A NEW GENERALIZED SIDELOBE CANCELLER WITH A COMPACT ARRAY OF MICROPHONES SUITABLE FOR MOBILE TERMINALS

Akihiko Sugiyama and Ryoji Miyahara†

Information and Media Processing Laboratories
NEC Corporation
† Internet Terminal Division, NEC Engineering
1753 Shimonumabe, Nakahara-ku, Kawasaki 211-8666, Japan
aks@ak.jp.nec.com

ABSTRACT

This paper proposes a new generalized sidelobe canceller with a compact array of microphones suitable for mobile terminals. The output of the fixed beamformer (FBF) is further processed by a newly introduced decorrelation unit which has an auxiliary input signal to improve poor interference suppression of FBF in low frequencies. The output of the decorrelation unit is used as the reference signal for the adaptive blocking matrix and the input for the multi-input canceller. Because low and high frequency components are cancelled or suppressed by the decorrelation unit and FBF, better output-signal quality is obtained. Output signal comparison confirms approximately 12 dB higher interference cancellation by the new beamformer.

Index Terms— Beamforming, Microphone array, Generalized sidelobe canceller, Low-frequency directivity, Compact array

1. INTRODUCTION

Microphone arrays based on beamforming have been attracting much attention for interference suppression in commercial products such as teleconference systems, hearing aids, TV receivers and personal computers (PCs) [1]–[4]. Adaptive microphone arrays are most widely studied for its good interference-suppression performance with a relatively small number of microphones. Griffiths-Jim beamformer (GJBF) [5], or the generalized sidelobe canceler (GSC), is one of the most popular structures.

Figure 1 shows a GSC structure with an adaptive blocking matrix (ABM) [6]. It has a fixed beamformer (FBF) and a multi-input canceller (MC) in addition to an ABM. FBF is designed to form a beam in the look direction, which is a known target source direction and, in most cases, is perpendicular to the array surface. The target signal is passed unattenuated while all other signals are attenuated with a factor corresponding to the signal direction of arrival (DOA). Contrary to FBF, ABM forms a null in the look direction with the FBF output as a reference signal to suppress the target and pass all other signals. MC finally removes all signal components correlated with the ABM output signals. Because the ABM output has strong correlation with signal components other than the target, the MC output consists of the enhanced target signal and a little residual interference.

Standard FBFs have a limitation from a viewpoint of constant beamwidth across frequency. The mainlobe in a low frequency is wider than that in a high frequency, leading to poor spatial selectivity. This is because a low-frequency wavelength is longer than its high-frequency counterpart and makes the effective microphone spacing shorter for the shared microphone units. It means that low-frequency signals are not attenuated as much as high frequency signals for arrays with a small microphone spacing. Insufficient attenuation means interference-contaminated ABM reference resulting in ABM output with a smaller interference power especially in low frequencies. As a result, interference cancellation in the MC output becomes insufficient.

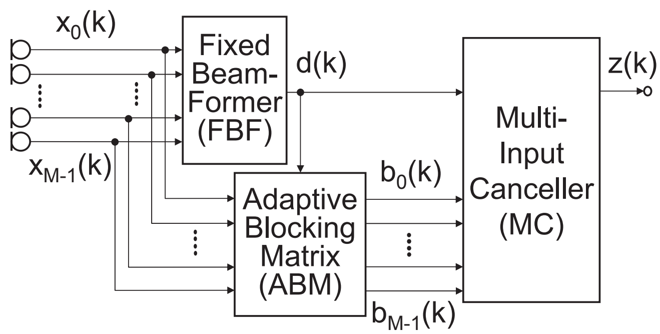

Fig. 1. Generalized sidelobe canceller with an adaptive blocking matrix (ABM).

Solutions to this problem are based on the idea that arrays of different size dedicated to different frequency ranges should be combined [7, 8]. A most common example is a harmonically-nested array [9]–[15]. However, as far as a nested technique is employed, increase in the array size and the number of microphones is inevitable. These requirements are not acceptable for compact microphone arrays on mobile terminals.

