7

Low Distortion Noise Cancellers – Revival of a Classical Technique

Akihiko Sugiyama

NEC Corporation, Japan

This chapter presents low-distortion noise cancellers with their applications to communications and speech recognition. This classical technique, originally proposed by Widrow et al. in mid 70's, is first reviewed from a view point of output-signal distortion to show that interference and crosstalk are the primary reasons. As a solution to the interference problem, a paired filter (PF) structure introduces an auxiliary adaptive filter for estimating a signal-to-noise ratio (SNR) that is used to control the coefficient-adaptation stepsize in the main adaptive filter. A small stepsize for high SNRs, when the desired signal seriously interferes the misadjustment, provides steady and accurate change of coefficients, leading to low-distortion. This PF structure is extended to more general cases in which crosstalk from the desired-signal source to the auxiliary microphone is not negligible. A cross-coupled paired filter (CCPF) structure and its generalized version are solutions that employ another set of paired filters. The generalized CCPF (GCCPF) is applied to speech recognition in a human-robot communication scenario where improvement in distortion is successfully demonstrated by evaluations in the real environment. This robot had been demonstrated for six months at 2005 World Exposition in Aichi, Japan.

7.1 Introduction

Adaptive noise cancellers (ANCs) were first proposed by Widrow et al. for two-microphone speech enhancement [22]. Widrow's ANC has two microphones. The primary microphone captures a mixture of a desired signal and noise. A secondary (or reference) microphone is placed sufficiently close to the noise source to pick up a reference noise. The reference noise drives an adaptive filter to generate a noise replica at the primary microphone. By subtracting the noise replica from the primary microphone signal, an enhanced speech is obtained as the output. The output contains residual noise and is used for coefficient adaptation. Due to the auxiliary information by the reference

microphone, Widrow's ANC is a more effective technique, especially in low SNR environments, than single-microphone noise suppressors based on, for example, spectral subtraction (SS) [2] and minimum mean squared error short time spectral amplitude (MMSE STSA) estimation [6]. However, the quality of the output speech may be degraded for two reasons, namely, interference and crosstalk.

Interference to coefficient adaptation is caused by the desired signal. Coefficients should be updated by the residual noise that is the difference between the noise at the primary microphone and the noise replica. However, the residual noise cannot be obtained separately from the desired signal. It serves as an interference when the error, which is composed of the residual noise and the desired signal, is used for coefficient adaptation of the ANC. As a result, the performance of the ANC is limited [22] with insufficient noise cancellation and distortion in the desired signal.

Crosstalk happens at the reference microphone. In Widrow's ANC, it is assumed that the reference microphone is placed sufficiently close to the noise source. It is necessary for the reference microphone to pick up only the noise. However, this assumption is often violated in reality. There are desired-signal components that leak into the reference microphone. Such a leak-in signal is called crosstalk. Crosstalk contaminates the reference noise that is the adaptive-filter input in Widrow's ANC. The adaptive-filter output is no longer a good replica of the noise and the ANC output may be distorted.

This chapter presents low-distortion noise cancellers as solutions to the interference and the crosstalk problems. In Sec. 7.2, interference and crosstalk are investigated from a viewpoint of distortion in the output signal. A solution to the interference problem is presented in Sec. 7.3. Some ANCs are described in Sec. 7.4 in search of a good structure for the crosstalk problem. Finally, Secs. 7.5 and 7.6 are devoted to more advanced ANC structures for successful applications.

7.2 Distortions in Widrow's Adaptive Noise Canceller

7.2.1 Distortion by Interference

Fig. 7.1 shows a block diagram of Widrow's ANC. $s(n)$, $n(n)$ and $n_1(n)$ are the signal, the noise, and the noise component in the primary-microphone signal, all with a time index n. $\boldsymbol{h}(n)$ represents the impulse response of the noise path from the noise source to the primary microphone. The primary signal $x_{\mathrm{P}}(n)$ and the reference signal $x_{\mathrm{R}}(n)$ can be written as

$$x_{\mathrm{P}}(n) = s(n) + n_1(n), \tag{7.1}$$

$$x_{\mathrm{R}}(n) = n(n). \tag{7.2}$$

The output $e_1(n)$ of the ANC is given by

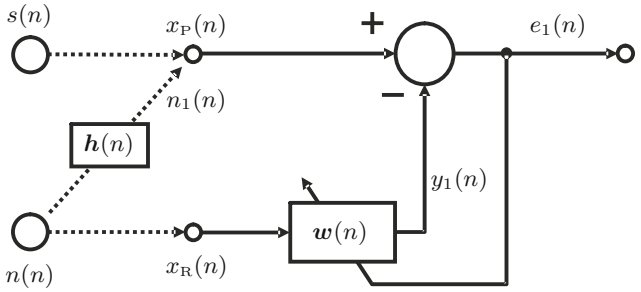


Fig. 7.1. Widrow's adaptive noise canceller.

$$e_1(n) = s(n) + n_1(n) - y_1(n), \tag{7.3}$$

$$n_1(n) = \boldsymbol{h}^{\mathrm{T}}(n)\,\boldsymbol{n}(n), \tag{7.4}$$

$$y_1(n) = \boldsymbol{w}^{\mathrm{T}}(n)\,\boldsymbol{x}_{\mathrm{R}}(n) = \boldsymbol{w}^{\mathrm{T}}(n)\,\boldsymbol{n}(n), \tag{7.5}$$

where $n_1(n)$ and $y_1(n)$ are the noise component in $x_{\mathrm{P}}(n)$ and the output of the adaptive filter. $\boldsymbol{w}(n)$ is the coefficient vector of the adaptive filter, and $\boldsymbol{h}(n)$ is the impulse-response vector of the noise path. $\boldsymbol{x}_{\mathrm{R}}(n)$ and $\boldsymbol{n}(n)$ are the reference signal and the noise vectors with a size of N. These vectors are defined by

$$\boldsymbol{h}^{\mathrm{T}}(n) = \big[h_0(n),\, h_1(n),\, \cdots,\, h_{N-1}(n)\big], \tag{7.6}$$

$$\boldsymbol{w}^{\mathrm{T}}(n) = \big[w_0(n),\, w_1(n),\, \cdots,\, w_{N-1}(n)\big], \tag{7.7}$$

$$\boldsymbol{n}^{\mathrm{T}}(n) = \big[n(n),\, n(n-1),\, \cdots,\, n(n-N+1)\big], \tag{7.8}$$

$$\boldsymbol{x}_{\mathrm{R}}^{\mathrm{T}}(n) = \big[x_{\mathrm{R}}(n),\, x_{\mathrm{R}}(n-1),\, \cdots,\, x_{\mathrm{R}}(n-N+1)\big] = \boldsymbol{n}^{\mathrm{T}}(n). \tag{7.9}$$

Eqs. 7.3 – 7.5 reduce to

$$e_1(n) = s(n) + \big[\boldsymbol{h}(n) - \boldsymbol{w}(n)\big]^{\mathrm{T}}\,\boldsymbol{n}(n). \tag{7.10}$$

Assuming that the noise path $\boldsymbol{h}(n)$ is estimated with the NLMS algorithm [9], the update of $\boldsymbol{w}(n)$ is performed by

$$\boldsymbol{w}(n+1) = \boldsymbol{w}(n) + \frac{\mu\, e_1(n)\, \boldsymbol{x}_{\mathrm{R}}(n)}{\big\|\boldsymbol{x}_{\mathrm{R}}(n)\big\|^2} = \boldsymbol{w}(n) + \frac{\mu\, e_1(n)\, \boldsymbol{n}(n)}{\big\|\boldsymbol{n}(n)\big\|^2}, \tag{7.11}$$

where μ is a stepsize. From Eqs. 7.3 and 7.10, it can be seen that $e_1(n)$ is close to $s(n)$ when $y_1(n) \approx n_1(n)$ or equivalently, $\boldsymbol{h}(n) \approx \boldsymbol{w}(n)$ near convergence. Because $e_1(n)$ is the output, the desired signal $s(n)$ is obtained at the output after convergence. However, the misadjustment $y_1(n) - n_1(n)$ which is needed for coefficient adaptation is severely contaminated by the desired signal $s(n)$, resulting in signal-distortion such as reverberation [3]. This distortion can be removed if the coefficient update is performed only in the absence of the desired signal $s(n)$. However, it requires an accurate speech detector.

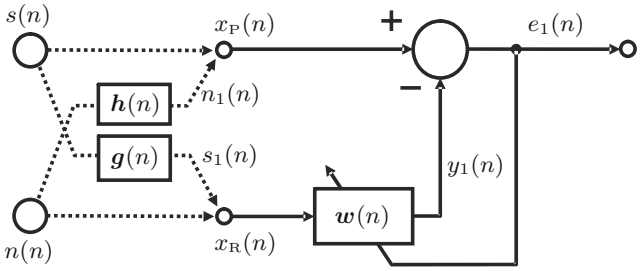


Fig. 7.2. Widrow's adaptive noise canceller with crosstalk.

7.2.2 Distortion by Crosstalk

When there is crosstalk, Eq. 7.2 does not hold anymore because a crosstalk term $s_1(n)$ should be considered as depicted in Fig. 7.2. In the presence of crosstalk, $x_{\mathrm{R}}(n)$ is expressed by

$$x_{\mathrm{R}}(n) = n(n) + s_1(n), \tag{7.12}$$

where

$$s_1(n) = \boldsymbol{g}^{\mathrm{T}}(n)\, \boldsymbol{s}(n), \tag{7.13}$$
$$\boldsymbol{g}^{\mathrm{T}}(n) = \big[g_0(n),\, g_1(n),\, \cdots,\, g_{N-1}(n)\big]. \tag{7.14}$$

In reference to Eqs. 7.2, 7.12, and 7.5, the output $y_1(n)$ with crosstalk is given by

$$\begin{aligned} y_1(n) &= \boldsymbol{w}^{\mathrm{T}}(n)\, \boldsymbol{x}_{\mathrm{R}}(n) + \boldsymbol{w}^{\mathrm{T}}(n)\, \boldsymbol{s}_1(n), \\ &= \boldsymbol{w}^{\mathrm{T}}(n)\, \big[\boldsymbol{n}(n) + \boldsymbol{s}_1(n)\big], \end{aligned} \tag{7.15}$$
$$\boldsymbol{s}_1^{\mathrm{T}}(n) = \big[s_1(n),\, s_1(n-1),\, \cdots,\, s_1(n-N+1)\big]. \tag{7.16}$$

Eqs. 7.3, 7.4, and 7.15 result in

$$e_1(n) = s(n) + \big[\boldsymbol{h}(n) - \boldsymbol{w}(n)\big]^{\mathrm{T}}\, \boldsymbol{n}(n) - \boldsymbol{w}^{\mathrm{T}}(n)\, \boldsymbol{s}_1(n). \tag{7.17}$$

When $\boldsymbol{w}(n)$ perfectly identifies $\boldsymbol{h}(n)$, *i.e.* $\boldsymbol{w}(n) = \boldsymbol{h}(n)$, Eq. 7.17 reduces to

$$e_1(n) = s(n) - \boldsymbol{h}^{\mathrm{T}}(n)\, \boldsymbol{s}_1(n), \tag{7.18}$$

which is not equal to the desired signal. In addition, Eqs. 7.16 and 7.18 suggest that past activities of $s(n)$ affect the output $e_1(n)$. Therefore, the output $e_1(n)$ is not equal to the desired signal $s(n)$, resulting in distortion, unless $\boldsymbol{h}(n)$ or $\boldsymbol{s}_1(n)$ is a zero vector.

