人工知能学会研究会資料
JSAI Technical Report
SIG-Challenge-B402-07

マイクアレイ伝達関数のオンライン校正とそのロボットへの適用

Online Calibration of Microphone Array Transfer Functions for Robots

中村圭佑, 中臺一博

Keisuke NAKAMURA, Kazuhiro NAKADAI

（株）ホンダ・リサーチ・インスティチュート・ジャパン

Honda Research Institute Japan Co., Ltd.

keisuke@jp.honda-ri.com, nakadai@jp.honda-ri.com

Abstract

本稿ではマイクアレイベースのロボット聴覚システムで事前情報として用いられるマイクアレイ伝達関数の校正について述べる．伝達関数は主に数値計算と計測の二つの方法によって得られるが，数値計算はロボットや部屋の形状に起因する反射や回折などを誤差なく模擬することは難しく，計測は正確であるものの時間がかかり，専門の機器を必要とする．そこで，本稿ではロボットや部屋の音響特性を含めたオンザスポット（ユーザーがその場で手軽にできる）オンライン伝達関数校正法を提案する．

1 序論

ロボットと人が自然なインタラクションを実現するには，ロボット聴覚 [1] を用いた周囲の音の聞き分けが不可欠である．実環境における人・ロボットインタラクションでは，ロボットに埋め込まれたマイクを用いることから，音源からの距離が遠く，信号対雑音比は接話マイクを使う場合よりも低い．それゆえ，音源定位や音源分離などのマイクアレイ処理はロボット聴覚において重要な役割を果たし，様々な応用がなされている [2; 3; 4; 5].

一般的にマイクアレイ処理は音源とマイク間の伝達関数を事前情報として使用するが，ロボット聴覚における応用では主に二つの方法で得られたものを使用している．一つは，マイクの位置から伝搬波モデルに基づいて幾何計算して求めた伝達関数である (幾何計算法) [2; 3]．幾何計算法はマイクが自由空間上に存在することが仮定されているが，ロボットに搭載されたマイクアレイを用いる場合は，ロボット形状に起因する反射や回折が誤差を生じるため，処理の性能を劣化してしまう．ロボットの形状を考慮した計算手法 [6; 7] も存在するが，計算コストが大きく，環境要因の音響特性 (壁からの反射等) を含めた計算ができない制約がある．もう一つは，*Time Stretched Pulse (TSP)* 信号や，M 系列信号を各方向に対して計測する方法である [8; 9]．この手法は上述の音響特性まで計測できるため，性能を確保できる [4; 5] ものの，計測に時間と手間がかかってしまう上，ロボットが移動すると音環境が計測した時と変化するため，その変化が誤差を生じ，性能が劣化してしまう．それゆえ，音環境が変化した時にユーザーが時間と手間をかけることなく伝達関数校正ができることが望ましい．

そこで，本稿ではマイクアレイ処理のロボット応用を考慮した，マイクアレイ伝達関数のオンライン校正について述べる．本稿は，特別な機器やマイク位置などの事前情報を必要とせず，十分に短いデータのみによって，マイクアレイをその場で高精度に校正する (オンザスポット校正) ことを目的とする．これまでにも，マイクアレイ周辺で移動する人の拍手音などの短い録音データを用いて，マイクアレイのオンザスポット校正が提案されてきた[10; 11; 12; 13; 14]．しかし，これらの手法は伝達関数ではなくマイク位置の校正となっており，伝達関数は伝搬波モデルを用いて幾何計算していた．この場合，上述のようにロボットや部屋の音響特性を考慮することができないため，その誤差がマイクアレイ処理性能を劣化してしまう．

そこで，本稿では，ロボットや部屋の音響特性を含めた伝達関数のオンザスポット校正を提案する．提案手法は *Frequency-domain Ordinary Least Squares (FOLS 法)* と *Frequency- and Time-Domain Linear Interpolation (FTDLI 法)*[15] から構成される．FOLS 法はマイクアレイ周囲を移動する人の拍手音から直接，伝達関数を推定する．既存手法とは違い，マイクアレイ位置を事前情報とせず，ロボットや部屋の音響特性を直接推定するため，ロボット応用に向いている．FOLS 法で得られる伝達関数は拍手音が観測された方向のみ得られるため，音源定位で使用されるような等間隔 (5◦ 毎など) に並んでいない．そこで，FTDLI 法を用いて伝達関数を補間することで，伝達関数を所望の解像度で整列させる．伝達関数を補間するため