This paper proposes a new generalized sidelobe canceller with a compact array of microphones suitable for mobile terminals. The following section is devoted to explain insufficient interference attenuation in low frequencies in the FBF. Section 3 presents a new beamformer structure with an auxiliary input. Finally, in Section 4, the enhanced signals with and without the auxiliary input signal are compared to demonstrate the superior performance of the proposed beamformer.

2. INTERFERENCE ATTENUATION IN FBF

Figure 2 shows an example plot of directivity for the FBF output signal $d(k)$. Four microphones are linearly arranged with a uniform spacing of 3.8 cm. A dashed and a solid line represent gains at 2 and 4 kHz. The mainlobe centered at 0 degrees is wider at 2 kHz

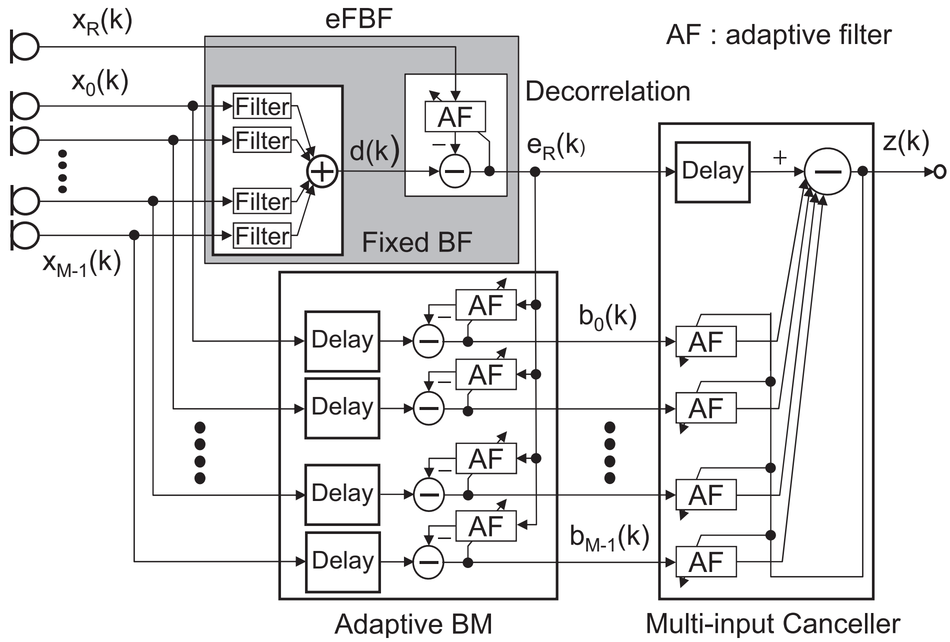

Fig. 4. New GSC with an extended FBF (eFBF).

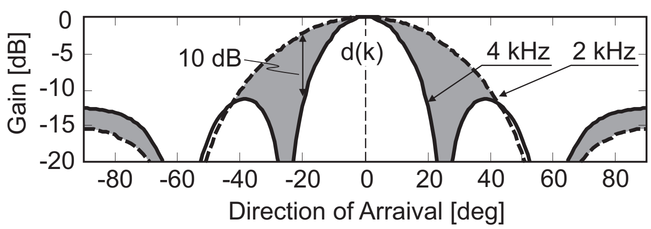

Fig. 2. Directivity of fixed beamformer (FBF).

