AN ADAPTIVE BEAMFORMING MICROPHONE ARRAY SYSTEM USING A BLIND DECONVOLUTION

Jin-Nam Park, Tsuyoshi Usagawa, Masanao Ebata

Department of Computer Science, Kumamoto University, Kumamoto 860-8555, Japan

ABSTRACT

This paper proposes an adaptive microphone array using blind deconvolution.

The method realizes an signal enhancement based on the combination with beamforming using blind deconvolution, synchronized summation and DSA method. The proposed method improves a performance of estimation by the iterative operation of blind deconvolution using a cost-function base on the coherency function.

The proposed method deconvolves the impulse response between microphones for certain direction to enhance the target signal. The output signals of blind deconvolution are fed into the conventional synchronized summation process assisted by the DSA method to obtain the DOA of target signal. An iterative operation both of the blind deconvolution and synchronized summation improves the total performance of signal enhancement. Simulations and experiments are conducted to show the performance of the proposed method.

1 INTRODUCTION

There are many researches to develop signal enhancement systems which can work under common workspace environment; concurrent speakers in reverberant room within ambient noise. The deterioration of quality of enhanced speech comes from many reasons such as reverberation and noise including concurrent speech. To obtain the better performance, we usually place microphones close to each speaker. However this approach has an obvious limitation, so that directional microphone systems are often considered. Directional microphone systems can be categorized into two types; a beamformer and an adaptive notch filter. An adaptive notch filter has very high directivity, however there is the strict limitation that the number of controllable noise sources and reflections must be less than the number of microphones. Also the directivity of a beamformer depends on the number of microphones and the physical spacing of microphones, and it is not always sufficient for lower frequency range.

As a beamforming algorithm, the synchronized summation is commonly used, but it is not effective to reduce the effect of reverberation and the low frequency ambient noise or is required large number of microphones to achieve sufficient performance. One of the main reasons for these degradations in the synchronized summation is the error in the DOA(Direction of Arrival) estimation [1]. To overcome this problem, we can use blind deconvolution to extract coherent signal from observed output signals of microphone array.

This paper proposes a method of signal enhancement combining blind deconvolution[2][3] with conventional array processing technique[1][4][5] using a cost-function[3][6]. The proposed method has an iterative process based on the DSA(Delay and Sum Array) method [4][7][8] for estimation of DOA and AMUSE(Algorithm for Multiple Unknown Signals Extraction) [8] for blind deconvolution.

2 A SIGNAL ENHANCEMENT USING THE BLIND DECONVOLUTION

An iterative algorithm for signal enhancement is detailed in this section. This algorithm implements an iterative operation combining the blind deconvolution with the synchronized summation. The blind deconvolution uses an iterative AMUSE and the synchronized summation produces the enhanced signal based on a DOA estimated by the DSA method.

2.1 The notation of output signals of a linear microphone array

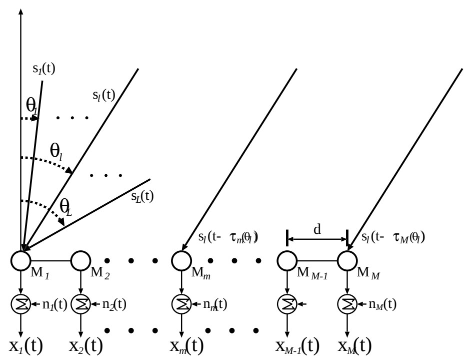


Figure 1: Schematic model of microphone array

Let us assume that a linear microphone array which consists of M omnidirectional microphones to observe L independent directions as shown Fig.1. The interval of microphones is d and all of the observed signals are assumed as plane waves.

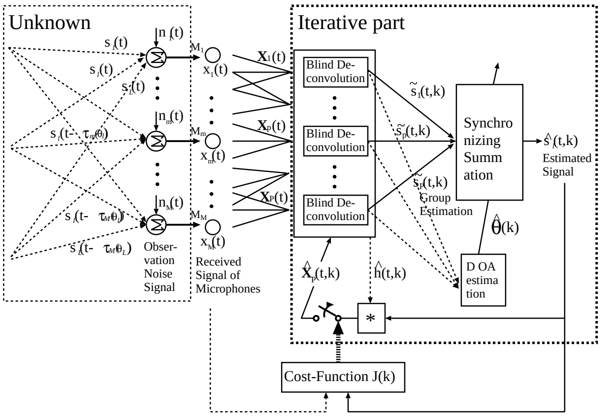


Figure 2: A block diagram of the proposed method

Let $s_l(t)$ be an observed signal component which arrived at the first microphone M_1 from a direction θ_l. When an impulse response $h_{m,l}(t)$ represents the impulse response from microphone M_1 to M_m for the direction θ_l, and an additive noise at the microphone M_m is $n_m(t)$, an output signal $x_m(t)$ of microphone M_m can be written as,

