音響モデルを利用したシングルチャネルによる音源方向推定の検討*

◎住田雄司, 滝口哲也, 有木康雄 (神戸大)

1　はじめに

音源方向を検出するために，これまで様々な方法が提案されてきた．しかし，その多くはマイクロホンアレーでの受信による信号の時間差を用いており，複数のマイクロホンという条件が必要不可欠であった [1].

単一マイクロホンで方向を推定することができれば，複数のマイクロホンの場合と比較して，コストを削減することができ，またマイクロホンの設置が容易であるといった様々な利点が存在する．

そこで本稿では，音響モデルを利用することにより，単一マイクロホンで音源方向を推定する方法を提案する．あらかじめクリーンな音声の音響モデルを作成しておき，各方向から到来する数単語の音声から EM アルゴリズムを用いることにより，音響伝達特性を推定する．これより得られた音響伝達特性の時系列データから，各方向における音響伝達特性モデルを作成する．そして実際の入力音声から同様にして音響伝達特性を推定し，これらのモデルとの尤度を求めることで方向の決定を行う．

2　音響伝達特性の推定

Fig. 1 に示すように，ある場所で発声されたクリーンな音声 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，

$$O(\omega;t) = H(\omega)S(\omega;t) \tag{1}$$

と表される．(1) 式の両辺に対数をとると，次式のように加算の形で表すことができる．

$$\log O(\omega;t) = \log H(\omega) + \log S(\omega;t) \tag{2}$$

(2) 式より，O と S が分かれば H を推定することができる．しかし S を観測することはできないので，S の代わりにクリーン音声モデルを用い，尤度最大基準に基づいて O から H を分離する．

まず，観測信号に対して，そのモデルの尤度が最大となるようにして音響伝達特性を求める．

$$\hat{\lambda}_{H_{cep}} = \underset{\lambda_{H_{cep}}}{\operatorname{argmax}} P(S+H|\lambda_{S_{cep}}, \lambda_{H_{cep}}) \tag{3}$$

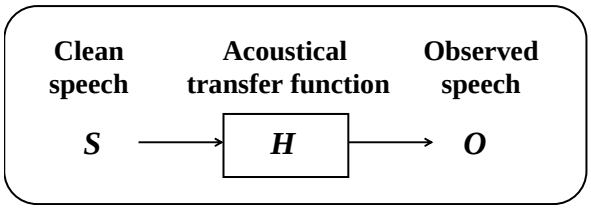

Fig. 1　対象とする環境のモデル

ここで，λ はモデルパラメータの集合を表す．Q 関数は，以下のように定義する [2].

$$Q(\mathbf{H},\mathbf{H}') = \sum_{t=1}^{T}\sum_{n=1}^{N}\sum_{m=1}^{M} \gamma_t(n,m)\left\{-\sum_{i=1}^{D}\frac{(O_{t,i}-H'_{t,i}-\mu_{n,m,i})^2}{2\sigma^2_{n,m,i}}\right\} \tag{4}$$

D は特徴量の次元数，T はフレーム数，N，M はそれぞれクリーン音声モデルの状態数と混合数を表し，平均を μ，分散を σ とおく．また，$\gamma_t(n,m)$ は，以下の式で表される．α は分布の重みである．

$$\gamma_t(n,m) = \frac{\alpha_{n,m}N[\mathbf{S_t};\mu_{n,m},\Sigma_{n,m}]}{\sum_{m=1}^{M}\alpha_{n,m}N[\mathbf{S_t};\mu_{n,m},\Sigma_{n,m}]} \tag{5}$$

これを $H'_{t,i}$ について偏微分して解くことにより，最終的に音響伝達特性の時系列パラメータとして $H'_{t,i}$ が得られる．

$$H'_{t,i} = \frac{\sum_{n=1}^{N}\sum_{m=1}^{M}\gamma_t(n,m)\dfrac{O_{t,i}-\mu_{n,m,i}}{\sigma^2_{n,m,i}}}{\sum_{n=1}^{N}\sum_{m=1}^{M}\dfrac{\gamma_t(n,m)}{\sigma^2_{n,m,i}}} \tag{6}$$

3　評価実験

提案手法を評価するために実験を行った．実験条件を Table 1 に示す．

音声が到来する方向は，30 °，90 °，130 °のいずれかであるとする．実験環境は，音源とマイクロホンの距離が 2 m，残響時間が 300 msec の環境で収録されたインパルス応答を畳み込むことにより，残響

*Studies on single-channel DOA estimation using acoustic model by Yuji Sumida, Tetsuya Takiguchi and Yasuo Ariki (Kobe Univ.)