Table 1: SLAM 問題とマイクアレイ校正問題の対応関係

SLAM(ロボット)	マイクアレイ校正
自己位置	音源位置
地図 (ランドマーク位置)	マイク位置
予測誤差最小	同期時刻ずれ推定付予測誤差最小

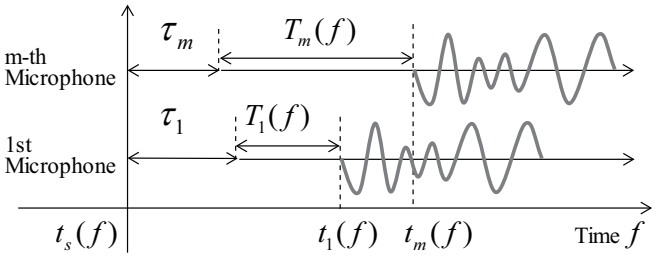


Figure 1: f 回目の拍手音到達時間

には，FOLS 法で得られた伝達関数の方向情報が必要となるため，*Simultaneous Localization And Mapping (SLAM)* に基づくオンラインマイク位置推定 (2 章参照 [13]) を導入し，幾何的に計算された伝達関数で拍手音を定位することで，方向情報を得る．最後に，提案法で推定された伝達関数を音源定位・音源分離に適用し有効性を確認する．

2 SLAM に基づくマイク位置推定

SLAM は，オンラインでロボットの自己位置と周囲の地図推定を同時に行う問題であり，ロボット分野で盛んに研究が行われている [15]．SLAM を解くため，様々なアルゴリズムが提案されているが，本稿では，*Extended Kalman Filter (EKF)* ベースの SLAM (EKF-SLAM) をマイクアレイ校正処理に適用する手法を導入する．この手法では，SLAM における地図推定を各マイクの位置推定，ロボットの自己位置推定を音源の位置推定に当てはめ，同期時刻のズレを含む推定誤差を最小になるよう更新を行うことによって校正処理を実現する (表 1 参照)．これによって，例えば，人が拍手をしながら，マイクアレイを周回することによって，マイクアレイのオンライン逐次校正を行うことができる．具体的には，M 個のマイクからなる非同期分散マイクアレイを考え，f 回目の拍手に対して，m 番目のマイクの状態ベクトル $\xi_m(f)$ (ただし，$1 \le m \le M$) と音源の状態ベクトル $\xi_s(f)$ を定義する．$\xi_m(f)$ は，2 次元のマイク位置情報と同期時刻ズレを含み，$\xi_s(f)$ は，2 次元の音源位置情報と進行方向を含む 3 次元ベクトルとして，$\xi_m(f) = [x_m(f), y_m(f), \tau_m(f)]^T$，$\xi_s(f) = [x_s(f), y_s(f), \theta_s(f)]^T$ と定義する．

2.1 観測モデル

m 番目のマイクで f 番目の拍手の到達時刻 $t_m(f)$ を観測する．m 番目のマイクの位置，および時刻ズレを (x_m, y_m)，τ_m，音速を c とすると，$t_m(f)$ は，図 1 に示されているように，音源が音を発した時刻 $t_s(f)$ を用いて，以下のように求めることができる．

$$t_m(f) = t_s(f) + T_m(f) + \tau_m \tag{1}$$

$$T_m(f) = \frac{\sqrt{(x_s(f) - x_m)^2 + (y_s(f) - y_m)^2}}{c} \tag{2}$$

音を発した時刻 $t_s(f)$ は未知であるため，基準マイク (マイク 1) での観測時刻との差をとると，観測モデルは，以下のように相対時刻で表すことができる．

$$\eta(f) = \begin{bmatrix} T_2(f) - T_1(f) + \tau_2 - \tau_1 \\ \vdots \\ T_M(f) - T_1(f) + \tau_M - \tau_1 \end{bmatrix} + \delta(f)\,, \tag{3}$$

観測誤差 $\delta(f)$ は白色雑音に従うものとする．

2.2 状態遷移モデル

音源，つまり人は，$\theta_s(l)$ 方向に等速 $v_s(l)$ で移動するとする．ここで，l は，歩数を表すインデックスであり，拍手の回数を表すインデックス f とは異なるものとする．状態は l，つまり 1 歩進むごとに更新されるため，音源の状態遷移モデルは以下のように表すことができる．

$$\xi_s(l+1) = \xi_s(l) + \begin{bmatrix} \sin(\theta_s(l)) & 0 \\ \cos(\theta_s(l)) & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} v_s(l) \\ \dot{\theta}_s(l) \end{bmatrix} \Delta t + \mathscr{W}_s(l)\,, \tag{4}$$

