

## 基礎行列と球面波拘束を併用するマイクロフォンの感度校正

川島 宏之<sup>†1</sup> 坂上文彦<sup>†1</sup> 本谷秀堅<sup>†1</sup>

本稿では音源定位をおこなうためのマイクロフォンアレイの感度校正に、カメラとマイクロフォンアレイの間に成り立つ幾何拘束を利用する手法を提案する。マイクロフォンアレイによる音源定位システムとカメラとを併用するシステムを構築する際には、座標系を校正する必要がある。音源定位システムは内部パラメータが既知の透視投影カメラと同様の扱いが可能であり、カメラとの座標系の校正には既存の手法が利用できる。しかし、個々のマイクロフォンの感度が未校正である場合には音源定位の精度が低下するだけでなく、座標系の校正も容易でなくなる。そこで提案法は音波が球面波であることを利用し、基礎行列も併用することによりマイクロフォンアレイの感度校正をおこなう。実験によりその効果を検証する。

### Calibration of Microphone Sensitivity using Fundamental Matrix and Spherical Wave Constraint

HIROYUKI KAWASHIMA,<sup>†1</sup> FUMIHIKO SAKAUE<sup>†1</sup>  
and HIDEKATA HONTANI<sup>†1</sup>

In this article, we propose a method that calibrates sensitivity of microphone array based on a geometric constraint between the microphone array and a camera. When we integrate a source localization system with a camera, we have to compute a map between their coordinates. For this computation, we can utilize classical stereo-view geometry because a source localization is identical with projective transformation. Though, when the sensitivity of each microphone is un-calibrated, the localization becomes inaccurate and the computation of the map becomes more difficult. Here, the proposed method auto-calibrates the sensitivity by using the stereo-view geometry and by considering a spherical wave constraint. We evaluated the performance of the proposed method by simulation experiments.

#### 1. はじめに

画像情報と音源の情報を融合して処理するためには、画像の座標系と音源定位システムの座標系との対応付けが不可欠である<sup>1)2)</sup>。対応付けができてはじめて、例えば画像中の複数の人物から発話者を同定したり、画像と音声の双方を使って人物をトラッキング<sup>3)</sup>したりすることが出来るようになる。

二台のカメラで撮影した画像間の対応付けは、基礎行列であらわすことができる。この基礎行列は空間中の点とそれぞれのカメラ中心に関する幾何的な拘束を表現するものであり、画像間の対応点の探索やカメラ校正などの基本となる行列であった。一方本稿では画像と音源の定位結果との間の対応付けを利用する。

音源定位は音声に基づいて音源の方向や距離を推定する処理のことであり、多くの手法が提案されている<sup>4)5)</sup>。音源定位により対象の方位だけが推定されるのであれば、音源定位は音源の三次元座標から方位ベクトルへの写像であり、射影カメラによる撮像と同等に扱える。また、方位と距離の双方を推定できるのであれば、音源定位は三次元座標のユークリッド変換と同等である。いずれの場合であっても、画像座標と音源定位結果の間に、基礎行列で表現可能な対応関係を求めることができる。

音源定位の手法の多くは、音速やマイクロフォンアレイの幾何的構成だけではなく、マイクロフォンの感度特性も既知であると仮定している。音速は気温により変化するが定位結果の質を大きく落とすほどの要因にはなりにくく、例えばマイクロフォン間の距離など、マイクロフォンアレイの幾何的構成もカメラの内部パラメータとは異なり、比較的容易に実測できる場合が多い。

一方マイクロフォンの感度特性については、逐次校正を必要とする場合が多い。なぜなら、同一規格のマイクロフォンの特性が個々に大きく異なることは稀ではあるが、アレイ状に配置したときに他のマイクや三脚・カメラによる隠蔽や反射の影響により、見かけ上の感度が大きく変化することがあるからである。

マイクロフォンアレイの感度の違いを、音波が球面波であることを利用して校正する手法が提案されている<sup>6)</sup>。この手法では、音の減衰と音場の勾配との間の成り立つ拘束式を満たすように、感度を逐次校正する。以上述べたように、カメラとマイクロフォンアレイを組み

<sup>†1</sup> 名古屋工業大学  
Nagoya Institute of Technology

合わせたシステムにおいては、カメラとマイクロフォンアレイの間に基礎行列であらわされる幾何的拘束が存在し、なおかつ各マイクロフォンで計測される音声の間には音波が球面波であることから要請される拘束が存在する。両者の拘束は、それぞれ異なる事象から導出されるものであり、それら事象の間は互いに独立である。

そこで本稿では両者の拘束を統合してマイクロフォンの感度補正に利用する手法を提案する。各拘束が異なる事象から導出されるものであるため、校正精度が向上することを期待できる。シミュレーション実験により、球面波拘束のみに基づく場合と比べて感度校正の精度が実際に向上することを確認したので報告する。

## 2. 音源定位法

本提案法で採用する音源定位法<sup>6)</sup>は音源拘束式と呼ばれる偏微分方程式に基づく。以下、音源拘束式の説明をおこなったあと、音源定位のアルゴリズムを説明する。

