

音響伝達特性を用いたシングルチャネル音源位置推定における 局所的回帰に基づく未学習位置の補間

高島 遼一[†] 滝口 哲也^{††} 有木 康雄^{††}

[†] 神戸大学大学院システム情報学研究科 〒 657-8501 兵庫県神戸市灘区六甲台町 1-1

^{††} 神戸大学自然科学系先端融合研究環 〒 657-8501 兵庫県神戸市灘区六甲台町 1-1

E-mail: [†]takashima@me.cs.scitec.kobe-u.ac.jp, ^{††}{takigu,ariki}@kobe-u.ac.jp

あらまし 我々はこれまで、観測音声の音響伝達特性が話者の位置に依存するという点に着目し、音響伝達特性を識別することにより、単一マイクロホンで音源位置推定を行う手法を提案してきた。しかしこの手法は、事前に想定される音源位置毎に音響伝達特性を学習させる必要があり、学習していない位置の推定が困難であった。そこで本稿では、限られた位置の音響伝達特性を用いて、音響伝達特性から位置への回帰モデルを学習し、その回帰モデルにより未学習位置の推定を行う手法について検討する。回帰モデルとして、線形回帰である重回帰分析、非線形回帰である GPR (Gaussian Process Regression), SVR (Support Vector Regression) を用い、さらにその学習方法として、評価データに類似した学習サンプルのみから回帰モデルを学習する局所的回帰を検討し、その性能を評価した。

キーワード 音源位置推定, 音響伝達特性, 重回帰分析, Gaussian process regression, support vector regression

Interpolation of unlearned position based on local regression for single-channel talker localization using acoustic transfer function

Ryoichi TAKASHIMA[†], Tetsuya TAKIGUCHI^{††}, and Yasuo ARIKI^{††}

[†] Graduate School of System Informatics, Kobe University
Rokkodaicho 1-1, Nada-ku, Kobe, Hyogo, 657-8501 Japan

^{††} Organization of Advanced Science and Technology, Kobe University
Rokkodaicho 1-1, Nada-ku, Kobe, Hyogo, 657-8501 Japan

E-mail: [†]takashima@me.cs.scitec.kobe-u.ac.jp, ^{††}{takigu,ariki}@kobe-u.ac.jp

Abstract This paper presents a sound source (talker) localization method using only a single microphone. In our previous work, we discussed the single-channel sound source localization method based on the discrimination of the acoustic transfer function. However, that method requires to train the acoustic transfer function for each possible position in advance, and it is difficult to estimate the position that have not been pre-trained. In this paper, we discuss a single-channel talker localization method based on a regression model, which predicts the position from the acoustic transfer function. For training the regression models, we use the local regression method that trains the regression model from only training samples being similar to the evaluation data. Considering both of linear and non-linear regression models, the effectiveness of this method has been confirmed by talker localization experiments performed in different room environments.

Key words talker localization, acoustic transfer function, multiple regression, Gaussian process regression, support vector regression

1. はじめに

これまでに、マイクロホンアレーを用いて音源方向や位置を推定する研究が多くなされている。MUSIC (Multiple Signal Classification) や CSP (Cross-power Spectrum Phase) といった従来手法では、マイクロホンアレーで収録される観測信号間の位相差を用いて音源方向や位置を推定している [1] ~ [4]。またパイノール信号を用いて、両耳間の音圧差や時間差から音源方向を推定する手法についても研究がされている [5], [6]。しかしマイクロホンアレーを用いた音声インターフェースは、システムが大規模になりやすく、特に小型な音声インターフェースが必要とされる環境では、単一マイクロホンで行える音声処理技術が必要であると言える。

単一マイクロホンによる音響処理手法は信号処理の分野において多く研究がされており、近年では雑音抑圧や音源分離の分野において、単一マイクロホンによる手法の提案がなされている [7] ~ [10]。また音源方向推定の分野においても単一マイクロホンによる手法がいくつか提案されており [11], [12]、それらの手法では耳介により得られる頭部伝達特性 (HRTF) を用いて音源位置を推定している。

