Journal of Southeast University (English Edition) Vol. 23, No. 4, pp. 507 – 511 Dec. 2007 ISSN 1003—7985

Speech enhancement based on leakage constraints DF-GSC

Zou Cairong[1,2] Chen Guoming[1] Zhao Li[1]

([1] School of Information Science and Engineering, Southeast University, Nanjing 210096, China)

([2] Foshan University, Foshan 528000, China)

Abstract: In order to improve the performance of general sidelobe canceller (GSC) based speech enhancement, a leakage constraints decision feedback generalized sidelobe canceller(LCDF-GSC) algorithm is proposed. The method adopts DF-GSC against signal mismatch, and introduces a leakage factor in the cost function to deal with the speech leakage problem which is caused by the part of the speech signal in the noise reference signal. Simulation results show that although the signal-to-noise ratio (SNR) of the speech signal through LCDF-GSC is slightly less than that of DF-GSC, the IS measurements show that the distortion of the former is less than that of the latter. MOS (mean opinion score) scores also indicate that the LCDF-GSC algorithm is better than DF-GSC and the Weiner filter algorithm.

Key words: speech enhancement; general sidelobe canceller (GSC); speech leakage

In speech communication applications, such as mobile phones, hands-free telephones and hearing aids, speech signals are often corrupted by acoustic background noise. Many speech enhancement methods have been introduced to solve these problems[1–2]. Since a multi-microphone system exploits spatial information in addition to temporal and spectral information of the desired signal and noise signal, it achieves much better performance than traditional single microphone-based speech enhancement algorithms such as Wiener filtering[1], subspace-based enhancement[2] etc.. Frost[3] provided an algorithm which deals with the problem of a broadband signal received by an array. The algorithm is capable of satisfying some desired signal in the look direction by using constrained minimization of the total output power. Griffiths and Jim[4] reconsidered Frost's algorithm and introduced the generalized sidelobe canceller (GSC) solution which is a widely used noise reduction algorithm for a multi-microphone system by adjusting the weights of a sensor array with adaptive filters.

The GSC algorithm achieves better performance assuming that the desired speaker location, the microphone characteristics and positions are known beforehand. However, in reality, these assumptions are often violated, resulting in speech leakage into the noise references which causes speech distortion[5]. Although some techniques[6–7] were introduced to reduce the amount of speech leakage, it can never be completely avoided. We consider speech leakage with leakage constraints in the GSC cost function and adopt the DF-GSC method against signal mismatch.

1 GSC-Based Optimal Filter

A GSC-based optimal filter is illustrated in Fig. 1. It consists of three parts. Part A is a fixed beamformer filter which is responsible for providing a speech reference signal. Part B is a blocking matrix which creates a noise reference signal by blocking the direction of the speech source, and part C is an adaptive filter which uses the noise reference signal as an input signal and the speech reference signal as a desired signal. The GSC attempts to recover the speech signal by constraining the array response to unity in the direction of the speech source and minimizes the total energy which is coming from all other directions.

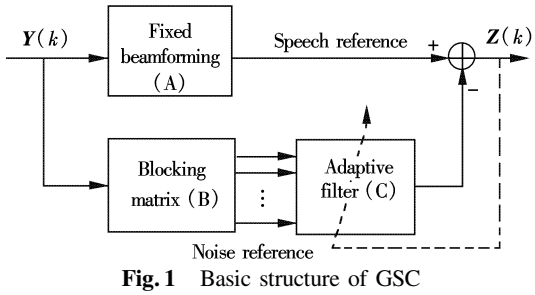

Fig. 1 Basic structure of GSC

Consider a uniform linear array(ULA) of N microphones, where each microphone signal $\boldsymbol{Y}_n(k)$, $n = 0, 1, \ldots, N-1$, at tap time k, consists of a filtered version of the clean speech signal and additive noise:

$$\boldsymbol{Y}_n(k) = \boldsymbol{x}_n(k) + \boldsymbol{v}_n(k) \tag{1}$$

Received 2006-11-28.

Foundation items: The National Natural Science Foundation of China (No. 60472058), the Ph. D. Programs Foundation of Ministry of Education of China(No. 20050286001), Program for New Century Excellent Talents in University(No. NCET-04-0483).

Biography: Zou Cairong(1963—), male, doctor, professor, zoucairong@seu.edu.cn.

where $\boldsymbol{x}_n(k)$ and $\boldsymbol{v}_n(k)$ are the speech component and the noise component received at the n-th microphone, respectively. The additive noise can be colored and is assumed to be uncorrelated with the speech signal.