$\mathscr{W}_s(l)$ はモデル誤差を表す白色雑音である．一方，マイクの位置は固定であるため，モデル誤差を $\mathscr{W}_m(l)$ とすれば，状態遷移モデルは以下のように表される．

$$\xi_m(l+1) = \xi_m(l) + \mathscr{W}_m(l)\,. \tag{5}$$

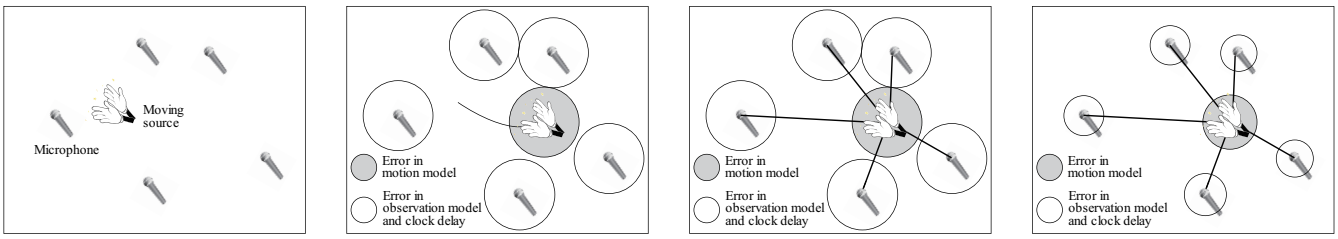
2.3 マイク位置推定

EKF-SLAM では，図 2(a) のように状態予測，観測予測，観測更新の 3 ステップを繰り返すことで校正を行う．まず，状態予測ステップでは，l 歩目での状態ベクトル $\xi_s(l)$，$\xi_m(l)$ を式 (4), (5) を用いて更新する (図 2(b))．観測予測ステップでは，状態予測ステップで更新された状態ベクトルと式 (3) を用いて f 回目の拍手の観測予測値 $\eta(f)$ を算出する (図 2(c))．観測更新ステップでは，f 回目の拍手で得られる実際の観測値と，観測予測値 $\eta(f)$ との誤差を最小にするようにカルマンゲインを導出し，状態ベクトルを更新する (図 2(d))．

3 伝達関数推定

伝達関数は室内における音源からマイクまでの音伝搬のモデルである．$S(\omega, f)$ と $X_m(\omega, f)$ をそれぞれ，短時間フーリエ変換後の f フレーム目の音源信号と m 番目のマイクでの観測信号とする．また，$A_m(\omega, \psi)$ を ψ 方向にある音源と m 番目のマイク間の周波数 ω での伝達関数とすると，ψ 方向にある音源 $S(\omega, f)$ は以下で表される．

$$X_m(\omega, f) = A_m(\omega, \psi) S(\omega, f) \tag{6}$$



(a) 初期状態: 基準マイクを原点とする．マイク位置は未知.

(b) 状態予測: 状態遷移モデルを用いて状態を更新する．

(c) 観測予測: 観測モデルを用いて観測を予測する．

(d) 観測更新: 予測誤差が最小となるようにカルマンゲインを更新する．

Figure 2: EKF-SLAM に基づく状態推定

ただし，フレーム長は十分長いとする．

伝達関数推定では $A_m(\omega,\psi)$ を推定することを目的とする．1 章で述べたように，伝達関数は数値計算もしくは計測によって求めるのが主流である．既存の伝達関数計測手法 [8; 9] では，$S(\omega,f)$ は既知である (TSP など) ことが前提であり，$A_m(\omega,\psi)$ は以下で得られた，

$$A_m(\omega,\psi) = X_m(\omega,f)/S(\omega,f)\,. \tag{7}$$

しかし，本稿では，拍手音から伝達関数を推定したいので，音源信号 $S(\omega)$，音源方向 ψ ともに未知として，伝達関数推定問題を解く必要がある．

3.1 FOLS 法による伝達関数推定

まず，拍手が行われた位置での音源とマイク間の伝達関数を推定する．拍手音の元信号 $S(\omega)$ は未知であり，事前情報として用いることができない．音源定位・分離では，あるチャンネルを基準にした相対的な伝達関数がわかれば処理上問題がないので，マイク 1 の観測信号 $X_1(\omega,f)$ を基準として，相対伝達関数を求める．すなわち，伝達関数は式 (7) の代わりに以下となる．

$$A_m(\omega,\psi) = X_m(\omega,f)/X_1(\omega,f) \tag{8}$$

しかし，$A_m(\omega,\psi)$ は，m 番目のマイクが基準マイクより拍手音を早く観測する場合，非因果成分を持ってしまう．そこで，$A_m(\omega,\psi)$ が非因果成分を持たないように．基準マイクの信号を T_o サンプルずらしたものを使う ($\tilde{X}_1(\omega,f)$ を表すこととする)．FOLS 法は，F フレームの $\tilde{X}_1(\omega,f)$ と $X_m(\omega,f)$ を用いて，回帰モデルを用いて雑音ロバストに伝達関数を推定する手法であり，以下で表される．

$$\underbrace{\begin{bmatrix} X_1(f+1) & \dots & X_M(f+1) \\ \vdots & & \vdots \\ X_1(f+F) & \dots & X_M(f+F) \end{bmatrix}}_{X_{[1:F]}} = \underbrace{\begin{bmatrix} \tilde{X}_1(f+1) \\ \vdots \\ \tilde{X}_1(f+F) \end{bmatrix}}_{\Omega_{[1:F]}} \underbrace{\begin{bmatrix} A_1(\psi) \\ \vdots \\ A_M(\psi) \end{bmatrix}^T}_{A^T(\omega,\psi)}$$

