

Bionic Wavelet Based Denoising Using Source Separation

M. Talbi, A.B. Aicha, L. Salhi, A. Cherif

Mourad Talbi, Lotfi Salhi, Adnene Cherif

Faculty of Sciences of Tunis,
Laboratory of Signal Processing,
University Campus, 2092 El Manar II, Tunis, Tunisia
E-mail: mouradtalbi196@yahoo.fr, lotfi.salhi@laposte.net,
adnane.cher@fst.rnu.tn

Anis Ben Aicha

Université de Carthage, Ecole Supérieure des Communications
Laboratoire de recherche COSIM
Route de Raoued 3.5 Km, Cité El Ghazala, Ariana, 2083,
Tunisie, Tél. : +216 71 857 000 - Fax : +216 71 856 829
E-mail: ben.aicha.anis@gmail.com

Abstract:

We consider the problem of speech denoising using source separation. In this study we have proposed a hybrid technique that consists in applying in the first step, the Bionic Wavelet Transform (BWT) to two different mixtures of the same speech signal with noise. This speech signal is corrupted by a Gaussian white noise with two different values of the Signal to Noise Ratio (SNR) in order to obtain those two mixtures. The second step consists in computing the entropy of each bionic wavelet coefficient and finds the two subbands having the minimal entropy. Those two subbands are used to estimate the separation matrix of the speech signal from noise by using the source separation. Our proposed technique is evaluated by comparing it to the denoising technique based on source separation in time domain.

Keywords: Bionic wavelet transform, Blinde Source Separation, entropy, speech enhancement.

1 Introduction

In signal processing, the source separation constitutes an attractive problem. Its goal is to extract from many signals mixture, the meaningful signals. This is performed with minimum a priori information on the mixture process. In the case of instantaneous mixture, many approaches employing the ICA algorithm can solve the problem of the source separation. One of those approaches permits to estimate the unmixing matrix by minimizing the mutual information between the separated sources [1, 2]. Others exploit the non-Gaussianity of the source signals and perform separation by maximizing this non-Gaussianity [2]. For example, a technique using a subband decomposing in combination with ICA, has been developed by Tanaka et al [3]. Kisilev et al [4] have employed geometric algorithms for separating mixed signals. Rachid Moussaoui et al [5] have proposed an algorithm using the idea of applying a preprocessing in the transformed domain but the separation is performed in the time domain. In this paper we have used the source separation with Bionic Wavelet Transform (BWT) for enhancement of speech signal corrupted by white noise. The source separation is performed with ICA and instead of using the wavelet packet transform as used in the technique proposed by Rachid et al [5], we have used in this work the BWT.

2 Restrictions of ICA

The ICA standard formulation needs at least as many sensors as sources. Therefore, we suppose in this paper that the source number is equals to the sensor number. In the instantaneous mixture case, the sources are not directly observed but as a linear combination such as:

$$x_i(t) = \sum_{j=1}^{j=N} a_{ij}S_j(t), \quad (1)$$

where x are the observed signals, s are the source signals and $A = [a_{ij}]$ is unknown full rank mixing matrix.

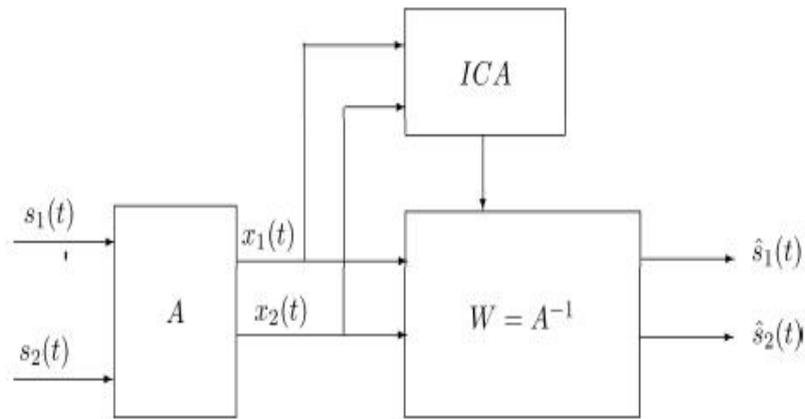


Figure 1: ICA Principle.

