A REALTIME ROBUST ADAPTIVE MICROPHONE ARRAY CONTROLLED BY AN SNR ESTIMATE

Osamu HOSHUYAMA†, Brigitte BEGASSE ‡, Akihiko SUGIYAMA†, Akihiro HIRANO †

†C&C Media Research Laboratories
NEC Corporation

‡Institut National des Sciences Appliquées

ABSTRACT

A robust adaptive microphone array (RAMA) using a new adaptation-mode control method (AMC) is proposed, and its evaluation results by hardware are presented. The adaptation of the RAMA is controlled based on an SNR (signal-to-noise) estimate using the output powers of the fixed beamformer and the adaptive blocking matrix. The RAMA is implemented on a multi-DSP realtime signal-processing system with a C-compiler. Simulation results with real acoustic data show that the AMC based on the SNR estimate causes less breathing noise than the conventional AMC and that it obtains 1.0-point higher score on a 5-point mean opinion score scale. Evaluation through a realtime signal-processing system demonstrates that noise reduction achieved by the RAMA is over 12 dB even in reverberant environments.

1. INTRODUCTION

Adaptive microphone arrays (AMAs) have been studied for teleconferencing, hands-free telephones, and speech enhancement, because they can reduce a great deal of noise through the use of a small number of microphones arranged in small space [1]–[8]. In an actual environment, target-signal cancellation caused by imperfections in the array is a serious problem. These include errors in the microphone position, the microphone gain, and the target DOA (direction of arrival). For teleconferencing and hands-free telephone conversation in a car, error in the target DOA is the dominant factor.

To cope with such a large target-DOA error, a robust AMA has been proposed [4]. This robust AMA (RAMA) can be implemented with just several microphones and it has high spatial selectivity, i.e. it is effective in reducing noise. This RAMA uses two-stage adaptive signal processing: the first stage is an adaptive blocking matrix (ABM) to reduce target-signal cancellation and the second stage is a multiple-input canceller (MC) with a norm constraint. The RAMA has both the ability to reduce noise effectively and good target-signal quality [4]. However, there is a trade-off between noise reduction and target-signal cancellation, because adaptations in the ABM and the MC may interfere with each other.

To avoid this trade-off, adaptation-mode control (AMC) is an important issue for RAMAs with an ABM (RAMA-ABM). However, only a few AMC methods have been reported. The AMC method proposed by Greenberg et al. is based on the cross-correlation of two microphone signals, and it is effective for RAMAs without an ABM [8]. However, it sometimes causes serious target-signal cancellation or undesirable breathing noise, when used for the RAMA-ABM. Therefore, a new AMC should be developed for the RAMA-ABM.

A part of this work was conducted as a training program between INSA and NEC.

Once the design of the signal processing algorithm has been completed in AMAs including the RAMA-ABM, evaluation by hardware is important. This is because the signal processing algorithms are complicated and some problems, unpredictable in the simulations, may occur in a real environment. The RAMA-ABM should also be evaluated by a realtime processor in real environments as conventional AMAs [1][2][5]–[7] have been.

This paper proposes a RAMA-ABM equipped with a new AMC and presents its evaluation on a multi-DSP realtime signal-processing system. The new AMC causes less breathing noise and it is able to obtain high mean opinion score. The RAMA-ABM with the new AMC is implemented on a multi-DSP realtime signal-processing system. Its directional response is measured in real environments.

2. STRUCTURE OF RAMA

2.1. RAMA-ABM

The structure of the RAMA-ABM [4] with the new AMC is shown in Fig. 1. The RAMA-ABM consists of a fixed beamformer (FBF), an MC, and an ABM. The FBF enhances the target signal. In Fig. 1, $x_m(k)$ is the output signal of the m-th microphone $(m = 0, ..., M-1)$ at a sample index k, and $d(k)$ is the output signal of the FBF. The MC adaptively subtracts the components correlated to the output signals $y_m(k)$ of the BM, from the output signal $d(k)$ of the FBF. The ABM is a spatial rejection filter. It rejects the target signal and passes the noise. When the input signals $y_m(k)$ of the MC, which are the output signals of the ABM, contain only the noise components, the MC rejects the noise and extracts the target signal.

If the target signal leaks into $y_m(k)$ in the ABM, target-signal cancellation occurs at the MC. Target-signal cancellation is recognized as the attenuation of high-frequency components. Sometimes, breathing noise can also be heard. To reduce the target-signal leakage, the ABM adaptively subtracts the components correlated to the output signal $d(k)$ of the FBF from the microphone signals $x_m(k)$. The coefficients in the MC and the ABM are updated by the NLMS (normalized LMS) algorithm with constraints [4].

