International Journal of Electronic Engineering Research
ISSN 0975 - 6450 Volume 1 Number 3 (2009) pp. 223–232

http://www.ripublication.com/ijeer.htm

Robust Preprocessing: Whitening in the Context of Blind Source Separation of Instantaneous Mixture of Audio Signals

[1]Sangeeta Nakhate, [2]R. P. Singh and [3]A. Somkuwar

[1,2,3]Department of Electronics and Communication Engineering, Maulana Azad Naional Institute of Technology, Bhopal-462007 (India)
E-mail: [1]sanmanit@gmail.com, [2]prof_rpsingh@rediffmail.com and [3]asomkuwar@rediffmail.com

Abstract

Prewhitening is often considered a necessary but not sufficient condition for stronger stochastic independence criteria. After prewhitening the task of Blind Source Separation (BSS) become somewhat easier. Robust preprocessing involves spectral whitening, which is done by transforming correlated signals to an uncorrelated flat spectrum signal. The maximum entropy power spectrum estimation has been used as a contrast function for Independent Component Analysis (ICA). To test the separation ability for the instantaneous mixtures of speech and acoustic noise, comparative assessment with and without spectral whitening have been presented using proposed ICA algorithm and other existing algorithms. Proposed model provides SNR 42.9dB and cross correlation 0.98 for mixture of two music signals.

Index Terms: BSS, ICA, Maximum entropy, Spectral whitening

Introduction

The speech enhancement is very important, because of the acoustic environment we are living in, is composed of noise and other atmospheric disturbances which makes it almost impossible to record a speech signal in pure form. In the real recording environment, signals reaching each microphone are not only direct-path signals, but also delayed and attenuated versions of the source signals. The Blind Source Separation (BSS) problem is defined as the procedure of estimating the original signals through the observed signals. The term blind indicates that these methods employ no prior information about the source signals and its mixing procedure. Independent Component Analysis (ICA) is major statistical tools for dealing with the

BSS problem can be directly employ to separate the instantaneous mixture of speech signals, exploiting only the assumption of mutual independence between the signals. Another fundamental restriction or assumption in ICA is that the independent component must be non-Gaussian or at most may have one Gaussian distribution for Independent Component Analysis to be possible [7,8,12]; Many algorithms for speech enhancement e.g. Wiener filtering [16], noise and speech both have been assumed to be Gaussian. However, many real world noise signals such as chair crack, clapping, object dropping, etc. are neither Gaussian nor exactly Laplacian [11], this is because of the joint probability densities of Gaussian random variables are completely symmetric. The whitening is an important pre-processing step in a variety of BSS methods. The idea of the robust whitening is to consider a linear combination of several time-delayed correlation matrices. In fact whitening is only half ICA. In terms of mathematical complexity; the mixing process can be divided into two types: instantaneous mixing and convolutive mixing. This paper deals with instantaneous mixture of speech signals. However, if the observed signals are noisy, it is required to preprocess the data before application of the ICA algorithm. BSS methods based on Fast-ICA and nonlinear PCA employ a preprocessing stage which essentially whitens the data [13,14]. Preprocessing can be done by denoising and spearing of the data [2]. In this paper robust preprocessing involves spectral whitening, which is done by transforming correlated signals to an uncorrelated flat spectrum signal followed by ICA based on the maximum entropy power spectrum estimation. The whitening procedure is based on higher order Blind Identification produces a whitening matrix. The proposed algorithm performs well at low input signal to noise ratio. The performance does not affect severely with increased noise level.

BSS model with Preprocessing

The proposed model is developed for extraction of statistically independent speech signals by maximizing entropy of power spectrum. Figure 1 shows Autoregressive model for speech enhancement with prewhitening. s_1 (m) to s_n (m) are n statistically independent unknown signals at time m. $x_1(m), x_2(m)....x_n(m)$ are instantaneous mixtures of speech signals observed at different n microphone, can be represented as

$$x(m) = As(m) \tag{1}$$

Where A is the matrix of mixing coefficients.

