Acoust. Sci. & Tech. **26**, 4 (2005)

ACOUSTICAL LETTER

Direction of arrival estimation using matching pursuit under reverberant condition

Yasuhiro Oikawa* and Yoshio Yamasaki†

Global Information and Telecommunication Institute, Waseda University, 1011 Okuboyama, Nishi-Tomida, Honjo, 367–0035 Japan

(*Received 17 January 2005, Accepted for publication 10 March 2005*)

Keywords: DOA estimation, Matching pursuit algorithm, Delay-and-Sum array, DOA re-optimization
PACS number: 43.38.Hz, 43.60.Fg, 43.60.Jn [DOI: 10.1250/ast.26.365]

1. Introduction

Many topics on microphone array processing, e.g., signal enhancement, noise reduction and direction of arrival (DOA) estimation, have recently been researched, and many methods have been suggested for DOA estimation. The number of estimated DOAs is generally limited by the number of microphones and we have spatial aliasing problem on the basis of the relationship between microphone distance and wave length [1].

We have proposed a new DOA estimation method using a matching pursuit algorithm [2], which can estimate more DOAs than microphones [3]. First, we obtain the array output of the Delay-and-Sum method and calculate the average array output in the frequency range considered. Next, we estimate DOAs using a matching pursuit algorithm. In this study, we performed DOA estimation in a real world using this method and estimated more DOAs than microphones.

2. DOA estimation

Our method of DOA estimation consists of the following steps. We first separately calculate the normalized power of the array output, $P(\theta)$, for each frequency bin using the Delay-and-Sum method [4]. We then obtain the average $P(\theta)$ over all frequency bins. Finally, we perform peak picking using a matching pursuit algorithm to estimate the DOA over all frequency bands. The matching pursuit algorithm includes, after each iteration step, a re-optimization of all DOAs found thus far. Its main characteristics is that it is possible to find true DOAs when the number of sources exceeds that of microphones. We will discuss these steps in more detail in the following subsections.

2.1. Array output

For K sounds (i.e., K different DOAs) and two microphones, the observed signals are

$$\boldsymbol{X}(\omega, t) = \begin{bmatrix} \sum_{k=1}^{K} H_{1k}(\omega) S_k(\omega, t) \\ \sum_{k=1}^{K} H_{2k}(\omega) S_k(\omega, t) \end{bmatrix}, \tag{1}$$

where ω is the angular frequency, t is the time, $H_{1k}(\omega)$ and $H_{2k}(\omega)$ are the respective transfer functions between the kth sound and each microphone, and $S_k(\omega, t)$ is the kth sound. The average $\hat{P}(\theta, \omega)$, which is a component that depends on θ in the power of the Delay-and Sum array output $P(\theta, \omega)$, over all frequency bins is

$$\begin{aligned} \hat{P}_{\mathrm{avrg}}(\theta) &\approx \frac{1}{N} \sum_{i=1}^{N} \Bigg[\sum_{k=1}^{K} 2 \cdot E\big[|S_k(\omega_i)|^2\big] \\ &\qquad \cdot \mathrm{Re}\{H_{1k}(\omega_i) H_{2k}(\omega_i)^* \exp(-j\omega_i \tau)\} \Bigg] \\ &= \sum_{k=1}^{K} \hat{P}_{\mathrm{avrg}}(\theta|\theta_k), \end{aligned} \tag{2}$$

where $\tau = d \sin\theta / c$, d is being the distance between microphones and c the velocity of sound. N is the number of frequency bins and $\hat{P}_{\mathrm{avrg}}(\theta|\theta_k)$ the frequency-averaged θ-dependent component of the array output power from the kth sound, i.e., the θ-dependent component of the Delay-and-Sum array output is approximately the sum of those from different sources. As we are only interested in finding the DOAs at this point, we let $H_{1k}(\omega) = \exp(-j\omega\tau_{1k})$, $H_{2k}(\omega) = \exp(-j\omega\tau_{2k})$, and $S_k(\omega_i) = S_k$. Thus,

$$\begin{aligned} \hat{P}_{\mathrm{avrg}}(\theta|\theta_k) &= 2E[|S_k|^2]/N \\ &\quad \cdot \sum_{i=1}^{N} \mathrm{Re}\{\exp(-j\omega_i(\tau_{1k} - \tau_{2k} + \tau)), \end{aligned} \tag{3}$$

where $\tau_{1k} - \tau_{2k} = d \sin\theta_k / c$, θ_k is being the true direction of the kth sound position, and θ is the steering direction.

