THE INCORPORATION OF MASKING THRESHOLD TO SUBSPACE SPEECH ENHANCEMENT

Jong Uk Kim, Sang G. Kim and Chang D. Yoo

Department of Electrical Engineering and Computer Science
Korea Advanced Institute of Science and Technology
373-1, Guseong-dong, Yuseong-gu, Daejon, Republic of Korea, 305-701
oribros@mail.kaist.ac.kr, zom@eeinfo.kaist.ac.kr, cdyoo@ee.kaist.ac.kr

ABSTRACT

A subspace enhancement algorithm using the masking property is proposed. The proposed algorithm minimizes the signal distortion while constraining the energy of the residual noise below the human psychoacoustic masking threshold. This requires simple transformation of the masking threshold from the Fourier domain to the Karhunen-Loève (KL) domain. The proposed method incorporates subband whitening filters in the KL domain in order to deal with colored noise. Performance test results show that the proposed algorithm is superior to the spectral subtraction and the subspace method suggested by Y.Ephraim et. al.

1. INTRODUCTION

Various methods have been proposed in the past to enhance speech degraded by additive noise. Most of them are implemented in the Fourier domain; however, a few methods such as the subspace method suggested by Y.Ephraim [1] are implemented in the KL domain, where the eigenvectors of the covariance matrix of a given signal form the basis. These eigenvectors are considered to be optimal in terms of energy compaction, and for this reason subspace method has been known to perform better than the Fourier domain based approach, such as spectral subtraction and the Wiener filter method.

Although subspace methods have been successful in the area of speech enhancement, psychoacoustic properties of the human auditory system have not been fully exploited in the subspace method. For this reason there have been a number of efforts to raise the performance of the subspace method by incorporating the masking threshold [2, 3]. However, these algorithms are found to be computationally inefficient and their usage of the masking threshold are found to be suboptimal. For these reasons, their performances leave much to be desired.

This work was supported by grant No. R01-2000-000-00259-0(2002) from the Korea Science & Engineering Foundation.

In speech enhancement, both noise reduction and signal distortion must be considered simultaneously, since both go hand-in-hand. In the proposed method, the masking threshold acts as a guideline as to how much noise should be reduced. Following this guideline, we can keep the signal distortion to its minimum, since reducing more noise than necessary (below masking threshold) introduces unnecessary signal distortion.

The organization of the paper is as follows. Section 2 briefly summarizes the subspace method. Section 3 presents the proposed subspace method incorporating the masking threshold. Section 4 summarizes the subband whitening method to deal with colored noise. Section 5 and 6 present the overall system and performance evaluation respectively. Section 7 concludes.

2. SUBSPACE SPEECH ENHANCEMENT

Assume that the clean speech y is contaminated by additive white noise w to give noisy speech z such that

$$z = y + w. \tag{1}$$

Clean speech y can be estimated using a linear estimator H so that the estimated clean speech is given by $\hat{y} = Hz$. Using H, the signal distortion r_y and residual noise r_w are given as $r_y = (H - I)y$ and $r_w = Hw$ respectively. Denoting the signal distortion energy by $\epsilon_y^2 = \mathrm{tr}E\{r_y r_y^{\#}\}$ and the eigenvector matrix of covariance matrix of y by $U = [u_1, u_2, \cdots, u_K]$, the spectral domain constrained (SDC) estimator H is obtained by

$$H = \arg\min_{H} \epsilon_y^2 \tag{2}$$

with constraint

$$\begin{array}{ll} E\{|u_k^{\#} r_w|^2\} \le \alpha_k \sigma^2 & k = 1, \cdots, M, \\ E\{|u_k^{\#} r_w|^2\} = 0 & k = M+1, \cdots, K, \end{array} \tag{3}$$

where σ^2 and M are noise variance and the dimension of signal subspace respectively. For notational convenience,

let the kth eigenvalue $\lambda_y(k)$ associated with the kth eigenvector u_k be arranged in descending order for $k = 1, \ldots, K$, that is $\lambda_y(1) \geq \lambda_y(2) \geq \cdots \geq \lambda_y(K)$. The estimator thus obtained is given by

$$H = UQU^{\#}, \tag{4}$$

where Q is a diagonal matrix with the kth diagonal element q_k given as the generalized Wiener filter of the form

$$q_k = \begin{cases} \alpha_k^{1/2} & k = 1, \cdots, M \\ 0 & k = M+1, \cdots, K, \end{cases} \tag{5}$$

and

$$\alpha_k = \exp\{-\nu\sigma^2/\lambda_y(k)\}. \tag{6}$$

3. THE USE OF MASKING THRESHOLD TO SUBSPACE METHOD

The performance of the subspace enhancement algorithm is shown to be superior to that of spectral subtraction. Furthermore, the subspace based methods do not suffer from musical tones, as does the spectral subtraction. However the subspace method suggested by Y.Ephraim can lead to the following undesirable results. First, it is possible to suppress noise far below the masking threshold and introduce excessive signal distortion. Second, if an insufficient amount of noise is removed so that the residual noise energy lies far above the masking threshold, the residual noise can be audible. The best approach is to minimize the signal distortion while constraining the energy of the residual noise so that it is just below the masking threshold. In this section, the masking threshold is incorporated in the subspace method.

