マイクロホン位置と音源スペクトルの確率モデルに基づくマイクロホンアレイのキャリブレーション

Calibration of a microphone array based on stochastic model of microphone position and sound source spectrum

段 雄啓[1]　糸山 克寿[1*]　西田 健次[1]　中臺 一博[1,2]
Katsuhiro Dan[1]　Katsutoshi Itoyama[1]　Kenji Nishida[1]　Kazuhiro Nakadai[1,2]

[1] 東京工業大学
[1] Tokyo Institute of Technology
[2] （株）ホンダ・リサーチ・インスティチュート・ジャパン
[2] Honda Research Institute Japan, Co., Ltd.

Abstract: 本稿は，マイクロホンアレイを構成するマイクロホンの位置ずれを校正（キャリブレーション）する手法について述べる．提案手法は，マイクロホンの位置，音源のスペクトル，録音された混合音のスペクトルに関する確率的生成モデルに基づいている．これにより，従来の手法では困難だった混合音を入力としたキャリブレーションを実現する．このモデルに基づき，観測した混合音に対する最大事後確率推定問題の解としてマイクロホン位置のキャリブレーションを実現する．3種類のマイクロホンアレイを用いたシミュレーション実験により，提案手法は混合音に対してマイクロホン位置のキャリブレーションを行えることが示された．

1　はじめに

近年，マイクロホンアレイはロボットなどの様々なデバイスに搭載されるようになっており，マイクロホンアレイの収録音を用いた音源定位や音源分離などの音響信号処理技術などが研究されている [1–4]．具体的な適用分野としては，スマートスピーカーやドローンを用いた災害救助などが挙げられる．前者は実際に存在する商品の一機能として既に実装されており，話者の方向をデバイスのランプで示す技術として用いられている．後者は災害時における要救助者発見のための技術として研究されており，暗所や瓦礫の中に要救助者がいる場合での運用デモンストレーションが既に行われている．上記のシステムのように，「現実世界の聴覚機能をロボットに対して実装すること」を目指しているロボット聴覚と呼ばれる分野が近年研究者の注目を集めている [5].

実際にマイクロホンアレイが実社会で活用される際，マイクロホンアレイ内のパラメタや外部環境を知ることは非常に重要な課題である．外部の環境は音源信号がマイクロホンアレイに収録するまでの過程に影響を及ぼし，内部のパラメタは音源定位などのアルゴリズムに既知として用いられるためその誤差が性能に大きな影響を及ぼすためである．主に音源定位などの研究分野では外部環境に対してロバストな推定を行うことができるようなアルゴリズムの開発が試みられている．一方で，マイクロホンアレイ内のパラメタ推定では主に以下の 2 つの分野が研究されている．

1. マイクロホン位置もしくは伝達関数の推定
2. 非同期信号の時刻ずれの推定

図 1 のように，マイクロホン位置や伝達関数が事前測定された値とずれてしまった場合，音源信号がマイクロホンアレイに収録される時間がずれてしまい，音響信号処理の性能に対して影響を与えることがある．信号が同期されていない場合も同様で，収録される時間のずれが以降の信号処理に対して悪影響をもたらす．同時に上記 2 点の同時解決を試みる手法などの提案も行われているが，音源信号の性質に関して制約を設けていることが多い．具体的には，単一音源であり，且つ，拍手音や TSP 信号などの立ち上がりの明確な音が主に用いられることが多い．立ち上がりの明確な音を用いることによって，音源信号がマイクロホンアレイに収録されるまでの時間が一意に定まるようになるためである．

*連絡先：東京工業大学工学院システム制御系
〒 152-8552 東京都目黒区大岡山 2-12-1 W8W-W310
E-mail: itoyama@ra.sc.e.titech.ac.jp
本稿は IEA/AIE2020 で採択された “Calibration of a Microphone Array Based on a Probabilistic Model of Microphone Positions” を和訳・一部修正したものである

