人工知能学会第二種研究会資料
AIチャレンジ研究会
SIG-Challenge-066-06

ガウス過程回帰を用いた音響伝達関数の環境変化適応

Online adaptation of acoustic transfer function using Gaussian process regression

藤田 侑樹[1]*　糸山 克寿[2]　西田 健次[1]　中臺 一博[1]

[1] 東京科学大学/ Institute of Science Tokyo
[2] (株) ホンダ・リサーチ・インスティチュート・ジャパン/HRI-JP
[3] 事業創造大学院大学/ Graduate Institute for Entrepreneurial Studies

Abstract: 本論文では，異なる部屋環境間のミスマッチを吸収するための音源定位技術における音響伝達関数（Acousitc Transfer Function, ATF）の適応法を提案する．従来手法では、入力音方向に対応する ATF の一部のみを更新するため，適応処理により更新方向と非更新方向の ATF 間で不整合が生じる問題があった．この課題を解決するため，モードフィルタとガウス過程回帰を組み合わせた新しい ATF 適応法を提案する．提案手法を用いた実験の結果，ほぼ全ての角度において平均二乗誤差が改善されることを確認した．

1　はじめに

音響伝達関数とは，音源・マイクロホン間の信号伝播を規定する関数である．音の方向を推定する音源定位や特定の音源を抽出する音源分離といったマイクロホンアレイ信号処理において，音響伝達関数は必須情報である [1]．しかし，音響伝達関数の収録や計算には，手間がかかり，音響環境が異なれば，異なる音響伝達関数を用意する必要がある．そこで，本稿では，ガウス過程回帰を用いて，音響伝達関数を観測信号からオンラインで適応的に推定する手法を提案する．

2　音響伝達関数のキャリブレーション

音響伝達関数は様々な音源方向からの伝達特性のセットとして定義される．音響伝達関数を取得するためには，自由音場を仮定し，様々な音源方向に対する伝達特性を幾何的に計算する，無響室での様々な方向から音響測定を行って取得する，といった手法が用いられる [2, 3]．しかし，音響伝達関数を取得する時点では，対象環境がわからない場合が多く，事前準備可能な音響伝達関数は，対象環境に即した音響伝達関数と差異が生じる．このため，音源定位・分離の性能が低下してしまう．仮に，事前に対象環境がわかる場合でも，差異の少ない音響伝達関数を得るためには，対象環境で手拍子や Time Stretched Pulse (TSP) などを用いた音響計測を時間をかけて行う必要がある [4]．また，せっかく計測しても，室内の家具の移動などで，音響環境が変化してしまえば，差異が生じ，再計測が必要となり，効率が悪い．

環境に即した音響伝達関数を事前準備する代わりに，対象環境における観測信号から音響伝達関数を推定することができれば，計測の手間を省きつつ，差異の小さい音響伝達関数が取得できる．この問題は，マイクロホンアレイのキャリブレーション問題として扱われてきた．Thrun ら [5] は複数のマイクロホンで観測した信号間の遅延情報から，各マイクロホンの位置を推定する手法を提案している．また，Miura ら [6] は手拍子音を用いて Simultaneous Localization And Mapping (SLAM) の枠組みを用いて，マイクロホン位置，音源位置，マイクロホン間のオフセット時間を同時に推定するキャリブレーション手法を提案している．さらに，Nakamura ら [7] は，この手法と音響伝達関数補間手法 [8] を利用して，音響伝達関数を推定する手法を提案している．この手法は，音響伝達関数を直接推定するわけではなく，マイクロホンと音源位置を推定し，それらの情報から幾何学的に音響伝達関数を計算している．この際に自由音場を仮定しているため，残響や反響を考慮することができない．また，キャリブレーションには，手拍子や Time Stretched Pulse (TSP) などの特定の音源が必要であり，こうした音源を用いたキャリブレーション処理を，事前に行っておく必要がある．システム運用の簡便さを考えれば，マイクロホンアレイ信号処理を行いながら，オンラインで適応的なキャ

