

Target Signal Detection Using MUSIC Spectrum in Noise Environment

Sangjun Park, Sangbae Jeong, Moonsung Han, and Minsoo hahn

Abstract—In this paper, a target signal detection method using multiple signal classification (MUSIC) algorithm is proposed. The MUSIC algorithm is a subspace-based direction of arrival (DOA) estimation method. The algorithm detects the DOAs of multiple sources using the inverse of the eigenvalue-weighted eigen spectra. To apply the algorithm to target signal detection for GSC-based beamforming, we utilize its spectral response for the target DOA in noisy conditions. For evaluation of the algorithm, the performance of the proposed target signal detection method is compared with that of the normalized cross-correlation (NCC), the fixed beamforming, and the power ratio method. Experimental results show that the proposed algorithm significantly outperforms the conventional ones in receiver operating characteristics(ROC) curves.

Keywords—Beamforming, direction of arrival, multiple signal classification, target signal detection.

I. INTRODUCTION

SPEECH is the most convenient way of communications. Thus, speech-based human-machine interface technologies have been widely studied. Due to the performance improvement of speech recognition, which is one of the key technologies of the speech interface, it has been adopted to smart phones, smart TVs, car navigation, etc. However, speech recognition performance is highly dependent on environmental noises. Therefore, noise reduction technologies are considered mandatory to guarantee its reliable performance.

Speech-corrupting noises can be categorized as stationary or nonstationary ones. Stationary noises of which the spectral characteristics are slowly varying are successfully reduced by single channel noise cancellers such as Wiener and Kalman filters[1][2]. However, most real environmental noises are nonstationary ones and more than two microphones are utilized to reduce them. Among many microphone array-based speech enhancement techniques, generalized sidelobe canceller (GSC) has been widely applied [3]. The performance of GSC beamforming is very sensitive to the accuracy of target signal detector. Insufficient noise reduction in pure noise intervals and target signal distortion can be caused by inaccurate target signal detector at the beamforming output [4][5].

In this paper, to improve the performance of target signal detectors for GSC, novel target signal detection using the

Sangjun Park and Minsoo Hahn are with the Department of Electrical Engineering, KAIST, Daejeon, Korea

Sangbae Jeong is with the Department of Electronics Engineering/ERI, Gyeongsang National University, Jinju, Korea (Corresponding author to provide phone: +82-55- 772-1727, e-mail: jeongsb@gnu.ac.kr).

Moonsung Han is with the IT Convergence Tech. Research Lab., ETRI, Daejeon, Korea.

time-domain contour of MUSIC spectrum is proposed. MUSIC is a kind of DOA estimation algorithms using sub-space analysis in the environment that exists multiple signal sources [6]. To construct the time-domain contour of its spectral response for the target DOA, the eigenvalue analysis is performed to the covariance matrix of multi-channel inputs for every analysis frame. For additional performance improvement, MUSIC spectral responses in a given analysis frame are weighted by the inverse of corresponding eigenvalues. Finally, target intervals are determined by simple decision logic which compares the estimated contour and a pre-selected threshold.

By using the estimated contour, target intervals are declared by a simple thresholding technique.

This paper is organized as follows. Section II introduces related researches. Section II describes the proposed target signal detection method. Experimental results are summarized in Section IV. Finally, conclusions are given in Section V.

II. RELATED WORKS

A. GSC-based Beamforming Algorithm

A basic structure of GSC-based beamforming algorithm is shown in Fig. 2. GSC algorithm is composed of three processes: fixed beamforming (FBF), blocking matrix (BM) and noise cancellation (NC). In the FBF, steering the microphone array to “a virtual front” is performed using time delay compensation to maximize the signal power to the target DOA. Usually, the delay-and-sum beamforming is used for time delay compensation. However, noise signals cannot be eliminated completely, thus additional noise reduction is performed in BM

