A ROBUST ADAPTIVE MICROPHONE ARRAY WITH IMPROVED SPATIAL SELECTIVITY AND ITS EVALUATION IN A REAL ENVIRONMENT

Osamu HOSHUYAMA, *Akihiko SUGIYAMA*, and *Akihiro HIRANO*

Information Technology Research Laboratories, NEC Corporation

E-mail: houshu@dsp.CL.nec.co.jp

ABSTRACT

This paper presents a new robust adaptive microphone array (AMA) and its evaluation in an echoic environment. The proposed AMA is a generalized sidelobe canceller equipped with a variable blocking matrix using coefficient-constrained adaptive filters, and a multiple-input canceller using norm-constrained adaptive filters (NCAFs). Because the NCAFs have selective nonlinearity in the relationship between coefficient norm and coefficient error, the proposed AMA has better spatial selectivity than the conventional AMA. Evaluation with real acoustic data captured in a room of 0.3-second reverberation time shows that the noise is suppressed by 19 dB. In subjective evaluation, the proposed AMA obtains 3.8 on a 5-point mean opinion score scale.

1. INTRODUCTION

Adaptive microphone arrays (AMAs) have been studied for teleconferencing, hands-free telephones, and speech enhancement for the reason that, in principle, they can attain high noise-reduction performance with a small number of microphones arranged in small space [1]–[7]. In actual environment, target-signal cancellation caused by array imperfections is a serious problem. Array imperfection includes errors in the microphone position, the microphone gain, and the target DOA (direction of arrival). For teleconference and hands-free telephone conversation in a car, the error in the target DOA is the largest factor. Target-signal cancellation is more serious in rooms with reverberation [4].

An AMA with robustness against a large target-DOA error has been proposed [7]. This AMA can be implemented with just several microphons and has high spatial selectivity, i.e. noise-reduction performance. However, this selectivity is not sufficient under severe conditions. The noise from a DOA near the allowable target-DOA range may not be attenuated enough.

This paper proposes a new robust AMA with improved spatial selectivity. The proposed AMA uses norm-constrained adaptive filters in its multiple-input canceller. The coefficients of the norm-constrained adaptive filters can converge to the optimum for noise suppression even if the noise DOA is near the target DOA. Therefore, the proposed AMA has improved spatial selectivity. The proposed AMA is evaluated in a real environment which has reverberation of 0.3 second.

2. CONVENTIONAL AMA

Structure of a conventional robust AMA in [7] with M microphones is shown in Fig.1. It is a generalized sidelobe canceller (GSC). A GSC consists of a fixed beamformer (FBF), a blocking matrix (BM), and a multiple-input canceller (MC). The FBF enhances the target signal. On the contrary, the BM rejects the target signal and passes noise components. The MC adaptively subtracts the components correlated to the output signals $y_m(k)$ of the BM, from the delayed output signal $d(k-Q)$ of the FBF, where k is the sample index and Q is the number of delay samples for causality. If the target signal leaks into $y_m(k)$ in the BM and filter coefficients in the MC grow excessively, target signal as well as noise components is cancelled at the MC. Target-signal cancellation is recognized as attenuation of some frequency components. Sometimes, breathing noise is also heard.

The conventional robust AMA inhibits undesirable target-signal cancellation using coefficient-constrained adaptive filters (CCAFs) in the BM, and leaky adaptive filters (LAFs) in the MC. In this paper, the conventional AMA is called AMA-LAF after the fact that it uses LAFs in the MC. The CCAFs of the BM are adapted to reduce the target-signal leakage caused by target-DOA errors, and the leakage of the LAFs in the MC also prevents the excess growth in the filter coefficients.

In the BM, the CCAFs behave like adaptive noise cancellers. The signal relationship in the BM with N-tap CCAFs is described by

$$y_m(k) = x_m(k-P) - H_m^T(k)D(k), \quad (1)$$

$$H_m(k) \triangleq [h_{m,0}(k), h_{m,1}(k), \ldots, h_{m,N-1}(k)]^T, \quad (2)$$

$$D(k) \triangleq [d(k), d(k-1), \ldots, d(k-N+1)]^T, \quad (3)$$

$$(m = 0, 1, \ldots, M-1),$$

where $[\cdot]^T$ denotes vector transpose. $x_m(k)$ is the m-th microphone signal, and P is the number of delay samples for causality. $H_m(k)$ is the coefficient vector of the m-th CCAF and $D(k)$ is the signal vector consisting of delayed samples of $d(k)$, which is the output signal of the FBF. In the output signal $y_m(k)$, the components correlated to $d(k)$ are cancelled by the CCAFs. This adaptive cancellation may be considered as target tracking.

