IJICT International Journal of Information & Communication Technology

ITRC

Volume 1- Number 1- March 2009 (pp. 21-27)

Noise Reduction from Speech Signal based on Wavelet Transform and Kullback-Leibler Divergence

Shima Tabibian
Computer Engineering Department
Audio and Speech Processing Lab
Tehran, Iran
Shimatabibian@iust.ac.ir

Ahmad Akbari
Computer Engineering Department
Audio and Speech Processing Lab
Tehran, Iran
Akbari@iust.ac.ir

Abstract— **A new method for speech enhancement based on Kullback-Leibler (K-L) divergence has been presented in this paper. First, the algorithm performs wavelet-packet transform to noisy speech and decomposes it into sub-bands; then we apply a threshold on coefficients in each sub-band to obtain enhanced speech. To determine the threshold, first the distributions of noisy speech, noise and clean speech coefficients are calculated; then a symmetric K-L divergence between the noisy speech and noise distributions is calculated. Finally a speech/noise decision is made based on the calculated distance. We conducted some tests using TIMIT database in order to assess the performance of the proposed method and to compare it to previous speech enhancement methods. The algorithm is evaluated using the Perceptual Evaluation of Speech Quality measure (PESQ) and the output SNR. We obtain an improvement of up to 2.2dB on SNR and 1.2 on PESQ for the proposed method in comparison to the results of the previous wavelet based methods.**

Keywords- Wavelet packet decomposition, K-L divergence, noise reduction, hard threshold.

I. INTRODUCTION

Speech enhancement methods have been widely used in noisy environments for many different applications such as mobile phone systems, speech coding systems and hearing aids devices. Classical speech enhancement methods can be divided into two main groups: FFT-based filtering and wavelet thresholding. FFT-based methods modify the frequency spectrum of noisy signal to reduce residual noise and speech artifacts in enhanced signal [1-6]. In wavelet thresholding methods, we reduce noise by thresholding the wavelet coefficients. The estimated threshold should define a boundary between the wavelet coefficients of noise and speech signal. Unfortunately, it is not always possible to separate the wavelet coefficients of speech and noise using a simple thresholding. For noisy speech, the energy of an unvoiced segment is comparable to the energy of noise. In such case, if we apply a unique threshold to all wavelet coefficients, we not only suppress additional noise, but also we probably remove unvoiced speech segments. As a result, the quality of the enhanced speech is degraded.

The main issue in wavelet thresholding is estimating an appropriate threshold value T. For I.I.D. zero-mean, normally distributed noise with variance $\hat{\sigma}_n^{\,2}$, Donoho and Johnstone [7] proposed the universal threshold. This threshold does not depend on the input data, but on the noise variance. It works well for the uncorrelated noise; but it oversmooths the noisy speech in many cases. The value of the threshold can be determined using the following formula [7]:

$$T = \hat{\sigma}_n^2 \sqrt{2\log(N\log_2 N)} \tag{1}$$

The study in [7] showed that when noise dominates the observed data, the universal threshold method performs well and when the underlying signal dominates the observed data, the SURE (Stein's unbiased risk estimator) method [8, 9] performs better than the universal method. This observation led to the heuristic SURE method, which selects either the universal threshold or the SURE threshold according to a test that finds out which of the noise or the underlying signal dominates the observed data [9]. Mini-max threshold [10] is another method for estimating the appropriate threshold value and is obtained very similar to the one obtained in Donoho's method.

After the threshold estimation, a thresholding function should be determined. Using this function noisy wavelet coefficients are compared with the threshold value in order to determine which part of them should be modified. Common thresholding functions are hard thresholding function, soft thresholding function and semi-soft thresholding function [11-17].

The thresholding methods use this idea that in the range of $-T$ and T, the noisy speech wavelet coefficients are similar to the noise wavelet coefficients and outside of this rang, the wavelet coefficients of the noisy speech are similar to that of the clean speech. So it is expected that in the range of –T and T [7] the probability distribution of the noisy speech coefficients is expected to be nearly similar to that of the noise coefficients. Furthermore, the probability distribution of the noisy speech coefficients is expected to be similar to that of the clean speech coefficients outside of this range. It is expected to obtain more exact threshold values using a suitable similarity measure between probability distributions of noisy speech and noise wavelet coefficients.