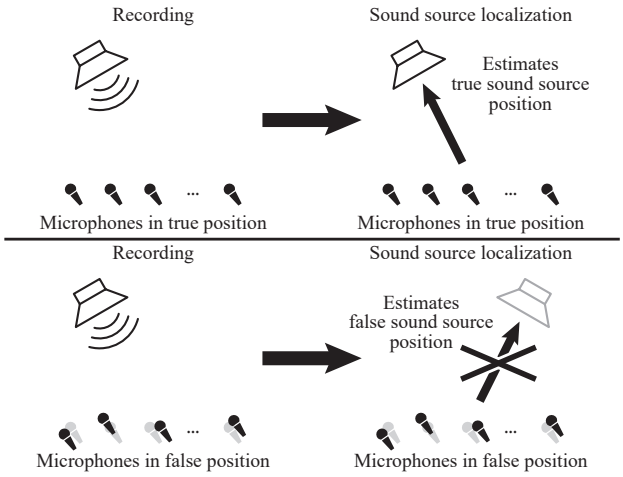


図 1: 真の位置にある各マイクロホンによって収録された音響信号を利用した場合，正しい音源定位結果が得られる．一方，誤った位置にあるマイクロホンによって収録された音響信号を利用した場合，音源定位結果は不正確なものとなる．

使いやすいキャリブレーションには，大きく分けて 2 つの条件が必要である．

1. 複数の音源が同時に収録された信号を用いたキャリブレーション：環境音などの音源は音の発生タイミングをコントロールできないため，同時収録された信号を利用できるキャリブレーションが必要である．
2. 任意の音源信号を使用したキャリブレーション：キャリブレーションの実施のために特定の音源を準備することなく実施できる手法が必要である．しかし，これまでのほとんどの研究においては，マイクロホンアレイ信号処理の前に面倒なキャリブレーション手順を必要とするキャリブレーション手法が提案されていた [6–9].

本論文では上記の 2 条件を実現したモデルを提案し評価することを目的とし，確率的生成モデルを用いた最大事後確率推定によるマイクロホン位置のキャリブレーション手法を提案する．従来の手法とは異なり，マイクロホン位置のずれや収録時に混ざるノイズ，未知の音源信号などの複数の乱雑さに対応するため，確率的なフレームワークからのアプローチを試みる．キャリブレーションには、環境音 (ホワイトノイズ) や音声などの複数の同時音源の収録信号を使用し，環境音などの立ち上がりが良くない音源信号を使用してもキャリブレーションを行うことができることを確認する．

2 関連研究

マイクロホンアレイの伝達関数の推定には主に二つの流れが存在する．同期マイクロホンアレイを対象とした手法と，非同期分散マイクロホンアレイを対象とした手法である．同期マイクロホンアレイの最初のテストは Thrun S [10] によって報告された．彼は音源信号の開始タイミングを使用したオンラインキャリブレーション手法を提案し，実際のマイクロホン装置を使用してその有用性を実証した．しかし，音源の位置が事前に決定されており，マイクロホンが完全に同期されているという制約が存在した．これらの制約を克服するために，三浦 *et al.* はロボットの自己位置とマップの同時推定を行う手法である SLAM(Simultaneous Localization And Mapping) に基づく非同期オンラインマイクロホン位置の推定法を提案した [7]．SLAM のロボット位置とマップ位置を，音源位置とマイクロホン位置に置き換えた手法である．拍手音を収録することにより，8ch マイクロホンアレイのマイクロホン位置を段階的にキャリブレーションすることに成功した．また，アドホックマイクロホンアレイについては，Raykar *et al.* と Ono *et al.* が別々に提案を行っている．彼らは，録音と再生が可能なデバイスを用意し，到来時間差 (TDOA: Time difference Of Arrival) [8,9] を使用して距離測定を行った．デバイス間の時間同期を使用し，機器間でお互いに発信した音を録音することによりマイクロホン位置のキャリブレーションを実現した．また，到来時間 (TOA: Time Of Arrival) でのキャリブレーションにおいて，近年提案された双線形アプローチなども応用されるようになった [11]．また，[8,12,13] で示されているように，MDS (Multidimentional Scaling) アルゴリズムを用いることで，マイクロホン位置を推定することをマイクロホン間の距離行列を推定することに帰着する手法も提案されている．[12] で提案されている手法は，Basis-point MDS と呼ばれる修正 MDS アルゴリズムを使用しており，推定すべき距離の数を減らすことができる．

上記のキャリブレーション方法では，いずれも音源信号に対して

- 音の鳴り始めの時刻が明確に分かる (すなわち TOA もしくは TDOA を波形から得ることができる)
- 音源信号のスペクトルがスパースである

という仮定が与えられていた．すなわち，鳴り始めの時刻が明示的には分からない音や複数の音源が同時に鳴動するような音などといった音響信号処理に用いる音でキャリブレーションを行うことができないという課題が存在している．

