한국음향학회지 제33권 제5호 pp. 326~334 (2014)
The Journal of the Acoustical Society of Korea Vol.33, No.5 (2014)
http://dx.doi.org/10.7776/ASK.2014.33.5.326

pISSN : 1225-4428
eISSN : 2287-3775

A Novel Covariance Matrix Estimation Method for MVDR Beamforming In Audio-Visual Communication Systems

오디오-비디오 통신 시스템에서 MVDR 빔 형성 기법을 위한 새로운 공분산 행렬 예측 방법

Gyeong-Kuk You, Jae-Mo Yang, Jinkyu Lee, and Hong-Goo Kang†

(유경국, 양재모, 이진규, 강홍구†)

Department of Electrical and Electronic Engineering, Yonsei University

(Received February 18, 2014; revised May 15, 2014; accepted July 16, 2014)

ABSTRACT: This paper proposes a novel covariance matrix estimation scheme for minimum variance distortionless response (MVDR) beamforming. By accurately tracking direction-of-sound source arrival (DoA) information using audio-visual sensors, the covariance matrix is efficiently estimated by adopting a variable forgetting factor. The variable forgetting factor is determined by considering signal-to-interference ratio (SIR). Experimental results verify that the performance of the proposed method is superior to that of the conventional one in terms of interference/noise reduction and speech distortion.

Keywords: Audio-visual sensors, MVDR, Beamforming, Covariance matrix, Forgetting factor
PACS numbers: 43.60.Fg

초 록: 본 논문은 MVDR 빔 형성 기법을 위한 새로운 공분산 행렬 예측을 제안한다. 오디오-비디오 센서를 이용하여 음원의 방향 정보를 정확히 추적함으로써, 공분산 행렬은 가변 적응 망각율을 적용하여 효과적으로 예측된다. 가변 적응 망각율은 신호 대 방해 신호 비를 고려하여 결정된다. 실험 결과에서는 제안하는 방법의 성능이 방해신호/잡음 감소 및 음성 왜곡의 면에서 기존의 방법의 성능보다 더 우수하다는 것을 보여준다.

핵심용어: 오디오-비디오 센서, MVDR, 빔 형성 기법, 공분산 행렬, 적응 망각율

I. Introduction

Microphone array (MA) based speech enhancement schemes are popularly used for various applications such as speech recognition and teleconferences[1]. For example, the linearly constrained minimum variance (LCMV) beamforming has a criterion of minimizing total signal power while preserving the target signal and eliminating interference signal[2]. Due to an additional constraint to the direction of interference signal, however, its noise reduction performance is poor. On the contrary, the MVDR beamforming only has a criterion of preserving a filter gain to the direction of target signal while minimizing the total signal power. Although its interference reduction performance is poor, it is simple to control. This paper focuses on the method to enhance the performance of the MVDR beamforming.

To successfully accomplish the MA processing with MVDR beamforming, the DoA of target speech should be provided accurately and a covariance matrix must include only interference and noise components. Provided that precise DoA is given, it is helpful for estimating the covariance matrix. Furthermore, an additional correction process for the beamforming is not needed[3]. However, it is very difficult to accurately estimate DoA information in reverberant environment using audio signal only because of its difficulty in phase estimation. In a situation where it

†**Corresponding author:** Hong-Goo Kang (hgkang@yonsei.ac.kr)
Department of Electrical and Electronic Engineering, Yonsei University, 134 Shinchon-Dong, Seodaemoon-Gu, Seoul 120-749, Republic of Korea
(Tel: 82-2-2123-2766, Fax: 82-2-364-4870)

is very hard to separate interference components from the received signal, the received speech is directly used for calculating the covariance matrix, which results in performance degradation. In order to solve the problem, an audio-visual (A/V) sensor based approach has been recently proposed which has excellent capability of DoA estimation capability[4]. Since the system still needs to estimate the covariance matrix with the acquired signals, thus the improvement of beamforming performance was minimal. If the interference component is speech and the accurate DoA information is given, the covariance matrix for MVDR beamforming can be constructed efficiently.