Figure 1 shows the ICA principle. The ICA aim is practically to find the inverse matrix of A , which is the unmixing matrix $W = A^{-1}$. To make an estimation of W , certain assumptions have to be made and some restrictions have to be imposed [2]: we assume that the individual components $s_i(t)$ are statistically independent over the observation time and the individual components must have non Gaussian distributions. In comparison to previous work, the novelty of the approach of Rachid Moussaoui et al [5] resides in the preprocessing implementation before the source separation process in to:

- Relax the previous restrictions by increasing the non-Gaussianity which is a pre-requirement for ICA.
- Initiate a preliminary separation by decreasing the mutual information between the resultant signals from the preprocessing.

The preprocessing transforms the observed signals to find an adequate representation where the signals distributions are non-Gaussian. For this reason, the wavelet transform is used in order to emphasize the non Gaussian nature of the observed signals. Once we have found the inverse matrix W with the wavelet packets based ICA then, the separation is performed in the time domain [5]. Figure 2 illustrated an overview of the system proposed by Rachid Moussaoui et al [5].

In this paper we have chosen $s_2(t)$ to be a white noise that corrupted the clean speech signal $s_1(t)$ with two different values of the Signal to Noise Ratio (SNR).

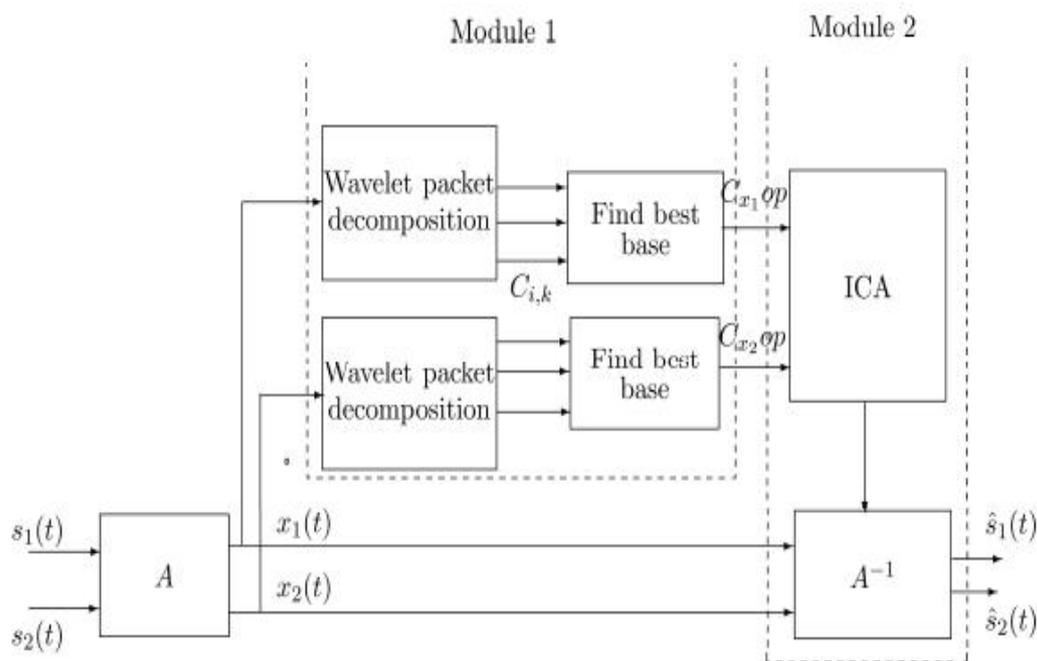


Figure 2: Overview of the source separation system proposed by Rachid Moussaoui et al [5].

3 The proposed technique

The proposed speech enhancement system, illustrated in Figure 4, is inspired from that of Rachid Moussaoui et al [5]. The latter is conceived for multi-channel source separation and is based on wavelet based independent component analysis. It comprises two modules shown in dotted boxes in Figure 4. The first module (pre-processing) extracts appropriate signals from the observed signals in order to facilitate the separation of the speech and noise signals. For this, the observed signals are projected on suitable bases, more specifically on bionic wavelet bases. The second module (speech and noise separation) performs the source separation using standard ICA [1]. The input of this module is the extracted signals from module 1 and the observed signals. Its output is the cleaned or the enhanced speech signal $\hat{s}_1(t)$. Figure 4 illustrated an overview of the proposed speech enhancement system which is summarized by the following steps:

- i Decompose the observed signals (the two noisy speech signals) into bionic wavelet subbands by applying the BWT.
- ii Compute the entropy value of each subband and select the two subbands having the minimum entropy.
- iii Use those two subbands as the inputs of the ICA system in order to estimate the separation matrix, A^{-1} .
- iv Estimate the enhanced speech signal $\hat{s}_1(t)$ by applying A^{-1} to the temporal mixtures $x_1(t)$ and $x_2(t)$.