In Fig. 2, a detailed fixed beamformer is provided. The speech signal from far field impinges on the array from a known DOA of θ_0 along with $M-1$ uncorrelated noise from unknown DOAs $\{\theta_1, \theta_2, \ldots, \theta_{M-1}\}$, so the received signal can be written as

$$\boldsymbol{Y}(k) = \boldsymbol{a}(\theta_0)x_0(k) + \sum_{m=1}^{M-1}\boldsymbol{a}(\theta_m)x_m(k) + \boldsymbol{v}(k) = \boldsymbol{x}(k) + \boldsymbol{v}(k) \quad (2)$$

where $\boldsymbol{a}(\theta_0) = \{1, \exp(\mathrm{i}\tau_{\theta_0}), \exp(\mathrm{i}2\tau_{\theta_0}), \ldots, \exp(\mathrm{i}(N-1)\tau_{\theta_0})\}^{\mathrm{T}}$, with $\tau_{\theta_0} = \dfrac{2\pi d}{\lambda}\sin\theta_0$, and λ is the signal wavelength. Similarly, $\boldsymbol{a}(\theta_m) = \{1, \exp(\mathrm{i}\tau_{\theta_m}), \exp(\mathrm{i}2\tau_{\theta_m}), \ldots, \exp(\mathrm{i}(N-1)\tau_{\theta_m})\}^{\mathrm{T}}$, with $\tau_{\theta_m} = \dfrac{2\pi d}{\lambda}\sin\theta_m$. The output signal is $z(k) = \sum_{n=0}^{N-1}\boldsymbol{w}_n^{\mathrm{T}}(k)\boldsymbol{y}_n(k)$.

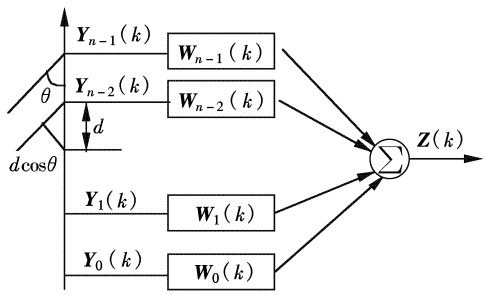

Fig. 2 Structure of fixed beamformer

The matrix $\boldsymbol{B}$ with dimension $N \times (N-1)$, blocking the direction of the speech source should satisfy the condition given as

$$\boldsymbol{B}^{\mathrm{H}}\boldsymbol{a}(\theta_0) = \boldsymbol{0} \quad (3)$$

Part C is depicted in Fig. 3. The output signal $\boldsymbol{z}(k)$ can be written as $\boldsymbol{z}(k) = \sum_{n=0}^{N-1}\boldsymbol{w}_n^{\mathrm{T}}(k)\boldsymbol{y}_n(k) = \boldsymbol{W}^{\mathrm{T}}(k)\boldsymbol{y}(k)$. In Fig. 3, $\boldsymbol{x}(k)$ is the desired response vector and $\boldsymbol{e}(k) = \boldsymbol{x}(k) - \boldsymbol{z}(k)$ is the estimation error vector. The mean square error (MSE) cost function leads to the well-known multidimensional Wiener filter $\boldsymbol{W}(k) = \boldsymbol{R}_{yy}^{-1}(k)\boldsymbol{R}_{yx}(k)$.

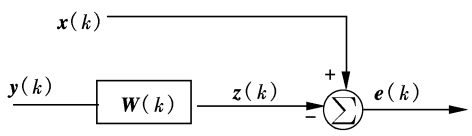

Fig. 3 Structure of MWF

2 Leakage Constraints Decision Feedback GSC (LCDF-GSC)

2.1 GSC solution

An effective approach to determining the weight $\boldsymbol{W}(k)$ in part C is based on the LCMV (linearly constrained minimum variance) criterion, which can be expressed as

$$\min_{\boldsymbol{W}}\boldsymbol{W}^{\mathrm{H}}\boldsymbol{R}_y\boldsymbol{W} \quad \text{subject to } \boldsymbol{C}^{\mathrm{H}}\boldsymbol{W} = \boldsymbol{f} \quad (4)$$

