同位置に置かれたマイクロホンによる音源定位手法

松本光春、橋本周司
早稲田大学理工学部応用物理学科

本論文では、音源定位に関する2つの手法について検討する。一つ目はマイクロホンアレイであり、二つ目はわれわれが提案している音源定位手法である。この手法は同位置に置かれたマイクロホンの伝達特性を用いるという特徴がある。まず、N 個のマイクロホンを利用した音響システムを定式化し、マイクロホンアレイと提案手法が同じ枠組みの中で扱い得ることを示す。
マイクロホンアレイは一般に2つのマイクロホンによって音源定位を行うことができない。しかし、提案手法では2つのマイクロホンによって音源定位を行うことが可能である。本論文では、その理論的な説明と実験結果について報告する。

SOUND LOCALIZATION BY AGGREGETED MICROPHONES

Mitsuharu Matsumoto, Shuji Hashimoto

Department of Applied Physics, Waseda University
Email:{matsu, shuji}@shalab.phys.waseda.ac.jp

In this paper, we discuss two methods for sound localization. The first one is the method using a microphone array. The second one is a method proposed by the authors to utilize the transfer characteristics of the microphones placed at the same position.

We formulate a sound system with N microphones at first. The microphone array and the proposed system can be described in the same framework, thus, simplify the theoretical analysis.

Microphone array cannot localize the sound by using only two microphones. However, it is possible to localize the sound with two microphones in the proposed method. This paper describes the theoretical foundation of the proposed methods together with the experimental results.

1. INTRODUCTION

Animals can localize sound based on the phase difference or the power differences sensed by their two ears. It can be considered that the ears are a microphone array system with two microphones. There exist researches on sound source localization inspired by living organisms [1][2]. However, it is known that it is difficult to localize the sound when the phase difference between the signal from the microphones exceeds π [rad]. It is also known that it is difficult to judge whether a sound comes from front or back direction using two ears.
In order to solve this problem, various sound localization methods using three or more microphones have been proposed in the past [3][4]. As the conventional direction presumption technique in a microphone array [5], the minimum variance method, the linear prediction method and the MUSIC method are known.
These methods use many microphones and it is necessary to set each microphone at a comparatively large distance [6]. As opposed to the existing method, a new sound system is proposed using the transfer characteristic of the microphones. Hanyu and Sekiguchi [7][8] proposed the method for the analysis of spatial information of sound field by using directional microphones at the same place and evaluate the method by simulations. The authors proposed the method for the sound localization and sound separation by using directional microphones at the same place and evaluate the method by real data.
In this paper, it is first shown that microphone array and the proposed system can be described in the same framework. The objective of the proposed method is to make the miniaturization easy by using not the difference of the position of microphones but the difference of the transfer characteristic of microphones. By using the transfer characteristics of the microphones, the microphones can be aggregated in one position, thus omitting the size constraint of the conventional microphone array system. A simple sound localization method is also proposed by using the transfer characteristics of two directional microphones. Finally, results of sound localization experiments conducted in an ordinary room are presented.

2. THEORY

2.1. Problem Formulation

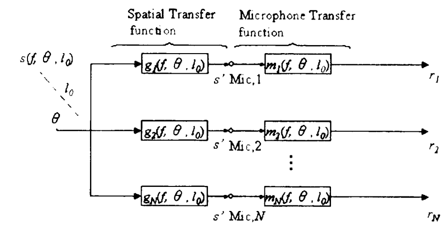


Fig. 1.Problem formulation.

A sound system with N microphones is shown in Fig.1.
In Fig.1, $g_i(f,\ \theta,\ l_0)$, i = 1,2,...,N, represents the spatial transfer characteristic from a sound source at the direction θ and the

distance l_0 to the i-th microphone where f and N represent frequency and the number of microphones, respectively.
$m_i(f, \theta)$ represents the transfer characteristic of the i-th microphone to the direction θ. When a sound is generated by the sound source at the direction θ and the distance l_0, the sound can be written as $s(f, \theta, l_0)$. $\boldsymbol{C}_i = (x_i, y_i)$ represents the coordinates of the i-th microphone. $\boldsymbol{C}_s = (x_s, y_s)$ represents the coordinates of the sound source.
According to the model in Fig.1, the output signal of the i-th microphone signal $r_i(f, \theta, l_0)$ can be expressed as
$r_i(f, \theta, l_0) = m_i(f, \theta) g_i(f, \theta, l_0) s(f, \theta, l_0) \quad (i = 1,2,...,N)$ (1)
The objective of sound localization is to calculate θ from $r_i(f, \theta, l_0)$ $(i = 1,2,...,N)$.

