A Robust Adaptive Microphone Array with Improved Spatial Selectivity and Its Evaluation in a Real Environment

空間選択性を改善したロバスト適応マイクロホンアレイとその実環境における評価

Osamu HOSHUYAMA Akihiko SUGIYAMA Akihiro HIRANO
宝珠山 治 杉山 昭彦 平野 晃宏
Information Technology Research Laboratories, NEC Corporation
NEC 情報メディア研究所

あらまし

本論文では、空間分離能力を改善したロバスト適応マイクロホンアレーを提案し、その実環境における評価について報告する。提案法は、一般化サイドローブキャンセラにおいて、タップ係数拘束適応フィルタによる可変ブロッキング行列と、ノルム拘束適応フィルタ (NCAF) による多入力キャンセラを用いている。NCAF において、フィルタ係数の大きさに対するフィルタ係数誤差の非線形性が選択的であるため、提案法の空間分離能力は、従来法を上回る。提案法を残響時間約 0.3 秒の部屋で取得した実際のデータで評価した場合、雑音を 19dB 抑圧することができる。また、MOS(Mean Opinion Score) による主観評価では、5 点満点中の 3.8 点が得られる。

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]–[8]. In actual environment, target signal cancellation caused by array imperfections is a serious problem [5]. 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.

An adaptive microphone array with robustness against large target-DOA error has been proposed [8]. 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 a severe condition. 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. Because the constraint does not prevent noise reduction outside the allowable target-DOA range, it improves noise-reduction performance.

In the following section, the conventional AMA and a problem this paper addresses are described. Section 3 presents the new AMA and its advantages. In Section 4, the proposed AMA is evaluated by simulations with sound data generated by computers and with sound data obtained in a real acoustic environment.

2 Conventional AMA

Structure of a conventional AMA in [8] is shown in Fig.1. It is a generalized sidelobe canceller (GSC). GSC consists of a fixed beaformer (FBF), multiple-input canceller (MC), and blocking matrix (BM). The FBF enhances the target signal. $d(k)$ is the output signal of the FBF at sample index k, and $x_m(k)$ is the output signal of the m-th microphone ($m = 0, ..., M-1$). The MC adaptively subtracts the components correlated to the output signals $y_m(k)$ of the BM, from the delayed output signal $d(k-Q)$

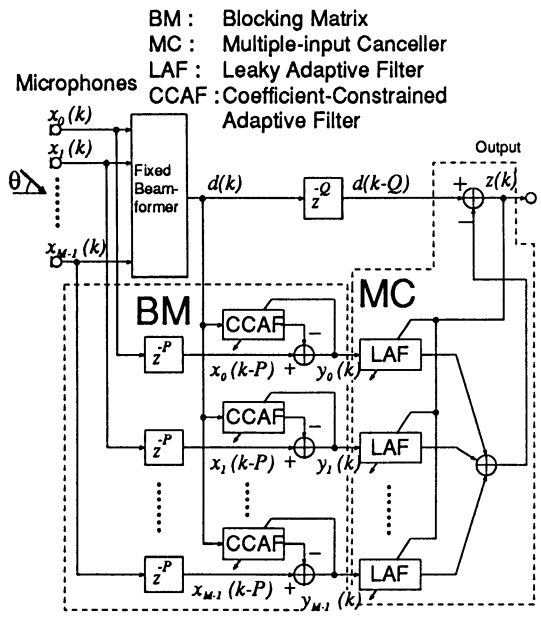

Figure 1: Structure of the Conventional Robust Adaptive Microphone Array in [8].

of the FBF, where Q is the number of delay samples for causality. The BM is a kind of spatial rejection filter. It rejects the target signal and passes noises. If the input signals $y_m(k)$ of the MC, which are the output signals of the BM, contain only noises, the MC rejects the noises and extract the target signal. However, if the target signal leaks into $y_m(k)$ in the BM and filter coefficients in the MC grow excessively, target-signal cancellation occurs at the MC. Target-signal cancellation is recognized as attenuation of high-frequency components. Sometimes, breathing noise is also heard.

The conventional AMA uses 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, thus inhibit undesirable target-signal cancellation. The leakage of the LAFs in the MC also prevents the undesirable target-signal cancellation in case of incomplete adaptation with the CCAFs.