我々はこれまで、観測された音声信号の音響伝達特性が、発話された位置によって異なるという点に着目して、位置毎に発話された音声から音響伝達特性を推定し、それらを識別することにより単一マイクロホンで音源位置を推定する手法を提案してきた [13]。この手法ではまず、ある位置から発話された音声からその音響伝達特性を、特定話者のクリーン音声 HMM (Hidden Markov Model) を用いて推定し、推定された音響伝達特性を位置毎に学習する。その後、ある位置から発話された評価音声についても同様に音響伝達特性を推定し、推定された音響伝達特性を識別することで音源の位置をパターン認識的に推定する。

この手法を用いることで、あらかじめ学習した位置であれば、単一マイクロホンでも音源位置を識別することが可能となる。しかし、この手法はパターン認識に基づく手法であるため、学習していない位置の推定が困難であった。そこで本稿では、限られた位置の音響伝達特性を用いて、音響伝達特性から位置への回帰モデルを学習し、その回帰モデルにより未学習位置の推定を行う手法について検討する。回帰モデルとして、線形回帰である重回帰分析、非線形回帰である GPR (Gaussian Process Regression) [14]、SVR (Support Vector Regression) [15] を用い、さらにその学習方法として、評価データに類似した学習サンプルのみから回帰モデルを学習する局所的回帰を検討し、その性能を評価した。

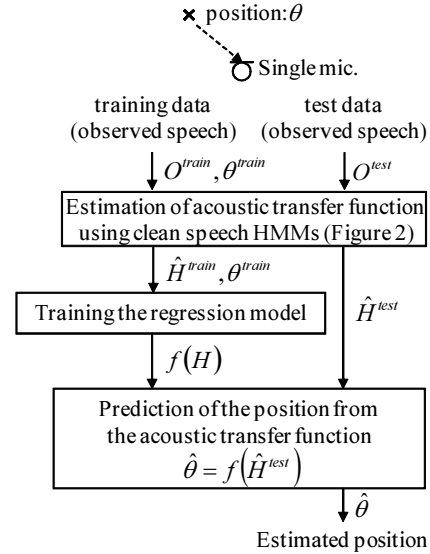


図1 提案手法の概要

2. 音源位置の推定

2.1 提案手法の概要

本研究では音響伝達特性を用いて音源の位置を推定している。音響伝達特性は音源の位置によって異なる値を持つため、これを用いて音源の位置を推定することができる。以前に提案していた手法では、あらかじめ位置毎に音響伝達特性を学習し、評価音声に対してその音響伝達特性を識別することで音源の位置を推定していた。しかしこの手法は伝達特性を学習した位置のみしか識別できないため、想定する位置の候補が増える度に、その位置の伝達特性を取得して学習する必要があった。

そこで本手法では、限られた位置の音響伝達特性を用いて、音響伝達特性から位置への回帰モデルを学習することで、学習データに含まれなかった位置についても、その回帰モデルを用いて推定を行う。

提案手法の概要を図1に示す。まず、学習用の音響伝達特性を得るために、特定の位置 θ^{train} で発話された音声 O^{train} を収録し、その音響伝達特性をクリーン音声の音素 HMM を用いて推定する。次に、学習データから推定された音響伝達特性のケプストラム \hat{H}^{train} と位置のラベル θ^{train} とのペアを用いて、音響伝達特性から位置への回帰モデル $f(H)$ を学習する。

そして評価したい音声 O^{test} についても学習データと同様にして音響伝達特性 \hat{H}^{test} を推定し、学習した回帰モデルを用いて音源位置 $\hat{\theta}$ を推定する。

2.2 音素 HMM による音響伝達特性の推定

本節では音素 HMM を用いて観測信号 O から音響伝達特性 H を推定する手法について述べる [13]。ある場所で発話されたクリーン音声 S は、音響伝達特性 H の影響を受けて観測される。このとき、フレーム n における観測信号 O のケプストラムを、