3.1. Masking Threshold Calculation

Masking threshold is the upper bound below which the human ear cannot perceive the presence of noise or any other sound. The calculation of the masking threshold is well summarized in [4]. The steps for obtaining the masking threshold are as follows :

1. *critical band analysis* : summing up the power spectrum in each critical band (Bark), where the power spectrum is obtained by magnitude squaring the FFT coefficient.
2. *spreading* : convolving with a spreading function to take into account the effect of adjacent critical band.
3. *offset* : subtracting the offset by considering the tone-like or noiselike nature of the speech.
4. *renormalization* : converting the spread spectrum back to Bark domain.
5. *absolute threshold* : comparing with absolute threshold and choosing maximum between them.

3.2. Masking Threshold Conversion

The masking threshold obtained by the above procedure leads to values in the Fourier domain. So it is necessary to convert the threshold of the Fourier domain into that of the KL domain. Let K-dimensional vector y be expressed by the Fourier domain representation

$$y = F^{-1}c \tag{7}$$

or by the KL domain representation

$$y = U\lambda, \tag{8}$$

where F, c, $U^{\#}$ and λ represent the Fourier transform matrix, the Fourier coefficient, the KL transform matrix, and the KL coefficient respectively. Equating (7) and (8), we get

$$\lambda = U^{\#}F^{-1}c, \tag{9}$$

which is the transformation from the Fourier domain to the KL domain. Now, obtain c_k by $m_k^{1/2}e^{-j\theta_k}$, where m_k (> 0) is the kth masking threshold in the Fourier domain, and θ_k is the kth phase component of the signal. Squaring each component of the vector λ, we finally obtain the masking threshold of the KL domain $\eta = [\eta_1, \cdots, \eta_K]^{\#}$. The use of θ_k is based on the fact that human ear is not insensitive to the phase, so the best phase can be obtained from the noisy speech signal itself. Henceforth, when we say masking threshold, it refers to the value of the KL domain.

3.3. The Use of Masking Threshold in Subspace Method

Considering the masking threshold, now the SDC estimator H is obtained by

$$H = \arg\min_{H} \epsilon_y^2 \tag{10}$$

with constraint

$$\begin{array}{ll} E\{|u_k^{\#}r_w|^2\} \leq \alpha_k\eta_k & k = 1, \cdots, M, \\ E\{|u_k^{\#}r_w|^2\} = 0 & k = M+1, \cdots, K, \end{array} \tag{11}$$

where η_k is the masking threshold associated with the kth eigenvector u_k. The linear estimator thus obtained satisfies

$$(I - Q)\Lambda_y - \sigma^2\Lambda_\mu Q = 0, \tag{12}$$

where $H = UQU^{\#}$, $Q = \text{diag}\{q_1, \cdots, q_K\}$ and $\Lambda_\mu = \text{diag}\{\mu_1, \cdots, \mu_K\}$ with μ_k being the Lagrangian multiplier. One possible solution of (12) is

$$q_k = \begin{cases} \frac{\lambda_y(k)}{\lambda_y(k)+\sigma^2\mu_k} & k = 1, \cdots, M \\ 0 & k = M+1, \cdots, K. \end{cases} \tag{13}$$

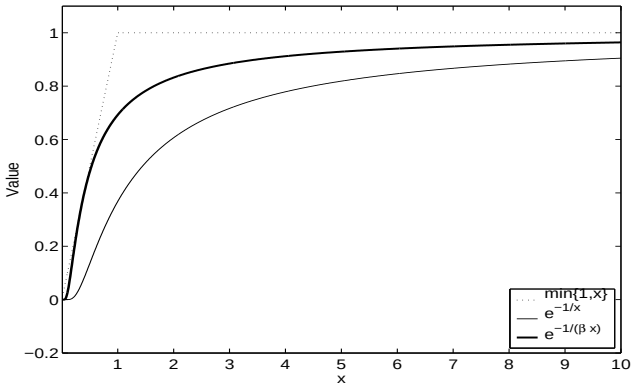


Fig. 1. Generalization of masking threshold constraint. The value $\beta = \exp(1)$ is used to satisfy marginal condition.