We propose an adaptive algorithm in order to choose an appropriate threshold value, based on the K-L divergence. First, our algorithm performs a wavelet packet transform on the noisy speech and decomposes it into sub-bands; then the distributions of noisy speech, estimated noise and estimated clean speech sub-bands are calculated; finally a symmetric K-L divergence is used for the estimation of the distance between speech and noise distributions. The decision rule is formulated in terms of the sub-band K-L divergence that is compared to a SNR (Signal to Noise Ratio)-adaptable threshold. The main idea of the algorithm is to remove the background noise almost completely, specially in the silent parts of the noisy speech, without any more distortion in the enhanced signal, compared to the classic speech enhancement methods.

The organization of this paper is as follows. In Section 2, the principles of the wavelet based speech enhancement approach are studied. In Section 3, some of the speech processing approaches are reviewed that exploit some ideas of the information theory. The proposed algorithm is presented in Section 4. In Section 5, the performance of the proposed approach is evaluated. The paper is concluded in Section 6.

II. WAVELET BASED SPEECH ENHANCEMENT METHOD

As Fig. 1 shows, a wavelet based speech enhancement system, is composed of five steps.

- Windowing and overlapping
- Wavelet packet decomposition
- Thresholding or filtering
- Wavelet packet reconstruction
- Adding and overlapping

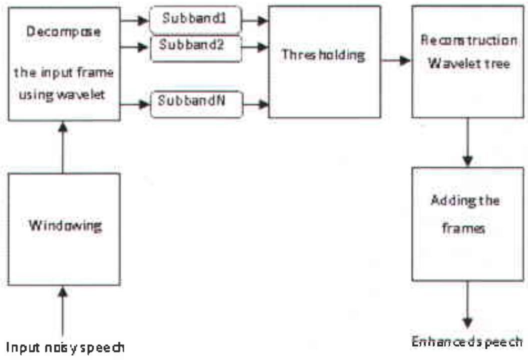


Figure 1. A wavelet based speech enhancement system

Some of the thresholding methods are listed in the following:

- Hard threshold
- Soft threshold
- Semi soft threshold
- Semi hard threshold
- Quintile based threshold

Hard threshold will be discussed in the following [14]. The standard form of hard thresholding has an input-output characteristic that is drawn by solid line in Fig. 2.

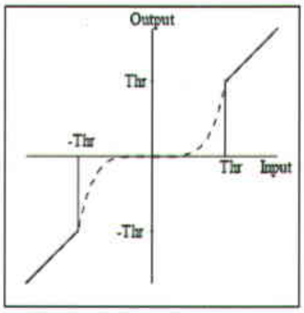


Figure 2. Input-output characteristics for hard thresholding

The hard threshold is defined using (2):

$$Thr_H(w_S^D, T) = \hat{w}_S^D = \begin{cases} 0 & |w_S^D| < T \\ w_S^D & |w_S^D| \geq T \end{cases} \quad (2)$$

Where T is the threshold value that is set due to Donoho's method, SURE method, heuristic SURE method or mini-max method and $\hat{w}_s^D$ stands for the wavelet coefficients.

In order to use hard threshold for noise reduction, we have to set a suitable tradeoff between the remained background noise and the distortion in the enhanced speech signal. In the present work, we use K-L divergence to determine suitable threshold in order to reduce background noise without any more distortion in the enhanced speech signal.

III. Some Applications of Information Theory Concepts in Speech Processing

Speech processing area contains a wide range of applications on the speech signals such as speech enhancement, speech recognition, speaker recognition, speaker validation, spoken dialog systems, text to speech systems, blind source separation, voice activity detection and voiced/ unvoiced decoder. Although information theory concepts [18, 19] may be used in almost all mentioned fields but they are mainly used in the last three applications.

