PRINCIPAL SUBSPACE MODIFICATION FOR MULTI-CHANNEL WIENER FILTER IN MULTI-MICROPHONE NOISE REDUCTION

Gibak Kim

Department of Electrical Engineering
University of Texas at Dallas
Richardson, TX 75080-3021
imkgb27@utdallas.edu

Nam Ik Cho

School of Electrical Engineering
Seoul National University
Seoul 151-744 Korea

ABSTRACT

In multi-microphone noise reduction for single desired speech signal, the principal subspace based multi-channel Wiener filter provides better performance compared with the conventional multi-channel Wiener filter. The principal subspace vector estimates the acoustic transfer function vector up to a scaling factor. However, as input SNR becomes lower, the error increases in the acoustic transfer function vector estimation. In this paper, we propose the principal subspace modification which is controlled by the angle between the principal subspace vector and the steering vector of the desired speech signal. In the simulation, the proposed method is evaluated with multi-channel speech data which are degraded by interfering noise coming from other direction. The simulation results show that the modification of principal subspace vector allows better performance compared to the conventional principal subspace based multi-channel Wiener filter.

***Index Terms*—** Microphone array, multi-channel filtering, noise reduction

1. INTRODUCTION

In multi-microphone system, noise can be reduced by spatial filtering such as beamforming techniques when the desired speech and noise signals arrive from different directions. Multi-microphone noise reduction using spatial filtering provides more noise reduction and less distortion compared to single microphone techniques. To further reduce the residual noise, the output of beamformer can be filtered by a single channel post-filter [1]. Recently, the multi-channel Wiener filter (MWF) has been developed, which can be analyzed as the combination of spatial filter and single channel spectral filter [2, 3].

For more efficient noise reduction, subspace based approach is applied to the MWF, which removes noise subspace and estimates the desired speech component from the remaining signal subspace. In this paper, subspace decomposition is performed in the frequency domain and the subspace vectors are obtained by joint diagonalization of the multi-channel input spatial correlation matrix and the noise spatial correlation matrix. The desired speech spatial correlation matrix can be estimated by subtracting the noise spatial correlation matrix from the input spatial correlation matrix under the assumption that the desired speech signal is uncorrelated with noise. Then, the principal subspace vector estimates the acoustic transfer function vector up to a scaling factor [4]. In practical situations, the cross correlation between the desired speech and noise may not be zero, but can be ignored when the cross correlation is much smaller than the desired speech power. However, in the case of low SNRs, the cross correlation is not much smaller than the desired speech power any more and cannot be ignored. Thus, the obtained principal subspace vector deviates from the acoustic transfer function vector and the performance is degraded at low SNRs. For better performance at low SNRs, we propose the principal subspace vector modification controlled by the angle between the principal subspace vector and the steering vector of the desired speech signal. The principal subspace vector is modified by the linear interpolation between the original principal subspace vector and the steering vector of the desired speech signal. Simulation results demonstrate that the modified subspace vector gets closer to the acoustic transfer function vector and the modified principal subspace based MWF provides better performance than the conventional principal subspace based MWF.

This paper is organized as follows. Section 2 describes the multi-channel signal model and reviews the frequency domain MWF for noise reduction. Section 3 discusses the principal subspace vector and proposes the modified principal subspace based MWF. Simulation results and performance evaluation are shown in Section 4.

2. FREQUENCY DOMAIN MWF AND SUBSPACE DECOMPOSITION

If there are M microphones and a single desired speech source, let us consider an M-channel signal model where the desired speech source is convolved with M acoustic transfer functions to every microphone. When additive noise degrades the multi-channel speech, the multi-channel signal model is given by

$$y_i[k] = h_i[k] * s[k] + n_i[k] = x_i[k] + n_i[k] \quad i = 1, \ldots, M \tag{1}$$

where $y_i[k]$ denotes the observed signal at the i-th microphone at time k, $x_i[k]$ and $n_i[k]$ are speech and additive noise component respectively, $s[k]$ is the desired speech source, and $h_i[k]$ is the acoustic transfer function from the desired speech source to the i-th microphone. Assuming infinite filter lengths, (1) is represented in the frequency domain as