*連絡先：東京科学大学
〒 152-8552 東京都目黒区大岡山 2-12-1
y.fujita@ra.sc.e.titech.ac.jp

リブレーション処理が可能な手法が望ましい．

このような手法として，Nakadai ら [9] は，観測信号の相対位相差を用いた音響伝達関数をオンラインで適応更新する手法を提案している．Fujita ら [10] は，この手法を拡張し，観測信号に含まれる外れ値を検出することで，オンライン適応性能を向上する手法を提案している．しかし，これらの手法では，一回の観測で更新する音響伝達関数は，音響伝達関数セット中の観測信号の音源方向に対応する音響伝達関数のみであり，その他の方向に対応する音響伝達関数は更新されない．このため，音響伝達関数全体の整合性が取れなくなる危険性がある．例えば，特定音源方向に対する音響伝達関数のみを更新した結果，更新した音響伝達関数が，全く異なる音源方向に対応する音響伝達関数と類似した関数となり，音源定位や音源分離に悪影響を及ぼしてしまうといった弊害が考えられる．

この問題を解決するには，観測に対して，特定方向の音響伝達関数のみを更新するのではなく，音響伝達関数セット全体を更新する手法が必要である．このような手法として，音響伝達関数の周波数領域位相情報を対象に，ガウス過程回帰 [11, 12, 13] を用いた更新を検討する．ベイズ推論に基づくガウス過程回帰を用いれば，観測信号を確率的に捉えることができるため，観測信号の方向以外の方向も扱うことが可能である．つまり，観測に対して，音響伝達関数セット全体を更新することが可能となり，整合性の取れた更新が期待できる．一方で，ガウス過程回帰は観測値を真値に近いと見なすため，逐次的に用いると分散が減少する．マイクロホンアレイでの観測値には揺らぎが発生することが多く，推定・更新手法としてガウス過程回帰を用いるためには分散が消失する問題を解決しなければならない．従って，以下の 2 つの課題を同時に解決可能なガウス過程回帰に基づく音響伝達関数更新手法を提案する．

- 音響伝達関数全体の整合性をとった更新が困難
- ガウス過程回帰適用時の分散消失

3　提案手法

本節では，全体で整合性を取るための音響伝達関数の環境変化適応手法について述べる．Fig. 2 に提案手法の全体図を示す．提案手法は最頻値フィルタを用いた音源方向における観測信号のフィルタリングとガウス過程回帰を用いた音響伝達関数全体の推定・更新の二つのブロックからなる．以下では，信号のモデルについて説明した後，二つのブロックについて，それぞれ説明する．

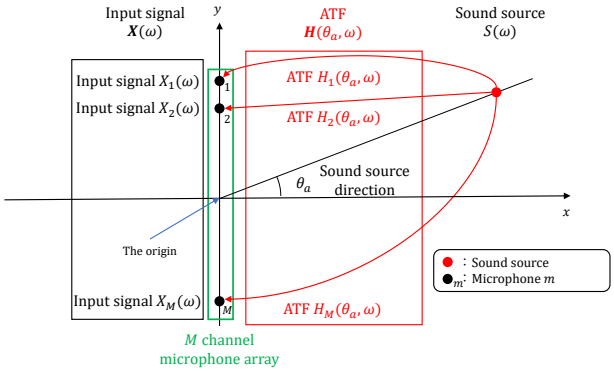

図 1: The model of signal propagation

3.1　信号のモデルと音源定位

Fig. 1 に音源からマイクロホンアレイへの信号伝播のモデルを示す．平面上に $x-y$ 座標を取り，周波数を ω として，音源方向 θ_a から，音源信号 $S(\omega)$ が到来するものとする．この音源信号を M チャネルのマイクロホンアレイで観測した信号を $\boldsymbol{X}(\omega) = [X_1(\omega),\cdots,X_M(\omega)]^T$ と表す．ここで，$X_m(\omega)$ は，m 番目のマイクロホンで観測信号を表す．観測信号 $\boldsymbol{X}(\omega)$ は音源方向 θ_a に対応する音響伝達関数 $\boldsymbol{H}(\theta_a,\omega) = [H_1(\theta_a,\omega),\cdots,H_M(\theta_a,\omega)]^T$ を用いて，式 (1) のように計算できる．ここで，$H_m(\theta_l,\omega)$ は，m 番目のマイクロホンにおける音源方向 θ_l に対応する音響伝達関数である．音響伝達関数はマイクロホンアレイに対する方向ごとに式 (2) のように定義できる．L は水平方向の離散化の総数 (音源方向数) である．