This paper proposes a novel covariance matrix estimation method for MVDR beamforming. The main application area of the proposed scheme is a multi-user A/V communication system. By accurately estimating DoA using an A/V sensor based approach, a reliable steering vector for the covariance matrix is created and the steering vector is a phase component of covariance matrix. Then, the magnitude component of covariance matrix in the current frame is recursively updated by taking a first-order recursive averaging with the previous obtaining covariance matrix and the acquired speech signals. The forgetting factor needed for the averaging process is adjusted depending on the estimated SIR. Unlike the LCMV beamforming approach, the proposed MVDR method handles the interference constraint within the minimization criterion of estimating the precise covariance matrix.

Experimental results show that the proposed beamforming successfully reduces both interference and noise components of which performance is much better than the conventional beamforming. Also, the proposed method shows better performance than conventional methods in reverberant environment.

This paper is organized as follows. Section 2 formulates the problem and explains the MVDR beamforming method. The proposed method for efficient beamforming is described in Section 3. Section 4 shows the experimental results as figures and tables and conclusion is described in Section 5.

II. Minimum Variance Distortionless Response (MVDR)

In a noisy acoustic environment where people have A/V communication, it is necessary to acquire target speech signal only using a beamforming. In this paper, it assumes that there are two speakers in a room. Given a target speech $t[n]$, an undesired interference speech $u[n]$, and a diffused noise $n[n]$ are captured by M microphone array sensors, the received signals can be represented by

$$\begin{aligned} y_m[n] &= h_{t,m}[n]*t[n] = h_{i,m}[n]*u[n]+n[n] \\ &= x_m[n] = i_m[n]+n[n], \end{aligned} \tag{1}$$

where $h_{t,m}[n]$ and $h_{i,m}[n]$ are transfer function of target and interference speech to each microphone, respectively. $x_m[n]$ and $i_m[n]$ are captured target and interference speech and, m is microphone index. It assumes that the speeches and diffused noise are uncorrelated. In the frequency domain, the received signals can be rewritten as

$$Y(\omega) = X(\omega) + I(\omega) + N(\omega), \tag{2}$$

and $Y(\omega)$, $X(\omega)$, $I(\omega)$ and $N(\omega)$ are the M-length vector which consists of acquired component in each microphone, respectively.

For simplicity, frequency index ω is omitted from now on. Using the far-field assumption, the signal acquired by each sensor is represented by the phase shifted version of the first sensor signal as follows:

$$\begin{aligned} X &= X_1[l \quad e^{-j\omega\tau_{i,1}} \quad e^{-1\omega\tau_{i,2}} \cdots e^{-j\omega\tau_{i,M-1}}]^T, \\ &= X_1 d(\theta_1, \omega), \end{aligned} \tag{3}$$

$$\begin{aligned} X &= X_1[l \quad e^{-j\omega\tau_{i,1}} \quad e^{-1\omega\tau_{i,2}} \cdots e^{-j\omega\tau_{i,M-1}}]^T, \\ &= X_1 d(\theta_2, \omega), \end{aligned} \tag{4}$$

where τ is a time delay component and $d(\theta, \omega)$ is a steering vector to the angle of direction of arrival (DoA).

θ_1 is the DoA of desired speech and θ_2 is that of interference speech.

Given an optimum weight vector, the output of the optimum beamforming or enhanced target speech is obtained by following equation

$$\hat{X} = w^H Y, \tag{5}$$

where H is Hermitian (complex conjugate) transpose and w is optimum weight vector.

The optimum weight vector w of MVDR beamforming is determined by solving the criterion of minimizing the output power while preserving the target speech signal such as[2]:

$$\min_{\omega} w^H R w \text{ subject to } d(\theta_1, \omega) w = 1 , \tag{6}$$

where R is the covariance matrix of acquired signal. The optimum solution is given by [2]

$$w = R^{-1} d(\theta_1, \omega)[d^H(\theta_1, \omega) R^{-1} d(\theta_1 \omega)]^{-1}. \tag{7}$$

Assuming that the DoA of target speech, θ_1, is given, how to obtain the covariance matrix R is very important. Note that the covariance matrix R must include interference and noise components only. Since it is difficult to extract interference component only, the conventional MVDR beamforming directly utilizes the acquired signal. After substituting R with R_{YY}, the covariance matrix is computed by taking a recursive averaging with a fixed value of forgetting factor [5] :

$$R_{YY}(l) = \lambda R_{YY}(l-1) + (l-\lambda)\, Y(l)\, Y^H(l), \tag{8}$$

where l and λ indicates the frame index and a forgetting factor, respectively. The phase of the covariance matrix R_{YY} consists of the phase terms in target and interference speech signal. Note that the mixed phase is not corresponded to phase of interference speech. Therefore, beamforming performance is significantly deteriorated.

