

Adaptive Speech Enhancement Using Partial Differential Equations and Back Propagation Neural Networks

¹Mojtaba Bandarabadi, ²MohammadReza Karami-Mollaei, ³Reza Ghaderi, ⁴Meysam Salahshoor

^{1,2,3,4}Department of ECE, DSP Lab., Babol Noshirvani Univ. of Tech., Babol, Iran

Abstract: In this work, we propose a new approach to improve the performance of speech enhancement technique based on partial differential equations. As we know, the real-world noise is highly random in nature. So we try for reduction of white Gaussian noise. The proposed method was evaluated on several speakers. The subjective and objective results show that the new method highly improves speech enhancement. Comparisons of several methods are reported.

Key word: Speech enhancements, Partial differential equations, Fast Fourier transform, Back propagation neural networks

INTRODUCTION

In many speech communication systems, background noise causes the quality of speech to degrade. In most of speech processing applications such as mobile communications, speech recognition and hearing aids, removing the background noise in a noisy environment is inevitable. So, speech enhancement as a necessity for related applications has been widely studied in recent years. There are several techniques such as spectral subtraction (Deller *et al.*, 2000; Boll, 1979; Berouti *et al.*, 1979; Kamath and Loizou, 2002; Ghanbari *et al.*, 2004; 2004; Donoho, 1995), Wiener filtering (Chen *et al.*, 2006), hidden Markov modeling (Sameti *et al.*, 1998), wavelet-based methods (Chen *et al.*, 2006; Sheikhzadeh *et al.*, 2001; Seok and Bae, 1997), adaptive filtering (Chang *et al.*, 2002) and signal subspace methods (Klein and Kabal, 2002). To solve the mentioned problems using Partial Differential Equations is a new method. In this method the existing changes in speech signal under research, like the model of air temperature oscillation, is considered in which the air current from warmer circumstances to colder one, is done until two circumstances are balanced with the same temperature. These temporal changes are shown as Gaussian function. While using PDE in de-noising signal, speech signal oscillation model as Gaussian Function are considered. In this accord, every sudden change in speech signal is known as noise. In using PDE a parameter is defined as propagation coefficient, whose value expresses oscillation intensity from one condition to another. In the other hand this parameter plays an important role in de-noising. The researches show propagation coefficient in speech signal often is like a nonlinear function.

One of the important parameters that exist in PDE is recursive coefficient. Our researches show that this parameter plays a determining role in de-noising the speech signal. Choosing low value for this parameter causes non-denoising in improper quantity on speech signal and choosing a large value for that will cause the details of speech signal to be eliminated. One method for de-noising based on PDE is calculating recursive coefficient according to try and error upon some speech signal and its exertion upon the other signals. The problem that comes into existence here is that the tissue of speech signal, are diverse and the applied noise upon them are different too and these constraints cause these choices not to de-noise effectively and with high efficiency. Our purpose in this work is to be able to obtain the recursive coefficient existing in PDE by considering researched speech signal tissue.

In section 2 of this work, PDE and previous methods weakness are investigated. In section 3, the proposed method will be presented in this article. In section 4, some standards are used to evaluate the efficacy of the proposed method and the result of experiments would be presented. Finally in section 5 the conclusion of the work is also included.

Partial Differential Equations:

Partial differential equations (PDEs) are one of the methods that have been initially used for image de-noising. The authors in (Wu *et al.*, 2007) developed a new approach for image de-noising based on PDE.

Corresponding Author: Mojtaba Bandarabadi, Department of ECE, DSP Lab., Babol Noshirvani Univ. of Tech., Babol, Iran
E-mail: m.bandarabadi@gmail.com

Results in this work show that PDEs are very powerful for image de-noising compared to other exiting approaches. These results prompted us to use PDE for signal de-noising. To the best of our knowledge, PDEs have not yet been used for signal de-noising. Results in this research indicate that PDEs are suitable for signal de-noising.