$$\boldsymbol{X}(\omega) = \boldsymbol{H}(\theta_a,\omega)S(\omega) \quad (1)$$

$$\boldsymbol{H}(\omega) = [\boldsymbol{H}(\theta_1,\omega),\boldsymbol{H}(\theta_2,\omega),\cdots,\boldsymbol{H}(\theta_L,\omega)] \quad (2)$$

以下では，信号についてはすべて周波数領域で考え，ω を省略する．実際には式 (1) における音源信号 S は未知であり，直接得ることはできず，観測信号 $\boldsymbol{X}$ として得る．この音源の方向 θ_a は，式 (3) のように空間スペクトル S_{sp} の最大化問題を解くことで得られる．

$$\theta_a = \underset{\theta_l}{\mathrm{argmax}}(S_{sp}(\boldsymbol{H},\boldsymbol{X},\theta_l)) \quad (3)$$

また，θ_l 方向 m 番目のマイクロホンにおける音響伝達関数 $H_m(\theta_l)$ はそのマイクの位相差 $f^m(\theta_l)$ を用いて，式 (4) の関係がある．ただし i は虚数単位を表し，本稿では位相差を 1 番目のマイクロホンから各マイクロホンへの周波数領域での位相遅れとして定義する．

$$H_m(\theta_l) = \exp\left(if^m(\theta_l)\right) \quad (4)$$

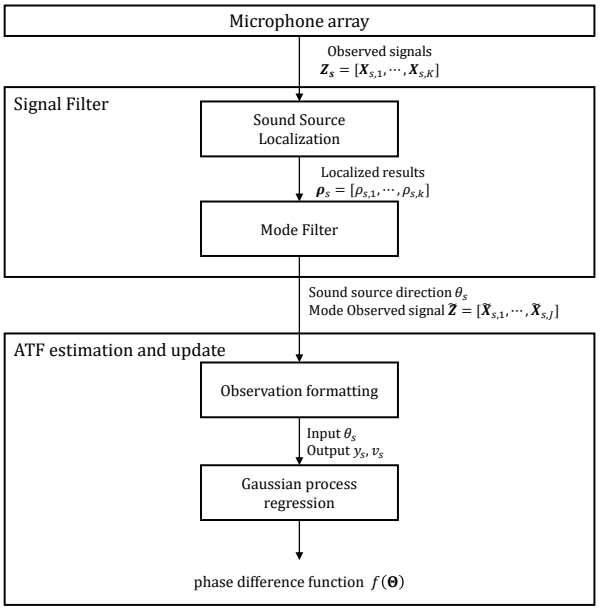

図 2: A process flow for the proposed method

3.2 最頻値フィルタ

観測信号をある時間単位ごとに区切り，1 つの区間には K 番目の観測信号を $\boldsymbol{X}_{s,k}$ と表す．M チャネルのマイクロホンアレイでの観測信号を用いる場合，観測信号 $\boldsymbol{X}_{s,k}$ は転置 T を用いて $\boldsymbol{X}_{s,k} = [X_{s,k,1}, \cdots, X_{s,k,M}]^T$ と表すことができる．ここで $X_{s,k,m}$ は，m チャンネル目のマイクロホンで観測する信号を表している．また，観測信号群 $\boldsymbol{Z}_s$ を $\boldsymbol{Z}_s = [\boldsymbol{X}_{s,1}, \cdots, \boldsymbol{X}_{s,K}]$ とする．s 区間目の音響伝達関数を $\boldsymbol{H}_s$ として，観測信号群 $\boldsymbol{Z}_s$ の K 個の観測信号について，音源定位を行う．式 (3) と同様，方向の分解数を N として，式 (5) により，$\boldsymbol{X}_{s,k}$ に対する音源定位結果を $\rho_{s,k}$ として得る．この処理によって，観測信号群 $\boldsymbol{Z}_s$ に対する定位方向群 $\boldsymbol{\rho}_s = [\rho_{s,1}, \cdots, \rho_{s,K}]$ が得られる．定位方向群 $\boldsymbol{\rho}_s$ に対して，最頻値を取る処理を式 (6) として行う．ここで MODE は，定位方向群 $\boldsymbol{\rho}_s$ の中の最頻値を音源方向 θ_s として得る処理である．その後，観測信号群 $\boldsymbol{Z}_s$ のうち，音源方向 θ_s と定位方向群 $\boldsymbol{\rho}$ の各要素が一致する $J(\leq K)$ 個の観測信号からなる最頻値観測信号群 $\tilde{\boldsymbol{Z}}_s = [\tilde{\boldsymbol{X}}_{s,1}, \cdots, \tilde{\boldsymbol{X}}_{s,J}]$ を取り出し，式 (7) のように「==」を用いて表す．