The RAMA-ABM has high spatial selectivity. However, it requires AMC based on target-signal detection both in the ABM and the MC. The adaptations in the ABM and in the MC need classification. This is because of the contrary relationships between the desired signal and the noise for the adaptation algorithm. For the adaptation algorithm in the ABM, the target signal is the desirable signal and the noises are the undesirable signals. Therefore, the SNR (signal-to-noise ratio) should be high in terms of the convergence speed and reducing target-signal leakage. In the MC, however, the noises are the desirable signals and the target signal is the

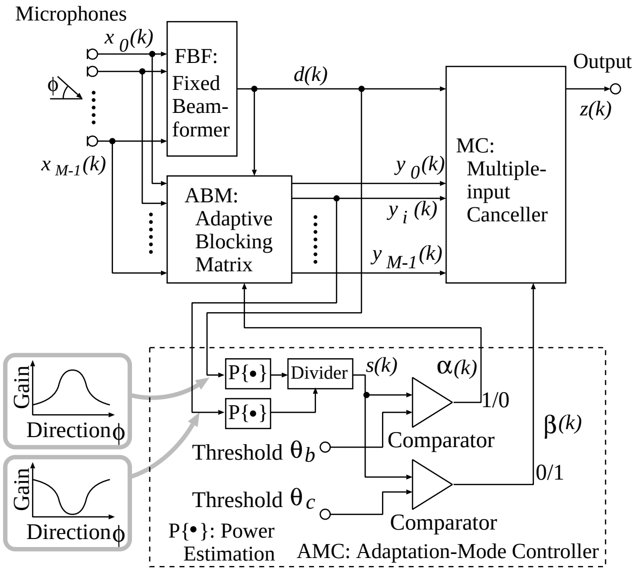

Figure 1: Structure of RAMA-ABM with AMC based on SNR estimate.

undesirable signal. Therefore, a low SNR is better for the MC. This is why the RAMA-ABM needs an AMC method similar to double-talk detectors for echo cancellation.

2.2. Adaptation-Mode Control

The RAMA in Fig. 1 uses a new AMC method based on an SNR estimate (AMC-SE). AMC-SE consists of two power estimators, a divider, and two comparators. $s(k)$, the index for AMC, is the power ratio of the FBF output signal $d(k)$ to an output signal $y_i(k)$ of the ABM. When the index is larger than the threshold θ_b, the adaptation of the ABM is performed. On the other hand, the MC is adapted when the index is smaller than another threshold θ_c.

The adaptation step size $\alpha(k)$ for adaptive filters in the ABM is controlled as follows:

$$\alpha(k) = \begin{cases} 1 & \text{for } s(k) > \theta_b \\ 0 & \text{otherwise} \end{cases}, \tag{1}$$

$$s(k) = \frac{p_d(k)}{p_y(k)}, \tag{2}$$

$$p_d(k) = (1-\gamma)\, p_d(k-1) + \gamma\, d^2(k), \tag{3}$$

$$p_y(k) = (1-\gamma)\, p_y(k-1) + \gamma\, y_i^2(k), \tag{4}$$

where $p_d(k)$ is a power estimate of $d(k)$, $p_y(k)$ is a power estimate of $y_i(k)$, and i is an integer satisfying $0 \le i \le M-1$. On the contrary, the step size $\beta(k)$ for the MC is controlled in reverse with the other threshold as

$$\beta(k) = \begin{cases} 0 & \text{for } s(k) > \theta_c \\ 1 & \text{otherwise} \end{cases}. \tag{5}$$

The index $s(k)$ can be considered to be a direct estimate of the SNR, because the main component at the FBF output is the target signal and the main components at the ABM output are noises. The power estimates $p_d(k)$ and $p_y(k)$ have large variances, however, the variances corresponding to the same components cancel each other out in the divider. Therefore, even when the SNR is low, the variance in the index $s(k)$ is much smaller than its average. This characteristic of the index leads to less misdetection.