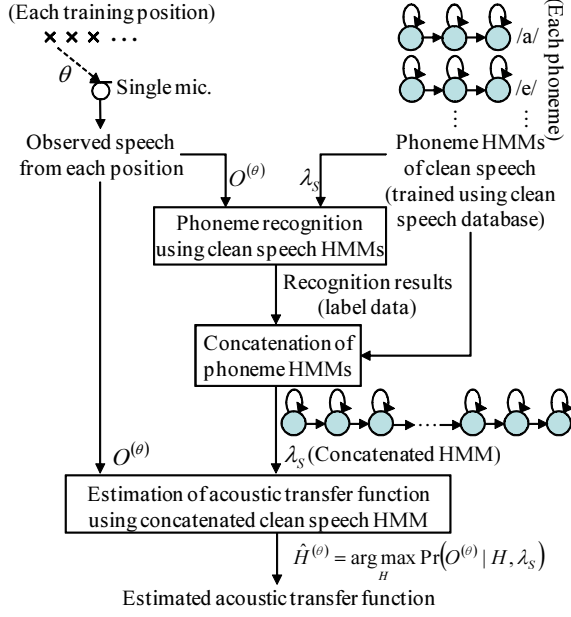


図 2 音素 HMM による音響伝達特性の推定

$$O_{cep}(d; n) \approx S_{cep}(d; n) + H_{cep}(d; n) \quad (1)$$

と近似する。 d はケプストラムの次元を表す。ケプストラムは、音声情報を効率よく表現できるパラメータの一つであり、音声認識などでよく用いられていることから、本手法においてもケプストラムを特徴量として用いている。仮に S が既知であれば、音響伝達特性 H は式 (1) から直接求めることができるが、実際の環境では S が未知であるため、直接 H を求めることはできない。そこで、 S の統計モデルをあらかじめ学習しておき、最尤推定法により O から H を推定する。

音響伝達特性の推定の流れを図 2 に示す。あらかじめ特定話者のクリーン音声の MFCC を音素 HMM でモデル化しておく。HMM を用いて音響伝達特性を推定するためには、その音声信号の音素ラベルが必要であるため、まず学習した音素 HMM を用いて観測信号を音素認識する。そして出力された音素認識結果をラベルとして音素 HMM を連結し、連結された HMM を用いて観測信号から最尤推定法により音響伝達特性の MFCC を推定する。

$$\hat{H} = \operatorname{argmax}_H \Pr(O | \lambda_S, H) \quad (2)$$

λ_S はクリーン音声のモデルパラメータを表す。(2) 式の解は EM アルゴリズムによって推定される。その際、 Q 関数は以下のように定義される。

$$\begin{aligned} Q(\hat{H} | H) &= E[\log \Pr(O, p, b_p, c_p | \hat{H}, \lambda_S) | H, \lambda_S] \\ &= \sum_p \sum_{b_p} \sum_{c_p} \frac{\Pr(O, p, b_p, c_p | H, \lambda_S)}{\Pr(O | H, \lambda_S)} \\ &\quad \cdot \log \Pr(O, p, b_p, c_p | \hat{H}, \lambda_S) \end{aligned} \quad (3)$$

b_p と c_p はそれぞれ音素 p における HMM の状態、混合要素を表す。 O, p, b, c の同時確率 $\Pr(O, p, b_p, c_p | \hat{H}, \lambda_S)$

は以下のように展開される。

$$\begin{aligned} &\Pr(O, p, b_p, c_p | \hat{H}, \lambda_S) \\ &= \prod_n a_{b_p(n-1), b_p(n)} w_{b_p(n), c_p(n)} \\ &\quad \cdot \Pr(O(n) | p, b_p(n), c_p(n); \hat{H}, \lambda_S) \end{aligned} \quad (4)$$

n, a, w はそれぞれフレーム番号、状態遷移確率、混合重みを表す。ここで、(1) 式より O は S と H の加算とみなされるため、 O の事後確率をクリーン音声 HMM を用いて以下のように表すことができる。

$$\begin{aligned} &\Pr(O, p, b_p, c_p | \hat{H}, \lambda_S) \\ &= \prod_n a_{b(n-1), b(n)} w_{b(n), c(n)} \\ &\quad \cdot \mathcal{N}(O(n); \mu_{p,j,k}^{(S)} + \hat{H}(n), \Sigma_{p,j,k}^{(S)}) \end{aligned} \quad (5)$$