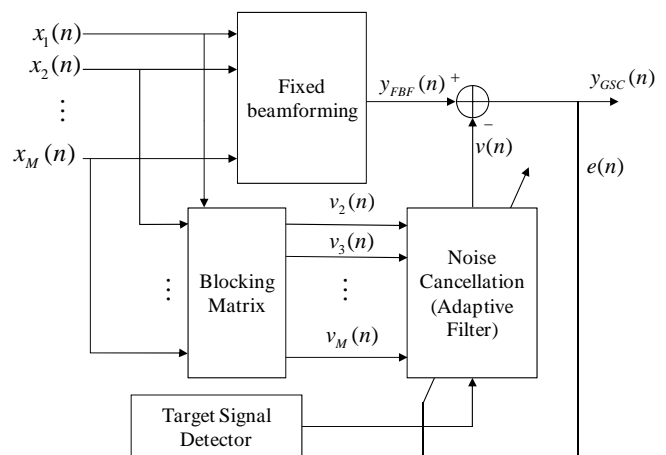


Fig. 1 The basic structure of the GSC-based beamforming algorithm

and NC. The NC reduces noise signals using the noise reference signals $v_i(n)$ estimated in the BM. Finally, the output of GSC algorithm is $y_{GSC}(n)$. The more noise can be eliminated by using adaptive filters in NC, however, the target signal would be distorted significantly. To solve this problem, the target signal detector is utilized. By using the target signal detector, filter coefficients are updated only at noise interval. Fewer noises are eliminated by using the target signal detector, but we can expect better noise reduction performance because it can prevent target signal distortion. Many researchers have been studied target signal detection methods[10][11]. Among them, the conventional methods using normalized cross-correlation (NCC), FBF output contour and power ratio of $v_i(n)$ are introduced, as follows.

B. Conventional target signal detection algorithm

1. Normalized Cross-Correlation (NCC) Method

In the target signal interval, two input signals, arbitrarily selected in the array, are highly correlated because of the time delay compensation. A NCC of the selected two channel signal becomes high in target intervals and low in noise intervals. The NCC between the i th channel and the j th channel is represented in (1).

$$P_{NCC} = \frac{\sum_{k=0}^{N-1} x_i(k - \Delta n_i) x_j(k - \Delta n_j)}{\sqrt{\sum_{k=0}^{N-1} x_i^2(k - \Delta n_i) \sum_{k=0}^{N-1} x_j^2(k - \Delta n_j)}} \quad (1)$$

where N is frame size and Δn_i is the number of delayed sample on the i th channel for FBF.

2. Fixed Beamforming Method

FBF maximizes the energy at the target DOA. Thus, only the noise signal is attenuated and the FBF output has enhanced signal-to-noise ratio (SNR) in target signal intervals[4]. The target signal can be detected using the FBF output with a proper threshold. The energy contour using the FBF output is denoted as (2).

$$P_{FBF} = \frac{1}{N} \sum_{k=0}^{N-1} |y_{FBF}(k)|^2 \quad (2)$$

$$y_{FBF}(n) = \frac{1}{M} \sum_{m=1}^M x_m(n - \Delta n_m) \quad (3)$$

where M is the number of channels.

3. Power ratio method

GSC algorithm estimates noise reference signals using BM[3][5]. The power ratio between the FBF output and the BM output can be effective to detect target signal. The BM estimates pure noise reference signals without target signals, thus the ratio

in target signal intervals would be large. Assume that the i th channel is a reference channel, the power ratio between the FBF output and the BM output is defined as (4).

$$P_{PR} = \frac{\sum_{k=0}^{N-1} |y_{FBF}(k)|^2}{\sum_{k=0}^{N-1} \frac{1}{M-1} \sum_{m \neq i} |v_m(k)|^2} \quad (4)$$

$$v_m(n) = x_m(n - \Delta n_m) - x_i(n - \Delta n_i) \quad (5)$$

where $v_m(n)$ is the BM output of the m th channel.

C. MUSIC Algorithm

The signal can be divided into a signal subspace and a noise subspace by eigenvalue analysis of covariance matrix. The subspace-based DOA estimation method is called multiple signal classification (MUSIC) algorithm. This method is one of the conventional DOA estimation methods proposed by Schmidt[6].

The multi-channel signals collected by M channels are linear combinations of the D incident waves and noises. The MUSIC algorithm begins with the following model in (6), (7).

$$\mathbf{X} = \mathbf{A}\mathbf{F} + \mathbf{W} \quad (6)$$

$$\begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_M \end{bmatrix} = \begin{bmatrix} \mathbf{a}(\theta_1) & \mathbf{a}(\theta_2) & \cdots & \mathbf{a}(\theta_D) \end{bmatrix} \begin{bmatrix} F_1 \\ F_2 \\ \vdots \\ F_D \end{bmatrix} + \begin{bmatrix} W_1 \\ W_2 \\ \vdots \\ W_M \end{bmatrix} \quad (7)$$