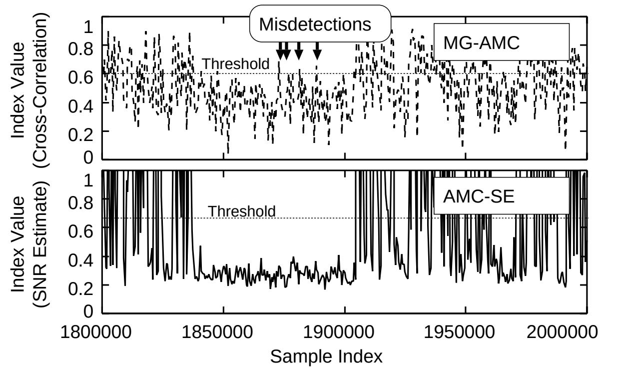

Figure 2: Behavior of the indices. MG-AMC: Modified Greenberg's AMC, AMC-SE: AMC based on SNR estimate.

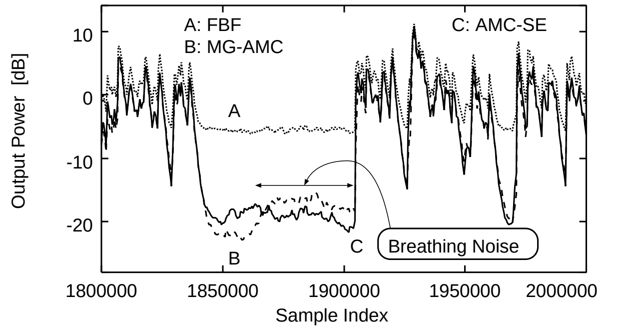

Figure 3: Output powers.

3. EVALUATION OF AMC-SE

3.1. Breathing Noise

AMC-SE was separately evaluated by simulations with real acoustic data. The behavior of the indices for AMC were compared between the AMC-SE and a conventional AMC method (Greenberg's AMC [8] modified for RAMA-ABM). The index in modified Greenberg's AMC method (MG-AMC) is a cross correlation and that in the AMC-SE is an SNR estimate. The data were acquired using a four-microphone linear array and they were sampled at 8 kHz. The reverberation time of the room was 0.3 second. A male-speech source as the target signal was located on a line perpendicular to the array surface, and a white-Gaussian signal as the noise source was located 45-degrees off the target direction. The SNR was about 6 dB. The step sizes selected were 0.02 for the BM and 0.006 for the MC. All the other parameters were the same as in [4].

The two microphones in the center were used for the MG-AMC. All the γ's for low-pass filters were 0.995 both for the AMC-SE and MG-AMC. The thresholds θ_b and θ_c were 0.6 for the MG-AMC, and 0.65 for the AMC-SE. They were selected so that the subjective

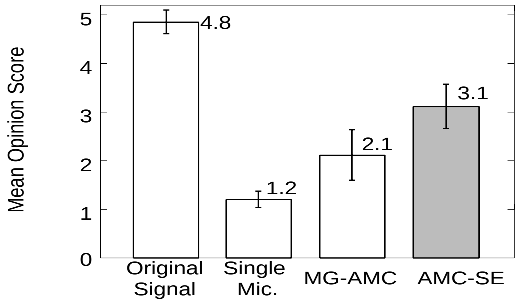

Figure 4: Mean Opinion Score.

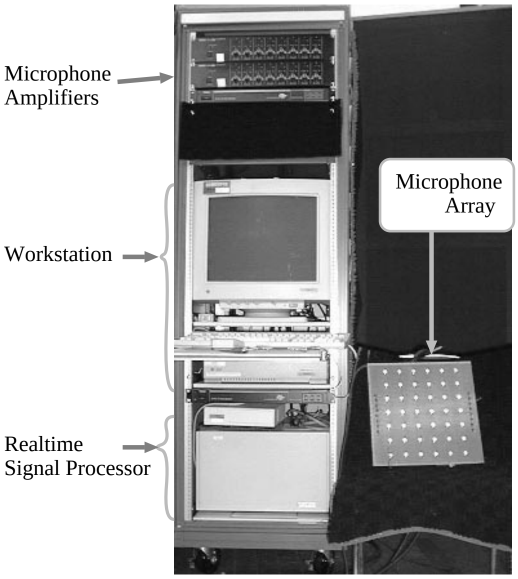

Figure 5: Realtime Signal-Processing System (S-RTP2000) and a Workstation.

degradations in the sound quality by target-signal cancellation were almost the same.

The behavior of the indices is shown in Fig. 2. Whereas the MG-AMC has many misdetections, the AMC-SE achieves almost perfect detection. The output power of the RAMA is also compared in Fig. 3. The AMC-SE avoids the serious breathing noise which occurs with the MG-AMC.