$\mathcal{N}(O; \mu, \Sigma)$ は多次元正規分布を表し、 $\mu_{p,j,k}^{(S)}$ 、 $\Sigma_{p,j,k}^{(S)}$ はそれぞれ S の状態 $b(n) = j$ 、混合要素 $c(n) = k$ における平均ベクトルと共分散行列 (対角行列) を表す。これらを用いて (3) 式を展開し、 H に関わる項のみを取り出すと以下ようになる。

$$\begin{aligned} &Q(\hat{H} | H) \\ &= - \sum_p \sum_j \sum_k \sum_n \gamma_{p,j,k}(n) \\ &\quad - \sum_{d=1}^D \left\{ \frac{1}{2} \log(2\pi)^D \sigma_{p,j,k,d}^{(S)2} \right. \\ &\quad \left. + \frac{(O(d;n) - \mu_{p,j,k,d}^{(S)} - \hat{H}(d;n))^2}{2\sigma_{p,j,k,d}^{(S)2}} \right\} \end{aligned} \quad (6)$$

$$\gamma_{p,j,k}(n) = \Pr(O(n), p, j, k | \lambda_S) \quad (7)$$

D は次元数、 $\mu_{p,j,k,d}^{(S)}$ 、 $\sigma_{p,j,k,d}^{(S)2}$ はそれぞれ平均ベクトルの d 次元目の値と、共分散行列の d 番目の対角要素の値を表す。(6) 式を最大にする H は、 $\partial Q(\hat{H} | H) / \partial \hat{H} = 0$ を解くことで求められる。

$$\hat{H}(d; n) = \frac{\sum_p \sum_j \sum_k \gamma_{p,j,k}(n) \frac{O(d;n) - \mu_{p,j,k,d}^{(S)}}{\sigma_{p,j,k,d}^{(S)2}}}{\sum_p \sum_j \sum_k \frac{\gamma_{p,j,k}(n)}{\sigma_{p,j,k,d}^{(S)2}}} \quad (8)$$

2.3 回帰モデルによる音源位置の推定

学習位置で得られた音響伝達特性 \hat{H}^{train} と、その位置のラベル θ^{train} を用いて、音響伝達特性から位置への回帰モデルを学習する。本研究では回帰モデルとして、線形回帰である重回帰分析、非線形回帰である GPR [14]、SVR [15] を用い、さらにその学習方法として、評価データに類似した学習サンプルのみから回帰モデルを学習する局所的回帰を用いた。

2.3.1 Gaussian Process Regression

学習データ集合として、説明変数である音響伝達特性と目的変数である位置の組 $Z_n^{train} = (\hat{H}_n^{train}, \theta_n^{train})$ 、 $(n = 1, \dots, N)$ が得られている場合において、評価データの音響伝達特性 \hat{H}^{test} についてその音源位置 θ^{test} を推定することを考える。GPR の枠組みでは、学習データの組と評

価データの説明変数を観測したときの目的変数の事後確率は、以下のような正規分布に従うと仮定されている．

$$\Pr(\theta^{test} | \hat{H}^{test}, Z_1^{train}, \dots, Z_N^{train}) \sim \mathcal{N}(\mathbf{K}_* \mathbf{K}^{-1} \Theta, \mathbf{K}_{**} - \mathbf{K}_* \mathbf{K}^{-1} \mathbf{K}_*^T) \quad (9)$$

ここで $\Theta = [\theta_1^{train}, \dots, \theta_N^{train}]^T$, \mathbf{K} はグラム行列

$$\mathbf{K} = \begin{bmatrix} k(\hat{H}_1^{train}, \hat{H}_1^{train}) & \dots & k(\hat{H}_1^{train}, \hat{H}_N^{train}) \\ \vdots & \ddots & \vdots \\ k(\hat{H}_N^{train}, \hat{H}_1^{train}) & \dots & k(\hat{H}_N^{train}, \hat{H}_N^{train}) \end{bmatrix}$$

$$\mathbf{K}_* = [k(\hat{H}^{test}, \hat{H}_1^{train}), \dots, k(\hat{H}^{test}, \hat{H}_N^{train})] \quad (10)$$

$$\mathbf{K}_{**} = k(\hat{H}^{test}, \hat{H}^{test})$$

を表す． $k(H, H)$ はカーネル関数であり、本研究では RBF カーネル [14] を使用している．

評価データに対する目的変数の推定値 $\hat{\theta}^{test}$ は、式 (9) を最大とする値、すなわち正規分布の平均値がとられる．

$$\hat{\theta}^{test} = f(\hat{H}^{test}) = \mathbf{K}_* \mathbf{K}^{-1} \Theta \quad (11)$$