The complex vector $\mathbf{a}(\theta_i)$ which compensates gain and phase is determined by arrangement of microphones and the input path of source signals. According to the MUSIC model, $\mathbf{a}(\theta)$ is multiplied by complex signals for phase-shifting of each channel. Thus, to apply $\mathbf{a}(\theta)$ to time delay compensation of signals, \mathbf{F} should be a narrow-band signal[6][7]. In the case that \mathbf{F} is the narrow-band signal with center frequency ω_c and microphones are arranged linearly, $\mathbf{a}(\theta)$ can be represented as (8).

$$\mathbf{a}(\theta) = \begin{bmatrix} a_1 e^{-j\omega_c \Delta t_1} & a_2 e^{-j\omega_c \Delta t_2} & \cdots & a_M e^{-j\omega_c \Delta t_M} \end{bmatrix}^T \quad (8)$$

where θ , a_i and Δt_i are DOA, gain and time delay of each channels, respectively. The covariance matrix of \mathbf{X} can be obtained as (9).

$$\mathbf{R} = E[\mathbf{X}\mathbf{X}^H] \quad (9)$$

In (9), $E[\cdot]$ denotes the mean and H is the hermitian transpose. By eigenvalue analysis of \mathbf{R} , it can be divided into the D -dimensional signal subspace and the

$(M - D)$ -dimensional noise subspace [6]. The noise subspace \mathbf{E}_N and signal subspace \mathbf{E}_S are defined as (10), (11).

$$\mathbf{E}_N = [\mathbf{e}_1 \quad \mathbf{e}_2 \quad \cdots \quad \mathbf{e}_{M-D}] \quad (10)$$

$$\mathbf{E}_S = [\mathbf{e}_{M-D+1} \quad \mathbf{e}_{M-D+2} \quad \cdots \quad \mathbf{e}_M] \quad (11)$$

$$(\lambda_1 < \cdots < \lambda_{M-D} \ll \lambda_{M-D+1} < \cdots < \lambda_M)$$

Where \mathbf{e}_i is the i th eigenvector and λ_i is the eigenvalue corresponding to \mathbf{e}_i . The Euclidean distance between the signal subspace and $\mathbf{a}(\theta)$ is same with the length of $\mathbf{a}(\theta)$ projected on the noise subspace[6][7]. Thus, the distance and the MUSIC spectrum, the inverse of (12), can be represented by (12) and (13), respectively.

$$|d(\theta)|^2 = \mathbf{a}(\theta)^H \mathbf{E}_N \mathbf{E}_N^H \mathbf{a}(\theta) = \sum_{i=1}^{M-D} |\mathbf{e}_i^H \mathbf{a}(\theta)|^2 \quad (12)$$

$$P_{MU}(\theta) = \frac{1}{|d(\theta)|^2} = \frac{1}{\sum_{i=1}^{M-D} |\mathbf{e}_i^H \mathbf{a}(\theta)|^2} \quad (13)$$

In many situations, the number of sources is unknown. According to (13), the number of sources is needed in MUSIC algorithm. Additionally, it exploits little information of eigenvalues by hard-decision of eigenvector (signal subspace or noise subspace). To cope with these problems, the eigenvalues are employed for weighting the MUSIC spectrum. This method uses all subspace by soft-decision, and the weighted MUSIC spectrum is denoted as (14).

$$P_{MU,Weighted}(\theta) = \frac{1}{\sum_{i=1}^M \lambda_i |\mathbf{e}_i^H \mathbf{a}(\theta)|^2} \quad (14)$$

III. PROPOSED TARGET SIGNAL DETECTION ALGORITHM

The conventional target signal detection methods mentioned in sub-section II-B have several problems. The NCC method cannot perform well when the time delay of two channel is computed and compensated inaccurately in real noisy environment. Another drawback of this method is that it cannot use entire information in multi-channel inputs because only two channels of them are utilized. The upper bound of the FBF method is highly dependent on the performance of FBF which has very simple structure and shows limited SNR improvement. Also, the noise reference signals extracted from BM are likely to be inaccurate when the speech inputs are corrupted by nonstationary noises. Lastly, the power ratio of the BM output would decrease due to target signal leakage caused by simple subtraction of blocking matrices[5]. In this paper, the target signal detection algorithm using the MUSIC spectrum is proposed to solve these problems. The proposed algorithm detects the target signal using the MUSIC spectrum contour at