第 1 章で述べたように，使いやすいキャリブレーショ

ンには満たすべき 2 つの要件がある．本論文では，複数のランダムな要因を同じフレームワークで扱うために，確率を用いてマイクロホン位置を推定することを試みる．具体的には，マイクロホン位置と収録スペクトル，音源スペクトルに事前分布を仮定することによって確率的生成モデルを構築する．そして，構築した確率的生成モデルを用いて，最大事後確立推定によりマイクロホン位置のキャリブレーション手法の提案を行う．

3 提案手法

マイクロホンアレイを用いた録音には複数の乱雑さや未知の要因が存在する．提案手法では以下の 3 点を単一の尺度で扱うべく，確率的なフレームワークにおいて確率的生成モデルの構築を行った．

- 音源信号が未知
- マイクロホン位置の想定位置からのずれ
- 観測ノイズがランダムに発生

データの流れは以下のようになる．

1. 収録音を短時間フーリエ変換 (short-time Fourier transform, STFT) し，観測スペクトルを取得
2. 観測スペクトルとマイクロホン位置の基準位置を用いて推定音源スペクトルを導出
3. 観測スペクトルと推定音源スペクトルを用いて推定マイクロホン位置を導出
4. 観測スペクトルと推定マイクロホン位置を用いて推定音源スペクトルを導出
5. 3. に戻り，推定マイクロホン位置が収束するまで反復

3.1 問題設定

$\boldsymbol{x}_m \in \mathbb{R}^d$ を d 次元空間（$d = 2$ or 3）で M チャネルマイクロホンアレイを構成する m 番目のマイクロホン位置の座標とする．これらのマイクロホンは所与の基準位置 $\bar{\boldsymbol{x}}_m \in \mathbb{R}^d$ に従って配置されるが，実際の位置は基準位置からずれている．提案手法の目標は，マイクロホンアレイによって収録された音響信号を利用して，各マイクロホン位置 $\boldsymbol{x}_m$ を推定することである．提案手法では以下の 2 点を仮定する．

1. マイクロホン位置および音源信号の位置は時不変である．
2. 伝達関数は時不変でマイクロホン位置の関数として与えられる（例えば [3, 14]）．

N を音源数とし，$s_{nft} \in \mathbb{C}$ を n 番目の音源信号を STFT して得られる，f 番目の周波数ビン（$f = 1, \ldots, F$），t 番目の時間フレーム（$t = 1, \ldots, T$）における複素スペクトルとする．$z_{mft} \in \mathbb{C}$ をマイクロホンアレイを構成する m 番目のマイクロホンで録音された混合音の f 番目の周波数ビン，t 番目の時間フレームにおける複素スペクトルとする．以下で表される伝達関数

$$\boldsymbol{r}_{nf} = (r_{n1f}, \ldots, r_{nMf})^T \tag{1}$$

によって，n 番目の音源の音源スペクトルと m 番目のマイクロホンの観測スペクトルの関係は以下の式で表現される．

$$z_{mft} = \sum_n r_{nmf} s_{nft} + \epsilon_{mft} \tag{2}$$

ϵ_{mft} は m 番目のマイクロホン，f 番目の周波数ビン，t 番目の時間フレームにおける観測ノイズを表す．

3.2 確率的生成モデル

音源信号の生成や伝達過程に関する複数の乱雑さを扱うために，乱雑さが確率的に生成されると考え，混合音スペクトルが観測される過程をモデル化する．以下の 3 つの乱雑さを含む要因を考える．

- マイクロホンの位置 $\boldsymbol{x}_m \in \mathbb{R}^d$, $\boldsymbol{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_M)$
- 音源スペクトル $s_{nft} \in \mathbb{C}$, $\boldsymbol{S} = (s_{nft})_{n=1,\ldots,N,\, f=1,\ldots,F,\, t=1,\ldots,T}$
- 観測ノイズを含む混合音スペクトル $z_{mft} \in \mathbb{C}$, $\boldsymbol{Z} = (z_{mft})_{m=1,\ldots,M,\, f=1,\ldots,F,\, t=1,\ldots,T}$

キャリブレーションの問題は同時事後確率 $p(\boldsymbol{X}, \boldsymbol{S}|\boldsymbol{Z})$[1] の MAP 推定として表現できる．提案手法の目的はマイクロホン位置のキャリブレーションであるため，音源スペクトル $\boldsymbol{S}$ の推定は必ずしも必要ではないが，MAP 推定によって副次的に得られる．すなわち提案手法では，キャリブレーションと同時に音源分離も行われることになる．

