Robust Microphone Array Algorithm for Reverberant Environments

Junho Lee, Yang-Won Jung, Il-Whan Cha, H. Kang, Chungyong Lee
MCSP Lab., Dept. of Electrical and Electronic Eng., Yonsei University, Seoul, KOREA
junholee@cyclon.yonsei.ac.kr

Abstract
In this paper, a microphone array algorithm that overcomes the problem of signal cancellation in reverberant environments is proposed. The structure presented in this paper adopts a spatial smoothing algorithm, which was originally proposed for narrowband cases, to solve the problem of a desired signal cancellation phenomenon. In order to evaluate the performance of the proposed algorithm, we build a data acquisition room that models a real living room. With real measured data from the experiment room, the proposed algorithm outperformed conventional microphone array methods.

1. Introduction
Human-Machine Interface (HMI) is the most comfortable means that people communicate and interact with machines. To realize the HMI, speech processing unit acquiring and recognizing user's speech must be developed.

Traditional speech recognition modules use single microphone to get user’s speech signal. These modules have been successful in high SNR and single-source environments. However, with existing speech recognition modules, the recognition performance can be drastically degraded in practical environments[1]. Therefore, effective speech acquisition modules must be developed. Microphone Array (MA) have distinct advantage as they enable hands-free acquisition of speech with little constraint on the user.

In general, wideband Minimum Variance (MV) scheme that is extended version of Minimum Variance Distortionless Response (MVDR) beamformer is used for the MA system. The performance of the wideband MV algorithm, however, is degraded in the real acoustic environment due to array imperfection, target direction error and etc. Moreover, desired signal would be cancelled when target signal comes from the direction that is not an angle of incident[2]. In room environment, there always exist reverberation and it causes target signal propagate via multipath other than target direction. To prevent this phenomenon, use of Voice Activity Detector (VAD) have been described in the literature[3]. In such cases, the performance of overall system is greatly dependent on the accuracy of VAD.

In this paper, we propose a MA algorithm for real acoustic environments. To eliminate reverberation effect, the proposed algorithm incorporates spatial smoothing technique, which used in narrowband application. With proposed algorithm, VAD is not required and stability of the performance can be guaranteed. To evaluate the performance of the proposed algorithm, we build a data acquisition room that models a real living room and use real measured data from the experiment room. Presented results show that the proposed algorithm can be used for robust hands-free speech recognition modules.

2. Conventional MA System

2.1. Signal Model
Consider D broadband far-field signals, from directions \( \theta_1, \ldots, \theta_D \), propagate into the MA with \( M \) sensors. Assume the direction of desired signal is \( \theta_1 \) and signals from other directions are interferences. The input vector \( \mathbf{x}_k \) is defined as discrete Fourier transform of the received array data.

\[
\mathbf{x}_k = \mathbf{A}_k \mathbf{s}_k + \mathbf{n}_k
\]  

where,

\[
\mathbf{x}_k = [X_{1,k} \cdots X_{m,k} \cdots X_{M,k}]^T
\]

\[
\mathbf{A}_k = [\mathbf{a}_k(\theta_1) \cdots \mathbf{a}_k(\theta_j) \cdots \mathbf{a}_k(\theta_D)]
\]

\[
\mathbf{s}_k = [S_{1,k} \cdots S_{d,k} \cdots S_{D,k}]^T
\]

\[
\mathbf{n}_k = [N_{1,k} \cdots N_{m,k} \cdots N_{M,k}]^T
\]

and \( k \) denotes the frequency bin index. \( X_{m,k}, N_{m,k} \) are the discrete Fourier transform of observed signal and background noise at the \( m \)-th microphone and \( S_{d,k} \) is the discrete Fourier transform of the \( d \)-th signal source. The steering vector \( \mathbf{a}_k(\theta_j) \) of \( k \)-th frequency bin can be expressed as.
Figure 1: Block diagram of conventional MA system

![Block diagram of conventional MA system](image)

(a) DFT Estimate R
(b) Wideband MUSIC
(c) Wideband MV
(d) IDFT

The objective of wideband MV beamformer is to find the \( \mathbf{w}_k \), weights of each frequency bin, to the following linearly constrained minimization problem.

\[
\min_{\mathbf{w}_k} \mathbf{R}_k \mathbf{w}_k \quad \text{subject to} \quad \mathbf{a}_k^H(\theta) \mathbf{w}_k = 1
\]  

By using Lagrange multiplier, we can get optimum solution \( \mathbf{w}_{k_{\text{opt}}} \).

\[
\mathbf{w}_{k_{\text{opt}}} = \frac{\mathbf{R}_k^{-1} \mathbf{a}_k(\theta)}{\mathbf{a}_k^H(\theta) \mathbf{R}_k \mathbf{a}_k(\theta)}.
\]  

The Fig. 1 shows the block diagram of the MA system proposed by Asano[3]. Signal, received by \( M \) microphones, is decomposed by FFT and the spatial correlation matrix for each frequency bin is estimated. Using the estimated \( \mathbf{R}_k \), the direction of desired signal can be estimated by the wideband Multiple Signal Classification (MUSIC) algorithm[4]. Based on the estimated location, the weight can be obtained by (9). This system is stable only when \( \mathbf{R}_k \) is estimated in the noise interval. Since speech signal is propagated via not only a direct path but also multipath in indoor environment, the interference is highly correlated with the desired one. Unfortunately, to minimize the power contributed by noise and any signals coming from other directions than \( \theta \), the MV algorithm uses the correlated interference to cancel desired signal component[2]. To prevent this problem, Asano used VAD unit by human operator. By virtue of VAD, Asano’s MA system can update the weight in section where desired signal does not exist. This method, however, has a shortcoming that the performance of the whole MA system is influenced by VAD, additional to the inconvenience of human operated system. Moreover, it is very difficult to distinct desired signal from other ones when they are both speech signals. Therefore, MA system without VAD must be developed.

