社団法人 情報処理学会 研究報告
IPSJ SIG Technical Report
2004－MPS－48 (2)
2004／3／1

同位置に置かれたマイクロホンの伝達特性を利用した音源定位手法

松本光春、橋本周司
早稲田大学理工学部応用物理学科

本論文では同位置に置かれたマイクロホンを用いた新しい音源定位手法(AMSOL)を提案する。マイクロホンアレイを用いた多くの音源定位手法ではマイクロホンを異なった場所に置き、その位置関係の違いを利用して音源定位を行う。本手法はマイクロホンを同位置に置き、マイクロホンの伝達特性を利用することで外部変数によらない音源定位を実現する。
AMSOL は次の３つの特徴を持つ。
1) 最低二つのマイクロホンで音源定位が可能である。
2) 事前に既知となる内部変数にのみ依存し、外部環境から独立である。
3) システムの小型化が容易である
本論文では、AMSOL の理論的な枠組みについて説明し、実験を通してその性能を評価する

AGGREGATED MICROPHONES METHOD FOR SOUND LOCALIZATION

Mitsuharu Matsumoto, Shuji Hashimoto

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

This paper introduces a new *a*ggregated *m*icrophones method for *so*und *l*ocalization (AMSOL). Most sound localizations are executed by solving the inverse matrix of the spatial transfer matrix from the sound sources to the multiple microphones located at different places. Contrastingly, AMSOL executes the localization by solving the inverse of the internal transfer matrix of the different microphones located at one particular place. The method is simpler than the existing method because the internal transfer function of each microphone can be known in advance.
The AMSOL has the following three features.
1) With at least 2 microphones of which frequency response with respect to the sound direction is known, the proposed system will be able to execute sound localization tasks.
2) The performance of the proposed system is independent of the external environment. It depends only on the internal parameters that are known in advance.
3) It is easy to miniaturize the system setup.
In this paper, we described the theoretical background of AMSOL. We also show the performance of sound localization using AMSOL, through some experiments.

1. INTRODUCTION

In an unregulated real world environment, the directions of sound sources are generally unknown, making the task of sound localization a very difficult one. Therefore, a sound localization method that could be applied in a realistic setting is of interest in many fields.

The microphone array is a well-known method for sound localization [1][2][3]. Most of these methods utilize the phase difference of the sound received by a number of microphone located at different places. Hence, the distances between the microphones have to be set considering the wave length of the sounds. This constraint makes the size of microphone array systems large, to limit the applications.

In this paper, a new sound localization method called "AMSOL" (*a*ggregated *m*icrophones method for *so*und *l*ocalization) is presented. A mathematical formulation for sound localization utilizing microphone array is explained in Section 2. In Section 3, we explain the method of designing the proposed AMSOL based on the formulation. In Section 4, the results of localization experiments using AMSOL are presented.

2. THEORY

2.1. PROBLEM FORMULATION

A sound localization filtering scheme with a microphone array is shown in Figure.1.

In Figure.1, $g_i(f, \theta_s)$, $i = 1,2,...,M$, represents the spatial transfer characteristic from a sound source at the direction θ_s to the i-th microphone where f and M represent frequency and the number of microphones, respectively.

$m_i(f, \theta_s)$ represents the transfer characteristic of the i-th microphone to the direction θ_s. $\boldsymbol{h}$ represents a N-input 1-output sound localization filter. When a sound is generated by the sound source at the direction θ_s, the sound can be written as $s(f, \theta_s)$. According to the model in Figure.1, the output signal vector of microphones signal $\boldsymbol{r}$ can be expressed as

$$\boldsymbol{r} = \boldsymbol{mg}s \qquad (1)$$

where

$$\boldsymbol{r} = [r_1(f, \theta_s), r_2(f, \theta_s), \ \dots \ r_M(f, \theta_s)]^t \qquad (2)$$

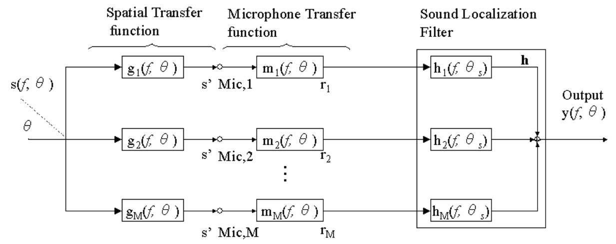

Figure.1.Block Diagram of sound localization system

$$\boldsymbol{mg} = \begin{bmatrix} m_1(f,\theta_s)g_1(f,\theta_s) \\ m_2(f,\theta_s)g_2(f,\theta_s) \\ \\ m_M(f,\theta_s)g_M(f,\theta_s) \end{bmatrix} \quad (3)$$