In the area of blind source separation, the mutual information [18] has been widely used .The blind source separation problem has been discussed in many references [20-24]. Suppose there are a (known) number of speakers talking at the same time. Using microphones placed at different locations that are able to record mixtures of the signals coming from all the speakers. The objective in the blind source separation problem is to recover the original speech patterns from the linear mixtures. Here, "blind" implies that neither the mixing coefficients nor the probability distributions of the original sources are known. The key feature that the problem solver has to explore is the statistical independence in the components of the original vector. It means that the problem solver has to impose as much statistical independence as possible on the individual components of the output vector. A good measure for the degree of statistical independence is to choose the mutual information I(Yi; Yj) between the random variables Yi and Yj (constituting any two components of the output vector Y). In the ideal case, when I(Yi; Yj) = 0, the components Yi and Yj are statistically independent. This would therefore suggest minimizing the mutual information between every pair of components of the output vector Y.

In the context of voiced/unvoiced decoding, the entropy measure [18] has been used for achieving a good performance in the wavelet based speech enhancement systems [25]. In order to remove the noisy coefficients with low distortion in the enhanced speech signal, the value of threshold has to be different in the voiced and unvoiced frames. The value of the threshold in the unvoiced frames is smaller than it in the voiced frames. The entropy measure is used to detect voiced/unvoiced frames. The entropy of the coefficients in the unvoiced frames is less than a predefined threshold. So by selecting a suitable value for the threshold, a good performance could be achieved for the voice/unvoiced decoder.

The K-L divergence [18] has been widely used in the field of voice activity detection [19, 26, 27]. The K–L divergence between two probability functions p(x) and q(x) is defined as:

$$D(p \| q) = \sum_{i=1}^{M} p(x_i) \log \frac{p(x_i)}{q(x_i)} \tag{3}$$

The K–L divergence is always nonnegative and is zero if and only if $p = q$. However, it is not a true distance between distributions since it is not symmetric and does not satisfy the triangle inequality. In order to have a symmetric distance, Symmetric K-L divergence has been defined as follows [19]:

$$SD(p \| q) = D(p \| q) + D(q \| p) \tag{4}$$

Voice activity detection based on K-L divergence is discussed here. First, the signal is pre emphasized and segmented into several frames with a pre determined window shift. Then a wavelet packet transform is applied to the input frames of the signal. A symmetric K-L divergence is used for the estimation of the distance between speech and noise distributions. The decision rule is formulated in terms of the average sub-band K-L divergence that is compared to a noise-adaptable threshold. In the silent parts of the input signal, the K-L divergence is less than the threshold. So by using a suitable threshold, the silent parts of the input signal are accurately detected.

In the present work, we use K-L divergence to determine suitable threshold in order to reduce background noise without any more distortion in the enhanced speech signal.

IV. The Proposed Method

Removing noise components by thresholding the wavelet coefficients is based on the observation that in many signals, energy is mostly concentrated in a small number of wavelet dimensions. The coefficients of these dimensions are relatively large compared to other dimensions or to any other signal that has its energy spread over a large number of coefficients. Hence, by setting smaller coefficients to zero, one can nearly optimally eliminate noise while preserving the important information of the original signal.

The probability distribution of the signal coefficients could be obtained by calculating the histogram of the coefficients. Due to the thresholding idea, in the range of –T and T, the symmetric K-L divergence between the noisy speech and the estimated noise coefficients is expected to be almost zero. Also we can expect that the symmetric K-L divergence between the noisy speech and clean speech coefficients would be zero outside this range. Fig. 3 shows the discussed idea.

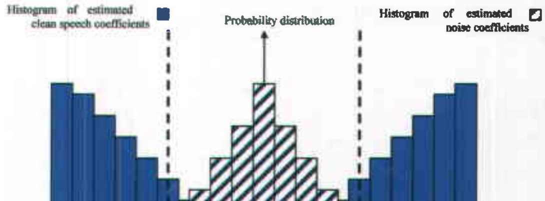


Figure 3. Wavelet coefficient histogram with response to the thresholding idea

The histogram of coefficients will be calculated for each sub-band. One of the critical parameter is the number of bins in the histogram. This parameter has been chosen equal to the square root of the number of samples divided by two [28]. The probability distribution of the coefficients could be estimated using (5).

$$P(x_i)=\frac{Number\ \ of\ \ Coefficients\ \ in\ \ ith\ \ bin}{The\ \ whole\ \ number\ \ of\ \ the\ \ coefficients} \tag{5}$$

As discussed there are two adaptive thresholds that satisfy equations (9-11).