環境をシミュレーションした．このインパルス応答はRWCP 実環境音声・音響データベースを用いた．

音声データとして，ATR 音声データベースより男性話者一名，クリーン音声の学習データには 2620 単語，音響伝達特性を推定するためのデータには 10 単語，評価用のデータには 1000 単語を用いた．

音響モデルとして，クリーン音声モデルは GMM（64 混合）で作成し，音響伝達特性モデルは GMM の混合数を変化させて比較を行った．

実験結果について述べる．30 °，90 °，130 °の方向から音声が到来したときの方向推定結果をそれぞれ Fig. 2，Fig. 3，Fig. 4 に示す．30 °，130 °の場合では混合数にかかわらず，それぞれ 90%，95% 程度の正解率が出ており，正確に方向が推定されているといえる．しかし，90 °の場合では，単一の正規分布において，80% が 130 °と誤識別されており，混合数を増やしても正確率は 60% 程度に留まる結果となった．

原因として，各方向における音響伝達特性の類似性，EM アルゴリズムにおける近似誤差や残響による過去の音声の重なりの影響があげられる．これらの誤差から値がばらつき，分散が増大することにより誤識別が起きたものと考えられる．

4　まとめ

音源方向の検出法として，音響モデルを利用した単一マイクロホンによる方向推定の方法について述べた．評価実験より，提案手法によって単一マイクロホンでも方向推定ができることを示した．

今後の課題としては，方向の間隔を狭くし，かつ数を増やすことにより全ての方位に対応していくことと，方向推定に他の手法を用いて，今回の手法との比較ならびに精度の向上を目指してことがあげられる．また，今回の実験は特定話者で行ったが，これを不特定話者に拡張していく必要がある．

謝辞　本研究の一部は，公益信託 小野音響学研究助成基金の助成により行われた．

参考文献

[1] C.H.Knapp and G.C.Carter, “The Generalized Correlation Method for Estimation of Time Delay, ” IEEE Trans. On Acoust., Speech and Signal Proc., ASSP-24, 4, pp.320-327, 1976.

[2] A.Sanker and C-H.Lee, “A maximum-likelihood approach to stochastic matching for robust speech recognition, ” IEEE Trans. Speech and Audio Processing, vol.4, no.3, pp.190-202, 1996.

Table 1　実験条件

サンプリング周波数	12 kHz
窓関数	Hamming
Frame length	32 msec
Frame shift	8 msec
特徴量	MFCC(order 16)

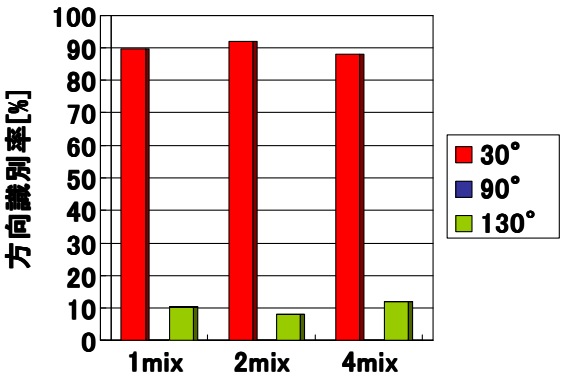

Fig. 2　30 °入力における方向識別率

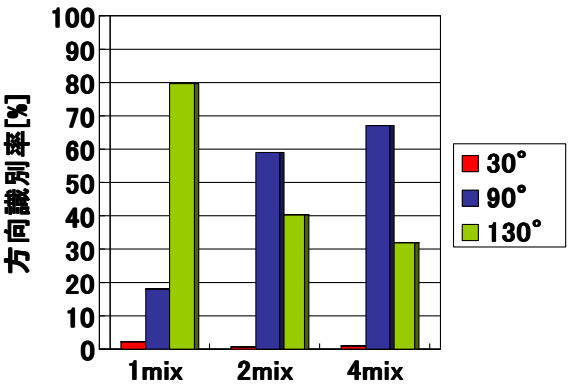

Fig. 3　90 °入力における方向識別率

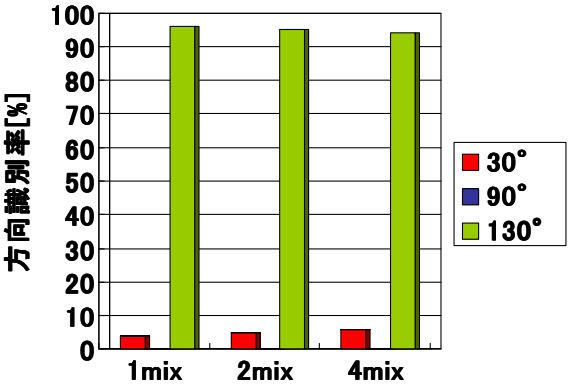

Fig. 4　130 °入力における方向識別率