Each coefficient of the CCAFs is constrained based on the fact that filter coefficients for target-signal minimization vary significantly with the target DOA. The constraints are designed to inhibit mistracking to noise sources. An example of filter-coefficient variation is illustrated in Fig.2. When the CCAF coefficients are constrained in the hatched region in Fig.2, up to 20-degree error in target DOA could be

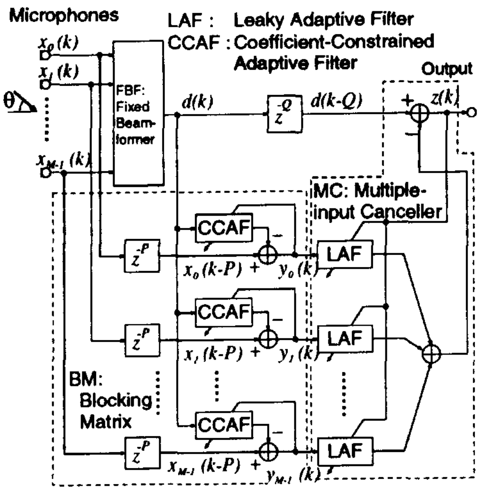

Figure 1. Structure of the Conventional Robust Adaptive Microphone Array in [7].

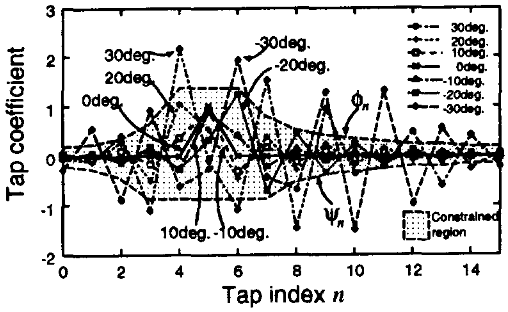

Figure 2. CCAF coefficients to minimize signals from different DOAs and an example of constraints.

allowed. Adaptation of the CCAFs using normalized-least-mean-squares (NLMS) algorithm is described as follows:

$$h'_{m,n} = h_{m,n}(k) + \alpha \frac{y_m(k)\, d(k-n)}{\|D(k)\|^2}, \quad (4)$$

$$h_{m,n}(k+1) = \begin{cases} \phi_{m,n} & \text{for } h'_{m,n} > \phi_{m,n} \\ \psi_{m,n} & \text{for } h'_{m,n} < \psi_{m,n} \\ h'_{m,n} & \text{otherwise} \end{cases}, \quad (5)$$

$$(m = 0, 1, \ldots, M-1), (n = 0, 1, \ldots, N-1),$$

where $\|\cdot\|$ denotes the Euclidean norm. $h'_{m,n}$ are temporal coefficients for limiting functions, and α is the step size. $\phi_{m,n}$ and $\psi_{m,n}$ are the upper and lower limits for each coefficient.

In the MC, the LAFs subtract the components correlated to $y_m(k)$ from $d(k-Q)$. Let L be the number of taps in each LAF, $W_m(k)$ and $Y_m(k)$ be coefficient and signal vectors of the m-th LAF, and $z(k)$ be the output signal of the MC, respectively. The signal processing in the MC is described by

$$z(k) = d(k-Q) - \textstyle\sum_{m=0}^{M-1} W_m^T(k) Y_m(k), \quad (6)$$

$$W_m(k) \triangleq [w_{m,0}(k), w_{m,1}(k), \ldots, w_{m,L-1}(k)]^T, \quad (7)$$

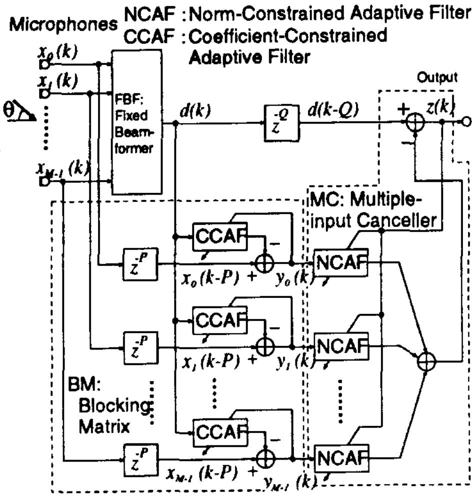

Figure 3. Structure of the Proposed Robust Adaptive Microphone Array.