$$SD(P_{NS},P_N)\approx 0, Where\ \ the\ \ number\ \ of\ \ bins\in[T_1,T_2] \tag{6}$$

$$SD(P_{NS},P_S)\approx 0, Where\ \ the\ \ number\ \ of\ \ bins\in[1,T_1-1] \tag{7}$$

$$SD(P_{NS},P_S)\approx 0, Where\ \ the\ \ number\ \ of\ \ bins\in[T_2+1, Number\ \ of\ \ bins] \tag{8}$$

Where P_{NS}, P_N and P_S are the probability distributions of noisy speech, estimated noise and estimated clean speech coefficients, respectively. By substituting (4) in the above equation, we obtain (9) and (10).

$$\sum_{i=T_1}^{T_2}[P_{NS}(i)-P_N(i)]\log\frac{P_{NS}(i)}{P_N(i)}\approx 0 \tag{9}$$

$$\sum_{i=1}^{T_1-1}[P_{NS}(i)-P_S(i)]\log\frac{P_{NS}(i)}{P_S(i)}+\sum_{i=T_2+1}^{BinNum}[P_{NS}(i)-P_S(i)]\log\frac{P_{NS}(i)}{P_S(i)}\approx 0 \tag{10}$$

Proof:

From (6) we have:

$$\sum_{i=T_1}^{T_2}P_{NS}(i)\log\frac{P_{NS}(i)}{P_N(i)}+\sum_{i=T_1}^{T_2}P_N(i)\log\frac{P_N(i)}{P_{NS}(i)}\approx 0 \tag{11}$$

The (11) could be rewrite as follows:

$$\sum_{i=T_1}^{T_2}P_{NS}(i)\log\frac{P_{NS}(i)}{P_N(i)}-\sum_{i=T_1}^{T_2}P_N(i)\log\frac{P_{NS}(i)}{P_N(i)}\approx 0 \tag{12}$$

From (12) we have:

$$\sum_{i=T_1}^{T_2}[P_{NS}(i)-P_N(i)]\log\frac{P_{NS}(i)}{P_N(i)}\approx 0 \tag{13}$$

In order to obtain (10), first, we have to compose (7) and (8). By substituting (4) in the resulted equation, (10) will be obtained.

The adaptive thresholds T_1 and T_2 will be obtained by solving (9) and (10). Also, we have to choose an accurate value instead of almost zero (≈0). This value could be chosen in order to the signal to noise ratio (SNR) and the type of the additive noise. In this work this value has been selected in the range of 0.01 and 0.03.

As discussed in the previous part, two adaptive thresholds T_1 and T_2 will be obtained by solving (9) and (10). These two thresholds satisfy equations (9-11). It means that in the range of $[T_1, T_2]$ the symmetric K-L divergence between histograms of noisy speech and noise wavelet coefficients is nearly equal to zero and outside of this range the symmetric K-L divergence between the noisy speech wavelet coefficients and clean speech coefficients is nearly equal to zero. So we can use these two thresholds to label the noisy speech wavelet coefficients outside the range of $[T_1, T_2]$, as clean speech coefficients; and mark the noisy speech wavelet coefficients in the range of $[T_1, T_2]$, as noise coefficients. After detecting the voiced/unvoiced segments, a general noise reduction filter could be applied to noisy speech coefficients in the range of $[T_1, T_2]$ to efficiently reduce the background noise. Fig. 4 shows the proposed enhancement method.

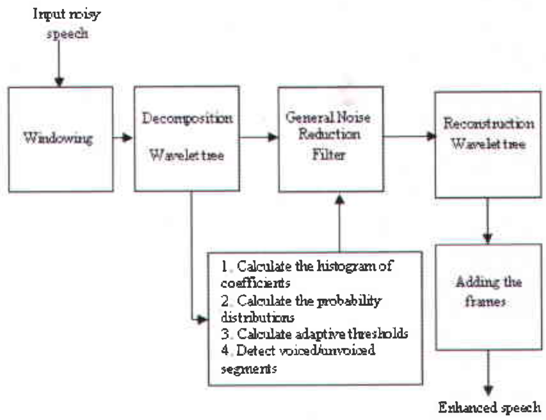


Figure 4. Proposed voiced/unvoiced detection system based on K-L divergence

Instead of using the proposed method just as a voice activity detector, we can use it in order to calculate suitable threshold values for thresholding noisy wavelet coefficients. In this new speech enhancement method, we need to only thresholding without computing and using a filter. Fig. 5 shows the proposed speech enhancement method due to this new idea. As the figure shows the proposed algorithm composed of 9 steps listed in the following:

- Divide the input signal into several segments due to the Frame Size.
- Perform wavelet packet transform to the input segment.
- Estimate the noise and clean speech in each segment.

