LP-Based Over-Sampled Subband Adaptive Noise Canceller for Speech Enhancement in Diffuse Noise Fields

Soheil Khorram Hossein Sameti Hadi Veisi
Computer Engineering Department, Sharif University of Technology, Tehran-Iran
khorram@ce.sharif.edu, sameti@sharif.edu, veisi@ce.sharif.edu

Abstract

Adaptive Noise Cancellers (ANCs) do not provide sufficient noise reduction in the diffuse noise fields. In this paper, a new hybrid structure is proposed as a solution to this problem. The proposed system is a combination of two subsystems, an ANC and a new multistage post-filter. The post-filter is based on linear prediction (LP) and attempts to extract speech component by using intermediate ANC signals. The system is implemented on an over-sampled DFT filterbank with different analysis and synthesis prototype filters. The experimental results using various quality measures show that the proposed system is superior to both the subband ANC and subband LP based speech enhancement systems.[1]

1. Introduction

Speech enhancement is a necessary part of almost every current speech processing system [1]. Adaptive noise cancellation is a common technique for enhancing the speech quality in automobiles and other moving vehicles [2]. Real environments need long tap length full-band Adaptive Filters (AFs) to be identified in noise cancellers. Usually Subband Adaptive Filters (SAFs) are used to reduce filter length and improve the convergence rate [3, 4].

Real-life noise fields in vehicular applications are correlated only at lower frequencies because they are approximately diffuse [4]. In this case, Adaptive Noise Cancellers (ANCs) do not work efficiently and a post-filtering is normally required [5, 6]. Abutalebi et al. [3, 4] proposed a hybrid system that integrates a SAF and a Wiener filter. They used a Voice Activity Detector (VAD) to control both the adaptation in the SAF and the noise spectrum estimation in the Wiener filter. Accuracy of VADs degrades in the noisy environments, so in this paper we are looking for a method to eliminate the VAD block.

We will demonstrate that the speech and noise component of ANC output are approximately separate in frequency domain, so speech enhancement can be performed by an appropriate post-filtering. We have designed a Linear Prediction (LP) based post-filter which follows the work of Kawamura et al. [7, 8].

In Section 2 we study the challenge of ANCs in diffuse noise conditions. Section 3 explains the fundamental linear predictor and its problems for speech enhancement. Section 4 gives the proposed multistage filtering approach for white noise suppression. The overall hybrid system which is extended to remove colored noises and its appropriate subband implementation are presented in Section 5 and 6. Experimental results are given in Section 7. Finally, we summarize our works and conclude in Section 8.

2. ANC challenge in diffuse noise fields

Fig. 1 shows the structure of ANC in a real noisy environment. Reflection of a noise signal from a hard surface can be considered as a new noise signal generated from a new noise source that located on the opposite side of the reflective boundary. When a sufficient number of noise sources are considered, the noise field approaches diffuse noise model [2].

For analyzing the noise reduction ability of ANCs in various conditions, several authors [2-4] have used noise reduction factor which is defined by

$$NR(f) = \frac{S_{\zeta_1\zeta_1}(f)}{S_{\zeta\zeta}(f)}. \quad (1)$$

In this Equation, $S_{\zeta\zeta}(f)$ is the Power Spectrum Density (PSD) of $\zeta(n)$, and $\zeta(n)$ and $\zeta_1(n)$ are named according to Fig. 1. In the ideal three dimensional diffuse noise field, when adaptive filter converges to its optimum solution (Wiener solution), the noise reduction factor is obtained by Eq. (2) [2-4] where d is the microphone spacing and c is the sound velocity.

$$NR(f) = \frac{1}{1-\mathrm{sinc}^2(2fd/c)} \quad (2)$$

According to this Equation, the most important deficiency of ANCs is that they can only eliminate the noise components in low frequencies. By this explanation, and considering that most of the speech signal lays in the low frequencies, it seems that the post-filtering method can improve the ANC noise reduction ability efficiently.