$$Y_m(k) \triangleq [y_m(k), y_m(k-1), \ldots, y_m(k-L+1)]^T, \quad (8)$$
$$(m = 0, 1, \ldots, M-1).$$

The coefficients of the LAFs are updated by the NLMS algorithm with leakage as follows:

$$W_m(k+1) = (1-\gamma) W_m(k) + \beta \frac{z(k)\, Y_m(k)}{\sum_{j=0}^{M-1} \|Y_j(k)\|^2}, \quad (9)$$

where β is a step size and γ is the constant for leakage. The leakage prevents excess growth in the tap coefficients, which is the cause of target signal cancellation.

The adaptation of the CCAFs and LAFs is controlled based on SNR (signal-to-noise ratio). Adaptation of the CCAFs is carried out when the SNR is high enough. On the contrary, the adaptation of the LAFs is performed during low-SNR periods.

This AMA has robustness against a large target-DOA error and can be implemented with small number of microphones. However, the amount of the leakage in the LAFs varies continuously with the coefficient value, which causes the sensitivity outside the allowable target-DOA range to drop gradually, as shown by line A in Fig.5. If a noise arrives from a DOA near the allowable range, it may not be suppressed enough.

3. PROPOSED AMA

The proposed AMA has improved noise-reduction performance outside the allowable direction range. Figure 3 shows the structure of the proposed AMA. It uses norm-constrained adaptive filters (NCAFs) in its MC instead of the LAFs in the AMA-LAF.

The coefficients of NCAFs are updated by the NLMS algorithm with the constraint as follows:

$$W'_m = W_m(k) + \beta \frac{z(k)\, Y_m(k)}{\sum_{j=0}^{M-1} \|Y_j(k)\|^2}, \quad (10)$$

$$\Omega^2 = \textstyle\sum_{m=0}^{M-1} \|W'_m\|^2, \quad (11)$$

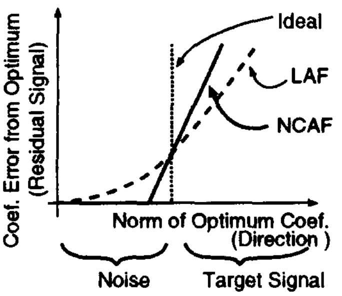

Figure 4. Comparison in Selectivity between LAF and NCAF. (Not quantitative)

$$W_m(k+1) = \begin{cases} \sqrt{\frac{K}{\Omega^2}}\, W'_m & \text{for } \Omega^2 > K \\ W'_m & \text{otherwise} \end{cases} , \qquad (12)$$

where W'_m is a temporal vector for the constraint. Ω^2 and K are the total squared-norm of $W_m(k)$ and a threshold. If Ω^2 exceeds K, $W_m(k+1)$ are restrained by scaling.

Figure 4 illustrates qualitative comparison between LAF and NCAF in the relationship between the norm of the optimum coefficients for signal rejection and the coefficient error from the optimum. Because the error helps the coefficients avoid excess growth, it prevents undesirable target-signal cancellation when the target signal slightly leaks into the MC inputs. The larger the error is, the less target signal the MC cancels. For an ideal spatial selectivity, all the target signal should be passed and only the noise components should be rejected.

Both the norm constraint and the leakage give norm-error characteristics approximating the ideal nonlinearlity. However, the nonlinearity of the NCAF is a better approximation to the ideal nonlinearity than that of the LAF as in Fig.4. The error from the optimum with the LAF varies continuously with the norm of the coefficients. On the other hand, the error with the NCAF becomes effective only when the norm of the coefficients exceeds the threshold, otherwise it has no effect. Therefore, the NCAF leads to sharper spatial selectivity.

The proposed AMA has another advantage over the AMA-LAF. The leakage constant γ, which defines the amount of leakage, should be settled depending on the step size β. However, the threshold of norm, K, can be settled independent of β. The step size is an important factor for extracted target signal quality. Therefore, the independent characteristics of the parameters with the proposed AMA are useful when the user wants to control the signal quality.