The used entropy in this work, is the Shannon entropy which is defined for each subband $w_{.j}$, $1 \leq j \leq 30$ as:

$$H(j) = - \sum p_i \log(p_i). \quad (2)$$

Note that in the expression $w_{\cdot,j}$, $1 \leq j \leq 30$, \cdot is replaced by 1 if we apply the BWT to $x_1(t)$ and is replaced by 2 if we apply the BWT to $x_2(t)$.

The probability p_i is expressed as: $p_i = \frac{w_{\cdot,j(i)}^2}{\|W\|^2}$, $1 \leq i \leq N$ and N is the number of samples in the subband $w_{\cdot,j}$ and W is obtained by concatenating all the subbands $w_{\cdot,j}$, $1 \leq j \leq 30$.

4 The bionic wavelet transform

J. Yao and Y. T. Zhang have proposed the bionic wavelet transform (BWT) as a new time-frequency technique by referring to the perceptual model [6]. The term "bionic" means that the BWT is guided by an active biological mechanism [7]. Moreover, the BWT decomposition is both perceptually scaled and adaptive [8]. The initial perceptual aspect of the transform comes from the logarithmic spacing of the baseline scale variables which are designed to match basilar membrane spacing [8]. Then, two adaptation factors control the time-support employed at each scale, based on a non-linear perceptual model of the auditory system [8]. The basis of this transform is the Giguere -Woodland non-linear transmission line model of the auditory system [9, 10], an active-feedback electro-acoustic model incorporating the auditory canal, middle ear, and cochlea [8]. The model yields estimates of the time-varying acoustic compliance and resistance along the displaced basilar membrane, as a physiological acoustic mass function, cochlear frequency-position mapping, and feedback factors representing the active mechanisms of outer hair cells. The net result can be seen as a technique for the estimation of the time-varying quality factor Q_{eq} of the cochlear filter banks as the input sound waveform function [8]. The references [6–9] give the complete details on the elements of this model. The BWT adaptive nature is ensured by a time-varying linear factor $T(a, \tau)$ which represents the scaling of the cochlear filter bank quality factor Q_{eq} at each scale over time [8]. For each scale and time, the adaptation factor $T(a, \tau)$ of BWT is computed by using the update equation [8]:

$$T(a, \tau + \Delta\tau) = \frac{1}{\left[1 - G_1 \frac{C_s}{C_s + |X_{BWT}(a, \tau)|}\right] \left[1 + G_2 \left|\frac{\partial}{\partial t} X_{BWT}(a, \tau)\right|\right]} \quad (3)$$

where C_s is a constant (typically $C_s = 0.8$) that represents non linear saturation effects in the cochlear model [6, 8].

The quantities G_1 and G_2 are respectively the active gain factor, which represents the outer hair cell active resistance function, and the active gain factor representing the time-varying compliance of the basilar membrane [8]. Practically speaking, the partial derivative in equation (3) can be approximated by using the first difference of the previous points of the BWT at that scale [8]. $X_{BWT}(a, \tau)$ represents the bionic wavelet transform (BWT) of the signal $x(t)$ and it is given by:

$$X_{BWT}(a, \tau) = \frac{1}{T(a, \tau)\sqrt{a}} \int x(t) \cdot \tilde{\varphi}^* \left[\frac{t - \tau}{a \cdot T(a, \tau)} \right] \cdot e^{-j\omega_0 \left(\frac{t - \tau}{a} \right)} dt, \quad (4)$$

where a denotes the parameter of scale, τ is the shifting parameter in time and $\tilde{\varphi}$ is the mother wavelet envelop given by [7]:

$$\varphi(t) = \frac{1}{T(a, \tau)\sqrt{a}} \tilde{\varphi} \left[\frac{t}{T(a, \tau)} \right] \cdot e^{j\omega_0 t} \quad (5)$$

where ω_0 is the base fundamental frequency of the unscaled mother wavelet.