As mentioned above, partial differential equations have been initially introduced for image de-noising. One of these equations which have been used in image processing applications is the heat equation. This equation is defined as follows:

$$\frac{\partial I(x, y, t)}{\partial t} = \nabla \cdot (c(x, y, t)) \nabla I(x, y, t) \tag{1}$$

Where $I(x,y,t)$ is the noisy image and $c(x,y,t)$ is the influence coefficient. In this method, gradient in four directions of any pixels are calculated and then their influencing coefficients are obtained to reduce the noise using (1). Then, with a number of iterations, the enhanced image is obtained.

But in the form of a signal, the gradient of each sample is computed using the samples before and after the current sample. Then, the influencing coefficients in each directions of the current sample, forward (c_f) and backward (c_b), are computed as follows:

$$S(x, t+ \Delta t) = S(x, t) + \Delta t(d_f c_f + d_b c_b) \tag{2}$$

$$d_f = S(x - \Delta x, t) - S(x, t) \quad c_f = \frac{1}{1 + \left(\frac{d_f}{k}\right)^2} \tag{3}$$

$$d_b = S(x + \Delta x, t) - S(x, t) \quad c_b = \frac{1}{1 + \left(\frac{d_b}{k}\right)^2}$$

In equations (2) and (3), $S(x,t)$ is the noisy signal, d_f, d_b are the gradients in forward and backward directions; c_f, c_b are the corresponding influencing coefficients for each of the directions; k is a constant value between 5 and 100, Δt is a coefficient between 0.1 to 0.3 representing the step of de-noising in each iteration, and Δx is the sampling rate.

The output signal is reapplied to the algorithm at the next iteration to gradually reduce the noise. This process is repeated for a number of times that lead to the favorite de-noised signal. The criteria to stop the iteration can be the SNR or MSE value.

Research Method:

As mentioned, our purpose in this work is calculating recursive coefficient in an adaptive form for speech signal de-noising in the way that no need to try and error experiment is for obtaining its proper value in the different segments of signal. As in eq.4 shown, in using PDE as much as the recursive coefficient value smaller in a procedure of de-noising is, PDE coefficient is Smaller and at the result the signal noise will be lessened and as much as the value greater is the obtained speech signal with de-noising will be more flattened. Fast Fourier transform (FFT) on a speech signal showing that the existing changes in signal can be an appropriate standard for choosing recursive coefficient. As much as a segment of speech signal possesses more details, the proper value of recursive coefficient is more for de-noising this part of speech signal enhancively. we use this property to obtain the proper value of the recursive coefficient in different segments of speech signal adaptively .The algorithm of the proposed method is in this order which at first speech signal is divided into 15ms intervals (256 samples with sampling rate 16khz), then in order to find P.D. coefficient enhanced for each segment of speech signal, FFT on that segment of signal applied and 30 frequencies at which the most signal energy density exists , would be selected in the order of the energy density as characteristic. In this stage the obtained properties will be applied into the trained MLPNN (multi layer Perceptron neural network) and in output of network, there would be the enhanced P.D. coefficient by placing this coefficient in PDE and its exertion upon this segment of the noisy signal, the signal noise would be enhanced and lessened.

Neural Network Architecture:

Back-propagation multi layer Perceptron neural network with one hidden layer is used for classification. It has 30 input units ($x_1 \dots x_{30}$), 20 hidden units (we use the several numbers of units in this layer and select the number that best performance is obtained) and 15 output units $y_1 \dots y_{15}$. The NN structure is shown in Figure 1.

