

音響伝達特性の識別に基づく話者の頭部回転方向の推定

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。この手法では、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 HMM (Hidden Markov Model) を用いて推定し、それらを学習・識別することで音源位置を推定している。本稿では、観測信号の音響伝達特性が話者の位置だけでなく、頭部の回転方向にも依存する点にも着目し、音源位置と頭部回転方向の推定について、提案手法の有効性を評価する。

キーワード 頭部回転方向, 音源位置推定, 単一マイクロホン, 残響, 音響伝達特性

Estimation of Head Orientation Based on Discrimination of Acoustic Transfer Functions

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 talker's head orientation estimation method using only a single microphone, where phoneme HMMs (Hidden Markov Models) of clean speech are introduced to separate the acoustic transfer function at the user's position and head orientation. The frame sequence of the acoustic transfer function is estimated by maximizing the likelihood of training data uttered from a given position with a given head orientation. Using the separated frame sequence data, the user's position and the head orientation are trained by Support Vector Machine (SVM) in advance. Then, for each test utterance, the frame sequence of the acoustic transfer function is separated based on the maximum likelihood estimation using the label sequence obtained from the phoneme recognition, and the user's position and head orientation are estimated by discriminating the separated acoustic transfer function using SVM. The effectiveness of this method has been confirmed by talker localization and head orientation estimation experiments performed in a real environment.

Key words head orientation, talker localization, single channel, reverberant speech, acoustic transfer function

1. はじめに

人と人のコミュニケーション、あるいはロボットとのコミュニケーションにおいて、話者の頭部の回転方向は

聞き手にとって重要な手がかりの一つであり、我々は話し手の頭部の向きから「誰が話しているのか」だけでなく、「誰に向かって話しているのか」という情報まで得ることができる。この「誰が誰に向かって話しているのか」

という情報は、特に複数のユーザーが会話をしている状況において有効であり、会議システム、ロボット対話、雑談とシステム要求の判別など、様々なタスクにおいて利用することができる期待される。

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、バイノーラル信号を用いて、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されている [5], [6]。

一方、話者の頭部回転方向の推定へ関心が向けられ出したのは比較的近年のことであり、いくつかの手法が提案されている [7] ~ [10]。これらの手法は複数組のマイクロホンアレーからなるネットワークを用いており、従来の音源位置推定のアルゴリズムを拡張することで話者の頭部回転方向を推定している。文献 [7] で提案されている手法は、従来の音源位置推定法の一つである SRP-PHAT (Steered Response Power with the PHase Transform) をベースとした手法であり、従来の SRP-PHAT の目的関数を、話者の頭部回転方向に依存する重み係数によって重み付けを行うことにより、話者の位置推定問題から頭部回転方向推定問題へ拡張している。文献 [8], [9] では話者の頭部の方向ごとに変化する観測信号の音圧のパターンに着目しており、文献 [9] では提案手法を文献 [7] の手法と組み合わせることでさらなる精度の向上を示している。また、文献 [10] では各マイクロホンアレーから算出された音源方向推定結果を用いて作成されるヒストグラムから、話者の頭部回転方向を推定する手法を提案している。

しかしながらこれらの手法は複数のマイクロホンアレーを、ユーザーを囲むようにして部屋の壁などに設置する必要があり、システムが大規模になってしまうという欠点がある。そのため、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の需要が高まっており、近年では雑音抑圧や音源分離の分野においても、単一マイクロホンで処理できる手法が多く提案されてきている [11] ~ [14]。

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方法を提案してきた [15]。この手法では、ある位置から発話された音声からその音響伝達特性を、特定話者の音素 HMM 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、それを識別することで音源の位置を推定する。