### 2.1 音源定位の基礎となる偏微分方程式

マイクロフォンアレイの位置ベクトルを  $\mathbf{r}_0 = (x, y, z)^T$  で表し、音源の位置ベクトルを  $\mathbf{r} = (x_0, y_0, z_0)^T$  で表す。また時刻  $t$  における音源の音圧を  $g(t)$  で表し、音速を  $c$  で表す。このとき、観測点  $\mathbf{r}_0$  における音圧は式 (1) のように表される。

$$f(\mathbf{r}, t) = \frac{1}{|\mathbf{r} - \mathbf{r}_0|} g\left(t - \frac{|\mathbf{r} - \mathbf{r}_0|}{c}\right). \quad (1)$$

ただし

$$|\mathbf{r} - \mathbf{r}_0| = \{(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2\}^{1/2}, \quad (2)$$

である。式 (1) の両辺の空間勾配を求めると次の式が得られる。

$$\begin{aligned} \nabla f(\mathbf{r}, t) &= \nabla \left\{ \frac{1}{|\mathbf{r} - \mathbf{r}_0|} g\left(t - \frac{|\mathbf{r} - \mathbf{r}_0|}{c}\right) \right\} \\ &= \frac{1}{|\mathbf{r} - \mathbf{r}_0|^2} g\left(t - \frac{|\mathbf{r} - \mathbf{r}_0|}{c}\right) \mathbf{n} + \frac{1}{c|\mathbf{r} - \mathbf{r}_0|} \dot{g}\left(t - \frac{|\mathbf{r} - \mathbf{r}_0|}{c}\right) \mathbf{n}. \end{aligned} \quad (3)$$

ただし、式 (3) の導出の過程では、マイクロフォンからの音源の方向を示す単位ベクトル  $\mathbf{n}$  に関する下記の式を利用している。

$$\nabla |\mathbf{r} - \mathbf{r}_0| = \frac{\mathbf{r}_0 - \mathbf{r}}{|\mathbf{r} - \mathbf{r}_0|} = -\mathbf{n}. \quad (4)$$

$$\nabla \frac{1}{|\mathbf{r} - \mathbf{r}_0|} = -\frac{\mathbf{r}_0 - \mathbf{r}}{|\mathbf{r} - \mathbf{r}_0|^3} = \frac{1}{|\mathbf{r} - \mathbf{r}_0|^2} \mathbf{n}. \quad (5)$$

また、 $\dot{f} = \partial f / \partial t$  であり、次式が成り立つ。

$$\dot{f}(\mathbf{r}, t) = \frac{1}{|\mathbf{r} - \mathbf{r}_0|} \dot{g}\left(t - \frac{|\mathbf{r} - \mathbf{r}_0|}{c}\right). \quad (6)$$

式 (1) と (3) を式 (6) に代入することにより、音源波形  $g$  が消去され観測点での音圧とその勾配のみを含む式 (7) を得る。

$$\begin{aligned} \nabla f(\mathbf{r}, t) &= \left\{ \frac{1}{|\mathbf{r} - \mathbf{r}_0|} f(\mathbf{r}, t) + \frac{1}{c} \dot{f}(\mathbf{r}, t) \right\} \mathbf{n} \\ &= \left\{ \frac{1}{R} f(\mathbf{r}, t) + \frac{1}{c} \dot{f}(\mathbf{r}, t) \right\} \mathbf{n}. \end{aligned} \quad (7)$$

ただし、式中の  $R$  は音源までの距離を表しており、 $R = |\mathbf{r} - \mathbf{r}_0|$  である。偏微分方程式 (7) を音源拘束式 (Sound Source Constraint) と呼ぶ。

### 2.2 荷重積分法

この音源拘束式が任意の時間区間  $[-T/2, T/2]$  で一様に成り立っていると仮定する。音源拘束式に複素正弦波荷重  $e^{-j\omega t}$  を乗じてこの時間空間内で積分をすると式 (8) を得る。

$$\begin{aligned} g_\omega(\mathbf{r}) &\equiv \int_{-T/2}^{T/2} f(\mathbf{r}, t) e^{-j\omega t} dt \\ &= \int_{-T/2}^{T/2} \left\{ \nabla f(\mathbf{r}, t) - \left( \frac{1}{R} f(\mathbf{r}, t) + \frac{1}{c} \dot{f}(\mathbf{r}, t) \right) \mathbf{n} e^{-j\omega t} \right\} dt \\ &= \nabla g_\omega(\mathbf{r}) - \frac{\mathbf{n}}{R} g_\omega(\mathbf{r}) - \frac{\mathbf{n}}{c} [f(\mathbf{r}, t) e^{-j\omega t}]_{-T/2}^{T/2} - \frac{j\omega \mathbf{n}}{c} g_\omega(\mathbf{r}) \\ &= 0. \end{aligned} \quad (8)$$

ここで  $\omega T = 2n\pi$  と選ぶと、部分積分によって生じた積分境界値項は式 (9) のようになることから、積分境界値項は周波数に依存しない値になることがわかる。

$$\begin{aligned} [f(\mathbf{r}, t) e^{-j\omega t}]_{-T/2}^{T/2} &= f(\mathbf{r}, T/2) e^{-j\pi n} - f(\mathbf{r}, -T/2) e^{j\pi n} \\ &= (-1)^n [f(\mathbf{r}, t)]_{-T/2}^{T/2}. \end{aligned} \quad (9)$$