In the BM, the CCAFs behave like adaptive noise cancellers. The input signal of each CCAF is the output signal of the FBF and the output of the CCAF is subtracted from the delayed microphone signal. The signal relationship in the BM with N-tap CCAFs is described by

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

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

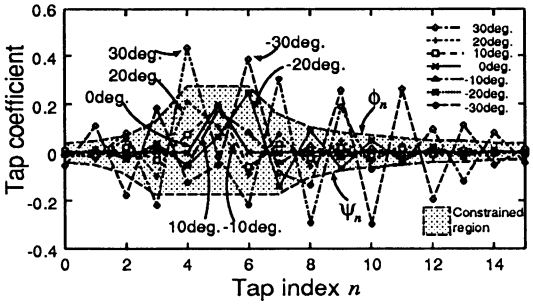

Figure 2: An example of CCAF coefficients to minimize signals from different DOAs and an example of CCAF constraints.

$$D(k) \triangleq [d(k), d(k-1), \ldots, d(k-N+1)]^T, \quad (3)$$
$$(m = 0, 1, \ldots, M-1),$$

where $\cdot^T$ denotes vector transpose. 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 signals 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 is a kind of 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 allowed. Only the signal that arrives from a DOA in the limited DOA region is minimized at the outputs of the BM and remains at the output of the MC. The CCAF coefficients $h_{m,n}(k)$ are adapted with coefficient constraints. Adaptation using normalized-least-mean-squares (NLMS) algorithm is described as follows:

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

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

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

where $\|\cdot\|$ denotes the Euclid 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 $y_m(k)$, $(m = 0, ..., M-1)$, 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

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

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

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

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

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

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

where 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 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:

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

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

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

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

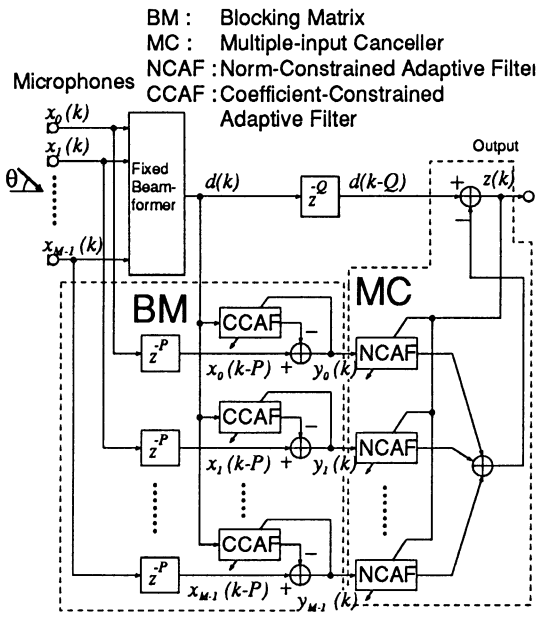

Figure 3: Structure of the Proposed Robust Adaptive Microphone Array.

where β is a step size, W'_m is a temporal vector for the constraint, and 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 means avoiding excess growth in the coefficients, it prevents undesirable target-signal cancellation when the target signal slightly leaks into the MC inputs. The larger the error is, the more target signal the MC saves. For an ideal spatial selectivity, all the target signal should be saved and only the noises should be rejected.

Both the norm constraint and the leakage give nonlinearities approximating the ideal nonlinearlity. However, the nonlinearity of the NCAF is the better approximation to the ideal nonlinearity than that of the LAF. 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 if 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 independently of β. As illustrated later, the step size is important factor for extracted target signal quality. Therefore, the in-

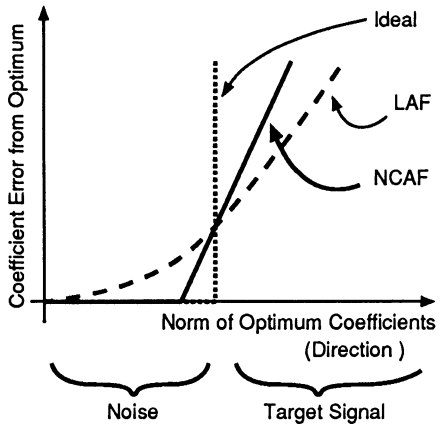

Figure 4: Comparison in Selectivity between LAF and NCAF. (Not quantitative)

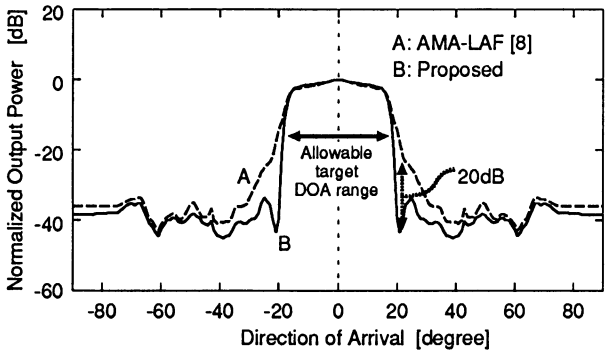

Figure 5: Sensitivity vs. DOA.

dependent characteristic of the parameters with the proposed AMA is useful if the user wants to adjust the signal quality.