[1] This work was supported by Iran Telecommunication Research Center (ITRC).

978-1-4244-2179-4/08/$25.00 ©2008 IEEE

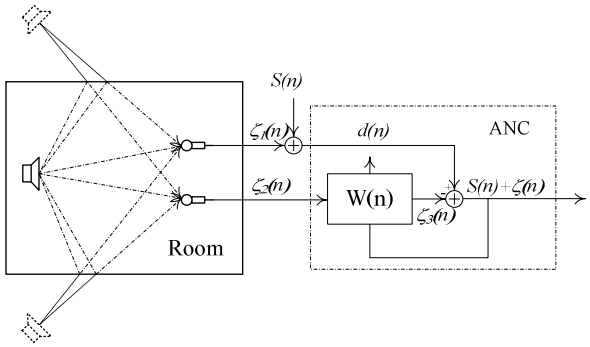


Fig. 1. ANC structure in real environment

3. LP-based speech enhancement problems

Fig. 2(a) shows the fundamental structure of linear predictors. Predictor coefficients, $W_{LP}(n)$, are determined in such a way that the prediction residual, $e(n)$, is minimized and whitened [8]. Therefore, white noise component, $\zeta(n)$, does not change the filter coefficients, and appears on $e(n)$ directly. On the other hand, if the number of filter taps is sufficient, i.e. 512 or more taps in 16 kHz sampling rate, adaptive filter becomes a comb filter for $s(n)$, and $\hat{s}(n)$ will be approximately equal to $s(n)$. This results a simple noise suppression. Following problems cause this structure to be inadequate for fruitful noise reduction:

1) The adaptive filter should be fast enough to track the statistical variations of speech signal.
2) The fundamental structure cannot estimate clean speech signal perfectly.
3) This method will encounter a difficulty if the additive noise is colored.

Long length adaptive filters can not converge rapidly to their optimum solution. To accelerate the convergence rate of adaptive filters, we have used this structure in the subband form, but subband implementation is not sufficient for solving the first problem. Because subband or fullband adaptive filters cannot track the variations of speech-like signals. authors in [7] suggested an exponential weighting coefficient which obtained by applying forgetting factor in adaptation rule.

Clean speech signal cannot be extracted completely even if fast adaptive filters are used in linear predictors (second problem). For example an unvoiced phoneme which is similar to noise is not predictable in this structure. Consequently, LP-based speech enhancement systems must make a compromise between the portion of remaining background noise and eliminating speech components. To manage this tradeoff we propose a multistage structure in Section 4.

4. A new multi-stage linear predictor for white noise suppression

In this section, we concentrate on the suppression of white noise by applying a multi-stage linear predictor as a solution for 2^{nd} problem. Fig. 2(b) shows the suggested structure with 4 stages. In this figure, each triangular is an amplifier with the gain of 0.5. Input and output of i^{th} stage are $d_i(n)$ and $d_{i+1}(n)$ respectively. Gradient descent adaptation algorithm for i^{th} stage is

$$W_{LPi}(n+1) = \lambda \cdot W_{LPi}(n) + \mu \cdot E\{D_1(n-1)e_i^*(n)\} \quad (3)$$

where $D_1(n-1)=[d_1(n-1),\ d_1(n-2),\ \ldots,\ d_1(n-N)]^T$. N and E denote filter length and expectation operator respectively. Each stage has the same reference input, $d_1(n-1)$, but its desired input, $d_i(n)$, is the output of the previous stage. The noise component of i^{th} stage reference input is $\zeta(n-1)$ and the noise component of the desired input is a factor of $\zeta(n)$. Due to the white noise assumption, these noise components are uncorrelated with each other. So, $\hat{s}_i(n)$ is the distorted speech which becomes more distorted as i increases. The output of i^{th} stage, $d_{i+1}(n)$, can be computed by the following equation:

$$d_{i+1}(n) = \frac{\zeta(n)}{2^i} + \frac{s(n)}{2^i} + \sum_{j=1}^{i} \frac{1}{2^{i-j+1}} \times \hat{s}_j(n) \cdot \quad (4)$$