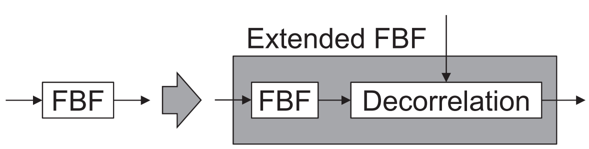

Fig. 3. FBF and its extension.

than at 4 kHz. A wider mainlobe indicates that more interference off the look direction remain unattenuated. For example, in the direction of 20 degrees, there is 10 dB less attenuation at 2 kHz than at 4 kHz. Likewise, in the shaded regions, low frequency attenuation is significantly smaller than that in high frequencies. In another word, interference attenuation in low frequencies is insufficient at the FBF output for arrays with a constant and small microphone spacing for wideband signals. This is a common scenario for easy-to-carry mobile terminals.

This problem causes partial-blocking of low-frequency interference at the ABM output or, equivalently, under-estimation of low-frequency interference power. Finally, use of too small low-frequency power as a reference causes insufficient interference attenuation in low frequencies compared to that in high frequencies.

3. PROPOSED GSC

The proposed GSC has an extended FBF structure which has a decorrelation unit after the FBF to compensate for insufficient interference attenuation in low frequencies. This new decorrelation unit works as postprocessing for the incomplete FBF as shown in Fig. 3 with a reference signal. Interference in high frequencies is attenuated by spatial selectivity (directivity) of FBF and that in low frequencies is further attenuated by decorrelation with the reference signal. The reference signal should be most correlated with the interference and least correlated with the target. Then, all signal components in the FBF output correlated with the interference are cancelled in this decorrelation process. Such a function is known as a noise canceller [16].

Figure 4 illustrates a block diagram of the new GSC. FBF, ABM, and MC may take any structure. In this paper, FBF, ABM, and MC are implemented as a delay-and-sum beamformer, ABM with coefficient-constrained adaptive filters (CCAF) [17, 18], and MC with norm-constrained adaptive filters (NCAF) [19], respectively.

The decorrelation unit has an adaptive filter (AF) driven by a reference signal $x_R(k)$ from the reference microphone, which collects signals other than the target signal. The adaptive filter output is a convolution of the reference signal and the filter coefficient vector $\mathbf{h}(k)$, which is adapted by the NLMS algorithm [20] with stepsize $\mu(k)$.

$$\mathbf{h}(k+1) = \mathbf{h}(k) + \frac{\mu(k)e_R(k)\mathbf{x}_R(k)}{\mathbf{x}_R(k)^T\mathbf{x}_R(k)}, \tag{1}$$

$$\mu(k) = \begin{cases} \mu_{min} & SNR(k) > SNR_{max} \\ \mu_{max} & SNR(k) < SNR_{min} \\ f(SNR(k)) & otherwise \end{cases}, \tag{2}$$

$$\mathbf{h}(k) = [h_0(k), h_1(k), \ldots, h_{L_D-1}(k)]^T, \tag{3}$$

$$\mathbf{x}_R(k) = [x_R(k), \ldots, x_R(k-L_D+1)]^T. \tag{4}$$

L_D is the number of taps of the decorrelation adaptive filter. $e_R(k)$ is the decorrelation result and provided with ABM as a reference signal. $SNR(k)$ is estimated as an average power ratio of $e_R(k)$ to the adaptive filter output. They approximate the target signal power and the interference power, respectively. Function $f(\cdot)$ is a monotonically decreasing function [21]. Equation (2) indicates that the stepsize $\mu(k)$ takes a small and a large value for a high and a low

SNR. This stepsize control [21] guarantees stable behavior of the adaptive filter with no "doubletalk control."

The output of the FBF is used as the common input to CCAFs in ABM. The output of each CCAF is subtracted from the delayed microphone signal. CCAF coefficient vectors $\mathbf{g}_m(k)$ $(0 \leq m \leq M-1)$ are adapted with constraints by the NLMS algorithm as follows:

$$\mathbf{g}'_m(k{+}1) \;=\; \mathbf{g}_m(k) + \alpha \frac{b_m(k)}{\mathbf{e}_R(k)^T \mathbf{e}_R(k)} \mathbf{e}_R(k), \qquad (5)$$