where $\boldsymbol{R}_y(k) = E\{\boldsymbol{y}(k)\boldsymbol{y}^{\mathrm{T}}(k)\}$ is the input correlation matrix; $\boldsymbol{C}$ is an $N \times p$ constraint matrix and can be expressed as $\boldsymbol{C} = \boldsymbol{I}_N \otimes \boldsymbol{C}^k$ with $\boldsymbol{C}^k = \{C_0, C_1, \ldots, C_{p-1}\}$; and $\boldsymbol{f} = \{\boldsymbol{f}^k, \boldsymbol{0}_p^{\mathrm{T}}, \ldots, \boldsymbol{0}_p^{\mathrm{T}}\}^{\mathrm{T}}$ is a $p \times 1$ response vector, in which $\boldsymbol{0}_p^{\mathrm{T}}$ is a $p \times 1$ zero vector and $\boldsymbol{I}_N$ is an $N \times N$ identity matrix. By decomposing $\boldsymbol{W} = \boldsymbol{W}_{\mathrm{q}} - \boldsymbol{B}\boldsymbol{W}_{\mathrm{a}}$, where $\boldsymbol{W}_{\mathrm{q}} = \boldsymbol{C}(\boldsymbol{C}^{\mathrm{H}}\boldsymbol{C})^{-1}\boldsymbol{f}$, $\boldsymbol{B} = \boldsymbol{I}_N \otimes \boldsymbol{B}^k$, in which $\otimes$ denotes the Kronecker product, the LCMV beamformer can be implemented via the GSC structure and can be formulated as an unconstrained optimization problem:

$$\min_{\boldsymbol{W}_{\mathrm{a}}}\boldsymbol{J} = \min_{\boldsymbol{W}_{\mathrm{a}}}(\boldsymbol{W}_{\mathrm{q}} - \boldsymbol{B}\boldsymbol{W}_{\mathrm{a}})^{\mathrm{H}}\boldsymbol{R}_y(\boldsymbol{W}_{\mathrm{q}} - \boldsymbol{B}\boldsymbol{W}_{\mathrm{a}}) \quad (5)$$

And the optimum solution of $\boldsymbol{W}_{\mathrm{a}}$ can be calculated as

$$\boldsymbol{W}_{\mathrm{a,\,opt}} = (\boldsymbol{B}^{\mathrm{H}}\boldsymbol{R}_y\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{H}}\boldsymbol{R}_y\boldsymbol{W}_{\mathrm{q}} \quad (6)$$

2.2 Decision feedback GSC

In Refs. [8 − 9], a new scheme to improve the performance of the traditional GSC algorithm was proposed. Fig. 4 depicts the whole structure of the proposed DF-GSC. Accordingly, the cost function (5) is changed to

$$\min_{\boldsymbol{W}_{\mathrm{a}}}\boldsymbol{J} = \min_{\boldsymbol{W}_{\mathrm{a}}}E\{|\boldsymbol{X}(k) - \boldsymbol{W}_{\mathrm{b}}^{*}\hat{\boldsymbol{S}}_0(k)|^2\} = \min_{\boldsymbol{W}_{\mathrm{a}}}E\{|(\boldsymbol{W}_{\mathrm{q}} - \boldsymbol{B}\boldsymbol{W}_{\mathrm{a}})^{\mathrm{H}}\boldsymbol{Y}(k) - \boldsymbol{W}_{\mathrm{b}}^{*}\hat{\boldsymbol{S}}_0(k)|^2\} \quad (7)$$

where $\boldsymbol{W}_{\mathrm{b}}$ is the feedback tap weight and $\hat{\boldsymbol{S}}_0(k)$ is the detected desired signal. The optimum solution of (7) is

$$\boldsymbol{W}_{\mathrm{a,\,opt}} = (\boldsymbol{B}^{\mathrm{H}}\boldsymbol{R}_Y\boldsymbol{B})^{-1}\boldsymbol{B}^{\mathrm{H}}\boldsymbol{R}_Y\boldsymbol{W}_{\mathrm{q}}, \quad \boldsymbol{W}_{\mathrm{b,\,opt}} = \boldsymbol{a}(\theta_0)\boldsymbol{W}_{\mathrm{q}} \quad (8)$$