III. Proposed Method

In this section, an estimation accuracy of covariance matrix is improved by including only the components of interference signal and diffused noise.

It can be realized by accurately obtaining DoA information using an audio-visual sensor based approach [6]. In addition, the performance enhances further by introducing a variable forgetting factor in Eq. (8).

A system block diagram of the proposed approach is shown in Fig. 1. In this system, precise DoA information is obtained by a depth camera and MA. Note that the DoA information of target/interference speech is estimated by employing a head detecting/tracking algorithm [6]. Then, the covariance matrix can be estimated by using the obtained DoA information. For the precise covariance matrix, the magnitude compensation method is utilized. In the covariance matrix, phase information is derived from the visual sensors and magnitude information is controlled by the variable forgetting factor.

3.1 Covariance matrices estimation

As is assumed above, speech signal and diffused noise are uncorrelated so that the covariance matrix of the received signal can be separated into three types of covariance matrices :

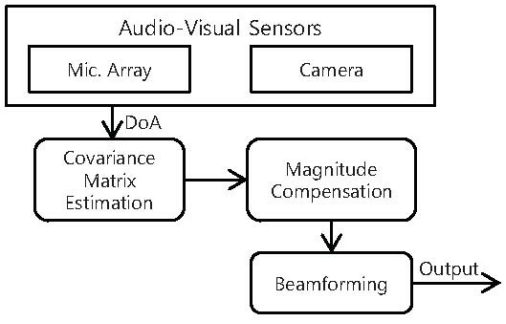


Fig. 1. Block diagram of the proposed system.

$$R_{YY}(l) = R_{XX}(l) + R_{II}(l) + R_{NN}(l), \quad (9)$$

where R_{XX}, R_{II}, and R_{NN} indicate target, interference speech, and noise covariance matrices. Since the desired covariance matrix should include only interference and noise components, R_{XX} term must be removed. In other words, the desired covariance matrix R_{NN+II} includes only the interference and noise terms [5]

$$\begin{aligned} R_{NN+II}(l) &= R_{II}(l) + R_{NN(l)} \\ &= E\{I(l)I^H(l)\} + E\{N(l)N^H(l)\} \\ &= \frac{1-\lambda}{1-\lambda^1}\sum_{l=1}^{K}\lambda^{K-1}\{|I_1(l)|^2 d(\theta_2, \omega), \\ &\quad d^H(\theta_2, \omega) + \sigma_{N^2} I_{M\times M}\} \end{aligned} \quad (10)$$

where K is the number of snapshot and $R_{II}(l)$ is derived by taking a product between scalar value $|I_1(l)^2|$ and steering vector $d(\theta_2, \omega)$, and a recursive averaging with a fixed forgetting factor λ. $d(\theta_2, \omega)$ can be constructed by the DoA obtained by audio-visual sensors and the unknown magnitude of interference signal $|I_1(l)^2|$ is estimated by modifying the known magnitude of output signal $|Y_1(l)^2|$.

In the proposed method, the estimated value of interference covariance matrix $R_{II}(l)$ is expressed as follows:

$$\begin{aligned} \hat{R}_{II}(l) &= \lambda \hat{R}_{II}(l-1) \\ &\quad + (1-\lambda)|Y_1(l)|^2 d(\theta_2\, \omega) d^H(\theta_2, \omega). \end{aligned} \quad (11)$$

The noise covariance matrix $R_{NN}(l)$ is also expressed as follows:

$$\hat{R}_{NN}(l) = \lambda \hat{R}_{NN}(l-1) + (l-\lambda)\sigma_{N^2} I_{M\times M}, \quad (12)$$

where $I_{M\times M}$ is an identity matrix and the power of noise term, σ_N^2, is updated in non-speech frames using a simple voice activity detection (VAD) algorithm.