$$\mathbf{g}_m(k{+}1) \;=\; \begin{cases} \boldsymbol{\phi}_m & \text{for } \mathbf{g}'_m(k{+}1) > \boldsymbol{\phi}_m \\ \boldsymbol{\psi}_m & \text{for } \mathbf{g}'_m(k{+}1) < \boldsymbol{\psi}_m \\ \mathbf{g}'_m(k{+}1) & \textit{otherwise} \end{cases}, \qquad (6)$$

$$\mathbf{g}_m(k) \;=\; [g_{m,0}(k), g_{m,1}(k), \ldots, g_{m,L_B-1}(k)]^T, \qquad (7)$$

$$\mathbf{e}_R(k) \;=\; [e_R(k), e_R(k{-}1), \ldots, e_R(k{-}L_B{+}1)]^T, \qquad (8)$$

where each CCAF is assumed to have L_B taps and M is the number of microphones. $\mathbf{g}'_m(k{+}1)$ is a temporal coefficient vector for limiting functions. $\boldsymbol{\phi}_m$ and $\boldsymbol{\psi}_m$ are the upper and the lower bound vectors for coefficients [17, 18]. In the output signal $b_m(k)$, components correlated with $\mathbf{e}_R(k)$ are cancelled by the CCAFs and nontarget components are maximized. By the constrained-region design of the CCAF coefficients, the maximum allowable look-direction error can be specified [17, 18].

MC subtracts M NCAF output signals correlated with $b_m(k)$ from a delayed version of $e_R(k)$ to generate the its output $z(k)$. Let L_M and $\mathbf{w}_m(k)$ be the number of taps in each NCAF and the m-th NCAF coefficient vector, respectively. NCAF coefficients are updated by the NLMS algorithm with a norm constraint as follows:

$$\mathbf{w}'_m(k+1) \;=\; \mathbf{w}_m(k) + \beta \frac{z(k)}{\mathbf{b}_m(k)^T \mathbf{b}_m(k)} \mathbf{b}_m(k), \qquad (9)$$

$$\mathbf{w}_m(k{+}1) \;=\; \begin{cases} \sqrt{\frac{K}{\Omega}}\, \mathbf{w}'_m(k+1) & \text{for } \Omega > K \\ \mathbf{w}'_m(k+1) & \textit{otherwise} \end{cases}, \qquad (10)$$

$$\Omega \;=\; \mathbf{w}'^T_m(k+1)\mathbf{w}'_m(k+1), \qquad (11)$$

$$\mathbf{w}_m(k) \;=\; [w_{m,0}(k), w_{m,1}(k), \ldots, w_{m,L_M-1}(k)]^T, \qquad (12)$$

$$\mathbf{b}_m(k) \;=\; [b_m(k), b_m(k{-}1), \ldots, b_m(k{-}L_M{+}1)]^T, \qquad (13)$$

where β and $\mathbf{w}'_m(k+1)$ are stepsize and a temporal vector for the constraint, respectively. Ω and K are the total squared-norm of $\mathbf{w}_m(k)$ and a threshold. If Ω exceeds K, $\mathbf{w}_m(k+1)$ are limited by scaling.

Norm constraint by scaling helps reduce excess growth of tap coefficients. Undesirable target cancellation becomes smaller even when the target signal leaks into the NCAF inputs for adverse conditions such as reflections, reverberations, insufficient number of taps with CCAFs, and low signal-to-interference ratios.

Coefficient adaptation in ABM and MC should be performed alternately. They have the opposite relationship with respect to "the desired signal" and "the interference" for the adaptation algorithm. For robustness in the real environment, this control should be achieved automatically with the help of an adaptation mode controller (AMC) [22]–[25]. In this paper, the AMC based on symmetric leaky blocking matrices (SLBMs) [24] is employed.

