14th European Signal Processing Conference (EUSIPCO 2006), Florence, Italy, September 4-8, 2006, copyright by EURASIP

A NEW ADAPTATION MODE CONTROLLER FOR ADAPTIVE MICROPHONE ARRAYS BASED ON NESTED AND SYMMETRIC LEAKY BLOCKING MATRICES

Thanh Phong HUA, Akihiko SUGIYAMA, Régine LE BOUQUIN JEANNES‡, Gérard FAUCON‡

Media and Information Research Laboratories, NEC Corporation, KAWASAKI 211-8666, JAPAN
‡LTSI, Inserm U642, Université de Rennes 1, Campus de Beaulieu, 35042 RENNES CEDEX, FRANCE

ABSTRACT

An adaptation mode controller (AMC) based on an estimation of signal-to-interference ratio using multiple blocking matrices for adaptive microphone arrays is proposed. A new nested blocking matrix enhances detection of the interference power. A normalized cross-correlation between symmetric leaky blocking-matrix outputs improves directivity. The detection of hissing sounds in the target speech is enhanced by modifying the high frequency components of the fixed beamformer output. Evaluations are carried out using a four-microphone array in a real environment with reverberations for different signal-to-interference ratios, interference directions of arrival, and target distances from the array. They show that the proposed AMC contributes to an enhanced output quality as well as an increased speech recognition rate by as much as 31% compared to the conventional AMC.

1. INTRODUCTION

In hands-free devices, a generalized sidelobe canceller (GSC) [1] based on adaptive beamforming is often used to capture the target signal coming from a specific direction of arrival (DOA) and attenuate all others. An advanced structure of GSC is the robust adaptive microphone array with an adaptive blocking matrix, RAMA-ABM [2]. It is composed of a fixed beamformer (FBF), and an adaptive path including an adaptive blocking matrix (ABM), and a multiple input canceller (MC). FBF is a spatial filter, which enhances the target. ABM adaptively blocks the target and passes the interference. The residual interference in the FBF output that is correlated with the ABM outputs is adaptively cancelled by MC. Adaptation of filter coefficients in ABM should be performed only during periods of target speech, and inversely in MC, to avoid target speech cancellation [3], [4]. An adaptation mode controller (AMC) conducts these alternate adaptations based on the predominance of the target speech over the interference.

An AMC for the multiple input canceller has been proposed by Greenberg and Zureck [3]. It is based on the cross-correlation of two adjacent microphone signals. The cross-correlation reflects the phase difference between the two microphone signals. Because the target speech is assumed to come from the perpendicular direction to the microphone array surface, the cross-correlation should be large during periods of target speech and small otherwise. The cross-correlation is compared with a threshold to control the adaptations. The problem of this AMC is the small bandwidth of speech detection, typically 0.6 - 1.2 kHz, to avoid aliasing caused by its large inter-microphone distance imposed by its original application, the hearing-aids. The average power of speech is not always dominant in this frequency range, which may cause failure of this AMC.

With larger frequency bands, the AMC based on a signal-to-interference ratio estimation, AMC-SE, gives better performance [4], [5]. The signal-to-interference ratio (SIR) estimate is the power ratio of the FBF output to a fixed blocking matrix output. The FBF is used as the power estimator for the target speech, whereas the fixed blocking matrix is the one for the interference. When the SIR estimate is larger than a threshold, the signal is considered as target speech and only the coefficients in ABM are adapted. Inversely, when it is below the threshold, only the coefficients in MC are adapted. However, both estimators have a non-flat frequency response and a gradual transition in the directivity response due to a limited number of microphones [6]. These non-ideal responses cause inaccurate SIR estimation depending on both the spectra of the target and the interference, and the interference DOA.

This paper proposes a new AMC based on multiple fixed blocking matrices. A new fixed blocking matrix with a nested structure serves as a more accurate power estimator for the interference for all DOAs. The symmetric leaky blocking matrices are introduced to help discriminate the target from the interference by means of a normalized cross-correlation.