3.2 Variable forgetting factor

To estimate the interference covariance matrix $R_{II}(l)$ given in Eq. (11), the forgetting factor, λ, needs to be determined. The forgetting factor compensates the inaccurate magnitude value $|Y_1(l)^2|$ of estimated covariance matrix. The update rate needs to be controlled dynamically to improve the accuracy. For example, if the magnitude of interference speech is much stronger than that of target speech, it is appropriate to use a small forgetting factor in order to rapidly track the interference speech. In this paper, a variable forgetting factor is introduced which varies depending on signal-to-interference ratio (SIR):

$$\begin{aligned} \hat{R}_{II}(l) &= \alpha(l) R_{II}(l-1) + [1-\alpha(l)] \\ &\quad |Y_1(l)^2| d(\theta_2, \omega) d^H(\theta_2, \omega), \end{aligned} \quad (13)$$

where

$$\alpha = \begin{cases} \alpha_{\max}, & \mathrm{SIR}(l) \geq \zeta_{\max}, \\ \alpha_{\min}, & \mathrm{SIR}(l) \leq \zeta_{\min}, \\ \min[\max[\dfrac{\log(\mathrm{SIR}(l)/\zeta_{\min})}{\log(\zeta_{\max}/\zeta_{\min})}, \alpha_{\min}], \alpha_{\max}], & \text{otherwise}, \end{cases} \quad (14)$$

$$\zeta_{\max} = 20\ \mathrm{dB}, \zeta_{\min} = -20\ \mathrm{dB}, \alpha_{\max} = 0.98,\ \alpha_{\min} = 0.6$$

In Eq.(14), the variable forgetting factor is varied from 0.98 to 0.6 depending on the SIR value. if the variable forgetting factor varies rapidly, the performance may be degraded. Therefore, upper and lower bounds of SIR value are introduced and the maximum and minimum values of the forgetting factor (0.98, 0.6) are empirically determined. Therefore, the variable forgetting factor is decreased in order to update the magnitude, if the power of the interference speech is high. Since the DoA information of target and interference speech is given already, SIR can be approximated by power ratio between θ_1 and θ_2 of steered response power (SRP). Note that the impact of $\hat{R}_{II}(l)$ and $\hat{R}_{NN}(l)$ to the desired covariance matrix $\hat{R}_{NN+II}(l)$ is different.

Therefore, an additional weighting term to each component may be desirable somehow. In tne reference 7, a generalized form of cost function was introduced to the multichannel Wiener Filter where μ is a tradeoff parameter between the noise reduction and the speech distortion. In this paper, a desired covariance matrix is finally expressed as

$$\hat{R}_{NN+II}(l) = \hat{R}_{II}(l) + \mu \hat{R}_{NN}(l), \quad (15)$$

where $\mu \geq 0$ is the tradeoff parameter between interference reduction and noise reduction. If $\mu > 1$, noise is more reduced by decreasing interference reduction and vice versa. μ is empirically determined to maximize SINR improvement in this paper. In this paper, μ is empirically selected as 6.

IV. Experiments

4.1 Experimental setup

Experiments are conducted to measure the effect of interference reduction and diffused noise reduction. 16 kHz sampled male and female speech are used as the target and interference speech and 512-point of FFT size is chosen. Diffused noise is created using white noise. Signal is mixed to have 5 dB signal-to-interference ratio (SIR) and white noise has 20 dB signal-to-noise ratio (SNR). A linear array has four microphones which are uniformly spaced with an interval of 4 - cm and it is located at the center of the room. The target/interference speech is fixed at the 0°/60° and located 1.5 m/2.5 m from the microphones in the room of which size is 6 m width, 6 m length, and 3 m height. The linear microphone array and speech have 1.7 m height. Fig. 2 depicts the simulation environment.

The first experiment is conducted in an artificial anechoic room. That is, there are only target, interference speech and diffused noise without reverberation. The second experiment is conducted in reverberant environment. By utilizing the image method[8], room impulse response (RIR) is artificially created with $T_{60} = 200\ ms$ and the RIR has 1000 taps.