2.3.2 Support Vector Regression

カーネルを用いた非線形回帰である SVR では目的変数は以下の回帰モデルにより推定される．

$$f(\hat{H}^{test}) = w^T \phi(\hat{H}^{test}) + b \quad (12)$$

ϕ は高次元空間への写像関数を表す．回帰モデルのパラメータ w, b は以下の式を満たすように計算される [15] ．

$$\begin{aligned} \min \quad & \frac{1}{2} \|w\|^2 + C \sum_{n=1}^N (\xi_n + \xi_n^*) \quad (13) \\ \text{s.t.} \quad & \begin{cases} \theta_n^{train} - w^T \phi(\hat{H}_n^{train}) - b \leq \epsilon + \xi_n \\ w^T \phi(\hat{H}_n^{train}) + b - \theta_n^{train} \leq \epsilon + \xi_n^* \\ \xi_n, \xi_n^* \geq 0 \end{cases} \quad (14) \end{aligned}$$

式 (14) の第 1 項は正則化項であり、第 2 項は許容誤差 ϵ を超える回帰誤差 ξ に対するペナルティ項、 C は第 1 項と第 2 項のトレードオフを決定する変数である．この双対問題は以下の二次計画問題に書き換えられる．

$$\begin{aligned} \max \quad & -\frac{1}{2} \sum_{n,m=1}^N (\alpha_n - \alpha_n^*)(\alpha_m - \alpha_m^*) k(\hat{H}_n^{train}, \hat{H}_m^{train}) \\ & - \epsilon \sum_{n=1}^N (\alpha_n + \alpha_n^*) + \sum_{n=1}^N \theta_n^{train} (\alpha_n - \alpha_n^*) \quad (15) \\ \text{s.t.} \quad & \sum_{n=1}^N (\alpha_n - \alpha_n^*) = 0 \text{ and } \alpha_n, \alpha_n^* \in [0, C] \quad (16) \end{aligned}$$

α, α^* はラグランジュ乗数であり、カーネル関数 $k(H, H)$ は本研究では RBF カーネルを用いている．この式を解くことで得られる α, α^* を用いて、式 (12) が以下のように書き換えられ、最終的な回帰式が得られる．

$$f(\hat{H}^{test}) = \sum_{n=1}^N (\alpha_n - \alpha_n^*) k(\hat{H}_n^{train}, \hat{H}^{test}) + b \quad (17)$$

2.3.3 局所的回帰

本研究では各学習位置の音響伝達特性から回帰モデル

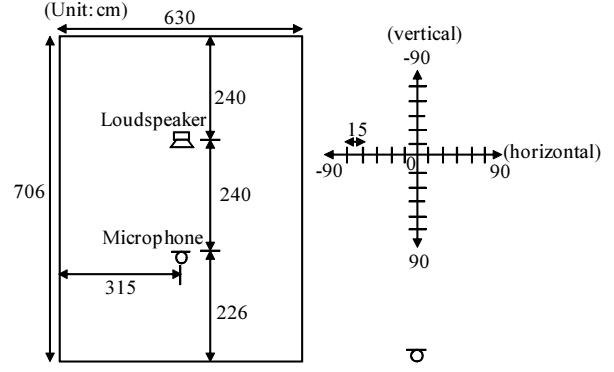


図 3 水平・垂直方向の位置推定実験における部屋環境 (左) とスピーカーの位置 (右)

を学習させることを検討している．これは例えば、近い位置の音響伝達特性は似たような値になるといった、位置と伝達特性の相関性を仮定していることになる．しかしながら部屋の形状によっては、近い位置同士であっても、それらの音響伝達特性が大きく異なるといった現象も起こり得る．そういった可能性を考慮した場合、部屋内の全ての位置の音響伝達特性を、一つの回帰関数でモデル化することは困難であると考えられる．

そこで本研究では、学習データサンプル全体から一つの回帰モデルを作成するのではなく、評価データに類似した学習サンプルのみで回帰モデルを作成する局所的回帰により、回帰モデルを学習させる．Chao ら [16] は顔画像からの年齢推定において、回帰モデルに SVR を用いた局所的回帰により、通常の SVR よりも高い推定精度を示している．