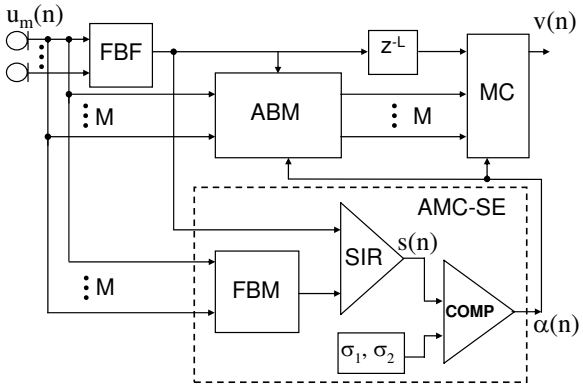

Figure 1: RAMA-ABM with the conventional AMC.

2. CONVENTIONAL AMC

The structure of AMC-SE in a robust adaptive microphone array, RAMA-ABM [2], is shown in Fig. 1 for M microphones. AMC-SE controls coefficient adaptation of the filters in ABM and in MC. It is composed of a fixed blocking matrix (FBM), an SIR calculator, and a comparator (COMP). An SIR estimate $s(n)$ at sample n is the output-power ratio of FBF to FBM. FBF is an estimator for the target because it forms a beam in the look direction of the microphone array, whereas FBM is the one for the interference because it forms a null in the look direction. The SIR estimate is compared to thresholds, σ_1 and σ_2, to obtain a control signal $\alpha(n)$, as

$$\alpha(n) = \begin{cases} 0, & s(n) \le \sigma_1, \\ \frac{s(n)-\sigma_1}{\sigma_2-\sigma_1}, & \sigma_1 < s(n) < \sigma_2, \\ 1, & s(n) \ge \sigma_2. \end{cases} \quad (1)$$

To achieve opposite adaptations, the step sizes of coefficient adaptation in ABM are multiplied by $\alpha(n)$, whereas those in MC are multiplied by $(1-\alpha(n))$. The weakness of AMC-SE resides in the dependency of the SIR estimate on both the spectra and the DOA of the sound sources. The frequency response is not flat and the DOA response has a gradual transition from the target to the interference directions.

[†]On leave from Université de Rennes 1.

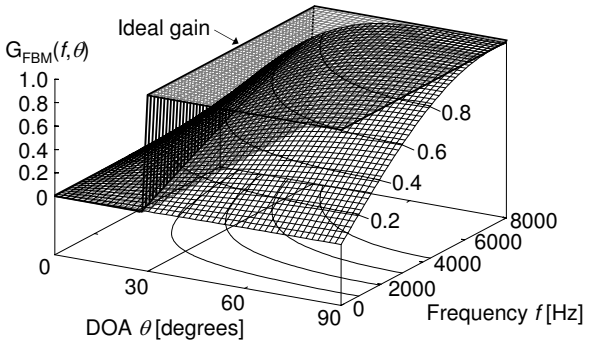

Figure 2: Comparison between the FBM gain and an ideal gain given a minimum interference DOA of 30°.

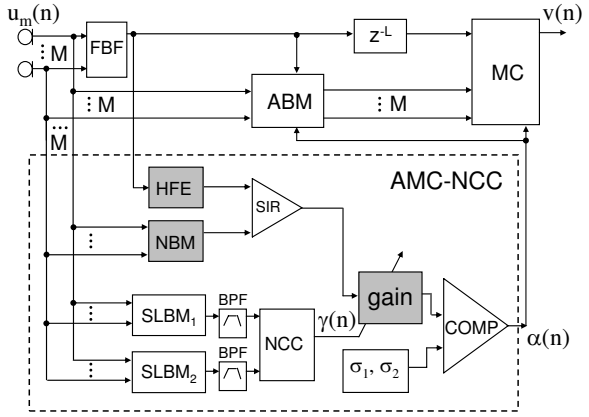

Figure 3: RAMA-ABM with AMC-NCC.

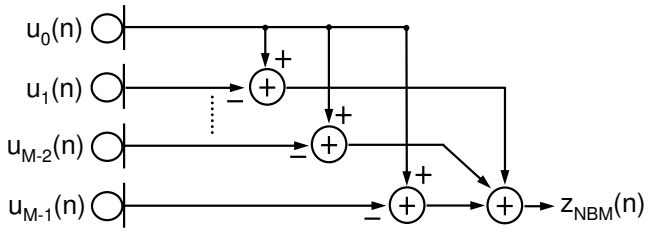

Figure 4: Structure of NBM.