$$\begin{aligned}\mathbf{Y}(f) = \begin{bmatrix} Y_1(f) \\ Y_2(f) \\ \vdots \\ Y_M(f) \end{bmatrix} &= S(f) \begin{bmatrix} H_1(f) \\ H_2(f) \\ \vdots \\ H_M(f) \end{bmatrix} + \begin{bmatrix} N_1(f) \\ N_2(f) \\ \vdots \\ N_M(f) \end{bmatrix} \\ &= S(f)\mathbf{H}(f) + \mathbf{N}(f) = \mathbf{X}(f) + \mathbf{N}(f) \end{aligned} \tag{2}$$

where $Y_i(f)$, $H_i(f)$, $S(f)$, $N_i(f)$, $X_i(f)$ are frequency domain representations of $y_i[k]$, $h_i[k]$, $s[k]$, $n_i[k]$, $x_i[k]$, respectively. With a multi-channel noise reduction filter $\mathbf{W}(f)$, the output $Z(f)$ can be written as

$$Z(f)=\mathbf{W}^{H}(f)\mathbf{Y}(f). \tag{3}$$

Hereafter the frequency index (f) is omitted for the sake of brevity. If we assume that the desired speech and noise signals are uncorrelated and estimate the desired speech component in the 1st microphone signal in the minimum mean square error (MMSE) sense, the frequency domain MWF is given by

$$\begin{aligned}\mathbf{W} &\simeq \mathbf{R}_{\mathbf{YY}}^{-1}\mathbf{R}_{\mathbf{XX}}\mathbf{e}_1\\ &\simeq \mathbf{R}_{\mathbf{YY}}^{-1}\left(\mathbf{R}_{\mathbf{YY}}-\mathbf{R}_{\mathbf{NN}}\right)\mathbf{e}_1\end{aligned} \tag{4}$$

with $\mathbf{R}_{\mathbf{YY}}=E\left\{\mathbf{Y}\mathbf{Y}^{H}\right\}$, $\mathbf{R}_{\mathbf{XX}}=E\left\{\mathbf{X}\mathbf{X}^{H}\right\}$, $\mathbf{R}_{\mathbf{NN}}=E\left\{\mathbf{N}\mathbf{N}^{H}\right\}$, and $\mathbf{e}_1=\left[\begin{array}{cccc}1 & 0 & \cdots & 0\end{array}\right]^{T}$.

By incorporating the subspace decomposition in the frequency domain, the spatial subspaces can be taken into consideration. The subspace decomposition can be performed by joint diagonalization of the spatial correlation matrices $\mathbf{R}_{\mathbf{YY}}$ and $\mathbf{R}_{\mathbf{NN}}$ which can be obtained by solving the generalized eigenvalue problem as

$$\mathbf{R}_{\mathbf{YY}}\mathbf{Q}=\mathbf{R}_{\mathbf{NN}}\mathbf{Q}\boldsymbol{\Lambda}, \tag{5}$$

$$\left\{\begin{array}{l}\mathbf{Q}^{H}\mathbf{R}_{\mathbf{YY}}\mathbf{Q}=\boldsymbol{\Lambda}_{\mathbf{Y}}\\ \mathbf{Q}^{H}\mathbf{R}_{\mathbf{NN}}\mathbf{Q}=\boldsymbol{\Lambda}_{\mathbf{N}}\end{array}\right. \tag{6}$$

where $\boldsymbol{\Lambda},\boldsymbol{\Lambda}_{\mathbf{Y}},\boldsymbol{\Lambda}_{\mathbf{N}}$ are diagonal matrices as

$$\boldsymbol{\Lambda}=\operatorname{diag}\left\{\lambda_1\ \ \lambda_2\ \cdots\ \lambda_M\right\} \tag{7}$$

$$\boldsymbol{\Lambda}_{\mathbf{Y}}=\operatorname{diag}\left\{\lambda_{Y,1}\ \ \lambda_{Y,2}\ \cdots\ \lambda_{Y,M}\right\} \tag{8}$$

$$\boldsymbol{\Lambda}_{\mathbf{N}}=\operatorname{diag}\left\{\lambda_{N,1}\ \ \lambda_{N,2}\ \cdots\ \lambda_{N,M}\right\} \tag{9}$$