In practice ω_0 is equals to 15165.4 for the human auditory system [6]. The discretization of the scale a is achieved by employing a pre-determined logarithmic spacing across the desired frequency range, so that at each scale the center frequency is expressed by [8]:

$$\omega_m = \frac{\omega_0}{(1.1623)^m}, m = 0, 1, 2, \dots \quad (6)$$

Based on Yao and Zhang's original work for cochlear implant coding [9], coefficients at 22 scales, $m = 7, \dots, 28$, are calculated employing numerical integration of the continuous wavelet transform. These 22 scales correspond to center frequencies logarithmically spaced from 225 Hz to 5300 Hz. (Although the scales used here match those from Yao and Zhang's original work, empirical variation of the number of scales and frequency placement showed minimal effect on the overall enhancement results). For this implementation, we have used coefficients at 30 scales. In the formula (4), the role of first factor $T(a, \tau)$ multiplying \sqrt{a} is to ensure that the energy remains the same for each mother wavelet. The role of second factor $T(a, \tau)$ is to adjust the envelop $\tilde{\varphi}(t)$ without adjusting the central frequency of $\varphi(t)$ [7]. Thus, the main difference between (BWT) and the continuous wavelet transform (CWT) is based on the fact that the time-frequency resolution achieved by (BWT) can be adjusted in an adaptive manner not only by frequency variation of the signal but also by instantaneous amplitudes of this signal. It is the mother wavelet which makes the continuous wavelet transform adaptive, while the adaptive characteristic of the BWT comes from the mechanism of active control in the human auditory model. which adjusts the mother wavelet associated to (BWT) according to the analyzed signal. Basically, the idea of the (BWT) is inspired from the fact that we need to make the mother wavelet envelop variable in time according to the signal characteristics.

The employed mother wavelet $\varphi(t)$ in [7] is the Morlet wavelet and its envelop $\tilde{\varphi}(t)$ is given by [8]:

$$\tilde{\varphi}(t) = e^{\left[-\left(\frac{t}{T_0}\right)^2\right]} \quad (7)$$

where T_0 denotes the initial time-support.

Figure 3 illustrated the real and the imaginary parts of the complex Morlet mother wavelet.

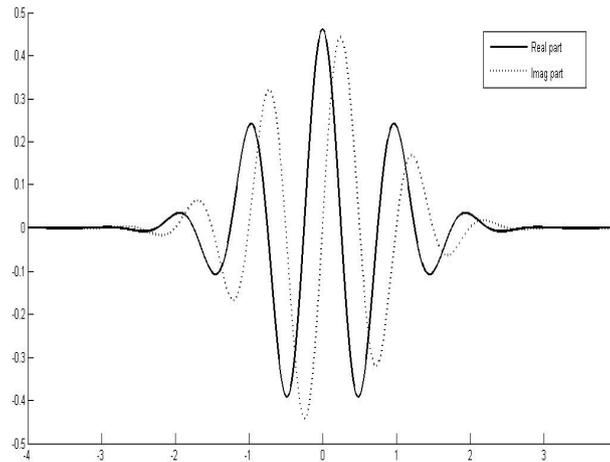


Figure 3: The Morlet wavelet.

It can be shown [7, 9] that the obtained BWT coefficients, $X_{BWT}(a, \tau)$ are derived by using the following formula [8]:

$$X_{BWT}(a, \tau) = K(a, \tau)X_{WT}(a, \tau), \quad (8)$$

where $K(a, \tau)$ is given by:

$$K(a, \tau) = \frac{\sqrt{\pi}}{C} \frac{T_0}{\sqrt{1 + T^2(a, \tau)}} \quad (9)$$

where C represents a normalizing constant calculated from the squared mother wavelet integral.

This representation yields to an effective computational technique for calculating in direct manner, the BWT coefficients from those of the wavelet transform. This is performed without using the BWT definition given by equation (4). There are some key differences between the discretized CWT employing the Morlet wavelet used for the BWT and a filterbank based WPT using an orthonormal wavelet. One of them is that the WPT provides a perfect reconstruction, while the discretized CWT is an approximation whose exactness depends on the number and placement of frequency bands selected [8].