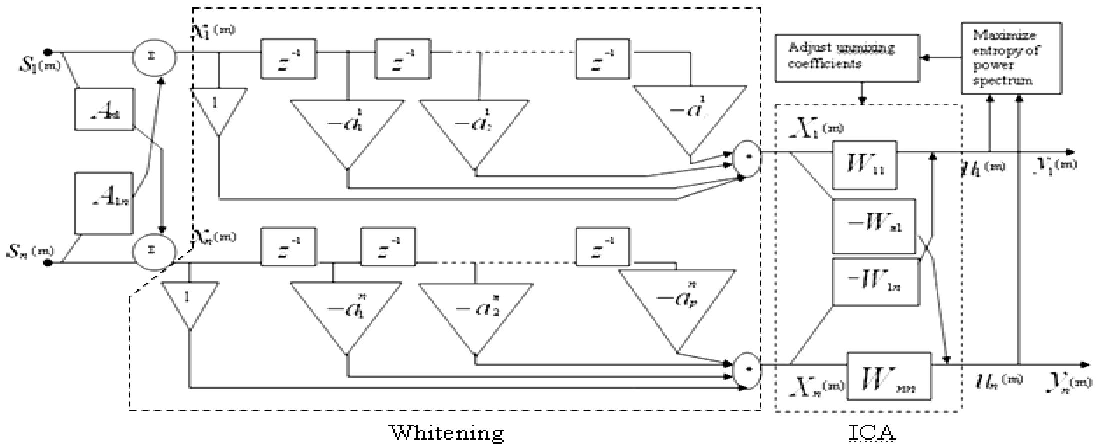


Figure 1: Autoregressive model for speech enhancement with prewhitening.

As the observed signals are noisy; it is required to preprocess the data before application of the ICA algorithm. Robust preprocessing is done by spectral whitening. Mixed signal are whitened to get X_1(m) to X_n(m). ICA based on the maximum entropy power spectrum estimation is applied on whitened signals. The entropy of the power spectrum of the intermidiate unmixed signal $u(m)$ is maximized, corresponding change in weights in the feed forward-ICA network is find out. Cross correlation coefficient of intermediate unmixed signals and test signals are check to conform the received signal is the original signal; for most approximately one cross correlation coefficient, the corresponding unmixed signal is referred as extracted signal $y(m)$ equivalent to the test speech signal.

(a) Spectral Whitening:

The model for spectral whitening transforms correlated signals $x_1(m), x_2(m)....x_n(m)$ to an uncorrelated flat spectrum signals X_1(m) to X_n(m). Uncorrelated signals can regarded as the information content of the sample $x(m)$ can represent

$$X\,(\mathrm{m}) = x\,(\mathrm{m}) - \hat{x}\,(\mathrm{m})$$

$$= x\,(\mathrm{m}) - \sum_{k=1}^{p} a_p x(m-k) \qquad (2)$$

$$= (a^{inv})^T x \qquad (3)$$

Where $\hat{x}$(m) is the prediction of x(m) and a_p are the predictor coefficients.

$(a^{inv})^T = [1, -a_1, \ldots.. -a_p]$

x^T is the combination of past p samples [$x(m)$,……… $x(m-p)$] .

(b) Calculation of Predictor coefficients:

Predictor coefficients are obtained by minimizing the mean Square prediction error as

$$E[X^2(m)] = E\{[x(m) - \sum_{k=1}^{p} a_p x(m-k)]^2\}$$

$$= r_{xx}(0) - 2r_{xx}^T a + a^T R_{xx} a \quad (4)$$

Where $R_{xx} = E[xx^T]$ is the autocorrelation matrix of the input vector.

$x^T = [x(m-1), \ldots\ldots\ldots x(m-k)]$, $r_{xx} = E[x(m)x(m-k)]$ is the autocorrelation vector and $a^T = [a_1, a_2, \ldots\ldots a_p]$ is the prediction coefficient vector. From equation (4) the gradient of the mean square prediction error with respect to the predictor coefficient vector a is given by

$$\frac{\partial}{\partial a} = E[X^2(m)] = -2r_{xx}^T + 2aR_{xx} \quad (5)$$

Where the gradient vector is defined as $(\frac{\partial}{\partial a_1}, \frac{\partial}{\partial a_2}, \ldots\ldots\ldots, \frac{\partial}{\partial a_p})^T$