4. EVALUATIONS

Evaluations were performed with a tablet PC placed on a table and a linear microphone array with four omnidirectional microphones in a room shown in Fig. 5. An auxiliary microphone was fixed on the lower rear side of the tablet PC. Female speech was radiated as the target signal from a pseudo mouth simulator located at 0.5m away from and perpendicular to the array surface. Male speech was played back from a loudspeaker 1.5m away from the array center and behind the target speaker with an angle of 30 degrees to the left. Two interfering independent music signal sources were played back from loudspeakers located at 1.5m away from the array center and behind the tablet PC with angles of 30 degrees to the left and the right. Speech and music signals have a bandwidth of 8 and 22.05kHz and microphone signals were sampled at 16kHz. CCAF was designed to allow up to 20 degree look-direction error [17, 18]. The number of taps for adaptive filters were L_D=256 and L_B=L_M=64 for eFBF, ABM, and MC.

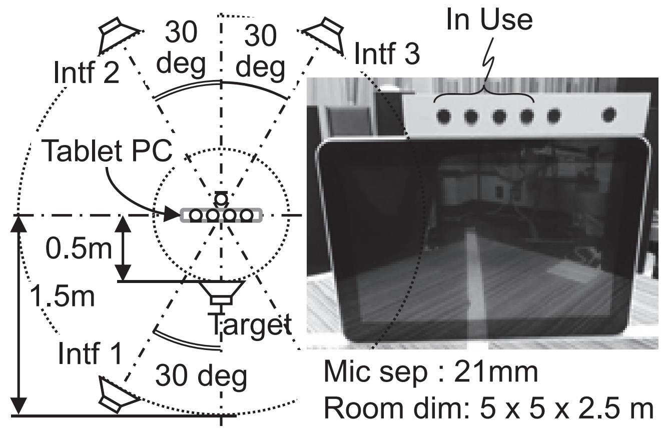

Fig. 5. Evaluation layout.

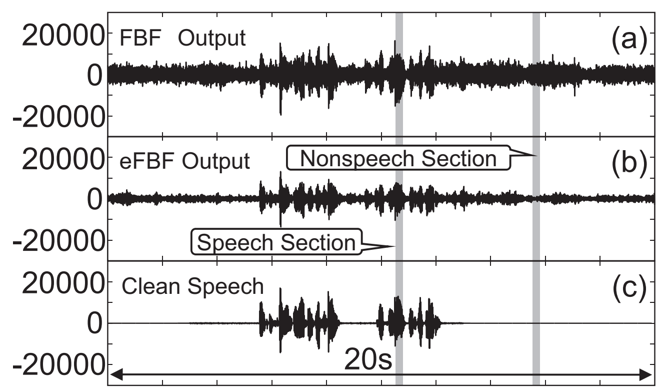

Fig. 6. Comparison of (a) FBF output, (b) extended FBF (eFBF) output, and (c) Clean speech at microphone 2.

Figure 6 compares FBF output, eFBF (extended FBF with decorrelation) output, and clean speech at microphone 2 in (a), (b), and (c). In speech sections, which can be identified in comparison with the clean speech, FBF output and eFBF output exhibit similar magnitude. However, in nonspeech sections, the latter has a much smaller magnitude due to decorrelation.

Figure 7 compares power spectra of the FBF output (gray solid line) and the eFBF output (thick black solid line) for the speech and the nonspeech section corresponding to the gray vertical lines in Fig. 6. The two curves are almost the same in the speech section. However, in the nonspeech section, the eFBF output has a much smaller power in low frequencies ($0-2$ kHz) than the FBF output showing

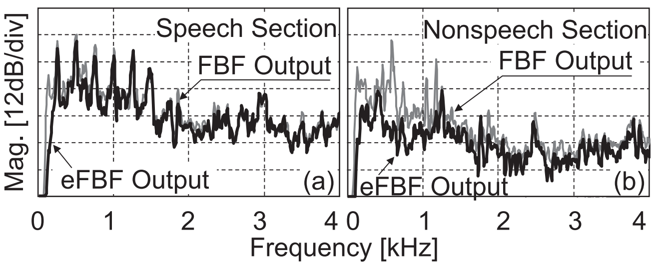

Fig. 7. Comparison of FBF and eFBF output spectra. (a) Speech section, (b) Nonspeech section.