with $\boldsymbol{\Lambda}=\boldsymbol{\Lambda}_{\mathbf{Y}}\boldsymbol{\Lambda}_{\mathbf{N}}^{-1}$, $\lambda_i=\frac{\lambda_{Y,i}}{\lambda_{N,i}}$, $\lambda_1>\lambda_2>\cdots>\lambda_M$ and $\mathbf{Q}$ is an invertible, but not necessarily orthogonal matrix. Then the spatial correlation matrices can be expressed by the subspace matrix $\bar{\mathbf{Q}}$ as

$$\left\{\begin{array}{l}\mathbf{R}_{\mathbf{YY}}=\bar{\mathbf{Q}}\boldsymbol{\Lambda}_{\mathbf{Y}}\bar{\mathbf{Q}}^{H}\\ \mathbf{R}_{\mathbf{NN}}=\bar{\mathbf{Q}}\boldsymbol{\Lambda}_{\mathbf{N}}\bar{\mathbf{Q}}^{H}\end{array}\right. \tag{10}$$

with $\bar{\mathbf{Q}}=\mathbf{Q}^{-H}$. By substituting (10) into (4) the frequency domain MWF is obtained as

$$\mathbf{W}=\mathbf{Q}\left(\mathbf{I}-\boldsymbol{\Lambda}_{\mathbf{Y}}^{-1}\boldsymbol{\Lambda}_{\mathbf{N}}\right)\bar{\mathbf{Q}}^{H}\mathbf{e}_1. \tag{11}$$

3. PRINCIPAL SUBSPACE VECTOR MODIFICATION

3.1. Principal Subspace Vector

When each of the frequency domain multi-channel speech components is the multiplication of each acoustic transfer function and a single desired speech source as shown in (2), the desired speech spatial correlation matrix can be written as

$$\mathbf{R}_{\mathbf{XX}}=E\left\{\mathbf{X}\mathbf{X}^{H}\right\}=E\left\{SS^{*}\right\}\mathbf{H}\mathbf{H}^{H} \tag{12}$$

and the rank of $\mathbf{R}_{\mathbf{XX}}$ is equal to 1. From (10) and the rank-1 property of $\mathbf{R}_{\mathbf{XX}}$, the estimate of the desired speech spatial correlation matrix is given by

$$\begin{aligned}\mathbf{R}_{\mathbf{XX}}&=\mathbf{R}_{\mathbf{YY}}-\mathbf{R}_{\mathbf{NN}}-\mathbf{R}_{\mathbf{XN}}-\mathbf{R}_{\mathbf{XN}}^{H}\\ &\simeq \mathbf{R}_{\mathbf{YY}}-\mathbf{R}_{\mathbf{NN}}\end{aligned} \tag{13}$$

$$\begin{aligned}\mathbf{R}_{\mathbf{XX}}&\simeq \bar{\mathbf{Q}}(\boldsymbol{\Lambda}_{\mathbf{Y}}-\boldsymbol{\Lambda}_{\mathbf{N}})\bar{\mathbf{Q}}^{H}\\ &\simeq (\lambda_{Y,1}-\lambda_{N,1})\bar{\mathbf{q}}_1\bar{\mathbf{q}}_1^{H}\end{aligned} \tag{14}$$

where $\mathbf{R}_{\mathbf{XN}}=E\left\{\mathbf{X}\mathbf{N}^{H}\right\}$ and the desired speech signal is assumed to be uncorrelated with noise and the M-dimensional principal subspace vector $\bar{\mathbf{q}}_1$ is the 1st column vector of $\bar{\mathbf{Q}}$. From (12) and (14), note that $\bar{\mathbf{q}}_1$ is an estimate of the acoustic transfer function vector $\mathbf{H}$ up to a scaling factor [4]. The principal subspace based MWF can be expressed as

$$\mathbf{W}=\lambda_{N,1}\mathbf{R}_{\mathbf{NN}}^{-1}\bar{\mathbf{q}}_1\left(\frac{\lambda_{Y,1}-\lambda_{N,1}}{\lambda_{Y,1}}\right)\bar{\mathbf{q}}_1^{H}\mathbf{e}_1 \tag{15}$$

$$\left\{\begin{array}{l}\lambda_{Y,1}=\left(\bar{\mathbf{q}}_1^{\mathbf{H}}\mathbf{R}_{\mathbf{YY}}^{-1}\bar{\mathbf{q}}_1\right)^{-1}\\ \lambda_{N,1}=\left(\bar{\mathbf{q}}_1^{\mathbf{H}}\mathbf{R}_{\mathbf{NN}}^{-1}\bar{\mathbf{q}}_1\right)^{-1}\end{array}\right. . \tag{16}$$