The non-flat frequency response of the SIR estimation can be caused by the interference power estimation. This is shown by calculating the FBM gain. The FBM output $z_{FBM}(t)$ is defined as the difference between two adjacent microphone signals as follows:

$$z_{FBM}(t) = \frac{1}{2}\left[u_{m+1}(t) - u_m(t)\right], \tag{2}$$

where m represents the m-th microphone index with $m < M-1$. Each microphone signal is a delayed version of another assuming plane waves. This delay t_0 is known as $t_0 = \frac{D\sin\theta}{c}$, where D is the inter-microphone distance, c the sound speed and θ the source DOA. It simplifies (2) to

$$z_{FBM}(t) = \frac{1}{2}\left[u_m(t-t_0) - u_m(t)\right]. \tag{3}$$

The transfer function $H_{FBM}(j\omega, t_0)$ of FBM, with respect to the input signal $u_m(t)$, is therefore

$$H_{FBM}(j\omega, t_0) = \frac{1}{2}(e^{-j\omega t_0} - 1), \tag{4}$$

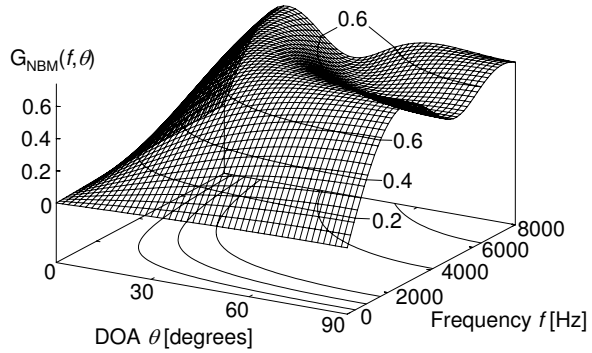

Figure 5: NBM gain using 4 microphones.

where $\omega = 2\pi f$, f the frequency, and $j^2 = -1$. The gain $G_{FBM}(f,\theta)$ of FBM is the norm of $H_{FBM}(j\omega, t_0)$ and is expressed as

$$G_{FBM}(f,\theta) = \frac{1}{2}\sqrt{2 - 2cos\left(2\pi f \frac{D\sin\theta}{c}\right)}. \tag{5}$$

The gain of FBM versus the interference DOA and frequency are shown in Fig. 2 with $D = 0.021\,m$ [1] and $c = 340\,m/s$. For any non-zero DOA, FBM is a high-pass filter. Therefore, the frequency response is not flat. The gradual transition from the target to the interference directions in the directivity response contrasts with the ideal gain in Fig. 2. The directivity response in microphone arrays relies on the inter-microphone spacing, *i.e.* the phase difference between microphone signals. This phase difference is small in low frequencies due to a small inter-microphone spacing to avoid aliasing of the highest frequency component. It makes the phase-based distinction between the target and the interference difficult. Therefore, the SIR estimation is likely to be inaccurate in low frequencies.

3. PROPOSED AMC

The proposed AMC contains a nested blocking matrix (NBM) for flat frequency response and an normalized cross-correlation (NCC) with symmetric leaky blocking matrices, $SLBM_1$ and $SLBM_2$, for sharp transition in the directivity response. NBM utilizes multiple lobes from different spacings. The symmetric structures of $SLBM_1$ and $SLBM_2$ increase the phase difference for the interference and guarantee equal gains, while the leakage ensures in-phase signals for the target. NCC is calculated between the outputs of $SLBM_1$ and $SLBM_2$ to measure their phase difference, which discriminates the sources. In a frequency range delimited by the bandpass filters (BPF), NCC controls the gain $g(\gamma)$, which improves the inaccurate SIR estimate, *i.e.* the output-power ratio of the high frequency enhancer (HFE) to NBM. HFE modifies the FBF output for a better detection of high frequency components in the target. The improved SIR is then compared by COMP with thresholds σ_1 and σ_2 to obtain the control signal $\alpha(n)$ using (1). The structure of the proposed AMC, AMC-NCC, is depicted in Fig. 3.