The least mean square error solution, obtained by setting equation (5) to zero is given by

$$R_{xx} a = r_{xx} \quad (6)$$

From equation (6) the prediction coefficient vector is given by

$$a = R_{xx}^{-1} r_{xx} \quad (7)$$

Equation (7) can be written in the expanded form as

$$\begin{bmatrix} a_1 \\ a_2 \\ a_3 \\ . \\ . \\ a_p \end{bmatrix} = \begin{bmatrix} r_{xx}(0) & r_{xx}(1) & r_{xx}(2) \ldots\ldots\ldots & r_{xx}(p-1) \\ r_{xx}(1) & r_{xx}(0) & r_{xx}(1) \ldots\ldots\ldots & r_{xx}(p-2) \\ r_{xx}(2) & r_{xx}(1) & r_{xx}(0) \ldots\ldots\ldots & r_{xx}(p-3) \\ . & . & . & . \\ . & . & . & . \\ r_{xx}(p-1) & r_{xx}(p-2) & r_{xx}(p-3) \ldots\ldots & r_{xx}(0) \end{bmatrix}^{-1} \begin{bmatrix} r_{xx}(1) \\ r_{xx}(2) \\ r_{xx}(3) \\ . \\ . \\ r_{xx}(p) \end{bmatrix} \quad (8)$$

(c) Maximum entropy spectral estimation

For simplicity number of microphones and the number of sources are considered to be same as n. The measured mixed signals at the n^{th} microphone at m^{th} sampling time may be represented as

$$x_n(m) = s_n(m) + \sum_{q \neq n}^{N} A_{nq} s_q(m) \quad (9)$$

Where n =1,2,…..N.

As per ICA intermediate unmixed signal can be written as

$$u_n(m) = W_{nn} X_n(m) + \sum_{q \neq n}^{N} W_{nq} X_q \quad (10)$$

s_n(m) denotes the n^{th} source signal at m^{th} time and A_{nq} is the mixing coefficient from n^{th} source to q^{th} microphone W_{nm} unmixing coefficient. To avoide further whitening W_{11} to W_{nn} are considered to be one and other weights in the unmixing matrix are considered to be negative because extracted signal u_n(m) must have more information of $X_n(m)$, the power spectrum of unmixed signal is defined as the Fourier transform of autocorrelation sequence.

$$p_{uu}(f) = \sum_{m=-\infty}^{\infty} r_{uu}(m) e^{-j2\pi fm} \tag{11}$$

The randomness or entropy of a unmixed signal is defined as

$$H[p_{uu}(f)] = \int_{-1/2}^{1/2} \ln p_{uu}(f)\,\mathrm{df} \tag{12}$$

To obtain the maximum entropy correlation estimate, differentiate (12) with respect to the unknown values of correlation coefficients and set the derivatives to zero.

The derivative of the logarithm of the power we have

$$\frac{\partial[\ln Puu(f)]}{\partial r_{uu}(m)} = p_{uu}^{-1}(f) e^{-j2\pi fm} \tag{13}$$

Assuming that $p_{uu}^{-1}(f)$ is integrable, it can be associated with an autocorrelation sequence $C(m)$

$$p_{uu}^{-1}(f) = \sum_{m=-\infty}^{\infty} C(m) e^{-j2\Pi fm} \tag{14}$$

Where $C(m)$ is autocorrelation sequence

Maximum entropy power spectrum is given by

$$p_{uu}^{\Lambda ME}(f) = \frac{1}{\sum_{m=-M}^{M} C(m) e^{-j2\Pi fm}} \tag{15}$$

Since the denominator polynomial in equation (15) is symmetric, it follows that for every zero of this polynomial situated at radius r, there is a zero at radius 1/r, Hence this symmetric .

Where $\frac{1}{\sigma^2}$ is a gain term

From equation (16) we can write Maximum entropy power spectrum as polynomial can be factorised and expressed as

$$\sum_{m=-M}^{M} C(m) e^{-j2\pi fm} = \frac{1}{\sigma^2}[U_n(X)] \tag{16}$$

$$p_{uu}^{\Lambda ME}(f) = \frac{\sigma^2}{[U(X)]} \tag{17}$$

$U(X)$ is a polynomial of order k (can be any natural value) defined as

$$U_n(X) = 1 + \sum_{k=1}^{n} W_{nq} X_q^{-q} \tag{18}$$

Equation (17) shows that the maximum entropy power spectrum estimate is the power spectrum of an autoregressive model shown in figure 1 which is obtained by maximising the entropy of the power spectrum with respect to unknown autocorrelation values.