従って、 $\omega_l = 2\pi l/T$  と  $\omega_m = 2\pi m/T$  の場合の二式を連立して積分境界値項を消去すると式 (10) を得る。

$$\begin{aligned} (-1)^{l+m} [f(\mathbf{r}, t)]_{-T/2}^{T/2} &= (-1)^m \nabla g_{\omega_l}(\mathbf{r}) - (-1)^m \frac{\mathbf{n}}{R} g_{\omega_l}(\mathbf{r}) - (-1)^m \frac{j\omega_l \mathbf{n}}{c} g_{\omega_l} \\ &= (-1)^l \nabla g_{\omega_m}(\mathbf{r}) - (-1)^l \frac{\mathbf{n}}{R} g_{\omega_m}(\mathbf{r}) - (-1)^l \frac{j\omega_m \mathbf{n}}{c} g_{\omega_m}. \end{aligned} \quad (10)$$

よって

$$\nabla p_\omega(\mathbf{r}) - \left( \frac{p_\omega(\mathbf{r})}{R} + \frac{j\omega_\omega(\mathbf{r})}{c} \right) \mathbf{n} = 0, \quad (11)$$

が得られる。ただし

$$\begin{aligned} p_\omega(\mathbf{r}) &= (-1)^l g_{\omega_m}(\mathbf{r}) - (-1)^m g_{\omega_l}(\mathbf{r}), \\ q_\omega(\mathbf{r}) &= (-1)^l \omega_m g_{\omega_m}(\mathbf{r}) - (-1)^m \omega_l g_{\omega_l}(\mathbf{r}), \end{aligned} \quad (12)$$

である。

### 2.3 定位アルゴリズム

マイクロフォンを六つ配置して、音圧の  $x, y, z$  微分を差分により求める。最小二乗法による単一音源定位をおこなう。最小化すべき評価関数は式 (13) で表される。

$$J = \int_\omega \left| \nabla p_\omega^n - \left( \frac{p_\omega^n(\mathbf{r})}{R} + \frac{j\omega_\omega(\mathbf{r})}{c} \right) \mathbf{n} \right|^2 d\omega. \quad (13)$$

すると式 (13) を最小にする  $\mathbf{n}, 1/R$  は式 (14), (15) で与えられる。

$$\mathbf{n} = \frac{\Re\{p_\omega^n/R + jq_\omega(\mathbf{r})/c\} \nabla p_\omega^n(\mathbf{r})^*}{|p_\omega^n/R + jq_\omega(\mathbf{r})/c|^2}. \quad (14)$$

$$\frac{1}{R} = \frac{\Re\{p_\omega^n(\mathbf{r})(\nabla p_\omega^n(\mathbf{r}))^H \mathbf{n}\} - \Re\{p_\omega^n(\mathbf{r})^* \frac{j}{c} q_\omega(\mathbf{r})\}}{|p_\omega^n(\mathbf{r})|^2}. \quad (15)$$

これを解くには、まず初期値として音源距離が無限遠  $1/R = 0$  とおいて式 (14) より単位ベクトル  $\mathbf{n}$  を推定する。そして求めた  $\mathbf{n}$  を用いて式 (15) より  $1/R$  を推定する。これを  $\mathbf{n}$  が収束するまで反復することにより  $\mathbf{n}, 1/R$  を推定することが出来る。また実験により  $\mathbf{n}$  は数回の反復で収束することが確認できた。

### 3. 感度パラメータの校正

周波数  $\omega$  の信号に対するマイクロフォン  $i$  の感度をあらわすパラメータを  $\beta_i$  で表す。同一の音圧信号がマイクロフォンに到達したとき、 $\beta_i$  の値が小さい時ほど当該マイクロフォンからの出力も小さくなる。感度パラメータの校正とは、全マイクロフォンの感度パラメー

タ  $\beta \equiv (\beta_1(\omega) \cdots \beta_n(\omega))^T$  を推定することである。ただし  $n$  はマイクロフォンの個数である。また  $\beta(\omega)$  は以下の凸拘束条件に繰り返し直交射影されることにより逐次更新される。

#### 3.1 球面波拘束条件

定位する単一音源の波面が球面波であることから、音源方向  $\mathbf{n}$  と距離  $R$  によらず常に球面波拘束式 (16) が成立しなければならない。

$$\nabla \mathbf{F}(\omega) - \left( \frac{j\omega}{c} + \frac{1}{R} \right) \mathbf{F}(\omega) \mathbf{n} = 0. \quad (16)$$

ただし  $\mathbf{F}(\omega)$  は

$$\mathbf{F}(\omega) = \int f(r, t) e^{-j\omega t} dt, \quad (17)$$

でアレイ中心の音圧を、 $\nabla \mathbf{F}(\omega)$  はその空間勾配を、 $\omega$  は角周波数、 $c$  は音速を表す。

ここで本実験で用いるマイクロフォンの数を図 1 のように原点を中心に  $x, y, z$  軸に二つずつ計六つとし、各対向マイクロフォンの間隔を  $D$  とする。各マイクロフォンにより観測される音圧信号のフーリエ変換を  $G_1(\omega), \dots, G_6(\omega)$  として表すと、式 (16) は式 (18) のように表せる。

$$\begin{pmatrix} (\beta_1(\omega)G_1(\omega) - \beta_2(\omega)G_2(\omega))/D \\ (\beta_3(\omega)G_3(\omega) - \beta_4(\omega)G_4(\omega))/D \\ (\beta_5(\omega)G_5(\omega) - \beta_6(\omega)G_6(\omega))/D \end{pmatrix} - \left( \frac{j\omega}{c} + \frac{1}{R} \right) \frac{\sum_i \beta_i(\omega)G_i(\omega)}{6} \begin{pmatrix} n_x \\ n_y \\ n_z \end{pmatrix} \simeq 0. \quad (18)$$

これは  $\beta(\omega)$  に関する一次式であり式 (19) のように書き換えることができる。

$$(\mathbf{J}_x(\omega), \mathbf{J}_y(\omega), \mathbf{J}_z(\omega))^H \beta(\omega) \simeq 0. \quad (19)$$

ただし  $\mathbf{J}_x(\omega), \mathbf{J}_y(\omega), \mathbf{J}_z(\omega)$  は音源位置によって定まる係数ベクトルである。

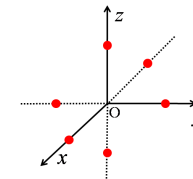


図 1 マイクロフォンの配置。赤:マイクロフォン。

Fig.1 Location of microphones. Red spots show the locations.