t: transposed matrix

mg is a vector decided by $g_i(f, \theta_s)$, the spatial transfer characteristic from the sound source to the *i*-th microphone, and $m_i(f, \theta_s)$, the transfer characteristic of the *i*-th microphone in the direction θ_s. In Figure.1, $\boldsymbol{h}(f, \theta_s)$ (= [$h_1(f, \theta_s)$, $h_2(f, \theta_s)$, … $h_M(f, \theta_s)$]) express a sound localization filter to detect the sound from the direction θ_s. When the sound is played from the direction of θ, the frequency response $y(f, \theta)$ of this microphone array system to the desired signal can be expressed as,

$$y(f, \theta) = \boldsymbol{h}(f, \theta_s)\boldsymbol{m}(f, \theta)\, g(f, \theta)s(f, \theta) \quad (4)$$

It is possible to localize the direction θ_s of the sound if the following condition is satisfied for every assumed direction θ_s.

$$y(f,\theta) = \begin{cases} 0(\theta = \theta_s) \\ nonzero(\theta \neq \theta_s) \end{cases} \quad (5)$$

The sound localization characteristic formed by this system is determined only by the characteristic of the sound localization vector $\boldsymbol{h}(f, \theta_s)$.
Hence the sound localization problem can be simplified to a problem of designing $\boldsymbol{h}(f, \theta_s)$ as the condition (5) is satisfied.

2.2.AMSOL SIGNAL PROCESSING

We quantize the sound position as θ_1, …, θ_N in the following discussion, which is practically acceptable as we can use sufficiently large *N*.

Let θ_i, *i* = 1,2, …, *N*, be the *i*-th direction under consideration. Let $\boldsymbol{m}(f, \theta_i)$, *i* = 1,2, …, *N*, as the transfer characteristic of microphones for the direction of θ_i where

$$\boldsymbol{m}(f, \theta_i) = [m_1(f, \theta_i), m_2(f, \theta_i), \ \ , m_M(f, \theta_i)]^t \quad (6)$$

The following two conditions are assumed in this method.
Assumption 1) *M* microphone elements are aggregated. "Aggregated" means that the microphones are intentionally arranged in the same place.
Assumption 2) The transfer characteristic $m_j(f, \theta_i)$ of the *j*-th microphone to the sound from the direction θ_i (*i* = 1,2,…,*N*) can be investigated in advance. According to Assumption 1), the spatial transfer characteristic from a sound source to each microphone is identical to be written as

$$g_j(f, \theta_i) = g(f, \theta_i) \quad \forall i, j\ (i = 1,2, …,N, j = 1,2, ...,M) \quad (7)$$

under the assumption 1)

Then formula (3) can be expressed as

$$\boldsymbol{m}(f,\theta_i)\boldsymbol{g} = \begin{bmatrix} m_1(f,\theta_i)g(f,\theta_i) \\ m_2(f,\theta_i)g(f,\theta_i) \\ \\ m_M(f,\theta_i)g(f,\theta_i) \end{bmatrix} \quad (8)$$

Therefore, if the new sound source *s'* is defined as

$$s'(f, \theta_i) = g(f, \theta_i)s(f, \theta_i) \quad (9)$$

then ***r***, the received sound, can be expressed as

$$\boldsymbol{r}(f, \theta_i) = \boldsymbol{m}(f, \theta_i)s'(f, \theta_i) \quad (10)$$

where

$$\boldsymbol{m}(f, \theta_i) = [m_1(f, \theta_i), m_2(f, \theta_i), \ … \ m_M(f, \theta_i)]^t \quad (11)$$

m: Microphone Transfer Characteristic vector, which is decided only by the transfer characteristic of the microphones.

According to formula (4) and (10), when the sound is played from the direction of θ_j, the frequency response $y_i(f, \theta_j)$ of AMSOL can be written as

$$y_i(f, \theta_j) = \boldsymbol{h}(f, \theta_i)\boldsymbol{m}(f, \theta_j)s'(f, \theta_j) \quad (12)$$

It is possible to localize the direction θ_j of the sound if the following condition is satisfied for every assumed direction θ_i.

$$y_i(f,\theta_j) = \begin{cases} 0(\theta_j = \theta_i) \\ nonzero(\theta_j \neq \theta_i) \end{cases} \quad (13)$$

Hence a sound localization problem can result in the problem which sets the microphones and designs $\boldsymbol{h}(f, \theta_i)$ as the condition (13) is satisfied.

