고품질 음성 취득을 위한 Small-Aperture 적응 마이크로폰 어레이 시스템

이준호* · 박영철** · 윤대희***

Small-Aperture Adaptive Microphone Array System for High Quality Speech Acquisition

Junho Lee* · Young-Cheol Park** · Dae-Hee Youn***

요 약

본 논문에서는 개인용 컴퓨터를 기반으로 한 실시간 마이크로폰 어레이 시스템을 제안한다. 제안한 시스템은 Generalized Sidelobe Canceller (GSC) 를 기반으로 하며, 새로운 적응 모드 제어기 (adaptation mode controller) 를 이용함을 특징으로 한다. 제안한 마이크로폰 어레이 시스템의 성능 평가를 위해, 일반 가정의 거실을 모델링한 실험 환경을 구축하여 실험을 수행하였다. 실험을 통해 제안된 시스템이 안정적으로 GSC의 적응 모드를 제어하고, 목적 신호의 왜곡을 최소화하면서도 간섭 신호 제거 성능이 뛰어남을 확인하였다.

ABSTRACT

In this paper, a PC-based real-time microphone array system with small aperture is presented. The microphone array system is based on the generalized sidelobe canceler (GSC) but it employs a new adaptation mode controller (AMC). The performance of the proposed system was evaluated in the Multimedia Room modeled on an office situation. Evaluation experiments show that the proposed system can perform with stable noise suppression.

키워드

microphone array, beamforming, generalized sidelobe canceler, adaptation mode controller.

Ⅰ. Introduction

With increasing maturity in speech and speaker processing technologies, and the prevalence of telecommunications, there is a need for effective speech acquisition devices. Microphone arrays have a distinct advantage as they enable hands-free acquisition of speech with little constraint on the user [1].

Incorporating a microphone array into a hand-held device presents a set of challenges. Size, power, and cost requirements dictate that a limited number of microphone elements be used and the total length of the microphone array should be small. In addition, the mobility of hand-held devices causes high variability in acoustic conditions. Because of these problems, we chose the generalized sidelobe canceller (GSC) as an adaptive beamforming algorithm. The GSC is widely used for a microphone array algorithm because of its computational efficiency and ability to suppress interference [1][2].

Two different approaches leading to two widely

* 연세대학교 전자공학과 박사과정
** 연세대학교 정보기술학부 교수 (young00@yonsei.ac.kr)
*** 연세대학교 전기전자공학과 교수

used algorithmic families have been previously adopted for the GSC. The first one is from the least mean square (LMS) family, and the latter is from the recursive least squares (RLS) family. Since adaptive beamforming has to be performed in real time, a low cost LMS family is more attractive. However, even using the LMS schemes, it is hard to achieve good results on a hand-held device due to heavy complexity of requirements.

Also, the conventional GSC suffers from target-signal cancellation due to target-signal leakage to the multiple input canceller (MIC) [1][2]. To reduce target-signal cancellation, the coefficient adaptation in the MIC should be carried out only when there is no target signal. Greenberg et al. proposed an adaptation mode controller (AMC) for an adaptive beamformer for hearing aids [3]. In their method, cross-correlations between input sensors determine target signal existence. However, this is applicable only to uncorrelated background noise environments. Later, a method that is able to distinguish a target signal from an interference signal was suggested [4], based on the power ratio between the outputs of a fixed beamformer (FBF) and a block matrix (BM). Unfortunately, this method has the shortcoming that it is difficult to set a threshold for both high and low signal-to-interference ratio (SINR) environments.

Research into a small-aperture microphone array system using the GSC algorithm is thus required to address these drawbacks, enabling the advantages of microphone arrays to be exploited in practical systems. In this paper, we focus on the realization of the adaptive microphone array system with a small number of sensors and size, which is suitable for a hand-held device. The paper is organized as follows: In Section II we cover the implementation issues related to the small-aperture microphone array on a hand-held device. Section III outlines a proposed microphone array system. Finally, in Section IV, the system is evaluated under various acoustic conditions.

II. IMPLEMENTATION ISSUES RELATED TO THE SMALL-APERTURE MICROPHONE ARRAY ON A HAND-HELD DEVICE

This section covers issues regarding implementation in relation to the small-aperture microphone array.

2.1 Generalized sidelobe canceler

The GSC comprises a FBF, a MIC, and a BM, as depicted in Fig. 1 [1]. The FBF enhances the target signal by summing the steered signal in the desired look direction. On the other hand, the BM forms a null in the look direction so that the target signal is suppressed and all other signals are passed through.