本稿では観測信号の音響伝達特性が、話者の位置だけではなく頭部の回転方向にも依存することに着目し、音響伝達特性の識別による話者の頭部回転方向を推定する手法を提案する。以前に提案した音源位置推定手法では、話者の位置ごとの音響伝達特性を学習・識別していたのに対し、本稿における提案手法では、各音源位置とその位置における頭部の各回転方向の音響伝達特性を学習・識別する。従来の頭部回転方向の推定法と異なり、本手法はあらかじめ音響伝達特性を学習しておく必要があるが、マイクの位置を任意の場所に設置することができるという利点がある。評価実験では実環境下において音源位置のみの推定、頭部回転方向のみの推定、音源位置及び頭部回転方向の推定の 3 つのタスクにおいて実験を行い、その有効性を示す。

2. 音源位置と頭部回転方向の推定

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と頭部の回転方向を推定する。音響伝達特性は音源の位置や頭部の回転方向によって異なる値を持つため、あらかじめこれを各音源位置とその位置における頭部の回転方向毎に学習しておけば、評価音声に対してもその音響伝達特性を識別することで音源位置及び頭部の回転方向を推定することができる。

提案手法の概要を図 1 に示す。まず、位置と頭部方向の組み合わせ毎の音響伝達特性を学習するために、それぞれの位置 θ において頭部を各方向 ϕ へ向けた状態で発話された音声 $O_{train}^{(\phi, \theta)}$ 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、位置と頭部方向毎に推定された音響伝達特性 $\hat{H}_{train}^{(\phi, \theta)}$ を SVM (Support Vector Machine) により学習する。そして評価したい音声 $O_{test}^{(\phi, \theta)}$ についても学習データと同様に、音素認識結果により得られるラベル情報を用いて音響伝達特性 $\hat{H}_{test}^{(\phi, \theta)}$ を推定し、それを SVM で識別することで、音源位置と頭部方向 $(\hat{\phi}, \hat{\theta})$ 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。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 のケプストラムは、

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

と近似される。 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

$$H_{cep}(d; n) \approx O_{cep}(d; n) - S_{cep}(d; n) \quad (2)$$

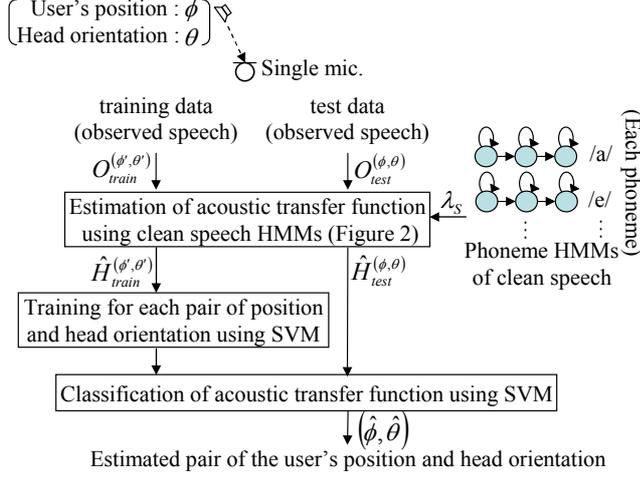


図 1 提案手法の概要

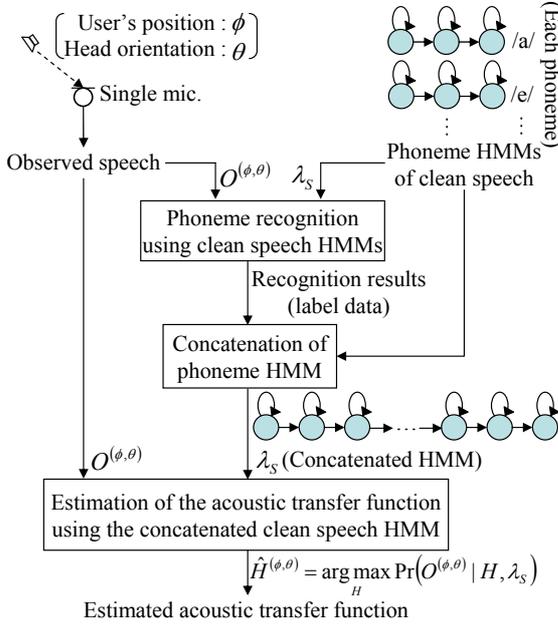


図 2 音素 HMM による音響伝達特性の推定

として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。

$$\hat{H} = \underset{H}{\operatorname{argmax}} \Pr(O | \lambda_S, H) \quad (3)$$