3.2. Subjective Evaluation

A mean opinion score (MOS) evaluation by 17 nonprofessional subjects through loudspeaker listening was performed based on [12]. As anchors, the signal captured by a single microphone with an SNR of 10dB was used for grade 1.0, and the original signal without noise was used for grade 5.0. Subjects were instructed to evaluate reductions in noise and speech quality which included breathing noise. Figure 4 compares the MOSs for output signals with different AMCs. The tops of the bars represent the MOS, and their variances are indicated by the vertical lines. The AMC-SE obtained 3.1 points, which is 1.0-point higher than that for the MG-AMC.

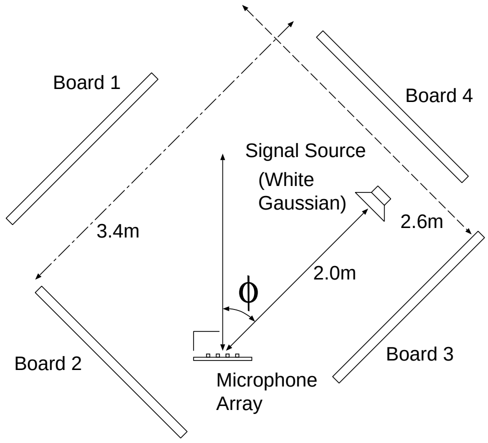

Figure 6: Experimental Set-up.

Table 1: Materials of Boards in Fig. 6.

Environment	Boards 1 & 3	Boards 2 & 4
A	Absorbing	Absorbing
B	Absorbing	Reflecting
C	Reflecting	Reflecting

4. EVALUATION BY REALTIME PROCESSING SYSTEM

4.1. Implementation

The RAMA including AMC-SE was implemented on a realtime signal-processing system, S-RTP2000 by the Systems Design Service Corporation [9]. Twelve DSPs (TI TMS320C40 [10]) were connected each other through communication ports. Sixteen-channel analog-to-digital converters and 16-channel digital-to-analog converters were mounted on the DSP boards. The computational capability of the system was more than adequate and it could be increased by adding DSP boards. The DSP boards were controlled through the VME bus from the workstation. The workstation transferred the program to the DSPs, and data on the DSP boards could be obtained on the workstation. Programming was performed using a C-compiler and assembly language. The system is shown in Fig. 5.

To reduce the delay due to inter-chip communications, data were transferred block by block, and a pipelined structure [11] was employed. The block length was 100 samples and the pipeline depth was 6 stages. The adaptation algorithm in the ABM was the normalized LMS (NLMS) algorithm, and that in the MC was a delayed NLMS algorithm [11].

4.2. Evaluation of Overall System

The noise-reducing performance of the overall system was evaluated on a realtime signal-processing system. An equi-spaced linear array with four omni-directional microphones was used. The microphone spacing was 4.1 cm, and the sampling frequency was 8 kHz. The experimental set-up is shown in Fig. 6. Reverberation in the room was controlled by changing the material for the boards illustrated in Fig. 6 as Table 1 shows. Environment B is more reverberant than environment A, and environment C is the most reverberant. The

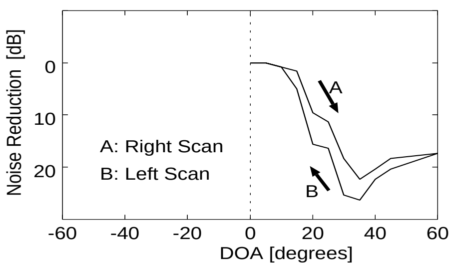

Figure 7: Noise Reduction in Environment A.

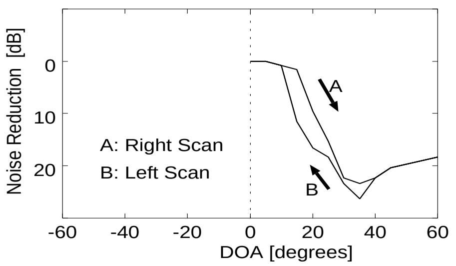

Figure 8: Noise Reduction in Environment B.

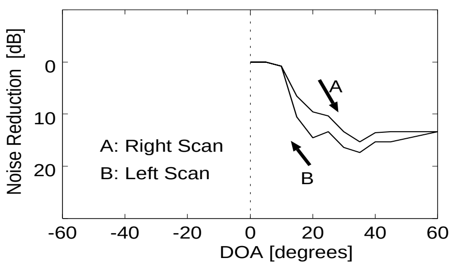

Figure 9: Noise Reduction in Environment C.

reverberation time for the room itself was 0.3 second. The white Gaussian source was scanned in two ways. The constraints of the ABM were set so that the allowable target-direction range was ± 15 degrees. All the other parameters used were the same as in Section 3.1.