For the environment with crosstalk, multi-stage ANCs [5, 8, 13] have been developed as extensions of Widrow's ANC. They all try to perform coefficient

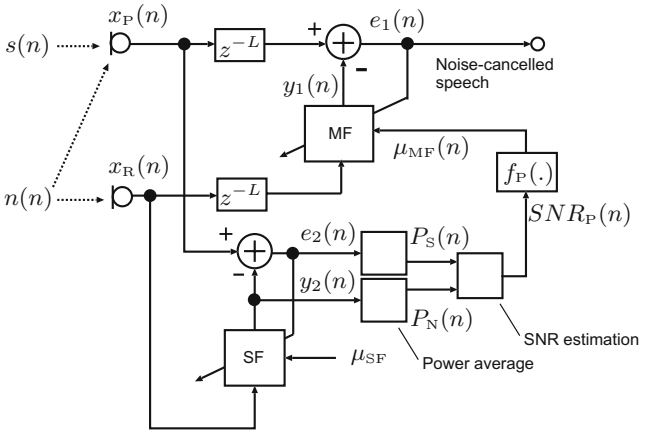


Fig. 7.3. ANC with a paired filter (PF) structure.

adaptation during absence of the desired signal so that the undesirable effect of the crosstalk does not appear neither in the input signal nor in the error that is used for coefficient adaptation. Eq. 7.2 holds in this case. It implies that a sufficiently accurate signal detector for $s(n)$ is essential, which is not available in reality. In addition, this strategy does not work in the presence of $s(n)$. The output is still expressed by Eq. 7.18. Therefore, multi-stage ANCs [5,8,13] are not effective for the crosstalk problem.

7.3 Paired Filter (PF) Structure

A paired filter (PF) structure [11] has been developed as a solution to the interference problem. Fig. 7.3 shows an ANC with a paired filter structure, which consists of two adaptive filters, namely, a main filter (MF) and a subfilter (SF). They operate in parallel to generate noise replicas. The SF is used for estimating the signal-to-noise ratio (SNR) of the primary input. The stepsize for the MF is controlled based on the estimated SNR.

7.3.1 Algorithm

7.3.1.1 SNR Estimation by Subfilters

The SF works in the same way as Widrow's ANC. Filter coefficients are updated by the NLMS algorithm [9]. The stepsize, μ_{SF}, is set large and fixed for fast convergence and rapid tracking of the noise-path change. A small value of μ_{SF} results in more precise estimation. For the estimation of SNR, an average power of the noise replica, $P_N(n)$, and the power of the error signal, $P_S(n)$, are calculated by

$$P_{\mathrm{N}}(n) = \sum_{j=0}^{M-1} y_2^2(n-j), \tag{7.19}$$

$$\begin{aligned} P_{\mathrm{S}}(n) &= \sum_{j=0}^{M-1} \big(x_{\mathrm{P}}(n-j) - y_2(n-j)\big)^2, \\ &= \sum_{j=0}^{M-1} e_2^2(n-j), \end{aligned} \tag{7.20}$$

where $y_2(n)$ is the SF output. M is the number of samples used for calculating $P_{\mathrm{N}}(n)$ and $P_{\mathrm{S}}(n)$. From $P_{\mathrm{S}}(n)$ and $P_{\mathrm{N}}(n)$, an estimated signal-to-noise ratio, $SNR_{\mathrm{P}}(n)$, of the primary signal is calculated by

$$SNR_{\mathrm{P}}(n) = 10\ \log_{10}\left\{\frac{P_{\mathrm{S}}(n)}{P_{\mathrm{N}}(n)}\right\}\ \mathrm{dB}. \tag{7.21}$$

7.3.1.2 Stepsize Control for the Main Filter

The stepsize for the MF is controlled by the estimated signal-to-noise ratio, $SNR_{\mathrm{P}}(n)$. If $SNR_{\mathrm{P}}(n)$ is low, the stepsize is set large for fast convergence because low $SNR_{\mathrm{P}}(n)$ means small interference for the coefficient adaptation. Otherwise, the stepsize is set small for smaller signal-distortion in the ANC output. The following equation shows a function which determines the stepsize, $\mu_{\mathrm{MF}}(n)$, based on $SNR_{\mathrm{P}}(n)$:

$$\mu_{\mathrm{MF}}(n) = \begin{cases} \mu_{\mathrm{M_{min}}}, & \text{if } SNR_{\mathrm{P}}(n) > SNR_{\mathrm{P_{max}}}, \\ \mu_{\mathrm{M_{max}}}, & \text{if } SNR_{\mathrm{P}}(n) < SNR_{\mathrm{P_{min}}}, \\ f_{\mathrm{P}}\big(SNR_{\mathrm{P}}(n)\big), & \text{otherwise}, \end{cases} \tag{7.22}$$

where $\mu_{\mathrm{M_{max}}}$, $\mu_{\mathrm{M_{min}}}$ and $f_{\mathrm{P}}(\cdot)$ are the maximum and the minimum stepsizes and a function of $SNR_{\mathrm{P}}(n)$, respectively. It is natural that $f_{\mathrm{P}}(\cdot)$ is a decreasing function since a small stepsize is suitable for a large SNR. For simplicity, let us assume that $f_{\mathrm{P}}(\cdot)$ is a first-order function of $SNR_{\mathrm{P}}(n)$. Then, it may be given by

$$f_{\mathrm{P}}\big(SNR_{\mathrm{P}}(n)\big) = A \cdot SNR_{\mathrm{P}}(n) + B, \tag{7.23}$$

where A ($A < 0$) and B are constants. $\mu_{\mathrm{M_{min}}}$ determines the signal distortion in the utterance. If $\mu_{\mathrm{M_{min}}}$ is set to zero, the adaptation is skipped when $SNR_{\mathrm{P}}(n)$ is higher than $SNR_{\mathrm{P_{max}}}$. In this case, this algorithm works as the adaptation-stop method [10] with a speech detector.

7.3.1.3 Delay Compensation for the Main Filter

The estimated SNR, $SNR_{\mathrm{P}}(n)$, is given with a time delay, which depends on the number of samples M used for calculation of $P_{\mathrm{N}}(n)$ and $P_{\mathrm{S}}(n)$. This time

delay directly raises the signal distortion in the processed speech because the stepsize remains large in the beginning of the utterance. To compensate for this delay, the delay unit z^{-L} is incorporated only for the MF. L is set to $M/2$ since the time delay is $M/2$.

7.3.2 Evaluations

The performance of the ANC with a paired filter structure was evaluated in comparison with that of a variable stepsize algorithm [18] assuming a communication scenario in a military tank. The tank operators wear headsets with a reference microphone on the earpiece. A diesel-engine noise recorded in a tank was used as a noise source. Shown in Fig. 7.4 at the top is a noise-path impulse response measured in a room with a dimension of 3.05 m (width) × 2.85 m (depth) × 1.80 m (height). In order to evaluate the tracking capability, the polarity of the noise path was inverted at 12.5 sec.[1] The noise component, which was generated by convolution of the noise source with the noise-path, was added to the speech source to obtain the noise-corrupted signal. This signal contains the uncorrelated noise component which should exist in the recording environment. The sampling frequency was 8 kHz and other parameter values are shown in Tab. 7.1. Parameters for the variable stepsize algorithm were adjusted such that fast convergence and the final misadjustment equivalent to that of the ANC with the PF structure are obtained.

Table 7.1. Parameters and corresponding values.

Parameter	Value
N	64
M	128
L	64
μ_{SF}	0.4
$\mu_{\mathrm{M}_{\max}}$	0.4
$\mu_{\mathrm{M}_{\min}}$	2^{-6}
$SNR_{\mathrm{P}_{\max}}$	$-10\,\mathrm{dB}$
$SNR_{\mathrm{P}_{\min}}$	$-50\,\mathrm{dB}$
A	-0.01
B	0.4

[1] This is nothing more than an example. An abrupt polarity change was imposed as an extreme example. Path changes in the real environment are slower and less significant, thus, easier to track.

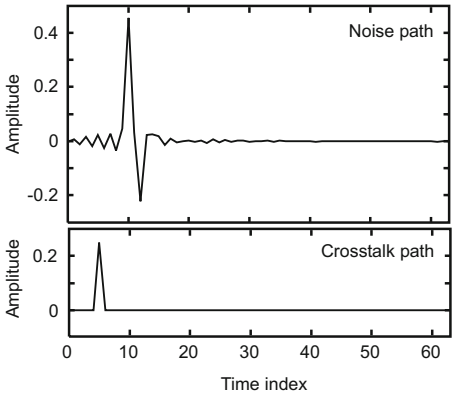


Fig. 7.4. Impulse responses of noise and crosstalk paths.

7.3.2.1 Objective Evaluations

Fig. 7.5 illustrates the desired signal, the primary signal and the output signal. The SNR in the primary signal was around 0 dB in the utterance. The ANC with the PF structure successfully cancels the noise and tracks the noise-path change at 12.5 sec.

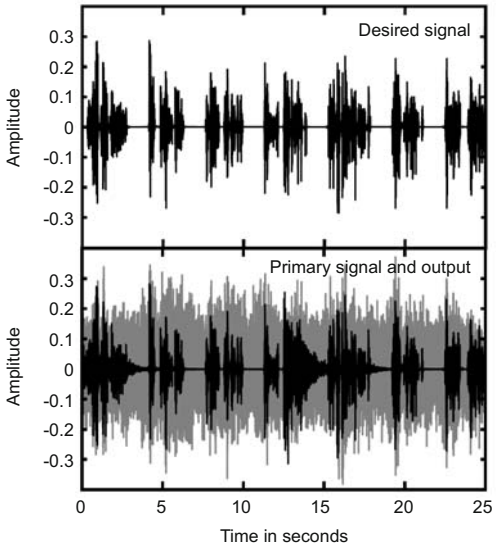


Fig. 7.5. Desired signal (upper diagram), primary signal (gray, lower diagram), and output speech signal (black, lower diagram).

The original SNR and the SNR estimated by the SF are compared in part (a) and (b) of Fig. 7.6. Since the peaks of the estimated SNR approximate those of the original SNR in a good manner, it is considered reliable. Part (c) of Fig. 7.6 exhibits the stepsize behaviors of the ANC with the PF structure and that of the variable stepsize algorithm. The stepsize of the ANC with the PF structure remains small in the utterance, while the other becomes larger in the beginning of the utterance.

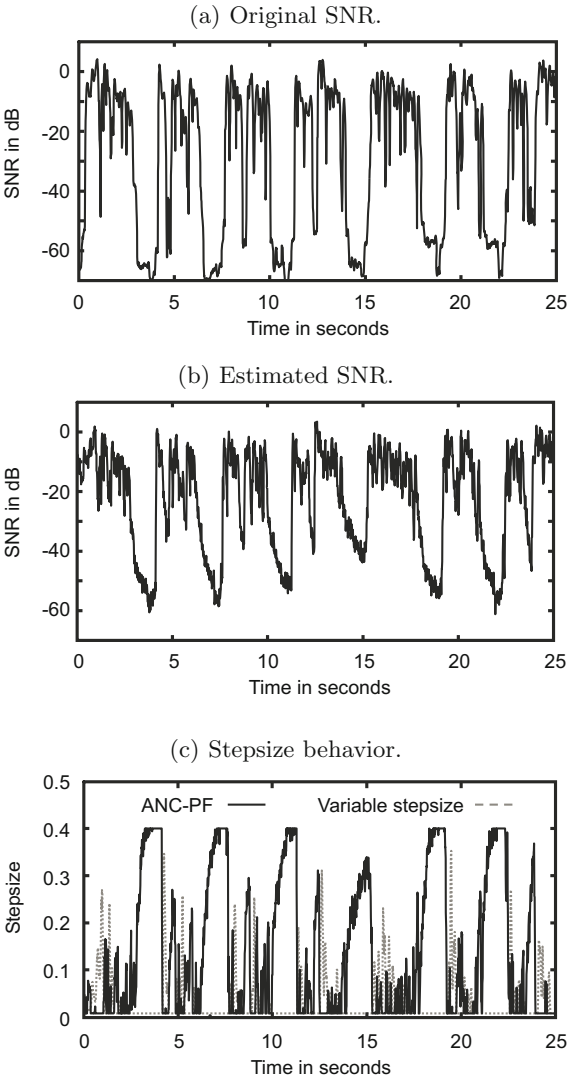


Fig. 7.6. Original (a) and estimated SNR (b), as well as the stepsize behavior (c).