ここで，$\Omega_{[1:F]}$ はリグレッサである．最後に，$A(\omega,\psi)$ は，以下で求められる．

$$A^T(\omega,\psi) = (\Omega^T_{[1:F]}\Omega_{[1:F]})^{-1}\Omega^T_{[1:F]}X_{[1:F]} \tag{9}$$

フレーム数 F を長く取れば雑音ロバスト性能を高くすることが可能である．FOLS 法によって推定された伝達関数はマイク位置から得られる直接音成分だけでなく，ロボットや部屋の音響特性も含めたものであるため，実環境下のロボット聴覚応用に向いていると言える．

3.2 補間による伝達関数の整列

次に，音源方向 ψ について考える．拍手を行った際の音源の方向は，2 章の EKF-SLAM によって得ることができる．しかし，実際には，人の移動は状態遷移モデルに完全には従っていないため，EKF-SLAM から得られる音源方向の誤差は大きい．一方，マイク位置は，そもそも移動しないため，モデル誤差が小さく，精度のよい結果が得られる．そこで，各拍手の ψ を，より精度よく推定するため，推定したマイク位置を用いて，伝搬波モデルを用いて伝達関数を計算し，各拍手音 $S(\omega,f)$ のビームフォーミングを用いた定位を行い，精度のよい拍手方向 ψ を得る．

得られる音源方向 ψ のセットは拍手の位置であるため，伝達関数が等間隔に並んでおらず，音源定位や分離で使い勝手が悪い．そこで，伝達関数の補間を行い，所望の間隔 (5° 毎など) に配置された伝達関数を得る．

具体的には，K を総拍手数，ψ_k を k 回目の拍手の音源方向とする．また，得たい伝達関数を水平各方向に一周を N 等分した ψ_n ($1 \le n \le N$) とする．各 ψ_n に対して，ψ^- と ψ^+ を ψ_k の中で ψ_n に最も近い近傍の 2 点とする ($\psi^- \le \psi_n < \psi^+$)．FTDLI 法 [15] を用いて，$\psi^-$ と ψ^+ における伝達関数 $A(\omega,\psi^-)$ および $A(\omega,\psi^+)$ から $\hat{\psi}_n$ における伝達関数 $A(\omega,\psi)$ を補間する。

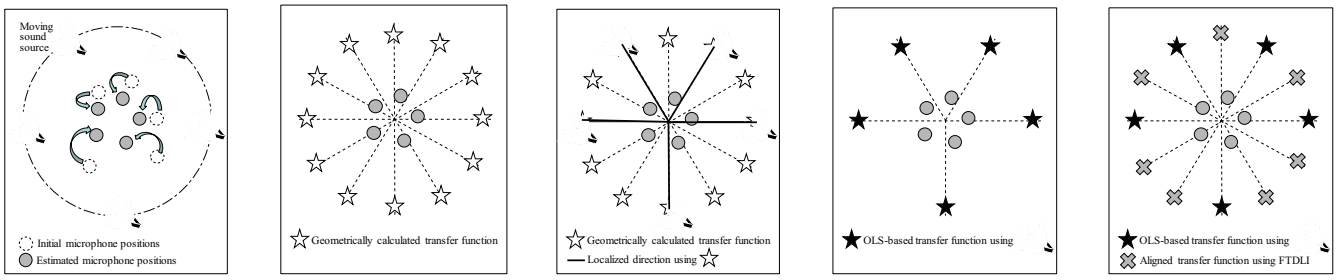
1) 周波数領域上と時間領域上で ψ_n が $[\psi^-\psi^+]$ の内分点となる α を算出し，線形補間を以下のように行う:

$$A_{m[F]}(\omega,\psi) = \alpha A_m(\omega,\psi^-) + (1-\alpha)A_m(\omega,\psi^+)$$
$$A_{m[T]}(\omega,\psi) = A^{\alpha}_m(\omega,\psi^-)A^{1-\alpha}_m(\omega,\psi^+),$$

ここで，$\psi^- \le \psi \le \psi^+$, $\alpha = \frac{\psi^+-\psi}{\psi^+-\psi^-}$ である．

2) 得られた伝達関数を振幅情報と位相情報に分離する:

$$A_{m[F]}(\omega,\psi) = \lambda_{m[F]}\exp(-j\omega t_{m[F]})$$
$$A_{m[T]}(\omega,\psi) = \lambda_{m[T]}\exp(-j\omega t_{m[T]})$$



(a) EKF-SLAM によるマイク位置推定　(b) 幾何計算法による伝達関数計算　(c) ビームフォーミングによる拍手音の定位　(d) 拍手音を用いた FOLS 法による伝達関数推定　(e) FTDLI 法による伝達関数の整列

