A Robust Adaptive Microphone Array Controlled Based on Output Signals of Different Beamformers

複数のビームフォーマの出力信号に基づいて制御するロバスト適応マイクロホンアレイ

Osamu HOSHUYAMA † Brigitte BEGASSE ‡ Akihiko SUGIYAMA † Akihiro HIRANO †
宝珠山 治 † ブリジット ベガッセ ‡ 杉山 昭彦 † 平野 晃宏 †
†C&C Media Res. Labs., NEC Corp. ‡Institut National des Sciences Appliquées
†NEC C&C メディア研究所 ‡INSA Rennes

Abstract

A robust adaptive microphone array (RAMA) with an adaptation mode control (AMC) and its evaluation by hardware are presented. The adaptation of the RAMA is controlled based on an SNR (signal-to-noise) estimate using output powers of the fixed beamformer and the adaptive blocking matrix. The RAMA is implemented on a real-time signal-processing system with multiple DSPs. Simulation results show that the AMC based on the SNR estimate causes less breathing noise than the conventional AMC. Evaluation by a real-time signal-processing system demonstrates that the noise reduction by the RAMA is over 12 dB even in reverberant environments.

あらまし

適応モードを制御するロバスト適応マイクロホンアレイ (RAMA) と、その実機評価について報告する。適応動作の制御は、適応ブロッキング行列 (ABM) の出力と、固定ビームフォーマ (FBF) の出力を用いた信号対雑音比の推定に基いて行われる。この RAMA は、マルチ DSP 実時間処理システム上に実現されている。適応モード制御をコンピュータシミュレーションによって評価し、出力信号における息づき雑音が、従来法より少ないことを示す。また実機評価によって、提案法の雑音除去能力が残響の強い環境においても 12dB 以上あることを示す。

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]–[15]. 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].

A robust AMA against a large target-DOA error (RAMA) has been proposed [9][10]. The RAMA can be implemented with just several microphons and has high spatial selectivity, i.e. noise-reduction performance. It has a trade-off between noise reduction and target-signal cancellation. The RAMA has both high noise-reduction performance and good target-signal quality [9][10]. This RAMA uses two-stage adaptive signal processing: one is an adaptive blocking matrix (ABM) to reduce target-signal cancellation and the other is a multiple-input canceller (MC) with norm constraint.

Because adaptations in the ABM and the MC may interfare each other, adaptation-mode control (AMC) is an important issue for the RAMA with the ABM (RAMA-ABM) [8]. However, only a few AMC methods have been reported. One of these is by Greenberg et al. This AMC is based on cross-correlation of two microphone signals, and is effective for the AMAs without the ABM [14]. However, it sometimes causes serious target-signal cancellation or undesirable breathing noise, when used for the RAMA-ABM.

For AMAs, once the design of the signal processing algorithm has been completed, evaluation by hardware is important. This is because the algorithms are complicated and some problems unpredictable in simulations may occur in the real environment. Most of the existing evaluations have been performed by off-line simulations. Only several evaluations have been performed with real-time processors [2][3][11][12][13].

This paper presents a RAMA-ABM equipped with a new AMC, and its evaluation on a real-time signal-processing system. In the next section, the structure of the RAMA-ABM and the AMC is explained. Implementation of the RAMA on a real-time signal processing system is described in Section 3. Section 4 shows simulation results for evaluating the AMC and experimental results of noise-reduction performance by a hardware system.

2 Structure of RAMA

2.1 RAMA-ABM

The structure of the RAMA which this paper addresses is shown in Fig. 1. This RAMA is a RAMA-ABM [10] with a new AMC. The RAMA-ABM consists of a fixed beamformer (FBF), a multiple-input canceller (MC), and an adaptive blocking matrix (ABM). The FBF enhances the target signal. $d(k)$ is the output signal of the FBF at a 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 output signal $d(k)$ of the FBF. The ABM is a spatial rejection filter. It rejects the target signal and passes noises. When the input signals $y_m(k)$ of the MC, which are the output signals of the ABM. contain only noises, the MC rejects the noises and extract 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 attenuation of high-frequency components. Sometimes, breathing noise is also 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 [9][10].