4. EVALUATION

4.1. Anechoic Environment

Simulations for an anechoic environment with a linear, 4.1cm equispaced, 4-channel broad-side array were performed. The signal of each microphone was bandlimited between 0.3 and 3.4kHz and sampled at 8kHz. The number of taps was 16 for both the CCAFs and NCAFs. The FBF

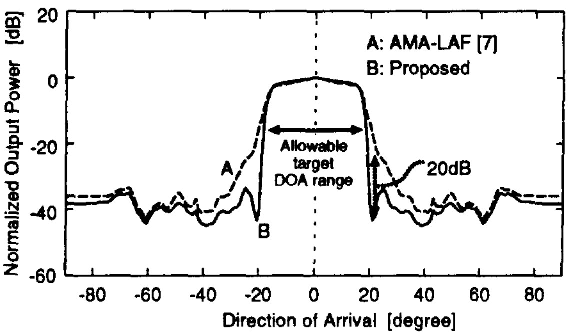

Figure 5. Sensitivity after Convergence as a function of DOA.

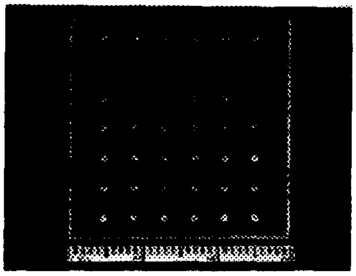

Figure 6. Microphone Array Used for Experiments.

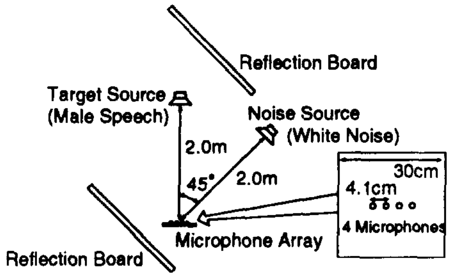

Figure 7. Arrangement for Experiments.

was a simple delay-and-sum beamformer as follows:

$$d(k) = \frac{1}{M}\sum_{m=0}^{M-1} x_m(k). \qquad (13)$$

The step size α for the CCAFs was 0.1 and β for the NCAFs was 0.2. These step sizes were selected so that breathing noise and cancellation of the target signal are sufficiently small subjectively. All other parameters were settled based on the microphone arrangement.

The sensitivity as a function of DOA for a band-limited white Gaussian signal is plotted in Fig.5. Total output power normalized by the power in the assumed target DOA was used as the measure of sensitivity. The sensitivity of the proposed AMA at $\theta = 22$ degrees is 20dB lower than that of the AMA-LAF. This difference indicates the improved spatial selectivity of the proposed AMA.

4.2. Echoic Environment

Simulations with real sound data captured in an echoic environment were also performed. The data were acquired with

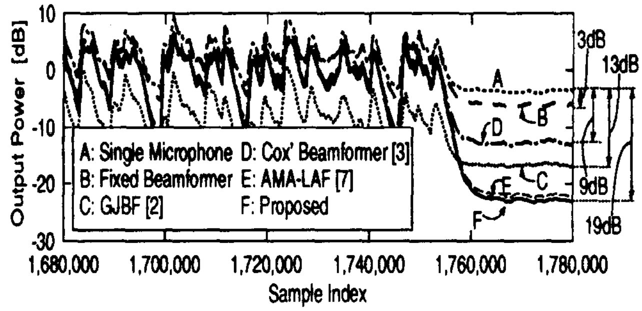

Figure 8. Output Powers for a Male Speech and a White Noise

broad-side linear array as shown in Fig.6. Forty-two omni-directional microphones without calibration are mounted on a universal printed circuit board with an equal spacing of 4.1cm (1.6 inch). Four microphones in the center of the board were used.

Figure 7 illustrates the arrangement of equipment for sound-data acquisition. The target source was in front of the array at a distance of 2.0m. A white noise source was placed about 45 degrees off the target DOA at a distance of 2.0m. The reverberation time of the room was about 0.3 second, which is common with actual small offices. All the parameters except the step sizes were the same as those in the previous section. The target source used was a male speech in English. For comparison, simulations of a fixed beamformer (FBF), a simple Griffiths-Jim beamformer (GJBF) [2], and a Cox' beamformer (Cox) [3] were also carried out.

