AN OPTIMUM MMSE POST-FILTER FOR ADAPTIVE NOISE CANCELLATION IN AUTOMOBILE ENVIRONMENT

Soheil Khorram, Hossein Sameti, and 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 Cancellation (ANC) is an effective dual-channel technique for background noise reduction. Due to the presence of uncorrelated noise components at the two inputs in vehicular environments, ANC does not provide sufficient background noise reduction. To alleviate this problem, a complementary linear filter is added to ANC structure. Filter coefficients are determined to make the enhanced signal an MMSE estimation of speech signal. Therefore, the ANC structure is modified to a dual-channel Wiener structure. We prove that this structure is identical to the LMS type ANC which is followed by a Wiener post-filter. A new method is proposed for the noise spectrum estimation in the Wiener post-filter. This method does not require Voice Activity Detectors (VADs) and performs better speech enhancement in nonstationary noisy environments. Experimental results show that the proposed system can overcome the problem efficiently, at the cost of more complexity and more speech distortion.

1. INTRODUCTION

Performance of speech processing systems degrades severally in noisy and reverberant environments such as the interior of a moving vehicle [1]. Several speech enhancement systems in automobiles have employed Adaptive Noise Cancellation (ANC) technique to improve the quality of recorded speech [1-5]. In this technique, the noisy speech is recorded by the desired microphone placed in front of the speaker at the inside of the automobile cabin and the background noise is recorded by the reference microphone placed at the outside of the cabin. Since the background noise is generated outside and penetrates to the inside, high correlation between noise components of two inputs are expected [2]. ANC improves the speech quality by removing this correlation [2, 3].

However, since various sources like mechanical friction within the engine compartment, air flow, road noise, street noise and etc, generate automobile noise field through many paths, there is no such high correlation between the inside and outside noises in all frequencies and they are correlated only at lower frequencies [2]. Experiments have shown that the noise field is approximately diffuse [3]. In this condition, adaptive noise cancellers and other multi-microphone speech enhancement techniques cannot work efficiently [3].

Authors in [1] introduced this challenge and applied an array of N microphones as a solution. Kuo et al. [4] proposed a multiple reference ANC to compensate the effect of multi-source problem, but they did not consider the multi-path problem. Due to the reverberation, even if one source generates the noise field, there are some uncorrelated noise components at particular frequencies. In our previous works [6, 7], two Linear Prediction (LP)-based post-filters were designed which provide appropriate noise reduction, but their speech distortion are not acceptable.

In this paper, a dual-channel Wiener filter is selected. It is demonstrated that if the reference input does not contain speech component, i.e. crosstalk is negligible; this structure can be implemented by cascading an LMS type ANC and a single-channel Wiener filter. Therefore, ANC with Wiener post-filter provides an MMSE estimation of speech signal. Abutalebi et al. [3, 5] examined a similar structure in the subband form. They used a Voice Activity Detector (VAD) based method to control the noise spectrum estimation in the Wiener filter. Clearly, VAD based methods cannot efficiently suppress nonstationary noises such as automobile noises [8]. As a remedy, we propose an iterative method for noise spectrum estimation which uses the reference channel for predicting statistical variation of the noise component.

In Section 2, the performance of ANC in automobile is discussed. Section 3 proves the equality of Dual-channel Wiener filter and Wiener post-filtered ANC. Section 4 gives the proposed noise estimation method. Experimental results are presented in Section 5 and finally, we summarize our work and conclude in Section 6.

2. ANC METHOD IN AUTOMOBILE

Figure 1 shows the conventional ANC structure. The reference signal, $\zeta_2(n)$, passes through an adaptive filter and its result is subtracted from the desired signal, $d(n)$. Adaptive filter parameters are changed to optimize a cost function, e.g., minimizing the mean of the squared residual signal, $E\{e^2(n)\}$, as small as possible.