3.1 Nested blocking matrix

Increasing the distance D in (5) moves the lobe of the FBM gain to lower frequencies. To achieve flatter response, NBM takes advantage of multiple lobes by combining different spacings. The structure of NBM is shown in Fig. 4. The output $z_{NBM}(t)$ of NBM is obtained by combining the outputs of $(M-1)$ blocking matrices

[1] Half of the shortest wavelength to satisfy the spatial Nyquist criterion.

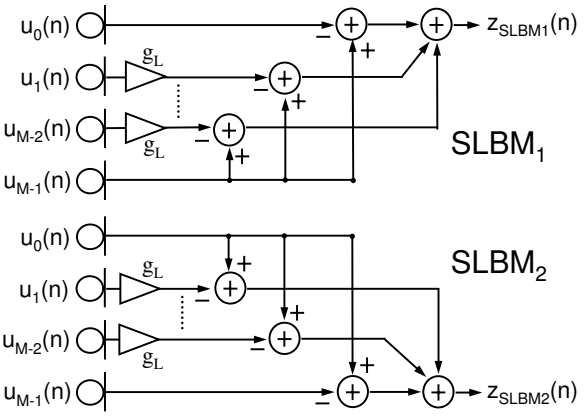

Figure 6: Symmetric structures of $SLBM_1$ and $SLBM_2$.

with different spacings as

$$z_{NBM}(t) = \frac{1}{2(M-1)} \{[u_0(t) - u_1(t)] + [u_0(t) - u_2(t)] + \ldots + [u_0(t) - u_{M-1}(t)]\}, \quad (6)$$

which is equivalent to

$$z_{NBM}(t) = \frac{1}{2(M-1)} \left[(M-1)\, u_0(t) - \sum_{m=1}^{M-1} u_m(t) \right]. \quad (7)$$

The transfer function of NBM, $H_{NBM}(j\omega, t_0)$, using the input signal $u_0(t)$, is calculated like in Section 2 as

$$H_{NBM}(j\omega, t_0) = \frac{1}{2(M-1)} \left[(M-1) - \sum_{m=1}^{M-1} e^{-j\omega t_0 m} \right]. \quad (8)$$

The gain $G_{NBM}(f, \theta)$ of NBM is the norm of $H_{NBM}(j\omega, t_0)$ and is expressed as

$$G_{NBM}(f, \theta) = \frac{1}{2(M-1)} \left[(M-1)^2 + (M-1) - 2 \sum_{m=1}^{M-1} m \cos\left(2\pi f m \frac{D \sin\theta}{c}\right) \right]^{1/2}. \quad (9)$$

Fig. 5 shows the gain of NBM for $M = 4$. Compared to the gain of FBM in Fig. 2, an additional lobe appears in the medium frequencies making the frequency response flat.

3.2 NCC with symmetric leaky blocking matrices

NCC is calculated between the outputs of the symmetric leaky blocking matrices $SLBM_1$ and $SLBM_2$ whose structures are shown in Fig. 6. They have the same gain, and ideally maximize their phase difference for the interference and minimize it for the target. Their outputs, respectively, $z_{SLBM1}(t)$ and $z_{SLBM2}(t)$, are defined by

$$z_{SLBM1}(t) = (M-1)\, u_{M-1}(t) - g_L \sum_{m=1}^{M-2} u_m(t) - u_0(t), \quad (10)$$

and

$$z_{SLBM2}(t) = (M-1)\, u_0(t) - g_L \sum_{m=1}^{M-2} u_m(t) - u_{M-1}(t), \quad (11)$$

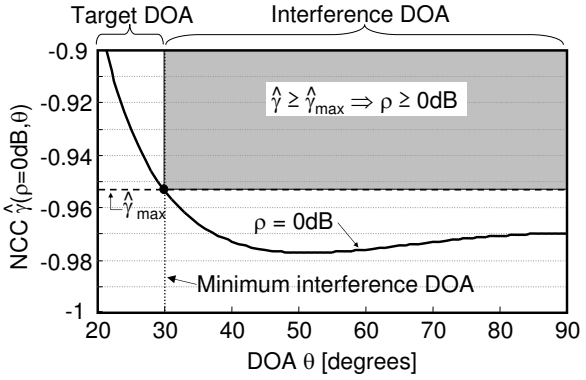

Figure 7: Directivity of NCC for uncorrelated white noises with $g_L = 0.92$, BPF passband = 500-1500 Hz, and $\rho = 1$.

where $|g_L| \neq 1$ is a parameter to control the phase difference and the target leakage. If $g_L = 1$, $SLBM_1$ and $SLBM_2$ become nested blocking matrices. The phase difference between their outputs is large in low frequencies and their gains are the same. However, the target, which is required for the discrimination, is blocked. This is the reason for introducing $|g_L| \neq 1$. It allows leakage of the in-phase target, whereas it still keeps a large phase difference for the interference.