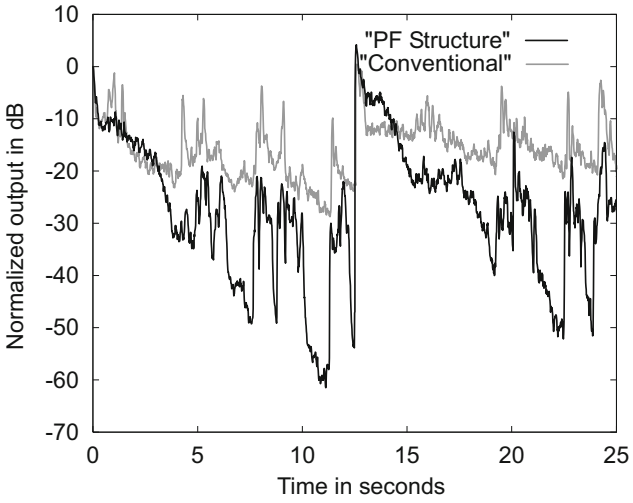


Fig. 7.7. Normalized output.

Fig. 7.7 shows a normalized output $\epsilon(n)$ defined by

$$\epsilon(n) = 10\ \log_{10} \left\{ \frac{\sum_{j=0}^{M_\epsilon - 1} e_{\mathrm{I}}^2(n-j)}{\sum_{j=0}^{M_\epsilon - 1} x_{\mathrm{P}}^2(n-j)} \right\} \text{dB}, \tag{7.24}$$

where M_ϵ is the number of samples for average and was set to 512. In case of good noise cancellation, $\epsilon(n)$ should take a large negative value in nonspeech sections. The normalized output of the ANC with the PF structure is approximately 10 dB smaller in the utterance compared with the variable stepsize algorithm.

Fig. 7.8 depicts signal distortion $\delta(n)$ in the output defined by

$$\delta(n) = 10\ \log_{10} \left\{ \frac{\sum_{j=0}^{M_\delta - 1} \big(e_1(n-j) - s(n-j)\big)^2}{\sum_{j=0}^{M_\delta - 1} s^2(n-j)} \right\}\ \text{dB}, \tag{7.25}$$

where M_δ is the number of samples for average and was set to 512. The ANC with the PF structure reduces the signal distortion by up to 15 dB compared with the variable stepsize algorithm in the utterance.

Similar performance of the ANC with the PF structure has been confirmed for ± 6 dB and 0 dB SNR with respect to the estimated SNR, the normalized

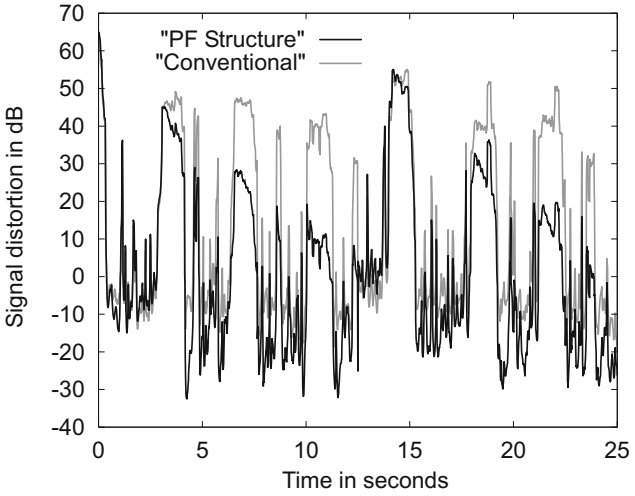


Fig. 7.8. Signal distortion.

output $\epsilon(n)$, and the signal distortion $\delta(n)$ [11]. The performance does not change for different SNRs.[2]

7.3.2.2 Subjective Evaluation

To evaluate the subjective performance of the ANC with a paired filter structure, a listening test was carried out. Widrow's ANC with the NLMS algorithm [9] and the ANC with the PF structure were compared for the case where the SNR in the primary signal is 0 dB. The 5-point mean opinion scores (MOS's) were given by 20 listeners. A noise-free (*i.e.* clean) speech sample and a noisy speech sample before noise-cancellation were included as the highest and the lowest anchors. The same speech source as in Sec. 7.3.2.1 was employed for the subjective test. Fig. 7.9 shows the subjective evaluation results. The vertical line centered in the shaded area and the numeral represent the mean value of the MOS. The width of the shaded area corresponds to the standard deviation. The mean values of the MOS for the ANC with the PF structure are higher than Widrow's ANC by about 1 point.

7.4 Crosstalk Resistant ANC and Cross-Coupled Structure

Crosstalk resistant ANC (CTRANC) [14, 15, 23] and an ANC with a cross-coupled structure [1] have been developed independently for crosstalk resistance.

[2] For hardware implementation and evaluation, please refer to [11].

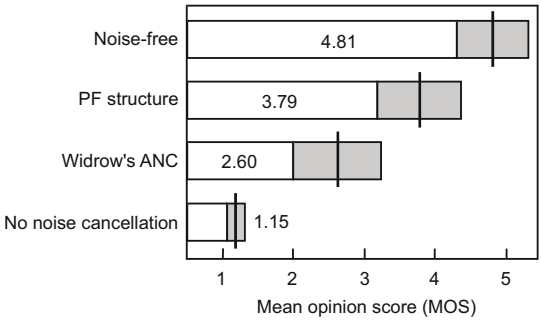


Fig. 7.9. Results of the subjective evaluation.

They both employ an auxiliary filter for crosstalk, however, its usage is different from that in the multi-stage ANCs [5, 8, 13].

7.4.1 Crosstalk Resistant ANC

Fig. 7.10 depicts a block diagram of CTRANC. The secondary adaptive filter, F2, is driven by the ANC output $e_1(n)$ to generate a crosstalk replica. When the ANC operation is ideal, its output should be equal to the desired signal. It suggests that the ANC output could be used as a replica of the desired signal. The output $y_3(n)$ of F2, approximating the crosstalk, is subtracted from the reference input $x_{\mathrm{R}}(n)$. The result $e_3(n)$ is used as the input of the primary adaptive filter, F1. Because the crosstalk components that were originally contained in the reference input are cancelled by F2, F1 has little crosstalk contamination in its input. Therefore, F1 works as if there were no crosstalk.

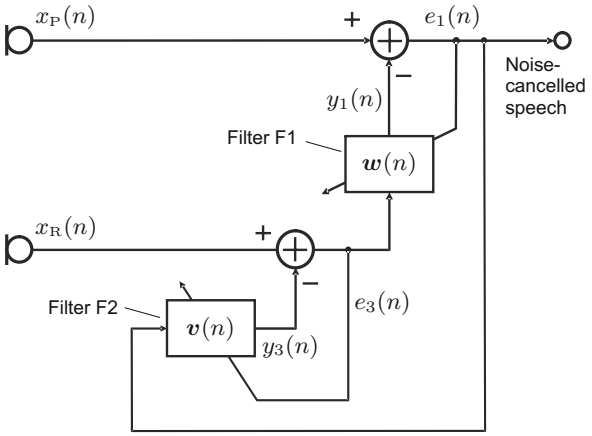


Fig. 7.10. Crosstalk resistant adaptive noise canceller (CTRANC).

Referring to Fig. 7.10, the input of F1 represented by $e_3(n)$ is given by

$$e_3(n) = x_{\mathrm{R}}(n) - \boldsymbol{v}^{\mathrm{T}}(n)\,\boldsymbol{e}_1(n). \tag{7.26}$$

$$e_1(n) = x_{\mathrm{P}}(n) - \boldsymbol{w}^{\mathrm{T}}(n)\,\boldsymbol{e}_3(n), \tag{7.27}$$

From Eqs. 7.1, 7.4, 7.12, 7.13, 7.26, and 7.27, the following equations are obtained:

$$e_1(n) = s(n) + \boldsymbol{h}^{\mathrm{T}}(n)\,\boldsymbol{n}(n) - \boldsymbol{w}^{\mathrm{T}}(n)\,\boldsymbol{e}_3(n), \tag{7.28}$$

$$e_3(n) = \boldsymbol{g}^{\mathrm{T}}(n)\,\boldsymbol{s}(n) + n(n) - \boldsymbol{v}^{\mathrm{T}}(n)\,\boldsymbol{e}_1(n), \tag{7.29}$$

where

$$\boldsymbol{e}_1^{\mathrm{T}}(n) = \big[e_1(n-1),\, \cdots,\, e_1(n-N+1),\; e_1(n-N)\big], \tag{7.30}$$

$$\boldsymbol{e}_3^{\mathrm{T}}(n) = \big[e_3(n),\, e_3(n-1),\, \cdots,\, e_3(n-N+1)\big], \tag{7.31}$$

$$\boldsymbol{v}^{\mathrm{T}}(n) = \big[v_0(n),\, v_1(n),\, \cdots,\, v_{N-1}(n)\big]. \tag{7.32}$$

$\boldsymbol{v}(n)$ is the coefficient vector of F2. It should be noted that the elements of $\boldsymbol{e}_1(n)$ are one-sample shifted to the left. This is because $e_1(n)$ is not available when $y_3(n)$ is calculated.[3]

For perfect cancellation of the crosstalk $s_1(n)$, $e_3(n) = n(n)$ should be satisfied. Applying this condition to Eqs. 7.28 and 7.29 leads to

$$e_1(n) = s(n), \tag{7.33}$$

$$\boldsymbol{w}(n) = \boldsymbol{h}(n), \tag{7.34}$$

$$\boldsymbol{v}(n) = \boldsymbol{g}(n). \tag{7.35}$$

Eq. 7.33 implies that the output $e_1(n)$ theoretically has no distortion.

7.4.2 Cross-Coupled Structure

An ANC with a cross-coupled structure [1] is an equivalent form to CTRANC. The cross-coupled structure is illustrated in Fig. 7.11. It has a paired structure with dedicated adaptive filters for noise and crosstalk paths. Combining Eqs. 7.1 and 7.2, with Fig. 7.11, it is straightforward to derive Eqs. 7.28 and 7.29. It is clearly seen in Fig. 7.11 that the filters F1 and F2 are cooperating to make the input of its counterpart less contaminated by crosstalk or noise. Therefore, if one works better with a clean input, the other also works better with its own input that is cleaner. Finally, both F1 and F2 operate with crosstalk-free and noise-free inputs, respectively, which is the ideal situation.

[3] This fact implies that $v_0(n)$, the first element of $\boldsymbol{v}(n)$, approximates $g(1)$ in Fig. 7.2. Although $g(0)$ cannot be modeled by $\boldsymbol{v}(n)$, it does not cause a problem as far as there is a one-sample delay in the crosstalk path. This is usually the case in practice.