3.2. Modification of Principal Subspace Vector

For the estimation of $\mathbf{R}_{\mathbf{XX}}$, the desired speech and noise signals are assumed to be uncorrelated in (13). In practical situations, the assumption may not be true, but the cross correlation between the desired speech and noise can be ignored at high SNRs where the absolute value of the cross correlation $\left|E\left\{X_iN_j^{*}\right\}\right|$ is much smaller than $\left|E\left\{X_iX_j^{*}\right\}\right|$. However, in the case of low SNRs, the cross correlation cannot be ignored any more, and a large error occurs in the estimation of $\mathbf{R}_{\mathbf{XX}}$ and consequently, $\bar{\mathbf{q}}_1$ deviates from $\mathbf{H}$. To illustrate the closeness between $\mathbf{H}$ and $\bar{\mathbf{q}}_1$ as a function of input SNR, the angle between $\bar{\mathbf{q}}_1$ and $\mathbf{H}$ is examined in Fig. 1. The angle between two vectors $\mathbf{v}_1$ and $\mathbf{v}_2$ is one possible measure of the closeness and can be defined as

$$\angle(\mathbf{v}_1,\mathbf{v}_2)=\cos^{-1}\left(\frac{\left|\mathbf{v}_1^{H}\mathbf{v}_2\right|}{\|\mathbf{v}_1\|\,\|\mathbf{v}_2\|}\right) \tag{17}$$

$$0\le\angle(\mathbf{v}_1,\mathbf{v}_2)\le\pi/2$$

where $\|\cdot\|$ denotes the vector norm. For the illustration of Fig. 1, $\mathbf{H}$ is approximated as the principal eigenvector of the desired speech spatial correlation matrix which is estimated using the noise-free speech data (refer to Section 4 for the simulation data). The angle between two vectors goes high as input SNR becomes low, which implies that $\bar{\mathbf{q}}_1$ deviates from $\mathbf{H}$. Consequently, the principal subspace based MWF does not perform well at low SNRs.

To obtain better performance of the principal subspace based MWF at low SNRs, we propose principal subspace modification method using the information on the direction of the desired speech signal. First, we assume that the direction of the desired speech signal is known. Then, the angle between $\bar{\mathbf{q}}_1$ and $\mathbf{H}$ is calculated to measure the closeness of the two vectors. The steering vector of the desired speech signal can be an estimate of $\mathbf{H}$. The steering vector is an M-dimensional vector as

$$\mathbf{v}_s=\left[\begin{array}{cccc}1 & e^{j\varphi_2} & \cdots & e^{j\varphi_M}\end{array}\right]^{T} \tag{18}$$

in which φ_i represents the phase of the i-th microphone signal with respect to the first microphone and can be obtained from the direction (angle) of the desired speech signal with respect to the microphones, the signal frequency, and the microphone array configuration [5]. Before calculating the angle between $\bar{\mathbf{q}}_1$ and $\mathbf{v}_s$, each element of $\bar{\mathbf{q}}_1$ is divided by its absolute value as

$$\bar{\bar{\mathbf{q}}}_1=\left[\begin{array}{cccc}\frac{\bar{q}_1}{|\bar{q}_1|} & \frac{\bar{q}_2}{|\bar{q}_2|} & \cdots & \frac{\bar{q}_M}{|\bar{q}_M|}\end{array}\right]^{T} \tag{19}$$

with $\bar{\mathbf{q}}_1=\left[\begin{array}{cccc}\bar{q}_1 & \bar{q}_2 & \cdots & \bar{q}_M\end{array}\right]^{T}$. By calculating the angle between $\bar{\bar{\mathbf{q}}}_1$ and $\mathbf{v}_s$ instead of the angle between $\bar{\mathbf{q}}_1$ and $\mathbf{v}_s$, we alleviate the error caused by the microphone gain mismatch. Finally, the