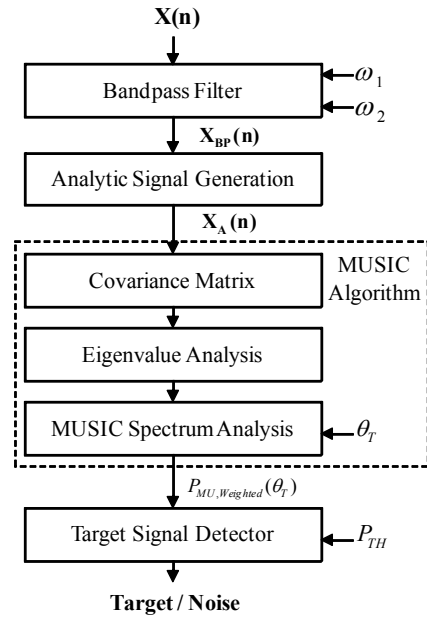


Fig. 2 The block diagram of proposed target signal detection algorithm

target DOA for beamforming. The MUSIC spectrum of the proposed method is obtained by the process in Fig. 2. In Fig. 2, resulting MUSIC spectrum for target DOA given by θ_T is $P_{MU,Weighted}(\theta_T)$. The target signal detector determines whether the signal is target or noise by comparing the music spectrum with threshold P_{TH} . The details about each block are summarized as follows.

A. Bandpass Filter

As mentioned in sub-section II-C, the input signals should be the narrow-band signal to compensate for the time delay by multiplying complex function $\mathbf{a}(\theta)$ [6][7]. In this paper, a band pass filter is used to generate the narrow-band signal. It is important to choose ω_1, ω_2 which are boundaries of the filter passband because the filtered signal $\mathbf{X}_{BP}(n)$ should be a narrow-band signal containing significant frequency components of the target signal in the passband.

B. Analytic Signal Generation

The input signal $X(n)$ should be the complex signal by the MUSIC algorithm model. However, we cannot obtain the complex signal from microphones in real environment. It is essential to express the input signal in complex form. In this way, to generate the complex signal from the real signal is called analytic signal generation (ASG) and the analytic signal means the generated complex signal[7]. The relation between the input signal $x(n)$ and the analytic signal $x_a(n)$ is shown in (15).

$$x_a(n) = \int_0^\pi 2X(\omega)e^{j\omega n} d\omega \quad (15)$$

where, $X(\omega)$ is the discrete fourier transform (DFT) of $x(n)$.

C. Target signal detector

The target signal detector uses $P_{MU,Weighted}(\theta_T)$ obtained by MUSIC spectrum analysis. The output of MUSIC spectrum is compared with empirically pre-selected threshold P_{TH} . When the output is larger than P_{TH} , the interval is determined as the target signal, and vice versa. The threshold is represented by

$$P_{TH} = \alpha_{TH} P_{MU,Weighted}^{(mean)}(\theta_T) \quad (16)$$

where α_{TH} is a gain multiplied to mean of initial parameters and $P_{MU,Weighted}^{(mean)}(\theta_T)$ is mean of the MUSIC spectrum on pure noise signal.