$$\begin{aligned} x_m(t) &= \sum_{l=1}^{L} s_l(t-\tau_m(\theta_l)) + n_m(t), \quad (1) \\ &= \sum_{l=1}^{L} x_{m,l}(t) + n_m(t), \end{aligned}$$

where
$$\begin{aligned} x_{m,l}(t) &= s_l(t-\tau_m(\theta_l)), \\ \tau_m(\theta_l) &= (m-1)d\ sin(\theta_l)/c, \end{aligned}$$

and c is the sound velocity. Because the impulse response $h_{m,l}(t)$ is defined as,

$$h_{m,l}(t) = \delta(t-\tau_m(\theta_l)), \quad (2)$$

Eq.(1) is rewritten as,

$$x_m(t) = \sum_{l=1}^{L} h_{m,l}(t) * s_l(t) + n_m(t), \quad (3)$$

where $*$ indicates a convolution and $\delta(t)$ is a delta function. Note that $h_{1,l}(t) = \delta(t)$.

To apply the blind deconvolution on sub-group of output signals of microphone array, output signals are divided into P groups in vector forms $\mathbf{x}_p(t)$ $(p = 1, \cdots, P)$ as

$$\begin{aligned} \mathbf{x}_p(t) &= [x_p(t),\ x_{p+1}(t),\ \cdots,\ x_{p+N-1}(t)], \quad (4) \\ N &= M-P+1. \end{aligned}$$

Note that each vector $\mathbf{x}_p(t)$ includes N output signals of microphone array.

2.2 Iterative AMUSE

Let $\hat{s}_l(t,k)$ denote an estimated signal and $\hat{h}_{p,l}(t,k)$ be an estimated impulse response at kth iteration.

An estimated output signal vector corresponding to the source θ_l of microphone M_p at the each iteration, $\hat{\mathbf{x}}_p(t,k)$, is given as,

$$\hat{\mathbf{x}}_p(t,k) = \hat{h}_{p,l}(t,k-1) * \hat{s}_l(t,k-1), \quad (5)$$

where an initial vector is defined as followed,

$$\hat{\mathbf{x}}_p(t,1) = \mathbf{x}_p(t).$$

Eigenvectors $\mathbf{u}_{p,n}(k)(n = 1,\cdots,N)$ and eigenvalues $\lambda_{p,n}(k)(n = 1,\cdots,N)$ are calculated using a covariance matrix $\hat{\mathbf{R}}_{\hat{\mathbf{x}}_p}(t,k)$ of $\hat{\mathbf{x}}_p(t,k)$.

$$\begin{aligned} \hat{\mathbf{R}}_{\hat{\mathbf{x}}_p}(t,k) &= E[\hat{\mathbf{x}}_p(t,k)\hat{\mathbf{x}}_p(t-\tau,k)] \quad (6) \\ SVD(\hat{\mathbf{R}}_{\hat{\mathbf{x}}_p}) &= [\mathbf{u}_{p,1}(k),\ \cdots,\ \mathbf{u}_{p,N}(k)], \\ &\quad diag(\lambda^2_{p,1}(k),\ \cdots,\ \lambda^2_{p,N}(k)) \\ &\quad [\mathbf{u}_{p,1}(k),\ \cdots,\ \mathbf{u}_{p,N}(k)]^T, \quad (7) \end{aligned}$$

where $SVD(\cdot)$ denotes the singular value decomposition[3][4][8].

The characteristic sequence of source $\mathbf{y}_p(t,k)$ is obtained by eigenvectors $\mathbf{u}_{p,n}(k)$ and eigenvalues $\lambda_{p,n}(k)$ as follows,

$$\mathbf{y}_p(t,k) = \mathbf{T}_p(k)\mathbf{x}_p(t), \quad (8)$$

where

$$\begin{aligned} \mathbf{T}_p(k) &= diag\left(\frac{1}{d_1(k)}, \frac{1}{d_2(k)}, \cdots, \frac{1}{d_{\hat{L}}(k)}\right)\mathbf{U}^T_{p,s}(k), \\ d_l(k) &= \sqrt{\lambda^2_{p,l}(k)-\sigma^2}, \quad l = 1,2,\cdots,\hat{L} \\ \mathbf{U}_{p,s}(k) &= [\mathbf{u}_{p,1}(k),\cdots,\mathbf{u}_{p,\hat{L}}(k)]. \end{aligned}$$

Note that ^{T} indicates a transpose and σ^2 is a variance corresponding to an additive noise. An estimated number of sources is assumed as $\hat{L}$ [7][8].