λ_S はクリーン音声のモデルパラメータを表す。(3) 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。

$$\begin{aligned} Q(\hat{H} | H) &= E[\log \Pr(O, p, b_p, c_p | \hat{H}, \lambda_S) | H, \lambda_S] \\ &= \sum_p \sum_{b_p} \sum_{c_p} \frac{\Pr(O, p, b_p, c_p | H, \lambda_S)}{\Pr(O | H, \lambda_S)} \\ &\quad \cdot \log \Pr(O, p, b_p, c_p | \hat{H}, \lambda_S) \end{aligned} \quad (4)$$

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。

$$\begin{aligned} \Pr(O, p, b_p, c_p | \hat{H}, \lambda_S) &= \prod_n a_{b_p(n-1), b_p(n)} w_{b_p(n), c_p(n)} \\ &\quad \cdot \Pr(O(n) | p, b_p(n), c_p(n); \hat{H}, \lambda_S) \end{aligned} \quad (5)$$

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。

$$\begin{aligned} \Pr(O, p, b_p, c_p | \hat{H}, \lambda_S) &= \prod_n a_{b(n-1), b(n)} w_{b(n), c(n)} \\ &\quad \cdot N(O(n); \mu_{p,j,k}^{(S)} + \hat{H}(n), \Sigma_{p,j,k}^{(S)}) \end{aligned} \quad (6)$$

$N(O; \mu, \Sigma)$ 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 (4) 式を展開すると [16]、

$$\begin{aligned} Q(\hat{H} | H) &= \sum_p \sum_i \sum_j \sum_n \\ &\quad \Pr(O(n), p, b_p(n) = j, b_p(n-1) = i | H, \lambda_S) \log a_{p,i,j} \\ &\quad + \sum_p \sum_j \sum_k \sum_n \\ &\quad \Pr(O(n), p, b_p(n) = j, c_p(n) = k | H, \lambda_S) \log w_{p,j,k} \\ &\quad + \sum_p \sum_j \sum_k \sum_n \\ &\quad \Pr(O(n), p, b_p(n) = j, c_p(n) = k | H, \lambda_S) \\ &\quad \cdot \log N(O(n); \mu_{p,j,k}^{(S)} + \hat{H}(n), \Sigma_{p,j,k}^{(S)}) \end{aligned} \quad (7)$$

となり、 H に関わる項のみを取り出すと以下のようになる。

$$\begin{aligned} Q(\hat{H} | H) &= - \sum_p \sum_j \sum_k \sum_n \gamma_{p,j,k}(n) \\ &\quad - \sum_{d=1}^D \left\{ \frac{1}{2} \log(2\pi)^D \sigma_{p,j,k,d}^{(S)2} \right. \\ &\quad \left. + \frac{(O(d;n) - \mu_{p,j,k,d}^{(S)} - \hat{H}(d;n))^2}{2\sigma_{p,j,k,d}^{(S)2}} \right\} \end{aligned} \quad (8)$$

