HMMを用いた音響伝達特性の推定と音源位置推定*

◎高島遼一，滝口哲也，有木康雄 (神戸大)

1 はじめに

これまでに提案されてきた音源方向や位置の推定方法は，マイクロホンアレーにおける各観測信号の位相差を用いた手法が多く，複数のマイクロホンが必要であった [1]．単一マイクロホンで音源位置を推定することができれば，コスト削減やシステムの縮小化など様々な利点がある．

住田らはこれまでに位置毎の音響伝達特性を判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 [2]．本稿では観測信号から音響伝達特性を HMM (Hidden Markov Model) を用いて推定し，それらを判別することで単一マイクロホンによる音源位置推定を行う手法を提案する．

2 音源位置の推定

2.1 HMMによる音響伝達特性の推定

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に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．そのために，まず観測された信号から音響伝達特性を推定する必要がある．ある場所で発声されたクリーン音声 s は，音響伝達特性 h の影響を受ける．このとき，観測信号 o はフーリエ変換を適用して以下のように表現される．

$$O(\omega;n) \approx H(\omega;n)S(\omega;n) \tag{1}$$

ここで，ω は周波数，n はフレーム番号を表す．(1) 式の両辺の対数をとり，逆フーリエ変換を適用することによりケプストラムが得られる．

$$O_{cep}(d;n) \approx H_{cep}(d;n) + S_{cep}(d;n) \tag{2}$$

ここで，d はケプストラムの次元を表す．ケプストラムは音声認識の分野で広く用いられていることから，音響伝達特性の特徴量として使用する．(2) 式より，O と S を観測することができれば H を推定することができる．しかし，実際には S を観測することはできないので，S の代わりにあらかじめクリーン音声のモデルを作成しておき，これを用いて最尤推定法により O から H を推定する．

提案手法における音響伝達特性の推定の流れを Fig. 1 に示す．あらかじめ特定話者のクリーン音声を音素 HMM でモデル化しておき，それを用いて観測信号を

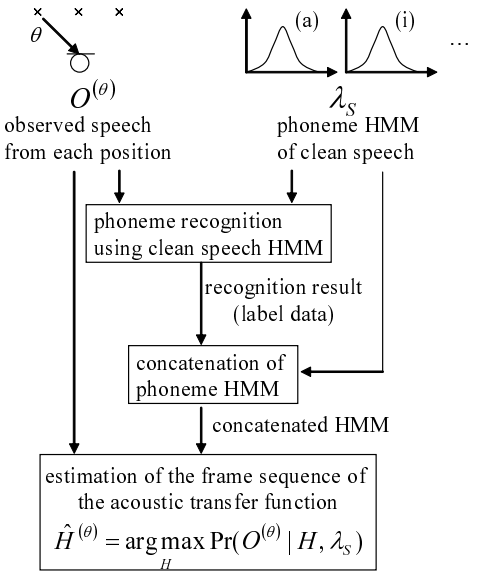

Fig. 1 音素 HMM を用いた音響伝達特性の推定

音素認識する．そして音素認識の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音響伝達特性を推定する．

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

ここで，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 アルゴリズムにより推定される．そのとき，Q 関数は次式のように導出される [3]．

$$\begin{aligned} Q(\hat{H}|H) \;=\; & -\textstyle\sum_p \sum_j \sum_k \sum_n \gamma_{p,j,k}(n) \\ & -\textstyle\sum_{d=1}^{D} \left\{ \frac{1}{2}\log(2\pi)^D \sigma^{(S)^2}_{p,j,k,d} \right. \\ & \left. + \frac{(O(d;n)-\mu^{(S)}_{p,j,k,d}-\hat{H}(d;n))^2}{2\sigma^{(S)^2}_{p,j,k,d}} \right\} \end{aligned} \tag{4}$$

$$\gamma_{p,j,k}(n) = \Pr(O(n), p, j, k|\lambda_S) \tag{5}$$

ここで，$\mu_{p,j,k,d}$ と $\sigma^{(S)^2}_{p,j,k,d}$ はそれぞれ音素 p，状態 j，混合要素 k における平均ベクトルと共分散行列の対角成分の d 次元目の値を表す．この Q 関数を最大にする $\hat{H}$ は，$\hat{H}$ について偏微分して解くことにより求めることができる．

$$\hat{H}(d;n) = \frac{\sum_p \sum_j \sum_k \gamma_{p,j,k}(n) \frac{O(d;n)-\mu^{(S)}_{p,j,k,d}}{\sigma^{(S)^2}_{p,j,k,d}}}{\sum_p \sum_j \sum_k \frac{\gamma_{p,j,k}(n)}{\sigma^{(S)^2}_{p,j,k,d}}}. \tag{6}$$

*HMM separation of acoustic transfer function for single-channel sound source localization

