

マイクロホンと音源位置に関する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の検討

段雄啓¹, 糸山克寿¹, 西田健次¹, 中臺一博^{1,2}

¹ 東京工業大学工学院システム制御系

² (株) ホンダ・リサーチ・インスティテュート・ジャパン

1 はじめに

近年、音源定位などマイクロホンアレイ音響信号処理技術が身近なものとなっている。マイクロホンアレイを用いた音源定位のアルゴリズムの多くは、音源からマイクロホンアレイへの音の伝わり方を表す伝達関数を事前に用意する必要がある。伝達関数はマイクロホンアレイの幾何的形狀から計算することができるが、アレイを都度設置する場合やアレイが剛性の低い物体上に実装されている場合などでは幾何的形狀の変化により伝達関数の精度が低下してしまう。本稿では、アレイの幾何的形狀が事前に設定したものからずれてしまう場合を想定し、アレイ位置、音源スペクトル、収録音スペクトルの確率モデルに基づくアレイ位置の校正手法を提案する。

2 関連研究

マイクロホンアレイの伝達関数推定の研究は、主に非同期分散マイクロホンアレイやアドホックマイクロホンアレイの分野で精力的に行われている。非同期分散マイクロホンアレイ研究では、ロボットの自己位置と地図の同時推定問題として知られる SLAM (Simultaneous Localization And Mapping) の手法をマイクロホンと音源の位置を同時推定する問題とみなして拡張した手法が報告されている [1]。アドホックマイクロホンアレイ研究では、機器から音を出力させることにより、機器自体を音源と見立て、同期後の録音信号の到達時間差による機器間距離の推定法が提案されている [2]。一方、各マイクロホンの完全同期収録を前提としたオンライン校正法も提案されている [3]。上記の研究 [1-3] では、拍手音などの立ち上がりが明確な音を用いることによってマイクロホン位置や伝達関数の校正を行っている。本稿で提案する手法は音の伝達過程に関する確率モデルを導入することにより、立ち上がりが明確な音を用いることなく校正を実現する。

3 問題設定

N 個の音源を M 本のマイクロホンからなるマイクロホンアレイで収録するものとし、 n 番目の音源からの信号を s_n ($1 \leq n \leq N$)、 m 番目のマイクロホンの収録音を z_m ($1 \leq m \leq M$) とする。 s_n, z_m を短時間フーリエ変換して得られる複素スペクトルを S_{nft}, Z_{mft} とする。ここで、添え字の f は周波数ビン、 t はフレームを表す。以降、時間周波数空間で議論を行う。音源信号 S_{nft} がマイクロホンアレイで観測されたときの複素スペクトルを $Z_{ft} = (Z_{1ft}, \dots, Z_{Mft})^T$ 、伝達関数を $R_{nft} = (R_{n1f}, \dots, R_{nMf})^T$ (ただし、 R_{nft} は m 番目のマイクロホンと音源間の伝達関数) とすると、 $Z_{ft} = \sum_n R_{nft} S_{nft}$ が成り立つ。また、マイクロホンアレイの m 番目のマイクロホン

の位置を $X_m = (x_m, y_m)^T$ 、マイクロホンアレイの位置情報を $X = (X_1^T, \dots, X_M^T)^T$ とする。マイクロホンアレイの収録音をもとに、マイクロホンアレイ位置情報 X を校正する。

4 提案手法

実際の音源スペクトルや信号伝達過程にはランダムなノイズやゆらぎが含まれるため、第3節で述べたスペクトル伝達過程の確率的な表現である確率的生成モデルを用いる。確率的生成モデルは、収録音スペクトル、音源スペクトル、マイクロホン位置のモデルからなる。定義された確率的生成モデルに基づいて、収録音スペクトルが与えられた際的事後確率を最大化するようなマイクロホン位置、すなわち最大事後確率 (MAP: Maximum A Posteriori) 推定値を出力する。