When the noise source moved, breathing noise could be observed, however, it disappeared within a few seconds by adaptation. For example, in environment B, when the noise-DOA was 45-degrees off the target direction, the noise could be reduced by 10 dB in about 3 seconds, and by 20 dB in about 20 seconds. The noise reduction after convergence is shown in Figs. 7, 8, and 9. In these figures, a lower curve means better noise reduction. In environments A and B, over 18-dB noise reduction could be achieved in a DOA over 30 degrees. Even in environment C, the poorest noise reduction was more than 12 dB in a DOA over 40 degrees. These results indicate that the RAMA is promising in applications such as voice communications at least in typical environments such as A and B.

Figures 7, 8, and 9 also indicate that noise reduction varies with reverberation. From Figs. 7 and 8, reverberation by reflecting boards barely degrades noise reduction. However, from Figs. 8 and 9, additional reverberation by reflecting boards decreases noise reduction. This characteristic in the relationship between reverberation and noise reduction agrees with the well-known fact that room reverberation degrades noise reduction.

As is clear from Figs. 7, 8, and 9, the ability of the RAMA to reduce noise shows hysteresis for source movement. In terms of spatial selectivity, hysteresis should be reduced. Analysis of the hysteresis and ways to reduce it will be be left for future work.

5. CONCLUSION

A robust adaptive microphone array with a new adaptation-mode control method and its evaluation by hardware have been presented. The adaptation mode is controlled based on an SNR estimate using the output powers of the fixed beamformer and the adaptive blocking matrix. The robust adaptive microphone array was implemented on a multi-DSP system with a C-compiler. Simulation results have demonstrated that the adaptation-mode control method based on the SNR estimate had less breathing noise than the conventional method and that it obtained a 1.0-point higher score on a 5-point MOS scale. Evaluation with a realtime signal-processing system has shown that the implemented adaptive microphone array reduced noise by over 12 dB even in reverberant environments.

ACKNOWLEDGEMENT

The authors would like to thank Mr. Takafumi Kikuchi of the Department of Electrical Engineering, Keio University for his assistance with the realtime evaluation. They would also like to thank Dr. Takao Nishitani, Deputy General Manager of Silicon System Research Laboratories, and Dr. Kazunori Ozawa, Senior Manager of C&C Media Research Laboratories, NEC Corporation for their valuable comments on this study.

REFERENCES

[1] J. L. Flanagan, D. A. Berkley, G. W. Elko, and W. M. M. Sondhi, “Autodirective Microphone Systems,” Acustica, vol.73, pp.58–71, Feb. 1991.

[2] Y. Kaneda and J. Ohga, “Adaptive Microphone Array System for Noise Reduction,” IEEE Trans. ASSP, vol.34, pp.1391–1400, Jun. 1986.

[3] J. Kates, “A Comparison of Hearing-Aid Array-Processing Techniques,” J.A.S.A, vol. 99, no. 5, pp. 3138–3148, May 1996.

[4] O. Hoshuyama, A. Sugiyama, and A. Hirano, “A Robust Adaptive Microphone Array with Improved Spatial Selectivity and Its Evaluation in a Real Environment,” IEEE Proc. of ICASSP’97, vol. I, pp.367–370, Apr. 1997.

[5] G. Elko and A. Pong, “A Steerable and Variable First-Order Differential Microphone Array,” IEEE Proc. of ICASSP’97, vol. I, pp.223–226, Apr. 1997.

[6] P. L. Chu, “Superdirective Microphone Array for a Set-Top Video Conferencing System,” IEEE Proc. of ICASSP’97, vol. I, pp.235–238, Apr. 1997.

[7] H. Silverman, W. R. Patterson III, J. L. Flanagan, and D. Rabinkin, “A Digital Processing System for Source Location and Sound Capture by Large Microphone Arrays,” IEEE Proc. of ICASSP’97, vol. I, pp.251–254, Apr. 1997.

[8] 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.

[9] “Real-Time Signal Processing System for LAN, S-RTPstation Series,” Systems Design Service Corporation, Product Catalog, 1994.

[10] *TMS320C4x User’s Guide*, Texas Instruments, Aug. 1993.

[11] N. R. Shanbhag and K. K. Parhi, *Pipelined Adaptive Digital Filters*, Kluwer Academic Publishers, 1994.

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