$p(\boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{S})$ を全てのランダム要素の同時確率分布とする．音源スペクトルとマイクロホン位置は独立な事象であるため，同時確率は以下のように分解することができる．

$$p(\boldsymbol{Z}, \boldsymbol{X}, \boldsymbol{S}) = p(\boldsymbol{Z}|\boldsymbol{X}, \boldsymbol{S}) p(\boldsymbol{X}) p(\boldsymbol{S}). \tag{3}$$

第 1 項 $p(\boldsymbol{Z}|\boldsymbol{X}, \boldsymbol{S})$ は，観測スペクトル $\boldsymbol{Z}$ の条件付確率を表す．各マイクロホンでの観測ノイズは時間，周波数，チャネルにおいて独立であるという仮定の下で，$p(\boldsymbol{Z}|\boldsymbol{X}, \boldsymbol{S})$ を以下のように分解する．

$$p(\boldsymbol{Z}|\boldsymbol{S}, \boldsymbol{X}) = \prod_m \prod_f \prod_t p(z_{mft}|s_{1ft}, \ldots, s_{Nft}, \boldsymbol{X}) \tag{4}$$

[1] $\boldsymbol{S}$ を周辺化し，$p(\boldsymbol{X}|\boldsymbol{Z})$ の MAP 推定を行うことがより望ましい．

観測ノイズ ϵ_{mft} は平均が 0 で分散が σ_z^2 の円対象複素ガウス分布に従うとする．式 (2) より，観測スペクトル z_{mft} は平均 $\sum_n r_{nmf} s_{nft}$，分散 σ_z^2 の複素ガウス分布に従う．すなわち，以下のように表される．

$$p(z_{mft}|s_{1ft},\ldots,s_{Nft},\boldsymbol{X}) \propto \exp\left(-\frac{|z_{mft}-\sum_n r_{nmf}s_{nft}|^2}{\sigma_z^2}\right) \quad (5)$$

$(\cdot)^\star$ は複素共役を表す．

第 2 項 $p(\boldsymbol{X})$ はマイクロホン位置の事前分布を表す．マイクロホンは所与の基準位置に配置されるが，実際の位置には製造誤差や取り付ける際の誤差などの不確実性による基準位置からのずれが含まれる．このずれは各マイクロホンに独立であるとすると，$p(\boldsymbol{X})$ は以下のように分解できる．

$$p(\boldsymbol{X}) = \prod_m p(\boldsymbol{x}_m) \quad (6)$$

各マイクロホンの位置のずれは等方向的な正規分布 (分散が σ_x^2) に従うとすると，各マイクロホン位置の事前分布は以下で表現される．

$$p(\boldsymbol{x}_m) \propto \exp\left(-\frac{\|\boldsymbol{x}_m-\bar{\boldsymbol{x}}_m\|_2^2}{2\sigma_x^2}\right) \quad (7)$$

第 3 項 $p(\boldsymbol{S})$ は音源スペクトルの分布を表す．音源スペクトルは音源ごとに独立しているとする．単一音源のスペクトルの事前分布の表現に関しては，様々な表現が提案されてきた．例えば，非負値行列因子分解を用いた低ランク表現 [15,16] や，深層学習 [17,18] を使用した非線形表現が提案されている．ここでは，時間周波数平面での音源スペクトルの独立性を仮定し，$p(\boldsymbol{S})$ を以下のように分解する．

$$p(\boldsymbol{S}) = \prod_n \prod_f \prod_t p(s_{nft}) \quad (8)$$

各時間周波数スロットにおける音源スペクトル s_{nft} は等方的な複素ガウス分布に従うとすると，その分布は以下のように表現される．

$$p(s_{nft}) \propto \exp\left(-\frac{|s_{nft}|^2}{\sigma_s^2}\right) \quad (9)$$

3.3 キャリブレーションアルゴリズム

前節で述べた確率的生成モデルに基づく，マイクロホンアレイのキャリブレーションアルゴリズムを構築する．与えられた観測スペクトル $\boldsymbol{Z}$ の最適なマイクロホン位置 $\boldsymbol{X}$ は対数事後確率の最大化により得られる．

$$\hat{\boldsymbol{X}},\hat{\boldsymbol{S}} = \arg\max_{\boldsymbol{X},\boldsymbol{S}} p(\boldsymbol{X},\boldsymbol{S}|\boldsymbol{Z}) = \arg\max_{\boldsymbol{X},\boldsymbol{S}} \log p(\boldsymbol{Z},\boldsymbol{X},\boldsymbol{S}) \quad (10)$$