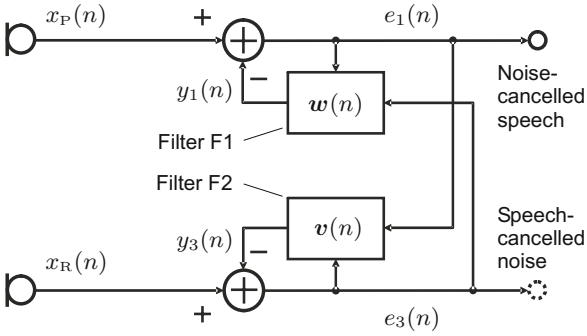


Fig. 7.11. ANC with a cross-coupled structure.

However, [1] points out that the output of the cross-coupled ANC may not be sufficiently good. This is because the desired signal in the output interferes the coefficient adaptation in F1. Similarly, the noise in $e_3(n)$ interferes the coefficient adaptation in F2. The ANC structure by itself does not make a good solution to the crosstalk problem. From this viewpoint, integration of a good coefficient adaptation algorithm and a good ANC structure is essential. Such integrations are discussed in the following sections.

7.5 Cross-Coupled Paired Filter (CCPF) Structure

To cancel crosstalk caused by the desired signal in the reference input, the cross-coupled paired filter (CCPF) structure [12] employs a cross-coupled structure [1] in which cross-coupled adaptive filters should cancel the noise component in the primary signal and the crosstalk in the reference signal simultaneously. For the interference problem the paired filter structure is extended to the cross-coupled structure.

7.5.1 Algorithm

Fig. 7.12 depicts a block diagram of an ANC with a CCPF structure. The filters MF1 and MF2 form the cross-coupled structure. The filters SF1 and SF2 make a pair with MF1 and MF2 for the interference resistance.

MF1 and SF1 take the roles of the main filter (MF) and the subfilter (SF) in the PF structure. Another set of paired filters, namely MF2 and SF2, operate in the same manner as MF1 and SF1 except that they try to cancel the crosstalk instead of the noise. The stepsizes of MF1 and MF2 are controlled with the help of SF1 and SF2 in a similar way to that in MF in the PF structure. When no crosstalk is present, this structure in Fig. 7.12 works as that in Fig. 7.3. Because there are no correlated components to $s(n)$ in either $e_3(n)$ or $e_4(n)$, the coefficients of SF2 and MF2 do not grow from the initial values (*i.e.* zero).

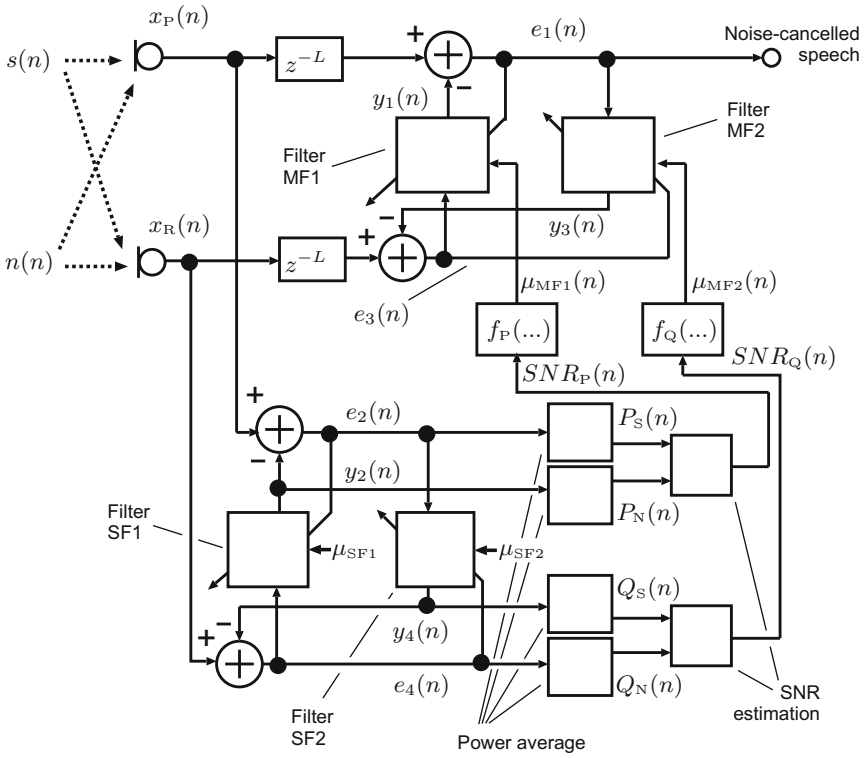


Fig. 7.12. ANC with a cross-coupled paired filter (CCPF) structure.

7.5.1.1 SNR Estimation by Subfilters

SF1 carries out SNR estimation for the primary signal in the same way as in the PF structure by Eqs. 7.19 – 7.21. SF2, on the other hand, estimates the SNR for the reference signal $x_R(n)$ by generating a replica of the desired signal component (crosstalk) in the reference signal. $SNR_R(n)$ of the reference signal can be estimated using a replica of the desired signal at the subtractor output and the estimated noise. The crosstalk estimated by SF2 is generally a speech signal, which naturally contains silent sections. When the crosstalk is small, the signal component to be estimated may be much smaller than the noise component. To perform a stable SNR estimation, the stepsize for the coefficient adaptation μ_{SF2} is therefore set smaller than the stepsize for SF1.

The average power $Q_S(n)$ of the desired-signal replica and the average power $Q_N(n)$ of the error signal of SF2 can be calculated using the following equations:

$$Q_{\mathrm{S}}(n) = \sum_{j=0}^{M-1} y_4^2(n-j), \tag{7.36}$$

$$\begin{aligned} Q_{\mathrm{N}}(n) &= \sum_{j=0}^{M-1} \Big(x_{\mathrm{R}}(n-j) - y_4(n-j)\Big)^2, \\ &= \sum_{j=0}^{M-1} e_4^2(n-j), \end{aligned} \tag{7.37}$$

where $y_4(n)$, $e_4(n)$ are respectively the desired-signal replica which is the output of SF2 and the reference signal. M is the number of samples averaged for calculation of $Q_{\mathrm{S}}(n)$ and $Q_{\mathrm{N}}(n)$. From $Q_{\mathrm{S}}(n)$ and $Q_{\mathrm{N}}(n)$, the signal-to-noise ratio $SNR_{\mathrm{Q}}(n)$ can be obtained by

$$SNR_{\mathrm{Q}}(n) = 10\ \log_{10}\left\{\frac{Q_{\mathrm{S}}(n)}{Q_{\mathrm{N}}(n)}\right\} \mathrm{dB}. \tag{7.38}$$

7.5.1.2 Stepsize Control in the Main Filters

The adjustment of the stepsize $\mu_{\mathrm{MF1}}(n)$ in MF1 is controlled based on the estimated $SNR_{\mathrm{P}}(n)$ in the same manner as in the PF structure. $\mu_{\mathrm{MF1}}(n)$ is determined by Eqs. 7.21 and 7.22 where $\mu_{\mathrm{MF}}(n)$, $\mu_{\mathrm{M}_{\max}}$, and $\mu_{\mathrm{M}_{\min}}$ should be replaced with $\mu_{\mathrm{MF1}}(n)$, $\mu_{\mathrm{M1}_{\max}}$, and $\mu_{\mathrm{M1}_{\min}}$. The stepsize $\mu_{\mathrm{MF2}}(n)$ is controlled by the estimated $SNR_{\mathrm{Q}}(n)$. When $SNR_{\mathrm{Q}}(n)$ is low, the stepsize is set small since there is a large noise component which interferes with MF2 in the crosstalk estimation. On the other hand, when $SNR_{\mathrm{Q}}(n)$ is high with a large crosstalk component, the stepsize is set large. $\mu_{\mathrm{MF2}}(n)$ is determined by Eq. 7.39 based on $SNR_{\mathrm{Q}}(n)$:

$$\mu_{\mathrm{MF2}}(n) = \begin{cases} \mu_{\mathrm{M2}_{\min}}, & \text{if } SNR_{\mathrm{Q}}(n) < SNR_{\mathrm{Q}_{\min}}, \\ \mu_{\mathrm{M2}_{\max}}, & \text{if } SNR_{\mathrm{Q}}(n) > SNR_{\mathrm{Q}_{\max}}, \\ f_{\mathrm{Q}}\big(SNR_{\mathrm{Q}}(n)\big), & \text{otherwise}, \end{cases} \tag{7.39}$$

where $\mu_{\mathrm{M2}_{\max}}$ and $\mu_{\mathrm{M2}_{\min}}$ are the maximum and the minimum values of the stepsize, respectively. It is desirable that $f_{\mathrm{Q}}(\cdot)$ is a monotonically increasing function. For simplicity, let us assume that $f_{\mathrm{Q}}(\cdot)$ is a first-order function. Then, it may be given by

$$f_{\mathrm{Q}}\big(SNR_{\mathrm{Q}}(n)\big) = C \cdot SNR_{\mathrm{Q}}(n) + D, \tag{7.40}$$

where C ($C > 0$) and D are constants.

7.5.1.3 Delay Compensation for Main Filters

As in the PF structure, the estimated $SNR_{\mathrm{P}}(n)$ and $SNR_{\mathrm{Q}}(n)$ generate time delays depending on M. To compensate for these delays, L-sample delay units

z^{-L} are incorporated into the input paths of the primary and the reference signals of MF1 and MF2. L is set to $M/2$ since the delay in the sense of the moving-average for M samples is a half of M.

7.5.2 Evaluations

The performance of the ANC with the CCPF structure was evaluated in comparison with that of the PF structure. The same recorded speech and the noise as in Sec. 7.3.2 as well as the impulse response of the noise path in Fig. 7.4 were used. Since the crosstalk path can be approximated by a delay when the primary and the reference microphones are located close to each other, like using a headset, a unit impulse response with an amplitude of 0.25 and a time delay of 5 samples was used.

The noise component, which was generated by convolution of the noise source with the impulse response of the noise path, was added to the desired signal to create the primary signal. The reference signal was generated by adding the noise to the crosstalk generated by convolution of the desired signal with the impulse response of the crosstalk path. SNRs of the primary and the reference signals in the utterance were 6 dB and −12 dB, respectively. To evaluate the influence by an SNR change, these SNRs were increased by 10 dB by decreasing the noise power by 10 dB after 15 seconds[4]. The sampling frequency was 8 kHz. Other parameter values are shown in Tab. 7.2.

Table 7.2. Parameters and corresponding values.

Parameter	Value	Parameter	Value
N	64	$SNR_{\mathrm{P}_{\mathrm{min}}}$	−30 dB
M	128	A	−0.01
L	64	B	0.1
μ_{SF}	0.1	$\mu_{\mathrm{M2}_{\mathrm{max}}}$	0.02
μ_{SF1}	0.1	$\mu_{\mathrm{M2}_{\mathrm{min}}}$	0.0
μ_{SF2}	0.002	$SNR_{\mathrm{Q}_{\mathrm{max}}}$	0 dB
$\mu_{\mathrm{M1}_{\mathrm{max}}}$	0.2	$SNR_{\mathrm{Q}_{\mathrm{min}}}$	−10 dB
$\mu_{\mathrm{M1}_{\mathrm{min}}}$	0.0	C	0.002
$SNR_{\mathrm{P}_{\mathrm{max}}}$	−10 dB	D	0.02

[4] A severer condition was selected on purpose. A smaller noise power means a stronger interference (*i.e.* stronger desired signal) for adaptation of SF1 and MF1.