$\mathbf{V}_p(t,k)$ is an eigenvector obtained from a covariance matrix from $\mathbf{R}\mathbf{y}_p(t,k)$ of $\mathbf{y}_p(t,k)$ by SVD method as follows,

$$\begin{aligned} \mathbf{V}_p(t,k)\mathbf{\Sigma}_p(k)\mathbf{V}^T_p(t,k) &= SVD((\mathbf{R}\mathbf{y}_p(t,k) \\ &\quad +\mathbf{R}^T_{\mathbf{y}_p}(t,k))/2). \quad (9) \end{aligned}$$

where $\mathbf{\Sigma}_p(k)$ is a singular value matrix. It is able to estimate a source signal $\tilde{s}_p(t,k)$ and an impulse response $\tilde{h}_{p,l}(t)$ using an eigenvector $\mathbf{V}_p(t,k)$.

$$\begin{aligned} \tilde{s}_p(t,k) &= \mathbf{V}^T_p(t,k)\mathbf{y}_p(t,k) \quad (10) \\ \tilde{h}_{p,l}(t) &= \mathbf{T}^{\dagger}_p(k)\mathbf{V}_p(t,k) \quad (11) \end{aligned}$$

where † is a pseudo-inverse defined in the literature [3].

2.3 DOA estimation and synchronized summation

A DOA is estimated by the DSA method using estimated source signals $\tilde{s}_p(t,k)$ from each group $\hat{\mathbf{x}}_p(t,k)$ for $p = 1,\cdots,P$.

The directivity function, $P_{DSA}(\theta)$, is estimated as

$$P_{DSA}(\theta) = \frac{1}{T}\sum_{t=1}^{T}|\sum_{p=1}^{P} e^{-j2\pi f(p-1)d\ \sin\theta/c} \cdot\tilde{s}_p(t,k)|^2, \quad (12)$$

$$
\begin{aligned}
&= \frac{1}{T}\sum_{t=1}^{T} |\mathbf{d}(\theta)^H \tilde{\mathbf{s}}(t,k)|^2, \\
&= \mathbf{d}(\theta)^H \hat{\mathbf{R}}_{\tilde{\mathbf{S}}}(k)\mathbf{d}(\theta), \\
&\qquad (-\pi/2 \text{ !e } \theta \text{ !e } \pi/2)
\end{aligned}
$$

In these equations, T denotes the observed time, ^{H} denotes a complex conjugate and transpose, $\mathbf{d}(\theta)$ is a complex s-canning vector, and $\hat{\mathbf{R}}_{\tilde{\mathbf{S}}}(k)$ is a covariance matrix of $\tilde{s}(t,k)$. The DOA $\hat{\theta}$ is defined at which $P_{DSA}(\theta)$ defined in Eq.(12) gives a maximum value.

We are able to obtain an estimated signal $\hat{s}_l(t,k)$ using the synchronized summation corresponding to the estimated direction $\hat{\theta}$ as follows,

$$
\hat{s}_l(t,k) = \sum_{p=1}^{P} \tilde{s}_p(t - \hat{\tau}_p(\hat{\theta}), k),
$$

where $\hat{\tau}_p(\hat{\theta}) = (p-1)d\ sin(\hat{\theta})/c$.

2.4 Performance evaluation using a cost-function based on a coherency function

Because the estimated signal $\hat{s}_l(t,k)$ can be treated as an enhanced signal, we can retrieve the more accurate information on $h_{m,l}(t)$ from the output signals of microphone array.

The information on $h_{m,l}(t)$ is obtained in the characteristic sequence $\mathbf{y}_p(t,k)$ and it can be estimated through the eigenvectors and eigenvalues of the covariance $\hat{\mathbf{R}}_{\tilde{\mathbf{X}}_p}(t,k)$ as shown in Eq.(6).