At the same time, from the constraint (11),

$$E\{|u_k^{\#} r_w|^2\} = q_k^2 \sigma^2 = \begin{cases} \alpha_k \eta_k & k = 1, \cdots, M \\ 0 & k = M+1, \cdots, K \end{cases} \quad (14)$$

or

$$q_k^2 = \begin{cases} \alpha_k \eta_k / \sigma^2 & k = 1, \cdots, M \\ 0 & k = M+1, \cdots, K \end{cases} \quad (15)$$

must hold at the boundary. To satisfy $0 \le q_k \le 1$, (15) becomes

$$q_k^2 = \begin{cases} \alpha_k \min\{1, \eta_k/\sigma^2\} & k = 1, \cdots, M \\ 0 & k = M+1, \cdots, K, \end{cases} \quad (16)$$

where $\alpha_k = \exp\{-\nu\sigma^2/\lambda_y(k)\}$ is generalized Wiener filter. For aggressive noise suppression, $\min\{1, \eta_k/\sigma^2\}$ is generalized to the exponential function as graphically explained in Fig. 1. Consequently we get the final gain function

$$q_k = \begin{cases} \exp\left(-\frac{1}{2}\frac{(1+\nu)\sigma^2}{\beta\eta_k + \lambda_k}\right) & k = 1, \cdots, M \\ 0 & k = M+1, \cdots, K, \end{cases} \quad (17)$$

with $0 \le \beta \le \exp(1)$. It should be noted that (17) becomes subspace method when $\eta_k = \infty$ for $k = 1, \cdots, M$.

4. SUBBAND WHITENING

To reduce signal distortion and computational complexity while improving the spectral resolution, subband whitening filter [5] is used. Define the nth whitening and inverse whitening filter as

$$W_n = \Sigma_n^{-1} U_n^{\#}, \quad n = 1, 2, \cdots, N, \quad (18)$$
$$W_n^* = U_n \Sigma_n, \quad n = 1, 2, \cdots, N, \quad (19)$$

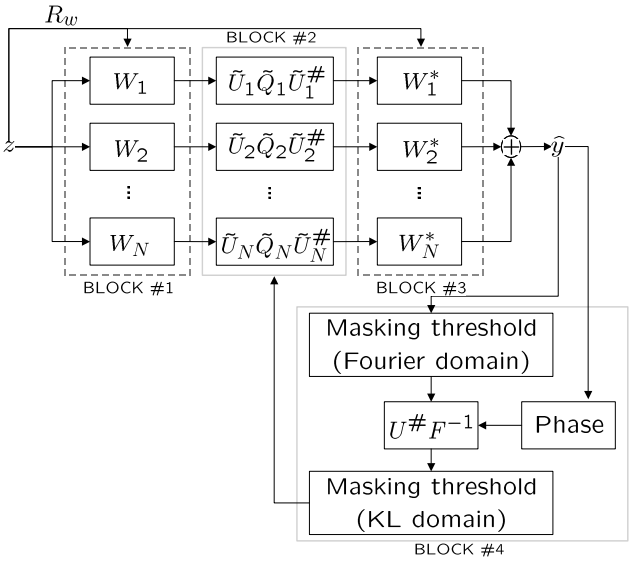


Fig. 2. Overall system. BLOCK #1 : subband whitening, BLOCK #2 : subspace enhancement, BLOCK #3 : inverse whitening, BLOCK #4 : masking threshold calculation.

where the columns of U_n are K eigenvectors of noise covariance matrix, and the diagonal elements of Σ_n^2 are K eigenvalues of noise covariance matrix. The size of both $U_w = [U_1, U_2, \cdots, U_N]$ and $\Sigma_w^2 = \text{diag}\{\Sigma_1^2, \Sigma_2^2, \cdots, \Sigma_N^2\}$ is $(NK) \times (NK)$. The output of nth subband whitening filter W_n is

$$\tilde{z}_n = W_n z = W_n y + W_n w \triangleq \tilde{y}_n + \tilde{w}_n, \quad (20)$$

On the other hand, the output of fullband whitening filter is

$$\tilde{z} = R_w^{-1/2} z = R_w^{-1/2} y + R_w^{-1/2} w \triangleq \tilde{y} + \tilde{w}, \quad (21)$$

where R_w is noise covariance matrix. Denoting the fullband and nth subband signal distortion by ϵ_y^2 and $\epsilon_{y_n}^2$ respectively, both quantities are upper-bounded as $\epsilon_y^2 \le \gamma$ and $\epsilon_{y_n}^2 \le \gamma_n$. It is derived [5] that

$$\sum_{n=1}^{N} \gamma_n \le \gamma, \quad (22)$$

which means that the upper bound on total signal distortion is smaller in subband than in fullband structure. The reduction of upper bound results in the reduction of overall signal distortion.