2.2. DOA estimation using matching pursuit

We used a modified matching pursuit algorithm that includes a re-optimization step [3] to decompose the signal into a set of direct and reflected sounds. We define the vector of the angles of i DOAs, which are estimated during i iterations, as

$$\Theta_i = [\hat{\theta}_1, \ldots, \hat{\theta}_i]^{\mathrm{T}}, \tag{4}$$

where Θ_0 is a vector without any elements. The matching pursuit algorithm for DOA estimation consists of the following steps:

Step:1) Define a dictionary as

$$\mathcal{D} = \{\hat{P}_{\mathrm{avrgn}}(\theta|\theta_k)\}_{-\pi/2 < \theta_k < \pi/2}, \tag{5}$$

i.e., an element of family $\mathcal{D}$ is defined as Eq. (3) normalized by its norm:

*e-mail: yoikawa@waseda.jp
†e-mail: y-yamasaki@waseda.jp

$$\hat{P}_{\mathrm{avrgn}}(\theta|\theta_k) = \frac{\hat{P}_{\mathrm{avrg}}(\theta|\theta_k)}{\sqrt{\frac{1}{\pi}\int_{-\pi/2}^{\pi/2} |\hat{P}_{\mathrm{avrg}}(\theta|\theta_k)|^2 d\theta}}. \quad (6)$$

Step:2) Initialization:

$$e_0(\theta) = \hat{P}_{\mathrm{observed}}(\theta) \quad \text{and} \quad (7)$$

$$i = 1. \quad (8)$$

Step:3) Calculate the residual for all θ_k values:

$$e_i(\theta|\theta_k) = e_{i-1}(\theta) - a_{i-1}(\theta_k)\hat{P}_{\mathrm{avrgn}}(\theta|\theta_k), \quad (9)$$

where $a_{i-1}(\theta_k)$ denotes the inner product of $e_{i-1}(\theta)$ and $\hat{P}_{\mathrm{avrgn}}(\theta|\theta_k)$.

Step:4) Select θ_k (estimate DOA $\hat{\theta}_i$):

$$\hat{\theta}_i = \underset{\theta_k}{\operatorname{argmin}} \sum |e_i(\theta|\theta_k)|^2. \quad (10)$$

Step:5) Re-optimize Θ_i (all DOAs) and calculate the residual $e_i(\theta)$:

$$e_i(\theta) = e_0(\theta) - \sum_{l=1}^{i} \hat{a}(\hat{\theta}_l)\hat{P}_{\mathrm{avrgn}}(\theta|\hat{\theta}_l), \quad (11)$$

where $\hat{a}(\hat{\theta}_l)$ is computed from Eq. (18).

Step:6) If

$$10 \log \frac{\int e_0{}^2(\theta) d\theta}{\int e_i{}^2(\theta) d\theta} < \delta, \quad (12)$$

where δ is the stopping criterion,

$$i = i + 1, \quad (13)$$

go to Step:3), or else end the procedure.