$z_{SLBM1}(n)$ and $z_{SLBM2}(n)$ are filtered by BPF to obtain $v_1(n)$ and $v_2(n)$ respectively. The NCC $\gamma(n)$ at sample n is calculated between the signals $v_1(n)$ and $v_2(n)$ as follows:

$$\gamma(n) = \frac{\sum_{p=0}^{N-1} v_1(n-p) \cdot v_2(n-p)}{\sqrt{\sum_{p=0}^{N-1} v_1^2(n-p) \cdot \sum_{p=0}^{N-1} v_2^2(n-p)}}. \quad (12)$$

N is the number of past values to calculate $\gamma(n)$. For uncorrelated white noise signals as the target at 0° and the interference at an angle θ, the relation between the approximate NCC $\hat{\gamma}(\rho, \theta)$ of $\gamma(n)$ and the actual SIR ρ is shown in [8] to be

$$\hat{\gamma}(\rho, \theta) = \frac{\sum_{i=0}^{N-1} G^2(i,\theta) \cos[\varphi(i,\theta)] + \rho G^2(i,0)}{\sum_{i=0}^{N-1} G^2(i,\theta) + \rho G^2(i,0)}, \quad (13)$$

where $G(i, \theta)$ is the gain of the symmetric leaky blocking matrices at the normalized frequency i and the interference DOA θ, and $\varphi(i, \theta)$ the phase difference between $v_1(n)$ and $v_2(n)$. The dependency on the time index n has been dropped in the left-hand side of (13) for simplicity. Fixing θ, the derivative of $\hat{\gamma}(\rho, \theta)$ with respect to ρ is always positive. Thus, $\hat{\gamma}(\rho, \theta)$ is an increasing function of ρ. It is shown in [8] that $\hat{\gamma}(\rho, \theta)$ and ρ have a strong correlation. As a result, NCC $\gamma(n)$ can be taken as an estimate of the actual SIR ρ.

3.3 Gain controlled by NCC

A large $\gamma(n)$ indicates predominance of the target over the interference. In that case, the SIR estimate should be amplified by a large gain $g(\gamma)$ to have a better distinction between the target and the interference. Otherwise, the SIR estimate should be attenuated by a small $g(\gamma)$. Therefore, $g(\gamma)$ should be an increasing function of $\gamma(n)$. For simplicity, $g(\gamma)$ should linearly modify the SIR estimate in the dB scale with respect to $\gamma(n)$ as

$$g(\gamma) = \delta(\gamma - \hat{\gamma}_T), \quad (14)$$

where δ (> 0) and $\hat{\gamma}_T$ are constants. Two conditions have to be found to determine δ and $\hat{\gamma}_T$. The first one is that no distinction is needed ($g(\gamma) = 0$ dB) when neither the interference nor the target is predominant ($\rho = 0$ dB). This condition links $g(\gamma)$ with $\gamma(n)$ if there is a direct link between $\gamma(n)$ and ρ. Although (13) links $\gamma(n)$ with ρ,

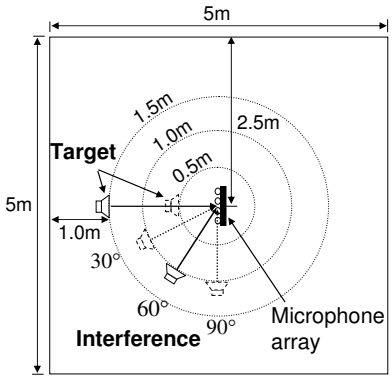

Figure 8: Experimental set-up.