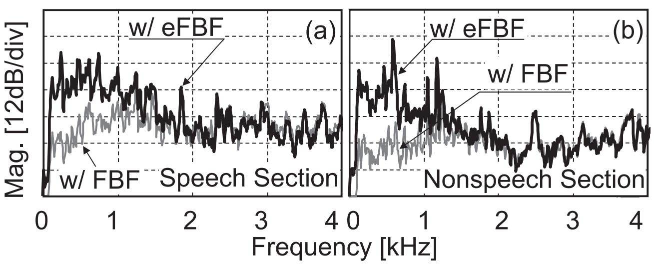

Fig. 8. Comparison of ABM output spectra. (a) Speech section, (b) Nonspeech section.

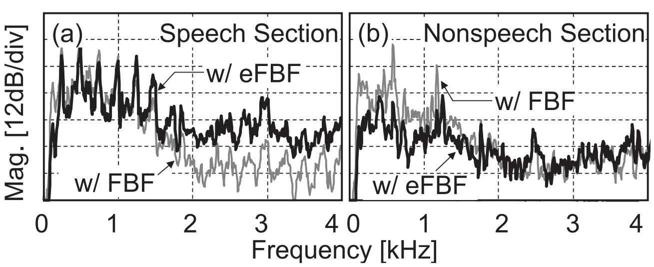

Fig. 9. Comparison of MC output spectra. (a) Speech section, (b) Nonspeech section.

additional low-frequency attenuation by decorrelation.

Shown in Fig. 8 are ABM output signals that correspond to Fig. 7. In this case, the two curves show the same relationship in speech and nonspeech sections. This is because the ABM output mainly consists of signal components other than target, thus, does not significantly change its power in both speech and nonspeech sections. The ABM output with FBF has a much smaller power than that with eFBF in low frequencies ($0 - 1$ kHz). This is a sign of partial cancellation/under-estimation of the interference by FBF in low frequencies. Because the ABM output is used as a reference to cancel the interference in MC, underestimation will cause insufficient interference cancellation.

Such insufficient cancellation, sometimes as much as 12 dB, in low-frequencies ($0 - 1.2$ kHz) is depicted in Fig. 9 (b). Insufficient interference cancellation in high frequencies with eFBF is caused by overadaptation of the MC to low frequencies as well as a small value of L_M. The MC has fullband adaptive filters. In the speech section in (a), they both keep the significant components of the target signal at low-frequency peaks. A dip of the MC output with eFBF is observed around 200 Hz. This is below the first Formant as in Fig. 11 and represents better noise cancellation.

Insufficient cancellation in the nonspeech section is also observed in the MC output in the time-domain as shown in Fig. 10. The MC output with eFBF in Fig. 10 (b) has better interference cancellation compared to that with FBF in (a) due to decorrelation in eFBF. The waveform shape is similar to that of clean speech in (c) with some residual interference. Some power loss is mainly due to highpass filtering in the system to compensate for DC bias removal of the recorded input signal.

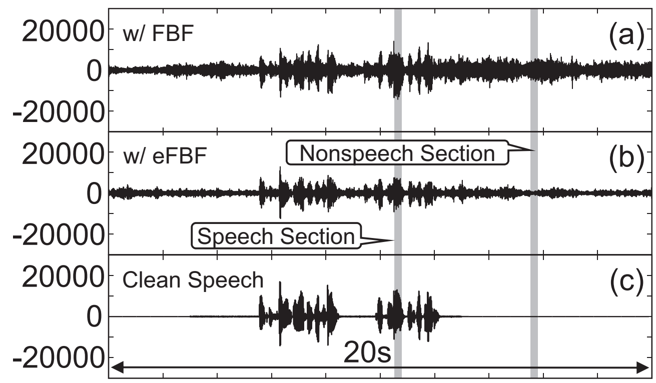

Fig. 10. Comparison of MC output with FBF and extended FBF (eFBF). (a) With FBF, (b) With extended FBF (eFBF), (c) Clean speech at Microphone 2.

