DIRECTION OF ARRIVAL ESTIMATION BASED ON NONLINEAR MICROPHONE ARRAY

Hidekazu Kamiyanagida†, *Hiroshi Saruwatari*†, *Kazuya Takeda* ‡, *Fumitada Itakura* ‡

† Graduate School of Information Science, Nara Institute of Science and Technology
8916-5, Takayama-cho, Ikoma, 630–0101, JAPAN
hide-ka@is.aist-nara.ac.jp
‡ Nagoya University / Center for Integrated Acoustic Information Research
Furo-cho, Chikusa-ku, Nagoya, 464–8603, JAPAN

ABSTRACT

This paper describes a new method for estimating the direction of arrival (DOA) using a nonlinear microphone array based on complementary beamforming. Complementary beamforming is based on two types of beamformers designed to obtain complementary directivity patterns each other. In this system, since the resultant directivity pattern is proportional to the product of these directivity patterns, the proposed method can be used to estimate DOAs even when the number of sound sources is equal to or exceeds that of microphones. First, DOA-estimation experiments are performed using actual devices in real acoustic environments. The results clarify that DOA estimation for two sound sources can be accomplished by the proposed method with only two microphones. Also, by comparing the resolutions of DOA estimation by the proposed method and by the conventional minimum variance method, we can show that the performance of the proposed method is superior to that of the conventional method.

1. INTRODUCTION

In acoustic signal processing, DOA estimation plays an essential part in microphone-array technology as a preprocess in speech enhancement or noise-robust speech recognition [1, 2]. The minimum variance (MV) method and the multiple signal classification (MUSIC) method are the conventional and popular DOA-estimation methods currently used in microphone array systems [1, 3, 4]. In general, the DOA-estimation results of conventional methods are accurate, especially when dealing with a small number of sound sources with many microphones. However, the performances of conventional methods are greatly degraded when the number of sound sources exceeds that of microphones. Thus, it is impossible to estimate DOAs using the practical microphone array system with a small number of elements.

In order to resolve this problem, we newly propose to utilize a nonlinear microphone array system based on complementary beamforming which has been proposed by one of the authors for efficient speech enhancement [5, 6]. Complementary beamforming is based on two types of beamformers designed to obtain complementary directivity patterns with respect to each other. In this system, since the resultant directivity pattern is proportional to the product of these directivity patterns, the proposed method can be used to estimate DOAs even when the number of sound sources is equal to or exceeds that of microphones. The proposed method enables the estimation of DOAs with a fairly small and practical microphone array.

In this paper, some experiments using actual devices are carried out to test the performance of the proposed method, and the resolutions of DOA estimation are measured and reported to quantify the performance.

2. PROPOSED ALGORITHM

In this section, the nonlinear microphone array based on the complementary beamforming proposed in Refs. [5, 6] is briefly described, and the new DOA-estimation algorithm for this array is proposed.

2.1. Nonlinear Microphone Array

In this study, a straight-line array is assumed. The coordinates of the elements are designated as $x_k (k = 1, \cdots, K)$, and $S_d(f)$ represents the signal arriving from the direction θ_d (see Fig.3).

First, using two types of complementary weight vectors of the element, $\boldsymbol{g} = [g_1, \cdots, g_K]$ and $\boldsymbol{h} = [h_1, \cdots, h_K]$ [6], we construct the signal spectra $S^{(g)}(f)$ and $S^{(h)}(f)$. The term, "complementary," implies one of the following conditions: "directivity pattern gain $|\boldsymbol{g}\boldsymbol{a}_d(f)| \gg$ directivity pattern gain $|\boldsymbol{h}\boldsymbol{a}_d(f)|$" or "$|\boldsymbol{g}\boldsymbol{a}_d(f)| \ll$ directivity pattern gain $|\boldsymbol{h}\boldsymbol{a}_d(f)|$" for an arbitrary direction, where $\boldsymbol{a}_d(f)$ is the steer vector given by

$$\boldsymbol{a}_d(f) \equiv [a_{1,d}(f), \cdots, a_{k,d}(f), \cdots, a_{K,d}(f)]^{\mathrm{T}}, \quad (1)$$

$$a_{k,d}(f) \equiv \exp[j2\pi f \cdot x_k \cdot \sin(\theta)_d / c]. \quad (2)$$

The exception is that the gain of both directivity patterns is unity with respect to the look direction θ_{d_0}. An example of directivity patterns using complementary beamforming is shown in Fig. 2, where the bold downnarrows indicate that the directional signal arrives from the corresponding direction θ_d.