The RAMA-ABM has high spatial selectivity. However, it requires an adaptation-mode control (AMC) based on target-signal detection both in the ABM and the MC. The adaptations in the ABM and the MC should be performed during almost opposite periods. This is because the relationships between the desired signal and the noise for the adaptation algorithm are contrary. 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 convergence speed and optimality of the coefficients. On the contrary, for the adaptation algorithm in the MC, the noises are the desirable signals and the target signal is the undesirable signal, therefore, low SNR is better for the MC. Therefore, the RAMA-ABM needs an AMC similar to double-talk detectors for echo cancellation.

2.2 Adaptation-Mode Control

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

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

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

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

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

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

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 a reversed way 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 as 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 out each other in the divider. Therefore, even when the SNR is low, the variance of the index $s(k)$ is much smaller than its average. This characteristic of the index leads to less mis-detection.

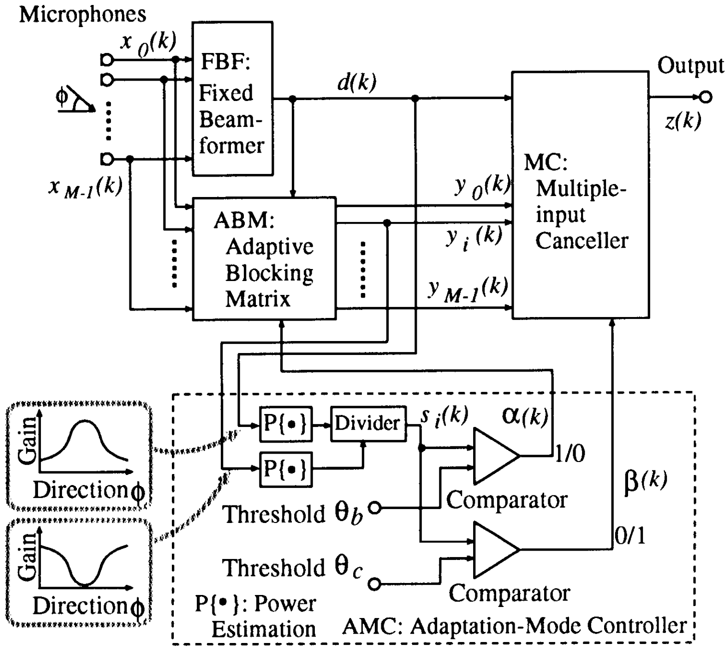

Figure 1: Structure of RAMA-ABM and AMC

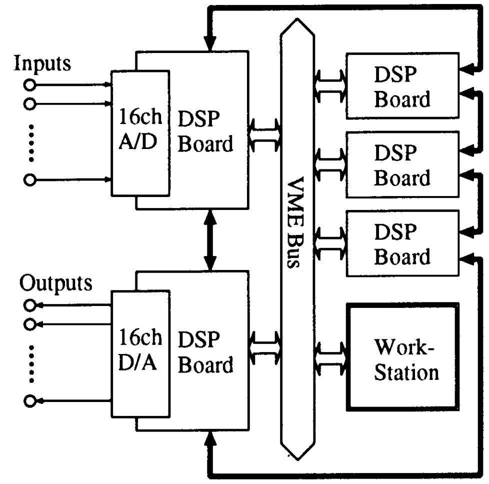

Figure 2: Structure of Real-Time Signal-Processing System

3 Implementation on Real-Time Processing System

In recent years, research of AMAs has made great progress with signal processing algorithms, however, only a few AMAs have been put in the market. Research of AMAs should continue, and the algorithms will be more complicated. In the development of the algorithms, a cycle of evaluation and improvement is important. Therefore, flexibility is the primary concern in the design of a real-time system for evaluation. Flexibility can be achieved with programmable digital signal processors (DSPs).

A multiple-DSP system has a vast amount of computations which is sufficient for the future sophisticated AMA algorithms with more complexity. Generally, programming of assembly language for a multiple-DSP system is still difficult. To accelerate the cycle of the research, the programming should be as easy as possible. Therefore, the programming should be carried out by a high-level language except for computationally critical parts. With these considerations, a real-time signal-processing system, S-RTP2000 by Systems Design Service Corporation [16] has been employed.