マイクロホン位置 $\boldsymbol{X}$ と音源スペクトル $\boldsymbol{S}$ の事後確率は独立ではないため，事後確率を最大化する $\boldsymbol{X}$ と $\boldsymbol{S}$ を同時に求めるのは困難である．本稿では $\boldsymbol{X}$ と $\boldsymbol{S}$ に関して反復的に事後確率を最大化することで，式 (10) を近似的に実現する．

マイクロホン位置 $\boldsymbol{X}$ に関する事後確率最大化はグリッドサーチによって実現する．未知の関数によって $\boldsymbol{X}$ から伝達関数 r_{nmf} への変換はなされると想定しているため，対数事後確率の $\boldsymbol{X}$ に関する導関数も得られない．グリッドサーチの範囲とグリッドの間隔は事前に定義する．

一方, 音源スペクトル $\boldsymbol{S}$ に関する事後確率最大化は解析的に導出可能である．対数事後確率は $\boldsymbol{S}$ に関して上に凸であるため, 偏導関数の零点を解くことで最適な音源スペクトル $\hat{s}_{nft}$ は得られる．観測スペクトルが $\boldsymbol{z}_{ft} = (z_{1ft},\ldots,z_{Mft})$，伝達関数が $\boldsymbol{r}_f = (\boldsymbol{r}_{1f},\ldots,\boldsymbol{r}_{Nf})$ と表されるとき，推定音源スペクトルは以下のように表される（$\boldsymbol{I}$ は $N\times N$ の単位行列）．

$$\hat{s}_{ft} = \left\{\left(\boldsymbol{r}_f^T\boldsymbol{r}_f^\star\right)^T + \frac{\sigma_z^2}{\sigma_s^2}\boldsymbol{I}\right\}^{-1}\left(\boldsymbol{z}_{ft}^T\boldsymbol{r}_f^\star\right)^T. \quad (11)$$

構築したアルゴリズムを Algorithm 1 に示す．

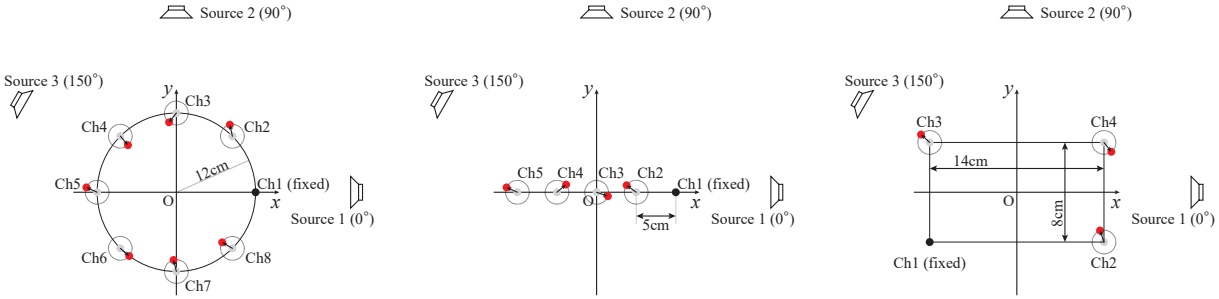
Algorithm 1 Iterative Estimation of $\boldsymbol{X}$ and $\boldsymbol{S}$

Initialize $\boldsymbol{X}^{(0)}$ and set $t \leftarrow 0$
repeat
  $\boldsymbol{S}^{(t+1)} \leftarrow \arg\max_S \log p(\boldsymbol{X}^{(t)},\boldsymbol{S}|\boldsymbol{Z})$ using Eq. (11)
  $\boldsymbol{X}^{(t+1)} \leftarrow \arg\max_X \log p(\boldsymbol{X},\boldsymbol{S}^{(t+1)}|\boldsymbol{Z})$ using grid search
  $t \leftarrow t+1$
until convergence

4 評価実験

提案手法を用いてマイクロホンアレイを構成するマイクロホンの位置のずれに対するキャリブレーションを行いその性能を評価した．性能評価の尺度には，キャリブレーションで推定された位置と真の位置の誤差を用いる．

実験はシミュレーション環境で行った．収録音のサンプリング周波数は 24 kHz，STFT の窓幅は 1024 点，シフト幅は 256 点とした．また，$\sigma_z^2 = 5\times10^{-10}$，$\sigma_s^2 = 5\times10^{-8}$ とした．σ_x^2 は与えたズレの大きさの二乗とした．グリッドサーチにおけるグリッドの大きさは 0.1 cm，グリッドサーチを行う範囲は与えた変位の大きさに比例して変化させた．音源信号については，ホ