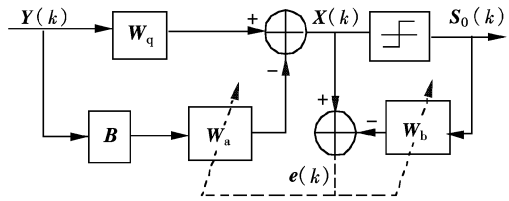

Fig. 4 Structure of DF-GSC

2.3 Leakage constraints decision feedback GSC

Theoretically, the output of the blocking matrix $\boldsymbol{B}$ is a noise reference signal; i. e., $d_i(k) = v_i(k)$, $k = 1, 2, \ldots, M-1$. In practical application, the vector $\boldsymbol{d}$ often contains part of the speech signal which is a so-called speech leakage; i. e., $d_i(k) = \underbrace{x_i(k)}_{\text{speech}} + \underbrace{v_i(k)}_{\text{noise}}$, $k = 1, 2, \ldots, M-1$; hence, the adaptive filter will also remove part of the speech signal from the speech reference signal. In this case speech distortion is introduced. In this section we consider a new beamformer which incorporates the new speech leakage constraints with the decision feedback GSC technique. The cost function is changed to

$$\boldsymbol{J}=\min_{\boldsymbol{W}_{\mathrm{a}},\boldsymbol{W}_{\mathrm{b}}}E\{\,|\,[(\boldsymbol{W}_{\mathrm{q}}-\boldsymbol{B}\boldsymbol{W}_{\mathrm{a}})^{\mathrm{H}}\boldsymbol{Y}(k)-\boldsymbol{W}_{\mathrm{b}}^{*}\boldsymbol{S}_0(k)]\,|^2\}+\mu E\{(\boldsymbol{W}_{\mathrm{a}}^{\mathrm{H}}\boldsymbol{X}(k))^2\} \tag{9}$$

This cost function consists of a term $E\{(\boldsymbol{W}_{\mathrm{a}}^{\mathrm{H}}\boldsymbol{X}(k))^2\}$ related to the speech leakage, then

$$\begin{aligned}\boldsymbol{J}=&\min_{\boldsymbol{W}_{\mathrm{a}},\boldsymbol{W}_{\mathrm{b}}}E\{\,|\,[(\boldsymbol{W}_{\mathrm{q}}-\boldsymbol{B}\boldsymbol{W}_{\mathrm{a}})^{\mathrm{H}}\boldsymbol{Y}(k)-\boldsymbol{W}_{\mathrm{b}}^{*}\boldsymbol{S}_0(k)]\,|^2\}+\\&\mu E\{(\boldsymbol{W}_{\mathrm{a}}^{\mathrm{H}}\boldsymbol{X}(k))^2\}=\\&\min_{\boldsymbol{W}_{\mathrm{a}},\boldsymbol{W}_{\mathrm{b}}}E\{[(\boldsymbol{W}_{\mathrm{q}}-\boldsymbol{B}\boldsymbol{W}_{\mathrm{a}})^{\mathrm{H}}\boldsymbol{Y}(k)-\boldsymbol{W}_{\mathrm{b}}^{*}\boldsymbol{S}_0(k)]\cdot\\&[(\boldsymbol{W}_{\mathrm{q}}-\boldsymbol{B}\boldsymbol{W}_{\mathrm{a}})^{\mathrm{H}}\boldsymbol{Y}(k)-\boldsymbol{W}_{\mathrm{b}}^{*}\boldsymbol{S}_0(k)]^{\mathrm{H}}\}+\\&\mu E\{(\boldsymbol{W}_{\mathrm{a}}^{\mathrm{H}}\boldsymbol{X}(k))^2\}=\min_{\boldsymbol{W}_{\mathrm{a}},\boldsymbol{W}_{\mathrm{b}}}\boldsymbol{W}_{\mathrm{q}}^{\mathrm{H}}\boldsymbol{R}_Y\boldsymbol{W}_{\mathrm{q}}-\\&\boldsymbol{W}_{\mathrm{q}}^{\mathrm{H}}[\boldsymbol{R}_Y\ \boldsymbol{B}\boldsymbol{P}]\,\boldsymbol{W}_{\mathrm{c}}-\boldsymbol{W}_{\mathrm{c}}^{\mathrm{H}}\begin{bmatrix}\boldsymbol{B}^{\mathrm{H}}\boldsymbol{R}_Y\\ \boldsymbol{P}^{\mathrm{H}}\end{bmatrix}\boldsymbol{W}_{\mathrm{q}}+\boldsymbol{W}_{\mathrm{c}}^{\mathrm{H}}\boldsymbol{R}_Y\boldsymbol{W}_{\mathrm{c}}+\\&\mu E\{(\boldsymbol{W}_{\mathrm{a}}^{\mathrm{H}}\boldsymbol{X}(k))^2\}\end{aligned}$$

