Directivity Steering Principle for Biomimicry Silicon Microphone

N. Ono[1], T. Arita[1], Y. Senjo[1], and S. Ando[1]
[1]Department of Information Physics and Computing, The University of Tokyo, Japan
e-mail: onono@alab.t.u-tokyo.ac.jp

ABSTRACT

In this paper, we propose a novel sound source separation principle particularly suitable for the biomimicry silicon microphone [1][2]. Conventionally, sound source separation has been performed by a wavelength-sized array of pressure-sensitive microphones. But it requires long observation interval and much computational cost due to the learning scheme of the algorithm, necessity for statistical properties of sound, subband decomposition or FFT, and so on. While, in our biomimicry microphone, tiny gimbal supported diaphragm detects sound pressure and its spatial gradients simultaneously. We show that their weighted summation generates null sensitivity zones invariant to the sound frequency, which enables us to exclude noise sources with high temporal resolution without distorting the target sound. We theoretically describe the principle and confirm the efficiency by some experiments.

Keywords: biomimicry, silicon microphone, directivity, sound separation, acoustic sensor

INTRODUCTION

Since a target signal like speech is often interfered by other speeches, various ambient noise, reverberation in real world, directivity control for sound source separation is one of the most concerned problems in the acoustic engineering. In conventional methods, the signals are obtained by an array of point pressure microphones and are combined in frequency domain. The frequency domain processing is inevitable since the weights for combination for the spread arrays are always frequency dependent. The use of modern signal processing methods like independent component analysis[3] also requires longer observation interval and additional knowledge on sounds due to learning adaptation, observation of statistical features, etc. All of them decrease temporal resolution and increase computational cost

Localization and separation of sound sources are also the important function for animals with auditory organs. They perform it instantaneously by only two ears. Among them, a parasitoid fly *Ormia ochracea* is known to have an excellent ability to localize the sound source in spite of its tiny tympanums[4]. In order to realize a novel directional microphone by mimicking the fly's ear, Tien *et al.* have fabricated a corrugated diaphragm by the DRIE bulk silicon micro-machining on a SOI wafer, and confirmed that it has a sensitivity to the gradient at one axis [5][6]. As reported in the last conference[1], we have extended it to two-dimensional structure, designed (Fig.1) and fabricated (Fig.2) a gimbal-supported circular diaphragm by Si surface micro-machining. The device detects the sound field $f(t)$ and the two-dimensional components of the gradients $f_x(t)$ and $f_y(t)$, simultaneously, by three orthogonal vibration modes. Although the first trial had problems with releasing the diaphragm, the second trial is now undergoing.

In this paper, based on the device, we describe a novel sound source separation system. Unlike conventional microphone array, the spatio-temporal gradient sensing by the biomimicry microphone generates null sensitivity zone with no frequency dependence, which enables us to realize high temporal resolution. We describe our principle theoretically, and confirm the efficiency by some experiments.

PRINCIPLE OF DIRECTIVITY STEERING

Let $f(t)$ be a sound pressure, $\dot{f}(t)$ be the time-differential, and $\nabla f(t) = (f_x(t), f_y(t), f_z(t))^t$ be the gradient vector. When $f(t)$, $f_x(t)$ and $f_y(t)$ are observed by the biomimicry microphone simultaneously, the weighted summation is represented as

$$s(t) = A(-uf(t) - u_t\dot{f}(t) + \boldsymbol{w} \cdot \nabla f(t)), \qquad (1)$$

where u and u_t are weights for $f(t)$ and $f_t(t)$, $\boldsymbol{w} = (w_x, w_y, 0)^t$ is an unit vector, and A represents a total gain. Note that the z component of $\boldsymbol{w}$ is fixed to 0 since $f_z(t)$ isn't observed.