Next, the sum of $S^{(g)}(f)$ and $S^{(h)}(f)$ is defined as the primary signal, $S^{(\mathrm{p})}(f)$, and the difference is defined as the

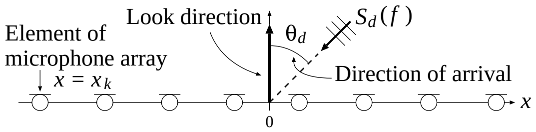

Figure 1: Configuration of a microphone array and acoustic signals.

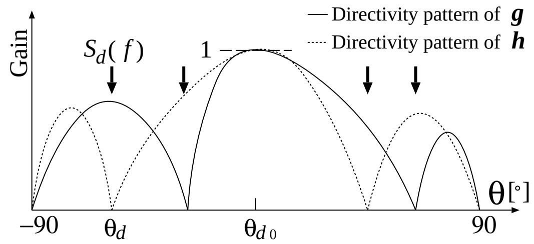

Figure 2: Example of directivity patterns using complementary beamforming.

reference signal, $S^{(\mathrm{r})}(f)$. These can be given as

$$S^{(\mathrm{p})}(f) = \sum_{d=1}^{D}\{\boldsymbol{g}\boldsymbol{a}_d(f)+\boldsymbol{h}\boldsymbol{a}_d(f)\}\cdot S_d(f), \tag{3}$$

$$S^{(\mathrm{r})}(f) = \sum_{d=1}^{D}\{\boldsymbol{g}\boldsymbol{a}_d(f)-\boldsymbol{h}\boldsymbol{a}_d(f)\}\cdot S_d(f). \tag{4}$$

Here, we use the minimized criterion proposed in Ref.[6] for the directional noise reduction. In this minimized criterion, a set of the frames used in the averaging process is regarded as a *block* in the time axis and this interframe-averaged power spectrum is designed to be $\langle|S^{(\mathrm{p})}(f)|^2\rangle_b$ or $\langle|S^{(\mathrm{r})}(f)|^2\rangle_b$, where the subscript b indicates that the interframe-averaged power spectra are obtained in the bth block.

Finally, applying this block-averaging technique, we calculate the array output in the bth block, $\left(|\hat{X}(f)|_b^2\right)^2$, over some blocks ($b=1,\ldots,B$), and define the squared sum of $\left(|\hat{X}(f)|_b^2\right)^2$ as a new minimized criterion. By assuming that the correlation among the signals is negligible in the averaged power spectra of each block, we can approximate the criterion as

$$\begin{aligned}&\sum_{b=1}^{B}\left(|\hat{X}(f)|_b^2\right)^2\\ &\equiv\sum_{b=1}^{B}(1/4)^2\cdot\left|\langle|S^{(\mathrm{p})}(f)|^2\rangle_b-\langle|S^{(\mathrm{r})}(f)|^2\rangle_b\right|^2\\ &\approx\sum_{b=1}^{B}\left|\sum_{d=1}^{D}\mathrm{Re}\big[\boldsymbol{g}\boldsymbol{a}_d(f)\cdot(\boldsymbol{h}\boldsymbol{a}_d(f))^*\big]\cdot\langle|S_d(f)|^2\rangle_b\right|^2. \quad (5)\end{aligned}$$

If each signal changes independently every block, Eq. (5) is minimized only when all directivity patterns $\mathrm{Re}[\boldsymbol{g}\boldsymbol{a}_d(f)\cdot(\boldsymbol{h}\boldsymbol{a}_d(f))^*]$ of each signal direction d are set to be zero. Thus, we can realize the directional nulls for each signal by minimizing $\Sigma_{b=1}^{B}\left(|\hat{X}(f)|_b^2\right)^2$ with respect to $\boldsymbol{g}$ and $\boldsymbol{h}$. Also,