Increasing the number of stages, results in more distorted speech signal and lower SNR. On the other hand, noise component power of $d_{i+1}(n)$ reduces exponentially; therefore it increases the SNR exponentially. Hence, the expected tradeoff can be controlled by the number of stages. Our experimental results show that increasing the number of stages more than 3 does not improve the speech quality.

The multistage method cannot solve the 3^{rd} problem. In the next section a solution based on a hybrid structure is proposed.

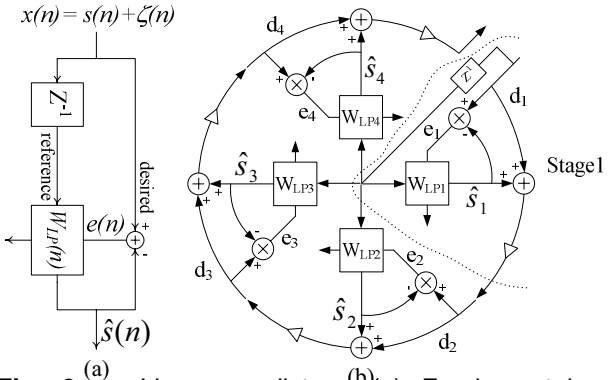


Fig. 2. Linear predictors, (a) Fundamental structure. (b) Multistage structure for speech enhancement based on linear prediction.

5. Proposed hybrid enhancement structure

A hybrid system is designed to overcome the deficiency of ANC and solve the 3^{rd} problem of fundamental structure mentioned in section 3. Fig. 3 shows the overall structure of this hybrid system. It is a combination of two stages; an ANC followed by Linear Prediction based Post-Filtering (LPPF) which is shown in Fig. 2(b). Input and output of ANC can be expressed by its speech and noise components as:

$$d^k(n) = s^k(n) + \zeta_1^k(n) \qquad k = 0,1,\cdots,K-1 \tag{5}$$

$$d_1^k(n) = s^k(n) + \zeta^k(n) \qquad k = 0,1,\cdots,K-1 \tag{6}$$

$$\zeta^k(n) = \zeta_1^k(n) - \zeta_3^k(n) \qquad k = 0,1,\cdots,K-1 \tag{7}$$

where k is the index of subbands. Therefore, the Eq. (3) in the previous section can be expressed for each subband as (8) where M is the number of stages.

$$W_{LPi}^k(n+1) = \lambda \cdot W_{LPi}^k(n) + \mu \cdot \nabla_i^k(W_{LPi}^k(n)) \quad i = 1,...,M \;\; k = 0,...,K-1$$

$$\nabla_i^k(W) = E\{D_1^k(n-1).[d_i^k(n) - W^H.D_1^k(n-1)]^*\} \tag{8}$$

Using Eq. (6), for the first stage gives

$$\nabla_1^k(W) = E\{D_1^k(n-1).[d_1^k(n) - W_{LP1}^k(n)^H.D_1^k(n-1)]^*\} = \tag{9}$$

$$E\{D_1^k(n-1).[s^k(n) - W_{LP1}^k(n)^H.D_1^k(n-1)]^* + Z^k(n-1).\zeta^k(n)^*\}$$

where $Z^k(n-1) = [\zeta^k(n-1), \zeta^k(n-2), \ldots, \zeta^k(n-N)]$.