この手法は K-NN (K-Nearest Neighbor) 法と回帰分析を組み合わせた手法で、学習データサンプルを全てデータベースとして保持しておく．そして評価データに対して、その K 近傍のサンプルを学習データから取り出す．その後、取り出した K サンプルのみで回帰モデルを学習し、その回帰モデルに従って評価データの音源位置を予測する．

3. 評価実験

3.1 実験環境

提案手法を評価するために、二種類の部屋で収録されたインパルス応答を、クリーン音声に積みこむことで残響音声信号をシミュレートし、音源位置推定の実験を行った．一方の部屋環境では水平方向、垂直方向 (奥行き) の位置推定を行った．インパルス応答の収録環境と、スピーカーの位置を図 3 に示す．スピーカーの位置は水平方向、垂直方向ともに -90cm から 90cm まで 15cm 間隔でスピーカーを移動させ、計 25 箇所インパルス応答を測定した．この内、水平・垂直方向どちらも、-90, -60, -30, 0, 30, 60, 90cm の 7 箇所を学習位置とし、評価位置は学習位置を含む全収録位置とした．よって、評価位置 13 箇所の

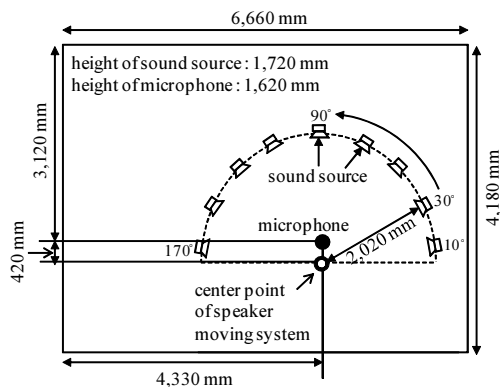


図 4 音源角度の推定実験における部屋環境

内 6 箇所が未学習位置となる．残響時間は約 1,220msec である．この実験において，位置の推定は水平方向のみ，垂直方向のみでそれぞれ独立に行っており，一方を評価するとき他方は 0cm に固定（既知と仮定）している．

もう一方の部屋環境では音源距離を固定（既知と仮定）し，角度の推定を行った．インパルス応答は RWCP 実環境音声・音響データベース [17] において収録されたものを用いた．この実験における部屋環境を図 4 に示す．音源の角度は 10° から 170° まで 20° 間隔で収録されており，この内 10°，50°，90°，130°，170° を学習位置とし，評価位置は学習位置を含む全収録位置とした．よって評価位置 9 箇所の内 4 箇所が未学習位置となる．残響時間は約 300msec である．

音声データは ATR 研究用日本語音声データベースセット A より男性話者 1 名の単語音声を用い，サンプリング周波数 12 kHz，窓幅 32 msec，フレームシフト 8 msec の分析条件で MFCC 16 次元を特徴量として使用した．音響伝達特性の推定におけるクリーン音声の音素 HMM は，2,620 単語を用いて学習した．音素数は 54，各音素 HMM の状態数は 3，混合数は 32 である．

回帰モデルの学習には，クリーン音声 50 単語に，学習位置のインパルス応答を畳みこむことで作成した残響音声 50 単語（× 学習位置数）を用いた．評価の際には text closed の場合と text open の場合に分けて評価した．まず text closed の実験では，回帰モデルの学習に用いたデータと同じクリーン音声 50 単語に，評価位置のインパルス応答を畳みこんで作成した残響音声 50 単語（× 評価位置数）を用いた．この場合，既学習位置の評価データについては，完全 closed データとなっている．一方 text open の実験では，回帰モデルの学習に用いたデータとは異なるクリーン音声 116 単語に，評価位置のインパルス応答を畳みこんで作成した残響音声 116 単語（× 評価位置数）を用いた．なお，クリーン音声 HMM の学習に用いたデータと残響音声の作成に使用したデータはそれぞれ異なる発話内容の単語である．