3. Proposed MA System

The desired signal cancellation phenomenon appears because of rank deficiency of spatial correlation matrix \( \mathbf{R} \) [5]. In order to overcome this problem, the rank of the matrix \( \mathbf{R} \) must be restored. This is generally achieved by Spatial Smoothing (SS) technique that utilize subarray averaging[5]. The SS was proposed in narrowband array signal processing to solve the coherence problem.

In this paper, we extended SS technique, so that application affords to broadband signal. After wideband input signal is decomposed into narrowband signal, SS scheme is applied in each narrowband signal. If we define \( p \) subsets among entire MA, input signal vector of each \( L \)-dimension subarray at \( k \)-th bin defined as

\[
\begin{align*}
\mathbf{x}_k^{(1)} &= [X_{1,k} \cdots X_{L,k}]^T \\
\mathbf{x}_k^{(2)} &= [X_{2,k} \cdots X_{L+1,k}]^T \\
\vdots \\
\mathbf{x}_k^{(p)} &= [X_{p,k} \cdots X_{L+p-1,k}]^T.
\end{align*}
\]  

Calculating spatial correlation matrix in each subarray, and averaging it, we get spatially smoothed correlation matrix \( \overline{\mathbf{R}}_k \).

\[
\overline{\mathbf{R}}_k = \frac{1}{p} \sum_{i=1}^{p} \mathbf{E} \left[ \mathbf{x}_k^{(i)} (\mathbf{x}_k^{(i)})^H \right]
\]  

With \( \overline{\mathbf{R}}_k \), desired signal cancellation problem in reverberant environment can be prevented. Fig. 2 illustrates the example in case the number of microphone and subarray are 4 and 3, respectively, is 4 and number of subarray microphone is 3.
The proposed MA system which employs the wideband SS is constructed. Fig. 3 illustrates the block diagram of the proposed MA system. Rather than estimating spatial correlation matrix in the noise interval, the proposed system estimates spatial correlation matrix in the entire interval. Therefore, the proposed MA system does not require VAD.

4. Experiment Result

To examine the performance of the suggested MA algorithm in reverberant environments, several experiments have been carried out. A experimental room was built, the size of which is 6m x 4m x 3m, that models a real living room. The reverberation time was 0.3s. Fig. 4 shows that experimental environment. Desired talker was located on sofa and spoke 2~5 syllable isolated words. Pseudo noise (PN) sequence and music were used for interference and played by loudspeaker. As MA, 9 microphones with linear, uniformly spaced with 6cm was used.

Desired signal cancellation phenomenon decreases as number of subarray increases, while the spatial resolution becomes poor due to the small size of effective aperture, vice versa. The subarray dimension is usually chosen from a compromise between the spatial resolution and the decorrelation effect[6]. The Table 1 shows the amount of improvement of SINR and spectral distance according to the number of subarray. For speech recognition modules, passing the desired signal without distortion is more important than interference rejection capability. Because of these relationships that mentioned above, we decided the subarray dimension by 6.

When the interference source was PN and music, the performance of the conventional MA system and the proposed system are shown in Fig. 5 and 6, respectively. In these figures, the output of the conventional and proposed MA system is shown when the spatial correlation matrix is estimated in the entire interval. From Fig. 5, it can be observed that the conventional MA system canceled the desired signal as well as interference signal. However, the proposed MA system preserves the desired signal with minimum distortion while the interference signal is minimized. The SINR (Signal to Interference and Noise Ratio) of the 1st channel input and the output is about 5dB and 15dB, respectively. Analogous result was obtained for the music interference. For comparison, the results of the single-microphone are also shown. The speech recognition score was summarized in Table 2. While speech recognition score was 62% for the single-microphone and was 69% for the conventional MA system, the recognition score was about 89% for the proposed MA system.
5. Conclusions

In this paper, the robust MA system that is applicable speech recognition modules is presented. The proposed MA system is constructed based on the wideband MV scheme, and it also employs a spatial smoothing algorithm to solve the problem of desired signal cancellation. With proposed system, no additional VAD logic is required to preserve characteristic of desired speech signal, while interference power minimization is achieved. SINR is improved more than 10dB and the recognition score is enhanced about 30% comparison with conventional single microphone system.

![Figure 6](image)

**Figure 6**: The performance of conventional and proposed MA system for music interference (a) reference signal (b) 1st mic. signal (c) output signal of the conventional MA system (d) output signal of the proposed MA system

**Table 1**: The amount of improvement for SINR, spectral distance according to subarray dimension

<table>
<thead>
<tr>
<th>Subarray dim. Improvement</th>
<th>9</th>
<th>8</th>
<th>7</th>
<th>6</th>
<th>5</th>
</tr>
</thead>
<tbody>
<tr>
<td>SINR improvement</td>
<td>3.2</td>
<td>8.6</td>
<td>11.9</td>
<td>10.1</td>
<td>8.0</td>
</tr>
<tr>
<td>Spectral distance</td>
<td>0.9</td>
<td>1.9</td>
<td>2.3</td>
<td>2.7</td>
<td>2.6</td>
</tr>
</tbody>
</table>

**Table 2**: Recognition score of the conventional MA system and the proposed MA system

<table>
<thead>
<tr>
<th></th>
<th>1st mic.</th>
<th>Conventional MA system</th>
<th>Proposed MA system</th>
</tr>
</thead>
<tbody>
<tr>
<td>Recognition Score</td>
<td>62.3%</td>
<td>68.8%</td>
<td>88.8%</td>
</tr>
</tbody>
</table>

6. References