Figure 3: 伝達関数のオンザスポット校正

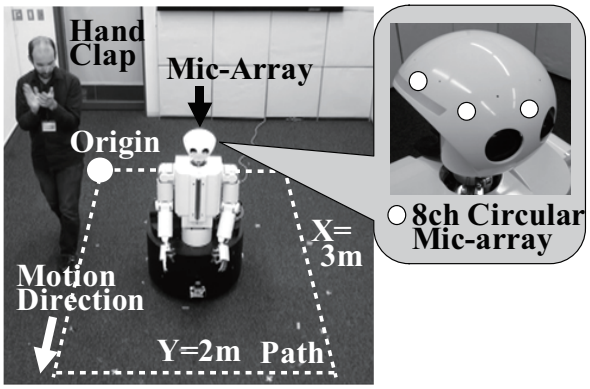


Figure 4: 実験環境

3) 振幅情報は時間領域，位相情報は周波数領域の補間を採用して最終的な伝達関数補間を行う:

$$A_m(\omega,\psi) = \lambda_{m[T]} \exp(-j\omega t_{m[F]})$$

この手法は，周波数領域の線形補間[16] と時間領域の線形補間[17] のハイブリッド法となっており，振幅と位相の両方が正しく補間できる．

3.3 システム構成

伝達関数推定の一連の流れを，図 3 に示す．EKF-SLAM によって得られるマイク位置 (図 3(a)) を用いて幾何計算法により，伝達関数を計算する (図 3(b))．得られた伝達関数を用いて，ビームフォーミングを行い，EKF-SLAM の際に観測した拍手の位置を推定する (図 3(c))．FOLS 法を用いて，拍手位置に対する伝達関数を推定する (図 3(d))．最後に FTDLI 法を用いて，伝達関数の補間を行い，等間隔に並んだ所望の方向の伝達関数を算出する (図 3(e))．

4 評価

本章では，まず，2 章のマイク位置推定の結果を示し，どのくらい短い計測でマイク位置が校正できるかを議論する．次に，3 章で推定された伝達関数をロボットを用いた音源定位・音源分離に適用し，既存手法で得られた伝達関数を用いた場合と比較を行う．

評価では提案手法をオープンソースのロボット聴覚ソフトウェア HARK [18] 上に実装し，2.0GHz の Intel Core i7 の CPU を持つ計算機で実時間動作することを確認した．本稿では，マイクアレイを搭載したロボットを残響時間が 0.2 秒 (RT20) の 7.0 m × 4.0 m の部屋の中央に設置した．マイクアレイは図 4 のように 8 チャネルの円状アレイを用いた．入力音響信号は 16kHz, 16 ビットでサンプリングした．音響信号処理のフレーム長とシフト長はそれぞれ，512, 160 サンプルとした．

4.1 マイク位置推定の評価

上述のように，ロボットは 7.0 m × 4.0 m の部屋の中央に設置されている．ロボットに搭載された円上アレイは半径約 0.1 m であった．人は，図 4 で示された原点からスタートし，3.0 m × 2.0 m の長方形の点線に沿って反時計回りに一定速度 $v_s(l)$ で移動した．拍手は 1 回/秒とした．

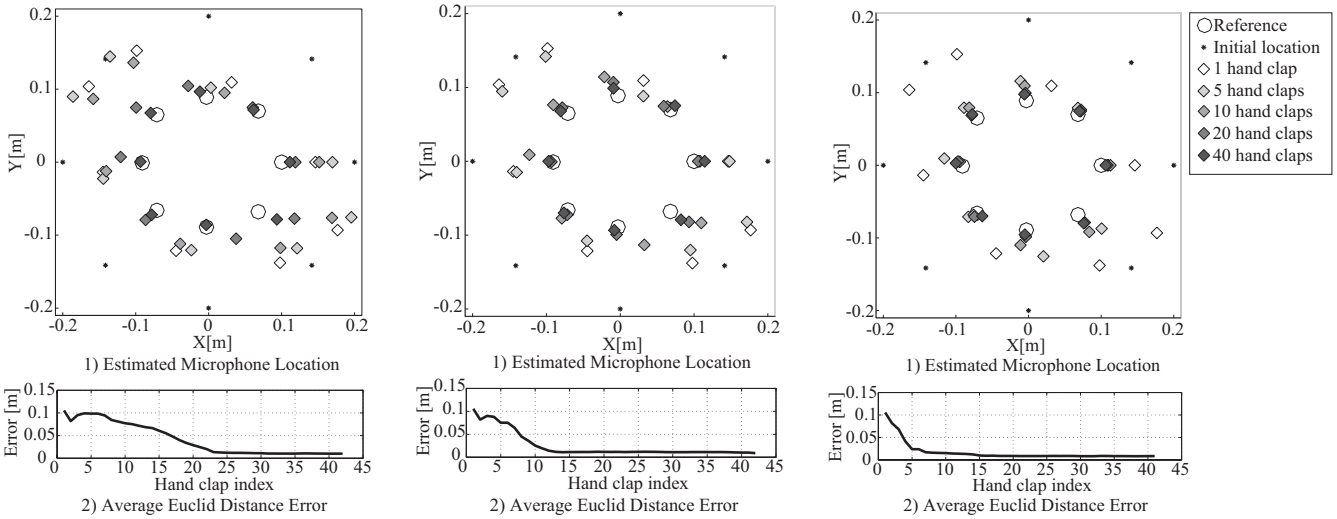
EKF-SLAM のため，人の初期位置と $v_s(l)$ は正解データを与えた．式 (3) の白色雑音の標準偏差 $\delta(f)$ は 0.0005 とした．式 (4) の白色雑音の標準偏差 $\mathscr{W}_s(l)$ は，$x_s(f), y_s(f)$ に対して 0.25，$\theta_s(f)$ に対して 1.0 とした．式 (5) の白色雑音の標準偏差 $\mathscr{W}_m(l)$ は，$x_m(f), y_m(f)$ に対して 0.25，$\tau_m(f)$ に対して 1.0 とした．収束速度を評価するため，マイクの初期位置はランダムに与えず，半径 0.2 m の円上アレイとなるように設定した．