4.2 Experimental result

Four performance evaluation methods such as noise reduction (NR), interference reduction (IR), speech distortion (SD), and signal-to-interference-plus-noise ratio improvement (SINRI) are measured to evaluate the performance [9].

Table 1 shows experimental result of conventional MVDR, LCMV, and other methods which are the method of covariance matrix estimation (Cov. Estimation), covariance matrix estimation with a variable forgetting factor (Cov. Estimation + Varable for. Fac.), and the proposed method which is covariance matrix estimation with a variable

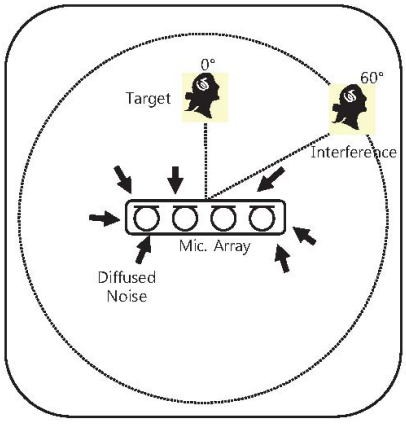
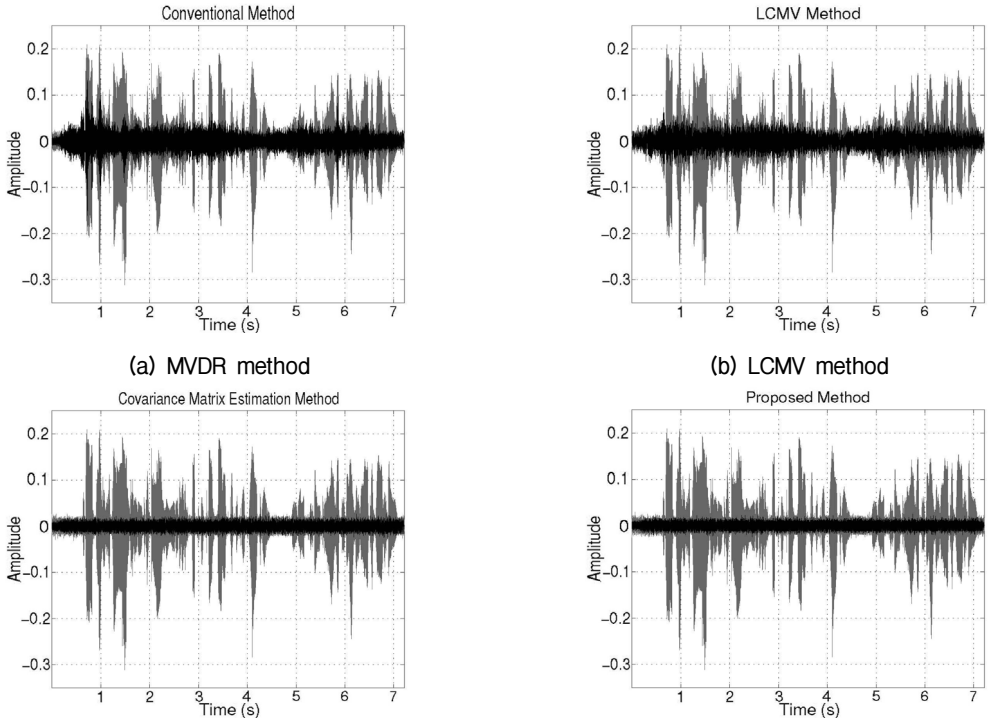


Fig. 2. Simulation environment.

Table 1. Performance of conventional and proposed beamforming in anechoic environment [dB].