### 3.2 感度パラメータの校正

音源定位により得られた  $\mathbf{n}$ ,  $R$  の推定値を用いて  $\mathbf{J}_x(\omega)$ ,  $\mathbf{J}_y(\omega)$ ,  $\mathbf{J}_z(\omega)$  を求める。これら推定値には推定誤差含まれるため、 $\mathbf{J}$  に直行する  $\beta$  を算出しても真の感度パラメータを推定することができない。文献<sup>6)</sup>では、 $\beta$  の推定に凸射影法<sup>7)</sup>が採用されている。すなわち、式(19)の左辺に誤差を許容して次式の条件へと書き換える。

$$|\mathbf{J}_x(\omega)^H \beta(\omega)| \leq \varepsilon. \quad (20)$$

凸射影法においては、仮に式(20)が満たされない場合には、式(20)が定義する凸空間へと  $\beta$  を射影して更新する。 $\beta$  の真値がこの凸空間に含まれているならば、更新を繰り返すことにより  $\beta$  は真値へと収束する。図2にその概要を示す。ここで現地点での最新の  $\beta$  を  $\beta_{now}$  とし、 $\beta_{now}$  を更新したものを  $\beta_{new}$  とする。また  $\mathbf{J}^T \beta_{now} < \varepsilon$  を満たす場合は既に  $\beta$  が直交射影されるべき空間内に入っているため更新はおこなわないものとする。つまり  $\mathbf{J}^T \beta_{now} > \varepsilon$  の場合のみ以下の更新の手続きをおこなう。

まず  $\mathbf{J}_{xk}$  は6次元の複素数ベクトルであるが、実部と虚部をそれぞれ長さ1で正規化し、それぞれを  $\mathbf{J}_{Re}$ ,  $\mathbf{J}_{Im}$  とおく。 $\mathbf{J}_{Re}$  に直交するベクトル五つを Gram-Schmidt の正規直交化法により求め、それぞれを  $\mathbf{u}_1, \dots, \mathbf{u}_5$  とする。 $\beta_{now}$  を  $\mathbf{u}_1, \dots, \mathbf{u}_5$  で張られる5次元空間  $\mathbf{U}$  へ直交射影し長さ1で正規化したベクトル  $\alpha$  を求める。

$\Delta = \beta_{now} - \alpha$  となるような  $\Delta$  を求め、式(21)により定まる  $\beta_{new}$  に  $\beta_{now}$  を更新する。ただし更新された  $\beta_{new}$  は長さ1で正規化する。そして同様の手順を  $\mathbf{J}_{Im}$  についても行う。

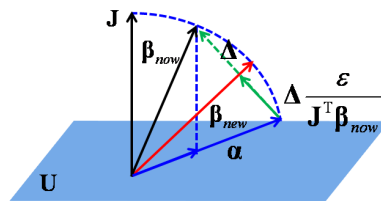


図2 の更新  
Fig.2 Update of  $\beta$

$$\beta_{new} = \alpha + \Delta \frac{\varepsilon}{\mathbf{J}^T \beta_{now}} \quad (21)$$

以上の手順により  $\beta$  を逐次更新することが出来る。

### 4. エピポーラ拘束と球面波拘束を併用するマイクロフォンアレイの感度校正

本章において、二台のカメラ間に成り立つエピポーラ拘束と球面波拘束を併用してマイクロフォンアレイの感度校正をおこなう提案法を説明する。

#### 4.1 音源定位と射影変換

音源定位で得られた球面拘束条件式(18)は次のように簡略に書き直すことができる。

$$\mathbf{G}\beta + k\mathbf{n} = 0. \quad (22)$$

ただし、行列  $\mathbf{G}$  と係数  $k$  は以下のようになる。

$$\mathbf{G} = \begin{bmatrix} G_1/D & G_2/D & 0 & 0 & 0 & 0 \\ 0 & 0 & G_3/D & G_4/D & 0 & 0 \\ 0 & 0 & 0 & 0 & G_5/D & G_6/D \end{bmatrix}, \quad (23)$$

$$k = \left( \frac{j\omega}{c} + \frac{1}{R} \right) \frac{\sum_i \beta_i(\omega) G_i(\omega)}{6}.$$

式(22)は両辺を定数倍しても成立し、感度パラメータ  $\beta_i$  は他のマイクロフォンの感度  $\beta_j$  との比にのみ意味することに注意する。式(22)に示されている通り、マイクロフォンの感度パラメータ  $\beta$  の校正は、基本的に音源の単位方向ベクトル  $\mathbf{n}$  に基づいておこなわれる。