ただし，最頻値フィルタを通した s 区間目 j 番目の観測信号 $\tilde{\boldsymbol{X}}_{s,j}$ は m 番目のマイクロホンでの観測信号 $\tilde{X}_{s,j,m}$ を用いて $\tilde{\boldsymbol{X}}_{s,j} = [\tilde{X}_{s,j,1}, \cdots, \tilde{X}_{s,j,M}]^T$ と表す．この最頻値観測信号群 $\tilde{\boldsymbol{Z}}_s$ と音源方向 θ_s をガウス過程回帰ブロックにおける入力とする．

$$\rho_{s,k} = \underset{\theta_l}{\mathrm{argmax}}(S_{sp}(\boldsymbol{H}_s, \boldsymbol{X}_{s,k}, \theta_l)) \quad (5)$$

$$\theta_s = \mathrm{MODE}(\boldsymbol{\rho}_s) \quad (6)$$

$$\tilde{\boldsymbol{Z}}_s = \boldsymbol{Z}_s[\rho_{s,k} == \theta_s] \quad (7)$$

3.3 ガウス過程回帰

音響伝達関数と位相差は式 (4) の関係がある．この位相差について音源方向を入力値とする関数と定義し，ガウス過程回帰を用いて位相差関数を推定する問題に置き換える．s 区間目について，まずは最頻値観測信号群 $\tilde{\boldsymbol{Z}}_s$ から j 番目の観測信号，m 番目のマイクロホンに関する位相差 $y_{s,j,m}$ を式 (8) のように複素数の偏角として求め，最頻値観測位相差群 $\boldsymbol{y}_{s,m} = [y_{s,1,m}, \cdots, y_{s,J,m}]$ を求める．

$$y_{s,j,m} = \mathrm{Arg}\frac{\tilde{X}_{s,j,m}}{\tilde{X}_{s,j,1}} \quad (8)$$

ガウス過程回帰はマイクロホン毎に行うため，以下ではマイクロホン番号 m を省略する．この最頻値観測位相差群 $\boldsymbol{y}_s$ の各要素から，観測値 y_s，観測不偏分散 v_s を式 (9)(10) のように求める．

$$y_s = \frac{\sum_j y_{s,j}}{J} \quad (9)$$

$$v_s = \frac{\sum_j (y_{s,j} - \mu_s)^2}{J-1} \quad (10)$$

s 区間目の音源方向 θ_s を入力値として，位相差の観測値 $y_{s,m}$ は位相差関数 $f(\cdot)$ とノイズ ϵ を用いて式 (11) のモデルで表す．この時，L 個の入力値からなるベクトルを $\boldsymbol{\Theta} = [\theta_1, \cdots, \theta_L]^T$ とし，位相差関数 $f(\cdot)$ とノイズ ϵ は式 (12)(13) で表す GP と仮定する．初期値 $\boldsymbol{\mu}_0^f, \boldsymbol{C}_0^f$ は初期値関数 $m(\cdot)$，カーネル関数 $k(\cdot,\cdot)$ を用いて式 (14)(15) のように表す．ここで，$m(\boldsymbol{\Theta}) = [m(\theta_1), \cdots, m(\theta_N)]^T$ であるようなベクトル，$k(\boldsymbol{\Theta}, \boldsymbol{\Theta})$ は (a,b) 要素が $k(\theta_a, \theta_b)$ であるような行列である．カーネル関数はハイパーパラメータ β, κ を用いて式 (16) とし，方向の周期性を考慮している．