3.2 実験結果

各方向毎の音源位置を，重回帰分析 (MR)，GPR，SVR

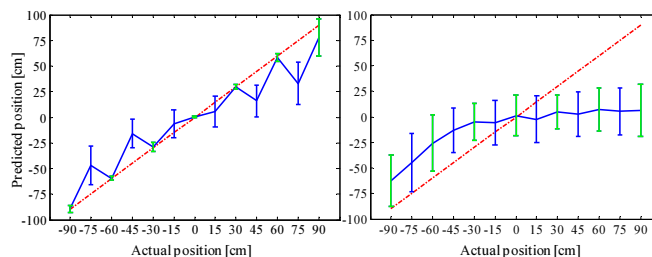


図 5 SVR に基づく局所的回帰による水平方向の位置推定結果（左図：text closed，右図：text open）

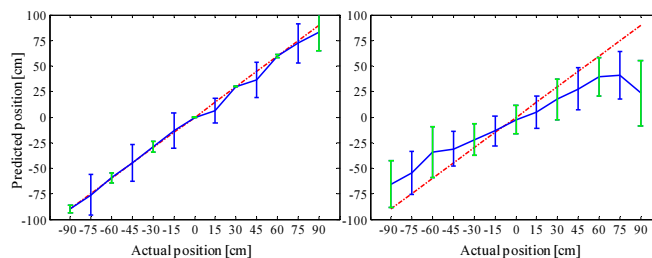


図 6 SVR に基づく局所的回帰による垂直方向の位置推定結果（左図：text closed，右図：text open）

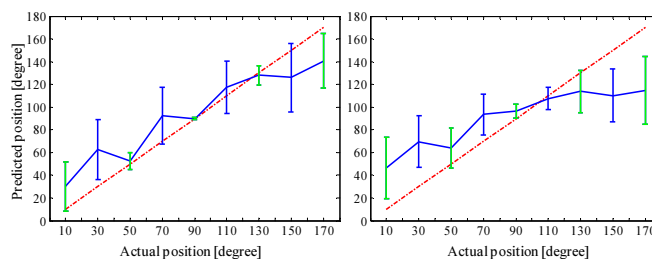


図 7 SVR に基づく局所的回帰による音源角度の推定結果（左図：text closed，右図：text open）

それぞれで推定した値と実測値との相関係数を表 1 に示す．局所的回帰における K は 75 とした．数値の斜線左側は未学習位置の評価データのみで，斜線右側は既学習位置の評価データも含めて，それぞれ相関係数を求めた場合の値を表す．表より，線形回帰モデルである重回帰分析よりも非線形回帰モデルである GPR，SVR の方が高い相関値を示している．また，text open の場合よりも text closed の方が未学習位置のみで相関係数を求めた場合でも高い値を示している．本来，音響伝達特性は発話内容に依存しないため，理想的には text open であっても text closed と同じ値になるはずである．しかし実際には音響伝達特性が完全に推定しきれておらず，クリーン音声成分の影響を受けているため，text open では相関値が低下してしまったと考えられる．

回帰モデルに SVR を用いた場合のそれぞれの位置推定結果を図 5，図 6，図 7 に示す．青線は推定された値の平均及び標準偏差を表しており，これが実測値である赤線に近いほど回帰モデルが優れていることを表している．また，既学習位置については緑線で表示している．図より，垂直（奥行き）方向の変化においては，比較的正確

表 1 水平方向 (horizontal), 垂直方向 (vertical), 音源角度 (direction) それぞれに対する音源位置の推定値と実測値との相関係数 (斜線左側は未学習位置の評価データのみで, 右側は既学習位置の評価データも含めてそれぞれ相関係数を求めた場合の値を表す.)

	horizontal		vertical		direction	
	text closed	text open	text closed	text open	text closed	text open
MR	0.57/0.72	0.39/0.48	0.89/0.90	0.70/0.69	0.68/0.74	0.53/0.56
GPR	0.77/0.92	0.49/0.56	0.94/0.98	0.82/0.80	0.62/0.86	0.58/0.66
SVR	0.82/0.92	0.51/0.58	0.94/0.97	0.86/0.83	0.67/0.83	0.61/0.70

に推定されているが, 一方, 水平方向や音源角度の変化においては, 未学習位置の推定誤差が大きくなっていることが分かる. また, text open の場合では既学習位置であっても推定値の偏差が大きくなっており, 全体的に性能が低下していることが分かる.