To focus on the spatial directivity pattern of eq.(1), we here fix $A = 1$ with loss of generality. When an acoustic wave arrives from a point source at $\boldsymbol{r} = (x, y, z)$, a source signal $g(t)$ and the observed signals have the following relationships:

$$f(t) \quad = \quad g(t) \qquad (2)$$

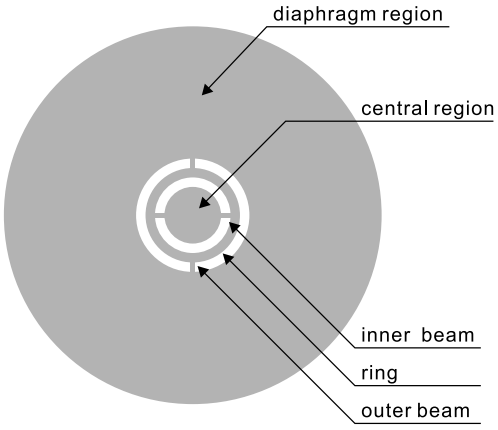

Figure 1: The design of a gimbal-supported circular diaphragm[1]: When the center region is fixed, torsions of narrow inner and outer beams permit the diaphragm region to incline to any direction.

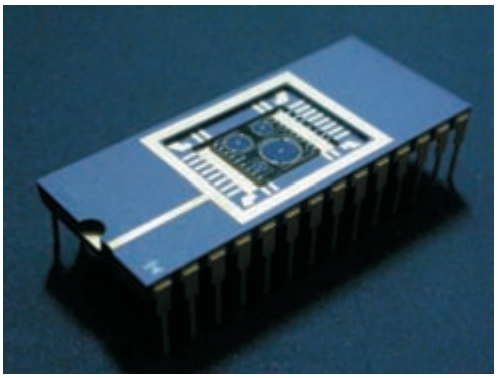

Figure 2: A photograph of a sensor mimicking a parasitoid fly's ear: gimbal-supported diaphragms are fabricated by Si surface micro-machining[1].

$$\dot{f}(t) = \dot{g}(t) \tag{3}$$

$$\nabla f(t) = \left(\frac{1}{|\boldsymbol{r}|}g(t) + \frac{1}{c}\dot{g}(t)\right)\frac{\boldsymbol{r}}{|\boldsymbol{r}|} \tag{4}$$

Then, $s(t)$ is represented as

$$s(t) = H(\boldsymbol{r})g(t) + H_t(\boldsymbol{r})\dot{g}(t), \tag{5}$$

where

$$H(\boldsymbol{r}) = \frac{\boldsymbol{r}\cdot\boldsymbol{w}}{|\boldsymbol{r}^2|} - u, \tag{6}$$

$$H_t(\boldsymbol{r}) = \frac{\boldsymbol{r}\cdot\boldsymbol{w}}{c|\boldsymbol{r}|} - u_t. \tag{7}$$

Hence, the weighted summation of spatio-temporal gradients consists of the sound source $g(t)$ and its time-differential $\dot{g}(t)$ with the gains $H(\boldsymbol{r})$ and $H_t(\boldsymbol{r})$, respectively. When $u+\alpha \neq 0$, the contour surface of $H(\boldsymbol{r}) = \alpha$ forms a sphere or a plane as

$$\left|\boldsymbol{r} - \frac{\boldsymbol{w}}{2(u+\alpha)}\right|^2 = \left|\frac{1}{2(u+\alpha)}\right|^2 \quad (u+\alpha \neq 0), \tag{8}$$