To control the iteration, the cost-function $J(k)$ based on the coherency function is introduced as follow,

$$
\begin{aligned}
J(k) = \frac{1}{M}\sum_{m=1}^{M} \cdot \frac{1}{Q_m}\sum_{q=1}^{Q_m} \| &\gamma^2_{x_m x_1}(f_q) \\
&- \gamma^2_{\hat{s}_l(k)(x_m - \hat{s}_l(k))}(f_q) \| \qquad (13)
\end{aligned}
$$

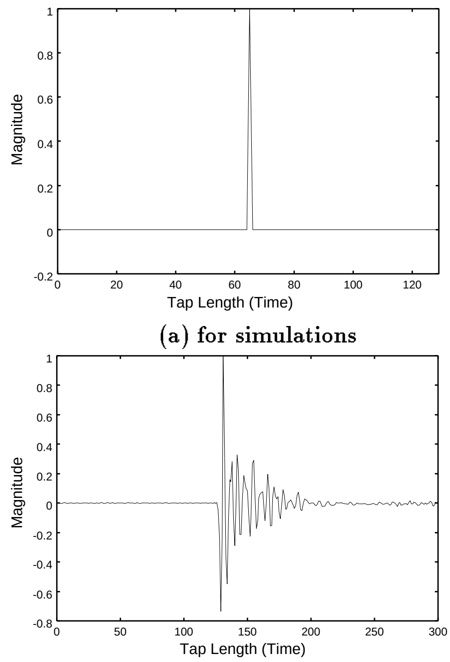
where $\gamma^2_{x_m x_1}(f_q)$ is the coherency frequency between $x_m(t)$ and $x_1(t)$. Discrete frequencies $f_q, (q = 1, \cdots, Q_m)$ are ones at which $\gamma^2_{x_m x_1}(f_q) \geq 0.99$.

The coherency function $\gamma^2_{\hat{s}_l(k)(x_m - \hat{s}_l(k))}$ denotes the relationship between an estimated signal $\hat{s}_l(t,k)$ and $x_m(t) - \hat{s}_l(t,k)$. The closer this coherency function to the value 0, the more accurate the estimated signal $\hat{s}_l(t,k)$ is.

3 SIMULATIONS AND EXPERIMENTS

Sources used in the simulations and experiments are five vowels, /a/,/i/,/u/,/e/,/o/,uttered by Japanese. A white noise is introduced as an ambient noise at $-40dB$ against the target signal. The sampling frequency is $10kHz$, the number of microphones M is 8, and the interval between microphones d is $0.04m$ for all simulations and experiments. Also the orthogonal direction of microphone array is assumed as $0°$.

In the simulations, an impulse response $h_{m,l}(t)$ is represented by a simple delay. On the other hand, an impulse response $h_{m,l}(t)$ used in experiments is a measured impulse response from a signal generator to the m-th microphone M_m of the array through amplifiers and loudspeaker in an anechoic room. Figure 3 (a) shows an example of impulse response used for the simulations and (b) is one used for the experiments. Both in the simulation and the experiments, only odd parts of output signal groups $\mathbf{x}_p(t)$ are used while $N = 4$, $P = 5$ and $M = 8$ in Eq.(4). The sta-



(a) for simulations

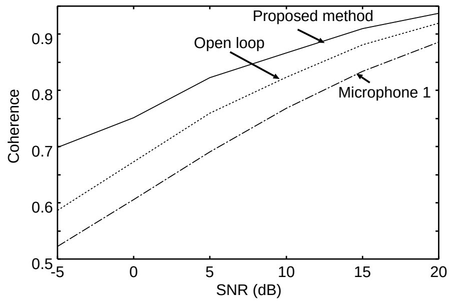
(b) for experiments

Figure 3: Examples of impulse response used for simulations and experiments

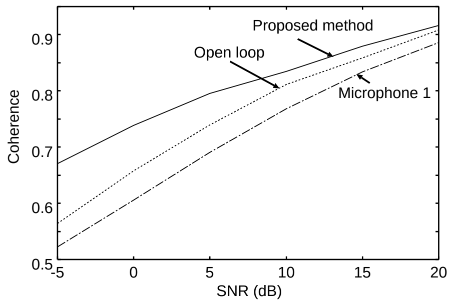
Table 1: The arrangement of two sources

	Target signal (s_1)	DOA	Another signal(s_2)	DOA
Case 1	Five Japanese vowels	$10°$	Five Japanese vowels	$-20°$
Case 2	Five Japanese vowels	$30°$	Five Japanese vowels	$-30°$

tistical performance of source estimation is evaluated using five vowels with the condition shown Table. 1. Because each vowel is uttered by the same Japanese, the signals used for the target $s_1(t)$ and ones for the other $s_2(t)$ are selected as different vowels. That is the total number of examinations are 20 not 25. For all combinations, SNR is varied from $-5dB$ to $+20dB$ by the $5dB$ step. Figures 4 and 5 show the performance of source estimation against various SNR. The dashed-and-point lines in Fig.4 and 5 show the averaged coherency between target source signal