The reference input, $\zeta_2(n)$, and the speech component of the desired input, $s(n)$, are independent, therefore the adaptive filter output cannot eliminate the speech component and $s(n)$ appears directly in the ANC output. On the

978-1-4673-0382-8/12/$31.00 ©2012 IEEE

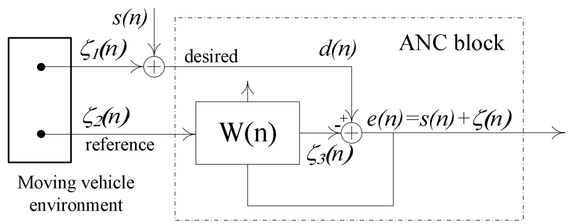


Figure 1 – Conventional ANC structure.

other hand, in an ideal non-reverberant noise field which is generated by single source, noise components of two channels are linearly correlated to each other and are eliminated from ANC output.

There are two main problems which affect the operation of conventional ANC in various applications.

1- Crosstalk problem: The presence of the speech component in the reference input degrades the performance of practical adaptive noise cancellers [9]. This problem is not considered in this paper.
2- Uncorrelated noise components: Due to the reverberation and the presence of the multiple noise sources, noise components are not linearly correlated in all frequencies. Therefore, a linear filter cannot compensate the effect of noise completely.

For analyzing the noise reduction ability of ANCs, several authors [1, 3, 5] have calculated Noise Reduction (NR) factor which is defined by

$$NR(f) = \frac{S_{\zeta_1\zeta_1}(f)}{S_{\zeta\zeta}(f)}, \tag{1}$$

where S denotes the cross power spectral density. $\zeta(n)$ and $\zeta_1(n)$ are the noise component of desired and reference inputs which are named according to figure 1.

There are Two categories of adaptation algorithms for determining adaptive filter (AF) coefficients, Least Mean Square (LMS) family and Higher Order Statistics (HOS) family. Following subsections discuss and calculate the performance of each family in the automobile noise field.

2.1 Performance of LMS Adaptive filters

In the LMS adaptive filtering, when AF converges to its optimum solution, NR factor can be written as [1, 3]

$$NR(f) = \frac{1}{1-\gamma^2_{\zeta_2\zeta_1}(f)}, \tag{2}$$

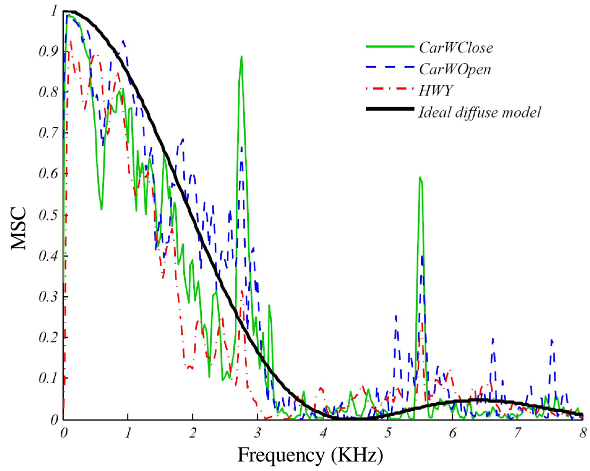


Figure 2 – MSC of the theoretical diffuse noise field (thick solid line) and three real-life automobile environments

where γ is the Magnitude Square Coherence (MSC) between the noise components of input signals and it is defined by the following equation [1].

$$\gamma^2_{\zeta_2\zeta_1}(f) = \frac{|S_{\zeta_2\zeta_1}(f)|^2}{S_{\zeta_2\zeta_2}(f)S_{\zeta_1\zeta_1}(f)}, \quad 0 \le \gamma^2_{\zeta_2\zeta_1}(f) \le 1 \tag{3}$$