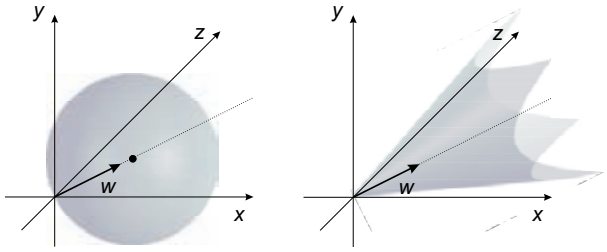

Figure 3: The examples of the contour shape of $H(\boldsymbol{r}) = \alpha$ (left) and $H_t(\boldsymbol{r}) = \alpha$ (right)

$$\boldsymbol{r}\cdot\boldsymbol{w} = 0 \qquad (u+\alpha = 0) \tag{9}$$

while the contour surface of $H_t(\boldsymbol{r}) = \alpha$ forms the corn

$$\cos\frac{\boldsymbol{r}\cdot\boldsymbol{w}}{|\boldsymbol{r}||\boldsymbol{w}|} = \cos\theta(\boldsymbol{r}) = c(u_t+\alpha) \tag{10}$$

where the $\theta(\boldsymbol{r})$ represents the angle between $\boldsymbol{r}$ and $\boldsymbol{w}$. From Eq.(6) and eq.(7), it is clear that the relationship between the directivity pattern and the weights as follows:

1. The directivity pattern of both $H(\boldsymbol{r})$ and $H_t(\boldsymbol{r})$ has a rotational symmetry and $\boldsymbol{w}$ determines the symmetry axis.
2. The diameter of the sphere $H(\boldsymbol{r}) = 0$ is given by $1/u$.
3. The vertical angle of the corn $H_t(\boldsymbol{r}) = 0$ is given by $\arccos(cu_t)$.

Examples of the contour shapes are shown in Fig.3.

Let $T(\boldsymbol{r},\omega)$ be a frequency response from a source at $\boldsymbol{r}$ to the weighted summation $s(t)$. When there exist N sound sources $G_n(\omega)$ at $\boldsymbol{r}_n (n = 1, 2, \cdots N)$. the Fourier transform of $s(t)$ is represented as

$$S(\omega) = \sum_{n=1}^{N} T(\boldsymbol{r}_n,\omega)G_n(\omega). \tag{11}$$

For sound source separation, the shape of null sensitivity zone $D_{null} = \{\boldsymbol{r} \mid T(\boldsymbol{r},\omega) = 0\}$ is the most important. When $\boldsymbol{r}_1 \notin D_{null}$ and $\boldsymbol{r}_n \in D_{null}$ for $n = 2, 3, \cdots, N$, the target signal can be recovered by compensating the frequency response like

$$G_1(\omega) = \frac{S(\omega)}{T(\boldsymbol{r}_1,\omega)}. \tag{12}$$

Note that $T(\boldsymbol{r}_1,\omega) \neq 0$ since $\boldsymbol{r}_1 \notin D_{null}$.

In conventional microphone array, D_{null} has frequency dependence, which makes subband decomposition or FFT essential for the separation. But in our case, since

$$T(\boldsymbol{r},\omega) = H(\boldsymbol{r}) + j\omega H_t(\boldsymbol{r}) \tag{13}$$

where $H(\boldsymbol{r})$ and $H_t(\boldsymbol{r})$ are real values, $T(\boldsymbol{r},\omega) = 0$ is equivalent to

$$H(\boldsymbol{r}) = 0 \quad \text{and} \quad H_t(\boldsymbol{r}) = 0. \tag{14}$$