- Perform wavelet packet transform to the estimated signals.
- In each sub-band, calculate the histogram of coefficients.
- Solve the (9) and (10) and calculate the adaptive thresholds.
- Apply hard thresholding on the noisy coefficients using adaptive thresholds.
- Perform inverse wavelet packet transform to the enhanced coefficients.
- Add the overlapped frames to construct the whole speech signal.

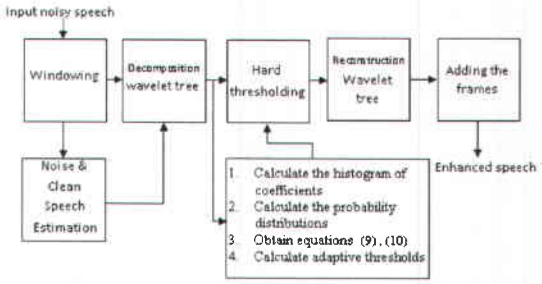


Figure 5. Proposed speech enhancement system in order to apply hard threshold

As T_1 and T_2 are the number of bins in the histogram of noisy speech coefficients, we have to modify the hard thresholding method. The coefficients with values in the range determined by T_1 and T_2, marked as noise coefficients and will be removed. The other coefficients marked as clean speech coefficients and will be remained with no change.

If we select an actual pdf for noisy speech, clean speech and noise wavelet coefficients, a general formula for the threshold value in each sub-band would be obtained by solving (9) and (10) in continuous mode. This new issue has been considered in our recent research.

V. IMPLEMENTATION AND EVALUATION

We carried out several experiments to evaluate the performance of the proposed method and to compare it with the classic wavelet based speech enhancement methods. We selected 10 utterances from TIMIT database (5 male speakers and 5 female speakers). Five different noises (white, pink, babble, car and factory from NATO RSG-10 database) were added to clean speeches at different SNR levels (0dB, 5dB, 10dB, 15dB, 20dB, 25dB). Three Different frame sizes (128 samples, 256 samples and 512 samples) were used for windowing the input signal. A 3-level wavelet decomposition tree with db10 bases function was exploited. Hard thresholding function were used for thresholding the wavelet coefficients. Our proposed method was compared to the classic speech enhancement method discussed in section 2. We used Donoho's method and mini-max method for setting the threshold value in the classic speech enhancement method. SNR and PESQ [29] were used for evaluating our proposed method. SNR is estimated using (14).

$$SNR = \sum_{i=1}^{SignalLength} S(i)^2 \Big/ \sum_{i=1}^{SignalLength} [\hat{S}(i) - S(i)]^2 \tag{14}$$

Where S and $\hat{S}$ are the source signal and the destination signal, respectively.

PESQ is an objective measurement tool that predicts the results of subjective listening tests on telephony systems. PESQ uses a sensory model to compare the original unprocessed signal with the degraded signal. The resulting quality score is analogous to the subjective "Mean Opinion Score" (MOS) [29]. In the speech processing tests the PESQ scores are in the range of 0 and 4.5. The higher the PESQ score, the higher the performance of the noise reduction method.

The results of implementing the proposed system are reported in this section. A comparison with the classic wavelet based speech enhancement method [17, 30] was provided.

Table 1 shows the average of results of evaluating the proposed method for all 10 speakers and all noise types and all input SNR values. Table 2 shows the total average results of evaluating the proposed method based on SNR measure.