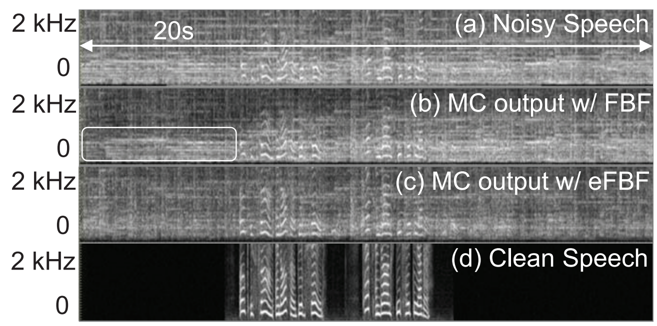

Fig. 11. Comparison of MC output with FBF and extended FBF (eFBF). (a) Noisy speech at Microphone 2, (b) With FBF, (b) With eFBF, (c) Clean speech at Microphone 2.

Finally, in Fig. 11, signal spectrogram of the input noisy signal at microphone 2 (a), MC output with FBF (b), MC output with eFBF (c), and clean speech at microphone 2 (d) are compared. A signal range of 0 to 2 kHz is depicted in the figure which is of interest to see the effect of eFBF. Again, in good agreement with Figs. 9 and 10, the MC output with eFBF has much better reproduction of important speech components in low frequencies. This fact can be understood from less horizontal lines that represent music. Although, in nonspeech sections, there are still some residual interference. However, the residual interference is relatively stationary and can be further suppressed by a noise suppressor [26]–[28].

5. CONCLUSION

A new generalized sidelobe canceller with a compact array of microphones suitable for mobile terminals has been proposed. A decorrelation unit with an auxiliary reference signal mainly composed of interference has been introduced to compensate for insufficient interference attenuation by FBF. Simulation results with recorded signals in the real environment has shown that the output signal with the proposed decorrelation unit has much better reproduction of important speech components with 12 dB higher interference cancellation in low frequencies.

6. REFERENCES

[1] M. Brandstein and D. Ward (Eds), “Microphone arrays,” Springer, 2001.

[2] W. Herbordt and W. Kellermann, “Adaptive beamforming for audio signal acquisition,” *Adaptive signal processing, Applications to real-world problems*, Chap. 6, Benesty and Huang, ed. Springer, 2003.

[3] G. W. Elko and J. Meyer, “Microphone arrays,” *Handbook of speech processing*, Chap. 50, Benesty, Sondhi, and Huang, ed. Springer, 2008.

[4] J. Benesty, Y. Huang, and J. Chen (Eds), “Microphone array signal processing,” Springer, 2008.

[5] L. J. Griffiths and C. W. Jim, “An alternative approach to linear constrained adaptive beamforming,” IEEE Trans. Antennas and Propagations, vol. AP-30, no. 1, pp.27–34, Jan. 1982.

[6] O. Hoshuyama and A. Sugiyama, “Robust adaptive beamforming,” Chap. 5, “Microphone arrays,” M. Brandstein and D. Ward (Eds), Springer, 2001.

[7] R. P. Smith, “Constant beamwidth receiving arrays for broadband sonar systems,” Acustica 23:21–26, 1970.

[8] E. L. Hixon and K. T. Au, “Wide-bandwidth constant bandwidth acoustic array,” J. of the Acoust. Soc. of America, 48(1):117, July 1970.

[9] M. M. Goodwin and G. W. Elko, “Beam dithering,” Proc. Audio Eng. Soc., Oct. 1992.