$$\gamma_{p,j,k}(n) = \Pr(O(n), p, j, k | \lambda_S) \quad (9)$$

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

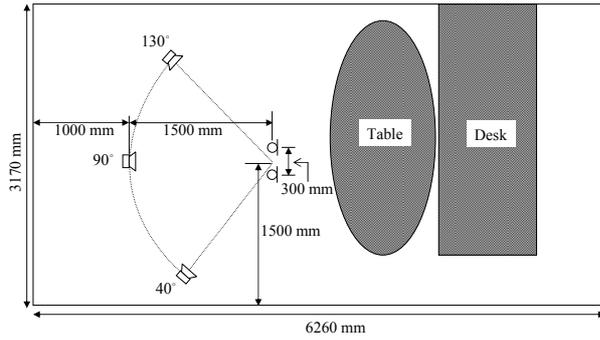


図3 実験環境

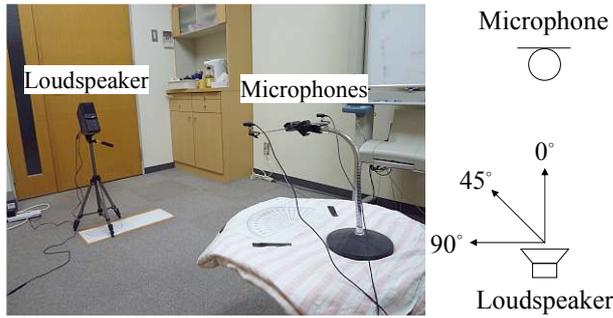


図4 収録環境のとスピーカークの回転方向

を表す。(8)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。

$$\hat{H}(d;n) = \frac{\sum_p \sum_j \sum_k \gamma_{p,j,k}(n) \frac{O(d;n) - \mu_{p,j,k,d}^{(S)}}{\sigma_{p,j,k,d}^{(S)2}}}{\sum_p \sum_j \sum_k \frac{\gamma_{p,j,k}(n)}{\sigma_{p,j,k,d}^{(S)2}}}. \quad (10)$$

(10)式によって、音源位置と頭部回転方向毎に音響伝達特性を推定し、それらをSVMによって学習・識別することで、音源位置と頭部回転方向の推定を行う。

3. 評価実験

3.1 実験環境

提案手法を評価するために、特定話者による実環境実験を行った。実験環境の図を図3に、収録環境とスピーカークの回転を表す図を図4に示す。約6.3 m × 3.2 m × 2.8 m (W × D × H)の部屋において、ある位置にスピーカークを設置し、スピーカークの向きを変えながら特定話者の音声を再生し、これを2chマイクロホンで収録した。ただし、提案手法では2chマイクロホンの内、一方のマイクロホンで収録された音声のみを用いた。部屋の残響時間は約350 msec、マイクロホンとスピーカークの距離は約1.5 mである。音源位置90°、スピーカークの回転方向0°におけるインパルス応答を図5に示す。スピーカークはBOSE Mediamate IIを、マイクロホンには指向性マイク(SONY ECM-66B)を使用した。音源位置の候補は40°、90°、130°の3種類、スピーカークの回転方向は0°、45°、90°の3種類で、これらの組み合わせは最大9通り存在する。

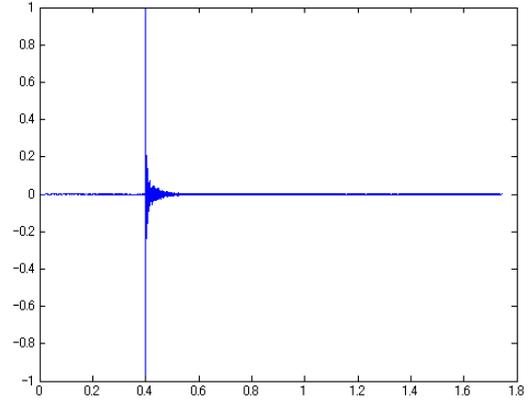


図5 音源位置90°、スピーカークの回転方向0°におけるインパルス応答

表1 それぞれの位置における音源位置推定精度と、CSP法により推定された音源方向の中央値

Position	40 deg.	90 deg.	130 deg.	average
Accuracy [%]	83.6	93.5	98.3	91.8
CSP [deg.]	40.9	90	131.4	-