(a) 8 チャネル円形アレイ (C8)　(b) 5 チャネル線型アレイ (L5)　(c) 4 チャネル長方形アレイ (R4)

図 2: シミュレーション実験に用いたマイクロホンアレイの形状と音源の配置

ワイトノイズ，音声データを用いた．音声は JVS corpus [19] からランダムに選択した．

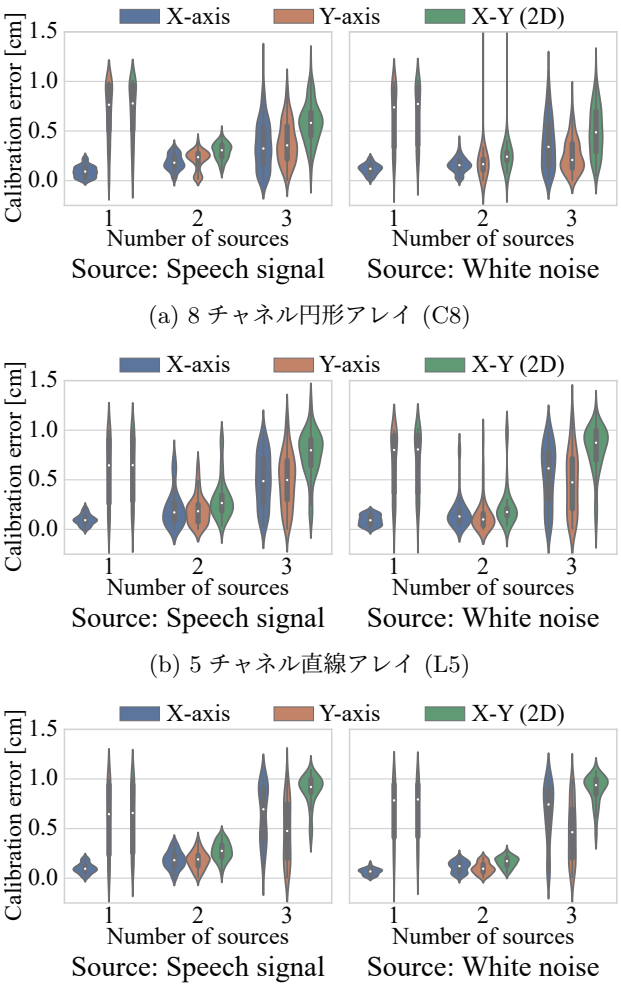
マイクロホンアレイは 8 チャネル円形，5 チャネル直線，4 チャネル長方形の 3 種類を用いた．以下では，それぞれのマイクロホンアレイを **C8**, **L5**, **R4** と呼ぶ．マイクロホンアレイの形状を図 2 に示す．マイクロホンアレイ C8 は半径 12 cm の円形で，45° おきにマイクロホンを配置した．マイクロホンアレイ L5 は長さ 20 cm の直線形で，5 cm おきにマイクロホンを配置した．マイクロホンアレイ R4 は 14 cm × 8 cm の長方形で，各頂点にマイクロホンを配置した．いずれのマイクロホンアレイも 2 次元平面上で構築され，音源もマイクロホンアレイと同一の平面上に配置したため，次元数は $d = 2$ と設定した．音源信号は平面波であると仮定し，到来方向は 0°, 90°, 150° とした．

本性能評価では，以下の二つの要素を変化させる．

- 音源数：ズレの大きさを 1 cm に固定した上で音源数を 1 (0°), 2 (0°, 90°), 3 (0°, 90°, 150°) とした際の性能の評価を行った．
- 与えるズレの大きさ：音源数は 2 つに固定し，Ch1 を除くマイクロホンに対して，1 cm, 2 cm, 3 cm のズレを与えた際の性能の評価を行った．

4.1　音源数の変化に対する評価

図 3 に音源数の変化に対するキャリブレーション誤差を示す．いずれのマイクロホン形状，音源の種類に関しても，2 音源を用いた場合が最もキャリブレーション誤差が小さかった．1 音源のみを用いた場合は，X 軸方向のキャリブレーション誤差は小さいものの，Y 軸方向のキャリブレーションがほとんど行われていない．この場合は音源が X 軸上に存在するため，その方向のキャリブレーションのみが行われ，直交する Y 軸方向のキャリブレーションは行われないためだと解釈できる．3 音源を用いた場合は，2 音源を用いた場合に比べ