2.2. Restrictions of microphone array

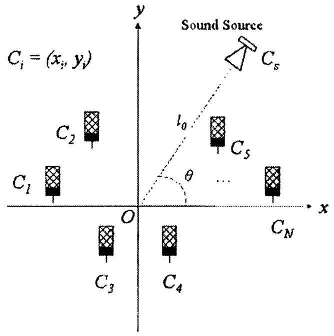


Fig. 2.General model of microphone array

N microphones are placed on x-y plane as shown in Fig.2. $\boldsymbol{O}$ represents the origin of the coordinate system. $g(f, l_0)$ represents the spatial transfer characteristic from a sound source $\boldsymbol{C}_s$ at distance l_0 from the origin $\boldsymbol{O}$. $g_i(f, \theta, l_0)$, $i = 1,2,...,N$, represents the spatial transfer characteristic from a sound source $\boldsymbol{C}_s$ to the coordinate of the i-th microphone $\boldsymbol{C}_i$. Generally the following conditions are satisfied in a microphone array system.
Condition 1) (Position differences)
$\forall i,j \quad \boldsymbol{C}_i \neq \boldsymbol{C}_j \qquad (i, j = 1, 2, ..., N, i \neq j)$ (2)
Condition 2) (Omnidirectional microphone)
$\forall i \forall \theta \ m_i(f, \theta) = C$: constant $(i = 1, 2, ..., N)$ (3)
Under these conditions, we can rewrite the problem formulation in equation (1) as follows.
ξ_i represents the difference of course from the i-th microphone to the coordinate origin. ξ_i is expressed as

$$\xi_i = \sqrt{x_s^2 + y_s^2} - \sqrt{x_i^2 + y_i^2} \tag{4}$$

And time delay τ_i can be expressed as

$$\tau_i = \frac{\sqrt{x_s^2 + y_s^2} - \sqrt{x_i^2 + y_i^2}}{v} \tag{5}$$

where v represents the velocity of the sound.
If the sound is a complex sinusoidal wave with the frequency f, $g_i(f, \theta, l_0)$ can be expressed as
$g_i(f, \theta, l_0) = \exp(-j2\pi f \tau_i) g(f, l_0)$ (6)
Then the received sound of the i-th microphone, $r_i(f, \theta, l_0)$ can be expressed as
$r_i(f, \theta, l_0) = \exp(-j2\pi f \tau_i)\, m_i(f, \theta) g(f, l_0) s(f, \theta, l_0)$ (7)

According to formula (3), $r_i(f, \theta, l_0)$ can be expressed as
$r_i(f, \theta, l_0) = C\exp(-j2\pi f \tau_i) g(f, l_0) s(f, \theta, l_0)$ (8)
According to formula (5), $\exp(-j2\pi f \tau_i)$ is independent of the spatial transfer characteristic $g(f, l_0)$ and the sound source $s(f, \theta, l_0)$. It depends on the relation between the position of the sound source and the position of the i-th microphone. $g(f, l_0)$ and $s(f, \theta, l_0)$ are common to all microphones. The characteristics of $\exp(-j2\pi f \tau_i)$ is controllable by change the position of the i-th microphone.
Microphone array realizes the sound focusing and the sound localization using the relation between the position of the sound source and the position of each microphone.