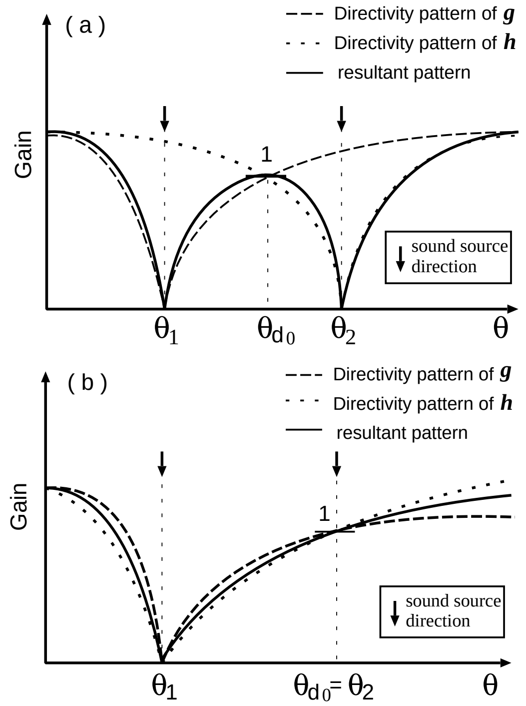

Figure 3: Example of directivity patterns.

to reduce the signal component in Eq. (5), it is not necessary to produce small $|\boldsymbol{g}\boldsymbol{a}_d(f)|$ and $|\boldsymbol{h}\boldsymbol{a}_d(f)|$ individually, but to design them so as to obtain small $\mathrm{Re}\big[\boldsymbol{g}\boldsymbol{a}_d(f)\cdot(\boldsymbol{h}\boldsymbol{a}_d(f))^*\big]$ in the directivity patterns. Taking advantage of this complementary characteristic, even when the directional nulls of $K-1$ directions are produced in each directivity pattern $|\boldsymbol{g}\boldsymbol{a}_d(f)|$ and $|\boldsymbol{h}\boldsymbol{a}_d(f)|$, we can realize the directional nulls of $2(K-1)$ directions in the directivity pattern of the proposed array.

More practically, $|\hat{X}(f)|_b^2$ is calculated using the observation signal vector $\boldsymbol{o}(f)$, i.e., this can be given as

$$\begin{aligned}|\hat{X}(f)|_b^2 &= (1/4)\cdot\big|(\boldsymbol{g}+\boldsymbol{h})\boldsymbol{R}_b(f)(\boldsymbol{g}+\boldsymbol{h})^{\mathrm{H}}\\ &\quad-(\boldsymbol{g}-\boldsymbol{h})\boldsymbol{R}_b(f)(\boldsymbol{g}-\boldsymbol{h})^{\mathrm{H}}\big|, \quad (6)\\ \boldsymbol{R}_b(f) &\equiv \langle\boldsymbol{o}(f)\boldsymbol{o}^{\mathrm{H}}(f)\rangle_b. \quad (7)\end{aligned}$$

Based on Eqs. (5) and (6), the following constrained minimization problem is solved.

$$\begin{aligned}\min_{\boldsymbol{g},\boldsymbol{h}}\quad &\sum_{b=1}^{B}\big|(\boldsymbol{g}+\boldsymbol{h})\boldsymbol{R}_b(f)(\boldsymbol{g}+\boldsymbol{h})^{\mathrm{H}}\\ &\quad-(\boldsymbol{g}-\boldsymbol{h})\boldsymbol{R}_b(f)(\boldsymbol{g}-\boldsymbol{h})^{\mathrm{H}}\big|^2 \quad (8)\\ \text{subject to}\quad &\boldsymbol{g}\boldsymbol{a}_{d_0}(f)=\boldsymbol{h}\boldsymbol{a}_{d_0}(f)=1 \quad (9)\end{aligned}$$

Equation (9) is the constraint in which the gain of both directivity patterns is unity with respect to the look direction.

2.2. DOA Estimation by Nonlinear Microphone Array

In solving the constrained minimization problem with respect to each look direction θ_{d_0}, we can classify the behavior of the array output into the following four cases.

1. In the case that each signal changes dependently every block,

   1(a) when the sound source does not exist in θ_{d_0}, since $\mathrm{Re}[\boldsymbol{ga}_d(f)\cdot(\boldsymbol{ha}_d(f))^*]$ may be both positive and negative, $\boldsymbol{g}$ and $\boldsymbol{h}$ are adjusted so as to cancel out $\mathrm{Re}[\boldsymbol{ga}_d(f)\cdot(\boldsymbol{ha}_d(f))^*]\cdot\langle|S_d(f)|^2\rangle_b$ for all θ_d, or produce directional nulls to reduce the gains for the DOAs of sound sources. Thus, the array output becomes quite small.

   1(b) when the sound source exists in θ_{d_0}, the array output becomes quite small for the same reason as described in 1(a).

2. In the case that each signal changes independently every block,

   2(a) when the sound source does not exist in θ_{d_0}, $\boldsymbol{g}$ and $\boldsymbol{h}$ are able to produce directional nulls which are steered to DOAs of sound sources complementarily to each other (see Fig. 3(a)). Thus, the array output becomes quite small.