A block diagram of the real-time signal-processing system is shown in Fig. 2. Each DSP board has 2 or 4 DSPs (TI TMS320C40 [17]), and total number of DSPs is 12. The DSPs are connected each

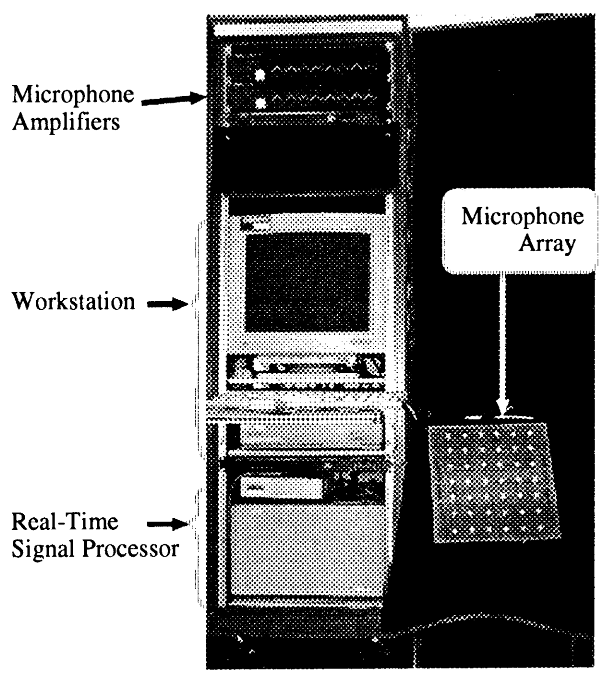

Figure 3: Real-Time Signal-Processing System (S-RTP2000) and a Workstation

other through communication ports. A daughter board with 16-channel analog-to-digital converters and another with 16-channel digital-to-analog converters are mounted on DSP boards. The computational capability of the system can be increased by additional DSP boards. The DSP boards are controlled through the VME bus from a workstation. The workstation transfers the program to the DSPs, and data on the DSP boards can be obtained on the workstation. Programming was performed using a C-compiler and an assembly language. The system is shown in Fig. 3.

To reduce the delay by inter-chip communications, data were transferred block by block, and pipelined structure [18] was used. The block length was 100 samples and the pipeline depth was 6 stages. Adaptation algorithm in the ABM was the normalized LMS (NLMS) algorithm, and that in the MC was a delayed NLMS algorithm [18].

4 Evaluations

4.1 Evaluation of AMC-SE

AMC-SE was independently evaluated by simulations. The behavior of the indices for AMC were compared between AMC-SE and a conventional AMC (Modified Greenberg's AMC: MG-AMC) [15]. The index in MG-AMC is cross correlation and that in AMC-SE is an SNR estimate. The data were acquired by a four-microphone linear array and sampled at 8 kHz. Reverberation time of the room was 0.3 second. A male-speech source as the target signal was located on the line perpendicular to the array surface, and a white-Gaussian signal as a noise source was located 45-degree 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 [10].

The two microphones in the center were used for MG-AMC. The thresholds $\theta_b = \theta_c = 0.6$ for MG-AMC and $\theta_b = \theta_c = 0.65$ for AMC-SE. They were selected so that subjective degradations of the sound quality by target-signal cancellation were almost the same. All the γ's for low-pass filters were 0.995 both for MG-AMC and AMC-SE.

The behavior of the indices are shown in Fig. 4. Whereas MG-AMC has many mis-detections, AMC-SE achieves almost perfect detection. The output power of the RAMA was also compared in Fig. 5.