	MVDR	LCMV	Cov. Estimation	Cov. Estimation + Varable for. Fac.	Proposed
NR	-4.6	-5.14	1.63	1.9	2.83
IR	12.56	40	37.8	36.8	30.68
SD	-13.46	-14.88	-21.6	-21.92	-22.65
SINRI	8.74	9.92	16.59	16.9	17.73



(a) MVDR method (b) LCMV method

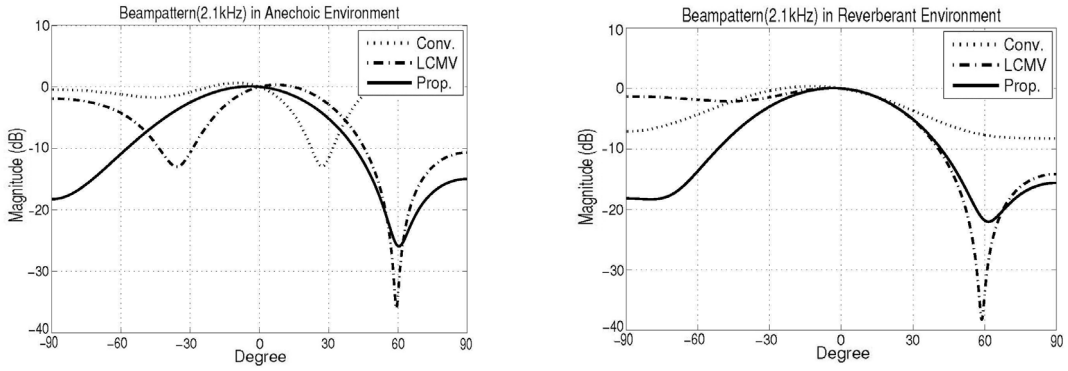
(c) Covariance matrix estimation (d) Proposed method

Fig. 3. Interference and noise reduction of conventional and proposed beamforming in anechoic environment.

forgetting factor and weighting factor in anechoic environment. MVDR and LCMV is referred from the reference 2. Also, third method creates the covariance matrix using the DoA information and fourth method creates it with variable forgetting factor. Finally, proposed method is additionally introduced the controlling weight factor. It can be seen that performance improves if the accurately estimated covariance matrix is used. Adopting a variable forgetting factor can attenuate the noise component and distortion without strong attenuation of the speech component though there is a little IR degradation. A controlling weight factor given in Eq. (15) also enhances beamforming performance.

Fig. 3 shows the interference and noise reduction of conventional and proposed beamformings in the anechoic environment. The gray signals indicate the original interference and noise components and the black signals indicate the reduced interference and noise components. The interference and noise components are substantially decreased by introducing the DoA information. In addition, noise component is more decreased in proposed method. Overall, noise and interference are decreased around 3 dB and 31 dB individually and speech is much less distorted than conventional methods in the first experiment. In terms of SINR improvement, there is around 18 dB gain. That is, the proposed MVDR method is much better performance than the LCMV one in terms of NR, SD, and SINR. The LCMV method has better performance than the proposed one in terms of IR because of the constraint on rejecting the interference speech. However, since the LCMV has a weak point of boosting the noise in low frequency region, the SINR improvement is small. On the other hand, the proposed MVDR beamforming simultaneously reduces interference and noise signal with less distortion. In addition, the ratio of interference/noise reduction is controllable by adjusting the weight factor.

Fig. 4 indicates a beampattern (2.1 kHz) of conventional MVDR, LCMV, and the proposed method in anechoic and reverberant environment. In order to attenuate interference speech while preserving the target speech, the beampattern



(a) Beampattern in anechoic environment (b) Beampattern in reverberant environment

Fig. 4. Beampattern (2.1 kHz) of conventional MVDR, LCMV, and the proposed method.

Table 2. Performance of conventional and proposed beamforming in reverberant environment [dB].

	MVDR	LCMV	Cov. Estimation	Cov. Estimation + Varable for. Fac.	Proposed
NR	-6.99	-6.38	1.6	1.88	2.76
IR	9.27	6.87	5.97	5.95	6.26
SD	-5.3	-5.54	-10.33	-10.35	-10.54
SINRI	0.39	0.63	5.43	5.43	5.73