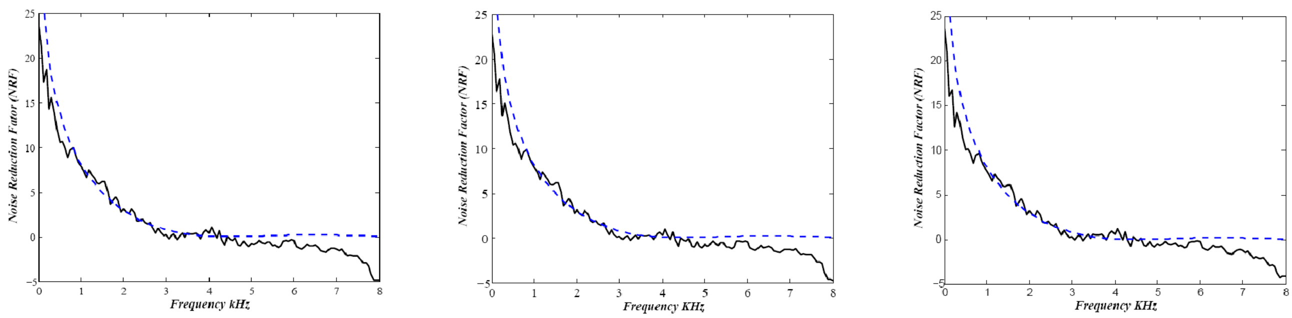
As a result of Eq. (2), LMS family can only eliminate the noise components with high MSC, i.e. correlated noise components.

Diffuse noise field is a good approximation of a number of practical noise fields such as the automobile noise field. The MSC of two signals, recorded in the ideal three-dimensional diffuse sound field, can be expressed as

$$\gamma^2_{\zeta_2\zeta_1}(f) = \frac{\sin^2\left(\frac{2\pi fd}{c}\right)}{\left(\frac{2\pi fd}{c}\right)^2} = sinc^2\left(\frac{2\pi fd}{c}\right), \tag{4}$$

where d is the microphone spacing and c is the sound velocity. MSC of the noise field in an automobile obey the diffuse noise field model closely [5]. The similarity of automobile noise filed to the defuse noise filed is demonstrated in figure 2. This figure shows the MSC of various automobile noise fields in comparison to the ideal diffuse noise field. Noise signals used in this figure were collected during the research reported in [3]. These signals were recorded in the following situations:

1- Inside a moving car with closed windows (CarWClose)
2- Inside a moving car with open windows (CarWOpen)



(a) (b) (c)

Figure 3 – NR factor of ANC that operates with two inputs recorded in automobile environment. Dashed line is the theoretical NR factor of LMS in diffuse noise field and solid lines are the NR factor of (a) LMS, (b) LMK and (c) ICA algorithms. Microphone spacing is *38 mm*.

3- Inside a working car parked next to a high way (HWY)

According to figure 2, the most important problem of LMS type ANC is that it can only eliminate the noise components in low frequencies.

2.2 Performance of HOS Adaptive filters

Least Mean Forth (LMF) [10], Least Mean Kurtosis (LMK) [10] and ICA-based AFs [11] are three examples of HOS adaptation algorithms. It is complicated to obtain a closed-formula for the NR factor of HOS adaptive filters. To evaluate their performances, an experiment is initiated using each AFs and NR factor is calculated from Eq. 1 directly. The results of our experiments are shown in figure 3 which demonstrate that HOS family does not improve the NR factor of LMS algorithm. Therefore, HOS and LMS adaptive filters cannot cope with the problem of uncorrelated noise components.

As a result of this section, the deficiency of ANC is not related to its adaptation algorithm, but its structure which uses a linear adaptive filter is imperfect to cope with the problem of uncorrelated noise components. From another point of view, we know that the MSC between the input and output of a linear filter is one in all frequencies. It means that if the amount of MSC between two signals is much smaller than one (uncorrelated signals), we cannot generate one of them by filtering the other. Therefore, ANC structure with a single linear AF, independent of its adaptation rule, is not sufficient for removing the uncorrelated noise components.