世界座標系における音源の斉次座標  $\tilde{\mathbf{X}} = [X, Y, Z, 1]$  と  $\mathbf{n}$  との間には次式が成り立つ。

$$s\mathbf{n} = [\mathbf{I}|\mathbf{R}|\mathbf{t}]\tilde{\mathbf{X}}. \quad (24)$$

ただし、 $s$  は係数、 $\mathbf{I}$  は  $3 \times 3$  の単位行列であり、 $\mathbf{R}$  は回転行列、 $\mathbf{t}$  は並進を表す3次元ベクトルである。音源定位により推定される  $\mathbf{n}$  と音源位置の関係は、カメラ行列が単位行列で表される射影カメラによる投影と同じ式で表せることがわかる。

スピーカや人の口などの音源の画像を撮影できたとして、その画像上の斉次表現を  $\tilde{\mathbf{m}}$  で表す。これまでの議論によりマイクロフォンアレイとカメラとの間の基礎行列  $\mathbf{F}$  が定義できることから、次式が成り立つ。

$$\mathbf{n}^T \mathbf{F} \tilde{\mathbf{m}} = 0. \quad (25)$$

基礎行列の計算には、例えば8点法<sup>8)</sup>が利用できる。すなわち、八箇所音源の画像座

標  $m_k (k = 1, 2, \dots, 8)$  を求めるとともに、それら八つの音源の定位をおこない  $n_k$  も取得すれば良い。

#### 4.2 球面波拘束式と基礎行列

式 (22) を式 (25) に代入することにより、音源の画像上での座標とマイクロフォンアレイの感度パラメータを直接結びつける式 (26) を得ることができる。

$$\beta^T G^T F \tilde{m} = 0. \quad (26)$$

ここで、式 (26) の右辺の要素  $G, F, \tilde{m}$  が既に得られているならば、次式のように書き直すことができる。 $\alpha$  と  $\beta$  は共にマイクロフォンの個数と同じ次元を持つベクトルである。

$$\beta^T \alpha = 0. \quad (27)$$

ただし、 $\alpha = G^T F \tilde{m}$  であり、複素数を要素とする、マイクロフォンの数を次元とするベクトルである。 $\alpha$  は、不確かな  $\beta$  に基づいておこなった音源定位結果の誤差の影響を受ける。このため、式 (27) に従って  $\alpha$  に直交する  $\beta$  を求めても真値を推定できない。そこで文献<sup>6)</sup>と同様に、凸射影法を採用し、感度パラメータの更新をおこなう。すなわち、適当に定められた  $\varepsilon$  に対して次式を満たす凸空間への射影をおこなう。

$$\beta^T \alpha \leq \varepsilon. \quad (28)$$

#### 4.3 感度校正アルゴリズム

基礎行列を 8 点法で求めるときの、本稿で提案するマイクロフォンアレイの感度校正アルゴリズムを示す。

1.  $\beta$  の初期値  $\beta^{k=0}$  を定める。
2. 八箇所の音源の画像と音声を逐次記録する。新たな音源からの音声を取得するたびに、式 (20) に基づく凸射影法により  $\beta$  の値を更新する。八箇所の音源すべてを用いて更新した感度パラメータを  $\beta^{k=1}$  であらわす。
3. 音源の画像座標  $m_k$  と音源の方向  $n_k$  を計算し、基礎行列  $F$  を 8 点法で計算する。ただし、 $n_k$  の推定には  $\beta^{k=1}$  を用いる。
4. 式 (28) の  $\alpha$  を計算し、凸射影法により  $\beta^{k=1}$  を更新し  $\beta^{k=2}$  を得る。
5.  $\beta^{k=2}$  が収束すれば終了。収束していなければ 3 に戻る。

### 5. 実験結果

#### 5.1 音源定位結果 (シミュレーション)

まず音源定位に関する実験結果を示し、次に感度校正に関する実験の結果を示す。定位する音源には周波数 300Hz の正弦波と周波数 400Hz の正弦波を足しあわせた音を用い、サ

ンプリング周波数は 44.1kHz とした。音速を求めるために気温は 23 度とし、式 (8) を計算するための積分区間を 0.0067 秒とした。これは荷重積分する複素正弦波の 2 周期分に相当する。

##### 5.1.1 音源定位の対ノイズ性能

荷重積分法による音源定位の結果を図 3 に示す。音声信号の S/N 比は 40dB に設定した。図 3 中の黒点は原点を表しており、その原点にマイクロフォンアレイを配置し、六つの赤い点は音源を配置した真の位置、その周りにプロットされている点は先述の音源定位を用いてそれぞれの音源の位置を推定した結果を示している。また図 3 の各軸はそれぞれ  $x, y, z$  の空間中の座標を示しており、単位は cm である。

音源の推定位置が真の位置の周りに分布していること、音源の位置が遠くなるほど推定精度が低下していることなどが確認できる。

図 4 に、各 S/N 比における音源までの距離と定位誤差のグラフを示す。横軸は距離 (cm) であり、縦軸は誤差 (cm) をあらわす。赤、緑、青のグラフはそれぞれ 46dB, 40dB, 33dB に対応している。例えば 33dB のとき、2メートル先の音源の定位精度がおおよそ  $\pm 5$ cm 程度であった。

##### 5.1.2 感度パラメータの変化の影響

図 5 に、六つのマイクロフォンのうち、ある一つのマイクロフォンの感度パラメータの値のみを変化させた場合の三つの音源位置の定位結果を示す。

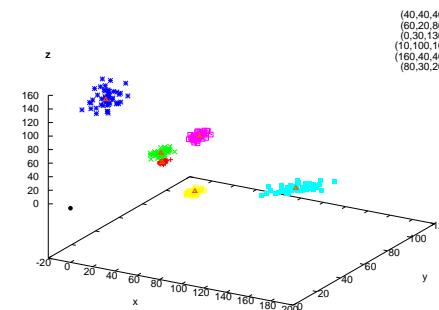


図 3 音源定位結果  
 Fig. 3 Results of source localization.