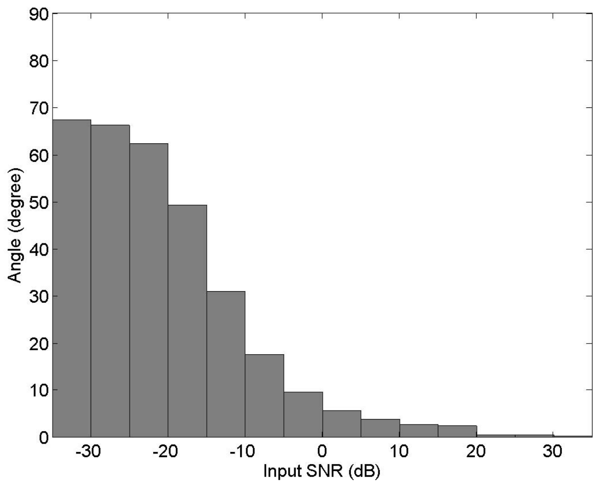

Fig. 1. Angle between the principal subspace vector $\bar{\mathbf{q}}_1$ and the acoustic transfer function vector $\mathbf{H}$.

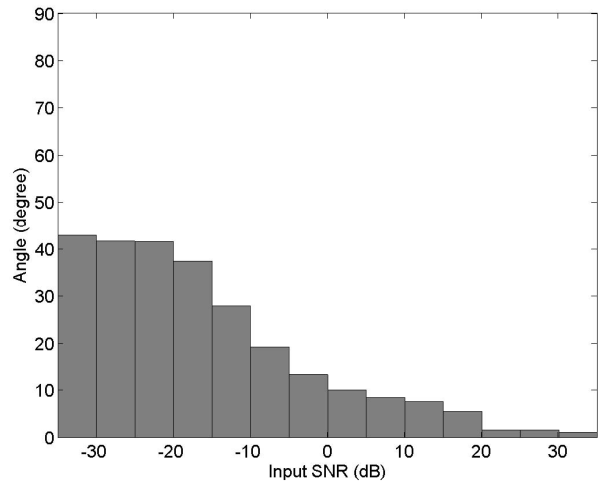

Fig. 2. Angle between the modified principal subspace vector $\bar{\mathbf{q}}_1'$ and the acoustic transfer function vector $\mathbf{H}$.

principal subspace vector $\bar{\mathbf{q}}_1$ is modified by the linear interpolation between $\bar{\bar{\mathbf{q}}}_1$ and $\mathbf{v}_s$ as

$$\bar{\bar{\mathbf{q}}}_1' = (1-\alpha)\frac{\bar{\bar{\mathbf{q}}}_1}{\|\bar{\bar{\mathbf{q}}}_1\|} + \alpha\frac{\mathbf{v}_s}{\|\mathbf{v}_s\|} \tag{20}$$

$$\alpha = \frac{\angle(\mathbf{v}_s, \bar{\bar{\mathbf{q}}}_1)}{\pi/2}. \tag{21}$$

After the interpolation controlled by α, each element of $\bar{\bar{\mathbf{q}}}_1'$ is multiplied by each absolute value of the element of $\bar{\mathbf{q}}_1$ to consider the channel gain as

$$\bar{\mathbf{q}}_1' = \bar{\bar{\mathbf{q}}}_1' \bullet |\bar{\mathbf{q}}_1| \tag{22}$$

where $\bullet$ denotes the elementwise product. Fig. 2 shows that the modified principal subspace vector $\bar{\mathbf{q}}_1'$ becomes closer to $\mathbf{H}$ at low SNRs. After the modification of the principal subspace vector, the MWF is calculated by (15) and (16).

4. SIMULATION RESULTS

The multi-channel signals are generated by the convolution of dry source (sound data measured in an anechoic room) with acoustic impulse responses from the RWCP Sound Scene Database [6]. The desired speech and noise sources are 2 m away from the center of microphones. The microphone array is a linear type and has 7 microphones located at 5.66 cm uniform intervals. In this simulation, desired speech signal is convolved with the impulse response measured at the fore side of the microphone array and added by white noise or a competing speech noise signal coming at the angle of $40°$ with reverberation time of 300 ms.