Now, our goal is to determine $W_{LP1}{}^k(n)$ such that make $\hat{s}_1^k(n)$ equivalent to speech signal $s^k(n)$ (not $d_1{}^k(n)$). So the second term of Eq. (9) must be removed. Then the new gradient becomes

$$\nabla_1^k(W_{LP1}^k(n)) = E\{D_1^k(n-1).e_1^k(n)^* - Z^k(n-1).\zeta^k(n)^*\}. \tag{10}$$

Unfortunately, $\zeta(n)$ is unknown and should be determined. Combining (7) and (10) gives

$$\nabla_1^k(W_{LP1}^k(n)) = E\{D_1^k(n-1).e_1^k(n)^* - Z_1^k(n-1).\zeta^k(n)^* . \tag{11}$$

$$+ Z_3^k(n-1).\zeta^k(n)^*\}$$

When ANC converges to its optimum filter, the last term of Eq. (11) becomes zero according to the principle of orthogonality. Also, Eq. (7) gives

$$\nabla_1^k(W_{LP1}^k(n)) = E\{D_1^k(n-1).e_1^k(n)^* - Z_1^k(n-1).\zeta_1^k(n)^*$$

$$+ Z_1^k(n-1).\zeta_3^k(n)^*\} \quad . \tag{12}$$

On the other hand, $s(n)$ and $\zeta_3^k(n)$ are uncorrelated. So

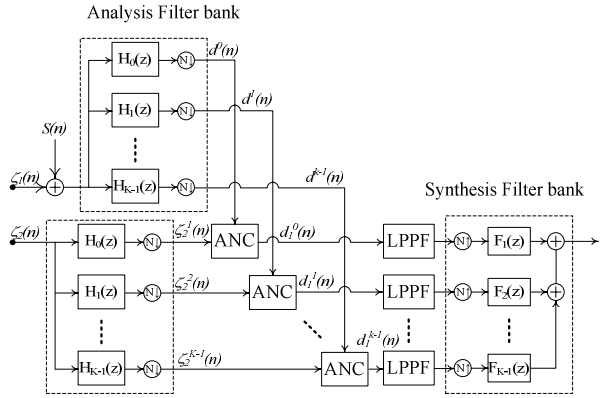


Fig. 3. Overall hybrid system

$$E\{Z_1^k(n-1).\zeta_3^k(n)^*\} = E\{D^k(n-1).\zeta_3^k(n)^*\} \tag{13}$$

and in the diffuse noise field we have

$$E\{Z_1^k(n-1).\zeta_1^k(n)^*\} = E\{Z_2^k(n-1).\zeta_2^k(n)^*\}. \tag{14}$$

Therefore, final updating algorithm is obtained as

$$W_{LP1}^k(n+1) = \lambda \cdot W_{LP1}^k(n) + \mu \cdot E\{D_1^k(n-1).e_1^k(n)^* - \tag{15}$$

$$Z_2^k(n-1).\zeta_2^k(n)^* + D^k(n-1).\zeta_3^k(n)^*\}$$

Removing expectation and approximating gradient with stochastic gradient gives the final practical adaptation rule for the first stage.

When ANC converges, for any kind of noise, $\hat{s}_1^k(n)$ approximates the speech signal. Therefore, output of first stage can be expressed as

$$d_2^k(n) = \frac{s^k(n) + \hat{s}_1^k(n)}{2} + \frac{\zeta^k(n)}{2} \tag{16}$$

In the previous section, Eq. (17) was used for computing the gradient of second stage. Similar to first stage, the last term of this Eq. should be eliminated. This modification results updating algorithm for second stage as Eq. (18).