2.3. DOA re-optimization

We define the vector of the angles of L DOAs as

$$\Theta = [\theta_1, \cdots, \theta_L]^{\mathrm{T}}. \quad (14)$$

The basis vectors and the observed vector are defined as

$$\hat{\boldsymbol{p}}_{\mathrm{avrgn}}(\theta_k) = \left[\hat{P}_{\mathrm{avrgn}}\left(-\frac{\pi}{2}\middle|\theta_k\right), \cdots, \hat{P}_{\mathrm{avrgn}}\left(\frac{\pi}{2}\middle|\theta_k\right)\right]^{\mathrm{T}}, \quad (15)$$

$$\boldsymbol{e}_0 = \left[e_0\left(-\frac{\pi}{2}\right), \cdots, e_0\left(\frac{\pi}{2}\right)\right]^{\mathrm{T}}, \quad (16)$$

where we have discretized the normalized frequency-averaged θ-dependent component of the array output power from the kth sound as a function of continuous steering direction variable θ.

For a given set of DOAs, the analysis matrix containing the basis vectors is constructed according to

$$\hat{\boldsymbol{P}}_{\mathrm{avrgn}\Theta} = [\hat{\boldsymbol{p}}_{\mathrm{avrgn}}(\theta_1), \cdots, \hat{\boldsymbol{p}}_{\mathrm{avrgn}}(\theta_L)]. \quad (17)$$

The projection of $\boldsymbol{e}_0$ onto a space that is defined by bases $\hat{\boldsymbol{p}}_{\mathrm{avrgn}}(\theta_1), \cdots$, and $\hat{\boldsymbol{p}}_{\mathrm{avrgn}}(\theta_L)$ is

$$\hat{\boldsymbol{a}} = (\hat{\boldsymbol{P}}_{\mathrm{avrgn}\Theta}^{\mathrm{T}} \cdot \hat{\boldsymbol{P}}_{\mathrm{avrgn}\Theta})^{-1} \cdot \hat{\boldsymbol{P}}_{\mathrm{avrgn}\Theta}^{\mathrm{T}} \cdot \boldsymbol{e}_0, \quad (18)$$

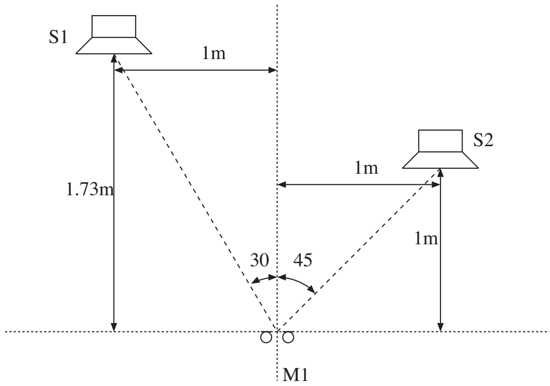

Fig. 1 Experimental conditions.

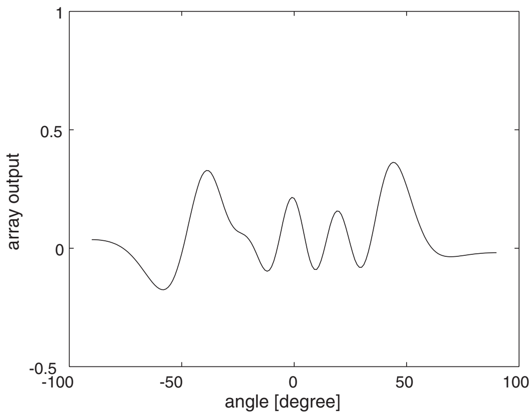

Fig. 2 Array output.

which is from the least-squares residual. Optimum DOAs are those for which the energy of the projection of $\boldsymbol{e}_0$ onto the column space of $\hat{\boldsymbol{P}}_{\mathrm{avrgn}\Theta}$ is maximized, that is,

$$\underset{\Theta}{\operatorname{argmax}}\{\hat{\boldsymbol{P}}_{\mathrm{avrgn}\Theta} \cdot \hat{\boldsymbol{a}}\}^{\mathrm{T}} \cdot \{\hat{\boldsymbol{P}}_{\mathrm{avrgn}\Theta} \cdot \hat{\boldsymbol{a}}\} = \underset{\Theta}{\operatorname{argmax}}\, \boldsymbol{e}_0^{\mathrm{T}} \boldsymbol{P}_{\Theta} \boldsymbol{e}_0. \quad (19)$$

3. Experiments

We estimated DOAs under real world experimental conditions, which were recorded in our office. The reverberation time was 0.94 s. The experimental conditions are shown in Fig. 1. The distance between microphones was 5 cm under these conditions. We used a data sampling frequency of 44.1 kHz, a frame length of 46 ms, and a frame update of 23 ms.