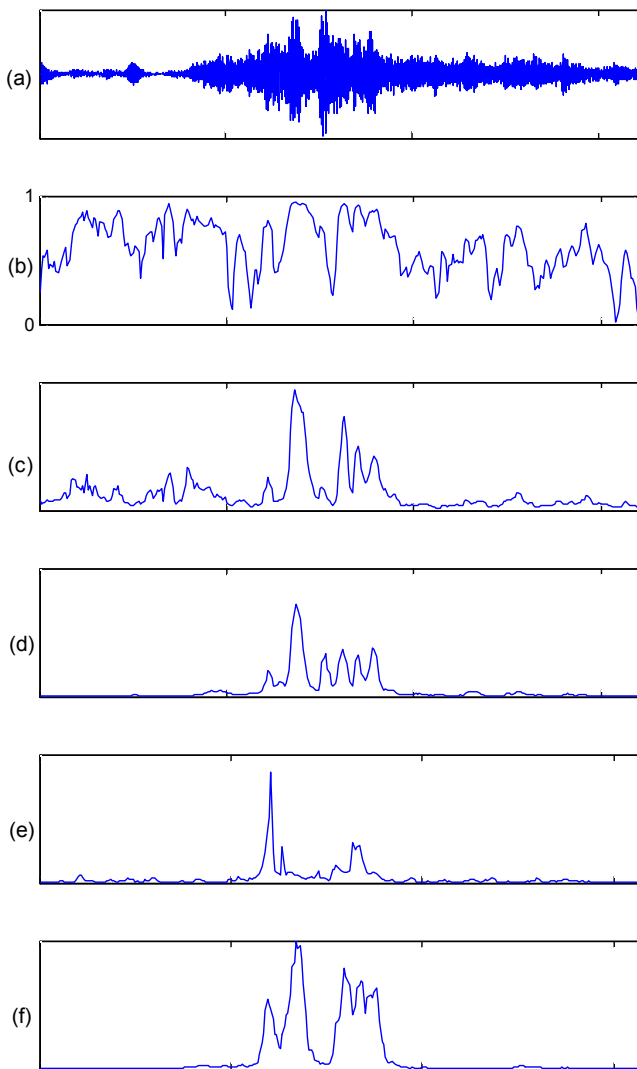


Fig. 3 (a) Input signal with 5dB input SNR (b) NCC contour (c) Power ratio contour (d) Power contour of FBF (e) MUSIC spectrum contour (f) Weighted MUSIC spectrum contour (*One second for a tick on the horizontal axis)

IV. EXPERIMENT AND RESULTS

A. Experiment condition

Test DB for performance evaluation was constructed by collecting target speech and noises separately in general office (6×5×3m) environment. They were synthesized artificially for each SNR and the pure noise signal was added on 1 second intervals before and after the target signal. SNR used in our experiments are 0, 5, 10, 15 and 20 dB. The total number of signal sources is two, a target source and a noise source. The distance between the microphone array and sources was 2m, the target source was in front of microphone array and the noise source was at 45° degree from the front. Four sets of 452 Korean isolated words were collected for target speech and a Korean talk show was played and recorded as noise. Thus, total (452×4×5=9040) samples were used for experiments. Microphones were placed linearly at the interval of 2 cm. All data were sampled with 16 kHz and 16 bit resolution.

The parameters in Fig. 2 are decided as follows. The $[\omega_1 = 100, \omega_2 = 500]$ Hz band where the energy of the speech signal is concentrated on was used for the filter passband. A 5th-order elliptic filter was employed for designing the band pass filter. θ_T was decided by 0° because the target source was placed in front of the microphone array. One hundred P_{TH} were computed by shifting α_{TH} from the minimum value to the maximum value which were derived in test DB by 100 steps. The initial 200 ms interval was assumed as pure noise to compute P_{TH} .

B. Performance Evaluation

The performance of the proposed algorithm is evaluated by comparing with conventional algorithms mentioned in sub-section II-B. The result waveforms of each algorithm are shown in Fig. 3. The proposed algorithm shows the discriminative contour. To validate the result, receiver operating characteristics (ROC) curves are illustrated in Fig. 4 using 100 P_{TH} derived by the method mentioned in sub-section IV-A. Horizontal axis of the figure means false acceptance rate (FAR), the ratio of detecting noise intervals as the target interval incorrectly, and vertical axis means true acceptance rate (TAR), the ratio of detecting target interval correctly. The area under the ROC curve (AUC), criterion of the performance is summarized in Table I.

TABLE I
 AUC FOR EACH ALGORITHM

Algorithm	AUC
NCC	0.6685
Power Ratio	0.7431
FBF	0.8051
MUSIC	0.7690
Weighted-MUSIC	0.8494

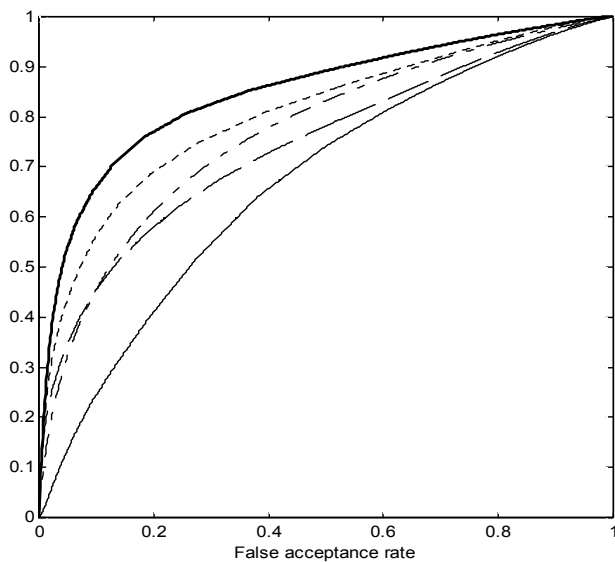


Fig. 4 ROC curves

(solid: NCC, dashed: Power ratio, dashed-dotted: MUSIC, dotted: FBF, thick solid: Weighted-MUSIC)

The closer ROC curve to upper left corner and the larger AUC means better performance. The proposed algorithm shows the best performance and NCC method is the worst. When TAR is fixed by 80%, general performance of target signal detector, the proposed method improves FAR about 15% comparing with the FBF method. The minimum error rates are summarized in Fig. 5. To measure the error rates, the threshold P_{TH} is decided empirically. The proposed algorithm shows 13.5% error rate, which decrease about 5% more than that of the FBF method.