(a) 8 チャネル円形アレイ (C8)

(b) 5 チャネル直線アレイ (L5)

(c) 4 チャネル長方形アレイ (R4)

図 3: 音源数の変化に対するキャリブレーション誤差．X 軸方向，Y 軸方向，X-Y 平面上での誤差の分布を表す．

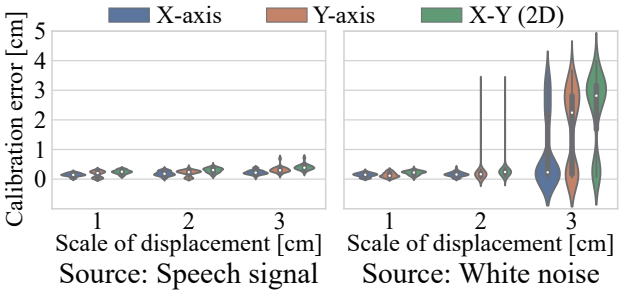


図 4: 8 チャネル円形アレイ (C8) を用いた場合のマイクロホンのずれの大きさに対するキャリブレーション誤差.

てキャリブレーション誤差が増大する傾向にある．音源数が増加すると，キャリブレーションに利用できる音源方向に関する情報は増えるものの，音源分離（音源スペクトルの推定）が困難になるため，誤差が増大したと考えられる．ただし，本実験ではすべての音源が常に発音していたが，実際には音源の発音区間は時間的にスパースになることが多いため，実環境である程度長時間の音響信号を用いればこの問題は解決される可能性がある．

マイクロホンアレイの形状を変化させても，キャリブレーション誤差の傾向に大きな違いは確認されなかった．したがって，提案手法は様々なマイクロホンアレイ形状に適用可能であるといえる．

4.2 ずれの大きさに対する評価

図 4 にマイクロホンアレイ C8 を用いた場合のマイクロホン位置のずれの大きさに対するキャリブレーション誤差を示す．音源信号に音声を用いた場合は，ずれの大きさ 1 cm, 2 cm, 3 cm に対してキャリブレーション後の誤差の中央値が 0.23 cm, 0.35 cm, 0.37 cm であり，ずれを 77 %, 83 %, 88 % 減少させることができた．一方ホワイトノイズを用いた場合は，ずれの大きさが 1 cm–2 cm のときは音声を用いた場合と同様の傾向を示しているが，ずれの大きさが 3 cm のときにキャリブレーション誤差が大きく増大した．図 4 を見ると，0.5 cm 付近と 3 cm 付近に誤差の分布が 2 極化していることが分かる．複数のホワイトノイズが重複するとスペクトログラムの時間周波数上でのスパース性がほぼ成り立たず，音源分離（音源スペクトルの推定）が困難になるため，誤差が増大したと考えられる．

4.3 音源定位性能の評価

マイクロホンアレイ C8 を用いてキャリブレーション前後で音源定位を行った際の推定誤差を図 5 に示す．

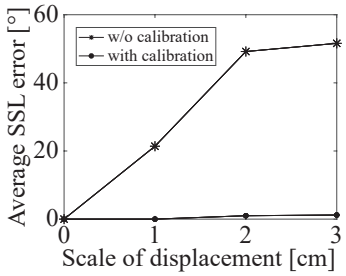


図 5: キャリブレーション前後の音源定位誤差の変化.

音源数は 2 に固定し，音源信号には音声を用いた．キャリブレーションを行わない場合はマイクロホン位置ずれの増大に伴い音源定位誤差も増大するが，キャリブレーションにより定位誤差は 2° 以下，キャリブレーション前の 4 % 以下に抑制された．上述の実験で示したように，提案手法はマイクロホン位置の真値を完璧に推定するわけではないが，音源定位性能を十分に改善することが示された．

5 まとめ

本論文では，マイクロホン位置に関する確率的生成モデルを用いたマイクロホンのキャリブレーションを提案した．確率的生成モデルを構築し，MAP 推定によるキャリブレーション手法を構築し，提案手法の性能評価のため，評価実験を行った．評価の結果，複数の同時音源を用いたキャリブレーションが実行可能であることが確認された．また，提案法によるキャリブレーションの結果，音源定位の性能向上を確認し，有効性を示した．

本論文ではグリッドサーチしか試していないため，他の最適化手法も検討してみる必要がある．また，マイクロホン位置の推定から伝達関数の推定にまで手法を拡張する必要性もある．

謝辞

本研究は JSPS 科研費 JP19K12017, JP19KK0260 および JP20H00475 の助成を受けた．

参考文献

[1] Zhang, C., Florencio, D., Ba, D. E. and Zhang, Z.: Maximum likelihood sound source localization and beamforming for directional microphone arrays in distributed meetings, *IEEE Trans. Multimedia*, Vol. 10, No. 3, pp. 538–548 (2008).