In this paper, a linear filter is added to the front of the desired channel of conventional ANC to overcome the problem. Both filter coefficients are determined such that the system output provides an MMSE estimation of clean speech. In this way, the structure of ANC is converted to a dual-channel wiener filter.

3. EQUALITY OF DUAL-CHANNEL WIENER FILTER AND WIENER POST-FILTERED ANC

Recently, multi-channel Wiener filtering technique has been proposed [12]. The structure of this technique which is depicted in figure 4(a) provides an MMSE estimation of the clean speech [12]. In this section, first, the optimum expressions for $W_1(f)$ and $W_2(f)$ in figure 4(a) are obtained. Then, we prove that by considering the ANC conditions this structure is identical to the cascade of LMS type ANC and a single channel Wiener filters when ANC converges to its optimum solution. It should be noted that all signals are named according to the figure 4.

For the optimum value of $W_1(f)$ and $W_2(f)$ we have

$$W_1(f), W_2(f) = \operatorname{argmin}_{W_1,W_2} E\{[s(n) - \hat{s}(n)]^2\}\,. \tag{5}$$

where $E\{.\}$ denotes the expectation operator and $\hat{s}(n)$, is obtained by Eq. (6).

$$\hat{s}(n) = \sum_{l=-\infty}^{\infty} w_1(l)d(n-l) + w_2(l)z_2(n-l) \tag{6}$$

Using the principle of orthogonality [12] gives

$$E\{(s(n) - \hat{s}(n)).z_2(n-i)\} = 0 \quad for\ i = 0, \pm 1, \ldots \tag{7}$$

$$E\{(s(n) - \hat{s}(n)).d(n-i)\} = 0 \quad for\ i = 0, \pm 1, \ldots \tag{8}$$

Replacing $\hat{s}(n)$ by using Eq. (6) and taking the Fourier transform results in a system of two equations with two variables. Solving this equation system gives the final expression for $W_1(f)$ and $W_2(f)$ as follows.

$$W_1(f) = \frac{S_{sd}(f)S_{z_2z_2}(f) - S_{sz_2}(f)S_{z_2d}(f)}{S_{z_2z_2}(f)S_{dd}(f) - S_{z_2d}(f)S_{dz_2}(f)} \tag{9}$$

$$W_2(f) = \frac{S_{sd}(f)S_{dz_2}(f) - S_{sz_2}(f)S_{dd}(f)}{S_{z_2d}(f)S_{dz_2}(f) - S_{z_2z_2}(f)S_{dd}(f)} \tag{10}$$

These equations can be more simplified by considering this fact that the speech component, $s(n)$, is uncorrelated with the reference noise, $z_2(n)$.

$$W_1(f) = \frac{S_{sd}(f)S_{z_2z_2}(f)}{S_{z_2z_2}(f)S_{dd}(f) - S_{z_2d}(f)S_{dz_2}(f)} \tag{11}$$

$$W_2(f) = \frac{S_{sd}(f)S_{dz_2}(f)}{S_{z_2d}(f)S_{dz_2}(f) - S_{z_2z_2}(f)S_{dd}(f)} \tag{12}$$

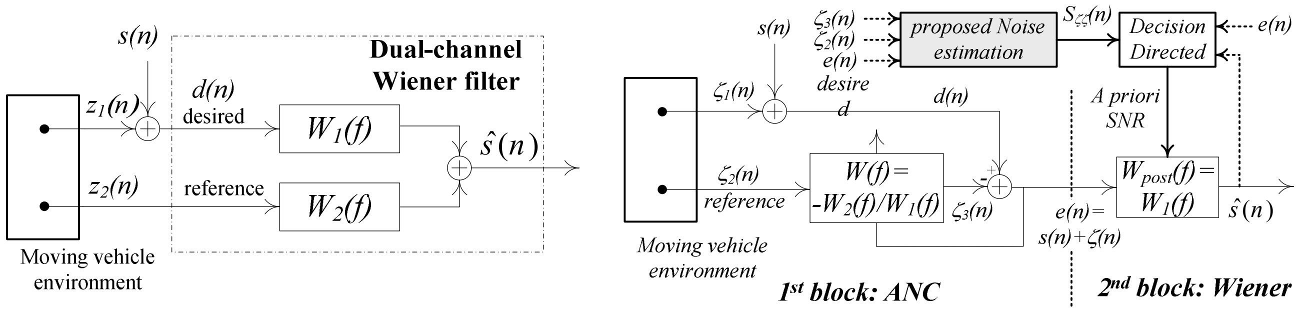
Figure 4(b) shows a structure which is identical to the dual-channel Wiener filter of figure 4(a), but with a view toward postfiltering. Our goal is to prove that the first and second blocks of figure 4(b) can be implemented by an ANC and a Wiener filter respectively.