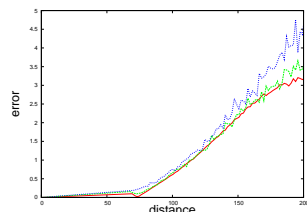


図 4 音源定位性能. 横軸: 距離. 縦軸: 定位誤差. 赤: S/N 比 46dB. 緑: S/N 比 40dB. 青: S/N 比 33dB.

Fig. 4 Performance of source localization. x- and y-axis show the distance and the estimation error, respectively. Red shows S/N ratio 46dB. Green shows S/N ratio 40dB. Blue shows S/N ratio 33dB.

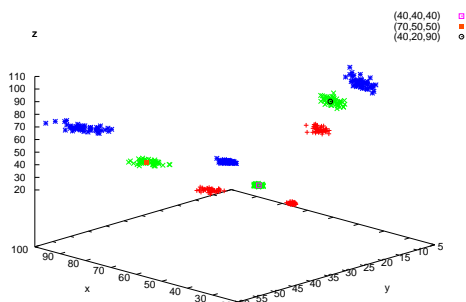


図 5 感度パラメータ  $\beta$  を変えた場合の三音源の定位結果. 緑:  $\beta$  の値が既知のときの推定結果. 赤:  $\beta$  の値が未校正で真値に  $\beta_k = 0.9$  を用いたときの推定結果. 青: 赤:  $\beta$  の値が未校正で真値に  $\beta_k = 1.1$  を用いたときの推定結果.

Fig. 5 Simulation results of source localization. Green results were obtained when the value of  $\beta$  is known. Red and blue show when  $\beta$  is un-calibrated.

図 5 中の緑色点の分布は,  $\beta_i = 1$  であることが既知の場合の定位結果を示す. 図 3 と同様に真値の周りに定位結果が分布していることが分かる. 一方, 六つのマイクロフォンのうち, ひとつの感度の真値が  $\beta_k = 0.9$  であるにも関わらず  $\beta_k = 1$  と仮定して定位した結果を赤色で示し,  $\beta_k = 1.1$  にも関わらず  $\beta_k = 1$  と仮定して得られた結果を青色で示す. 感度校正をおこなわないとき, 定位結果にバイアスがのることが分かる.

五つのマイクロフォンの感度を  $\beta_i = 1$  に固定し, 残り一つの  $\beta_k$  の値を変化させる. このときに,  $\beta_k = 1$  と仮定して音源定位した結果のバイアスの変化を図 6 に示す. 横軸が  $\beta_k$  の値であり, 真値と推定値の差をあらわし, 縦軸が定位誤差である. この誤差は音源の位置にも依存するが,  $\beta$  の値がずれるほど定位結果が次第に不正確になる様子を観察することができる.

### 5.2 音源定位結果 (実機)

本節では実機による実験結果について示す. ただし, 実験環境は無響室である. 定位する音源には録音した女性の声をスピーカにより再生したものを利用し, サンプリング周波数

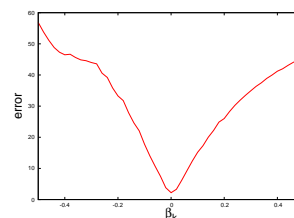


図 6 感度パラメータ  $\beta$  の変化に対する定位誤差. 横軸: 真の値と推定に用いた値の差. 縦軸: 定位誤差.

Fig. 6 Error of localization. x-axis shows the difference between the true value of  $\beta$  and the estimated one. y-axis shows estimated error.

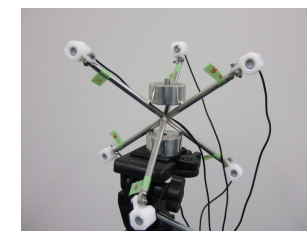


図 7 マイクロフォンアレイ  
Fig. 7 Microphone array.

は 44.1kHz とした. なおシミュレーション同様に, 積分区間 0.0067 秒としておこなった. 図 7 に本実験で用いたマイクロフォンアレイを示す. 対向マイクロフォンの間隔は 20cm である. 実験として, 音源の位置を変化させたときの定位位置の変化を観察した. すなわち, まずしかるべき位置に音源を配置して録音・定位し, 次に音源の位置をマイクロフォンアレイからみて 90 度ずらしたあとに再度録音・定位した.

まず感度パラメータ  $\beta_i \equiv 1$  を仮定して音源定位し, 次に  $\beta_i$  を校正したあとで同じ音声に対して定位をなおした. その結果を図 8 に示す. 図 8(a) に示されている通り,  $\beta_i \equiv 1$  を仮定したときには二つの音源位置を推定できていない. 一方, 図 8(b) に示すように,  $\beta_i$  を校正したあとは二つの音源を明瞭に区別出来ていることが分かる. 校正後の  $\beta$  の値は  $\beta = [1.17, 0.69, 1.00, 1.00, 1.10, 0.88]^T$  となった.