where

$$\boldsymbol{W}_{\mathrm{c}}=\begin{bmatrix}\boldsymbol{W}_{\mathrm{a}}\\ \boldsymbol{W}_{\mathrm{b}}\end{bmatrix}$$

$$\boldsymbol{R}_{\mathrm{c}}=E\left\{\begin{bmatrix}\boldsymbol{B}^{\mathrm{H}}\boldsymbol{Y}(k)\\ \boldsymbol{S}_0(k)\end{bmatrix}[\,\boldsymbol{Y}^{\mathrm{H}}(k)\ \ \boldsymbol{B}\quad\boldsymbol{S}_0^{*}(k)\,]\right\}=\begin{bmatrix}\boldsymbol{B}^{\mathrm{H}}\boldsymbol{R}_Y\boldsymbol{B}&0\\ \boldsymbol{0}^{\mathrm{H}}&\sigma_{S_0}^2\end{bmatrix}$$

$$\boldsymbol{P}=E\{\boldsymbol{Y}(k)\boldsymbol{S}_0^{*}(k)\}=\sigma_{S_0}^2\,\boldsymbol{a}(\theta_0)$$

in which $\sigma_{S_0}^2$ is the power of the transmitted desired signal.

From $\dfrac{\partial\boldsymbol{J}}{\partial\boldsymbol{W}_{\mathrm{a}}^{*}}=-\boldsymbol{B}^{\mathrm{H}}\boldsymbol{R}_Y\boldsymbol{W}_{\mathrm{q}}+\boldsymbol{B}^{\mathrm{H}}\boldsymbol{R}_Y\boldsymbol{B}\boldsymbol{W}_{\mathrm{a}}+\mu\boldsymbol{R}_X\boldsymbol{W}_{\mathrm{a}}=\boldsymbol{0}$, we obtain

$$\boldsymbol{W}_{\mathrm{a,\,opt}}=[\mu\boldsymbol{R}_X+\boldsymbol{B}^{\mathrm{H}}\boldsymbol{R}_Y\boldsymbol{B}]^{-1}\boldsymbol{B}^{\mathrm{H}}\boldsymbol{R}_Y\boldsymbol{W}_{\mathrm{q}} \tag{10}$$

where assuming that the speech and noise signals are uncorrelated, $\boldsymbol{R}_X$ can be estimated as

$$\boldsymbol{R}_X=\boldsymbol{R}_Y-\boldsymbol{R}_V$$

where $\boldsymbol{R}_Y$ is estimated during the periods of speech + noise and $\boldsymbol{R}_V$ during the periods of noise only.

From $\dfrac{\partial\boldsymbol{J}}{\partial\boldsymbol{W}_{\mathrm{b}}^{*}}=-\boldsymbol{W}_{\mathrm{q}}\boldsymbol{P}+\sigma_{S_0}^2\,\boldsymbol{W}_{\mathrm{b}}=\boldsymbol{0}$, we obtain

$$\boldsymbol{W}_{\mathrm{b,\,opt}}=\frac{\boldsymbol{P}^{\mathrm{H}}\boldsymbol{W}_{\mathrm{q}}}{\sigma_{S_0}^2}=\boldsymbol{a}^{\mathrm{H}}(\theta_0)\,\boldsymbol{W}_{\mathrm{q}} \tag{11}$$

The parameter μ gives the trade off between distortion of the speech reference and noise reduction. The greater the amount of speech leakage, the more attention is paid to speech distortion. For $\mu=0$, all emphasis is placed on the noise reduction and the speech leakage is not taken into account, which corresponds to the GSC-solution. Hence, the LCDF-GSC encompasses the GSC algorithm as a special case. For $\mu=\infty$, all emphasis is placed on speech leakage, $\boldsymbol{W}_{\mathrm{a}}=\boldsymbol{0}$, so $\boldsymbol{X}(k)$ is equal to the output of the fixed beamformer and no speech leakage is calculated, which corresponds to the delay-and-sum(DS) beamformer solution.