$$\nabla_2^k(W_{LP2}^k(n)) = E\{D_1^k(n-1).[d_2^k(n) - W_{LP2}^k(n)^H.D_1^k(n-1)]^*\} \tag{17}$$

$$= E\{D_1^k(n-1).[\frac{s^k(n) + \hat{s}_1^k(n)}{2} - W_{LP2}^k(n)^H.D_1^k(n-1)]^*$$

$$+ \frac{1}{2} Z^k(n-1).\zeta^k(n)^*\}$$

$$W_{LP2}^k(n+1) = \lambda \cdot W_{LP2}^k(n) + \mu \cdot [D_1^k(n-1).e_2^k(n)^* - \tag{18}$$

$$0.5 \times Z_2^k(n-1).\zeta_2^k(n)^* + 0.5 \times D^k(n-1).\zeta_3^k(n)^*]$$

Generally, the updating equation for i^{th} stage is obtained by induction as

$$W_{LPi}^k(n+1) = \lambda \cdot W_{LPi}^k(n) + \mu \cdot [D_1^k(n-1).e_i^k(n)^* - \tag{19}$$

$$\frac{1}{2^{i-1}} \times Z_2^k(n-1).\zeta_2^k(n)^* + \frac{1}{2^{i-1}} \times D^k(n-1).\zeta_3^k(n)^*]$$

Applying these adaptation rules makes the hybrid structure robust to colored noise.

6. Prototype filters with different bandwidth for analysis and synthesis filterbank

We employ even DFT filterbank because of their simple implementation and to avoid aliasing over-sampling method is selected. [9, 10] reported that the signal attenuation in band edges of the analysis prototype filters reduces the convergence rate of the conventional oversampled SAFs. Moreover, the noise components of each subband are predictable even for white noise input. Because over-sampling limits the bandwidth of each subband signal. Fortunately, taking the prototype filter bandwidth of analysis wider than synthesis can solve these problems. Fig. 4 shows the selected prototype filters. Analysis prototype filters are designed to generate white output for white inputs in each subband. Synthesis prototype filters are designed to avoid aliasing. The prototype filters are selected for 16 subbands and the over sampling factor of 2.

7. Experimental results

In our experiments, we have simulated diffuse noise field by two noise signals recorded in the non-reverberant room where four loudspeakers generated four independent pink noises from various directions. Microphone spacing and sampling rate are *d=38 mm* and *16 kHz* respectively. Fig. 5 illustrates measured Magnitude Square Coherence (MSC) [2-4] for recorded noises (solid line). The curve is extremely close to the theoretical MSC of diffuse noise field (dashed line), so our noise field can be considered as a diffuse noise field. A test set includes 10 sentences (5 males and 5 females) from the FARSDAT database [11] is used as the speech material.

Fig. 6 shows the results of our tests with their spectrograms. LPF is the fundamental LP based filtering speech enhancement (Fig. 2(a)). ANCFPF is a combination of ANC followed by fundamental structure and ANCMPF-i is a combination of ANC and multistage structure (Fig. 2(b)) with i stages.

For objective evaluation of speech quality, we employed the Perceptual Evaluation of Speech Quality (PESQ) [1] and segmental SNR measures in *0*, *5* and *10 dB* input SNRs. Fig. 7(a) demonstrates the results of our evaluations with PESQ measure. For all input SNR levels, LPF and ANC provide the worst results and the proposed methods have improved the performance of these methods. In the case of lower input SNR levels ANCFPF which uses the fundamental structure as a post-filter is the most effective method. In this condition, the ANCMPF has lower performance because multistage structure needs a large number of stages to acceptably reduce the noise power. As the number of stages increase, the speech signal becomes more distorted. This proves the fundamental structure more efficient than multistage especially in low SNR conditions. In higher input SNR levels using two or three stages results in remarkable noise reduction and furthermore does not distort speech signal very much. So we suggest the fundamental structure for high noisy environments and multistage structure for mild situations as an appropriate post-filtering approach.

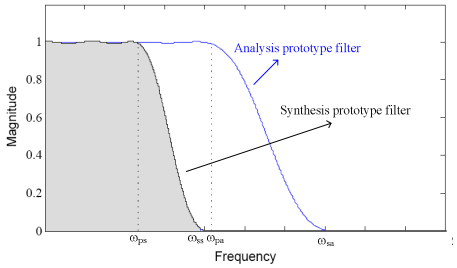


Fig. 4. Simulated analysis and synthesis prototype filters, with $\omega_{pa}=\pi/9-\pi/36$, $\omega_{ps}=\pi/16-\pi/64$ as passband and $\omega_{sa}=\pi/9+\pi/36$, $\omega_{ss}=\pi/16+\pi/64$ as stopband frequencies.