TABLE I. EVALUATING THE PROPOSED METHOD BASED ON SNR MEASURE FOR 10 SPEAKERS AND 5 NOISE TYPES

	Output SNR (dB)														
	White			*Pink*			*Factory*			*Babble*			*Car*		
Input SNR (dB)	***P***	***D***	***M***	***P***	***D***	***M***	***P***	***D***	***M***	***P***	***D***	***M***	***P***	***D***	***M***
0	6.9	5	5.1	7.7	5.4	3.6	8	5.6	4.2	8.6	5.8	5	8	5.6	4.3
5	10.9	9.4	9.8	11.7	9.9	8.4	11.9	10	8.9	12.4	10.4	9.8	12.1	10.1	8.9
10	14.9	14.2	14.3	15.9	14.7	13.37	16.4	14.9	13.7	16.6	15.1	14.7	16.6	14.8	13.8
15	19.7	18.9	19	20.2	19.5	18.2	21.4	19.8	18.6	21.5	19.9	19.6	21.1	19.6	18.7
20	24.8	23.8	23.7	25.3	24.3	23.1	26.3	24.7	23.5	26	24.7	24.6	25.7	24.4	23.7
25	29.4	28.7	28.5	29.5	29.3	28.1	30.5	29.6	28.4	30	29.6	29.5	30.2	29.2	28.6
Average	17.7	16.6	16.7	18.3	17.1	15.8	19.0	17.4	16.2	19.1	17.5	17.2	18.9	17.2	16.3

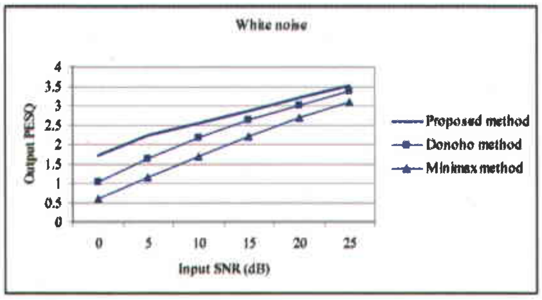
P: Proposed method, D: Donoho's method, M: Mini-max method

TABLE II. TOTAL AVERAGE OF RESULTS OF EVALUATING THE PROPOSED METHOD BASED ON OUTPUT SNR

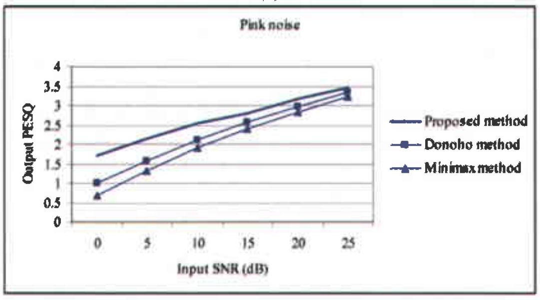
	Proposed method	Donoho's method	Mini-max method
Total Average	18.7dB	17.2dB	16.5dB

As the results show, in average, our proposed method has improved the input SNR up to 1.5 dB and 2.2 dB in comparison to the Donoho's method and Mini-max method, respectively, for different input SNR values and different noise types.

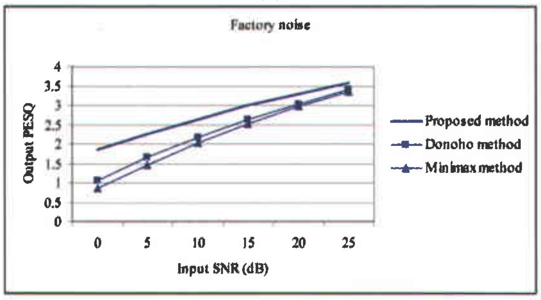
Fig. 6 (a-e) shows the average of results of evaluating the proposed system, based on PESQ measure, for 10 speakers and for 5 different noise types.



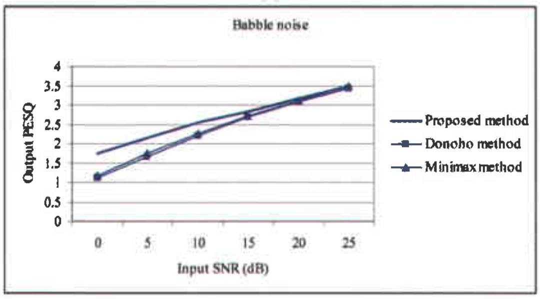
(a)



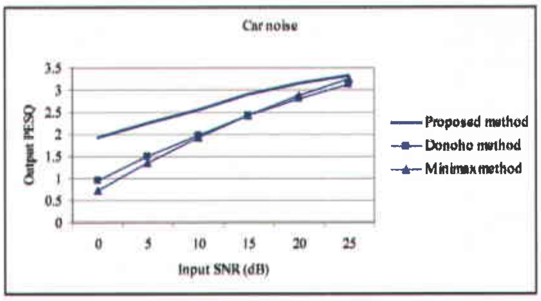
(b)



(c)



(d)



(e)

Figure 6. Evaluating the proposed method based on the PESQ measure for *10* speakers and *5* additive noise types

As Fig. 6 shows, in average, our proposed method has improved the PESQ measure up to 1.2, 0.7 in comparison to Donoho's method and Mini-max method, respectively.