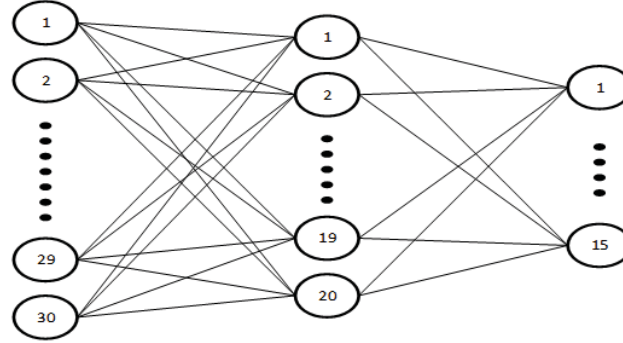


Fig. 1: Structure of BPNN used.

The feature vectors that obtained from the speech signal using fast Fourier transform are normalized using the following equation:

$$x_{i,new} = \frac{x_i - \min(x)}{\max(x) - \min(x)} \tag{4}$$

The response of output unit (y^j) considered +1 if its activation is equal or greater than zero, otherwise the output unit value is -1. The learning rate of neural network is 0.05. The Nguyen-Widrow algorithm is used to initialize the weights and bipolar sigmoid function is used as activation function (Laurene Fausett, 1994). We train the BPNN with momentum. The momentum value is 0.95.

Performance Evaluation:

In this part of the proposed method, noise reduction of speech signal is applied upon about 60 samples of TIMIT standard speech signals to evaluate the efficacy of its performance. The proposed algorithm has been tested on the spoken English sentence. The sentence is about 2.7 sec with the sampling rate of 16 kHz and spoken by a male speaker. In this evaluation, we benefited from some of the proper standards which are used often in the signal processing and the obtained results from applying these standards on the proposed method are illustrated.

Signal to Noise Ratio Metric:

The signal to noise ratio (SNR) is a well known measure in signal processing. It is defined as below:

$$(SNR)_{db} = \frac{Signal\ Power}{Noise\ Power} \tag{5}$$

This criterion indicates that how the noise was degraded in the de-noised signal. In other words, the larger the SNR value represents the de-noised signal is closer to the original. The global signal-to-noise ratio (SNR) values at this table were determined by the following equation as the objective evaluation criterion:

$$SNR = 10 \times \log_{10} \left(\frac{\sum_{n=1}^N s^2(n)}{\sum_{n=1}^N |s(n) - \hat{s}(n)|^2} \right) \tag{5}$$

Where N is the number of the samples in the clean and enhanced signals. The average SNRs of 60 enhanced speech signals are shown in Table 1. Figures 2(a), 2(b) and 2(c) show the results in the time domain of clean, noisy and enhanced speech by the proposed algorithm.

Table 1: Average SNRs of 60 enhanced speech signals

Input SNRs (dB)	Output SNRs (dB)
-10	2.7
-5	6.2
0	9.6
5	12.3
10	15.7
15	19.7

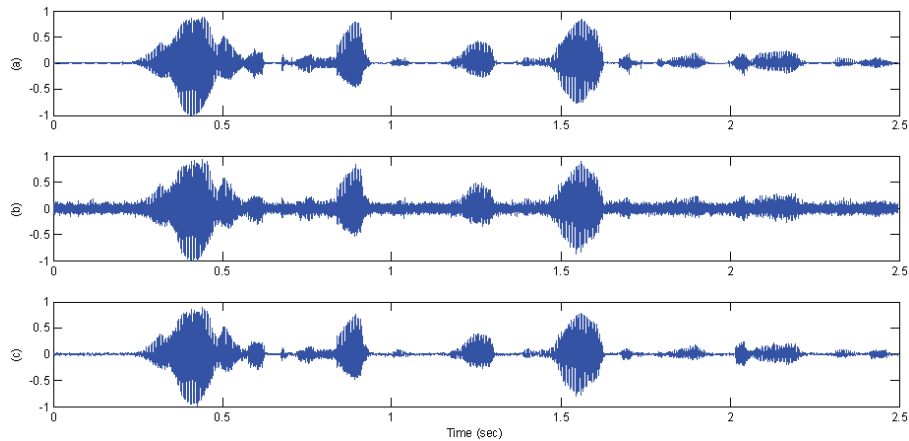


Fig. 2: The time domain results for: (a) Clean speech. (b) Noisy speech corrupted by WGN (SNR=10dB). (c) Enhanced speech by the proposed method (SNR=15.83dB).