Differentiating equation (18) with respect to W_{nq} for getting change in weight ΔW_{nq}

$$\Delta W_{nq} = \eta \frac{X_q^{-q}}{[u_n(X)]^2} \tag{19}$$

Where η is the praportionality constant.

If power spectral density of test signal matched with the power spectral density of intermediate unmixed signal corresponding intermediate unmixed signal is recorded to find out the reconstructed signal

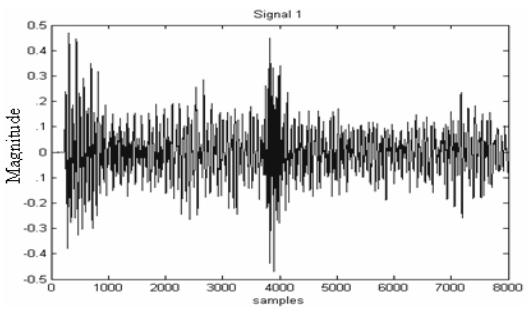
$$y_n(m) = \frac{1}{1 + e^{-u_n(m)}} \tag{20}$$

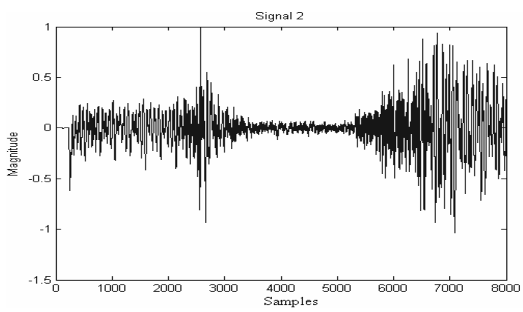
Which is most approximately equal to the original test signal $s_n(m)$. After every iteration SNR and cross correlation between intermediate unmixed signals and test signal is checked to set different parameters in the algorithm so as to verify model for unknown signals

Experimental Results

Two music signals are recorded independently for 10 seconds with sampling rate of 16 KHz these audio signals are taken as test signals, mixing matrix is generated as per the distance between microphones.