VI. Conclusion

In this paper we proposed a wavelet based speech enhancement that exploits K-L divergence for setting its threshold values. In each sub-band, two suitable threshold values have been obtained by solving equations based on symmetric K-L divergence between probability distributions of noisy speech and noise wavelet coefficients. A comparison with classic speech enhancement system that exploits Donoho's threshold value and mini-max threshold value was provided. We conducted some tests using the TIMIT database in order to assess the performance of the proposed method and to compare it to the previous speech enhancement methods. The algorithm is evaluated using the PESQ and the gain on input SNR. We obtained an improvement of up to 2.2dB on SNR and 1.2 on PESQ score for the proposed method in comparison to the results of the previous wavelet based methods for tested noises.

We need to compute the adaptive threshold values in each sub-band based on K-L divergence. This cause that our method has more computational complexity in comparison to other wavelet based speech enhancement methods; but its complexity is acceptable with attention to its performance.

References

[1] S., Vaseghi, Advanced Digital Signal Processing and Noise Reduction, 3rd ed., John Wiley & Sons, 2000.

[2] S.F., Boll, "Suppression of Acoustic Noise Speech Using Spectral Subtraction," IEEE Trans. On Acoustic, Speech and Signal Processing, vol.29, pp. 113-120, 1979.

[3] B., Chen, Ph. C., Loizou, "A Laplacian-based MMSE estimator for speech enhancement," Journal on Speech Communication , vol. 49, no. 2, pp. 134–143, 2007.

[4] J. H., Chang, "Complex Laplacian Probability Density Function for Noisy Speech Enhancement," IEICE Electronic Express, vol. 4, no. 8, pp. 245-250, 2007.

[5] T. H., Dat, K., Takeda, F., Itakura, "Multichannel Speech Enhancement Based on Generalized Gamma Prior Distribution with Its Online Adaptive Estimation," IEICE TRANS. INF. & SYST., Special Section on Statistical Modeling for Speech Processing, vol. E91–D, no.3, pp. 439-437, 2008.

[6] T. H., Dat, K., Takeda, F., Itakura, "Gamma Modeling of Speech Power and Its On-Line Estimation for Statistical Speech Enhancement," IEICE TRANS. INF. & SYST., Special Section on Statistical Modeling for Speech Processing, vol. E89–D, no.3, pp. 1040-1049, 2006.

[7] D. L., Donoho, "Denoising by soft thresholding," IEEE Trans. On Infromation Theory, vol. 41, no. 3, pp. 613-627, 1995.

[8] D. L., Donoho, I. M., Johnstone, "Adapting to unknown smoothness via wavelet shrinkage," J. Amer. Statist. Assoc. , vol. 90, pp. 1200–1224, 1995.

[9] Y., Hu, Ph. C., Loizou, " Speech Enhancement Based on Wavelet Thresholding the Multitaper Spectrum," IEEE Transactions on Speech and Audio Processing, vol.12, no.1, pp. 59-67, 2004.

[10] M. Jansen, Noise Reduction by Wavelet Thresholding, Springer, vol. 161, 2001, pp. 10-140.

[11] J., Weaver, D. M., Healy, L.D., Cromwell, "Filtering noise from images with wavelet transforms," Magnetic Resonance in Medicine, vol. 21, no. 2, pp. 288-295, 1991.

[12] I. M., Johnston, B. W., Silverman, "Wavelet threshold estimators for data with correlated noise," J. Roy. Statist. Soc. B, vol. 59, pp. 319-351, 1997.

[13] J. W., Seok, K.S., Bae, "Speech enhancement with reduction of noise components in the wavelet domain," ICASSP, vol. 2, pp. 1323 – 1326, 1997.