Dividing Eq. (11) by Eq. (12) results in W(f) according to Eq. (13) which is equal to the optimum solution of LMS type ANC. It is proved that the first block is nothing but an adaptive filter with LMS adaptation algorithm.

$$W(f) = -\frac{W_2(f)}{W_1(f)} = \frac{S_{dz_2}(f)}{S_{z_2z_2}(f)} \tag{13}$$

Now, we want to prove that the second block i.e. $W_1(f)$ is a single channel Wiener filter. From Eq. (11) we have

$$W_1(f) = \frac{S_{ss}(f)}{S_{ss}(f) + S_{z_1z_1}(f) - S_{z_2z_1}(f)S_{z_1z_2}(f)/S_{z_2z_2}(f)}\,. \tag{14}$$



(a) (b)

Figure 4 – Illustration of (a) dual-channel Wiener structure and (b) its identical structure with ANC and post-filter

Replacing $z_1(n)$ by the sum of $z(n)$ and $z_3(n)$, and considering that $z(n)$ and $z_2(n)$ are orthogonal to each other when ANC converges, results in a final expression of $W_1(f)$ as

$$W_1(f) = W_{post}(f) = \frac{S_{ss}(f)}{S_{ss}(f)+S_{zz}(f)}, \qquad (15)$$

Equations (13) and (15) show that an optimum ANC followed by a Wiener post-filter is identical to the dual channel Wiener structure and provide an optimum MMSE estimation of clean speech signal.

4. PROPOSED NOISE SPECTRUM ESTIMATION FOR WIENER POST-FILTER

Noise spectrum estimation is critical for the performance of Wiener filter and other single channel speech enhancement methods. This section focuses on this topic. Automobile noise field is non-stationary; therefore a new method is designed to track the variations of non-stationary noise signals. This method employs the information gathered from the first block of the system.

Wiener filter defined by Eq. (15) is a non-causal filter and is not realizable, because $S_{ss}(f)$ and $S_{zz}(f)$ are unknown [8]. By defining a priori SNR

$$\zeta(f) = \frac{S_{ss}(f)}{S_{zz}(f)}, \qquad (16)$$

the Wiener filter expression of Eq. (15) can be expressed as

$$W_1(f) = \frac{\zeta(f)}{\zeta(f)+1}. \qquad (17)$$

Several methods were proposed for estimating a priori SNR [8]. The majority of these methods are extensions to the decision-directed method proposed by Ephraim and Malah [14]. Noise spectrum estimation is one of the most critical requirements for estimating a priori SNR [8]. In this section, a new algorithm for calculating $S_{zz}(f)$ in our system is proposed. When ANC converges, we have [13]

$$S_{zz}(f) = S_{z_1z_1}(f) - S_{z_3z_3}(f). \qquad (18)$$

The second term of this equation is known and can be obtained from the reference signal directly. We have to obtain the first term. Characteristics of various points in an ideal diffuse noise field are identical, hence by the diffuse model assumption of automobile noise fields, $S_{z_1z_1}(f)$ is equal to $S_{z_2z_2}(f)$ [1]. Our experiments reveal that this assumption is true approximately when the distance between two microphones is low. We consider a more general case that there is a gain factor between the power spectrums of two channels. This gain factor changes slowly in all frequencies. The assumption can be expressed by the following equation.