2.4 Analysis of output SNR

Let $J(\infty)$ denote the MSE in steady state of the adaptive algorithm:

$$J(\infty)=J_{\min}+J_{\mathrm{ex}}(\infty) \tag{12}$$

where $J_{\mathrm{ex}}(\infty)$ is the excess MSE of the adaptation.

$$\begin{aligned}\boldsymbol{J}_{\min}=&\min_{\boldsymbol{W}_{\mathrm{a}},\boldsymbol{W}_{\mathrm{b}}}E\{\,|\,[(\boldsymbol{W}_{\mathrm{q}}-\boldsymbol{B}\boldsymbol{W}_{\mathrm{a,\,opt}})^{\mathrm{H}}\boldsymbol{Y}(k)-\boldsymbol{W}_{\mathrm{b}}^{*}\boldsymbol{S}_0(k)]\,|^2\}+\\&\mu E\{(\boldsymbol{W}_{\mathrm{a,\,opt}}^{\mathrm{H}}\boldsymbol{X}(k))^2\}=\boldsymbol{W}_{\mathrm{q}}^{\mathrm{H}}\boldsymbol{R}_Y\boldsymbol{W}_{\mathrm{opt}}-\\&\sigma_{S_0}^2\,|\,\boldsymbol{W}_{\mathrm{opt}}^{\mathrm{H}}\boldsymbol{a}(\theta_0)\,|^2+\mu\boldsymbol{W}_{\mathrm{a,\,opt}}^{\mathrm{H}}\boldsymbol{R}_X\boldsymbol{W}_{\mathrm{a,\,opt}}=\\&\boldsymbol{W}_{\mathrm{q}}^{\mathrm{H}}\boldsymbol{R}_Y\boldsymbol{W}_{\mathrm{opt}}-\boldsymbol{P}_{\mathrm{S}}+\boldsymbol{P}_{\mathrm{L}}\end{aligned} \tag{13}$$

where $\boldsymbol{P}_{\mathrm{S}}=\sigma_{S_0}^2\,|\,\boldsymbol{W}_{\mathrm{opt}}^{\mathrm{H}}\boldsymbol{a}(\theta_0)\,|^2$, $\boldsymbol{P}_{\mathrm{L}}=\mu\boldsymbol{W}_{\mathrm{a,\,opt}}^{\mathrm{H}}\boldsymbol{R}_X\boldsymbol{W}_{\mathrm{a,\,opt}}$.

The output signal-to-noise ratio (SNR) can be written as

$$\mathrm{SNR}(k)=\frac{E\{\,|\,\boldsymbol{W}^{\mathrm{H}}(k)\boldsymbol{a}(\theta_0)\boldsymbol{S}_0(k)\,|^2\}}{E\{\,|\,\boldsymbol{W}^{\mathrm{H}}(k)\boldsymbol{Y}(k)-\boldsymbol{W}^{\mathrm{H}}(k)\boldsymbol{a}(\theta_0)\boldsymbol{S}_0(k)\,|^2\}+\mu E\{(\boldsymbol{W}_{\mathrm{a}}^{\mathrm{H}}\boldsymbol{X}(k))^2\}}=\frac{\boldsymbol{P}_{\mathrm{S}}}{\boldsymbol{W}_{\mathrm{q}}^{\mathrm{H}}\boldsymbol{R}_Y\boldsymbol{W}_{\mathrm{opt}}-\boldsymbol{P}_{\mathrm{S}}+\boldsymbol{P}_{\mathrm{L}}+\boldsymbol{J}(\infty)}$$

such that

$$\mathrm{SNR}_{\mathrm{LCDF\text{-}GSC}}=\lim_{k\to\infty}\mathrm{SNR}(k)=\frac{\boldsymbol{P}_{\mathrm{S}}}{\boldsymbol{W}_{\mathrm{q}}^{\mathrm{H}}\boldsymbol{R}_Y\boldsymbol{W}_{\mathrm{opt}}-\boldsymbol{P}_{\mathrm{S}}+\boldsymbol{P}_{\mathrm{L}}} \tag{14}$$

The output signal-to-noise ratio of DF-GSC and GSC can be respectively written as

$$\mathrm{SNR}_{\mathrm{DF\text{-}GSC}}=\frac{\boldsymbol{P}_{\mathrm{S}}}{\boldsymbol{W}_{\mathrm{q}}^{\mathrm{H}}\boldsymbol{R}_Y\boldsymbol{W}_{\mathrm{opt}}-\boldsymbol{P}_{\mathrm{S}}} \tag{15}$$

$$\mathrm{SNR}_{\mathrm{GSC}}=\frac{\boldsymbol{P}_{\mathrm{S}}}{\boldsymbol{J}_{\min(\mathrm{GSC})}-\boldsymbol{P}_{\mathrm{S}}}=\frac{\boldsymbol{P}_{\mathrm{S}}}{\boldsymbol{f}^{\mathrm{H}}(\boldsymbol{C}^{\mathrm{H}}\boldsymbol{R}_X^{-1}\boldsymbol{C})^{-1}\boldsymbol{f}-\boldsymbol{P}_{\mathrm{S}}} \tag{16}$$