The performance is evaluated by the objective measures such as SNR gain, speech distortion, and log spectral distortion, and mel-frequency cepstral coefficient (MFCC) distortion. When the noise reduction is applied to the speech recognition system, the most important measure is the MFCC distortion which is widely used as a feature vector for the speech recognition. The log spectral distortion and MFCC distortion are distortions between the desired speech component in the 1st microphone signal and the output of MWF. The modified principal subspace based MWF (PS-MOD) is compared with following MWFs:

1) PS: the principal subspace based MWF without modifying the principal subspace vector;

2) PS-SV: the principal subspace based MWF where the principal subspace vector is replaced with the steering vector of the desired speech;

3) PS-SP: the principal subspace based MWF where the principal subspace vector is replaced with the principal eigenvector of $\mathbf{R_{XX}}$ which is estimated from noise-free speech signal.

The PS-SV MWF is equal to the minimum variance distortionless response (MVDR) beamformer followed by a single channel Wiener filter [3, 5]. The PS-SP MWF cannot be implemented in practical situations since the noise-free speech signal is not accessible. It is just evaluated for estimating the performance of the principal subspace based MWF when the principal subspace vector is equal to the acoustic transfer function vector up to a scaling factor. Fig. (3) and Fig. (5) describe the evaluation in the case of white noise, and the results for competing speech noise case are shown in Fig. (4) and Fig. (6). Fig. (3) and Fig. (4) show the SNR gain and the speech distortion of each MWF. The proposed method (PS-MOD) shows better SNR gain than other methods at the cost of similar amount of speech distortion in the PS MWF. The log spectral distortion and MFCC distortion are shown in Fig. (5) and Fig. (6). The PS-MOD MWF yields better performance than the PS MWF. Although the PS-SV MWF provides similar or slightly better performance in the log spectral distortion than the PS-MOD MWF, the PS-SV MWF provides less improvement in terms of the MFCC distortion. The reason is that the PS-SV MWF is more susceptible to the microphone gain mismatch which causes more distortion at low frequencies and the mel-scale filter bank for the MFCC has higher resolution at lower frequencies.

5. CONCLUSIONS

In this paper, we have proposed a principal subspace modification for the MWF. The principal subspace vector is modified by the interpolation between the principal subspace vector and the steering vector of the desired speech signal, which reduces the estimation error of

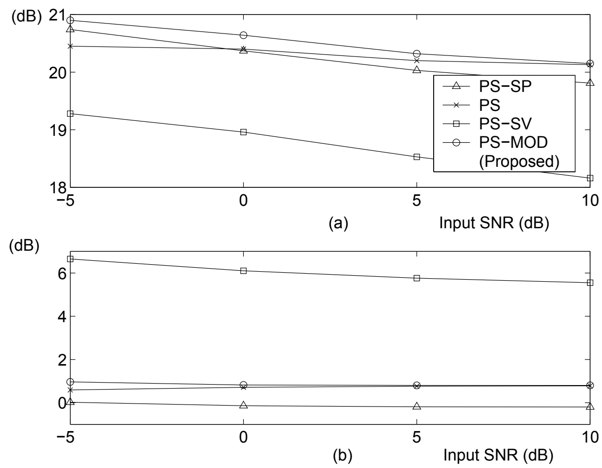

Fig. 3. (a) SNR gain and (b) speech distortion as a function of input SNR : white noise.

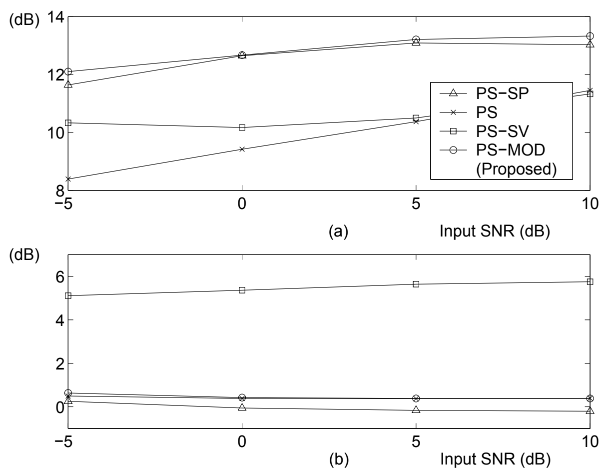

Fig. 4. (a) SNR gain and (b) speech distortion as a function of input SNR : competing speech noise.

the acoustic transfer function vector at low SNRs. The simulation results demonstrate the improvement of the modified principal subspace based MWF.