図 5 に，各拍手回数での推定マイク位置，および全マイクの推定位置のユークリッド距離平均誤差の変化を示す．$v_s(l)$ を 0.1 m/s, 0.2 m/s, 0.6 m/s と変化させて評価した．

まず，図 5(a)-2), 5(b)-2), 5(c)-2) を比較すると，$v_s(l)$ が速いほど，収束速度が速いことがわかる．人はまず x 方向に 3 m 移動し，一辺を終えるのに $\frac{3}{v_s(l)}$ 回の拍手を必要とする (例えば，$v_s(l) = 0.1$ m/s の場合は 30 回)．収束速度はその一辺を終えるまでにかかる時間と相関があることから，観測時間差に大きな分散があるように移動すれば収束が速いことがわかる．

また，図 5(a)-1), 5(b)-1), 5(c)-1) に共通して，マイクの x 方向の位置から校正されている傾向が見られる．これは人が x 方向に最初に移動するためだと考えられる．このことからマイクをより早く校正するための最適な移動方法があると考えられる．

いずれの場合もマイク位置は高精度に校正されている



(a) $v_s(l) = 0.1$ m/s (b) $v_s(l) = 0.2$ m/s (c) $v_s(l) = 0.6$ m/s

Figure 5: マイク位置推定の結果

ことから提案法の有効性を確認することができた．また，$v_s(l) \geq 0.2$ m/s の場合は 20 回の拍手でマイク位置が十分に校正できていることから，20 秒ほどの録音でマイクを校正することができることがわかり，オンザスポット校正が十分可能であることが示された．4.2 章の評価では 20 回の拍手で推定されたマイク位置を用いることとした．

4.2 伝達関数推定の評価

伝達関数推定の有効性を音源定位・音源分離を通じて評価する．提案する伝達関数推定法に加え，2 種類の手法を比較した。TSP 法 (**TSP**) は，TSP 信号を用いて実際に測定した伝達関数を用いる手法であり，最も精度が良いことが期待できる [8]．幾何計算法 (**Calc**) は，マイク位置は EKF-SLAM で得られた位置を用い，自由音場を仮定した幾何計算によって算出した伝達関数を用いた手法である．

4.2.1 音源定位の性能比較

音源定位には，適応ビームフォーマの一種である，*MUltiple SIgnal Classification (MUSIC)*[15]を用いた．MUSIC では，M チャネルの入力音響信号の空間相関行列を計算し，その固有値展開を行う．$E(\omega, f) = [e_1(\omega, f), \ldots, e_M(\omega, f)]$ を f フレーム目で得られた固有ベクトルとする．定位では以下で表される MUSIC スペクトルを算出し，ここで提案法で推定された伝達関数 $A(\omega, \psi)$ を用いた．

$$P(\omega, \psi, f) = \frac{|A^*(\omega, \psi)A(\omega, \psi)|}{\sum_{m=L+1}^{M} |A^*(\omega, \psi)e_m(\omega, f)|}, \qquad (10)$$

ここで，$()^*$ は共役転置作用素を，L は音源数を表す．

音源定位では，図 6a) に示すように，ロボットの周囲で 10 cm ごとに 100 箇所でスピーカから 10 秒間ずつ白色雑音を出力し，水平方向の定位を行った．なお，TSP 法は，100 箇所すべてにおいて測定した伝達関数を用いた．提案法，幾何計算法では，20 回の拍手を用いてマイク位置を推定し，伝達関数は，5° ごとに推定・算出した．得られた伝達関数を式 (10) に適用し，評価を行った．