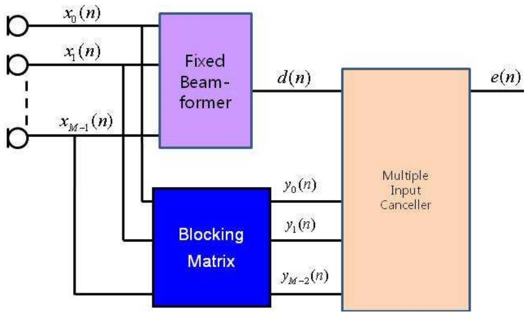


Fig. 1 Block diagram of the GSC

$$d(n) = \sum_{m=0}^{M-1} x_m(n - r_m), \tag{1}$$

$$y_m(n) = x_m(n - r_m) - x_{m+1}(n - r_{m+1}), \tag{2}$$

$$e(n) = d(n) - \sum_{m=0}^{M-2} \boldsymbol{w}_m^T(n) \boldsymbol{y}_m(n), \tag{3}$$

$$\boldsymbol{w}_m(n+1) = \boldsymbol{w}_m(n) + \frac{\mu e(n)}{\| \boldsymbol{y}_m(n) \|_2} \boldsymbol{y}_m(n), \tag{4}$$

where $\boldsymbol{w}_m(n)$ and $\boldsymbol{y}_m(n)$ denote $1 \times L$ weight vector, $1 \times L$ output signal vector of the BM, respectively, and T denotes the transpose operator. The numerical complexity of the LMS GSC algorithm is $2(M-1) \times L$ multiplication and $2(M-1) \times L$ addition [5]. When M=4, L=256, and the sampling frequency is 44.1 KHz, the computational complexity is about 67 million instructions per second (MIPS) not including AMC. To realize this on a small hand-held device with a small number of sensors, it is necessary to have low-cost implementation for the GSC.

2.2 Adaptation mode controller

If target signal leakage to the noise reference signal occurs, the adaptive filter in the MIC cancels the correlated part of the target signal [1][2]. In the GSC, it is implicitly assumed that adaptive filters in the MIC are adapted only when the target signal is inactive. In a real environment, however, we do not know in advance if there is a target signal. To solve the problem, an AMC based on the FBF output signal, $d(n)$, and the output signal of the BM, $y_m(n)$, was suggested [4]. The AMC, controlled by the signal-to-noise ratio (SNR) estimate, is represented as

$$p_d(n) = (1-\alpha)p_d(n-1) + \alpha d^2(n), \tag{5}$$

$$p_y(n) = (1-\alpha)p_y(n-1) + \alpha y_m^2(n), \tag{6}$$

$$\begin{cases} MIC: adaptation & p_d(n)/p_y(n) > T_p \\ MIC: freeze & otherwise \end{cases}, \tag{7}$$

where α denotes a forgetting factor and T_p denotes the pre-determined threshold. When only the target signal exists, the ratio has a large value, while the ratio becomes a small value when only the interference signal exists. The power ratio method determines the target signal existence by comparing the power ratio with the threshold. This method is easy to implement and shows reasonable performance in a well-defined simulation environment, but its threshold is greatly dependent on the SINR of the environments. Because of the variability in acoustic conditions where a hand-held device is used, a more robust AMC is also required.

III. PROPOSED MICROPHONE ARRAY SYSTEM

In this section, we present a microphone array system employing AMC that offers a better solution to the issue of discriminating between active/inactive periods of the target during implementation.

3.1 Frequency domain realization of LMS GSC

Fast transforms such as Discrete Fourier Transform (DFT) are employed in this study in order to obtain low computational complexity. The block diagram of the proposed microphone array system is shown in Fig. 2. Let $X_m(k)$ denote the DFT of the received signal at the m-th microphone; the time-aligned signal $X_{m,ta}(k)$ can then be written as

$$d(m) = \frac{d sin\theta}{C}(m-1) \times Fs, \tag{8}$$

$$X_{m,ta}(k) = X_m(k) e^{-j2\pi k d(m)/K}, \tag{9}$$

where d, C, θ, and Fs denote sensor spacing, wave speed, look direction, and sampling frequency, respectively. The transform domain GSC is described by the following set of equations:

$$Y_{FB}(k) = \frac{1}{M}\sum_{m=0}^{M-1} X_{m,ta}(k), \tag{10}$$

$$Y_{m,B}(k) = X_{m,ta}(k) - X_{m+1,ta}(k), \tag{11}$$

$$e(k,n) = Y_{FB}(k) - \sum_{m=0}^{M-2} W_m^H(k,n)\, Y_{m,B}(k), \tag{12}$$

$$W_m(k,n+1) = W_m(k,n) + \mu \frac{e^*(k,n)}{Y_{m,B}^2(k)} Y_{m,B}(k), \tag{13}$$

where H denotes a Hermitian (conjugate and transpose) operator. When M is 4, block size is 256, and the sampling frequency is 44.1 KHz, the complexity is about 8 MIPS not including the AMC.