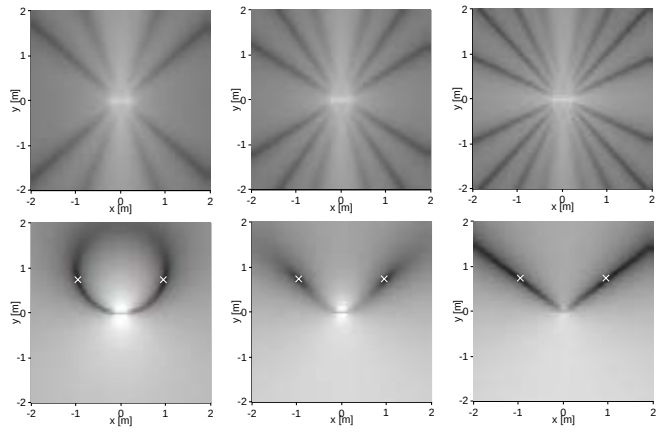

Figure 4: The examples of the distribution of $|T(\boldsymbol{r},\omega)|$. The upper figures represent a conventional case (a linear array of five microphones with 10cm interval and all weights are identical to 1.0) at 1800Hz(left), 2400Hz(center), and 3000Hz(right). The lower figures represent our case ($u = 0.6$, $cu_t = 0.5$, $w_x = 0$, $w_y = 1$) at 10Hz(left), 100Hz(center), and 1000Hz(right).

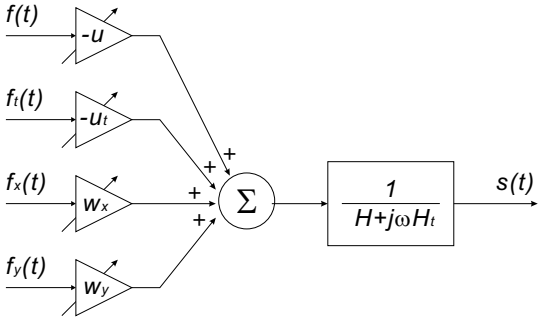

Figure 5: The sound separation scheme based on our biomimicry microphone

Namely, D_z has no frequency dependence in our system, which makes the separation very simple. The comparison of the distribution of $|T(\boldsymbol{r},\omega)|$ between a conventional microphone array and our system is shown in Fig.4. The conventional microphone array has frequency dependence, while in our system $|T(\boldsymbol{r},\omega)|$ is exactly equal to 0 at the location represented by white $\times$ marks for any ω. Therefore, the essential processing in our system is only 1)instantaneous weighted summation and 2)compensation filter (Fig.5).

SOUND SOURCE SEPARATION VIA NULL SENSITIVITY ZONE STEERING

In order to separate a target signal at a known position from other undesired signals, minimum variance (MV) beamformer is widely used in conventional microphone array processing[7], which automatically steer the null D_z to the noise direction by minimizing the variance (square power) of the weighted summation of observation signals under the gain constraint for the target. We show that the spatio-temporal gradients sensing by our biomimicry microphone makes the conventional MV method more useful and practical.

1) Optimal weight adaptation

Let $\boldsymbol{r}_1 = R_1\boldsymbol{n}_1$ ($R_1 = |\boldsymbol{r}_1|$, $\boldsymbol{n}_1 = \boldsymbol{r}/R_1$) be a known target position. In our scenario, the conventional MV method is deformed as follows: Minimize square power of the weighted summation

$$J =< s(t)^2 >= \int_\Gamma s(t)^2 dt \tag{15}$$

(Γ: observation interval) subject to the gain constraints for the target signal

$$H(\boldsymbol{r}_1) = \frac{1}{R_1}, \tag{16}$$

$$H_t(\boldsymbol{r}_1) = \frac{1}{c}. \tag{17}$$

More detailed analysis tells us that the above gain constants are optimum in the sense of the freedom of the null. By solving it, the optimal weights are obtained by

$$\begin{pmatrix} w_x \\ w_y \end{pmatrix} = \begin{pmatrix} B_{xx} & B_{xy} \\ B_{xy} & B_{yy} \end{pmatrix}^\dagger \begin{pmatrix} B_{ax} \\ B_{ay} \end{pmatrix}, \tag{18}$$

$$\begin{pmatrix} u \\ u_t \end{pmatrix} = (n_{1x}w_x + n_{1y}w_y - 1)\begin{pmatrix} 1/R_1 \\ 1/c \end{pmatrix}, \tag{19}$$

where $\dagger$ represents pseudo inverse and B_{ij} $(i,j = a,x,y)$ is defined as follows.

