng, Hua Zhang and Wei Xing Zheng, 2011. Bi-Iterative Algorithm for Extracting Independent Components From Array Signals, IEEE Transactions on Signal Processing, 59(8): 3636-3646.
4. Deville, Y., 2003. Temporal and time frequency correlation-based blind source separation methods. Proc. ICA2003, pp: 1059-1064.
5. Douglas, S.C. sansd M. Gupta, 2007. Scaled natural gradient algorithms for instantaneous and convolutive blind source separation, IEEE Int. Conf. Acoust., Speech, Signal Processing, 2: 637-640.
6. Clark Felipe, S.P., Mariane R. Petraglia and Diego B. Haddad, 2011. A New Initialization Method for Frequency-Domain Blind Source Separation Algorithms, IEEE Signal Processing Letters, 18(6): 343-346.
7. Nsesta Francesco and Maurizio Omologo, 2012. Generalized State Coherence Transform for Multidimensional TDOA Estimation of Multiple Sources, IEEE Transactions on Audio, Speech and Language Processing, 20(1): 246-261.
8. Nesta Francesco, Ted S. Wada and Biing-Hwang (Fred) Juang, 2011. Batch-Online Semi-Blind Source Separation Applied to Multi-Channel Acoustic Echo Cancellation. IEEE Transactions on Audio, Speech and Language Processing, 19(3): 583-599.
9. Nakajima Hirofumi, Kazuhiro Nakadai, Yuji Hasegawa and Hiroshi Tsujino, 2010. Blind Source Separation with Parameter-Free Adaptive Step Size Method for Robot Audition. IEEE Transactions on Audio, Speech and Language Processing, 18(6): 1476-1485.
10. Gunther Jake, 2012. Learning Echo Paths during Continuous Double-Talk Using Semi-Blind Source Separation, IEEE Transactions on Audio, Speech and Language Processing, 20(2): 646-664.
11. Chien Jen-Tzung and Hsin-Lung Hsieh, 2012. Convex Divergence ICA for Blind Source Separation, IEEE Transactions on Audio, Speech and Language Processing, 20(1): 302-315.
12. Marin-Hurtado Jorge, I., Devangi N. Parikh and David V. Anderson, 2012. Perceptually Inspired Noise-Reduction Method for Binaural Hearing Aids, IEEE Transactions on Audio, Speech and Language Processing, 20(4): 1372-1382.
13. Wang Lin, Heping Ding and Fuliang Yin, 2011. A Region-Growing Permutation Alignment Approach in Frequency-Domain Blind Source Separation of Speech Mixtures. IEEE Transactions on Audio, Speech and Language Processing, 19(3): 549-557.
14. Liu, J., J. Xin, Y. Qi and F.G. Zeng, 2009. A time domain algorithm for blind separation of convolutive sound mixtures and l1 constrained minimization of cross correlations. Commun. Math. Sci., 7(1): 109-128.
15. Murata, N., S. Ikeda and A. Ziehe, 2001. An approach to blind source separation based on temporal structure of speech signals. Neurocomp., 41: 1-24.
16. Parra, L. and C. Spence, 2000. Convolutive blind separation of nonstationary sources. IEEE Trans. Speech Audio Processing, 8(3): 320-327.
17. Rickard, S. and O. Yilmaz, 2002. On the approximate w-disjoint orthogonality of speech. Proc. ICASSP2002, 1: 529-532.
18. Smaragdis, P., 1998. Blind separation of convolved mixtures in the frequency domain. Neurocomp, 22(1-3): 21-34.
19. Syed Mohsen Naqvi, Miao Yu and Jonathon A. Chambers, 2010. A Multimodal Approach to Blind Source Separation of Moving Sources, IEEE Journal of Selected Topics in Signal Processing, 4(5): 895-910.
20. Sun Tsung-Ying, Chan-Cheng Liu, Shang-Jeng Tsai, Sheng-Ta Hsieh and Kan-Yuan Li, 2011. Cluster Guide Particle Swarm Optimization (CGPSO) for Underdetermined Blind Source Separation with Advanced Conditions, IEEE Transactions on Evolutionary Computation, 15(6): 798-811.
21. Vielva, L., D. Erdogmus, C. Pantaleon, I. Santanmaria, J. Pereda and J.C. Principe, 2002. Underdetermined blind source separation in a time-varying environment, Proc. ICASSP2002, 3: 3049-3052.
22. Zibulesky, M., B.A. Pearlmutter, P. Bofill and P. Kisilev, 1999. Blind source separation by sparse decomposition in a signal dictionary. TR No. CS99-1, University of New Mexico, Albuquerque.