(a) Case 1



(b) Case 2

Figure 4: The performance by mean value of coherence in simulation

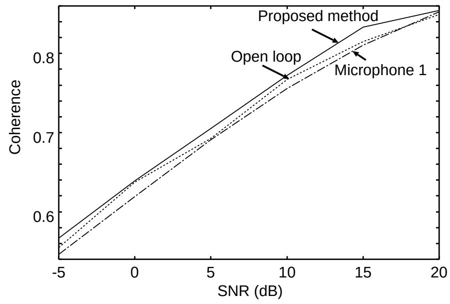
and the output signal of the first microphone $x_1(t)$. The dashed lines are the coherency between the target source signal and the estimated signal $s_1(t,1)$ without iteration, labeled as Open Loop in figures. The solid lines are the results obtained by the proposed method.

These results show the improvement of the accuracy of estimated target signal by the iterative processes. By comparing with the results of Case 1 and 2, the performance depends on the DOA of target and other signals. Also by comparing with the results of Fig.4 and Fig.5, the complexity of the impulse responses directly affects the performance of the proposed method. Note that the degradation of the performance is caused not only by the error of microphone arrangement but also the characteristics of loudspeaker and reflections of measured environment.

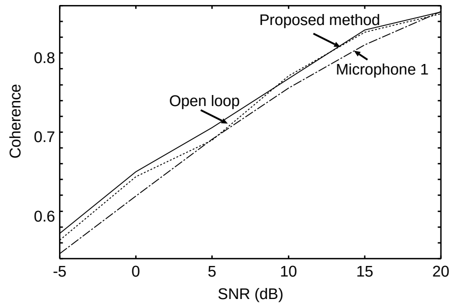
4 CONCLUSION

This paper proposed the signal enhancement method using blind deconvolution for microphone array system. This method deconvolves the impulse responses between microphones for certain direction to estimate an enhanced signal. The outputs of blind deconvolution are fed into the conventional synchronized summation process assisted by the DSA method to obtain a DOA of target signal. An iterative operation both of the blind deconvolution and synchronized summation improves the total performance of signal enhancement.

Simulations show the effectiveness of the proposed method. The results obtained by experiments are not as good as ones obtained simulations, but still the proposed method is effective in some extent.



(a) Case 1



(b) Case 2

Figure 5: The performance as a mean value of coherence in experiment

5 REFERENCES

1. Simom Haykin, James H. Justice, Normon L. Owesley, J.L.Yen, and Avi C. Kak " Array Signal Processing," pp.115-150, Prentice hall, Englewood Cliffs, NJ, 1985.
2. Simon Haykin, "Blind Deconvolution," Prentice-Hall, Englewood Cliffs, NJ, 1994
3. Simon Haykin, "Adaptive Filter Theory," pp.524-525 and 772-816, 2nd Ed., Prentice-Hall, Englewood Cliffs, NJ, 1991.
4. Jin-Nam park, Chun-Duck Kim, Tsuyoshi Usagawa, Masanao Ebata, "Characteristics comparison of the direction estimation method for a source signal in the noisy room," vol.53, no.4, pp.285-289, J.Acoust.Soc.Jpn.(J), 1997.
5. J.Ohga, Y.Yamasaki, Y.Kaneda, "Acoustic system and Digital process," pp.173-218, IEICE, Tokyo, 1995(in Japanese).
6. Bellini S, "Blind Equalization," vol.LVII, pp.445-450, Alta Frequenza, 1988.
7. Jin-Nam Park, Takashi Honda, Yoshifumi Chisaki, Tsuyoshi Usagawa, Masanao Ebata, "An estimation of source using Beamforming and Blind Deconvolution," pp.37-44(EA97-67), Technical Report of IEICE(Engineering Acoustics), 1997(in Japanese).
8. Lang Tong, Ruey-wen Liu, "Indeterminacy and Identifiability of Blind Identification," vol.38, no.5, pp.499-509, IEEE Trans. on CAS, 1991.