ACKNOWLEDGEMENT

This research was performed for the Intelligent Robotics Development Program, one of the 21st Century Frontier R&D Programs funded by the Ministry of Commerce, Industry and Energy of Korea.

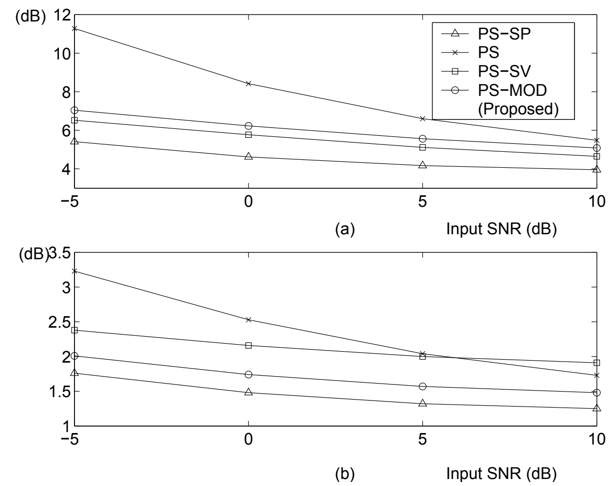

Fig. 5. (a) Log spectral distortion and (b) MFCC distortion as a function of input SNR : white noise.

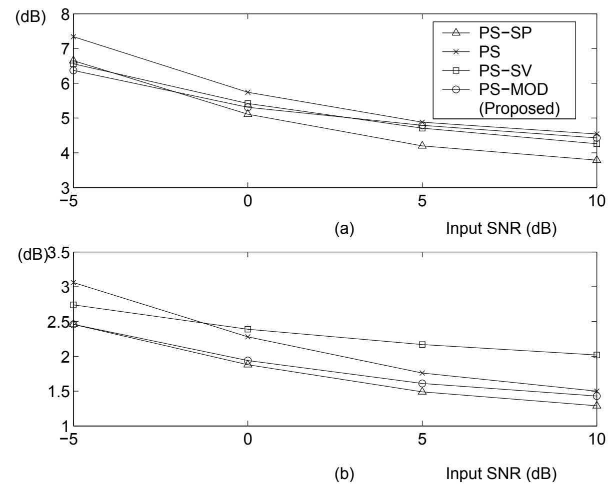

Fig. 6. (a) Log spectral distortion and (b) MFCC distortion as a function of input SNR : competing speech noise.

6. REFERENCES

[1] C. Marro, Y. Mahieux, and K. U. Simmer, "Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering," *IEEE Trans. on Speech and Audio Processing*, vol. 6, no. 3, May 1998.

[2] S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," *IEEE Trans. on Signal Processing*, vol. 50, no. 9, pp. 2230-2244, September 2002.

[3] A. Spriet, M. Moonen, and J. Wouters, "Robustness analysis of multichannel Wiener filtering and generalized sidelobe cancellation for multimicrophone noise reduction in hearing aid applications," *IEEE Trans. on Speech and Audio Processing*, vol. 13, no. 4, pp. 487-503, July 2005.

[4] S. Doclo and M. Moonen, "Combined frequency-domain dereverberation and noise reduction technique for multi-microphone speech enhancement," *International Workshop on Acoustic Echo and Noise Control*, Darmstadt, Germany, pp. 31-34, September 2001.

[5] H. Van Trees, *Detection, Estimation and Modulation Theory, Part IV: Optimum Array Processing*, New York: Wiley, 2002.

[6] "RWCP Sound Scene Database in Real Acoustical Environments," Real World Computing Partnership, (c)1998-2001.