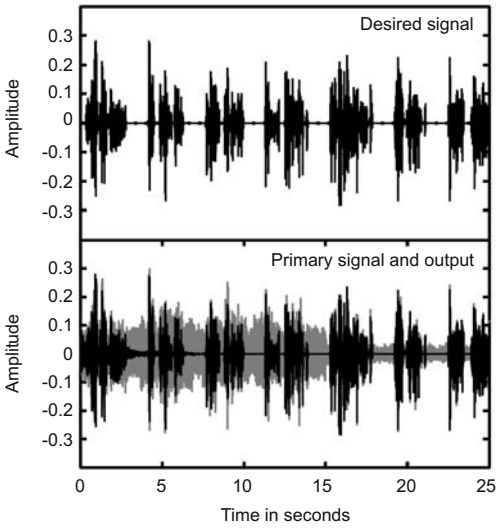


Fig. 7.13. Desired signal (upper diagram), primary signal (gray, lower diagram) and output (black, lower diagram) speech.

Fig. 7.13 shows the desired signal (upper diagram), the primary signal (gray, lower diagram), and the output (black, lower diagram) of the ANC with the CCPF structure. They indicate that the CCPF structure cancels the noise to a satisfactory level.

Fig. 7.14 shows the SNR of the primary signal, its estimate using SF1, the SNR of the reference signal, and its estimate utilizing SF2. Where the SNR is high, the estimated SNR agrees well with the actual one.

Fig. 7.15 shows the stepsizes of MF1 and MF2. The stepsize of MF1 is large when the speech signal is absent. The stepsize is generally small after 15 seconds since the SNR was increased by 10 dB. On the other hand, the stepsize of MF2 is large in those sections where the speech is present. After 15 seconds, the stepsize is generally large due to the increased SNR.

Part (a) of Fig. 7.16 depicts the normalized output, $\epsilon(n)$, in Eq. 7.24. A large negative value of $\epsilon(n)$ in nonspeech sections represents good noise cancellation. The ANC with the CCPF structure achieves as much as 10 dB lower output level than the ANC with the PF structure.

Part (b) of Fig. 7.16 exhibits the signal distortion $\delta(n)$ in Eq. 7.25 at the ANC output. The ANC with a CCPF structure reduces the distortion by as much as 15 dB in utterances compared to that with the PF structure. Although there are sections where the CCPF structure creates larger distortion, it does not result in the degradation of the subjective voice quality since these sections are limited to silent sections with small signal power. In addition, the CCPF structure creates no increase in distortion even when the SNR is increased by 10 dB.

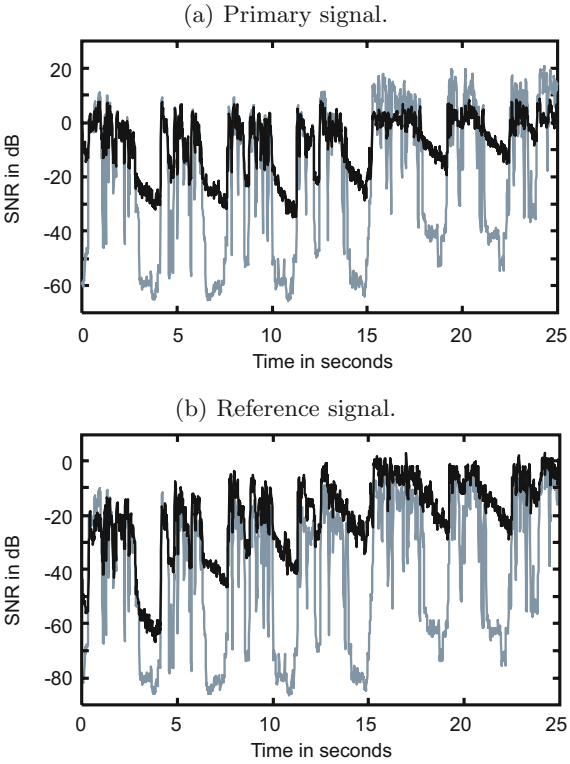


Fig. 7.14. True (gray) and estimated (black) SNRs of the primary (upper diagram) and the reference signals (lower diagram).

7.6 Generalized Cross-Coupled Paired Filter (GCCPF) Structure

The CCPF structure has two pairs of cross-coupled adaptive filters, each of which consists of a main filter and a subfilter. The subfilters serve as pilot filters whose output is used to estimate the signal-to-noise ratios (SNRs) of the primary and the reference signals. The stepsizes for the adaptation of the main filters are controlled according to the estimated SNRs. Since the stepsize is controlled by the subfilters, good noise cancellation and low signal-distortion in the output are simultaneously achieved. However, the CCPF structure was developed for communication headsets in noisy environment. The fixed stepsize for each subfilter may not provide satisfactory performance for a wide range of SNRs that are encountered in other applications. Actually, application to human-robot communication is attracting more interests from a viewpoint of noise and interference cancellation [19]. It is possible to follow the path from the PF structure to the CCPF structure by introducing yet

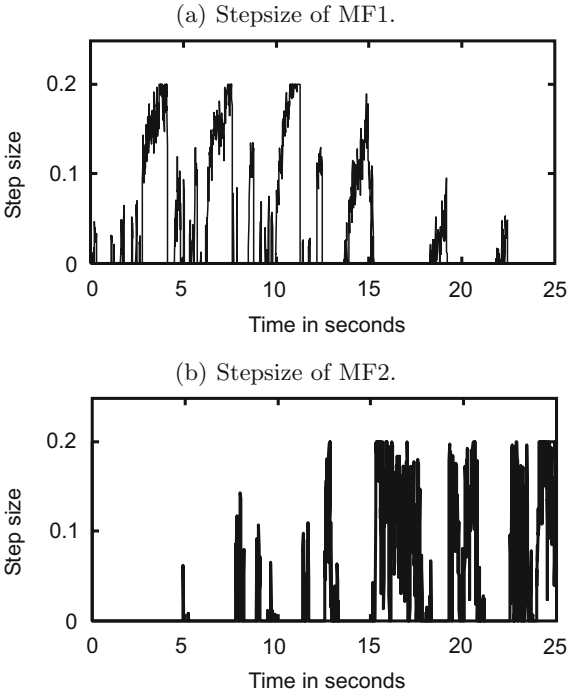


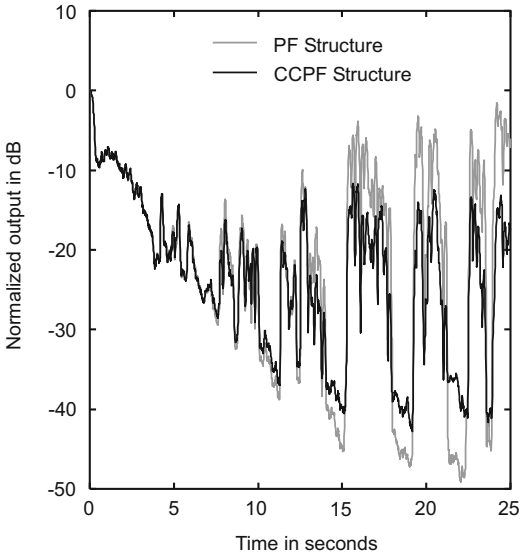
Fig. 7.15. Stepsizes of MF1 and MF2.

another set of subfilters for stepsize control of SF1 and SF2. However, this path may lead to an infinite chain of such extensions.

A generalized cross-coupled paired filter (GCCPF) structure [16] utilizes the primary and the reference signals for approximating their SNRs instead of another pair of pilot filters for SF1 and SF2. Fig. 7.17 shows a block diagram of an ANC with a GCCPF structure. Four adaptive filters, namely, the main adaptive filters (MF1, MF2) and the sub adaptive filter (SF1, SF2) generate noise and crosstalk replicas as in the CCPF structure. Adaptive control of the stepsizes for SF1 and SF2 forms the most significant difference from the CCPF structure. Average powers $R_{\mathrm{S}}(n)$ and $R_{\mathrm{N}}(n)$ of the primary signal $x_{\mathrm{P}}(n)$ and the reference signal $x_{\mathrm{R}}(n)$ are first calculated. A ratio of $R_{\mathrm{S}}(n)$ to $R_{\mathrm{N}}(n)$, representing a rough estimate of the SNR at the primary input, is used for controlling the stepsizes of SF1 and SF2.

The SF1 output $y_2(n)$ and the subtraction result $e_2(n)$ are used to estimate a more precise SNR at the primary input. $e_2(n)$ serves as an approximation to the desired signal, and $y_2(n)$ is used as that to the noise. The stepsize for MF1 is controlled based on the estimated SNR calculated from the output signal of SF1. SF2 works for crosstalk instead of noise in a similar way to

(a) Normalized output $\epsilon(n)$.



(b) Signal distortion $\delta(n)$.

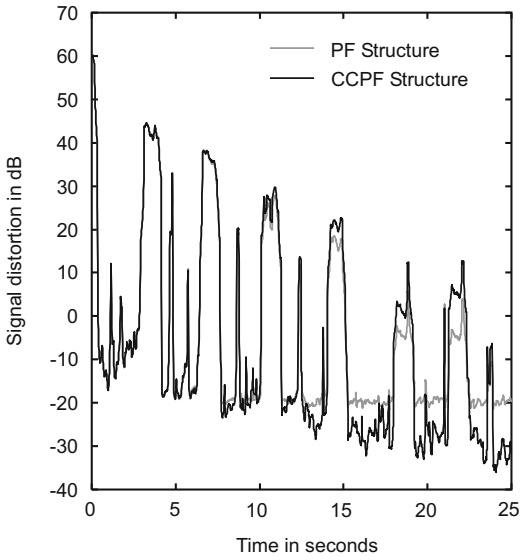


Fig. 7.16. Normalized output (a) and signal distortion (b).

that of SF1. The resulting SNR estimate from SF2 output signals is used to control the MF2 stepsize.

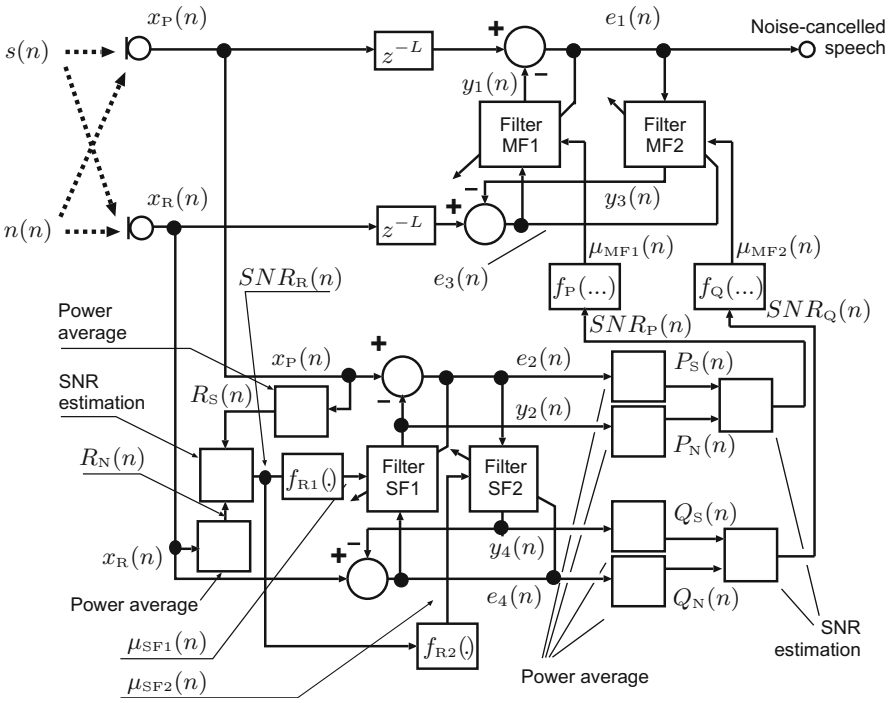


Fig. 7.17. ANC with a generalized cross-coupled paired filter (GCCPF) structure.