Since in DF-GSC, $\boldsymbol{W}_{\mathrm{q}}^{\mathrm{H}}\boldsymbol{R}_Y\boldsymbol{W}_{\mathrm{opt}}-\boldsymbol{P}_{\mathrm{S}}\cong\boldsymbol{0}$, the MMSE of DF-GSC is smaller than that of the GSC. As a result, the output SNR of DF-GSC is higher than that of the GSC. For $\mu=0$, Eq. (10) shows that the solution of $\boldsymbol{W}_{\mathrm{a,\,opt}}$ is the same as that of DF-GSC, and so is the output SNR of LCDF-GSC. For $\mu=0$, $\boldsymbol{W}_{\mathrm{a}}=\boldsymbol{0}$, $\boldsymbol{X}(k)$ is equal to the output of the fixed beamformer, and the output SNR corresponds to the weight $\boldsymbol{W}_{\mathrm{q}}$.

3 Results

This section discusses the performance (SNR improvement and speech distortion) of the LCDF-GSC algorithm.

Segmental SNR is defined as

$$G_{\mathrm{SEGSNR}}=\frac{1}{L}\sum_{l=1}^{L-1}10\log\frac{\dfrac{1}{N}\sum_{n=0}^{N-1}s^2(n+Nl)}{\dfrac{1}{N}\sum_{n=0}^{N-1}[s(n+Nl)-\hat{s}(n+Nl)]^2} \tag{17}$$

where L is the frame length, $s(n)$ is the original signal, and $\hat{s}(n)$ is the estimated signal.

Speech distortion will be analyzed by considering

Itakura-saito distance between the speech component of the first microphone signal and the speech component of the considered signal. Itakura-saito distance is defined as

$$d_{\mathrm{IS}}(\vec{\alpha}_{\mathrm{d}}, \vec{\alpha}_{\phi}) = \left[\frac{\sigma_{\phi}^2}{\sigma_{\mathrm{d}}^2}\right]\left[\frac{\vec{\alpha}_{\mathrm{d}}\boldsymbol{R}_{\phi}\vec{\alpha}_{\mathrm{d}}}{\vec{\alpha}_{\phi}\boldsymbol{R}_{\phi}\vec{\alpha}_{\phi}}\right] + \log\left(\frac{\sigma_{\mathrm{d}}^2}{\sigma_{\phi}^2}\right) - 1 \quad (18)$$

where $\vec{\alpha}_{\phi}$ is an original clean frame of speech with linear prediction (LP) coefficient vector, and $\vec{\alpha}_{\mathrm{d}}$ is the processed speech coefficient vector. σ_{d}^2 and σ_{ϕ}^2 represent the all-pole gains for the processed and clean speech frames, respectively. We have calculated this distance with an LPC-order of 12.

The white noise from the Noisex-92 database was used for the objective evaluation of the proposed algorithm. The speech signals were four sentences taken from two female speakers and two male speakers, respectively. We compare the performance of the LCDF-GSC algorithm with DF-GSC and the single channel speech enhancement algorithm of the Weiner filter. In the experiment, we choose $N=4, \mu=0.01$ in Eq. (10).

Fig. 5 depicts the SNR improvement in white noise. It can be seen that the SNR improvement of multi-channel speech enhancement is higher than that of single-channel speech enhancement, and the SNR improvement of DF-GSC is higher than that of LCDF-GSC due to the introduction of speech leakage. However, from Fig. 6, it can be seen that the IS distance of DF-GSC is higher than the proposed algorithm and lower than that of the Wiener filter. So the proposed algorithm has lower distortion than the other two algorithms. Finally, we assume that there is a 2° difference between the estimated and the actual desired signal's DOA[9]. From Fig. 7 we can see that the mismatch affects conventional GSC more than LCDF-GSC in SNR improvement.

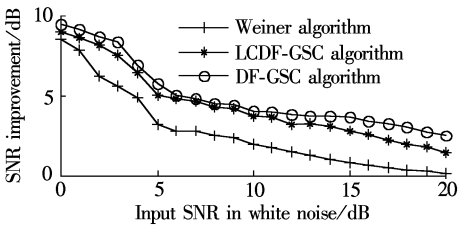

Fig. 5 SNR improvement in white noise

We adopted the MOS score to demonstrate the informal listening test results. Everyone listened to the four enhanced speech signals two times and gave eight scores. The results of the average score show that the speech enhanced by LCDF-GSC has more quality improvement than the other two algorithms.