Table 1: Parameter values for the evaluations

Parameter	Value	Parameter	Value
f_s	16 kHz	g_L	0.92
σ_1, σ_2	18 dB, 21 dB	δ	70
Passband of BPF	500 - 1500 Hz	$\hat{\gamma}_T$	-0.95
N	256	V_{HFE}	9

there is also a dependency on the interference DOA θ. $\hat{\gamma}$ versus θ is depicted in Fig. 7, fixing ρ to 0 dB, which is a part of the condition. Correction by $g(\gamma)$ is more necessary in small DOAs, where the SIR estimate lacks accuracy, than in large DOAs. Thus, the gain should be designed for the minimum interference DOA θ_{min}. As a result, the value $\hat{\gamma}(\rho = 0dB, \theta = \theta_{min})$ should be taken for $\hat{\gamma}_T$ to satisfy the condition that $g(\gamma) = 0$ dB for $\rho = 0$ dB. The second condition for δ determination is a trade-off between degrees of modification. It depends on how much NCC should influence the final SIR estimate.

3.4 High frequency enhancer

Speech usually has smaller power in high frequencies than in low frequencies except during hissing sounds, where these powers are comparable. To detect hissing sounds, the FBF signal is decomposed into a few frequency bins by a Fourier transform. If the maximum power is in the highest frequency bin, HFE multiplies the power of the FBF output by a constant V_{HFE} to improve the detection performance of hissing sounds.

4. EVALUATIONS

4.1 Environment

A uniform linear array of four microphones was placed in the middle of a reverberant room of 5 m in width and length, and 3 m in height, to acquire the data as shown in Fig. 8. The target speech, and the TV-commercial interference (comprising child, female and male voices, advertisements, music, stationary noise...), presented by loudspeakers, had been recorded separately before they were digitally mixed to obtain specific SIRs. The microphone gains were automatically calibrated with the method described in [7].

4.2 Parameters

5-th order elliptic filters for BPF are used to band-limit NCC in the range of 500-1500 Hz. The thresholds σ_1 and σ_2 are two SIR estimates obtained in a simulated environment using speech for both the interference and the target at an SIR of 0 dB. A sharp transition in the phase difference between the transfer functions of $SLBM_1$ and $SLBM_2$ can be achieved in low frequencies with $g_L = 0.92$ [8]. The theoretical value of $\hat{\gamma}(\rho = 0dB, \theta)$ at the minimum interference DOA $\theta_{min} = 30^\circ$ is taken for $\hat{\gamma}_T$, which corresponds to

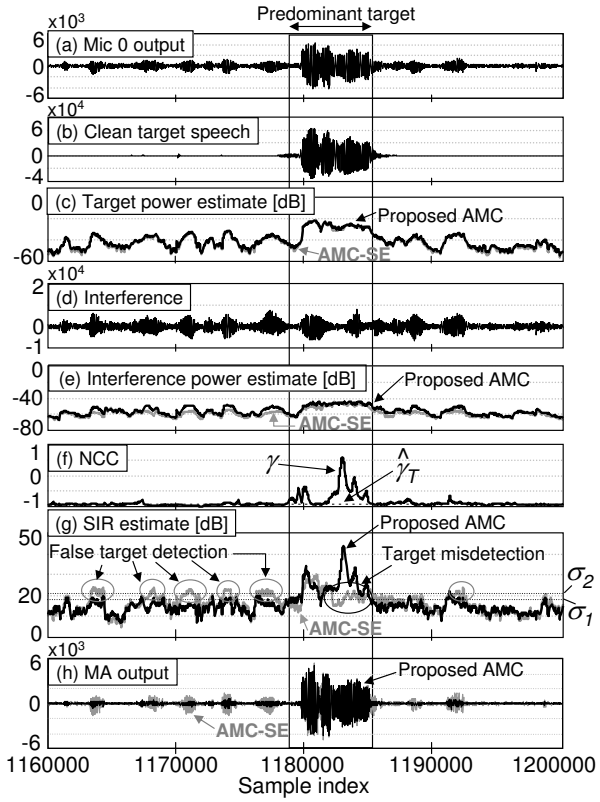

Figure 9: Signals for a target speech at 0.5 m away, an interference DOA of 30° and an average SIR of 0 dB.