7.6.1 Algorithm

The stepsize for the coefficient adaptation should be kept smaller when there is more interference in the error. In the structure in Fig. 7.17, the stepsize, $\mu_{SF1}(n)$, for SF1 and the stepsize, $\mu_{MF1}(n)$, for MF1 should be set to a small value when the SNR at the primary input is high to avoid distortion at the ANC output. On the other hand, $\mu_{SF1}(n)$ and $\mu_{MF1}(n)$ can be set large when this SNR is low for fast convergence and rapid tracking of noise-path changes. A similar rule applies to $\mu_{SF2}(n)$ and $\mu_{MF2}(n)$ for coefficient adaptation in SF2 and MF2 with respect to the SNR at the reference input. All these stepsizes can be controlled appropriately once the SNRs for the adaptive filters become available.

The SNR for the primary signal, $SNR_R(n)$, is approximated by

$$SNR_{\mathrm{R}}(n) = 10\ \log_{10}\left\{\frac{R_{\mathrm{S}}(n)}{R_{\mathrm{N}}(n)}\right\}\mathrm{dB}, \tag{7.41}$$

$$R_{\mathrm{S}}(n) = \sum_{j=0}^{M-1} x_{\mathrm{P}}^2(n-j), \tag{7.42}$$

$$R_{\mathrm{N}}(n) = \sum_{j=0}^{M-1} x_{\mathrm{R}}^2(n-j). \tag{7.43}$$

$\mu_{\mathrm{SF1}}(n)$ and $\mu_{\mathrm{SF2}}(n)$ are controlled by the estimated SNR, $SNR_{\mathrm{R}}(n)$, as in the following equations:

$$\mu_{\mathrm{SF1}}(n) = \begin{cases} \mu_{\mathrm{S_{min}}}, & \text{if } SNR_{\mathrm{R}}(n) > SNR_{\mathrm{R_{max}}}, \\ \mu_{\mathrm{S_{max}}}, & \text{if } SNR_{\mathrm{R}}(n) < SNR_{\mathrm{R_{min}}}, \\ f_{\mathrm{R1}}\big(SNR_{\mathrm{R}}(n)\big), & \text{otherwise}, \end{cases} \tag{7.44}$$

$$\mu_{\mathrm{SF2}}(n) = \begin{cases} \mu_{\mathrm{S_{min}}}, & \text{if } SNR_{\mathrm{R}}(n) < SNR_{\mathrm{R_{min}}}, \\ \mu_{\mathrm{S_{max}}}, & \text{if } SNR_{\mathrm{R}}(n) > SNR_{\mathrm{R_{max}}}, \\ f_{\mathrm{R2}}\big(SNR_{\mathrm{R}}(n)\big), & \text{otherwise}. \end{cases} \tag{7.45}$$

$\mu_{\mathrm{S_{max}}}$ and $\mu_{\mathrm{S_{min}}}$ are the maximum and the minimum stepsizes for $\mu_{\mathrm{SF1}}(n)$ and $\mu_{\mathrm{SF2}}(n)$. $f_{\mathrm{R1}}(\cdot)$ and $f_{\mathrm{R2}}(\cdot)$ are functions of $SNR_{\mathrm{R}}(n)$. $f_{\mathrm{R1}}(\cdot)$ should be a decreasing function because a small stepsize is suitable for a large SNR. On the other hand, it is desirable that $f_{\mathrm{R2}}(\cdot)$ is an increasing function. Eqs. 7.44 and 7.45 enable the ideal stepsize control described earlier, leading to small residual error and distortion in the noise-cancelled signal. $\mu_{\mathrm{MF1}}(n)$ and $\mu_{\mathrm{MF2}}(n)$ are controlled by the estimated SNRs, $SNR_{\mathrm{P}}(n)$ and $SNR_{\mathrm{Q}}(n)$, in the same way as in the CCPF structure based on Eqs. 7.19 – 7.22 and Eqs. 7.36 – 7.39. MF1 and MF2 are equipped with L-sample delay units z^{-L} for time-delay compensation.

7.6.2 Evaluation by Recorded Signals

7.6.2.1 Noise Reduction and Distortion

The performance of the ANC with the GCCPF structure was evaluated by computer simulations from the viewpoints of noise reduction and distortion in comparison with the ANC with the CCPF structure [12]. TV sound and a male voice were recorded in a carpeted room with a dimension of 5.5 m (width) × 5.0 m (depth) × 2.4 m (height) in a human-robot communication scenario. The primary microphone was mounted on the forehead and the reference microphone was attached to upper back of a robot, whose height is approximately 0.4 m. The impulse responses of the noise path and the crosstalk path were measured for a direction of noise arrival of 180 degrees with this set-up. An example with a speaker distance of 0.5 m is depicted in

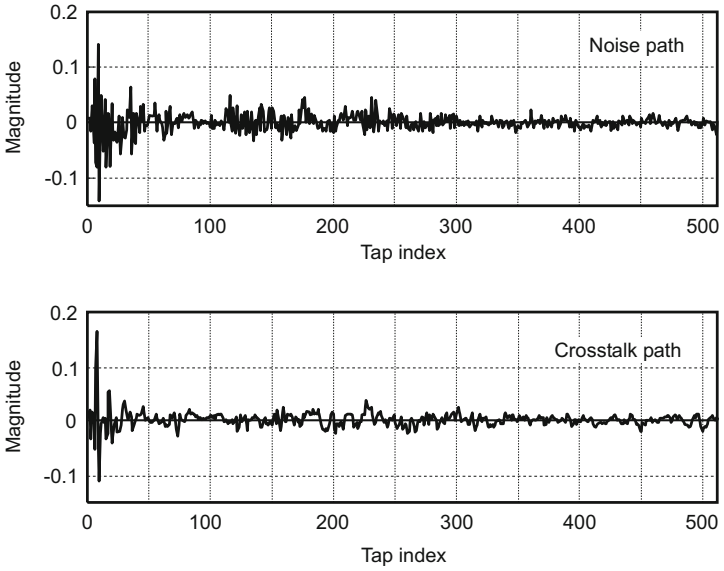


Fig. 7.18. Impulse responses of the noise and the crosstalk path.

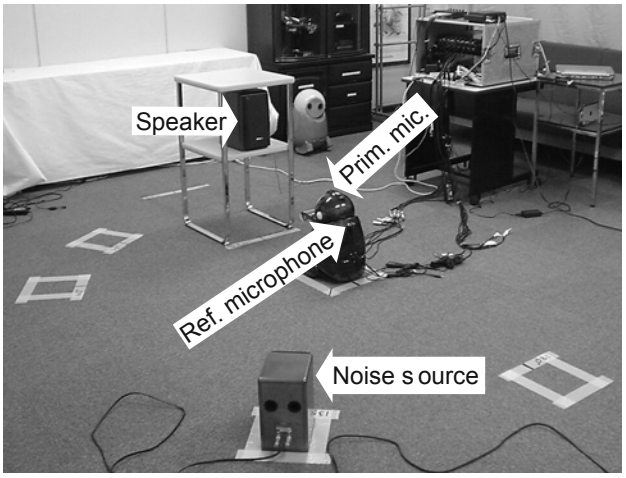


Fig. 7.19. Evaluation environment.

Fig. 7.18. The robot was placed on the straight line between the speaker and the noise source, facing the speaker. This environment is shown in Fig. 7.19.[5]

[5] This figure shows an example where the direction of noise arrival is set to 135 degrees.

Table 7.3. Speaker and noise-source layout.

SNR [primary, reference]	Speaker distance	Noise level
[35, 15] dB	0.5 m	low
[25, 5] dB	1.0 m	low
[20, 0] dB	0.5 m	high
[10, -10] dB	1.0 m	high

The noise component, which is obtained by convolution of the noise source with the impulse response of the noise path, was added to the speech signal to create the primary signal. The reference signal was generated by adding the noise to the crosstalk generated by convolution of the speech signal with the impulse response of the crosstalk path. SNRs of the primary and the reference signals in the utterance were set to [35, 15], [25, 5] dB, [20, 0] dB, and [10, −10] dB, which correspond to four different speaker and noise-source layouts summarized in Tab. 7.3. The sound level for "low" and "high" respectively correspond to 55 − 60 dB and 65 − 70 dB. The sampling frequency was 11.025 kHz. Other parameters are show in Tab. 7.4. These specific layouts have been selected for evaluations because they represent typical scenarios in human-robot communication at home. Moreover, they include some difficult situations such as 35 dB primary SNR where the interference is significant.

Figs. 7.20 and 7.21 show the stepsize for MF1 (upper diagram) and that for MF2 (lower diagram) in the cases of [35, 15] dB, [25, 5] dB, [20, 0] dB, and [10, -10] dB SNRs. Dips in the upper figure and peaks in the lower figure both correspond to speech sections. To highlight speech and nonspeech sections, a rectangular waveform is added to the top of each figure. The waveform has two levels: SP and NSP. SP represents speech sections and NSP corresponds to nonspeech sections.

The stepsizes of the ANC with the GCCPF structure represented by the black solid line show better match with speech sections than those of the conventional ANC expressed in a gray dotted line. Such characteristics, which are closer to the ideal behavior already described in Sec. 7.6.1, are achieved by newly introduced stepsize control for SF1 and SF2 based on the estimated primary and the reference signal powers.

The normalized output (upper diagram) and distortion (lower diagram) at the output are illustrated in Figs. 7.22 and 7.23 for SNR settings of [35, 15] dB, [25, 5] dB, [20, 0] dB, and [10, -10] dB. Speech and nonspeech sections are specified by the same rectangular waveform to that in Figs. 7.20 and 7.21. The results by the ANCs with the GCCPF and the CCPF structures are represented by a black solid and a gray dotted lines. The normalized output, $\epsilon(n)$, and the distortion, $\delta(n)$, were calculated by Eqs. 7.24 and 7.25. In case of

Table 7.4. Parameters and corresponding values.

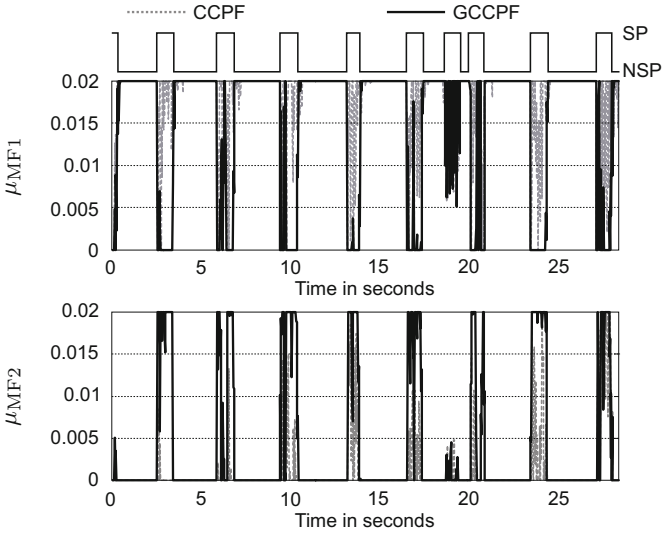
ANC structure	Parameter	Selected value
Common	N	512
	M	128
	L	64
CCPF	μ_{SF1}	0.02
	μ_{SF2}	0.002
	$SNR_{\mathrm{P}_{\mathrm{min}}}$	-7 dB
	$SNR_{\mathrm{P}_{\mathrm{max}}}$	5 dB
GCCPF	$SNR_{\mathrm{P}_{\mathrm{min}}}$	0 dB
	$SNR_{\mathrm{P}_{\mathrm{max}}}$	10 dB
	$\mu_{\mathrm{S}_{\mathrm{min}}}$	0.002
	$\mu_{\mathrm{S}_{\mathrm{max}}}$	0.02
	$SNR_{\mathrm{Q}_{\mathrm{min}}}$	-10 dB
	$SNR_{\mathrm{Q}_{\mathrm{max}}}$	0 dB
	$\mu_{\mathrm{M1}_{\mathrm{min}}}$, $\mu_{\mathrm{M2}_{\mathrm{min}}}$	0.002
	$\mu_{\mathrm{M1}_{\mathrm{max}}}$, $\mu_{\mathrm{M2}_{\mathrm{max}}}$	0.02

good noise cancellation, $\epsilon(n)$ should take a small value in nonspeech sections. When the SNR is low, it takes a lower value than 0 dB even in speech sections. It goes without saying that a smaller distortion, represented by a smaller value of $\delta(n)$, is desirable. Peaks in the normalized output and dips in the distortion correspond to speech sections.