Figures 2–5 show the DOA estimates using the matching pursuit algorithm. The x-axis indicates the direction of arrival, and the y-axis the power of the arriving sound. Figure 2 shows the array output. Figures 3–5 show the first, second and third estimated DOAs, respectively. We also estimated the big initial reflected sound.

4. Conclusions

In this study, we estimated more DOAs than microphones using a matching pursuit algorithm under reverberant conditions. Source separation has recently been studied by many researchers to separate sources effectively under real world

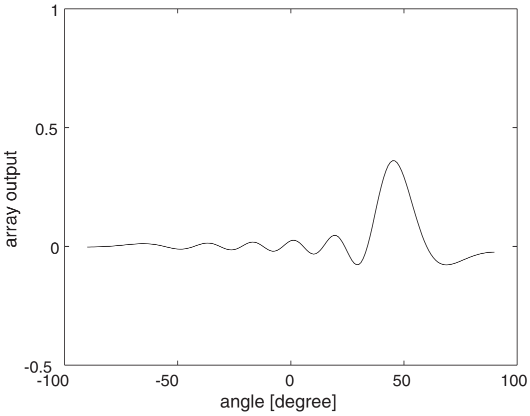

Fig. 3 First DOA.

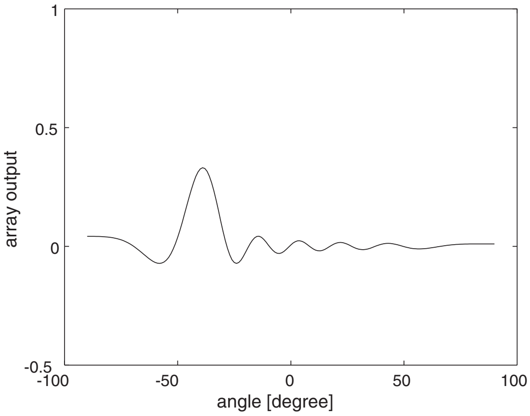

Fig. 4 Second DOA.

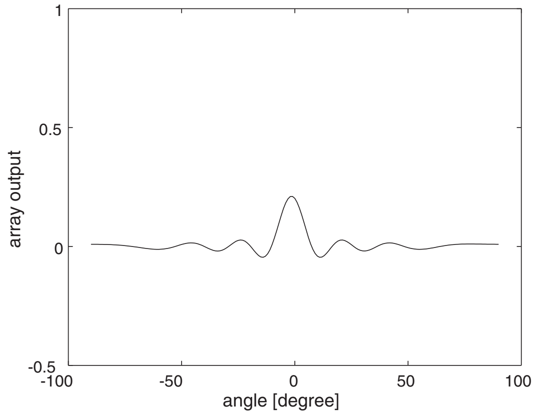

Fig. 5 Third DOA.

conditions. Thus, one of the research topics is to separate sources for convoluted mixtures consisting of long impulse responses in acoustics. We expect that additional improvement can be achieved in actual cases by considering the spatial information from this method.

References

[1] J. Ohga, Y. Yamasaki and Y. Kaneda, *Acoustic System and Digital Processing* (Korona Pub. Co., Ltd., Tokyo, 1995).

[2] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Process.*, **41**, 3397–3415 (1993).

[3] Y. Oikawa and Y. Yamasaki, "Direction of arrival estimates by two microphones using matching pursuit," *Proc. Autumn Meet. Acoust. Soc. Jpn.*, pp. 501–502 (2003).

[4] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques* (Prentice-Hall, Englewood Cliffs, 1993).