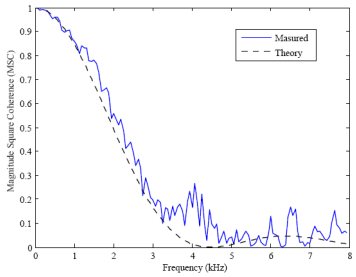


Fig. 5. Magnitude square coherence of the generated (solid line) and the theoretical (dashed line) diffuse field.

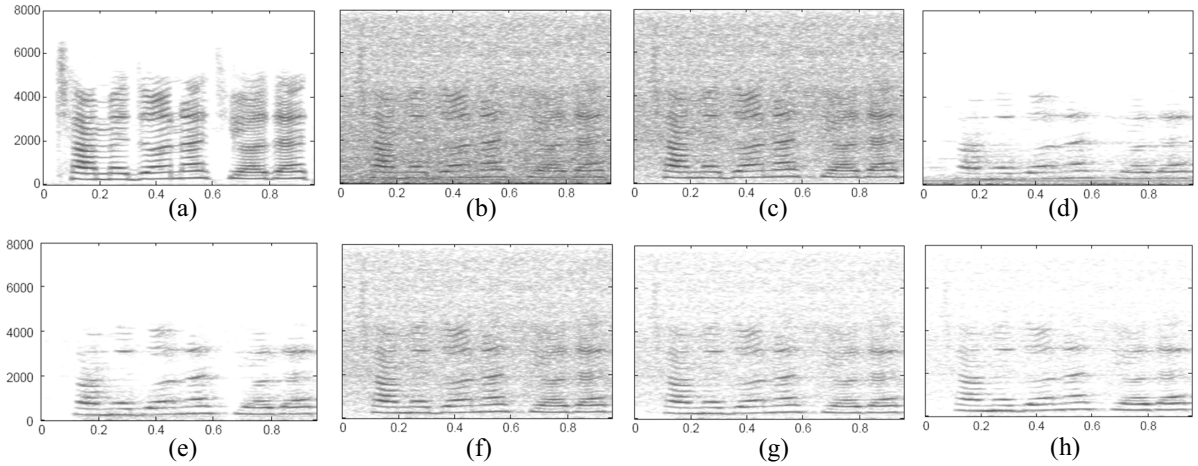
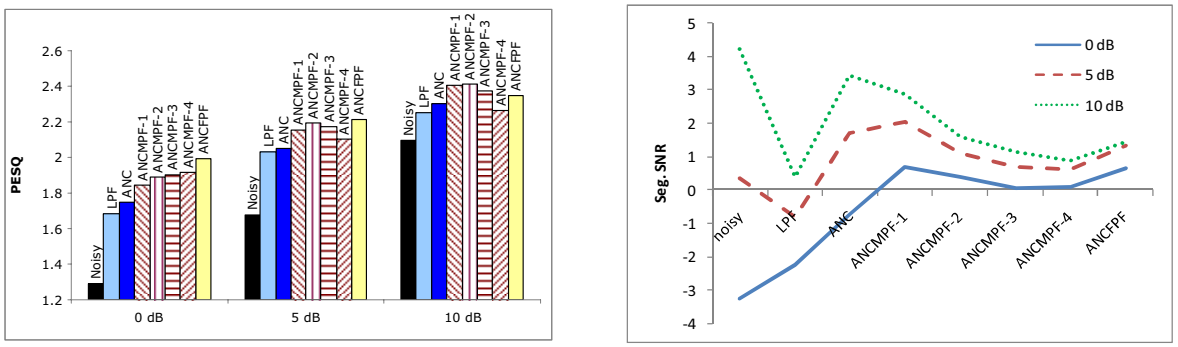