a





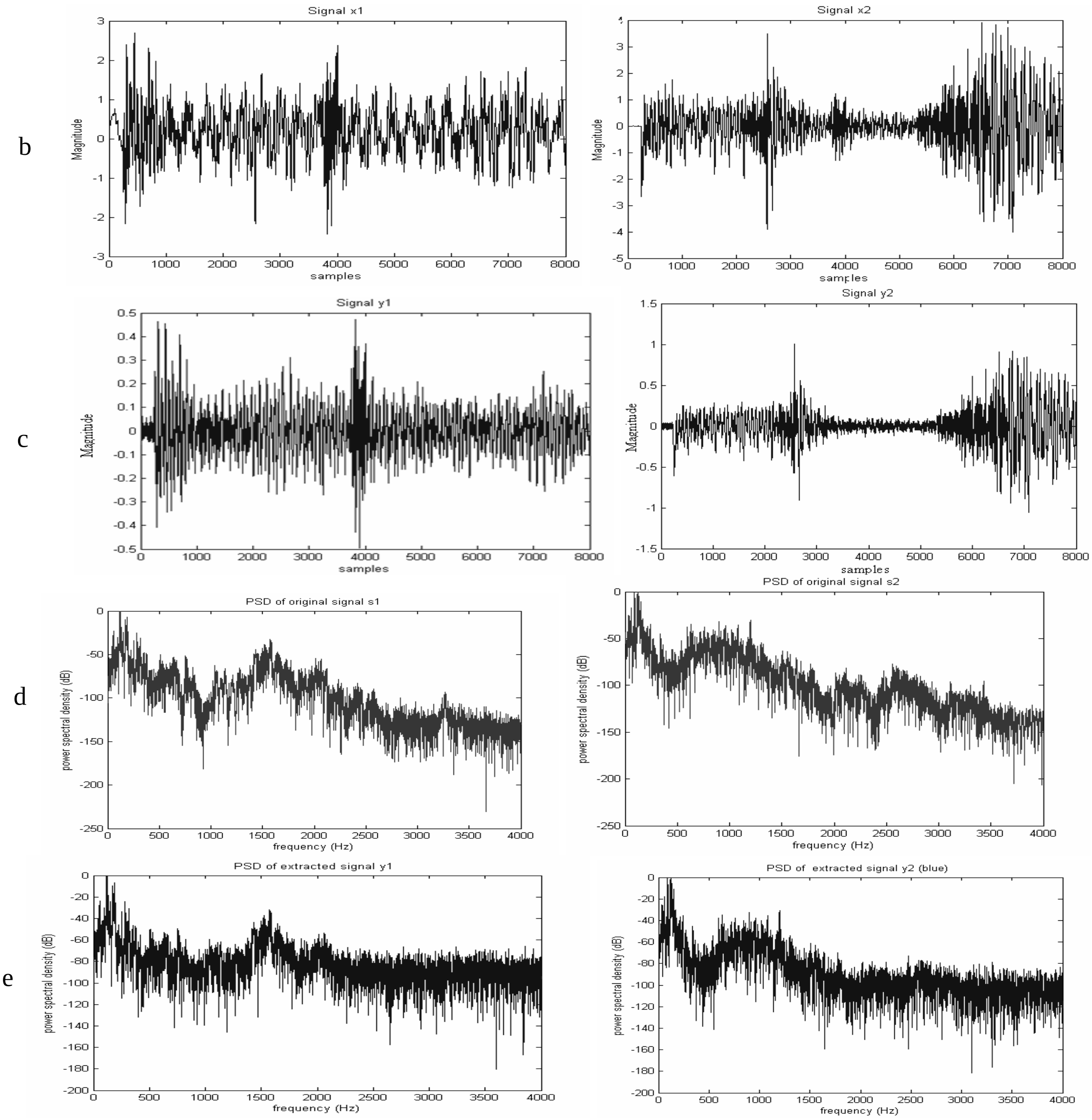


Figure 2: Separation of two mixed music signals.

a- original signals, b- mixed signals, c-extracted music signals, d- Power spectral density (PSD) of original signal and e- PSD of extracted signal

Signal to noise ratio is find out by using following formula

$$\text{SNR (dB)} = 10\log_{10}\left(\frac{\sum_{m}|s|^{2}}{\sum_{m}|y-s|^{2}}\right)$$

Where s is test signal and y is extracted signal. Output SNR obtained corresponding to input SNR is shown in the table I

Table I: SNR obtained while extraction two music signals.

Input SNR(dB)	-9.3	-5.3	0.7	4.8	8.9
Output SNR (dB)	8.7	11.3	23.8	32.7	42.9

From figure 3 it can observe that, the average correlation coefficient improves with increase in the SNR and reduces with a decrease SNR. Under high SNR condition all schemes show perfect separation ability with average correlation coefficient approaching towards 1. But for low SNR, the proposed algorithm with spectral whitening performs the better than Fast-ICA technique with denoising schemes like wavelet and Eigen filter as described in [2]. It can be observe that denoising by spectral whitening yields best results under low SNR condition compared to wavelets and eigenfilter scheme. A robust preprocessing technique helps to improve the performance.