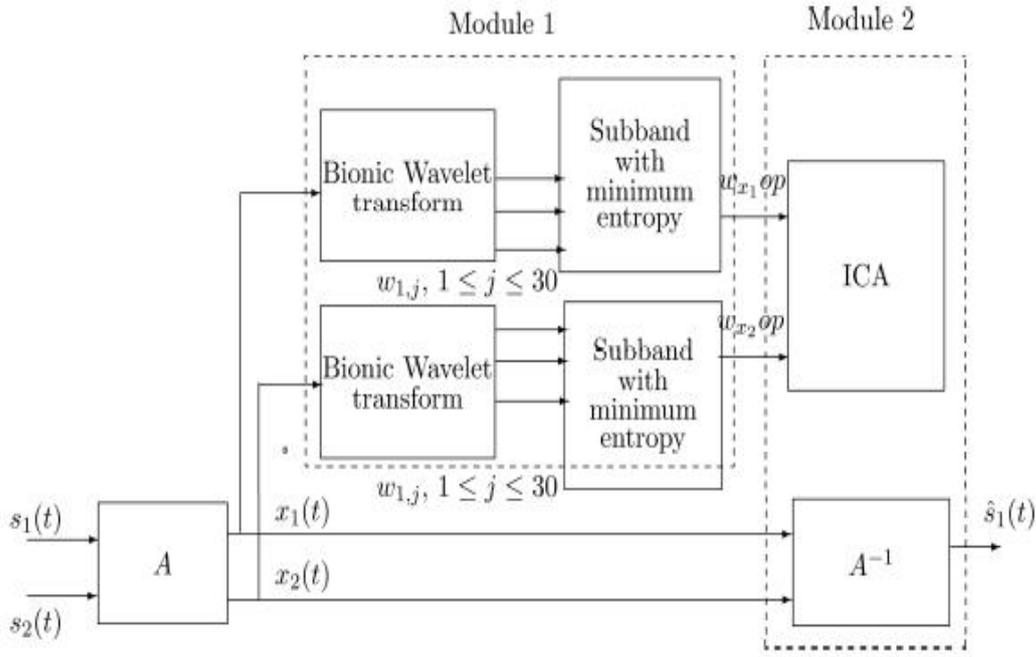


Figure 4: Overview of the proposed system.

5 Criterion of evaluation

For evaluating our proposed technique, we have compared it to the temporal technique based on runica. The evaluation is based on SNR, SSSNR, ISD and PESQ computation. These parameters are defined as follow:

- Signal-to-noise ratio

The signal-to-noise ratio (SNR) of the enhanced speech signal is defined by:

$$SNR_{dB} = 10 \log_{10} \left[\frac{\sum_{n=0}^{N-1} x(n)^2}{\sum_{n=0}^{N-1} (x(n) - \hat{x}(n))^2} \right], \quad (10)$$

where $x(n)$ and $\hat{x}(n)$ represent respectively the original and the enhanced speech signals, and N is the samples number per signal.

- Segmental signal to noise ratio

The segmental signal-to-noise ratio (segSNR) is calculated by averaging the frame based SNRs over the signal:

$$segSNR_{dB} = \frac{1}{M} \sum_{m=0}^{M-1} 10 \log_{10} \left[\frac{\sum_{n=N_m}^{N_m+N-1} x(n)^2}{\sum_{n=N_m}^{N_m+N-1} (x(n) - \hat{x}(n))^2} \right], \quad (11)$$

where M designates the number of frames, N is the size of frame, and N_m is the beginning of the m -th frame. As the SNR can become negative and very small during silence periods, the segSNR values are limited to the range of [-10dB, 35dB] as per [10].

- Itakura-Saito distance

The distance of Itakura-Saito (ISd) measures the spectrum changes and can be computed employing the coefficients of linear prediction (LPC) and this according to the following equation:

$$ISD(a, b) = \frac{(a - b)^T R(a, b)}{a^T R a}, \quad (12)$$

where a represents the LPC vector of the original speech signal $x(n)$. R is the matrix of autocorrelation and b is the LPC coefficients vector of the enhanced speech signal $\hat{x}(n)$. In this work, a 10^{th} order LPC based measure is employed.