4. おわりに

本稿では, 音源位置毎に異なる音響伝達特性に着目し, 音声信号から音響伝達特性をクリーン音声 HMM を用いて推定し, 推定された音響伝達特性から回帰モデルにより音源位置を推定する手法について検討した. 水平方向・垂直方向 (奥行き)・音源角度の 3 つの場合で音源位置の推定性能を評価した結果, 特に垂直方向の推定が高い値を示した. 一方, 水平方向や音源角度では推定誤差が大きく, 今回用いた学習位置の情報だけでは回帰モデルを学習することは困難であると言える. また, text open では推定性能が低下するため, 音響伝達特性のより正確な推定が必要である. 今後は垂直方向・水平方向の両方を動かした場合の評価と, より少ない位置での回帰モデルの学習について検討を行う.

5. 謝 辞

本研究は日本学術振興会特別研究員奨励費 (23-2495) の助成を受けたものである.

文 献

- [1] D. Johnson and D. Dudgeon, Array Signal Processing, Prentice Hall, Upper Saddle River, NJ, USA, 1996.
- [2] M. Omologo and P. Svaizer, "Acoustic source location in noisy and reverberant environment using CSP analysis," Proc. ICASSP96, vol.2, pp.921-924, Atlanta, Ga, USA, May 1996.
- [3] F. Asano, H. Asoh, and T. Matsui, "Sound source localization and separation in near field," IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences, vol.E83-A, no.11, pp.2286-2294, 2000.
- [4] Y. Denda, T. Nishiura, and Y. Yamashita, "Robust talker direction estimation based on weighted CSP analysis and maximum likelihood estimation," IEICE Trans. on Information and Systems, vol.E89-D, no.3, pp.1050-1057, 2000.
- [5] F. Keyrouz, Y. Naous, and K. Diepold, "A new method for binaural 3-D localization based on HRTFs," Proc. ICASSP06, vol.5, pp.341-344, Toulouse, France, May 2006.
- [6] M. Takimoto, T. Nishino, and K. Takeda, "Estimation of a talker and listener's positions in a car using binaural signals," Proc. the 4th Joint Meeting ASA and ASJ (ASA/ASJ06), p.3216, Honolulu, Hawaii, USA, Nov. 2006.
- [7] T. Kristjansson, H. Attias, and J. Hershey, "Single microphone source separation using high resolution signal reconstruction," Proc. ICASSP04, vol.2, pp.817-820, Montreal, Canada, May 2004.
- [8] B. Raj, M.V.S. Shashanka, and P. Smaragdis, "Latent dirichlet decomposition for single channel speaker separation," Proc. ICASSP06, vol.5, pp.821-824, Toulouse, France, May 2006.
- [9] G.-J. Jang, T.-W. Lee, and Y.-H. Oh, "A subspace approach to single channel signal separation using maximum likelihood weighting filters," Proc. ICASSP03, vol.5, pp.45-48, Hong Kong, April 2003.
- [10] T. Nakatani, B.-H. Juang, K. Kinoshita, and M. Miyoshi, "Speech dereverberation based on probabilistic models of source and room acoustics," Proc. ICASSP06, vol.1, pp.821-824, Toulouse, France, May 2006.
- [11] A. Fuchs, C. Feldbauer, and M. Stark, "Monaural sound localization," Proc. Interspeech 2011, pp.2521-2524, Florence, Italy, Aug. 2011.
- [12] R. Kliper, H. Kayser, D. Weinsall, I. Nelken, and J. Anemuller, "Monaural azimuth localization using spectral dynamics of speech," Proc. Interspeech 2011, pp.33-36, Florence, Italy, Aug. 2011.
- [13] R. Takashima, T. Takiguchi, and Y. Ariki, "HMM-based separation of acoustic transfer function for single-channel sound source localization," ICASSP2010, pp.2830-2833, Dallas, Texas, USA, 2010.
- [14] C.E. Rasmussen and C.K.I. Williams, Gaussian Processes for Machine Learning, The MIT Press, Cambridge, MA, 2006.
- [15] A.J. Smola and B. Scholkopf, "A tutorial on support vector regression," Journal Statistics and Computing, vol.14-3, pp.199-222, Aug. 2004.
- [16] W.L. Chao, J.Z. Liu, and J.J. Ding, "Facial age estimation based on label-sensitive learning and age-specific local regression," Proc. ICASSP 2012, pp.1941-1944, 2012.
- [17] S. Nakamura, "Acoustic sound database collected for hands-free speech recognition and sound scene understanding," Proc. International Workshop on Hands-Free Speech Communication (HSC01), pp.43-46, Kyoto, Japan, April 2001.