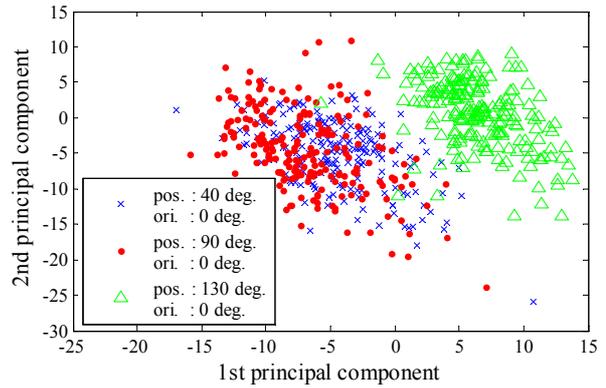


図6 各音源位置における音響伝達特性

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。音響伝達特性の学習には50単語を、評価には166単語を、組み合わせを変えて4-foldのクロスバリデーションにより推定精度を算出した。なお、クリーン音声の学習、位置の学習、評価に用いたデータはそれぞれ異なる発話内容の単語を使用している。SVMには SVM^{light} [17]を、カーネル関数にRBF (Gaussian)カーネルを使用し、one-vs-rest法によりマルチクラス識別を行った。

3.2 実験結果

3.2.1 音源位置推定における実験結果

まず、それぞれの音源位置において、スピーカークの回転方向を0°に限定して、音源位置推定の精度の評価を行っ

表 2 それぞれの位置 (pos.) における, スピーカーの回転方向 (ori.) の推定精度. 上表は回転方向の候補数が 2(0°, 90°) における推定精度を, 下表は候補数が 3(0°, 45°, 90°) における推定精度を表す.

pos. \ ori.	0 deg.	90 deg.	average
40 deg.	81.6	84.3	83.0
90 deg.	96.1	92.6	94.4
130 deg.	94.4	93.7	94.1
average	90.7	90.2	90.5

pos. \ ori.	0 deg.	45 deg.	90 deg.	average
40 deg.	73.0	20.0	86.7	59.9
90 deg.	97.1	10.2	90.1	65.8
130 deg.	82.8	33.7	97.1	71.2
average	84.3	21.3	91.3	65.7

た. 音源位置毎の推定精度と, CSP 法 [2] によって単語毎に推定された音源方向の中央値を表 1 に示す. CSP 法は GCC-PHAT (Generalized Cross-Correlation PHase Transform) アルゴリズムとも呼ばれる手法で, 音源方向や音源位置推定によく用いられる手法の一つである. CSP 法による音源方向の推定には, 30 cm の間隔で設置された 2ch マイクロホンを用いて, 以下の式によって計算された CSP 係数ピークを探索することで, 観測信号間の位相差を計算し, 位相差から音源方向を推定した.

$$CSP(\tau) = DFT^{-1} \left\{ \frac{DFT(o_l(t)) \cdot DFT^*(o_r(t))}{|DFT(o_l(t))| \cdot |DFT(o_r(t))|} \right\} \quad (11)$$

$$\hat{\tau} = \underset{\tau}{\operatorname{argmax}} CSP(\tau) \quad (12)$$

$$\text{degree} = \cos^{-1} \left(\frac{\hat{\tau} \cdot C}{ds \cdot fs} \right) \cdot \frac{180}{\pi} \quad (13)$$

$o_l(t)$ と $o_r(t)$ はそれぞれ左のチャンネルと右のチャンネルで観測された信号で, $\hat{\tau}$ は推定されたこれらの信号間の位相差, C , ds , fs はそれぞれ音速, マイクの間隔, サンプリング周波数を表す.

表より, CSP 法ではいずれの音源位置においても安定して音源方向が推定されているのに対し, 提案手法では推定精度が音源の位置毎に差があることが分かる. 図 6 は各音源位置における音響伝達特性の単語毎の平均値をプロットしたものである. 音響伝達特性は (2) 式に正解のクリーン音声を与えることで算出し, 主成分分析により 2 次元に圧縮したものをプロットしている. 図より, 130° における音響伝達特性は識別しやすい分布をしているのに対し, 40° における音響伝達特性は比較的識別しにくい分布をしていることが分かる.