Both noise reduction and distortion are improved by as much as 20 dB in part (a) of Fig. 7.22. Part (b) of Figs. 7.22 and part (a) of Fig. 7.23 also exhibit as much as 15 and 10 dB improvement in both measures. In the case of part (b) of Fig. 7.23, noise reduction is improved by as much as 8 dB. The improvement in distortion in part (b) of Fig. 7.23 is not as evident as that in part (a) of Fig. 7.22. This is because the parameters for the ANC with the CCPF structure are optimal for $[10, -10]$ dB SNR. However, an improvement close to 10 dB between the dotted and the solid lines can be observed in circled areas.

These lower residual-noise levels and smaller distortions for a wide range of SNRs are both due to the adaptive control of the stepsizes for SF1 and SF2. The SF1 stepsize takes relatively small values in speech sections and large

(a) SNR=[35, 15] dB.



(b) SNR=[25, 5] dB.

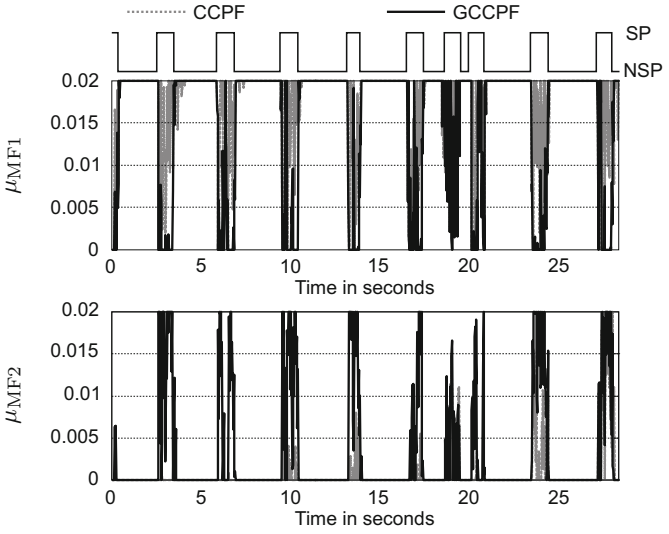
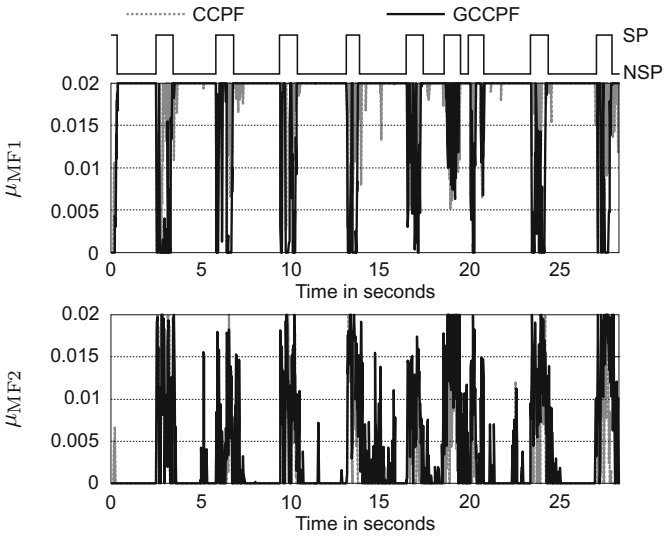


Fig. 7.20. Stepsize of MF1 and MF2 ((a) SNR=[35, 15] dB, (b) SNR=[25, 5] dB).

ones in nonspeech sections to implement the ideal behavior as was described earlier in Sec. 7.6.1. The SF2 stepsize takes the opposite pattern.

(a) SNR=[20, 0] dB.



(b) SNR=[10, -10] dB.

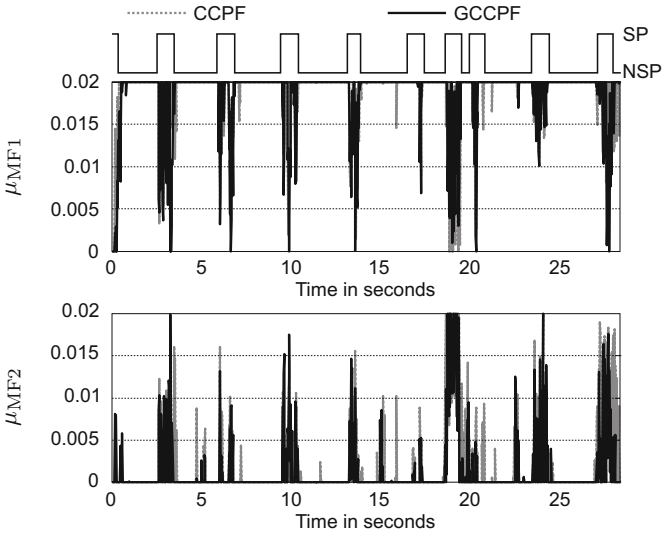
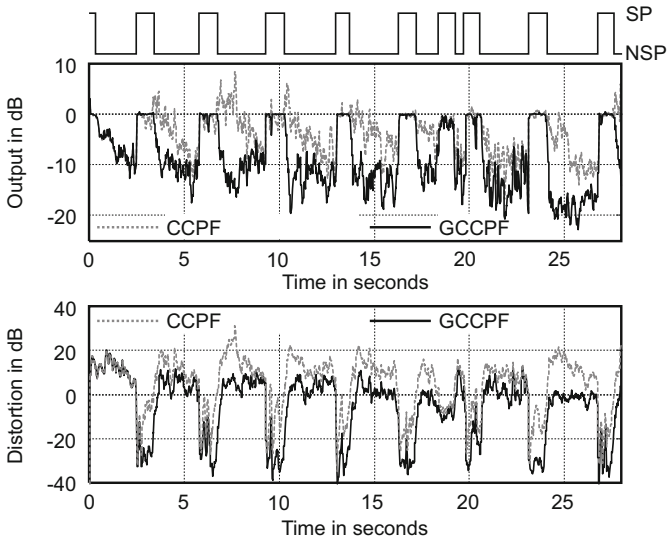


Fig. 7.21. Stepsize of MF1 and MF2 ((a) SNR=[20, 0] dB, (b) SNR=[10, -10] dB).

7.6.2.2 Speech Recognition

Speech recognition was performed with noise-cancelled speech by the ANC with the GCCPF structure. This is because the conventional ANC does not

(a) SNR=[35, 15] dB.



(b) SNR=[25, 5] dB.

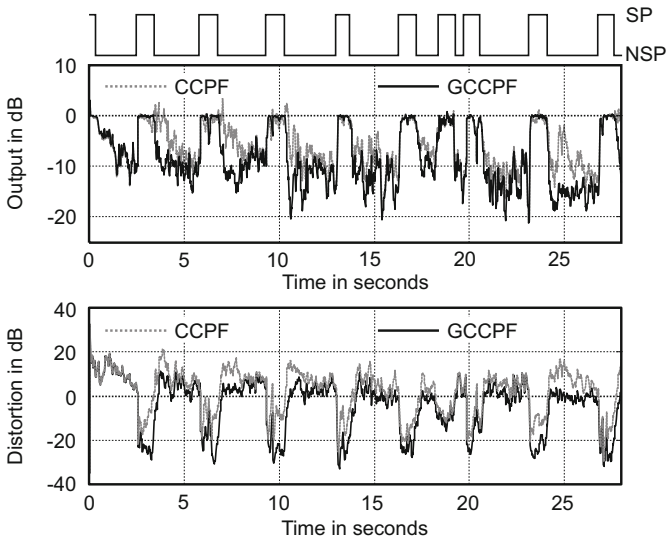


Fig. 7.22. Normalized output and distortion ((a) SNR=[35, 15] dB, (b) SNR=[25, 5] dB).

achieve sufficiently low residual noise nor low distortion for a wide range of SNRs. Distortion in the noise-cancelled speech degrades the speech recognition rate because its acoustic characteristics are less likely to match the HMM

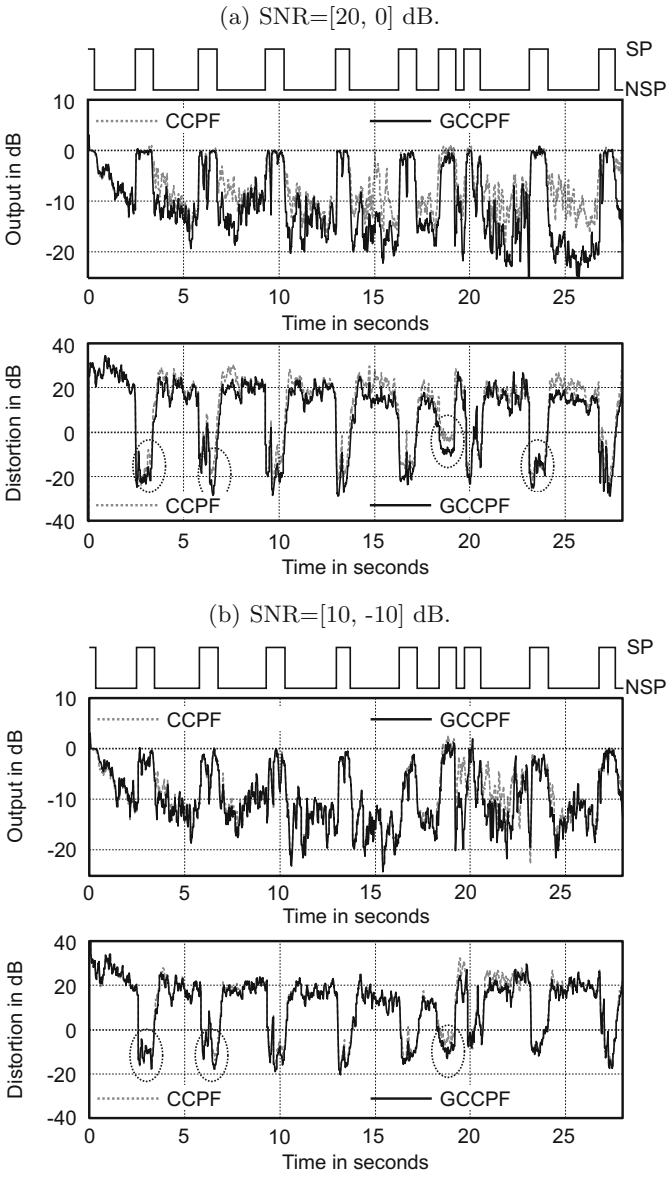


Fig. 7.23. Normalized output and distortion ((a) SNR=[20, 0] dB, (b) SNR=[10, -10] dB).