Fig. 6. Spectrograms of (a) Clean speech (b) 0 dB noisy speech signal and the outputs of (c) ANC (d) LPF (e) ANCFPF (f) ANCMPF-1 (g) ANCMPF-2 (h) ANCMPF-3



(a) (b)

Fig. 7. quality measure evaluation with (a) PESQ (b) Segmental SNR

Fig. 7(b) shows the evaluation results using segmental SNR. As it is shown, the proposed methods outperform the ANC and LPF in *0* and *5 dB* input SNRs, but the results of this measure are not consistent with PESQ and informal subjective quality assessment in the case of *10* dB input SNR. As described in [7], the LP-based enhancement filtering provides the improvement in the sense of SNR measure only in low SNR conditions.

8. Conclusion

Performance of ANCs degrades extremely in the diffuse noise fields. In this paper we have designed a set of post-filters to alleviate the challenge of ANC in diffuse noise fields. The adaptation algorithms for these post-filtering methods are modified such that they can eliminate colored noises too. Post-filtering with fundamental linear predictor is proposed for high noisy environments and multistage structure is suggested for mild situations. Proposed structures are implemented in the subband domain. To avoid aliasing problem we have used over-sampling technique and to prevent over-sampling problems we employed different analysis and synthesis prototype filters in our filterbanks. The evaluation results show the high performance of the proposed methods in comparison with standard ANC and LP based filtering.

Acknowledgment

The authors wish to thank Dr. Hamid Reza Abutalebi for his noisy data preparation and helpful discussions in the course of this work.

References

[1] Loizou P. C., Speech Enhancement Theory and Practice, CRC Press, 2007.

[2] Goulding M. M. and Bird J. S., “Speech enhancement for mobile telephony”, IEEE Trans. Vehicular Technology, vol. 39, no. 4, pp. 316-326, Nov. 1990.

[3] Sheikhzadeh H., Abutalebi H. R., Brennan R. L. and Freeman G. H., “Reduction of diffuse noise in mobile and vehicular applications”, chapter 10 of DSP for In-Vehicle and Mobile Systems, Springer, pp 153-168.

[4] Abutalebi H. R., Sheikhzadeh H., Brennan R. L. and Freeman G. H., “A hybrid subband adaptive system for speech enhancement in diffuse noise fields”, IEEE Sig. Proc. Letters, vol. 11, no. 1, Jan. 2004.

[5] Cohen I., and Berdigo B., “Microphone-array post-filtering for non-stationary noise suppression”, Proc. ICASSP, pp. 901-904., Orlando, Florida, May 13-17, 2002.

[6] Marro C., Mahieux Y., and Simmer K. U., ”Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering”, IEEE Tras. SAP, vol. 6, no.1, pp. 240-259, May 1998.

[7] Kawamura A., Fujii K., Itoh Y. and Fukui Y., “A noise reduction method based on linear prediction analysis”, Electronics and Communications in Japan, Part 3, vol.86, no. 3, pp. 1-10, Apr. 2003.

[8] Kawamura A., Iiguni Y. and Itoh Y., “A noise reduction method based on Linear prediction with variable step-size”, IEICE Trans. Fundamentals, vol. E88-A, no.4, Apr. 2005.

[9] Farhang –Boroujeny B., Adaptive filters theory and applications, John Wiley and Sons, New York, 1998.

[10] Farhang –Boroujeny B. and Wang Z., “Adaptive filtering in subband: design issues and experimental results for acoustic echo cancellation” Signal Processing, vol. 61, pp. 213-223, May 1997.

[11] Bijankhan, M., Sheikhzadegan, MJ “FARSDAT: The speech database of Farsi spoken language”, Proceedings of SST’94, Sydney, pp. 826-831, 1994.