$\hat{\gamma}_{max} \approx -0.95$. The parameter values are summarized in Table 1, where f_s denotes the sampling frequency.

4.3 Quality of the microphone array output

The improved output quality by the proposed AMC is shown in Figs. 9 and 10, where the interference was located 1.0 m away at 30° with an average SIR of 0 dB, and at 90° with an average SIR of 10 dB, respectively. The proposed AMC gives larger power estimate during bursts of interference as seen in Fig. 9 (e) around sample index 1178000. It implies a smaller SIR estimate in interference sections. The proposed AMC detects the target speech when it is predominant over interference, *i.e.* when $\hat{\gamma} > \hat{\gamma}_T$, as seen in Fig. 9 (f) around sample index 1183000. Thus, the SIR estimate is larger in target speech sections. The SIR estimate in Fig. 9 (g) is therefore improved compared to AMC-SE, which results in stronger interference suppression and less target cancellation in the microphone array output as in Fig. 9 (h). It should be noted that the computational complexity of the proposed AMC with RAMA-ABM is increased by 15% compared to the conventional system in case of 16 taps for the adaptive filters in ABM and MC.
The SIR estimate is also improved by better hissing sound detection. The target power estimate in Fig. 10 (c) is higher at sample indexes around 920000 which corresponds to a hissing sound in the target speech. The SIR estimate in Fig. 10 (g) is above the threshold of detection σ_1 in the proposed AMC. Consequently, the hissing sound is more preserved in the output signal of the microphone array with the proposed AMC as in Fig. 10 (h).

4.4 Speech recognition rates

The target and the interference were located at typical locations for human-robot communications in household environment [10]. The target was put at 0.5 m and 1.5 m away and the interference at 1.0 m in 3 different DOAs, 30°, 60°, and 90° as shown in Fig. 8. The tar-

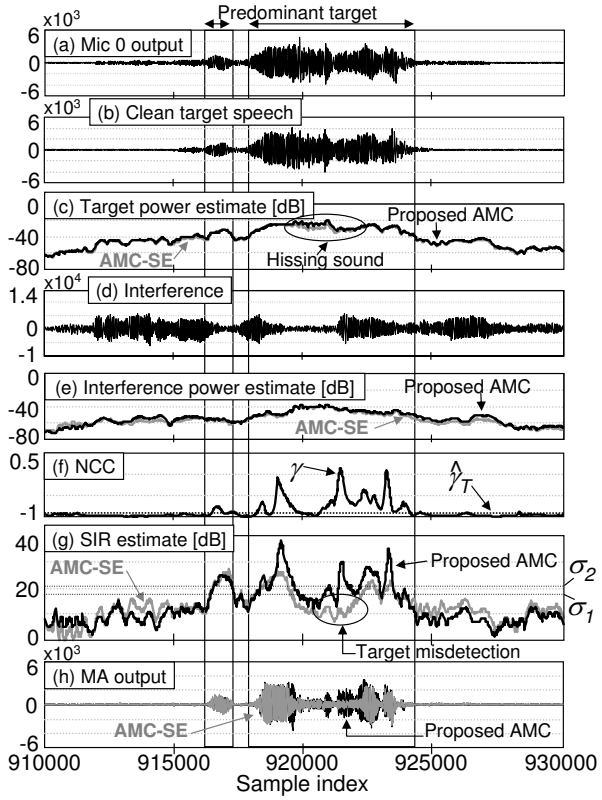

Figure 10: Signals for a target speech at 0.5 m away, an interference DOA of 90° and an average SIR of 10 dB.

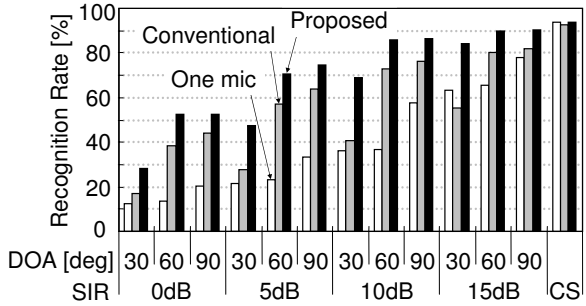

Figure 11: Speech recognition rates with the target at 0.5 m away.