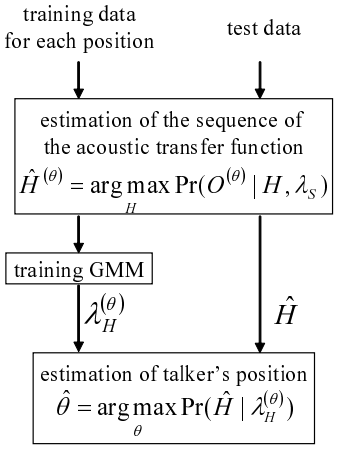

Fig. 2　音源位置の推定

2.2　GMM による音源位置の判別

音源位置 θ 毎に観測された学習用の発話データを用いて音響伝達特性を (6) により推定し，それらを GMM(Gaussian mixture model) でモデル化しておく．そして未知の位置で発話されたテストデータに対しても同様に音響伝達特性を推定し，学習した GMM との尤度を比較することで位置の推定を行う．

$$\hat{\theta} = \underset{\theta}{\operatorname{argmax}} \Pr(\hat{H} | \lambda_H^{(\theta)}) \tag{7}$$

ここで，$\lambda_H^{(\theta)}$ は位置 θ に対応する音響伝達特性 GMM を表す．Fig. 2 に音源位置推定の概要を示す．

2.3　評価実験

提案手法を評価するためにシミュレーション実験を行った．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た．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．クリーン音声のモデルは 2620 単語を用いて，54 種類の音素 HMM を学習しており，各音素 HMM の状態数は 5，混合数は 2，4，8 の場合で実験を行っている．推定された音響伝達特性の学習には 50 単語を用いて，混合数が 1，2，4，8，16 の GMM でモデル化し，1000 単語を用いて評価を行った．なお，クリーン音声の学習データ，音響伝達特性の学習データ，評価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．音響伝達特性の学習データと評価データは，RWCP 実環境音声・音響データベースより音源とマイクロホンの距離が 2 m，残響時間が 300 msec のインパルス応答をクリーン音声に畳み込むことで作成した．音源位置は 30°，90°，130° の 3 種類の場合と，10°，50°，90°，130°，170° の 5 種類の場合で実験を行った．

実験結果を Fig. 3，Fig. 4 に示す．全ての場合において，クリーン音声 HMM の混合数が 8 の場合が最も高い精度が得られており，位置が 3 種類の場合では最高 84.5% の精度で音源位置の推定が行えていた．しかしながら位置が 5 種類に増えると精度は大幅に下がり，最高でも 60 % 程度の推定精度となった．

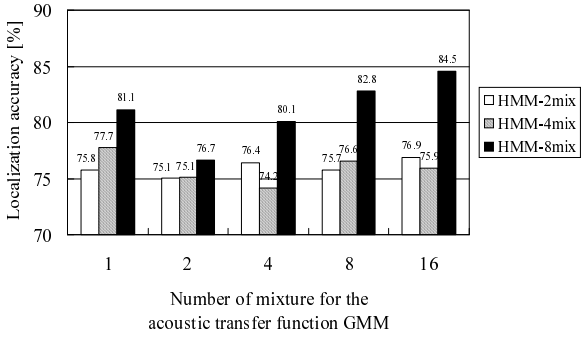

Fig. 3　音源位置が 3 種類の場合の位置推定精度

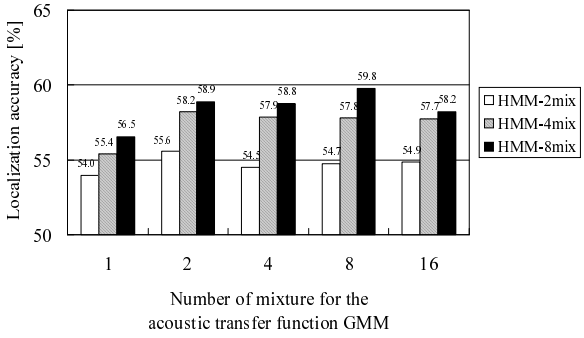

Fig. 4　音源位置が 5 種類の場合の位置推定精度

3　おわりに

本稿では，単一マイクロホンのみによる音源位置推定の方法として，HMM で分離された音響伝達特性を用いた尤度比較による推定方法を提案した．今後は HMM の学習条件を変えての実験や，クリーン音声を HMM でなく GMM で学習した場合などとの比較を行っていく予定である．

参考文献

[1] D. Johnson and D. Dudgeon, “Array Signal Processing,” Prentice Hall, 1996.

[2] T. Takiguchi, Y. Sumida, R. Takashima, Y. Ariki, “Single-Channel Talker Localization Based on Discrimination of Acoustic Transfer Functions,” EURASIP Journal on Advances in Signal Processing Vol. 2009, 9 pages, 2009.

[3] T. Takiguchi, M. Nishimura, “Improved HMM Separation for Distant-talking Speech Recognition,” IEICE TRANS. INF. & SYST., VOL.E87-D, NO.5 MAY 2004.