[14] A., Bruce, H.Y., Gao, "Wave Shrink: Shrinkage Functions and Thresholds," Technical report, Research Report 39, Statistical Sciences Division, MathSoft, 1996.

[15] I. J., Kim, S. I., Yang, Y., Kwon, "Speech enhancement using adaptive wavelet shrinkage," Proceedings of IEEE International Symposium on Industrial Electronics, ISIE-2001, vol. 1, pp. 501-504, 2001.

[16] A., Lallouani, M., Gabrea, C.S., Gargour, "Wavelet Based Speech Enhancement Using Two Different Threshold-Based Denoising Algorithms," IEEE Transaction, CCECE, Niagara Falls, pp. 315-318, 2004.

[17] S., Ayat, M.T., Manzuri, R., Dianat, "Wavelet based speech enhancement using a new thresholding algorithm," IEEE-Intelligent Multimedia, Video and Speech Processing, pp. 238-241, 2004.

[18] R. B., Ash, Information Theory, Urbana Illinois, 1965, pp. 1-24.

[19] J., Ramírez, J. C., Segura, C., Benítez, A., de la Torre, A. J., Rubio, "A New Kullback–Leibler VAD for Speech Recognition in Noise," IEEE Signal Processing Letters, vol. 11, no. 2, pp. 266-269, 2004.

[20] M., Babaie-Zadeh, Ch., Jutten, "A general approach for mutual information minimization and its application to blind source separation," www.elsevier.com/locate/sigpro, Journal on Signal Processing, vol. 85, no. 5, pp. 975-995, 2005.

[21] D., Smith, J., Lukasiak, I. S., Burnett, "An analysis of the limitations of blind signal separation application with speech," www.elsevier.com/locate/sigpro,, Journal on Signal Processing, vol. 86, no. 3, pp. 353-359, 2006.

[22] M., Babaie-Zadeh, Ch., Jutten, K., Nayebi, "Using Multivariate Score Functions in Source Separation: Application to Post Non-Linear Mixtures," Scientia-Iranica, vol. 9, no. 4, pp. 409-418, 2002.

[23] D. A., Iancu, "Separating Speech Signals Using Independent Component Analysis," www.mit.edu/~daniancu/Old_proj/H_es257/report.pdf, 2005.

[24] H. Ch., Wu, J. C., Principe, "Minimum Entropy Algorithms for Source Separation," Proceedings. of Midwest Symposium on Circuits and Systems, vol. 9, no. 12, pp. 242-245. 1999.

[25] B., Zamani Dehkordi, M., Mohammadi, Sh., Tabibian, B., Nassersfarif, A., Akbari, "Speech Enhancement using Wavelet Transform based on the voiced/unvoiced detection," 13th CSICC Kish Island, 2008.

[26] J., Shuojun, G., Haitao, Y., Fuliang , "A New Algorithm for Voice-Activity Detection Based on the Wavelet Transform," Proceeding of International Symposium on Intelligent Multimedia, Video and Speech Processing , pp. 222-225, 2004.

[27] Sh. H., Chen, J. F., Wang, "A Wavelet Based Voice Activity Detection in Noisy Environment," IEEE Trans., pp. 995-998, 2002.

[28] E., Gosse, R., Johnson, Pattern Recognition and Image Analysis, Prentice Hall, 1994, pp. 150-160.

[29] "Objective Measurement of Active Speech Level," ITU-T, Geneva, Recommendation P.56, 1995.

[30] M., Bahoura, J., Rouat, "Wavelet speech enhancement based on time–scale adaptation," NOLISP-Speech Communication, vol. 48, pp. 1620-1637, 2006.

Shima Tabibian is a Ph.D. student in Artificial Intelligence and Robotics at Iran University of Science and Technology. Her research interests include Speech Processing. Her work in speech processing focuses on Speech Enhancement and Keyword Spotting.

Ahmad Akbari is an Associated Professor of Computer Engineering at Iran University of Science and Technology. He received his Ph.D from university of Rennes, France, in 1995. His work focuses on Speech signal analysis, Discrete speech recognition, Speech enhancement, Spoken dialog systems, Speech-based human-computer interfaces, Voice over IP and Voice applications on web.