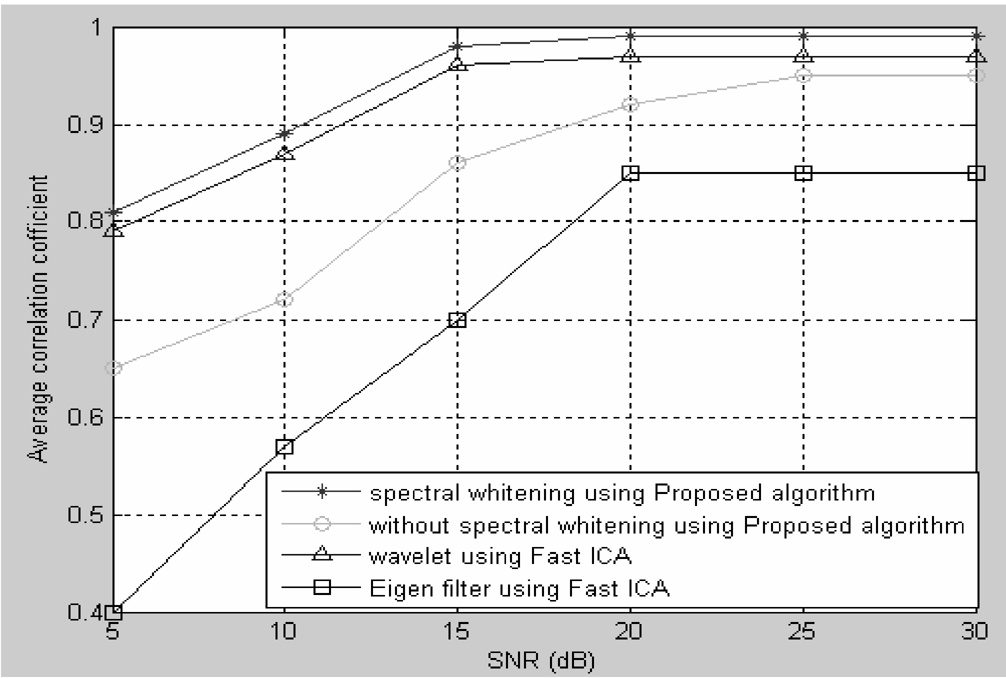


Figure 3: comparison of proposed algorithm with other methods in terms of average correlation coefficient.

Conclusion

In this paper robust preprocessing scheme is suggested for the blind separation of unknown sources, shows significant increase in performance while compared with Fast-ICA without preprocessing [2, 11, and 12]. The spectral whitening scheme is particularly suitable for noisy speech. Results will not affect severely for the increase of number of microphones. The proposed model performs well for SNR up to 42.9dB, with deteriorating performance upon increase in the noise level where a robust preprocessing technique helps to improve the performance.

References

[1] Yu Takahashi, Tomoya Takatani, Keiichi Osako, Hiroshi Saruwatari, and Kiyohiro Shikano, "Blind Spatial Subtraction Array for Speech Enhancement in Noisy Environment" IEEE Transactions on audio, speech, and language processing, vol. 17, no. 4, may 2009,pp 650- 664.

[2] Aurobinda Routray, Niva Das, P.K. Dash "Robust Preprocessing: Denoising and Whitening in the context of Blind Source Separation of Instantaneous Mixtures" IEEE transaction on audio, speech and language processing, vol.14, No 4, 2007,pp377-380.

[3] Jen-Tzung Chien, and Bo-Cheng Chen, "A new Independent Component Analysis for speech recognition and separation" IEEE transaction on audio, speech and language processing, vol. 14, No. 4, July 2006,pp1245-1254.

[4] Jen-Tzung Chien, and Bo-Cheng Chen "A New Independent Component Analysis for Speech Recognition and Separation", IEEE Transactions on audio, speech and language processing,Vol. 14, No. 4, July 2004, pp 1245-1254.

[5] A.Tkacenko, P.P. Vaidyanathan and T.Q Nguyen., "On the Eigenfilter Design Method and its Application: ATutorial", IEEE Trans. on Circuits and Systems II: Analog and Digital Signal Processing, vol.50,no.9,September- 2003.

[6] Aapo Hyvärinen, Juha Karhunen and Erikki Oja, "Independent Component Analysis,"John Wiley & Sons, Incorporation, 2001.

[7] Aapo Hyvärinen and Erkki Oja, "Independent Component Analysis: Algorithms and Applications" Neural Networks Research Centre, Helsinki University of Technology, http://www.cis.hut.fi/projects/ica, 2000.

[8] Belouchrani and A. Cichocki, "Robust whitening procedure in blind source separation context, IEE Electronics letters, 36 (24):2050-2053, 2000.

[9] G.G. Panayotis, T. Panagiotis, K. Chris, "Alpha-stable modeling of noise and robust time delay estimation in the presence of impulsive noise," IEEE Trans. On Multimedia, Vol.1, No.3, 1999 pp 291-301.

[10] Aapo Hyvärinen,"Fast and Roust Fixed-Point Algorithms for Independent Component Analysis", IEEE Transactions on Neural Networks, Volume 10, NO. 3, May 1999.

[11] J. Karhunen, P. Pajunen, and E. Oja., "The nonlinear PCA criterion in blind source separation: Relations with other approaches", *Neurocomputing*, 22:5-20, 1998.

[12] Hyvärinen and E. Oja, "A fast fixed-point algorithm for independent component analysis", Neural Computation, 9(7):1483-1492, 1997.

[13] David L. Donoho., "Denoising via soft thresholding.",IEEE Transaction on information theory , 41:613–627, May 1995.

[14] A.Ephraim, D. Malah, “Speech enhancement using a minimum mean square error short-term spectral amplitude estimator,” IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-32, No. 6,Dec.1984, pp 1109-21.