   2(b) when the sound source exists in θ_{d_0}, since the improved optimizing algorithm based on a block-averaged power spectrum is used, the cancellation which is described in case 1(a) is restricted. Thus, the resultant directivity patterns are produced as depicted in Fig. 3(b), and Eq. (5) yields a relatively large value.

Based on the above example, since Eq. (5) is extremely minimized in every angle, we cannot detect the existence of signals if the signals change every block dependently of each other. However, we can detect the existence of signals by steering the look direction θ_{d_0} to every direction and minimizing Eq. (5) if the signals change every block independently of each other, i.e., cases 2(a) and 2(b). Therefore, the DOA of $2(K-1)$ signals can be estimated using only K microphones in this case.

In this study, the constrained minimization problem given by Eqs. (8) and (9) can be solved using the method of Lagrange multipliers[3].

3. EXPERIMENTS OF DOA ESTIMATION IN REAL ACOUSTIC ENVIRONMENT

3.1. Experimental Conditions

Acoustic signals were recorded under the condition that the reverberation time (RT) was 150 msec. A loudspeaker was placed at each of the left side ($\theta = -40°$) and the right side ($\theta = 30°$) of the array.

All sound data prepared in all experiments were sampled at 12 kHz with 16-bit resolution. Two- and three-element arrays with the interelement spacing of 4 cm were used. In both the conventional MV[4] method and the proposed method, we use sound samples of 2 sec duration to

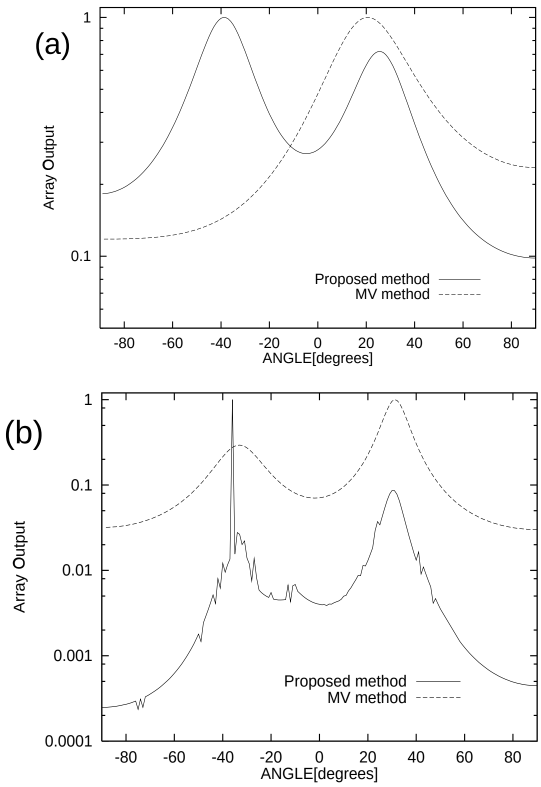

Figure 4: The experimental results in the case of (a) two- and (b) three-element array with 150 msec reverberation time.

perform the DOA estimation. In the proposed method, the optimization procedure is conducted under the following conditions: the frame length is 21.3 msec, the frame shift is half the frame length, the window function is rectangular, the block size is 50-frame lengths, and the block shift is 1-frame length. The analysis conditions in the MV method are the same as those in the proposed method with respect to frame length, frame shift and window function. The analysis frequency is set to be 3.0 kHz. We calculate and plot the resultant estimated DOA every 1° angle.

3.2. Experimental Results

First, the experimental results are shown in Fig. 4(a) for the case of a two-element array . In this figure, "Proposed" indicates the result of the proposed method, and "MV method" indicates that of the conventional MV method. This figure shows that the DOA-estimation result of the proposed method exhibits two peaks corresponding to the two sound sources, even when the conventional method fails. Thus, the DOA of $2(K-1)$ signals, where $K = 2$, can be estimated by the proposed method, and this results agrees with that derived theoretically in Sect.2.2. Next, the experimental results obtained with a three-microphone array are shown in Fig. 4(b). This figure shows that both the proposed method and MV method yield two peaks corresponding to the two sound sources because the number of microphones exceeds that of sound sources. It is, however, interesting that the