V. CONCLUSION

In this paper, a new target signal detection algorithm using the MUSIC-based DOA estimation method was proposed. Experimental results show that the error rate of the proposed algorithm is 5% lower than that of the fixed beamforming method. As future works, we are planning to integrate the proposed algorithm to the GSC-based beamforming and to evaluate performances in the aspect of noise reduction and speech recognition.

ACKNOWLEDGMENT

This work was supported by the IT R&D program of MKE & KEIT [10041610, The development of the recognition technology for user identity, behavior and location that has a performance approaching recognition rates of 99% on 30 people by using perception sensor network in the real environment]

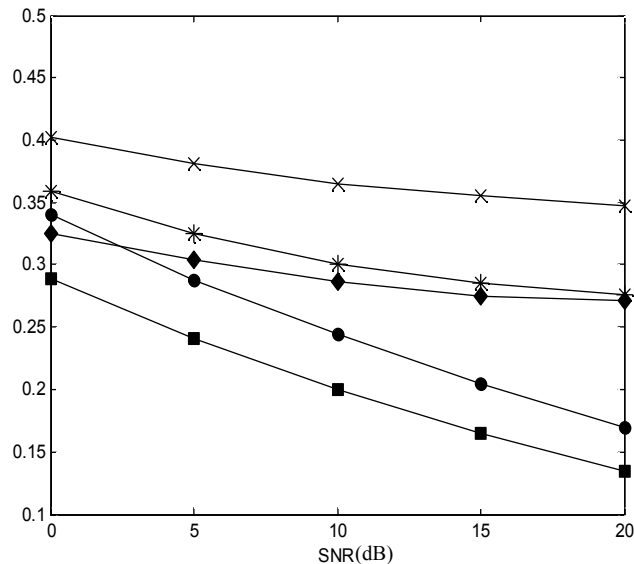


Fig. 5 Minimum error rates for each input SNR

(‘x’: NCC, ‘*’: Power ratio, ‘◆’: MUSIC, ‘●’: FBF, ‘■’: Weighted MUSIC)

REFERENCES

- [1] ETSI ES 202 212, Speech processing transmission and quality aspects (STQ), v.1.1.2., 2005.
- [2] S. Jeong and M. Hahn, "Speech quality and recognition rate improvement in car noise environments," *Electronics Letters*, vol. 37, no. 12, 2001, pp. 801-802.
- [3] J. Benesty, J. Chen, and Y. Huang, *Microphone array signal processing*, Springer, 2008.
- [4] Y. Lee, S. Kim, S. Han, M. Hahn, Y. Kim, and S. Jeong, "Probabilistic target speech detection and its application to multi-input-based speech enhancement," *Phonetics and speech sciences*, vol. 1, no. 3, 2009, pp. 97-104.
- [5] O. Hoshuyama, *et al.*, "A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters," *IEEE Trans. Signal Proc.*, vol 47, no. 10, 1999, pp. 2677-2688.
- [6] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas and Propagation.*, vol 34, no. 3, 1986, pp. 276-280.
- [7] K. Varma, "Time-delay-estimate based direction-of-arrival estimation for speech in reverberant environments," *M. S. thesis, dept. of electrical and computer engineering, Virginia Polytechnic Institute and State Univ.*, Virginia, USA, 2002.
- [8] M. Hayes, *Statistical digital signal processing and modeling*, John Wiley & Sons, 1996.
- [9] H. K. Hwang, *et al.*, "Direction of arrival estimation using a root-MUSIC algorithm," *IMECS 2008*, vol. 2, 1996, pp. 19-21.
- [10] Y. Jung, H. Kang, C. Lee, D. Youn, C. Choi, and J. Kim, "Adaptive microphone array system with two-stage adaptation mode controller," *IEICE Trans. Fund.*, vol. E88-A, no.4, pp.972-977, Apr.2005.
- [11] J. Hong, S. Hahn, S. Jeong, and M. Hahn, "Adaptive Microphone Array Processing for High-Performance Speech Recognition in Car Environment," *IEEE Trans. Consumer Electronics*, vol. 56, no. 2, 2010.