We have also compared the performances of our proposed algorithm with six algorithms including the basic spectral subtraction algorithm proposed by Boll (1979) (named “BSS”), the proposed algorithm by Kamath & Loizou (2002) (named “MBSS”), the proposed algorithm by Ghanbari & Karami (2004) (named “SSWD”), the basic wavelet thresholding algorithm proposed by Donoho (1995) (named “BWT”), the proposed algorithm by Sheikhzadeh, & Abutalebi (2001) (named “IWBSE”) and the proposed algorithm by Soek & Bae (1997) (named “SERNCWD”). We have implemented mentioned algorithms and tested them on 60 various speech signals spoken by several speakers and chosen from TIMIT database. The average global SNR results for the performance on the noisy signal by WGN are depicted in Figure 3. As can be seen, the proposed algorithm has considerable performance improvements. Another noteworthy point which should be noticed is that the six algorithms which have been compared with our proposed algorithm have relatively bad spectrograms in comparison to the new algorithm. So, the new algorithm was shown to be much better in comparison to the other algorithms.

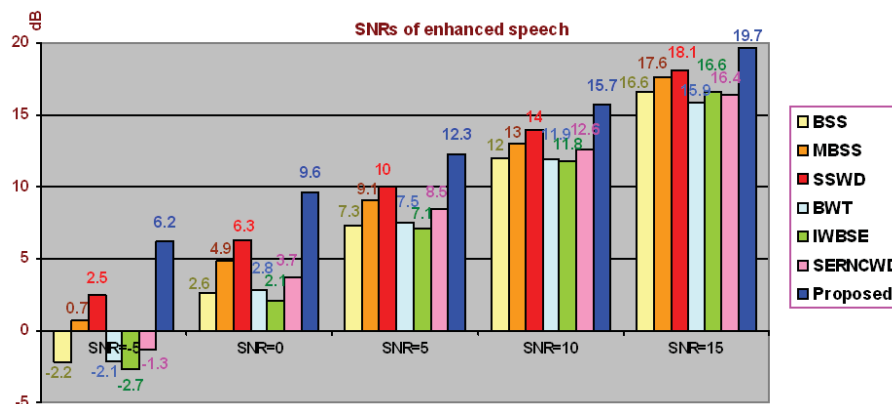


Fig. 3: The average performance of seven algorithms for sixty noisy signals by WGN

Speech Spectrograms:

Objective measures do not give indications about the structure of the residual noise. Speech spectrograms constitute a well-suited tool for observing this structure. The speech spectrogram for SNR of 10 dB is obtained by using a Hanning window of 128 samples with 50 % overlap. Fig. 4 shows the speech signals and its corresponding spectrograms.

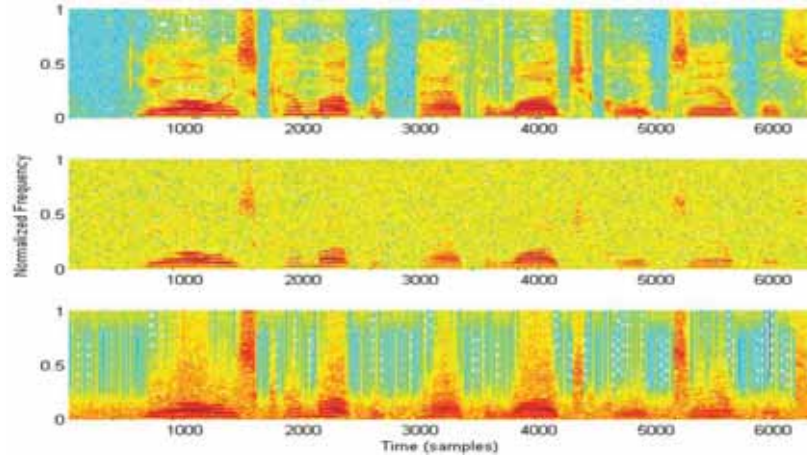


Fig. 3: Spectrograms for: (a) Clean speech. (b) Noisy speech corrupted by WGN (SNR=10dB). (c) Enhanced speech by the proposed method (SNR=16.25dB)