3.2.2 頭部回転方向の推定における実験結果

次に, それぞれの場所にスピーカーの位置を固定し, スピーカーの回転方向のみを変えることで, 頭部回転方向

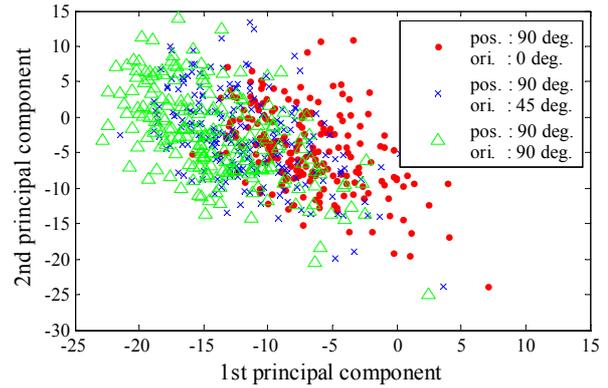


図 7 音源位置 90° におけるスピーカーの回転方向ごとの音響伝達特性

表 3 各音源位置, スピーカーの回転方向における CSP 法による推定結果の中央値

pos. \ ori.	0 deg.	45 deg.	90 deg.
40 deg.	40.9	40.9	40.9
90 deg.	90	90	90
130 deg.	131.4	131.4	100.9

の推定における精度の評価を行った. 回転方向の候補が 2 種類 (0°, 90°) における推定精度と, 3 種類 (0°, 45°, 90°) における推定精度を表 2 に示す. これらの表より, スピーカーの回転方向が 0° と 90° の場合において, 84% 以上の精度で回転方向を推定することができているが, 45° の場合, ほとんど推定できていないことが分かる. 音源位置 90° におけるスピーカーの回転方向ごとの音響伝達特性を図 7 に示す. 図より, スピーカーの回転方向による音響伝達特性の変化は, 音源位置による音響伝達特性の変化ほど顕著ではないことと, 特に回転方向 45° の音響伝達特性が判別しづらい分布になっていることが分かる.

スピーカーの各回転方向における, CSP 法によって推定された音源方向の中央値を表 3 に示す. 結果より, スピーカーの回転方向は CSP 法による音源方向の推定結果にあまり影響しないことが分かる.

3.2.3 音源位置と頭部回転方向の推定における実験結果

最後に, 音源の位置とスピーカーの回転方向の両方を変化させて, 音源位置と頭部回転方向の推定の精度を評価した. スピーカーの回転方向の候補が 2 種類の場合における推定精度と, 3 種類の場合における推定精度を表 4 に示す. 表より, これまでの実験結果と同様に, 音源の位置によって推定精度の差が存在することと, スピーカーの回転方向 45° において推定精度が低いことが分かる. しかしながら, 回転方向の候補が 2 種類の場合において約 80%, 3 種類の場合において約 60% の精度で音源位置と頭部回転方向の両方を推定できていることが分かった.

表 4 音源位置とスピーカーの回転方向の推定精度．上表は回転方向の候補数が $2(0^\circ, 90^\circ)$ における推定精度を，下表は候補数が $3(0^\circ, 45^\circ, 90^\circ)$ における推定精度を表す．

pos. \ ori.	0 deg.	90 deg.	average
40 deg.	48.6	70.8	59.7
90 deg.	87.2	93.4	90.3
130 deg.	95.0	84.5	89.8
average	77.0	82.9	79.9

pos. \ ori.	0 deg.	45 deg.	90 deg.	average
40 deg.	44.3	15.7	68.2	42.7
90 deg.	83.7	29.8	84.9	66.2
130 deg.	76.8	50.8	87.5	71.7
average	68.3	32.1	80.2	60.2

4. おわりに

本稿では，音響伝達特性が話者の位置や頭部の回転方向によって異なる特性を持つ点に着目し，各音源位置・頭部回転方向において発話された音声信号から，その音響伝達特性をクリーン音声の音素 HMM を用いて推定し，推定された音響伝達特性を SVM により学習・識別することで，音源の位置と頭部の回転方向をシングルチャンネルで推定する手法について検討を行った．実環境における実験では，音源位置の候補が 3 種類，頭部回転方向の候補が 2 種類の場合において，最大で約 95 %，頭部回転方向の候補が 3 種類の場合において，最大で約 87.5 % の精度で音源の位置と頭部の回転方向の推定が行えた．しかしながら，音源の位置によって推定精度に差があることや，頭部の回転方向が 45° の場合，ほとんど回転方向を推定できていなかった．そのため，今後は音源位置や頭部の回転方向のパリエーションを増やし，またマイクの位置も変化させることで，どのような条件であれば推定しやすくなるのかという点について検証する必要がある．また，既に提案されたマイクロホンアレーのネットワークによる頭部回転方向の推定法との比較についても今後検討していく．