4.1 確率的生成モデル

音源スペクトルを S (S_{nft} の n, f, t 全てに関する集合)、収録音のスペクトルを Z (Z_{ft} の f, t 全てに関する集合) とする。次に、 X の事前確率を $P(X)$ 、 S の事前確率を $P(S)$ とする。 X, S が与えられたときの Z の条件付き確率を $P(Z|S, X)$ とすると、 $P(X, S, Z)$ で X, S, Z の同時確率が表される。

$$\log P(X, S, Z) = \log P(Z|S, X) + \log P(S) + \log P(X) \quad (1)$$

式 (1) の第一項について説明する。収録音は S_{nft} と伝達関数 R_{nft} の積を平均とした複素正規分布に従うと仮定すると、 $\log P(Z|S, X)$ は以下のように表現される。 σ_Z^2 はマイクロホンアレイの収録音の分散を表す。

$$\log P(Z|S, X) \stackrel{c}{=} - \sum_m \sum_f \sum_t \frac{(Z_{mft} - \sum_n R_{nft} S_{nft})^* (Z_{mft} - \sum_n R_{nft} S_{nft})}{\sigma_Z^2} \quad (2)$$

なお、 $*$ は複素共役を表す演算子、 $\stackrel{c}{=}$ は定数項を除いて右辺と左辺が等しいことを表す等号である。

式 (1) の第二項について説明する。 S は平均 0 の複素正規分布に従うと仮定すると、 $\log P(S)$ は以下のように表現される。 σ_S^2 は音源スペクトルの分散を表す。

$$\log P(S) \stackrel{c}{=} - \sum_n \sum_f \sum_t \frac{S_{nft}^* S_{nft}}{\sigma_S^2} \quad (3)$$

式 (1) の第三項について説明する。マイクロホン相対位置は各マイクロホンで独立であり、マイクロホン位置の変位は等方的であると仮定する。すなわち、 \bar{X} をマイクロホンアレイの基準位置情報と定義し、 X は \bar{X} を平均とした正規分布に従うとすると、 $\log P(X)$ は以下のように表現される。ここで、 σ^2 はマイクロホンアレイ位置の分散、 \bar{X}_m はマイクロホン m の基準位置である。

$$\log P(X) = \sum_m \log P(X_m) \stackrel{c}{=} - \sum_m \frac{(X_m - \bar{X}_m)^T (X_m - \bar{X}_m)}{2\sigma^2} \quad (4)$$

A study on calibration of microphone array based on probabilistic model on microphone and sound source position.

¹ Department of Systems and Control Engineering, School of Engineering, Tokyo Institute of Technology

² Honda Research Institute Japan Co., Ltd.

4.2 マイクロホンアレイの校正

Z に対する S と X の対数事後確率 $\log P(X, S|Z)$ を最大化する \hat{X} を推定する. \hat{X} と S の事後確率は独立ではないため, 以下のアルゴリズムを用いて, S との反復推定により MAP 推定を実現する. ベイズの定理により, このアルゴリズムの 2. と

Algorithm 1 X, S の反復推定

```

1.  $X^{(0)}, S^{(0)}$  をランダムに初期化.
repeat
  2.  $X^{(t+1)} \leftarrow \operatorname{argmax}_X \log P(X, S^{(t)}|Z)$ 
  3.  $S^{(t+1)} \leftarrow \operatorname{argmax}_S \log P(X^{(t+1)}, S|Z)$ 
until 収束するまで
    
```

3. の計算は $\log P(X, S, Z)$ を最大化する X と S を求めることに帰着される. $\log P(X, S, Z)$ を S_{nft} の関数とみなすと, 凸性が保証され極値が一つに定まるため, 音源スペクトル S_{nft} の極値は $\log P(X, S, Z)$ の S_{nft} についての偏導関数の零点を求めることで得られる. X の極値を解析的に求めることは困難であるため, X についての推定を行う必要がある. そこで, グリッドサーチにより

$$\hat{X} = \operatorname{argmax}_X \log P(X, S, Z) \tag{5}$$

$$= \operatorname{argmax}_X (\log P(Z|S, X) + \log P(X)) \tag{6}$$

を最大化するような X を求める.

5 評価実験

提案手法による X の推定精度をシミュレーションで評価する. 8 ch マイクロホンアレイのうち 1 ch のみに変位を与えたときの収録音を用いて, 提案手法によりマイクロホン位置を推定する. 音源数を変化させて実験を行う. 変位を与えたマイクロホンの位置と提案手法による推定位置の距離を用いて音源数が推定結果に与える影響を評価する.

5.1 実験環境

実験環境を Fig. 1 に示す. マイクロホンアレイの基準位置は, 半径 $r = 0.12[m]$ の円形状に 45° おきに配置した. チャンネルの番号は $\theta = 0^\circ$ を 1 ch とし, 反時計回りに 2 ch, ..., 8 ch とした. 音源の方向は 0° と 90° の 2 通りで音源位置は無限遠, つまり平面波を仮定している.