- Perceptual evaluation of speech quality

The perceptual evaluation of speech quality (PESQ) algorithm [11, 12] is an objective quality measure, that is approved as the ITU-T recommendation P.862. It is a tool of objective measurement conceived to predict the results of a subjective Mean Opinion Score (MOS) test. It was proved [13, 14] that the PESQ is more reliable and correlated better with MOS than the traditional objective speech measures.

6 Results and discussion

From Table 1 to Table 8, we report the obtained results from the application of our proposed speech enhancement technique and the temporal technique based on runica on eight noisy sentences taken from the Timit database.

Table 1: Sentence 1.

Parameters	Proposed method	Temporal method
SNRi of the first mixture	0.7672	0.7672
SNRi of the second mixture	-3.4638	-3.4638
SNRf(dB)	69.5528	58.7147
SSNRi of the first mixture	-3.5715	-3.5715
SSNRi of the second mixture	-6.3509	-6.3509
SSNRf(dB)	34.8736	34.1366
PESQi of the first mixture	1.3110	1.3110
PESQi of the second mixture	1.0983	1.0983
PESQf	4.4989	4.4936
ISdi of the first mixture	2.4029	2.4029
ISdi of the second mixture	3.9475	3.9475
ISdf	$4.181 \cdot 10^{-10}$	$4.948 \cdot 10^{-8}$

These results show clearly that our proposed technique outperforms the temporal technique of source separation using standard ICA [1].

Fig. 5 and Fig. 6 illustrate an example of speech enhancement using our proposed technique.

Table 2: Sentence 2.

Parameters	Proposed method	Temporal method
SNRi of the first mixture	-2.1038	-2.1038
SNRi of the second mixture	-6.6325	-6.6325
SNRf(dB)	66.5296	48.9184
SSNRi of the first mixture	-5.8749	-5.8749
SSNRi of the second mixture	-7.8815	-7.8815
SSNRf(dB)	34.2126	30.7198
PESQi of the first mixture	1.4577	1.4577
PESQi of the second mixture	1.2445	1.2445
PESQf	4.4981	4.4535
ISdi of the first mixture	2.8500	2.8500
ISdi of the second mixture	4.6523	4.6523
ISdf	$1.4043 \cdot 10^{-9}$	$3.1847 \cdot 10^{-6}$

Table 3: Sentence 3.

Parameters	Proposed method	Temporal method
SNRi of the first mixture	4.2400	4.2400
SNRi of the second mixture	-0.2606	-0.2606
SNRf(dB)	77.4719	49.5281
SSNRi of the first mixture	-1.4316	-1.4316
SSNRi of the second mixture	-4.6465	-4.6465
SSNRf(dB)	34.7279	31.6486
PESQi of the first mixture	1.9873	1.9873
PESQi of the second mixture	1.7460	1.7460
PESQf	4.4999	4.4621
ISdi of the first mixture	1.8622	1.8622
ISdi of the second mixture	3.1204	3.1204
ISdf	$2.5837 \cdot 10^{-10}$	$8.2210 \cdot 10^{-6}$

7 Conclusion

In this paper, we have proposed a new speech enhancement technique that consists in applying in the first step, the Bionic Wavelet Transform (BWT) to two different mixtures of the same speech signal with gaussian white noise with two different values of Signal to Noise Ratio (SNR). The second step consists in computing the entropy of each bionic wavelet coefficient and finds the two subbands having the minimal entropy. Those two subbands are used to estimate the separation matrix of the speech signal from noise by employing the source separation. The obtained results from the SNR, SSNR, ISd and PESQ computation, show clearly that the proposed speech enhancement technique outperforms the temporal technique of source separation using standard ICA.

Table 4: Sentence 4.

Parameters	Proposed method	Temporal method
SNRi of the first mixture	1.6366	1.6366
SNRi of the second mixture	-2.7143	-2.7143
SNRf(dB)	61.8042	55.4325
SSNRi of the first mixture	-4.3027	-4.3027
SSNRi of the second mixture	-6.3345	-6.3345
SSNRf(dB)	31.4223	29.4299
PESQi of the first mixture	1.5414	1.5414
PESQi of the second mixture	1.2476	1.2476
PESQf	4.4816	4.4506
ISdi of the first mixture	2.5180	2.5180
ISdi of the second mixture	4.0632	4.0632
ISdf	$7.4591 \cdot 10^{-8}$	$8.9142 \cdot 10^{-7}$

Table 5: Sentence 5.