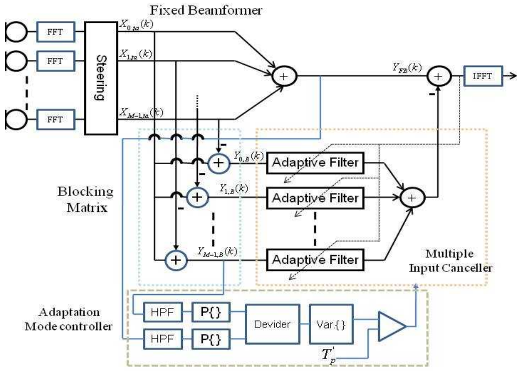


Fig. 2 Block diagram of the proposed system

3.2 Proposed adaptation mode controller

Fig. 2 shows the block diagram of the proposed AMC. Before calculating the power ratio of the FBF output and the BM output, the signals were filtered by high-pass filter (HPF), since the low-frequency component of the FBF output signal cannot be blocked completely. Although this strategy gives a more reliable result when determining the existence of the target signal, its threshold is greatly dependent on the SINR of the environments. To reduce the dependency, we don't use the power ratio; instead we use its variance.

$$P_r = \sum_{k=f_L}^{K/2} |Y_{FBF}(k)|^2 / \sum_{k=f_L}^{K/2} |Y_{m,BM}(k)|^2, \qquad (14)$$

$$\begin{cases} MIC: adaptation\ Var(P_r) > T_p^{'} \\ MIC: freeze \qquad otherwise \end{cases}, \qquad (15)$$

where f_L is the frequency index for the cut-off frequency of the HPF.

Additionally, to prevent misdetection in the voiced part of the target signal, the adaptation is performed in the past frame, whereas the filtering is performed on the current frame. This is an effective strategy for adding robustness to the process of discriminating between active/inactive periods of the target signal. To achieve it, we replace $X_{m,ta}(k)$ of the (10)-(13) with $X_{m,ta}^{'}(k)$, which is delayed $X_{m,ta}(k)$ with P.

3.3 PC.Realization on a PC platform

3.3.1 Platform description

A multichannel recording system as shown in Fig. 3-(a) was used for microphone array recordings. The acquisition system consisted of four linearly spaced omni-directional microphones with inter-element spacing of 2cm and an 8-channel A/D board. The total aperture size of the microphone array is only 6cm, so this makes the microphone array system applicable to a hand-held device.

3.3.2 Graphic user interface

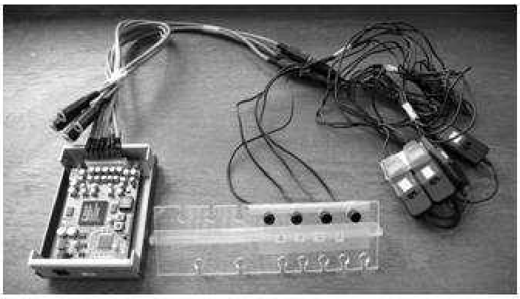
The proposed system was programmed with Microsoft Visual C++ 6.0. Acoustic data was sampled at 16 KHz and processing was performed on a laptop in real-time on Windows XP. The experiments were conducted on an Intel Pentium 4 1.5GHz processor. The acquisition module was implemented as a dynamic link library (DLL) using a low-level ASIO application programming interface (API). To synchronize between the acquisition module and the enhancement module using GSC, the method of thread programming was used. Fig. 3-(b) illustrates the graphic user interface (GUI) of the implemented system. Through a GUI, the parameters of the GSC algorithm can be controlled with various options.