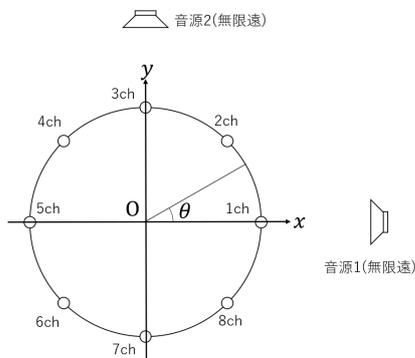


Fig. 1 実験環境 (数値シミュレーション)

5.2 手順

提案手法のシミュレーションによる評価手順について説明する. 今回は 1 ch に $(0.02, 0.02)[m]$ の変位を与えた時の校正を試みる. Fig. 1 に示すマイクロホンアレイ座標系でマイクロホン位置の推定を行う. 変位を与えられた時の収録音に対して, $0.0005[m]$ おきに各点における対数事後確率を導出し, MAP 推定を行う. また, 2 ch, ..., 8 ch のマイクロホン位置は固定したため, $P(X_2) = \dots = P(X_8) = 1$ とする. このため $P(X) = P(X_1)$ となる. 単一音源 (音源 1) の場合をケース 1, 複数音源 (音源 1, 音源 2) の場合をケース 2 としてシミュレーションを行った. また, 音源信号は同時に収録されないものとした. 入力音はホワイトノイズを用い, 音源スペクトルの分散 σ_s^2 とマイクロホンの収録音の分散 σ_z^2 は入力音の分散と同じ 5×10^{-6} とした. $0.02[m]$ 程度の精度での配置を仮定し, 事前確率の分散 σ^2 は $4 \times 10^{-4}[m^2]$ とした.

5.3 結果

実験結果を Fig. 2 に示す. 収録音に対して事後確率が最大となるマイクロホン位置を推定位置とした.

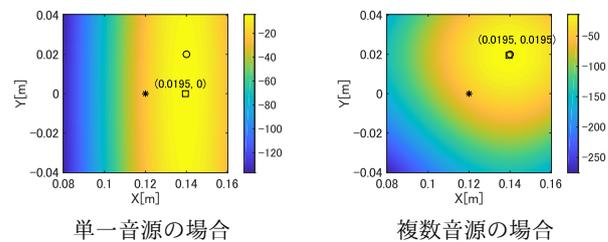


Fig. 2 収録音に対するマイクロホン位置の対数事後確率. 四角のマーカーは推定位置を, 丸のマーカーは真値 $(0.02, 0.02)$ を, *のマーカーはマイクロホンの基準位置を表す. また, 推定された座標を四角のマーカーの上に示す.

音源 1 の収録音を用いたケース 1 では x 軸方向の変位は検出できたが, y 軸方向の変位は上手く検出できなかった. それに対し, 音源 1 と 2 の収録音を用いたケース 2 では x, y 軸両方向の変位が検出できた. 0° から到来する音源 1 を用いた時に x 軸方向が推定できたことから, 単一の音源を用いた場合, その音源の方向の変位が検出可能と考えられる. 以上より, 様々な方向の音源からの収録音を用いれば精度がより向上することが予想される.

6 まとめ

本研究では, マイクロホンと音源位置に関する確率モデルに基づくマイクロホンアレイの校正の検討を行った. 音源方向の数がマイクロホン位置の推定精度に与える影響について考察を行った. 今後の課題としては, 実データでの検証や複数マイクロホンに変位がある場合での検討などが考えられる.

謝辞 本研究は, JSPS 科研費 16H02884, 16K00294, 17K00365 および JST ImPACT タフ・ロボティクス・チャレンジの助成をうけた.

参考文献

[1] H. Miura et al., "SLAM-based Online Calibration of Asynchronous Microphone Array for Robot Audition," IROS, pp.524-529, 2011.
 [2] 柴田他, "音の発信を利用したキャリブレーションに基づくアドホックマイクロホンアレイによる音源定位," 音講論集, pp.707-710, 2014.
 [3] S. Thrun, "Affine Structure from Sound," NIPS, vol.18, pp.1353-1360, 2006.