[10] M. M. Goodwin and G. W. Elko, “Constant beamwidth beamforming,” Proc. ICASSP’93, pp.169–172, Apr. 1993.

[11] D. Tucker, “Arrays with constant beam-width over a wide frequency range,” Nature, 180:496–497, Sep. 1957.

[12] T. Chou, “Frequency independent beamformer with low response error,” Proc. ICASSP’93, pp.169–172, Apr. 1993.

[13] J. L. Flanagan, D. A. Berkeley, G. W. Elko, J. E. West, and M. M. Sondhi, “Autodirective microphone systems,” Acustica, Vol. 73, pp.58–71, 1991.

[14] W. Kellermann, “A self-steering digital microphone array,” Proc. ICASSP’91, pp.3581–3584, Apr. 1991.

[15] F. Khalil, J. P. Jullien, and A. Gilloire, “Microphone array for sound pickup in teleconference systems,” J. Audio Eng. Soc., Vol. 42, No. 9, pp.691-700, Sep. 1994.

[16] B. Wodrow, J. R. Grover, Jr., J. M. McCool, J. Kaunitz, C. S. Williams, R. H. Hearn, J. R. Zeidler, E. Dong, Jr., and R. C. Goodlin, “Adaptive noise cancelling: Principles and applications,” Proc. of IEEE, Vol. 63, No. 12, pp.1692–1716, Dec. 1975.

[17] O. Hoshuyama, A. Sugiyama, A. Hirano: “A Robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters,” in Proc. of IEEE ICASSP’96, pp. 925–928, May 1996.

[18] O. Hoshuyama, A. Sugiyama, A. Hirano, “A robust adaptive beamformer for microhone Arrays with a blocking matrix using constrained adaptive filters,” IEEE, TRANS. SP, Vol.47, No.10, pp.2677–2684, Oct. 1999.

[19] H. Cox, R. M. Zeskind, M. M. Owen: “Robust adaptive beamforming,” IEEE, TRANS. ASSP, pp.1365–1376, Oct. 1987.

[20] G. C. Goodwin, K. S. Sin, “Adaptive filtering, prediction and control,” (Prentice-Hall, Inglewood Cliffs 1984)

[21] S. Ikeda and A. Sugiyama, “An adaptive noise canceller with low signal-distortion for speech codecs,” IEEE Trans. Sig. Proc., pp.665–674, Mar. 1999.

[22] J. E. Greenberg and P. M. Zurek, “Evaluation of an adaptive beamforming method for hearing aids,” J.A.S.A., vol. 91, no. 3, pp. 1662–1676, Mar. 1992.

[23] O. Hoshuyama, B. Begasse, A. Sugiyama, A. Hirano, “A new adaptation mode control based on cross correlation for a robust adaptive microphone array,” IEICE Trans. Fundamentals, vol.E84-A, No.2, pp.406–413, Feb. 2001.

[24] T. P. Hua, A. Sugiyama, G. Faucon, and R. L. Bouquin-Jeannes, “A new adaptation mode controller for adaptive microphone arrays based on nested and symmetric leaky blocking matrices,” Proc. EUSIPCO2006, Sep. 2006.

[25] T. P. Hua, A. Sugiyama, G. Faucon, and R. L. Bouquin-Jeannes, “A new robust adaptation mode controller based on normalized cross-correlation with multiple symmetric leaky blocking matrices for adaptive microphone arrays,” Proc. EUSIPCO2008, P10-2, Aug. 2008.

[26] S. F. Boll, “Suppression of acoustic noise in speech using spectral subtraction,” IEEE Trans. ASSP, vol.27, no. 2, pp. 113–120, Apr. 1979.

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

[28] M. Kato, A. Sugiyama and M. Serizawa, “Noise suppression with high speech quality based on weighted noise estimation and MMSE STSA,” Proc. IWAENC2001, pp. 183–186, Sep. 2001.