2.3. Restrictions of aggregated microphone method

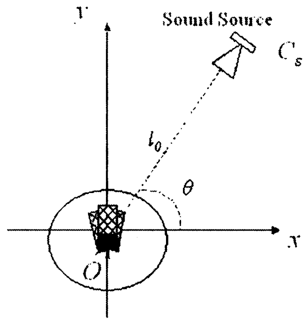


Fig. 3.General model of the proposed method

In the proposed method, N small microphones are located at the origin $\boldsymbol{O}$ with slightly rotated postures with regard to each other as shown in Fig.3. $g(f, l_0)$, $i = 1,2,...,N$, represents the spatial transfer characteristic from a sound source $\boldsymbol{C}_s$ to the origin $\boldsymbol{O}$. The size of the prepared microphone system is significantly small compared with the wavelength of sound because N microphones are located at the same position.Then $r_i(f, \theta, l_0)$ can be expressed as
$r_i(f, \theta, l_0) = m_i(f, \theta) g(f, l_0) s(f, \theta, l_0)$ (9)
In microphone array, if all microphones are located at the same place in microphone array, then
$\forall i,\ g_i(f, \theta, l_0) = const$
$\forall i,\ m_i(f, \theta, l_0) = const$
and $r_i(f, \theta, l_0)$ in the Equation (1) is identical for $i = 1,2,...,N$.
Hence, in microphone array, the sound system cannot be realized if all microphones are located at the same place.
However, if we use the directional microphones and regard $m_i(f, \theta)$ as the complex function, we can realize the sound system similar to the microphone array.
Comparing equation (8) and equation (9), $C\exp(-j2\pi f \tau_i)$ can be considered as substitution for $m_i(f, \theta)$ in Equation (9), thus the proposed method can be written in the same theoretical framework as in the conventional microphone array system. The proposed method realizes the sound focusing and the sound localization using the difference of the transfer characteristic of each microphone. It is necessary for this method to satisfy the following conditions.

Condition 1) (Observability)

$$\forall\theta \quad \exists m_i(f,\theta) \neq 0 \qquad (i = 1, 2, ..., N) \tag{10}$$

Condition 2) (Independency of the direction)

$$\forall\theta_i, \forall\theta_j \ (\theta_i \neq \theta_j)$$

$$(m_1(f,\theta_i), m_2(f,\theta_i), \dots, m_N(f,\theta_i)) \neq \alpha(m_1(f,\theta_j), m_2(f,\theta_j), \dots, m_N(f,\theta_j)) \tag{11}$$

2.4.The method of sound localization with two microphones

In this section, we introduce a new sound localization method that utilizes the transfer characteristics of the microphones.

Let $m_E(f,\theta)$ be a relative transfer characteristic defined as follows:

$$m_E(f,\theta) = \frac{m_1(f,\theta)}{m_2(f,\theta)} \tag{12}$$

According to formula (10), $m_E(f,\theta)$ can be defined for all θ values. $m_E(f,\theta)$ can be measured using $r_1(f,\theta,l_0)$ and $r_2(f,\theta,l_0)$ in advance as follows:

$$\begin{aligned} m_E(f,\theta) &= \frac{r_1(f,\theta,l_0)}{r_2(f,\theta,l_0)} \\ &= \frac{m_1(f,\theta)g_1(f,\theta,l_0)s(f,\theta,l_0)}{m_2(f,\theta)g_2(f,\theta,l_0)s(f,\theta,l_0)} \\ &= \frac{m_1(f,\theta)g(f,l_0)s(f,\theta,l_0)}{m_2(f,\theta)g(f,l_0)s(f,\theta,l_0)} \\ &= \frac{m_1(f,\theta)}{m_2(f,\theta)} \end{aligned} \tag{13}$$

It should be noted that $m_E(f,\theta)$ is not dependent on $g(f,\theta)$ and $s(f,\theta,l_0)$.

Suppose the positions (θ_s, l_s) of the sound source are unknown. We introduce the evaluation function, $M_E(f,\theta)$ defined as follows:

$$M_E(f,\theta) = \left| m_E(f,\theta) - \frac{r_1(f,\theta_s,l_s)}{r_2(f,\theta_s,l_s)} \right| \tag{14}$$

According to formula (11), the following relations stands.

1) $\forall\theta \neq \theta_s, M_E(f,\theta) \neq 0$ (15)

2) $\forall\theta = \theta_s, M_E(f,\theta) = 0$ (16)

$M_E(f,\theta)$ becomes minimum when θ is equal to θ_s. Therefore the sound localization is realized as follows.