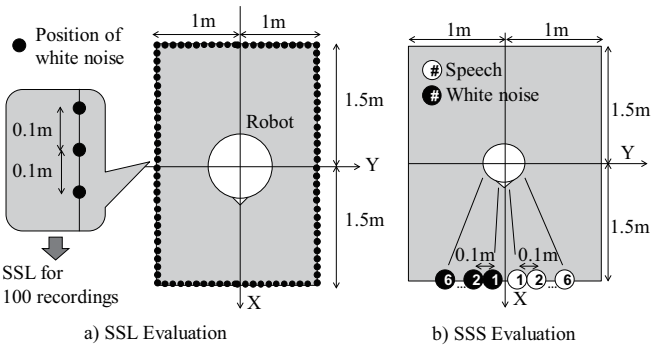


Figure 6: a) 音源定位評価のための白色雑音録音位置, b) 音源分離評価のための白色雑音・音声録音位置

Table 2: 音源定位評価結果

	TSP	Calc	Proposed
平均誤差 [deg]	5.11±2.01	6.93±2.04	6.82±2.13

表 2 に，音源定位の水平角度推定誤差の平均と標準偏差を示す．TSP 法が最も良い結果を示した．提案法 (Proposed) が，Calc に対して定位誤差が改善していることが確認でき，伝達関数にロボットや部屋の音響特性を考慮できたことの有効性が示された．

4.2.2 音源分離の性能比較

音源分離には，幾何拘束とブラインド分離のハイブリッドアルゴリズムである *Geometric High-order Decorrelation-based Source Separation (GHDSS)* [19] を用いた．GHDSS では，パーミュテーションとスケーリング問題を解決するために幾何拘束を用いており，伝達関数が使われている．音源分離は以下で表される．

$$Y(\omega, \psi) = W(\omega, \psi)X(\omega, \psi), \qquad (11)$$

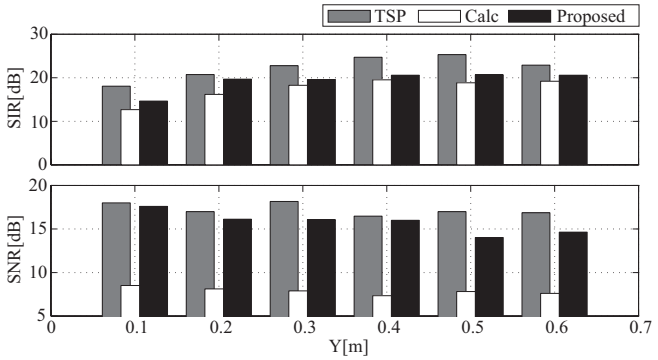


Figure 7: 音源分離評価結果

ここで，$Y(\omega,\psi)$ は分離音，$W(\omega,\psi)$ は分離行列，$X(\omega,\psi)$ は M チャネルの入力音響信号を表す．コスト関数 $J(W(\omega,f))$ を最小化するように $W(\omega,\psi)$ を更新することで音源分離が行われるが，GHDSS ではコスト関数を以下のように設定している．

$$J(W(\omega,f)) = \alpha J_1(W(\omega,f)) + \beta J_2(W(\omega,f)), \qquad (12)$$

ここで，$J_1(\cdot)$ はブラインド音源分離のためのコスト関数を，$J_2(\cdot)$ は幾何拘束のためのコスト関数を表す．α と β は，$\alpha+\beta=1$ を満たす重みを表す．提案法で推定された伝達関数 $A(\omega,\psi)$ は，以下のように $J_2(\cdot)$ で用いられる．

$$J_2(W(\omega,f)) = \|diag[W(\omega,f)A-I]\|^2. \qquad (13)$$

評価では，*Signal-to-Interference Ratio (SIR)* と *Signal-to-Noise Ratio (SNR)* の 2 つの指標を用いた．SIR の評価には，BSS EVAL Toolbox [20] を用いた．Toolbox では，分離音は $y_i(t) = s_t(t) + s_i(t) + s_n(t)$ としてモデル化されている．ここで，$s_t(t)$ は目的音のみがある場合の分離信号を，$s_i(t)$ は非目的音のみがある場合の分離信号を，$s_n(t)$ 背景雑音のみがある場合の分離信号を表す．SIR は $SIR = 10\log_{10}\frac{|s_t|^2}{|s_i|^2}$ として，SNR は $SNR = 10\log_{10}\frac{|s_t+s_i|^2}{|s_n|^2}$ として計算される．

図 6(b) に示すように，ターゲット音源の位置を白丸の 1〜6 から選択，雑音源 (白色雑音) は，ターゲット音源と対称になる位置の黒丸を選択した。各手法によって得られた伝達関数を GHDSS の式 (13) の D に用い，２音源からの混合音に対し音源分離を行った．