Parameters	Proposed method	Temporal method
SNRi of the first mixture	-0.7340	-0.7340
SNRi of the second mixture	-5.7034	-5.7034
SNRf(dB)	59.5033	57.0129
SSNRi of the first mixture	-5.6047	-5.6047
SSNRi of the second mixture	-7.6901	-7.6901
SSNRf(dB)	31.4463	30.8201
PESQi of the first mixture	1.3907	1.3907
PESQi of the second mixture	1.1391	1.1391
PESQf	4.4814	4.4748
ISdi of the first mixture	2.6384	2.6384
ISdi of the second mixture	4.2292	4.2292
ISdf	$2.4528 \cdot 10^{-8}$	$1.0487 \cdot 10^{-7}$

Table 6: Sentence 6.

Parameters	Proposed method	Temporal method
SNRi of the first mixture	2.5544	2.5544
SNRi of the second mixture	-2.1241	-2.1241
SNRf(dB)	62.4473	62.2521
SSNRi of the first mixture	-3.3643	-3.3643
SSNRi of the second mixture	-5.8605	-5.8605
SSNRf(dB)	31.4816	31.4398
PESQi of the first mixture	1.3805	1.3805
PESQi of the second mixture	1.0564	1.0564
PESQf	4.4929	4.4926
ISdi of the first mixture	2.6047	2.6047
ISdi of the second mixture	4.1036	4.1036
ISdf	$9.3330 \cdot 10^{-8}$	$1.0324 \cdot 10^{-7}$

Table 7: Sentence 7.

Parameters	Proposed method	Temporal method
SNRi of the first mixture	-0.0411	-0.0411
SNRi of the second mixture	-4.9428	-4.9428
SNRf(dB)	69.0167	67.9430
SSNRi of the first mixture	-5.7362	-5.7362
SSNRi of the second mixture	-7.8201	-7.8201
SSNRf(dB)	33.9383	33.8069
PESQi of the first mixture	1.6015	1.6015
PESQi of the second mixture	1.4326	1.4326
PESQf	4.4978	4.4974
ISdi of the first mixture	2.5867	2.5867
ISdi of the second mixture	4.1275	4.1275
ISdf	$4.6205 \cdot 10^{-10}$	$6.7472 \cdot 10^{-10}$

Table 8: Sentence 8.

Parameters	Proposed method	Temporal method
SNRi of the first mixture	-1.7018	-1.7018
SNRi of the second mixture	-6.5995	-6.5995
SNRf(dB)	68.1109	52.2155
SSNRi of the first mixture	-5.8031	-5.8031
SSNRi of the second mixture	-7.8610	-7.8610
SSNRf(dB)	33.2544	29.8972
PESQi of the first mixture	1.3814	1.3814
PESQi of the second mixture	1.1889	1.1889
PESQf	4.4968	4.4504
ISdi of the first mixture	3.3273	3.3273
ISdi of the second mixture	5.1147	5.1147
ISdf	$3.1514 \cdot 10^{-9}$	$5.4382 \cdot 10^{-6}$

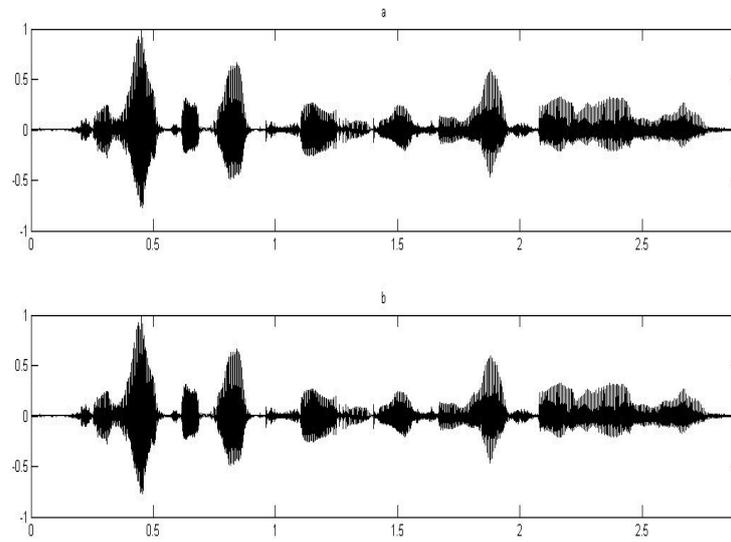


Figure 5: (a) clean speech signal, (b) enhanced speech signal.