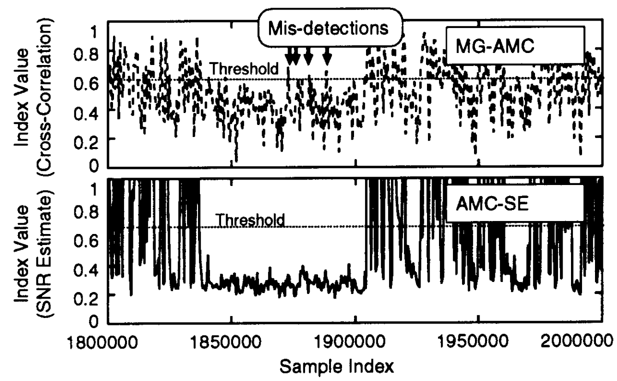

Figure 4: Behavior of the indices

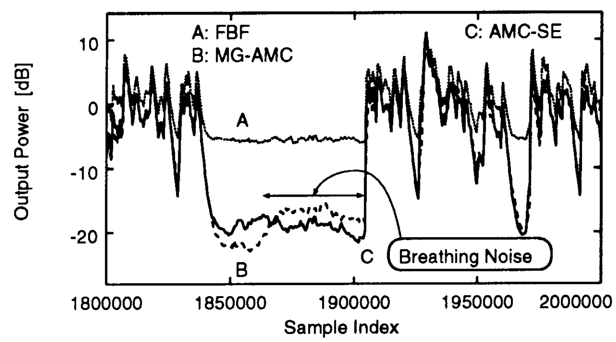

Figure 5: Output powers

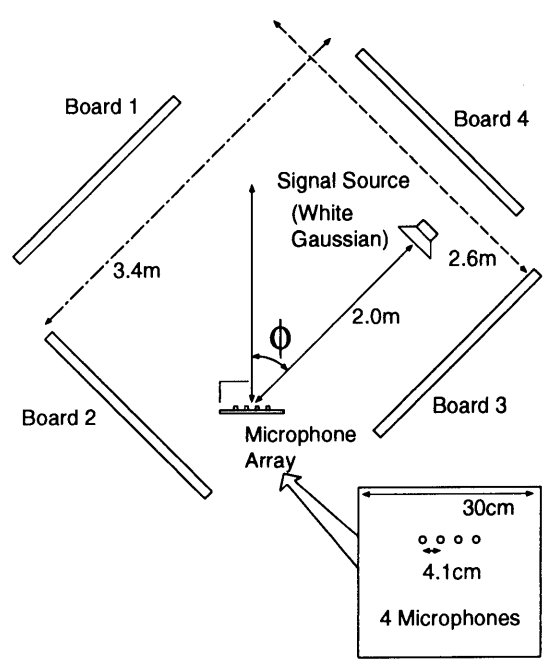

Figure 6: Experimental Set-up

AMC-SE avoids serious breathing noise which occurs with MG-AMC.

4.2 Evaluation of the Overall System

Noise-reduction performance of the overall system was evaluated on a real-time signal-processing system. An equi-spaced linear array with four omnidirectional microphones was used. Microphone spacing was 4.1cm, and the sampling frequency was 8 kHz. The experimental set-up is illustrated in Fig. 6. Reverberation of the room was controlled by changing the material of the boards in Fig. 6 as shown in Table 1. Environment B is more reverberant than environment A, and environment C is the most reverberant. Reverberation time of the room itself was 0.3 second. The white-Gaussian source was scanned in two ways. The constraints of the ABM was set so that allowable target-direction range was ±15 degrees. All the other parameters used were the same as in Section 4.1.

When noise source moved, breathing noise was observed, however, it disappeared in a few seconds by adaptation. For example, in environment B, when the noise-DOA was 45-degree off the target direction, the noise was reduced by 10 dB in about 3 seconds, and by 20 dB in about 20 seconds. The noise reduction after convergence are 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 is observed in the DOA over 30 degrees. Even in environment C, the worst noise reduction is more than 12 dB in the DOA over 40 degrees. These results indicate that the RAMA is promising for applications such as voice communications at least in typical environments like A and B.

Figures 7, 8, and 9 also indicate that noise reduction varies with reverberation. Between Figs. 7 and 8, reverberation by reflecting boards causes almost no degradation in noise reduction. However, between 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 [1].

As is clear form Figs. 7, 8, and 9, the noise-reduction performance of the RAMA has hysteresis for the source movement. In terms of spatial selectivity, the hysteresis should be reduced. Analysis of the hysteresis and its reduction will be left for future works.

Table 1: Materials of Boards in Fig. 6

Environment	Boards 1 & 3	Boards 2 & 4
A	Absb.	Absb.
B	Absb.	Refl.
C	Refl.	Refl.