さらに, 表 1 には音源までの距離と推定位置の標準偏差を示す. マイクロフォンアレイまでの距離が 50cm 程度近いときは, 推定位置の標準偏差は 11cm 程度であった.

図 9 には, マイクロフォンの感度パラメータの校正をおこなった後, 直線上で 10cm ずつ音源を遠ざけたときの定位結果を示す. 図中の水色の点が原点であり, 原点にマイクロフォンアレイを配置して定位をおこなった. また赤色, 緑色, 青色, 紫色はそれぞれ基準点, 基準点から 10cm, 基準点から 20cm, 基準点から 30cm の定位結果を表している. なお  $x$  軸,  $y$  軸,  $z$  軸ともに単位は cm である. 図 9 に示す通り, 音源定位は距離の推定値と比べると, 方向の推定結果のほうが信頼度が高いことがわかる.

### 5.3 提案法による $\beta$ の校正

本節では, シミュレーションにより基礎行列  $\mathbf{A}$  の校正をおこなった実験の結果を報告する.

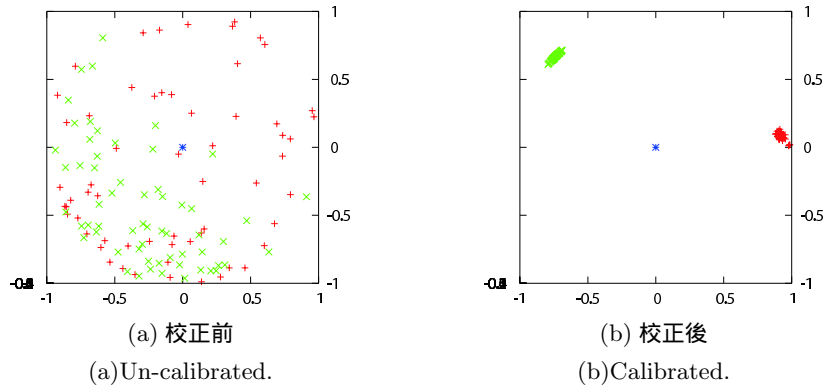


図 8 基準点での定位結果  
Fig. 8 Comparison of localization performance.

表 1 距離ごとの定位結果の標準偏差 (cm)  
Table 1 Standard deviation of estimated location.(cm)

	x 軸方向	y 軸方向	z 軸方向
基準点	0.285	0.055	0.078
10cm	0.954	0.236	0.627
20cm	2.938	0.978	2.966
30cm	6.077	2.475	8.232
40cm	6.891	2.677	10.403
50cm	6.017	2.512	9.476

### 5.3.1 シミュレーション実験

定位する音源には 100Hz の正弦波と 130Hz の正弦波を足し合わせた音を用い、ガウスノイズを付与した。また、サンプリング周波数は 44.1kHz、シミュレーション気圧は 26 度、積分区間は 0.0067 秒とした。カメラを  $[10.0, 0.0, 0.0]^T$  に配置し、マイクロフォンアレイを原点に配置し、音源はランダムに 8 点配置した。図 10 にカメラ、マイクロフォンアレイと

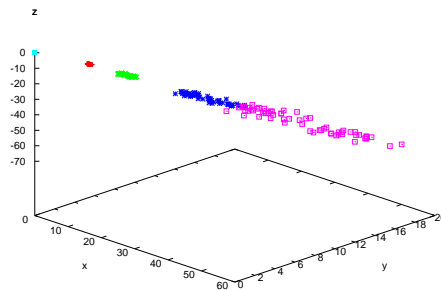


図 9 校正後、距離ごとの定位結果  
Fig. 9 Distribution of estimated locations. Each color corresponds to one sound source.

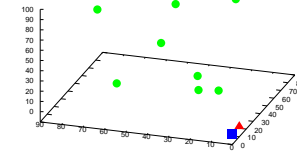


図 10 カメラとマイクロフォンアレイと音源の配置。赤:カメラ。青:マイクロフォンアレイ。緑:音源。  
Fig. 10 Location of camera, microphone array and sound sources. Red shows camera. Blue shows microphone array. Green shows sound sources.

表 2 音源の座標とマイクロフォンアレイとの距離  
Table 2 Distance between sound sources and microphone array.

音源番号	1	2	3	4	5	6	7	8
音源の座標	22.0	30.0	15.0	26.0	82.0	11.0	64.0	42.0
距離	21.0	28.0	24.0	99.0	64.0	32.0	53.0	46.0

音源の位置を示す。赤色の点はカメラを表しており、青色の点はマイクロフォンアレイを、緑色の点は配置した音源 8 点を表している。また表 2 に各音源の座標とマイクロフォンアレイから音源の位置までの距離を示す。

真の  $\beta = [0.9, 0.2, 0.4, 0.6, 0.5, 0.7]$  に対して、 $\beta$  の初期値  $\beta_0 = [1, 1, 1, 1, 1, 1]$  とした。このとき、球面波拘束式のみを用いて校正したところ最終的に得られた  $\beta$  の値は  $\beta = [0.86, 0.76, 0.89, 0.78, 0.87, 0.79]$  であった。一方、提案法により画像も併用したときに得られた  $\beta$  の値は  $\beta = [0.95, 0.93, 1.02, 1.00, 1.30, 1.43]$  であった。これらの結果に基づき八つの音源の位置を推定した結果を図 11, 図 12 に示す。図 10 と同様に、赤色の点がカメラ、青色の点がマイクロフォン、緑色の点が配置した音源を示し、紅色の点が音源の定位結果を表している。また、水色の線で結ばれた緑色と紅色の点に対応した点である。図 11 は、球面波拘束によって校正した  $\beta$  を用いた結果であり、図 12 は提案法により校正した  $\beta$  を用いた結果である。図に示したとおり、球面波拘束単独の校正よりも、提案法のほうが定位誤差が大幅に減少していることがわかる。