[2] Nishiura, T., T., Y., S., N. and K., S.: Localization of multiple sound sources based on a CSP analysis with a microphone array, in *ICASSP 2000*, Vol. 2, pp. 1053–1056 (2000).

[3] Nakamura, K., Nakadai, K., Asano, F., Hasegawa, Y. and Tsujino, H.: Intelligent sound source localization for dynamic environments, in *IROS 2009*, pp. 664–669 (2009).

[4] Nakadai, K., Nakajima, H., Hasegawa, Y. and Tsujino, H.: Sound source separation of moving speakers for robot audition, in *ICASSP 2009*, pp. 3685–3688 (2009).

[5] Nakadai, K., Okuno, H. G. and Mizumoto, T.: Development, deployment and applications of robot audition open source software HARK, *J. Robot. Mechatron.*, Vol. 29, No. 1, pp. 16–25 (2017).

[6] Su, D., Vidal-Calleja, T. and Miro, J. V.: Simultaneous asynchronous microphone array calibration and sound source localisation, in *IROS 2015*, pp. 5561–5567 (2015).

[7] Miura, H., Yoshida, T., Nakamura, K. and Nakadai, K.: SLAM-based online calibration of asynchronous microphone array for robot audition, in *IROS 2011*, pp. 524–529 (2011).

[8] Raykar, V. C., Kozintsev, I. V. and Lienhart, R.: Position calibration of microphones and loudspeakers in distributed computing platforms, *IEEE Trans. Speech and Audio Process.*, Vol. 13, No. 1, pp. 70–83 (2005).

[9] Ono, N., Shibata, K. and Kameoka, H.: Self-localization and channel synchronization of smartphone arrays using sound emissions, in *APSIPA ASC 2016*, pp. 1–5 (2016).

[10] Thrun, S.: Affine structure from sound, in *NIPS'05*, pp. 1353–1360 (2005).

[11] Crocco, M., Del Bue, A. and Murino, V.: A bilinear approach to the position self-calibration of multiple sensors, *IEEE Trans. on Signal Process.*, Vol. 60, No. 2, pp. 660–673 (2012).

[12] Birchfield, S. T. and Subramanya, A.: Microphone array position calibration by basis-point classical multidimensional scaling, *IEEE Trans. on Speech and Audio Process.*, Vol. 13, No. 5, pp. 1025–1034 (2005).

[13] Birchfield, S. T.: Geometric microphone array calibration by multidimensional scaling, in *ICASSP 2003*, Vol. 5, pp. 157–160 (2003).

[14] Valin, J.-M., Rouat, J. and Michaud, F.: Enhanced robot audition based on microphone array source separation with post-filter, in *IROS 2004*, Vol. 3, pp. 2123–2128 (2004).

[15] Févotte, C., Bertin, N. and Durrieu, J.-L.: Nonnegative matrix factorization with the Itakura-Saito divergence: With application to music analysis, *Neural Computation*, Vol. 21, No. 3, pp. 793–830 (2009).

[16] Smaragdis, P., Févotte, C., Mysore, G. J., Mohammadiha, N. and Hoffman, M.: Static and dynamic source separation using nonnegative factorizations: A unified view, *IEEE Signal Process. Mag.*, Vol. 31, No. 3, pp. 66–75 (2014).

[17] Uhlich, S., Giron, F. and Mitsufuji, Y.: Deep neural network based instrument extraction from music, in *ICASSP 2015*, pp. 2135–2139 (2015).

[18] Nugraha, A., Liutkus, A. and Vincent, E.: Multichannel audio source separation with deep neural networks, *IEEE/ACM Trans. Audio, Speech, Language Process.*, Vol. 24, No. 9, pp. 1652–1664 (2015).

[19] Takamichi, S., Mitsui, K., Saito, Y., Koriyama, T., Tanji, N. and Saruwatari, H.: JVS corpus: free Japanese multi-speaker voice corpus, *arXiv:1908.06248 [cs.SD]* (2019).