Ⅳ. EXPERIMENTAL EVALUATION

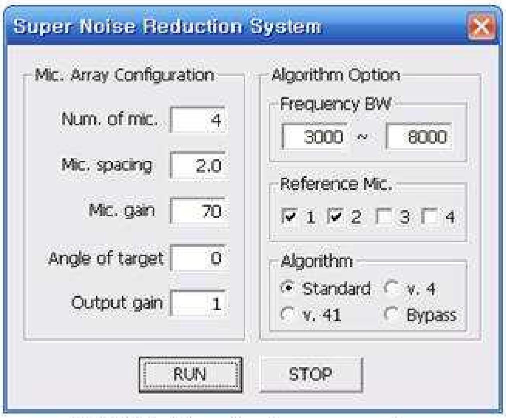
4.1 A. Experimental environment

The performance of the system was evaluated in the *Multimedia Room* of the *Center for Signal Processing Research* in *Yonsei Engineering Research Complex*, which has been purpose-built for various acoustic experiments. The dimension of the room is about $4.0m \times 6.5m \times 2.5m$ with reverberation time of approximately 0.3s. Fig. 4 is a picture of the experiment setups. We considered single competing speaker existence situation, so two loudspeakers were used to playback the target at (0.5m, 90°) with the interference signal at (1m, 30°) in a polar coordinate system. The target signal source comprised four sentences from the Texas Instruments and Massachusetts Institute of Technology (TIMIT) database and the interference signal was speech from the TIMIT database and various music signals. The SINR was varied from −10 to 10 dB.

Three objective quality measures were used to assess the system′s performance. The first criterion was the miss and false alarm rates of the proposed AMC. Detection rates were defined for frames (256 samples) relative to true target activity detection. The second criterion was the directivity pattern of the implemented system. Thirdly, we used the noise level (NL) during nonactive target periods for evaluation of noise reduction.



(a) A experimental apparatus.



(b) A GUI of the microphone array system.

Fig. 3 Realization on a PC platform

4.2 Experimental Results

Table 1 shows the miss and false alarm rates for various SINR conditions. False alarms cause the adaptive filters of the MIC to converge slower, but this effect is not critical in terms of performance. From this point of view, in our study the AMC outperformed the conventional power ratio method irrespective of SINR levels.

Beampatterns of the proposed system at different frequencies of interest are depicted in Fig. 5. From the figure, it can be observed that the system passed the target signal without distortion, while it formed a null in the direction of the interference.

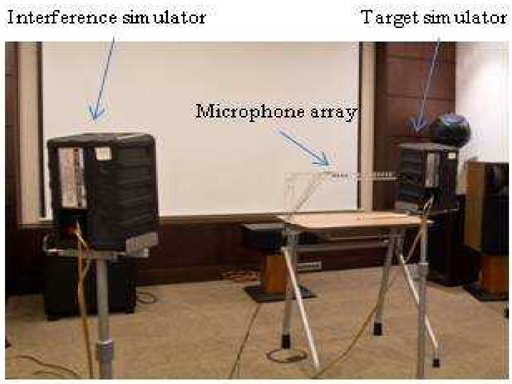


Fig. 4 The experimental setup

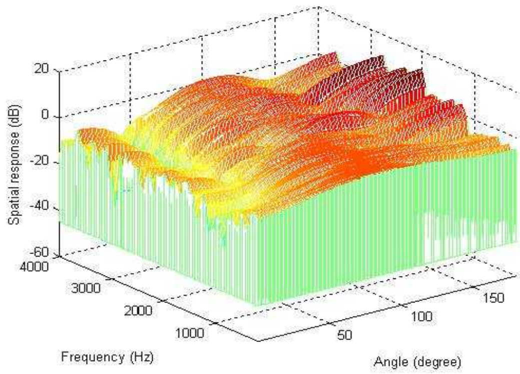


Fig. 5 Beampatterns of the proposed system

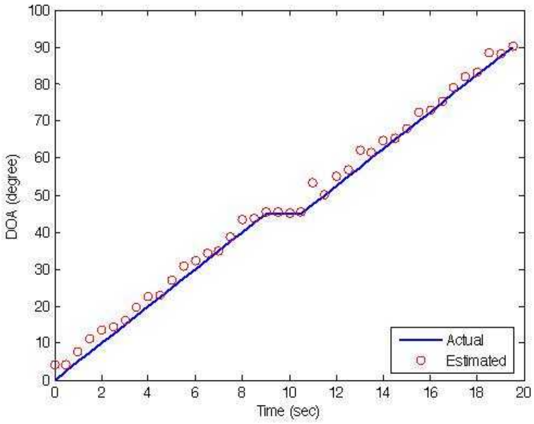


Fig. 6 The actual DOA and the estimated DOA

Through use of beampatterns, the system as implemented performed reasonably well for nonstationary interference, in spite of a small aperture size. Note that the degradation of array directivity at low frequencies is caused by the small aperture size.

To assess tracking capability an experiment to test moving interference was performed. In this experiment the speaker started moving from the 0° position to the 90° position; the speaker also stopped once at 45° for 2 seconds during the experiment.