(hidden Markov model) in the recognition system. The residual noise, on the other hand, leads to wrong detection of speech sections. Because speech segmentation is important in speech recognition, it also results in low recognition

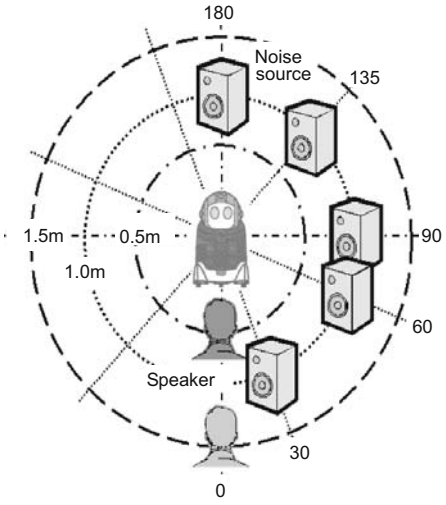


Fig. 7.24. Experimental set-up for speech recognition.

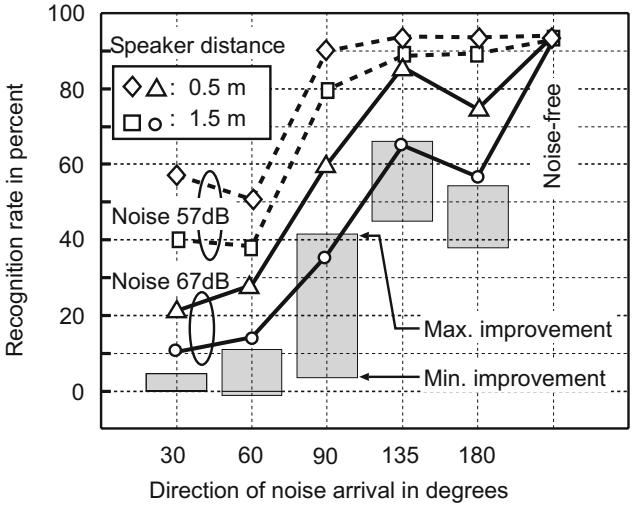
rates. $10 - 20$ dB decrease by the ANC with the CCPF structure compared to the ANC with the GCCPF structure apparently predicts its inferior recognition rates.

The experimental set-up is illustrated in Fig. 7.24. 150 utterances by 30 different male, female, and child speakers were presented at a distance of 0.5 and 1.5 m. The noise source was placed at a distance of 1.0 m in a direction of 30, 60, 90, 135, or 180 degrees. Speaker independent speech recognition based on semi-syllable hidden Markov models [21] was used with a dictionary of 600 robot commands.

Fig. 7.25 depicts the speech recognition rate for a commercial and a news TV-programs as the noise source. Shaded columns represent improvements, *i.e.* the difference in the recognition rate with and without the ANC. For the commercial program, the recognition rate is equivalent to that in the noise-free condition when the speaker distance is 0.5 m with a 57 dB noise in directions of 90 to 180 degrees. The maximum improvement reaches 65%. The recognition rate is degraded accordingly for off-direction noise placement, longer distance of the speech source, and/or a higher noise level of 67 dB.

For the news program, the recognition rate is slightly degraded compared to that in part (a) of Fig. 7.25 for noise directions of 135 and 180 degrees. However, with a noise directions of 30, 60, and 90 degrees, the recognition rate is significantly lower. The improvement is also degraded accordingly. This degradation is caused by similar spectral components in the the speech to be recognized and the news program. It should be noted that such a difference have some variance and its maximum and the minimum are shown. A transition from significant effect to moderate effect is observed in the direction of

(a) Commercial program.



(b) News program.

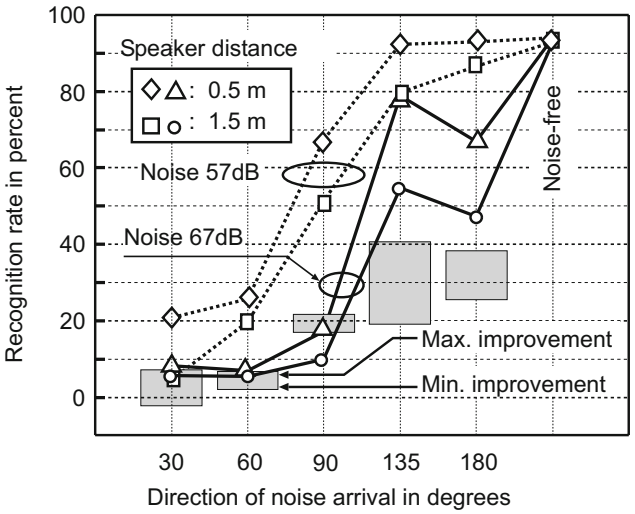


Fig. 7.25. Speech recognition results (top: commercial program, bottom: news program).

arrival of 90 degrees in part (a) and 135 degrees in part (b) of Fig. 7.25. In the case of a distant speaker, the recognition rates in the direction of arrival of 180 degrees are degraded. It is caused by significant contribution by increased reverberation and decreased SNR due to the attenuated speech.

Fig. 7.26. Child-care robot, PaPeRo.

7.7 Demonstration in a Personal Robot

In speech recognition for robots, microphone arrays [4] have become popular [20]. However, they are not as effective for diffuse noise as for directional interference. On the contrary, ANCs produce no directivity and are useful for diffuse noise.

The ANC with the GCCPF structure is equipped with in a personal robot, PaPeRo [7], whose child-care model is depicted in Fig. 7.26. PaPeRo has a laptop PC inside to perform all necessary computations and controls.[6] The primary microphone and the reference microphone are mounted on the forehead and the back of the neck. PaPeRo had been exhibited at the 2005 World Exposition (EXPO 2005), Aichi, Japan, as a childcare robot. Several children at a time with a dedicated instructor played a variety of games with the robot as shown in Fig. 7.27. 27000 children aged 3 to 12 enjoyed playing with PaPeRo. The total number of visitors reached 780000. In such a noisy environment, the speech recognition was successful due to the ANC with the GCCPF structure. This success demonstrates a revival of a classical technique originally proposed by Widrow et al.

7.8 Conclusions

Low-distortion noise cancellers and their applications have been presented. Distortion in Widrow's adaptive noise canceller (ANC) has been investigated

[6] A more compact implementation based on an embedded processor accomodating three ARM9 and a DSP cores is also available [17].

Fig. 7.27. PaPeRo demonstration with children and instructors.

to show that interference in coefficient adaptation and crosstalk are problems. As a solution to the interference problem, a paired filter (PF) structure has been described. For the crosstalk problem, it has been pointed out that CTRANC and a cross-coupled structure without a reliable adaptation control are not sufficiently good. As a good solution to both interference and crosstalk, an ANC with a cross-coupled paired filter (CCPF) structure has been presented. For more adverse environment, the CCPF structure has been extended to a generalized cross-coupled paired filter (GCCPF) structure. Evaluation results of the GCCPF structure have demonstrated its superior performance with respect to residual noise and distortion in a human-robot interaction scenario.

Although Widrow's adaptive noise canceller is a classical technique and has found a few applications, its descendants have found their ways in robotics where nondirectional interference plays a significant role. A successful demonstration of a partner-type robot PaPeRo at 2005 World Exposition in Aichi, Japan, for six months tells us that it is a revival of a classical technique.

References

1. M. J. Al-Kindi, J. Dunlop: A low distortion adaptive noise cancellation structure for real time applications, *Proc. ICASSP '87*, 2153–2156, Apr. 1987.
2. S. F. Boll: Suppression of acoustic noise in speech using spectral subtraction, *IEEE Trans. Acoust., Speech, Signal Processing*, **ASSP-27**(2), 113–120, Apr. 1979.
3. S. F. Boll, D. C. Pulsipher: Suppression of acoustic noise in speech using two microphone adaptive noise cancellation, *IEEE Trans. Acoust., Speech, and Signal Processing*, **ASSP-28**, 752–753, 1980.
4. M. Brandstein, D. Ward (eds.): *Microphone Arrays,* Berlin, Germany: Springer, 2001.
5. J. Dunlop, M. J. Al-Kindi, L. E. Virr: Application of adaptive noise cancelling to diver voice communications, *Proc. ICASSP '87*, 1708–1711, Apr. 1987.
6. Y. Ephraim, D. Malah: Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator, *IEEE Trans. Acoust., Speech, Signal Processing*, **ASSP-32**(6), 1109–1121, Dec. 1984.
7. Y. Fujita: Personal robot PaPeRo, *J. of Robotics and Mechatronics*, **14**(1), Jan. 2002.
8. W. A. Gardner, B. G. Agee: Two-stage adaptive noise cancellation for intermittent-signal applications, *IEEE Trans. IT*, **IT-26**(6), 746–750, Nov. 1980.
9. G. C. Goodwin, K. S. Sin: *Adaptive Filtering, Prediction and Control,* Englewood Cliffs, NJ, USA: Prentice-Hall, 1985.
10. W. A. Harrison, J. S. Lim, E. Singer: A new application of adaptive noise cancellation, *IEEE Trans. Acoust., Speech, and Signal Processing*, **ASSP-34**, 21–27, 1986.
11. S. Ikeda, A. Sugiyama: An adaptive noise canceller with low signal-distortion for speech codecs, *IEEE Trans. Sig. Proc.*, 665–674, Mar. 1999.
12. S. Ikeda, A. Sugiyama: An adaptive noise canceller with low signal-distortion in the presence of crosstalk, *IEICE Trans. Fund*, 1517–1525, Aug. 1999.
13. H. Kubota, T. Furukawa, H. Itakura: Pre-processed noise canceller design and its performance, *IEICE Trans. Fund.*, **J69-A**(5), 584–591, May 1986 (in Japanese).
14. G. Mirchandani, R. L. Zinser, J. B. Evans: A new adaptive noise cancellation scheme in the presence of crosstalk, *IEEE Trans. CAS-II*, 681–694, Oct. 1992.
15. V. Parsa, P. A. Parker, R. N. Scott: Performance analysis of a crosstalk resistant adaptive noise canceller, *IEEE Trans. Circuits and Systems*, **43**, 473-482, 1996.
16. M. Sato, A. Sugiyama, S. Ohnaka: An adaptive noise canceler with low signal-distortion based on variable stepsize subfilters for human-robot communication, *IEICE Trans. Fund.*, **E88-A**(8), 2055–2061, Aug. 2005.
17. M. Sato, T. Iwasawa, A. Sugiyama: A noise-robust speech recognition on a compact speech dialogue module, *Proc. SIG AI-Challenge 2007*, Nov. 2007 (in Japanese).
18. A. Sugiyama, M. N. S. Swamy, E. I. Plotkin: A fast convergence algorithm for adaptive FIR filters, *Proc. ICASSP '89*, 892–895, 1989.
19. A. Sugiyama, M. Sato: Robust speech recognition in noisy environment for robot applications, *J. of Acoust. Soc. Japan*, **63**(1), 47–53, Jan. 2007 (in Japanese).
20. J.-M. Valin, J. Rouat, F. Michaud: Enhanced robot audition based on microphone array source separation with post-filter, *Proc. ICRSJ 2004*, **3**(28), 2123–2128, Oct. 2004.

21. T. Watanabe: Problems in the design of a speech recognition system and their solution, *Trans.*, **J.79-D-II**(12), 2022–2031, Dec. 1996 (in Japanese).
22. B. Widrow, J. R. Glover, Jr., J. M. McCool, J. Kaunitz, C. S. Williams, R. H. Hearn, J. R. Zeidler, E. Dong, Jr., R. C. Goodlin: Adaptive noise cancelling: principles and applications, *Proc. IEEE*, **63**(12), 1692–1716, 1975.
23. R. L. Zinser, G. Mirchandani, J. B. Evans: Some experimental and theoretical results using a new adaptive filter structure for noise cancellation in the presence of crosstalk, *Proc. ICASSP '85*, 1253–1256, Mar. 1985.