$$B_{ij} = \int_\Gamma b_i(t)b_j(t)dt \quad (i,j = a,x,y), \tag{20}$$

$$b_a(t) = \frac{1}{R_1}f(t) + \frac{1}{c}\dot{f}(t), \tag{21}$$

$$b_x(t) = f_x(t) - b_a(t)n_{1x}, \tag{22}$$

$$b_y(t) = f_y(t) - b_a(t)n_{1y}. \tag{23}$$

2) Compensation filter

Since the gains $H(\boldsymbol{r})$ and $H_t(\boldsymbol{r})$ for the target are fixed, the compensation filter is time-invariant although the weights u, u_t, w_x, w_y are time-varying. The frequency response of the filter is represented as

$$I(\omega) = \frac{1}{(1/R_1) + (1/c)j\omega}. \tag{24}$$

By using the bilinear transformation:

$$s = \frac{2}{T_s}\cdot\frac{1-z^{-1}}{1+z^{-1}}, \tag{25}$$

(T_s: the sampling period), it can be easily realized as the first order IIR filter represented by

$$I(z) = \frac{1+z^{-1}}{(1/R_1 + 2/cT_s) + (1/R_1 - 2/cT_s)z^{-1}}. \tag{26}$$

Note that the cut-off frequency ω_c is much less than the sampling frequency $2\pi/T_s$ in many applications (e.g. $\omega_c/2\pi = c/2\pi R_1 = 180$Hz when $R_1 = 30$cm, while the sampling frequency $1/T_s$ is usually more than 16kHz), the distortion due to the nonlinear warping of eq.(25) is negligible.

EXPERIMENTS AND RESULTS

Since the biomimicry microphone is under fabrication, we performed experiments by using four microphones (EM133; Primo) positioned at $(2.5, 0, 0)$cm, $(0, 2.5, 0)$cm, $(-2.5, 0, 0)$cm, $(0, -2.5, 0)$cm. The observation signals are simultaneously captured by A/D board (TD-BD-8CSUSB; Tokyo Electron Device). The spatial gradients were obtained from the difference between them. We placed a target source (male speech) and three noise sources (burst noise) at about 40cm distance from the microphones. The sampling rate was 44.1kHz, the observation time Γ was 2.5ms (110 samples), the analysis window was a Hanning window, and the frame shift was a half of Γ (55 samples). The source signals, the observation signal, and the results are shown in Fig.6. It is clear that the parameters are time-varying with high temporal resolution depending on the noise direction in each observation interval, and the noises are much suppressed in the separated signal. The S/N ratio improved by 9.76dB by the separation process. The results indicate that our system successfully separates the target signal with no frequency decomposition, no adaptation, and high temporal resolution in real environment.

REFERENCES

[1] N. Ono, A. Saito, and S. Ando, "Design and Experiments of Bio-mimicry Sound Source Localization Sensor with Gimbal-Supported Circular Diaphragm," Proc. Transducers'03, pp. 939-942, Boston, Jun. 2003

[2] N. Ono, A. Saito, and S. Ando, "Bio-mimicry Sound Source Localization with Gimbal Diaphragm," IEEJ, vol.123-E, no.3, pp.92-97, 2003

[3] A. Hyv̈arinen, J. Karhunen, and E. Oja, "Independent Component Analysis," Wiley, 2001.

[4] R. N. Miles, D. Robert, and R. R. Hoy, "Mechanically coupled ears for directional hearing in the parasitoid fly *Ormia ochracea*," J. Acoust. Soc. Am, vol.98, no.6, pp. 3059–3070, Dec. 1995

[5] C. Gibbons and R. N. Miles, "Design of a biomimetic directional microphone diaphragm," Proc. IMECE, Nov. 2000

[6] K. Yoo, J. -L. A. Yeh, N. C. Tien, C. Gibbons, Q. Su, W. Cui, and R. N. Miles, "Fabrication of a biomimetic corrugated polysilicon diaphragm with attached single crystal silicon proof masses," Proc. Transducers'01, 2001