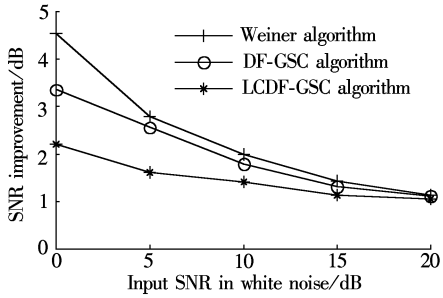

Fig. 6 IS distance measurement

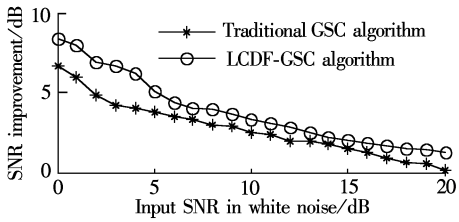

Fig. 7 SNR improvement in white noise with mismatch

4 Conclusion

In this paper, a new LCDF-GSC algorithm has been proposed for speech enhancement. With the introduction of leakage constraints, LCDF-GSC achieves less speech distortion than the DF-GSC algorithm and Wiener filtering, and simulation results also show that LCDF-GSC can maintain high SNR values even in mismatch than the traditional GSC algorithm.

References

[1] Sreenivas T V, Kimapure Pradeep. Codebook constrained Wiener filtering for speech enhancement [J]. *IEEE Transactions on Speech and Audio Processing*, 1996, **4**(5): 383 – 389.

[2] Hu Yi, Loizou Philipos C. A generalized subspace approach for enhancing speech corrupted by colored noise[J]. *IEEE Transactions on Speech and Audio Processing*, 2003, **11**(4): 334 – 340.

[3] Frost O L Ⅲ. An algorithm for linearly constrained adaptive array processing[J]. *Proceedings of IEEE*, 1972, **60**(8): 926 – 935.

[4] Griffiths L J, Jim C W. An alternative approach to linearly constrained adaptive beamforming[J]. *IEEE Trans Antennas Propag*, 1982, **30**(1): 27 – 34.

[5] Spriet Ann, Moonen Marc, Wouters Jan. Stochastic gradient-based implementation of spatially preprocessed speech distortion weighted multichannel Wiener filtering for noise reduction in hearing aids[J]. *IEEE Trans Signal Process*, 2005, **53**(3): 911 – 925.

[6] Hoshuyama O, Sugiyama A, Hirano A. A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters [J]. *IEEE Trans Signal*

Process, 1999, **47**(10): 2677 – 2683.

[7] Nordebo S, Claesson I, Nordholm S. Adaptive beamforming: spatial filter designed blocking matrix[J]. *IEEE Journal of Oceanic Engineering*, 1994, **19**(4): 583 – 590.

[8] Lee Yinman, Wu Wen-Rong. An LMS-based adaptive generalized sidelobe canceller with decision feedback[C]// *IEEE International Conference on Communications*. Seoul, Korea, 2005, **3**: 2047 – 2051.

[9] Lee Yinman, Wu Wen-Rong. A robust adaptive generalized sidelobe canceller with decision feedback[J]. *IEEE Trans Antennas Propag*, 2005, **53**(11): 3822 – 3832.

基于泄漏约束的 DF-GSC 语音增强

邹采荣[1,2]　陈国明[1]　赵　力[1]

([1] 东南大学信息科学与工程学院, 南京 210096)

([2] 佛山科学技术学院, 佛山 528000)

摘要: 为了改善广义旁瓣抵消(GSC)语音增强方法的性能,提出了一种带有泄漏约束的判决反馈旁瓣抵消(LCDF-GSC)方法. 采用 DF-GSC 方法以解决 GSC 对波达方向敏感的问题,在代价函数中引入泄漏因子,以此改善语音失真的问题,而这种问题是由于噪声参考信号中含有语音成份造成的. 试验结果表明,尽管经过 LCDF-GSC 处理后的语音信号信噪比要略低于 DF-GSC,IS 测度表明这时前者的语音信号失真度要小于后者. MOS 分也表明 LCDF-GSC 方法要优于 DF-GSC 和单通道 Weiner 滤波算法.

关键词: 语音增强;广义旁瓣抵消;语音泄漏

中图分类号: TN912.35