$$y_s = f(\theta_s) + \epsilon \quad (11)$$

$$f(\boldsymbol{\Theta}) \sim \mathcal{N}(\boldsymbol{\mu}_s^f, \boldsymbol{C}_s^f) \quad (12)$$

$$\epsilon \sim \mathcal{N}(0, \sigma_\epsilon^2) \quad (13)$$

$$\boldsymbol{\mu}_0^f = m(\boldsymbol{\Theta}) \quad (14)$$

$$\boldsymbol{C}_0^f = k(\boldsymbol{\Theta}, \boldsymbol{\Theta}) \quad (15)$$

$$k(\theta, \theta') = \beta \exp\left(-\frac{\kappa}{2}(\theta - \theta')^2\right) \quad (16)$$

以上を用いて逐次的にガウス過程回帰を用いる．式 (17) ~ (20) で $s-1$ 番目での事後分布を用いて s 番目での

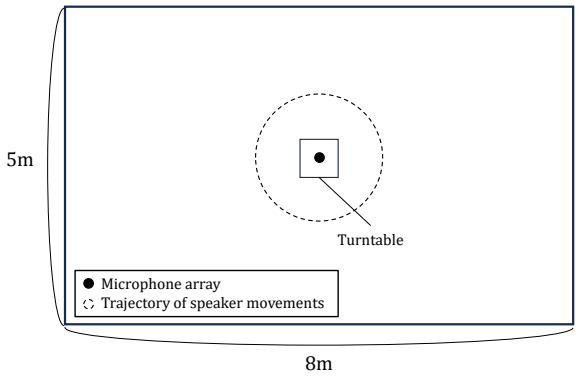

図 3: Layout of RoomA

予測分布を求め，式 (21) ∼ (23) で s 番目での事後分布を計算している．式 (22) が示すように，不偏分散を $\boldsymbol{G}_s v_s \boldsymbol{J}_s$ の項に反映し，全体の角度に加算することで分散の消失を止める役割を担っている．

$$\mu_s^p = m(\theta_s) + \boldsymbol{J}_s\left(\boldsymbol{\mu}_{s-1}^f - m(\boldsymbol{\Theta})\right) \quad (17)$$

$$C_s^p = B_s + \boldsymbol{J}_s \boldsymbol{C}_{s-1}^f \boldsymbol{J}_s^T \quad (18)$$

$$B_s = k(\theta_s, \theta_s) - \boldsymbol{J}_s k(\boldsymbol{\Theta}, \theta_s) \quad (19)$$

$$\boldsymbol{J}_s = k(\theta_s, \boldsymbol{\Theta}) \cdot k(\boldsymbol{\Theta}, \boldsymbol{\Theta})^{-1} \quad (20)$$

$$\boldsymbol{\mu}_s^f = \boldsymbol{\mu}_{s-1}^f + \boldsymbol{G}_s(y_s - \mu_s^p) \quad (21)$$

$$\boldsymbol{C}_s^f = \boldsymbol{C}_{s-1}^f - \boldsymbol{G}_s \boldsymbol{J}_s \boldsymbol{C}_{s-1}^f + \boldsymbol{G}_s v_s \boldsymbol{J}_s \quad (22)$$

$$\boldsymbol{G}_s = \boldsymbol{C}_{s-1}^f \boldsymbol{J}_s^T \left(C_s^p + \sigma_\epsilon^2\right)^{-1} \quad (23)$$