Absb.: Absorbing, Refl.: Reflecting

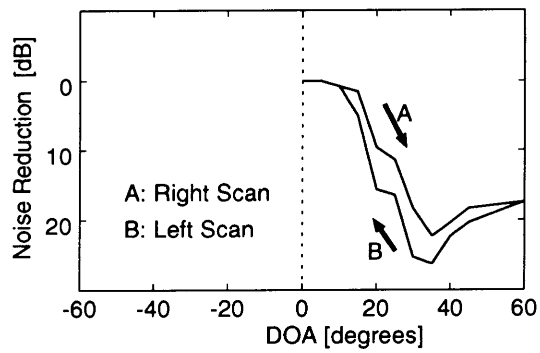

Figure 7: Noise Reduction in Environment A

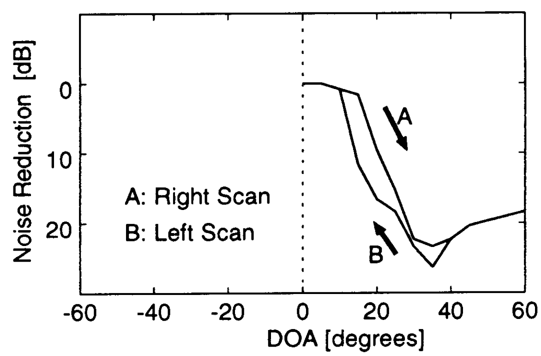

Figure 8: Noise Reduction in Environment B

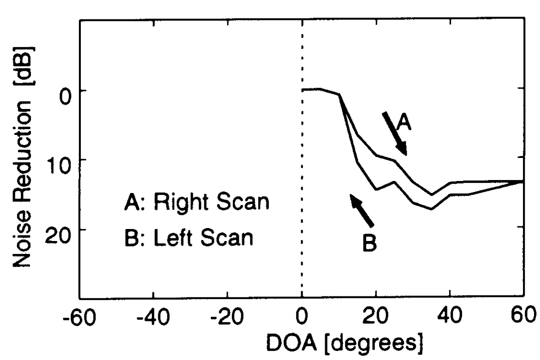

Figure 9: Noise Reduction in Environment C

5 Conclusion

A robust adaptive microphone array with a new adaptation-mode control and its evaluation by hardware have been presented. The adaptation mode is controlled based on an SNR estimate using output powers of the fixed beamformer and the adaptive blocking matrix. The robust adaptive microphone array has been implemented on a flexible multiple-DSP system with C-compiler. Simulation results demonstrated that the adaptation-mode control method based on the SNR estmate has less breathing noise than the conventional method. Evaluation with a real-time signal-processing system has shown that the implemented adaptive microphone array has over 12 dB noise-reduction performance even in reverberant environments.

Acknowledgement

The authors would like to thank Takafumi Kikuchi, Department of Electrical Engineering, Faculty of Science and Technology, Keio University for his assistance with the real-time 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 for their encouragement on this study.

References

[1] Y. Kaneda, "Adaptive Microphone Arrays," Trans. IEICE, Vol.J75-BII, pp.742–748, Nov. 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, Feb. 1991.

[3] Y. Kaneda, J. Ohga, "Adaptive Microphone Array System for Noise Reduction," IEEE, Trans. ASSP, Vol.34, pp.1391–1400, Jun. 1986.

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

[5] 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, Aug. 1994.

[6] S. Affes, Y. Grenier, "A Signal Subspace Tracking Algorithm for Microphone Arrays," IEEE, Trans. SAP, 1997. (to appear)

[7] J. Kates, "A Comparison of Hearing-Aid Array-Processing Techniques," J.A.S.A, Vol. 99, No. 5, pp. 3138–3148, May 1996.

[8] O. Hoshuyama, A. Sugiyama, "A Robust Generalized Sidelobe Canceller with a Blocking Matrix Using Leaky Adaptive Filters," Trans. IEICE, Vol.J79-A, No.9, pp.1516–1524, Sep. 1996. (in Japanese)

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

[10] O. Hoshuyama, A. Sugiyama, 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.

[11] G. Elko, A. Pong, "A Steerable and Variable First-Order Differential Microphone Array," IEEE, Proc. of ICASSP'97, Vol. I, pp.223–226, Apr. 1997.

[12] 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.

[13] H. Silverman, W. R. Patterson III, J. L. Flanagan, 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.

[14] J. E. Greenberg, 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.

[15] B. Begasse, O. Hoshuyama, A. Sugiyama, "A New Adaptation-Mode Control Based on Cross Correlation for a Robust Adaptive Microphone Array," IEICE, 12th DSP Symposium, A3-4, Nov. 1997. (to appear)

[16] "Real-Time Signal Processing System for LAN, S-RTPstation Series," Systems Design Service Corporation, Product Catalog, 1994. (in Japanese)

[17] *TMS320C4x User's Guide*, Texas Instruments, Aug. 1993.

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