It should be noted that $\boldsymbol{m}(f, \theta_i)$ is independent of the external environment. It is decided only by the design of microphones.

3. THE DESIGN METHOD OF AMSOL

3.1. REQUIRED CONDITION

Sound localization vector that correspond to $\boldsymbol{m}(f, \theta_i)$ is expressed as $\boldsymbol{h}(f, \theta_i)$. ***M***, microphone characteristic matrix, and ***H***, sound localization matrix, are made as follows.

$$\boldsymbol{M} = [\boldsymbol{m}(f, \theta_1), \boldsymbol{m}(f, \theta_2), …, \boldsymbol{m}(f, \theta_N)] \quad (14)$$

$$\boldsymbol{H} = [\boldsymbol{h}(f, \theta_1), \boldsymbol{h}(f, \theta_2), …, \boldsymbol{h}(f, \theta_N)]^t \quad (15)$$

As is mentioned, it is possible to localize the sound when the condition (13) is satisfied, that is

$^\exists\boldsymbol{H}$, s.t.

$$\boldsymbol{T} = |t_{ij}| = \boldsymbol{HM} = \begin{bmatrix} 0 & & \\ & 0 & nonzero \\ & nonzero & \\ & & 0 \end{bmatrix} \quad (16)$$

where t_{ij} is the element of ***T***.
What is necessary for satisfying the condition (16) is to prepare the microphones, which satisfy

$$\text{Rank } \boldsymbol{M}^t = N \quad (17)$$

t: transposed matrix
The condition (17) means that $\boldsymbol{m}(f, \theta_i)$ is linearly independent for θ_i, (*i = 1,2, …,N*). It is proved below.

3.2. MATHEMATICAL PROOF

Theorem 1)

If

Rank $\boldsymbol{M}^t = N$

then

$^\exists \boldsymbol{H'}$, s.t. ($\boldsymbol{T'}=\boldsymbol{H'M} \wedge \boldsymbol{T'}$ is nonsingular) (18)

Proof)

Let $\boldsymbol{T'}$ as

$\boldsymbol{T'}=|t'_{ij}| =\boldsymbol{H'M}$ (19)

where t'_{ij} is the element of $\boldsymbol{T'}$.

And choose $\boldsymbol{H'}$ where

Rank $\boldsymbol{H'} = N$ (20)

Then

$$t'_{ik} = \sum_l h_{il} m_{lk} \quad (21)$$

$$t'_{jk} = \sum_l h_{jl} m_{lk} \quad (22)$$

$^\forall i, j, k$ (i, j, k = 1,2, …, N)

where t'_{ik} is the i-th row and the k-th colomn element of $\boldsymbol{T'}$, and t'_{jk} is the j-th row and the k-th colomn element of $\boldsymbol{T'}$.

If t'_{ik} and t'_{jk} are linearly independent, then $^\exists\boldsymbol{T'}$. It is proved by the reduction to absurdity.

It is assumed that the i-th row and j-th row are linearly dependent. Then

$^\exists\alpha$, s.t. $t'_{ij} = \alpha t'_{jk}$, $^\forall k$ (k = 1,2, …, N) (23)

because of the definition of linearly dependence.

Therefore

$$\sum_l h_{il} m_{lk} = \alpha \sum_l h_{jl} m_{lk} \quad (24)$$

According to the formula (17)

Rank $\boldsymbol{M}^t = N$ (25)

thus,

$^\forall i,j$, $h_{il} = \alpha h'_{jl}$ (26)

However, according to formula (20)

Rank $\boldsymbol{H'} = N$ (27)

Therefore

$^\exists l$(1 = 1,2, …,N), s.t. $h_{il} \neq \alpha\, h_{jl}$ (28)

This is contradictory. Therefore the i-th row and j-th row are linearly independent.

Q.E.D)

Theorem 2)

If

$^\exists\boldsymbol{H'}$, s.t. ($\boldsymbol{T'}=\boldsymbol{H'M} \wedge \boldsymbol{T'}$ is nonsingular)

then

$^\exists\boldsymbol{H}$, s.t.

$$\boldsymbol{T} = |t_{ij}| = \boldsymbol{HM} = \begin{bmatrix} 0 & & & \\ & 0 & nonzero & \\ & nonzero & & \\ & & & 0 \end{bmatrix}$$

Proof)

It is shown that

$^\exists\boldsymbol{H''}$, s.t. $\boldsymbol{T}=\boldsymbol{H''}\,\boldsymbol{T'}$ (29)

where $\boldsymbol{T'}$ is nonsingular.