1) Calculate $M_E(f,\theta)$ for all θ

2) Find θ when $M_E(f,\theta)$ is the minimum.

3. EXPERIMENTAL RESULT

Fig.4 shows the block diagram of the experimental sound localization system. Several experiments were conducted to evaluate the performance of the sound localization system in the case that N=2 as shown in Fig.5. We used two directional capacitor microphones, RP-VC200 made by Panasonic. Each microphone was connected to the computer through the amplifier and A/D converter. We used AT-MA2 made by Audio-technica as the amplifiers and DAQCard-AI-16E-4 made by National Instruments as the A/D converter. All the amplifier gains were unified into +20dB. Above process was executed in real time for the sounds sampled with the sample frequency 22050 Hz and 16bit resolution. The window size of FFT was 1024/22050 [ms].

All experiments were conducted in a room with a reverberation time of 212 [ms] and with length 427 [cm], width 345 [cm] and height 240 [cm].

The following sounds are used as the sound source.

1) The continuous sinusoidal waves with 500Hz, 1000Hz, 2000Hz, 4000Hz and 8000Hz

2) Band noise from upto 10000Hz

The sound is sounded from the distance l_s = 1[m].

$m_E(f,\theta)$ is measured as follows.

1) Acquire $r_1(f,\theta,l_0)$ and $r_2(f,\theta,l_0)$ for θ in the range of 0[deg] to 350[deg] with 10[deg] interval.

2) Calculate $m_E(f,\theta)$ from $r_1(f,\theta,l_0)$ and $r_2(f,\theta,l_0)$ in advance.

The calculated $m_E(f,\theta)$ of various θ are then used as reference.

The sound localization is realized as follows.

1) Acquire another $r_1(f,\theta_s,l_s)$ and $r_2(f,\theta_s,l_s)$.

2) Calculate $M_E(f,\theta)$ for all θ.

3) Find θ such as $M_E(f,\theta)$ is the minimum.

Fig.6 shows the directional patterns of two microphones. The directional patterns are expressed as the same direction used in the practical experiment.

Fig.7 shows the output of the evaluation function for the continuous sinusoidal wave with 500Hz, 1000Hz, 2000Hz, 4000Hz and 8000Hz when the sound was generated from 90 [deg] and 180 [deg], respectively.

$ORM(f,\theta)$, the output ratio of $M_E(f,\theta)$, was calculated with dB value on the basis of the minimum value as follow.

$$ORM(f,\theta)[dB] = 10\log\left(\frac{M_E(f,\theta)}{\min(M_E(f,\theta))}\right) \tag{17}$$

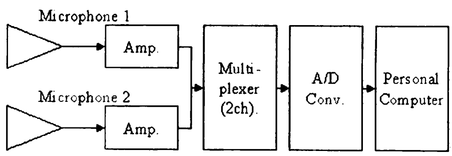


Fig. 4. Block diagram of the signal processing

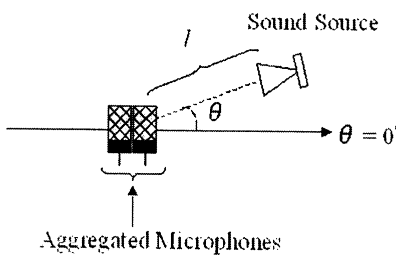
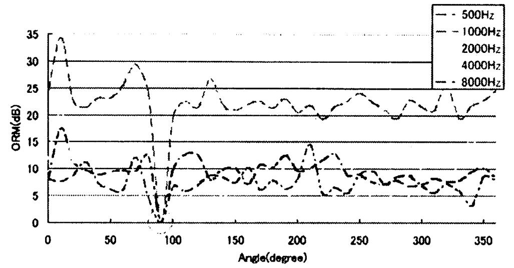
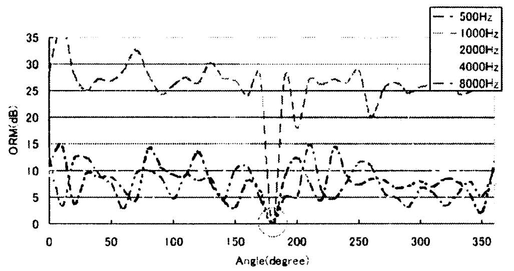


Fig. 5. Experimental environment



(1). The output of evaluation function when continuous sinusoidal wave was sounded from 90 degree.