[7] K. M. Buckley, "Spatial/spectral filtering with linearly-constrained minimum variance beamformers," IEEE Trans. Acoust., Speech., Signal Processing, vol.ASSP-35, pp.249-266, Mar. 1987

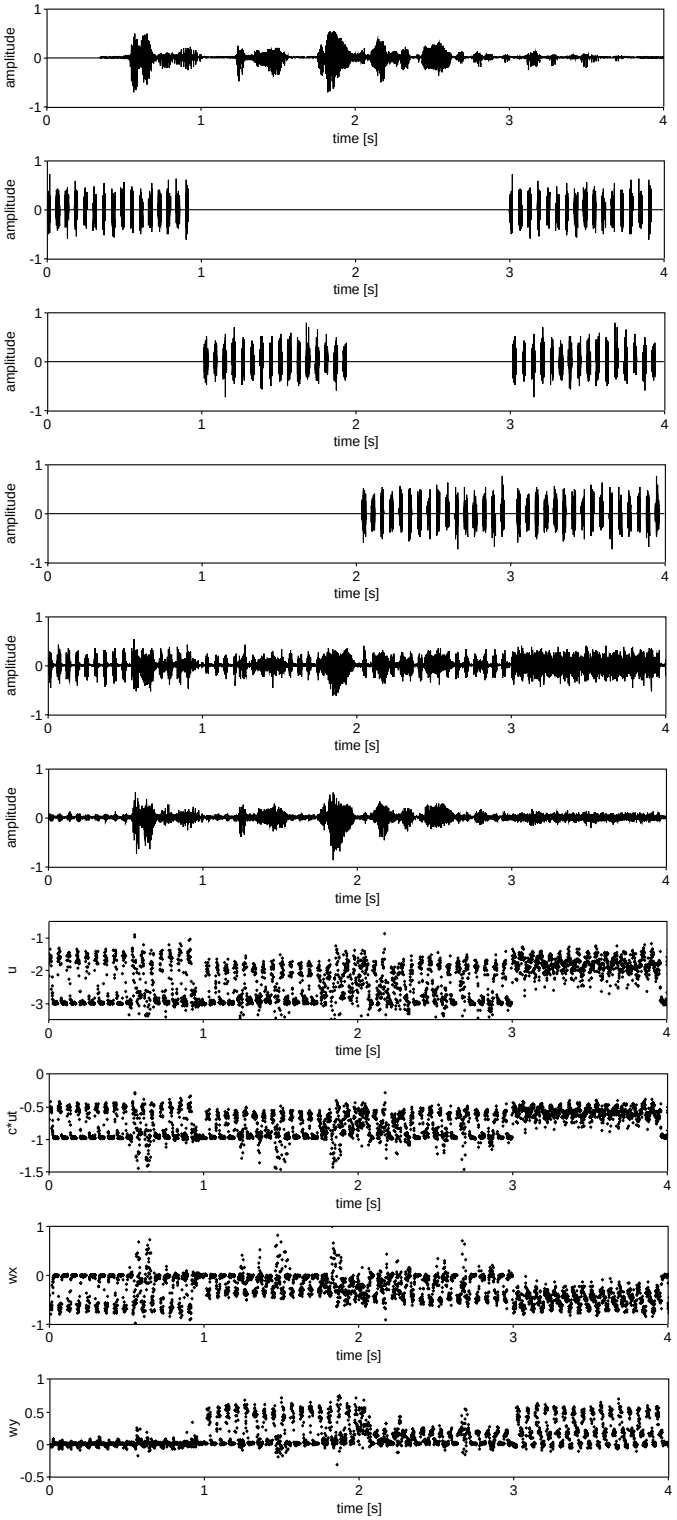

Figure 6: The experimental results of sound source separation: From top to bottom, the target signal, the burst noise 1, the burst noise 2, the burst noise 3, the mixed observation signal, the separated signal, the time-varying weight parameters u, cu_t, w_x and w_y.