文 献

- [1] D. Johnson and D. Dudgeon, "Array Signal Processing," Prentice Hall, 1996.
- [2] M. Omologo and P. Svaizer, "Acoustic event localization in noisy and reverberant environment using CSP analysis," Proc. ICASSP96, pp. 921-924, 1996.
- [3] F. Asano and H. Asoh and T. Matsui, "Sound source localization and separation in near field," IEICE Trans. on Fundamentals of Electronics, Communications and Computer Sciences, E83-A, pp. 2286-2294, 2006.
- [4] Y. Denda and T. Nishiura and Y. Yamashita, "Robust talker direction estimation based on weighted CSP analysis and maximum likelihood estimation," IEICE Trans. on Information and Systems, E89-D, pp. 1050-1057, 2000.
- [5] F. Keyrouz and Y. Naous and K. Diepold, "A

- new method for binaural 3-D localization based on HRTFs," Proc. ICASSP06, pp. V-341-V-344, 2006.
- [6] M. Takimoto and T. Nishino and K. Takeda, "Estimation of a talker and listener's positions in a car using binaural signals," The Fourth Joint Meeting ASA and ASJ, pp. 3216, 2006.
- [7] A. Brutti, M. Omologo, and P. Svaizer, "Oriented global coherence field for the estimation of the head orientation in smart rooms equipped with distributed microphone arrays," Proc. Interspeech05, pp. 2337-2340, 2005.
- [8] J. M. Sachar and H. F. Silverman, "A baseline algorithm for estimating talker orientation using acoustical data from a large-aperture microphone array," Proc. ICASSP04, vol. 4, pp. 65-68, 2004.
- [9] C. Segura, A. Abad, J. Hernando and C. Nadeu, "Speaker orientation estimation based on hybridation of GCC-PHAT and HLBR," Proc. Interspeech08, pp. 1325-1328, 2008.
- [10] M. Togami and Y. Kawaguchi, "Head orientation estimation of a speaker by utilizing kurtosis of a DOA histogram with restoration of distance effect," Proc. ICASSP10, pp. 133-136, 2010.
- [11] T. Kristjansson and H. Attias and J. Hershey, "Single microphone source separation using high resolution signal reconstruction," Proc. ICASSP04, pp. 817-820, 2004.
- [12] B. Raj and M. V. S. Shashanka and P. Smaragdis, "Latent dirichlet decomposition for single channel speaker separation," Proc. ICASSP06, pp. 821-824, 2006.
- [13] G.-J. Jang and T.-W. Lee and Y.-H. Oh, "A subspace approach to single channel signal separation using maximum likelihood weighting filters," Proc. ICASSP03, pp. 45-48, 2003.
- [14] T. Nakatani and B.-H. Juang, "Speech dereverberation based on probabilistic models of source and room acoustics," Proc. ICASSP06, pp. I-821-I-824, 2006.
- [15] R. Takashima, T. Takiguchi, Y. Ariki, "HMM-based Separation of Acoustic Transfer Function for Single-channel Sound Source Localization," ICASSP2010, pp. 2830-2833, 2010.
- [16] B.-H. Juang, "Maximum-likelihood estimation of mixture multivariate stochastic observations of Markov chains," AT&T Tech. J., Vol. 64, pp. 1235-1249, 1985.
- [17] T. Joachims, "Making large-scale SVM learning practical," in Advances in Kernel Methods - Support Vector Learning, B. Scholkopf, C. Burges and A. Smola, Ed. MIT Press, 1999.