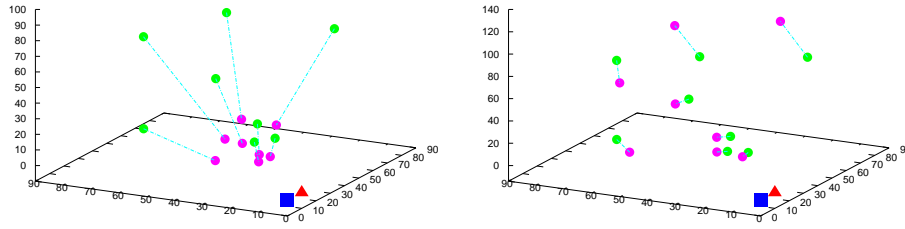


図 11 定位結果．球面波拘束による校正．赤:カメラ．青: マイク  
マイクロフォンアレイ．緑:音源．紅色:定位結果．  
Fig. 11 Experimental results of localization. Re-  
sults obtained when  $\beta$  was calibrated  
based only on spherical wave constraint.  
Red shows camera. Blue shows micro-  
phone array. Green shows sound sources.  
Pink shows result of localization.

図 12 基礎行列を併用した校正．赤:カメラ．青:マイク  
ロフォンアレイ．緑:音源．紅色:定位結果．  
Fig. 12 Results obtained when the spherical wave  
constraint and the fundamental matrix  
are used for the calibration. Red shows  
camera. Blue shows microphone array.  
Green shows sound sources. Pink shows  
result of localization.

また、八つの音源の位置を推定する処理をノイズを変化させつつ5回繰り返した時の平均定位誤差を図13に示す．横軸が音源の番号であり、縦軸が定位誤差を表す．緑色が従来の球面波拘束のみを用いたときの定位誤差であり、赤色が提案法による定位誤差である．図に示した通り、定位精度がすべての音源の位置において高くなっていることが分かる．

## 6. おわりに

本研究では、基礎行列によりマイクロフォンアレイの感度校正をおこなう手法を提案した．まず、マイクロフォン単独で音源定位をおこない、同時に感度パラメータ  $\beta$  の校正をおこなった．続いて、カメラとマイクロフォンアレイの間に成り立つエピソード拘束をもとに、8点法から基礎行列を推定し、感度パラメータ  $\beta$  の校正をおこなった．

また、球面波拘束のみの校正により推定した  $\beta$  を用いた音源定位よりも、画像情報から求めた基礎行列を併せた校正により推定した  $\beta$  を用いた音源定位のほうが精度が良いことをシミュレーションにより確認した．今後、式(19)の  $J$  と式(27)の  $\alpha$  の関係を調べるとともに、実機による  $\beta$  の校正の効果を検証する．

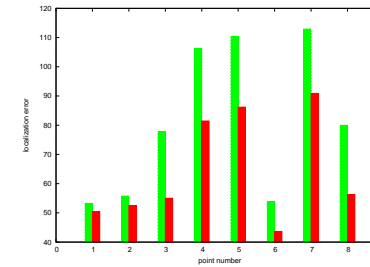


図 13 定位誤差の平均 (40dB) . 横軸:音源の番号 . 縦軸:定位誤差 . 緑:従来法 . 赤:提案法 .  
Fig. 13 Average of locate error(40dB). x-axis shows sound source number. y-axis shows locate  
error. Green shows past method. Red shows proposed method.

## 参 考 文 献

- 1) Barzelay, Z. and Schechner, Y.: Onsets coincidence for cross-modal analysis, *IEEE Transactions on Multimedia* 12 (2), Vol.12, No.5340552, pp.108–120 (2010).
- 2) Dupont, S. and Luetin, J.: Audio-visual speech modeling for continuous speech recognition, *IEEE Transactions on Multimedia* 2 (3), Vol.2, pp.141–151 (2000).
- 3) Wang, C. and Brandstein, M.: Hybrid real-time face tracking system, *ICASSP, IEEE International Conference on Acoustics*, pp.3737–3740 (1998).
- 4) Chen, J., Yao, K. and Hudson, R.: Source localization and beamforming, *IEEE Signal Processing Magazine*, Vol.24, No.4, pp.30–39 (2002).
- 5) Valin, J., Michaud, F., Rouat, J. and Letourneau, D.: Robust Sound Source Localization Using a Microphone Array on a Mobile Robot, *Proceedings Intelligent Robots and Systems*, Vol.2, pp.1228–1233 (2003).
- 6) Fujita, Y., Ono, N. and Ando, S.: Partial-differential-equarray-based sound source localization:Finite Fourier integral approach and its application to multiple source localization, *4th Joint Meeting ASA and ASJ* (2006).
- 7) Stark, H. and Oskoui, P.: High resolution image recovery from image-plane arrays, using convex projections, *J. Opt. Soc. Am. A*, Vol.6, pp.1715–1726 (1989).
- 8) 佐藤 淳: コンピュータビジョン -視覚の幾何学-, コロナ社 (1999).