4.2.1. Objective Evaluation

Output powers for all the methods after convergence are shown in Fig.8. The step sizes used were 0.02 for α and 0.004 for β. If there is any difference between trajectory A and any of B, C, D, E, or F when voice is active (e.g. sample index from 1,720,000 to 1,740,000), the target signal corresponding to the trajectory is partially cancelled. The FBF (B) causes almost no target-signal cancellation. With the GJBF (C), cancellation of the target signal is serious. With the Cox' beamformer (D), the AMA-LAF (E), and the proposed AMA (F), the cancellation of target signal was just 2dB, which is subjectively negligible.

The output powers during voice absence (after 1,760,000-th sample) indicate noise-reduction ratio (NRR). The NRR of the FBF is just 3dB, and that of the Cox' beamformer is 9dB. On the other hand, With the proposed AMA (F), the NRR is as much as 19dB. The NRR of the AMA-LAF (E) is 18dB which is almost the same as the proposed AMA. In this scenario, there is not significant difference between the AMA-LAF and the proposed AMA.

4.2.2. Subjective Evaluation

MOS (Mean Opinion Score) evaluation by 10 nonprofessional subjects with loudspeaker listening was performed based on [8]. As anchors, the signal captured by a single microphone was used for grade 1, and the original male speech without noise, for grade 5. Subjects were instructed that target-signal cancellation should obtain low score.

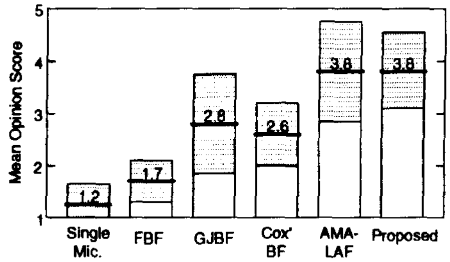

Figure 9. Mean Opinion Score

Evaluation results are shown in Fig.9. The AMA-LAF and the proposed AMA obtained 3.8 point, which is the highest of all the AMAs.

5. CONCLUSION

A new AMA with improved spatial selectivity and its evaluations with real acoustic data have been presented. The proposed AMA is equipped with an adaptive blocking matrix using coefficient-constrained adaptive filters and a multiple-input canceller using norm-constrained adaptive filters. In the direction near the allowable target-DOA range, the proposed AMA has shown maximaly 20dB higher noise-reduction ratio than the AMA-LAF. In a room with 0.3-second reverberation time, the proposed AMA suppresses a noise by 19 dB. MOS evaluation has shown that the proposed AMA obtained 3.8 point on a 5-point scale.

Acknowledgement

The authors would like to thank Dr. Takao Nishitani, Deputy General Manager of Information Technology Research Laboratories, NEC Corporation, for his guidance and valuable comments.

REFERENCES

[1] J. L. Flanagan, et al., "Autodirective Microphone Systems," Acustica, vol.73, pp.58–71, 1991.

[2] L. J. Griffiths, C. W. Jim, "An Alternative Approach to Linear Constrained Adaptive Beamforming," IEEE, Trans. AP, vol.AP-30, no.1, pp.27–34, Jan. 1982.

[3] H. Cox, et al., "Robust Adaptive Beamforming," IEEE, Trans. ASSP, vol.35, no.10, pp.1365–1376, Oct. 1987.

[4] Y. Grenier, "A Microphone Array for Car Environments," Speech Communication, vol.12, no.1, pp.25–39, Mar. 1993.

[5] M. W. Hoffman, et al., "Robust Adaptive Microphone Array Processing for Hearing Aids: Realistic Speech Enhancement," J.A.S.A., vol.96, pp.759–770, 1994.

[6] A. Wang, et al., "A High Performance Microphone Array System for Hearing Aid Applications" Proc. ICASSP'96, pp.3197–3200, 1996.

[7] O. Hoshuyama, A. Sugiyama, "A Robust Adaptive Beamformer for Microphone Arrays with a Blocking Matrix Using Constrained Adaptive Filters," Proc. ICASSP'96, pp.925–928, 1996.

[8] H. R. Silbiger, "Audio Subjective Test Methods for Low Bit Rate Codec Evaluations," ISO/IEC JTC1/SC29/WG11/N0981, Jul. 1995.