4 Evaluation

The proposed AMA was evaluated in anechoic environment simulated by computers and in a real environment which has echoes. In the former environment, it was compared with the AMA-LAF in sensitivity pattern. In the latter environment, it was evaluated objectively with SNR and subjectively with MOS.

4.1 Evaluation in 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 from 0.3 to 3.4kHz and sampled at 8kHz. The number of taps was 16 for both the CCAFs and NCAFs. The step size α for the CCAFs was 0.02 and

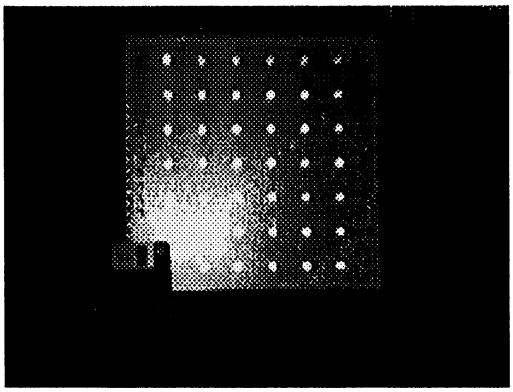

Figure 6: Microphone Array Used for Experiments.

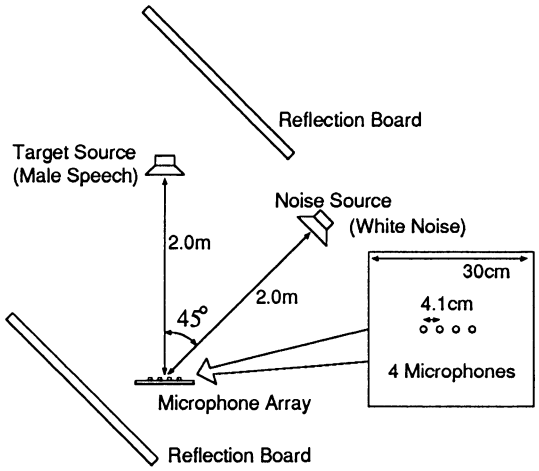

Figure 7: Equipment Arrangement in Experiments.

β for the NCAFs was 0.004. 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 bandlimited 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 Evaluation in Echoic Environment

Simulations with real sound data captured in an echoic environment were also performed. The data were acquired with broad-side linear array as shown in Fig.6. Forty-two omni-directional microphones without calibration are attached to a universal printed circuit board with an equal spacing of 4.1cm (1.6 inch). Four microphones on the center of the