4 評価実験

前節で述べた提案手法が，整合性を取れた更新に有効であるかを確認する．そのために，ある環境で収録した音声データに対して環境適応を行い，既存手法 [10] から得られる音響伝達関数と更新を行わない初期の音響伝達関数を用いて比較を行う．比較は，音響伝達関数が示す位相差を平均二乗誤差を用いて比較する．

4.1 データ準備

Fig. 3 に RoomA のレイアウトを示す．RoomA の広さは $5 \times 8 \times 3$ m，残響時間は $RT_{60} = 0.2$[s] である．部屋には，中央にマイクロホンアレイを固定する台と，スピーカを回転させる机が設置されている．円形マイクロホンアレイは部屋の中央の台に固定し，床からの高さは 1.0 m である．円形マイクロホンアレイから 1.0 m，床からの高さを 1.0 m にスピーカを配置する．音声の収録は 8 チャネルの円形マイクロホンアレイを用いてを行う．収録時のサンプリング周波数は 16 kHz である．

表 1: STFT parameters

Frame length	512 points
Shift width	256 points
Window function	Hann window

表 2: SSL parameters

Localization method	DS
Direction resolution	5°
Min frequency	300Hz
Max frequency	1250Hz

まずは TSP 信号を収録して，音響伝達関数 $\boldsymbol{H}_R$ を作成する．次に，日本語話し言葉コーパス (CSJ)[14] から選んだ男性の音声を収録する．この時マイクロホンアレイとスピーカの位置関係を保ったまま，スピーカを反時計回りに 2 周し，約 12 分間の長さの音声データとして収録した．

Table 1 の条件で，収録されたデータに対して短時間フーリエ変換 (short-time Fourier transform, STFT) を行う．無音区間を除くため平均音圧が -24 dB 以上のフレームを抽出し，これらに対して音源定位を行う．Table 2 に音源定位の条件を示す．

4.2 実験内容

マイクロホンの位置関係から幾何的に計算した音響伝達関数を，初期の音響伝達関数 $\boldsymbol{H}_I$ とする．$\beta = 5, \kappa = 1, \sigma_\epsilon = 0.1$ として提案手法を用いて，音響伝達関数 $\boldsymbol{H}_P$ を作成する．また，初期の音響伝達関数 $\boldsymbol{H}_I$ から既存手法 [10] で作成した音響伝達関数 $\boldsymbol{H}_E$ を作成する．以上 3 つの音響伝達関数 $\boldsymbol{H}_I$,$\boldsymbol{H}_P$,$\boldsymbol{H}_E$ と，計測した正確な音響伝達関数 $\boldsymbol{H}_R$ を比較する．最頻値フィルタの適用区間における K の値は 60 とした．

4.3 評価指標

以上 3 つの音響伝達関数 $\boldsymbol{H}_I$,$\boldsymbol{H}_P$,$\boldsymbol{H}_E$ が示す位相差と，計測した正確な音響伝達関数 $\boldsymbol{H}_R$ が示す位相差を，式 (24) で定義される音源方向毎の平均二乗誤差 $MSE(\theta_j, H_*)$ で比較する．ただし，$* = I, P, E$ であり，N は周波数の分解数を表す．また，$f_*^{n,m}$ は音響伝達関数 $\boldsymbol{H}_*$ の周波数 n, マイク m の位相差，$f_R^{n,m}$ は音響伝達関数 $\boldsymbol{H}_R$ が示す位相差を表す．

$$MSE(\theta_j, H_*) = \frac{1}{MN}\sum_m \sum_n (f_*^{n,m} - f_R^{n,m})^2 \quad (24)$$