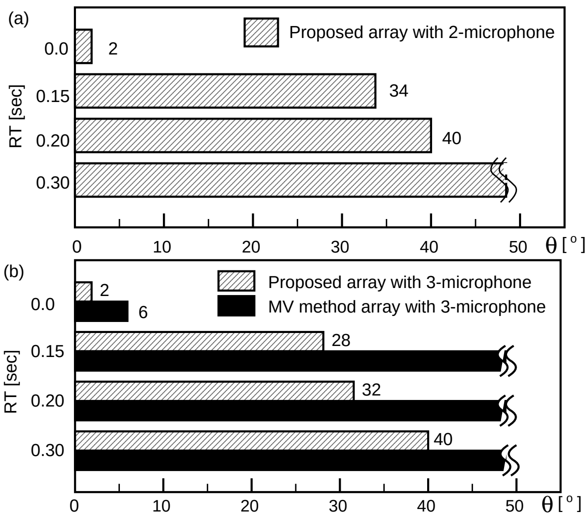

Figure 5: The directional resolutions of the proposed method and MV method with (a) two and (b) three microphones.

DOA resolution of the proposed method is different from that of the conventional method, so we will quantify the directional resolution in the next section.

4. DOA RESOLUTION MEASUREMENT

4.1. Definition of Directional Resolution

In this study, we use the directional resolution for the measurement of DOA-estimation quality, which is defined based on the following procedure. First, we set two loudspeakers with a narrow directional interval. Next, DOA estimations are performed for every additional two degrees with respect to the directional interval. Finally, when we obtain an accurate DOA of the loudspeakers, then the minimum directional interval is defined as the directional resolution of our experiments. In these measurements, RT is set to be 0, 150, 200, and 300 msec.

4.2. Experimental Results

Figure 5(a) shows the relation between the directional resolution of the proposed method and RT in the case of two microphones. We do not show the directional resolution of the MV method because the DOA estimation for two sound sources is inherently impossible with only two microphones[3]. Figure 5(b) shows the relation between the directional resolution of the proposed method and RT in the case of three microphones. We can summarize the results as follows.

- The proposed method can be used to distinguish two DOAs with a directional interval of at least 2° when two microphones are used and there is no reverberation. However, the directional resolution increases considerably, i.e., the accuracy of the DOA estimation deteriorates, as the RT increases.
- The directional resolution of the proposed method is superior to that of the conventional MV method under all reverberant conditions when three-element array are used. Also, the directional resolution of both the proposed method and the conventional method increases considerably as the RT increases.
- Increasing the number of elements in the proposed method prevents the deterioration of DOA estimation caused by reverberation.

5. CONCLUSIONS

In this paper, a new method for DOA estimation using the nonlinear microphone array based on complementary beamforming was described. To evaluate its effectiveness, DOA-estimation experiments were performed using actual devices in real acoustic environments. From the experiments of DOA estimation, comparing the conventional MV method and the proposed method, it was shown that the proposed method yields two peaks corresponding to the two sound sources with a two-element array, even when the conventional method fails. By comparing the resolutions of DOA estimation by the proposed method and by the conventional MV method, we could show that (1) the proposed method can distinguish two DOAs with a directional interval of at least 2° when there is no reverberation. However, the accuracy of the DOA estimation deteriorates as the RT increases; (2) the directional resolution of the proposed method is superior to that of the conventional MV method under reverberant conditions when a three-element array is used; (3) increasing in the number of elements in the proposed method prevents the deterioration of DOA estimation caused by reverberation.

6. ACKNOWLEDGEMENT

This work was partly supported by Grant-in-Aid for COE Research (No. 11CE2005) and CREST (Core Research for Evolutional Science and Technology) in Japan.

7. REFERENCES

[1] S. U. Pillai, "Array Signal Processing," Springer-Verlag Inc., 1989.

[2] G.W. Elko, "Microphone array systems for hands-free telecommunication," Speech Communication, vol.20, pp.229–240, 1996.

[3] D.H. Johnson and D.E. Dudgeon, "Array Signal Processing: Concepts and Techniques," Prentice-Hall, New Jersey, 1993.

[4] J. Capon, "High resolution frequency-wavenumber spectrum analysis," Proc. IEEE, vol.57, pp.1408–1418, 1969.

[5] H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, "Speech Enhancement Using Nonlinear Microphone Array with Complementary Beamforming," Proc. ICASSP'99, vol.1, pp.69–72, 1999.

[6] H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, "Speech Enhancement Using Nonlinear Microphone Array with Noise Adaptive Complementary Beamforming," Proc. ICASSP2000, vol.2, pp.1049–1052, 2000.