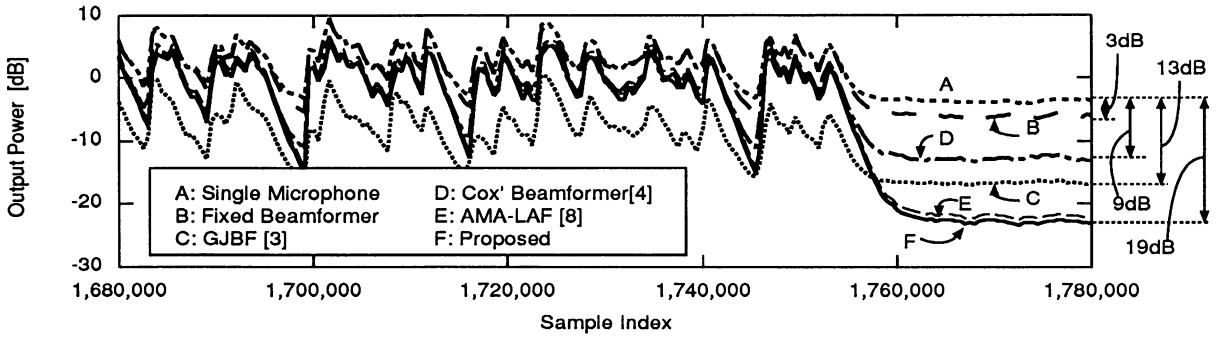

Figure 8: Output Powers for a Male Speech and a White Noise

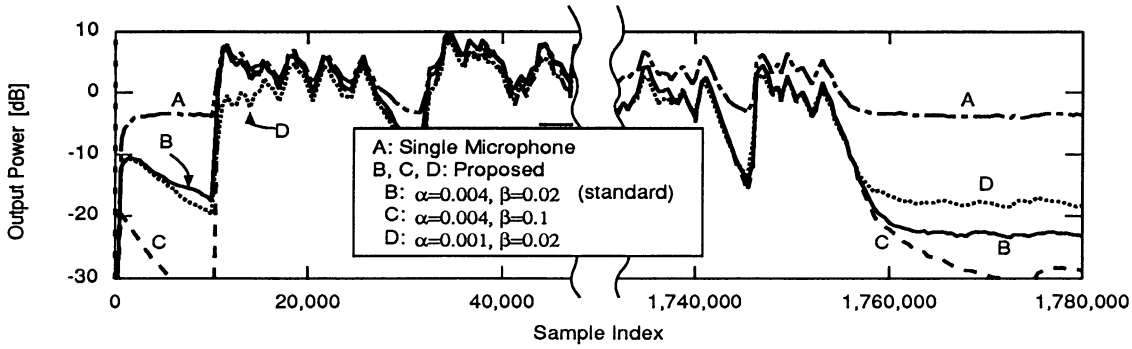

Figure 9: Output Powers for Different Step Sizes

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) [3], and a Cox' beamformer (Cox) [4] were also carried out. The output signals were evaluated objectively comparing the output powers in noise-reduction performance and extracted target-signal quality. Subjective evaluation in MOS (Mean Opinion Score) with loudspeaker listening was 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 (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.

The output powers for different step sizes are shown in Fig.9. In the left half of the figure, convergence and target-signal cancellation are compared. NRRs are compared in the right half of the figure. When the step size β in the NCAFs was larger (C) than the standard (B), the noise could be suppressed more rapidly (before 20,000-th sample). However, the breathing noise was increased subjectively. In contrast, when the step size α in the CCAFs was smaller (D), the breathing noise was decreased. However, cancellation of the target signal is increased in the beginning of adaptation (sample index from 10,000 to 20,000) and the final NRR became smaller

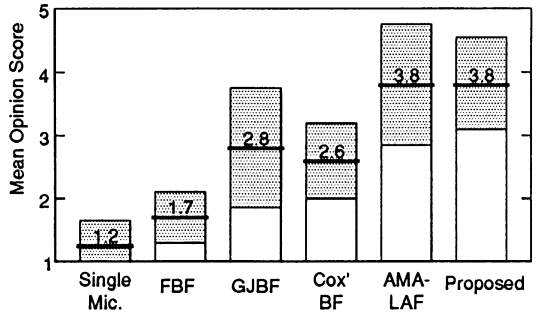

Figure 10: Mean Opinion Score

(sample index from 1,760,000 to 1,780,000). Therefore, step-size selection is important for the proposed AMA in terms of extracted signal quality. A step-size-control method or a new adaptation algorithm may be useful.

4.2.2 Subjective Evaluation

MOS evaluation by 10 nonprofessional subjects was performed based on [9]. As an anchor, 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.

Evaluation results are shown in Fig.10. The FBF obtained 1.7 point because the number of microphone is so small that its NRR is low. The GJBF reduced noises very much but target signal cancellation is serious, thus it was scored 2.8 point. The Cox' beamformer marked 2.6 point for its 9dB noise-reduction capability. 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 has been proposed and its evaluations with real acoustic data have been presented. The proposed AMA is equipped with an adaptive blocking matrix using CCAFs and a multiple-input canceller using NCAFs. 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. Two step sizes rule the performances: noise reduction, breathing noise, and target-signal cancellation. Therefore, step-size selection is important for the proposed AMA. 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 Signal Processing Research Laboratory, Information Technology Research Laboratories, NEC Corporation, for his guidance and valuable comments.

References

[1] Y. Kaneda, "Adaptive Microphone Arrays," Trans. IEICE, Vol.BII, pp.742–748, 1992. (in Japanese)

[2] J. L. Flanagan, D. A. Berkley, G. W. Elko, W. M. M. Sondhi, "Autodirective Microphone Systems," Acustica, Vol.73, pp.58–71, 1991.

[3] 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.

[4] H. Cox, R. M. Zeskind, M. M. Owen, "Robust Adaotive Beamforming," IEEE, Trans. ASSP, vol.35, no.10, pp.1365–1376, oct. 1987.

[5] Y. Grenier, "A Microphone Array for Car Environments," Speech Communication, Vol.12, No.1, pp.25–39, Mar. 1993.

[6] M. W. Hoffman, T. D. Trine, K. M. Buckley, D. J. Tasell, "Robust Adaptive Microphone Array Processing for Hearing Aids: Realistic Speech Enhancement," J.A.S.A., vol.96, pp.759–770, 1994.

[7] A. Wang, K. Yao, R. E. Hudson, D. Korompis, S. F. Soli, S. Gao, "A High Performance Microphone Array System for Hearing Aid Applications" Proc. ICASSP'96, pp.3197–3200, 1996.

[8] 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.

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