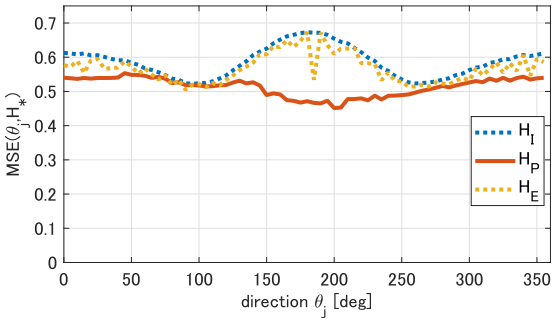

図 4: MSE of ATF's

4.4 実験結果

Fig. 4 に 3 つの音響伝達関数の音源方向毎の位相差の平均二乗誤差を示す．音声データを収録した環境で実際に測定した音響伝達関数 $\boldsymbol{H}_R$ に比べ，初期の音響伝達関数 $\boldsymbol{H}_I$ は大きく異なる．区間ごとに音源定位方向の音響伝達関数を更新する既存手法 [10] を用いた音響伝達関数 $\boldsymbol{H}_E$ は 180° 付近が大きく更新したが，その周りの角度は初期値とほぼ同じであり，更新されていないことがわかる．このような音響伝達関数は整合性の取れた正確な音響伝達関数を得ることはできない．一方で，提案手法によって得られる音響伝達関数 $\boldsymbol{H}_P$ はほぼすべての方向について $\boldsymbol{H}_R$ に近づいており，提案手法が整合性の取れた環境適応に有効であることを示している．

5 おわりに

本稿では，音響伝達関数の環境適応における整合性が及ぼすマイクロホンアレイ信号処理性能の低下問題を扱い，最頻値フィルタとガウス過程回帰に基づいた手法を提案した．この手法に対して，幾何学的に求められる音響伝達関数を環境に適応させる実験を行った．実験では，音声収録環境下での正確な音響伝達関数と初期，既存手法 [10]，提案手法の音響伝達関数が示す位相差を平均二乗誤差で比較し，提案手法の有効性を示した．今後は，このガウス過程回帰におけるパラメータについて考察を行う予定である．

謝辞

本研究は JSPS 科研費 JP22F22769，JP22KF0141，立石財団研究助成 (A) 2241011 および福島国際研究教育機構 (F-REI) の委託研究費 (JPFR23010102) の助成を受けた．

参考文献

[1] 浅野太. 音のアレイ信号処理. コロナ社, 2011.

[2] Y. Suzuki, F. Asano, H.-Y. Kim, and T. Sone. 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.

[3] 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.

[4] N. Aoshima. Computer-generated pulse signal applied for sound measurement. *J. Acoust. Soc. Am*, Vol. 69, No. 5, pp. 1484–1488, 1981.

[5] S. Thrun. Affine structure from sound. *Advances in Neural Information Processing Systems*, Vol. 18, pp. 1353–1360, 2006.

[6] H. Miura, T. Yoshida, K. Nakamura, and K. Nakadai. Slam-based online calibration for asynchronous microphone array. *Advanced Robotics*, Vol. 26, No. 17, pp. 1941–1965, 2012.

[7] K. Nakamura, S. Ambrose, and K. Nakadai. On-thespot calibration of microphone array transfer functions for robot audition. In *IEEE International Conference on Robotics and Automation*, pp. 3354–3359, 2015.

[8] K. Nakamura, K. Nakadai, and G. Ince. Realtime super-resolution sound source localization for robots. In *IEEE/RSJ International Conference on Robots and Intelligent Systems (IROS)*, pp. 694–699, 2012.

[9] K. Nakadai, M. Takigahira, Y. Kawai, and H. Nakajima. Fully-online always-adaptation of transfer functions and its application to sound source localization and separation. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 2100–2105, 2021.

[10] Y. Fujita, K. Itoyama, K. Nishida, and K. Nakadai. Adapting acoustic transfer functions to environmental changes with mode filter. bachelors thesis, 2022.

[11] C.E. Rasmussen and C.K.I. Williams. *Gaussian processes for machine learning*. The MIT Press, 2006.

[12] M.F. Huber. Recursive gaussian process regression. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pp. 3362–3366, 2013.

[13] K. Seo and M. Yamakita. Nonlinear time-varying system identification with recursive gaussian process. In *2017 American Control Conference (ACC)*, pp. 825–830, 2017.

[14] K. Maekawa, H. Koiso, S. Furui, and H. Isahara. Spontaneous speech corpus of japanese. In *the Second International Conference on Language Resources and Evaluation (LREC'00). ELRA*, 2000.