5. OVERALL SYSTEM

The overall system is shown in Fig 2. As shown, the system is comprised of four blocks. BLOCK #1 is a subband whitening block that projects the observation z onto each

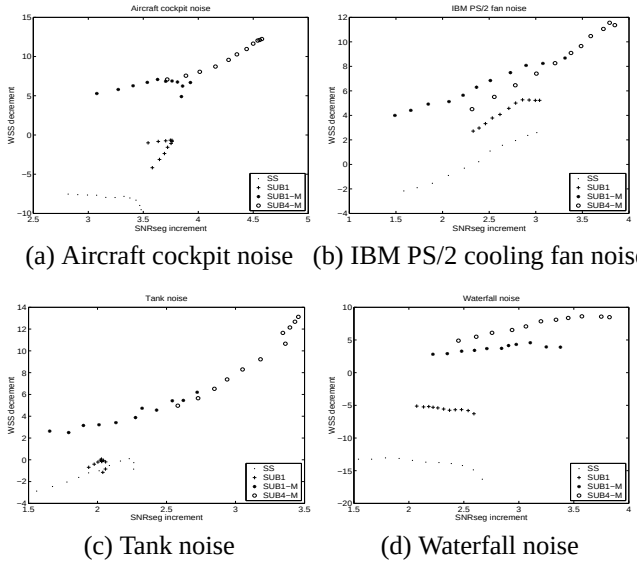
subspace and normalizes the energy. BLOCK #2 is an enhancement block. In applying the subspace enhancement method, we make use of the masking threshold to improve the performance. In order to calculate the masking threshold, an initial estimate of the clean speech $\hat{y}$ is necessary and this is done by making a rough estimate of clean speech. This is shown in BLOCK #4. BLOCK #3 is an inverse whitening block. In this block each subband estimates are inversely filtered and summed together to get the final estimate $\hat{y}$.

6. PERFORMANCE EVALUATION

Segmental signal-to-noise ratio (SNRseg) and weighted spectral slope (WSS) are used in the evaluation. The definitions of the two measures are well summarized in [6]. Performance results based on SNRseg increment and WSS decrement for 4 types of noise (aircraft cockpit noise, IBM PS/2 cooling fan noise, tank noise and waterfall noise) are given in Fig. 3. Each dot in the figure is associated with an input SNR ranging from 0dB to 20dB. The clean speech data are comprised of 10 Korean people's (5 women's and 5 men's) and are sampled at 8KHz. In the figure the results of spectral subtraction (SS), Y.Ephraim's method (SUB1), subspace method using masking threshold based on N band whitening filter (SUBN-M) are plotted. Since there is little difference between clean speech and the estimate at high input SNR, we can say that the speech quality corresponding to the dots in the upper right direction is better than that corresponding to the dots in the lower left direction. The test results show that 4 band method performed the best.

7. CONCLUSION

In this paper, subspace enhancement method using the masking threshold is proposed. By constraining the energy of the residual noise to be just below the masking threshold, we can minimize the signal distortion while making the residual noise inaudible. In order to deal with colored noise effectively, the proposed algorithm incorporates subband whitening filters. The experimental results based on both SNRseg and WSS show that the performance of the proposed algorithm is superior to that of the subspace method proposed by Y.Ephraim.



(a) Aircraft cockpit noise (b) IBM PS/2 cooling fan noise

(c) Tank noise (d) Waterfall noise

Fig. 3. SNRseg increment vs. WSS decrement in dB scale for various noise environment with varying input SNR. Each dot in the figure is associated with input SNRs from 0dB to 20dB with step 2dB.

8. REFERENCES

[1] Y.Ephraim and H.L. Van Trees, "Signal subspace approach for speech enhancement," *IEEE Trans. Speech, Audio Processing,* vol. 3, no. 4, pp. 251–266, Jul. 1995.

[2] F.Jabloun and B.Champagne, "A perceptual signal subspace approach for speech enhancement in colored noise," *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing,* (Orlando, FL, U.S.A.), pp. I-567–I-572, May 2002.

[3] M.Klein and P.Kabal, "Signal subspace speech enhancement with perceptual post-filtering," *Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing,* (Orlando, FL, U.S.A.), pp. I-537–I-540, May 2002.

[4] J.D.Johnston, "Transform coding of audio signals using perceptual noise criteria," *IEEE Journal on Selected Areas in Communications,* vol. 6, no. 2., pp. 314–323, Feb. 1988.

[5] J.U.Kim and C.D.Yoo, "Subspace speech enhancement using subband whitening filter," *Proc. on International Conference on Spoken Language Processing,* (Denver, CO, U.S.A.), vol. 3, pp. 1805–1808, Sep. 2002.

[6] J.H.L.Hansen and B.L.Pellom, "An effective quality evaluation protocol for speech enhancement algorithms," *Proc. on International Conference on Spoken Language Processing,* Sydney Australia, vol. 7, pp. 2819–2822, Nov. 1998.