Summary and Concluding Remarks:

In this work, a method for enhancing and improving the existing methods of de-noising is presented using PDE in which the existing value of recursive coefficient in PDE coefficient is a constant number for different segments of a signal.

This coefficient is obtained by using FFT. The FFT of the signal is one of the parameters that express the changes of speech signal segment. The obtained results from the experiment (test, trial are evaluated with the various standards). The results show the considerable enhancement of speech signal versus the previous methods.

REFERENCES

Deller, J.R., J.H.L. Hansen and J.G. Proakis, 2000. *Discrete-time processing of speech signals*, 2nd edition, IEEE Press.

Boll, S.F., 1979. "Suppression of acoustic noise in speech using spectral subtraction", *IEEE Trans. on Acoust. Speech & Signal Processing*, 27: 113-120.

Berouti, M., R. Schwartz, and J. Makhoul, 1979. "Enhancement of speech corrupted by acoustic noise", *Proc. IEEE ICASSP*, Washington DC, April., pp: 208-211.

Kamath, S. and P. Loizou, 2002. A multi-band spectral subtraction method for enhancing speech corrupted by colored noise, *Proceedings of ICASSP-2002*, Orlando, FL.

Ghanbari, Y., M. Karami, B. Amelifard, 2004. "Improved Multi-band Spectral Subtraction Method for Speech Enhancement", *Proc. of the 6th IASTED Int. Conf. on Signal and Image Processing*, USA, pp: 225-230.

Ghanbari, Y., M. Karami, 2004. "Spectral subtraction in the wavelet domain for speech enhancement", *International Journal of Software and Information Technologies (IJSIT)*, 1: 26-30.

Donoho, D.L., 1995. "De-noising by soft-thresholding", *IEEE Transactions on Information Theory.*, 41(3): 613-627.

Chen, J., J. Benesty, Y. Huang and S. Doclo, 2006. "New Insights into the Noise Reduction Wiener Filter". *IEEE Transactions on Audio, Speech and language Processing*, 14: 4.

Chang, S., Y. Kwon, S. Yang, I. Kim, 2002. "Speech enhancement for non-stationary noise environment by adaptive wavelet packet", *Proceedings of the 2002 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP 2002)*, 1: 561-564.

Sheikhzadeh, H., H.R. Abutalebi, 2001. "An Improved Wavelet-Based Speech Enhancement System," in Proc. 7th European Conference on Speech Communication and Technology (EuroSpeech), Aalborg, Denmark, Sep.

Seok J., K. Bae, 1997. "Speech enhancement with reduction of noise components in the wavelet domain", *Proceedings of the 1997 IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '97)*, 2(21-24): 1323-1326.

Laurene Fausett, 1994. "Fundamentals of Neural Networks", Prentic Hall Intenational, Inc.,

Wu, Y.-D., Y. Sun, H.-Y. Zhang, S.-X. Sun, 2007. "Variational PDE based image restoration using neural network", *IET Image Process.*, 1(1): 85–93.

Klein, M. and P. Kabal, 2002. "Signal subspace speech enhancement with perceptual post-filtering", *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing*, Orlando, FL, pp: 1-537-1-540.

Sameti, H., H. Sheikhzadeh, Li Deng, R.L. Brennan, 1998. "HMM-Based Strategies for Enhancement of Speech Signals Embedded in Nonstationary Noise", *IEEE Transactions on Speech and Audio Processing*, 6:5.