$^\exists(\boldsymbol{T'})^{-1}$ s.t. $\boldsymbol{I} = (\boldsymbol{T'})^{-1}\boldsymbol{T'}$ (30)

$\boldsymbol{I}$: Identity matrix

because $\boldsymbol{T'}$ is nonsingular.

Let matrix $\boldsymbol{A} = |A_{ij}|$ as

$$A_{ij} = \begin{cases} 0 (i = j) \\ 1 (i \neq j) \end{cases} \quad (31)$$

And let $\boldsymbol{H''}$as

$\boldsymbol{H''}= \boldsymbol{A}(\boldsymbol{T})^{-1}$ (32)

then

$$\begin{aligned} \boldsymbol{T} &= \boldsymbol{H'Y'} \\ &= \boldsymbol{A}(\boldsymbol{T'})^{-1}\boldsymbol{T'} \\ &= \boldsymbol{AI} \end{aligned} \quad (33)$$

Choose $\boldsymbol{H}$ as

$\boldsymbol{H}=\boldsymbol{H''H'}$ (34)

then

$$\begin{aligned} &\boldsymbol{HM} \\ &= \boldsymbol{H''}\,\boldsymbol{H'}\,\boldsymbol{M} \\ &= \boldsymbol{A}(\boldsymbol{T'})^{-1}\boldsymbol{T} \\ &= \begin{bmatrix} 0 & & & \\ & 0 & nonzero & \\ & nonzero & & \\ & & & 0 \end{bmatrix} \end{aligned} \quad (35)$$

Therefore $^\exists\boldsymbol{H}$. s.t.

$$\boldsymbol{T} = |t_{ij}| = \boldsymbol{HM} = \begin{bmatrix} 0 & & & \\ & 0 & nonzero & \\ & nonzero & & \\ & & & 0 \end{bmatrix}$$

Q.E.D)

Theorem 3)

If

$^\exists\boldsymbol{H}$. s.t.

$$\boldsymbol{T} = |t_{ij}| = \boldsymbol{HM} = \begin{bmatrix} 0 & & & \\ & 0 & nonzero & \\ & nonzero & & \\ & & & 0 \end{bmatrix} \quad (36)$$

then it is possible to localize the sound source.

Proof)

Design $\boldsymbol{H}$ as

$$\boldsymbol{T} = |t_{ij}| = \boldsymbol{HM} = \begin{bmatrix} 0 & & & \\ & 0 & nonzero & \\ & nonzero & & \\ & & & 0 \end{bmatrix}$$

Consider the output of $\boldsymbol{h}(f, \theta_i)$ when $s(f, \theta_j)$, the sound from θ_j, is played. According to formula (12),

$y_i(f, \theta_j) = \boldsymbol{h}(f, \theta_i)\boldsymbol{m}(f, \theta_j)s'(f, \theta_j)$ (37)

Formula (37) can be described as follow.

$$y_i(f, \theta_j) = t_{ij}\, s'(f, \theta_j) \tag{38}$$

According to the assumption,

$$t_{ij} = \begin{cases} 0 & i = j \\ nonzero & i \neq j \end{cases} \tag{39}$$

Hence, if $\boldsymbol{h}(f, \theta_i)$, $i = 1, 2, \ldots, N$, are prepared, it is possible to localize the sound. Because $y_i(f, \theta_j)$ is as follow.

$$y_i(f, \theta_j) = \begin{cases} 0 & (SoundDirection) \\ nonzero & (otherwise) \end{cases} \tag{40}$$

while the sound $s(\theta_i)$ is generated. Q.E.D)

The actual system is as follow.

i) Check all the output of $\boldsymbol{h}(f, \theta_i)$ filter ($i = 1,2,\ldots,N$)

ii) Find $\boldsymbol{h}(f, \theta_i)$ as

$y_i(f, \theta_j) = 0$

iii) Localize the sound using ii)

3.3 FILTER'S INITIALIZATION AND LOCALIZATON PROCEDURE

It is not necessary to follow the calculation procedure in proof when $\boldsymbol{H}$ is calculated. What is necessary is to prepare the microphones satisfied with the condition (17) and to find $\boldsymbol{h}(f, \theta_i)$ which outputs 0 to each direction in advance.

At first, the sound localization filter $\boldsymbol{h}(f, \theta_i)$ over each direction is decided as follow.

i) A reference sound from a known direction is generated.

ii) Calculate $\boldsymbol{h}(f, \theta_i)$ which gives minimum output, using $\boldsymbol{r}(f, \theta_i)$, the output of each microphones.