$$S_{z_1z_1}(f,k) = G(f,k).S_{z_2z_2}(f,k) \qquad (19)$$

Where $G(f,k)$ is the gain factor in frequency f and k^{th} time frame. This factor can be computed by low-pass filtering of its previous values. Eq. (20) shows a one pole low pass filtering which is used in our experiments.

$$G(f,k) = \beta\, G(f,k-1) + (1-\beta)\,\frac{S_{z_1z_1}(f,k-1)}{S_{z_2z_2}(f,k-1)} \qquad (20)$$

Post-filter algorithm

Inputs: $z_2(n)$, $z_3(n)$, $e(n)$
Output: $\hat{S}(n)$
Framing the inputs with overlap;
Initialization;
for *k=1 to number of frames*
 Taking the Fourier transform of inputs and compute the spectrum of them.
 for *f in all frequency bins*
  $G(f,k) = \beta\, G(f,k-1) + (1-\beta)\,\frac{S_{z_3z_3}(f,k-1)+(1-W_{post}(f,k-1))S_{ee}(f,k-1)}{S_{z_2z_2}(f,k-1)}$;
  $S_{zz}(f,k) = G(f,k).S_{z_2z_2}(f,k) - S_{z_3z_3}(f,k)$;
  $\zeta(f,k) = \alpha\frac{\hat{S}(f,k-1)}{S_{zz}(f,k-1)} + (1-\alpha)P\left\{\frac{|E(f,k)|^2}{S_{zz}(f,k)} - 1\right\}$;
  / P is an operator witch replace the nonnegative numbers by zero.* $E(f,k)$ *is the* k^{th} *frame of* $e(n)$ *Fourier transform.*/*
  $W_{post}(f,k) = \frac{\zeta(f,k)}{\zeta(f,k)+1}$;
  $\hat{S}(f,k) = W_{post}(f,k).E(f,k)$;
 end
end
Calculating enhanced speech by taking the inverse Fourier transform of $\hat{S}(f,k)$ *and adding the overlapped version of the output of transform;*

figure 5 – Pseudo code of the proposed post-filter algorithm

In this equation, $S_{z_2z_2}(f,k-1)$ is known and $S_{z_1z_1}(f,k-1)$ can be computed easily as follow by using enhanced signals instead of clean speech.

$$S_{z_1z_1}(f,k-1) = S_{z_3z_3}(f,k-1) + (1-W_1(f,k-1))S_{ee}(f,k-1) \qquad (21)$$

Therefore, in each iteration, first the gain factor $G(f,k)$ is computed by Eq. (22), then Eq. (23) estimates the noise spectrum by using this gain factor.

$$G(f,k) = \beta\, G(f,k-1) + (1-\beta)\,\frac{S_{z_3z_3}(f,k-1)+(1-W_1(f,k-1))S_{ee}(f,k-1)}{S_{z_2z_2}(f,k-1)} \qquad (22)$$

$$S_{zz}(f,k) = G(f,k).S_{z_2z_2}(f,k) - S_{z_3z_3}(f) \qquad (23)$$

The enhancement algorithm of proposed post-filter is completely described in figure 5.

5. EXPERIMENTS

Our speech test set includes a set of 10 sentences (5 males and 5 females) from the FARSDAT database [15]. 3 sets of dual channel noise signals recorded in the real-life automobile environment are used as the noise materials which introduced in section 2. The sampling rate and microphone spacing are 16 KHz and 38 mm (a typical value for boomless headset), respectively.

ANC was parameterized using 512 taps and Least Mean Square (LMS) adaptation algorithm with constant step-size set to 0.01. The Wiener filter is implemented according to the algorithm illustrated in figure 5.