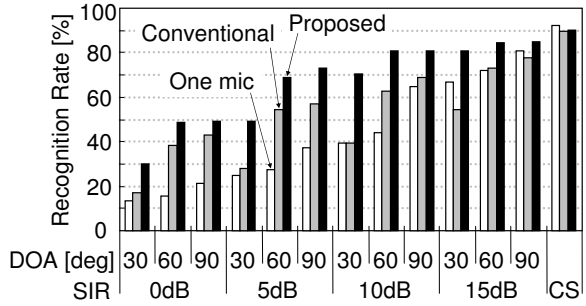

Figure 12: Speech recognition rates with the target at 1.5 m away.

get was composed of 30 Japanese speakers (10 males, 10 females and 10 children) with 50 utterances each. Signals with low to high average SIRs of 0, 5, 10, and 15 dB as well as clean speech (CS) signals formed 780 signals processed by RAMA-ABM with AMC-SE, and that with the proposed AMC. Their outputs are evaluated by a Japanese speech recognition system, *Julius* [9]. The results are shown in Figs. 11 and 12 for the target speech at 0.5 m and 1.5 m away, respectively. Compared to AMC-SE, the proposed AMC always achieves better recognition rates with an increase of up to 31% and an average increase of 14%. Compared to the one microphone case, it is increased by up to 49%, except for the clean speech sources at 1.5 m away with a 2% degradation. This degradation was caused by target cancellation at high SIRs due to reverberations [3], whose effects are amplified by the distance of the target source from the array. These recognition rates were obtained without adaptation in the speech recognition system to the microphone array. Thus, the proposed AMC requires no training of the recognition system to achieve good performance.

5. CONCLUSION

An AMC for adaptive microphone arrays has been proposed. It estimates the SIR to detect periods of target speech to control coefficient adaptation in the adaptive path of the beamformer. A nested blocking matrix based on a combination of microphone spacings ensures better interference power estimation. A gain controlled by the normalized cross-correlation between symmetric leaky blocking-matrix signals improves the SIR estimate in low frequencies for better directivity. An enhancer of high frequency-component powers serves as a hissing sound detector. Evaluations have shown that the estimated SIR by the proposed AMC is more accurate than the one by AMC-SE, which gives an enhanced output quality. The speech recognition rate with sound data recorded in an actual environment has been increased by as much as 31% compared to the conventional AMC.

REFERENCES

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

[2] O. Hoshuyama and A. Sugiyama, "Robust adaptive beamforming," *Microphone arrays*, chap. 5, Brandstein and Ward, ed. Springer, 2001.

[3] J. Greenberg and P. Zureck, "Evaluation of an adaptation beamforming method for hearing aids," *J.A.S.A*, vol.91, no.3, pp.1662–1676, Mar. 1992.

[4] O. Hoshuyama, B. Begasse, A. Sugiyama, A. Hirano "A real-time robust adaptive microphone array controlled by an SNR estimate," *ICASSP98*, pp. 3605–3608, 1998.

[5] O. Hoshuyama, A. Sugiyama, "An adaptive microphone array with good sound quality using auxiliary fixed beamformers and its single-chip implementation," *Proc. of IEEE ICASSP99*, pp. 949–952, Mar. 1999.

[6] D. Ward, R. Kennedy, R. Williamson, "Constant directivity beamforming," *Microphone arrays*, chap. 1, Brandstein and Ward, ed. Springer, 2001.

[7] T. P. Hua, A. Sugiyama, G. Faucon, "A new self-calibration technique for adaptive microphone arrays," *IWAENC 2005*, pp. 237–240, Sep. 2005.

[8] T. P. Hua, A. Sugiyama, R. Le Bouquin Jeannes, G. Faucon, "Estimation of the signal-to-interference ratio based on normalized cross-correlation with symmetric leaky blocking matrices in adaptive microphone arrays," Tech. Rep. of IEICE, SIP2006-11, pp. 61–66, Apr. 2006.

[9] A. Lee, T. Kawahara and K. Shikano, "Julius – An open source real-time large vocabulary recognition engine," *Proc. EUROSPEECH*, pp. 1691–1694, Sep. 2001.

[10] Y. Fujita, "Personal Robot PaPeRo," J. of Robotics and Mechatronics, vol.14, No.1, Jan. 2002.