The sound localization system is as follow.

i) Observe sound sources $\boldsymbol{r}(f, \theta_j)$.

ii) Execute FFT for the sound $\boldsymbol{r}(f, \theta_j)$.

iii) Input FFT data to all $\boldsymbol{h}(f, \theta_i)$.

iv) Choose the minimum value of the output of $\boldsymbol{h}(f, \theta_i)$

v) Output the angle corresponding to $\boldsymbol{h}(f, \theta_i)$

If the condition (17) is satisfied, it is possible to localize the sound to all the assumed angles.

4. EXPERIMENTAL RESULTS

Several experiments were conducted to evaluate the performance of the AMSOL in the case that N=2. Two microphones were prepared in the experiment. Each microphone was connected to the computer through the amplifier. We used two directional microphones, ECM-TS125 made by SONY (11 × 26mm). 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.

As the sound source, a loudspeaker was placed at distances 0.5 m from the microphones. The sound source was placed in one of eight directions. In this case, when the sound was generated from the loudspeaker, every outputs of $\boldsymbol{h}(f, \theta_i)$ were recorded. $\boldsymbol{h}(f, \theta_i)$ is the filter which outputs 0 only when the sound is generated

Table 1.The output of each $\boldsymbol{h}(f, \theta_i)$ (i=0,1,...,8)

(a) 500Hz

	0°	45°	90°	135°	180°	225°	270°	315°
h_1[dB]	-50.65	-36.06	-35.92	-34.77	-32.26	-47.54	-40.48	-41.21
h_2[dB]	-40.15	-51.4	-38.26	-34.08	-30.55	-43.32	-43.95	-39.59
H_3[dB]	-34.06	-41.46	-53.25	-44.67	-30.6	-36.02	-46.03	-41.93
H_4[dB]	-41.48	-30.96	-32.18	-56.46	-41.9	-36.7	-32.8	-41.98
h_5[dB]	-39.17	-29.93	-26.23	-33.97	-58.42	-32.45	-27.12	-33.04
h_6[dB]	-36.03	-45.86	-41.5	-45.86	-39.73	-57.68	-42.93	-45.02
h_7[dB]	-42.51	-33.78	-35.32	-31.69	-23.97	-31.29	-54.43	-32.21
h_8[dB]	-44.96	-47.6	-48.92	-46.53	-46.33	-47.46	-46.59	-52.05

(b) 1000Hz

	0°	45°	90°	135°	180°	225°	270°	315°
h_1[dB]	-59.71	-41.75	-42.48	-38.71	-48.55	-41.39	-39.47	-39.84
h_2[dB]	-43.23	-57.99	-36.63	-33.01	-35.83	-45.23	-38.43	-40.79
h_3[dB]	-42.25	-39.96	-57.31	-46.48	-36.56	-36.64	-47.16	-49.27
h_4[dB]	-42.45	-33.06	-46.54	-58.91	-34.33	-31.56	-42.39	-49.27
h_5[dB]	-47.41	-36.77	-39.32	-40,07	-58.28	-39.65	-37.13	-45.36
h_6[dB]	-45.44	-51.93	-47.65	-40.68	-40.7	-58.5	-47.47	-47.47
h_7[dB]	-38.07	-39.23	-51.44	-32.8	-37.51	-33.19	-56.92	-36.12
h_8[dB]	-44.99	-38.72	-36.63	-33.63	-40.53	-40	-36.89	-51.67

from the direction of 45× i. The experiment was conducted in an ordinary quiet room.

The actual output of the $\boldsymbol{h}(f, \theta_i)$ (i=1,2,...,8) is shown as Table.1. Table.1 shows the result of the experiment used sinusoidal wave as sound sources.

The frequency of the sound using in this experiment was 500Hz, 1000Hz. The sound source was sounded continuously. It is considered that the output of the diagonal ingredient of Table.1 is minimized if $\boldsymbol{h}(f, \theta_i)$ is designed appropriately.

Table.1 shows that the output of the diagonal ingredient is minimized. The experiment shows that it is possible to localize the sound using 2 microphones in AMSOL.

AMSOL is able to execute the sound localization tasks for every frequency. Hence, it is possible for AMSOL to localize the sound including different frequency, white noise, human speech and so on.

5. REFERENCES

[1] Masakiyo Fujimoto and Yasuo Ariki: ``Noise Robust Hands-Free Speech Recognition Using Microphone Array and Kalman Filter as Front-End System of Conversational TV", MMSP2002, In CD-ROM Proceedings, 2002-12.

[2] D. Giuliani, M. Omologo, P. Svaizer " Talker Localization and Speech Recognition using a Microphone Array and a Cross-power-Spectrum Phase Analysis"., Proceedings of ICSLP 94, Yokohama, 1994

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