For objective assessment of speech quality, we employed the Perceptual Evaluation of Speech Quality (PESQ) [8] and Frequency-Weighted segmental SNR (FW. SNR) measures in 0, 5 and 10 dB input SNRs. High correlations with subjective listening tests were reported using the PESQ score [8]. The FW. SNR measure of [16] is implemented with different weighting function.

Figure 6 demonstrates the results of our evaluations. VAD_ANCWF and DNE_ANCWF are combinations of ANC followed by Wiener post-filter with VAD based and proposed Dual-channel Noise Estimation (DNE) methods respectively. As it is shown, the proposed method outperforms other methods especially in CarWOpen environment since in this environment statistical variation of background noise is more than the other. On the other hand, CarWClose noise field is more stationary, therefore the results of proposed method is descending to the VAD_ANCWF level.

6. CONCLUSION

The uncorrelated noise components presented in automobile environment make the adaptive noise cancellers inefficient for background noise reduction. In this paper, we apply a dual-channel Wiener structure as a solution for the uncorrelated noise conditions. This structure is implemented by combining ANC and Wiener post-filter. Also, new noise estimation method is proposed for the Wiener post-filter.

The experimental results show that the proposed system improves the performance of previous systems. Also, suggested noise estimation method increases the average value of PESQ and FW. SNR by 1.2 and 0.5 respectively.

7. REFERENCES

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

[2] Goubran R. A., Hafez H. M., “Background acoustic noise reduction in mobile telephony”, IEEE Vehicular Technology Conference, vol. 36, pp. 72-76, May 1986.

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

[4] Kuo S. M., Kunduru J. “Multiple reference subband adaptive noise canceler for hands-free cellular phone applications”, Journal of the Franklin Institute, published by Elsevier, vol. 333(B), no. 5, pp. 669-686, 1996.

[5] Sheikhzadeh H., Abutalebi H. R., Brennan R. L., Freeman G. H., “Reduction of diffuse noise in mobile and vehicular applications”, chapter 10 of DSP for In-Vehicle and Mobile Systems, published by Springer US, pp. 153-168, 2005.

[6] Khorram S., Sameti H., Veisi H., “LP-Based Over-Sampled Subband Adaptive Noise Canceller for Speech Enhancement in Diffuse Noise Fields”, ICSP, Beijing, Oct.08.

[7] Khorram S., Sameti H., Veisi H., Abutalebi H. R., “Lattice LP-based Post-filter for Adaptive Noise Cancellers in Mobile and Vehicular Applications”, ISSPIT, Sarajevo, Dec. 2008.

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

[9] Madhavan G., Bruin H., “Crosstalk Resistant Adaptive Noise Cancellation”, Annals of Biomedical Engineering, Springer, vol. 18, no. 1, Jan. 1990.

[10] Tanrikulu O., Constantinides A. G., “Least-mean kurtosis: a novel higher-order statistics based adaptivefiltering algorithm”, IEEE Transaction on Electronics Letters, vol. 30, no. 3, pp. 189-160, Feb. 1994.

[11] Yang J. M., Sakai H., "A new adaptive filter algorithm for system identification using independent component analysis", ICASSP, Honolulu, Hawai'i, U.S.A., vol. 3, pp. 1341-1344, Apr. 2007.

[12] Doclo S., Marc M., “GSVD-based optimal filtering for single and multi-microphone speech enhancement”, IEEE Trans. on signal processing, vol. 50, no. 9, pp. 2230-2244, Sept. 2002.

[13] Farhang –Boroujeny B., Adaptive Filters Theory and Applications, John Wiley and Sons, New York, 1998.

[14] Ephraim Y., Malah D., “Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator”, IEEE Trans. on ASSP, vol. 32, no. 6, pp. 1109-1121, Dec. 1984.

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

[16] Tribolet J., Noll P., McDermott B., Crochiere R. E., “A study of complexity and quality of speech waveform coders.” ICASSP, pp. 586-590, 1978.