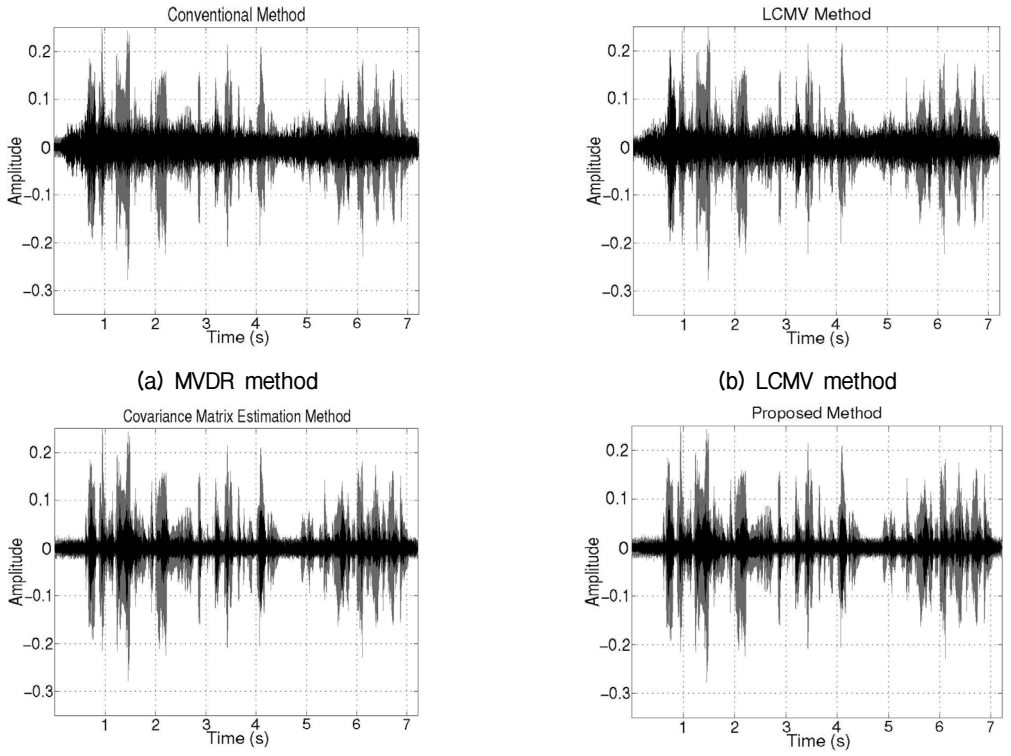
should have a characteristic of unity/low magnitude value at 0°/60°. LCMV can construct a null at 60° by using an additional criterion and the null does not exist in the beampattern of conventional MVDR so it can not reduce the interference speech as much as LCMV. However, since the proposed method estimates the covariance matrix accurately, its attenuation performance of the undesired interference speech is similar to LCMV. In addition, Fig. 4 shows that the proposed method attenuates even in reverberant environment. That is to say, look and null directions of beampatterns are not changed.

Table 2 shows experimental result of same beamforming methods in reverberant environment. Unlike anechoic environment, beamforming performance is considerably degraded because reverberation distorts phase information of target and interference speech. In comparison with Table 1, performance of all beamforming methods is degraded, especially interference reduction. Conventional MVDR and LCMV method also have a phenomenon of boosting the noise, however, the proposed method still reduces interference and noise signal. Overall, it can be analyzed that noise and interference are decreased around 3 dB and 6 dB individually. In the proposed method, the performance of NR and IR is increased and it can be considered as attenuation of the late reverberation of speech component. Therefore, the proposed method can attenuate both noise and interference signal even in reverberant environment.

Fig. 5 shows the interference and noise reduction of conventional and proposed beamforming in the reverberant environment. Similar to Fig. 3, the gray/black signals indicate the original/reduced interference and noise component, individually. Due to the reverberant, the interference and noise components are quite in case of the conventional MVDR and LCMV. However, the proposed method attenuates the interference and noise signal more than the conventional methods.

V. Conclusion

This paper has suggested an improved MVDR beamforming with A/V sensors in anechoic/reverberant environ-



(a) MVDR method (b) LCMV method

(c) Covariance matrix estimation (d) Proposed method

Fig. 5. Interference and noise reduction of conventional and proposed beamforming in reverberant environment.

ment. The proposed beamforming estimates a covariance matrix using DoA information and compensates inaccurate magnitude by a variable forgetting factor. The proposed method can be interpreted as a processing that handles the constraint of an interference component within a minimization criterion of estimating the precise covariance matrix. Using the proposed method, the beamforming system reduces both noise and interference signal not only in anechoic environment but also in reverberant environment. Though the performance is slightly degraded in reverberant environment, it still shows that the proposed method can be applied to real room environment.