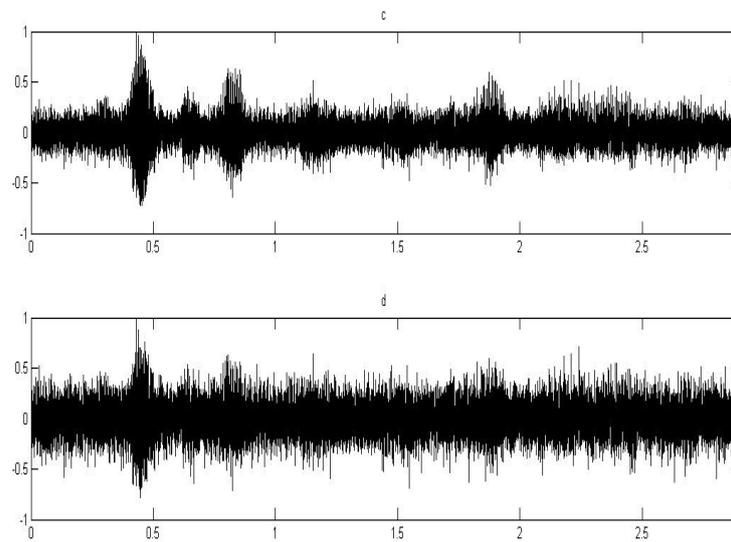


Figure 6: (c) first mixture, (d) second mixture.

Bibliography

- [1] A. J. Bell, T. J. Sejnowski, An information maximization approach to blind separation and blind deconvolution, *Neural Computation*, Vol.7, pp.1004-1034, 1995.
- [2] A Hyvarinen, J. Karhunen, E. Oja, *Independent component analysis*, Wiley and Sons, 2001.
- [3] T. Tanaka, A. Cichocki, Subband decomposition independent component analysis and new performance criteria, *ICASSP*, pp.541-544, 2004.
- [4] P. Kisilev, M. Zibulevsky, Blind source separation using multinode sparse representation, *ICIP*, 2001.
- [5] R. Moussaoui, J. Rouat, R. Lefebvre, Wavelet Based Independent Component Analysis for Multi-Channel Source Separation, *ICASSP*, pp.645-648, 2006.
- [6] J. Yao, Y. T. Zhang, Bionic wavelet transform: a new timefrequency method based on an auditory model, *IEEE Trans. on Biomedical Engineering* Vol.48, No.8, pp.856-863, 2001.
- [7] Xiaolong Yuan, B.S.E.E. A THESIS, ħ Auditory Model-based Bionic Wavelet Transform for speech enhancement. Electrical and computer engineering.
- [8] M. T. Johnsona, X. Yuanb, Y. Rena, Speech signal enhancement through adaptive wavelet thresholding. *in conference Elsevier*, pp.123-133, 2007.
- [9] J. Yao, Y. T. Zhang, The application of bionic wavelet transform to speech signal processing in cochlear implants using neural network simulations, *IEEE Trans. Biomed. Eng.*, Vol.49, No.11, pp. 1299-1309, 2002.
- [10] B. Chen, P. C. Loizou, A Laplacian-based MMSE estimator for speech enhancement, *Speech Communication*, Vol.49, No.2, pp.134-143, 2007.
- [11] ITU-T P.862. Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs, *ITU Recommendation P.862*, 2001.
- [12] A. W. Rix, J. G. Beerends, M. P. Hollier, A. P. Hekstra, Perceptual evaluation of speech quality (pesq) - a new method for speech quality assessment of telephone networks and codecs, *ICASSP*, pp.749-752, 2001.
- [13] Y. Hu, P. C. Loizou, Evaluation of objective measures for speech enhancement, *IEEE Trans. Speech, Audio Processing*, Vol.16, No.1, pp.229-238, 2008.
- [14] E. Zavarehei, S. Vaseghi, Q. Yan. Inter- frame modeling of DFT trajectories of speech and noise for speech enhancement using Kalman filters, *Speech Communication*, Vol.48, No.11, pp.1545-1555, 2006.