Table. 1 Detection results

SINR	Conventional		Proposed	
	% miss	% false	% miss	% false
-10 dB	5.8 %	17.8 %	3.9 %	15.1 %
0 dB	3.1 %	21.5 %	1.2 %	15.3 %
10 dB	2.2 %	13.2 %	1.1 %	14.9 %

Table. 2 Noise reduction results

SINR	Conventional	Proposed
-10 dB	12.0 dB	11.2 dB
0 dB	14.2 dB	15.1 dB
10 dB	17.1 dB	18.3 dB

Fig. 6 illustrates the actual direction of arrival (DOA) for the interference, and the null position calculated by the proposed MA system, as a function of time. From the figure, it can be observed that the proposed system adaptively formed a null in the direction of the interference. Note that the average absolute error for this example is about 3.1°. Table 2 gives the NL in terms of SINR; results were obtained by averaging 20 independent trials. The improvement achieved by the system was 15 dB on average. Consequently, the proposed system efficiently reduced the interference signal while preserving the target signal component. Note that in the Multimedia Room environment with a 4-element microphone array, the system was running with 18% CPU load.

4.2 Discussion

For a sound source $S(f)$ at a location p, the signal captured by each microphone can be represented as:

$$X_m(f,p_m) = \frac{e^{-j2\pi\lambda^{-1}\| p-p_m \|}}{\| p-p_m \|} U_m(f,p)S(f), \quad (16)$$

where the first term in the right-hand side represents the propagation delay and the decay. In a hand-held device scenario, the signal decay can be ignored. The term $U_m(f,p)$ in (16) is a directivity pattern, which is a complex function, providing the spatio-temporal transfer function of the channel. An ideal omni-directional microphone, $U_m(f,p)$ is constant for all frequencies and source locations. However, a commercial microphone has a limited frequency response, particularly at low frequencies, so that the microphone input signal is correspondingly poor. We can therefore say that, although the microphone array system has wider beamwidth due to its small aperture size, the proposed MA system has a reasonable performance.

Ⅴ. CONCLUSION

We implemented a PC-based real-time microphone array system with a small aperture using the newly developed AMC module; we then assessed the system in terms of various criteria. In a real environment, the system shows stable interference rejection capability for non-stationary interference.

References

[1] M. Brandstein and D. Ward, Microphone Arrays, Springer, 2001.

[2] O. Hoshuyama, A. Sugiyama, and A. Hirano, "A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters," IEEE Trans. on Signal Processing, vol. 47, pp. 2677-2684, Oct. 1999.

[3] J. E. Greenberg and P. M. Zurek, "Evaluation of an adaptive beamforming method for hearing aids," Journal of Acoustic Society of America, vol. 91, pp. 1662-1676, Mar. 1992.

[4] O. Hoshuyama, B. Begasse, A. Sugiyama, and A. Hirano, "A real-time robust adaptive microphone array controlled by an SNR estimate," in IEEE Inf. Conf. Acoust., Speech, Signal Processing, Seattle, WA, pp. 3605-3608, May 1998.

[5] G. Glentis, "Implementation of Adaptive Generalized Sidelobe Cancellers Using Efficient Complex Valued Arithmetic," Int. J. Appl. Math. Comp. Sci.,vol. 13, no. 4, pp. 549-566, 2003.

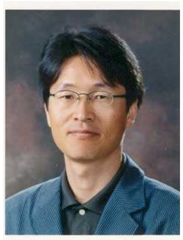
저자약력

이준호 (Junho Lee)

2001년 숭실대학교 정보통신전자공학과 학사
2003년 연세대학교 전자공학과 석사
2008년-현재 연세대학교 전자공학과 박사과정

<관심분야> 디지털 신호처리, 오디오 신호처리, 음성 신호처리, 적응 신호처리

박영철 (Young-Cheol Park)



1986년 연세대학교 전자공학과 학사
1988년 연세대학교 전자공학과 석사
1993년 연세대학교 전자공학과 박사
2008년-현재 연세대학교 정보기술학부 교수

<관심분야> 디지털 신호처리, 오디오 신호처리, 음성 신호처리, 적응 신호처리

윤대희 (Dae-Hee Youn)

1977년 연세대학교 전자공학과 학사
1979년 Kansas State Univ. 석사
1982년 Kansas State Univ. 박사
2008년-현재 연세대학교 전기전자공학과 교수

<관심분야> 디지털 신호처리, 오디오 신호처리, 음성 신호처리, 적응 신호처리