図 7 に結果を示す. 音源分離でも TSP 法が一番良い結果を示している。提案法 (Proposed) は，SIR, SNR 共に TSP 法に近い性能を示しており，Calc と比較すると良好な結果が得られていることがわかる。いずれも伝達関数にロボットや部屋の音響特性を考慮したことが効果を示した．

TSP 法は音源定位・音源分離の両者でもっとも良い性能を示したが，100 箇所の伝達関数の計測には特別な機器を用いても 60 分を要している．一方，提案法は 20 回の拍手 (20 秒) のみで伝達関数を校正できており，結果として TSP 法と同様な性能を得られた．このことから，提案法は初心者にも簡易に実現可能な実用的なマイクアレイ校正手法であると考えている．

5 結論

本稿ではロボットや部屋の音響特性を考慮したマイクアレイ伝達関数のオンザスポット (その場で簡易に可能な) 校正について述べた．EKF-SLAM に基づくマイク位置推定を導入し，FOLS 法と FTDLI 法による伝達関数推定を提案した．評価では，20 回の拍手 (20 秒の録音) でマイクアレイが精度良く校正でき，音源定位・音源分離を通して提案法で得られた伝達関数がマイクアレイ処理性能を向上していることが確認できた．今後の課題は音声などの一般音を用いた校正や三次元モデルへの拡張である．

参考文献

[1] K. Nakadai *et al.*, “Active Audition for Humanoid”, in *Proc. of 17th AAAI*, pp. 832–839, 2000.

[2] Y. Sasaki *et al.*, “Nested igmm recognition and multiple hypothesis tracking of moving sound sources for mobile robot audition,” in *IROS*, pp. 3930–3936, 2013.

[3] J.-M. Valin *et al.*, “Enhanced robot audition based on microphone array source separation with post-filter,” in *IROS*, pp. 2123–2128, 2004.

[4] K. Nakamura *et al.*, “Intell. sound source localization for dynamic environments,” in *IROS*, pp. 664–669, 2009.

[5] F. Asano *et al.*, “Sound source localization and signal separation for office robot Jijo-2,” in *Proc. of IEEE Int ’l Conf. Multisensor Fusion and Integ. for Intell. Sys. (MFI)*, pp. 243–248, 1999.

[6] K. Yamamoto *et al.*, “An acoustic simulation for speech interface of humanoid robot,” in *Proc. of Acoustical Society of Japan Autumn Meeting*, pp. 815–818, 2009.

[7] K. Nakadai *et al.*, “Applying scattering theory to robot audition system: robust sound source localization and extraction,” in *IROS*, pp. 1147–1152, 2003.

[8] Y. Suzuki *et al.*, “An optimum computer-generated pulse signal suitable for the measurement of very long impulse responses”, *J. Acoust. Soc. Am.*, vol. 97, no. 2, pp. 1119-1123, 1995.

[9] G. B. Stan, J. J. Embrechts and D. Archambeau, “Comparison of different impulse response measurement technique,” *Journal of the Audio Engineering Society*, vol. 50, pp. 249–262, 2002.

[10] S. Thrun, “Affine structure from sound,” *Advances in Neural Information Processing Systems*, vol. 18, pp. 1355–1362, 2005.

[11] Y. Kuang and K. Astrom, “Stratified Sensor Network Self-Calibration From TDOA Measurements,” in *EUSIPCO*, 2013.

[12] N. Ono *et al.*, “Blind Alignment of Asynchronous Recorded Signals for Distributed Microphone Array,” in *WASPAA*, pp. 161–164, 2009.

[13] H. Miura *et al.*, “SLAM-based Online Calibration of Asynchronous Microphone Array for Robot Audition,” in *IROS*, pp. 524–529, 2011.

[14] Y. Bando *et al.*, “Posture estimation of hose-shaped robot using microphone array localization,” in *IROS*, pp. 3446–3451, 2013.

[15] K. Nakamura, K. Nakadai and G. Ince, “Real-time Super-resolution Sound Source Localization for Robots,” in *IROS*, pp. 694–699, 2012.

[16] T. Nishino *et al.*, “Interpolating head related transfer functions in the median plane,” in *WASPPA*, pp. 167–170, 1999.

[17] M. Matsumoto *et al.*, “A method of interpolating binaural impulse responses for moving sound images,” *Acoust. Sci. Tech.*, Vol 24, pp. 284–292, 2003.

[18] K. Nakadai *et al.*, “Design and Implementation of Robot Audition System HARK”, *Advanced Robotics*, vol. 24, pp. 739–761, 2009.

[19] H. Nakajima *et al.*, “Blind Source Separation with parameter-free adaptive step-size method for robot audition,” *IEEE TASLP*, vol.18, no. 6, pp. 1476–1485, 2010.

[20] E. Vincent, R. Gribonval and C. Fevotte, “Performance measurement in blind audio source separation,” *IEEE TASLP*, vol. 14, no. 4, pp. 1462–1469, 2006.