Acknowledgement

This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science, and Technology (2010-0013394).

References

1. M. Brandstein and D. *Ward, Microphone Arrays − Signal Processing Techniques and Applications* (Springer-Verlag, Berlin, 2001), pp. 3-17, pp. 229-378.
2. B.D. Van Veen and K.M. Buckley, "Beamforming: a versatile approach to spatial filtering," ASSP Magazine, IEEE. **5**, 4-24 (1988).
3. J. Zhuang, P. Huang, and W. Huang, "Matched direction beamforming based on signal subspace," in IEEE ICASSP, 2585-2588 (2012).
4. H.K. Maganti, D. Gatica-Perez, and I. McCowan, "Speech enhancement and recognition in meetings with an audio−visual sensor array," IEEE Trans. on ASLP. **15**, 2257-2269 (2007).
5. J. Gu and P.J. Wolfe, "Robust adaptive beamforming using variable loading," in Workshop on Sensor Array and Multich. Proc. IEEE, 1-5 (2006).
6. J.S. Lee, G.K. You, J.M. Yang, and H.G. Kang, "Unified framework for user tracking and sound beamforming with audio/depth sensors in Kinect," in Workshop on Kinect in Pervasive Computing, Pervasive 2012, 1-4 (2012).
7. S. Doclo and M. Moonen, "On the output snr of the speech-distortion weighted multichannel wiener filter," IEEE

Signal Proc. Letters **12**, 809-811 (2005).
8. J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," J. Acoust. Soc. Am. **65**, 943-950 (1979).
9. J.Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing* (Springer-Verlag, Berlin, 2009), pp. 86-89.

Profile

▸ Gyeong-Kuk You(유경국)

Gyeong-Kuk You received the B.S. and M.S. degrees in electrical and electronic engineering from Yonsei University, Seoul, Korea, in 2012 and 2014, respectively. His research interests include speech/audio signal processing, speech enhancement, and microphone arrays.

▸ Jae-Mo Yang(양재모)

Jae-Mo Yang received the B.S., M.S., and Ph.D. degrees in electrical and electronic engineering from Yonsei University, Seoul, Korea, in 2007, 2008, and 2014. respectively. He served his internships at Microsoft Research Asia, Beijing, China, from 2010 to 2011, and Microsoft Research, Redmond, WA, in 2011, respectively. His research interests include speech/audio signal processing, speech enhancement, adaptive filters, and microphone arrays.

▸ Jinkyu Lee(이진규)

Jinkyu Lee received the B.S. in electrical and electronic engineering from Yonsei University, Seoul, Korea, in 2010 He is currently pursuing his M.S. and Ph.D. degree in electrical and electronic engineering at Yonsei University. He served his internships at Microsoft Research Asia, Beijing, China, from 2013 to 2014, and Microsoft Research, Redmond, WA, in 2014, respectively. His research interests include speech recognition, speech enhancement, and mahcine learning.

▸ Hong-Goo Kang(강홍구)

Hong-Goo Kang (M'02) received the B.S., M.S., and Ph.D. degrees from Yonsei University, Seoul, Korea, in 1989, 1991, and 1995, respectively. From 1996 to 2002, he was a senior technical staff member at AT&T Labs-Research, Florham Park, NJ. In 2002, he joined the Department of Electrical and Electronic Engineering, Yonsei University, where he is currently a Professor. He actively participated in international collaboration activities for making new speech/audio coding algorithms standardized by ITU-T and MPEG. His research interests include speech/audio signal processing, array signal processing, pattern recognition, and human computer interface. Dr. Kang was an associate editor of the IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING from 2005 to 2008. He served on numerous conferences and program committees. He was a vice chair of technical program committee in INTERSPEECH 2004 held in Jeju island, Korea. He is a technical reviewing committee member of the ICASSP and INTERSPEECH conferences.