(2). The output of evaluation function when continuous sinusoidal wave was sounded from 180 degree.

Fig. 6. The output of evaluation function related to the sinusoidal wave with 500Hz, 1000Hz, 2000Hz, 4000Hz and 8000Hz

4. CONCLUSION

In this paper, we proposed a sound localization system that uses the transfer characteristics of microphones located at the same place. It was shown that microphone array and the proposed system can be described in the same framework.

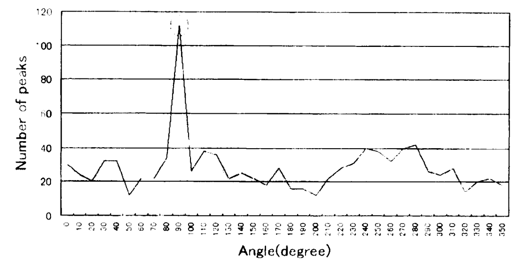
It was also shown that the microphone array and the method which uses the directivity of the microphones cannot localize the sound by using only two microphones. It is proved that the proposed system can localize the sound by using only two microphones.

The experiments were carried out in an ordinary room. Although the accuracy of the sound localization is decreased by the echo and noise in the room, the obtained results were satisfactory for practical application.

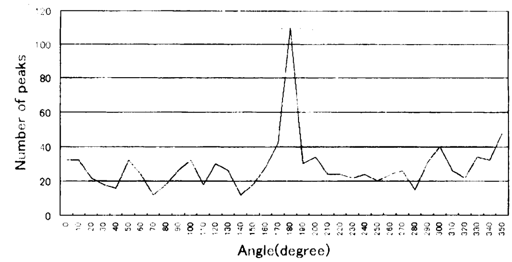
A future study will be to design an aggregated microphone chip for the proposed algorithm.

ACKNOWLEDGEMENTS

A part of this research was done by research support of the 21st century COE program "The innovative research on symbiosis technologies for human and robots in the elderly dominated society"



(1). The peak histogram with regard to various frequencies when white noise was sounded from 90 degree.



(2). The peak histogram with regard to various frequencies when white noise was sounded from 180 degree.

Fig. 7. The peak histogram with regard to various frequencies when white noise is generated from 90[deg] and 180[deg]

5. REFERENCES

[1] J. Huang, N. Ohnishi, N. Sugie: "Research on a sound source localization system suggested by living organisms - Localization of a single sound source in an anechoic", IEICE trans, Vol.J71-A-II, No.10, pp.1780-1789 (1988)

[2] J. Huang, N. Ohnish and N.Sugie, "Spatial localization of sound sources:azimuth and elevation estimation" Proc. of IEEE IMTC, pp.330-333,(1998).

[3] K. Sasaki, K Hirata:"3D-localization of a stationary random acoustic source in near-field by using 3 point-detectors", Trans of SICE, Vol.34, No.10, pp1329-1337 (1998)

[4] Y. Yamasaki and T. Itow "Measurement of spatial information in sound fields by closely located four point microphone method", J. of Acoustic Society Japan, 10-2, 101-110, (1990)

[5] S. U. Pillai, "Array signal processing", New York, Springer-Verlag,1989

[6] H.F.Silverman, "An algorithm for determining talker location using a linear microphone array and optical hyperbolic fit" proc. of speech and natural language workshop, pp.151-156 (1990)

[7] T.Hanyu K. Sekiguchi, "A New Algorithm for Directional Microphone System based on Spatial Deconvolution", Proceedings of ICA2004 (The 18th International Congress on Acoustics, Kyoto, April 2004), vol. IV, pp. 2759-2762 (2004)

[8] T.Hanyu, K. Sekiguchi, "Directional Microphone System based on Spatial Deconvolution for Analysis of Spatial Information of Sound Field" RADS2004 (2004)