

単一マイクロホンを用いた音響伝達特性の尤度判定による音源位置推定*

◎住田雄司, 滝口哲也, 有木康雄 (神戸大)

1 はじめに

音源の方向や位置を推定するためにこれまで様々な手法が提案されている。しかし、その多くはマイクロホンアレーを用いた受音により生じる音源の時間差 [1] や音圧差 [2] の情報を利用しており、複数のマイクロホンという条件が前提となっていた。

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。筆者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いる [3, 4]。本稿では、音響伝達特性を用いた提案手法と観測信号を用いた手法を比較し、提案手法の有効性を示す。

2 音源位置の推定

2.1 音響伝達特性の推定

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。

$$O(\omega; n) \approx H(\omega; n)S(\omega; n) \quad (1)$$

ここで、 ω は周波数、 n はフレーム番号を表す。(1) 式の両辺の対数を取り、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。

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

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、これを用いて尤度最大基準に基づき O から H を抽出する。

(2) 式の H の時系列データを、 O に対して、そのモデルの尤度が最大となるようにして求める。

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

ここで、 λ はモデルパラメータの集合を表し、添え字の S はケプストラム領域におけるクリーン音声を表

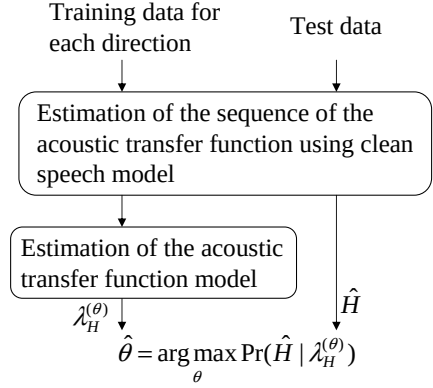


Fig. 1 音響伝達特性の尤度判定による音源位置推定

す。 Q 関数は、次式のように導出される [5]。

$$Q(\hat{H}|H) = \sum_m \sum_n \gamma_{m,n} \left\{ - \sum_d \frac{(O_{n,d} - \hat{H}_{n,d} - \mu_{m,d})^2}{2\sigma_{m,d}^2} \right\} \quad (4)$$

$$\gamma_{m,n} = \Pr(O_n, m|H, \lambda_S) \quad (5)$$

ここで、 m はクリーン音声モデルの混合系列を表し、平均を μ 、分散を σ とおく。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。

$$\hat{H}_{n,d} = \frac{\sum_m \gamma_{m,n} \frac{O_{n,d} - \mu_{m,d}}{\sigma_{m,d}^2}}{\sum_m \frac{\gamma_{m,n}}{\sigma_{m,d}^2}} \quad (6)$$

2.2 モデルによる位置の判別

評価発話毎に、(6) 式より $\hat{H}$ を推定し、尤度最大基準に基づき位置の推定を行う。

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

ここで、 $\lambda_H^{(\theta)}$ は位置 θ に対応する音響伝達特性モデルを表す。Fig. 1 に音源位置推定の概要を示す。各位置の発話から音響伝達特性の時系列データを (6) 式より計算し、音響伝達特性モデルを作成する。評価発話毎に推定された $\hat{H}$ に対して各モデルの尤度を計算し、最大尤度を与えるモデルを到来音源の位置として出力する。

*Single-channel sound source localization based on discrimination of acoustic transfer functions.
by Yuji Sumida, Tetsuya Takiguchi and Yasuo Ariki (Kobe University)

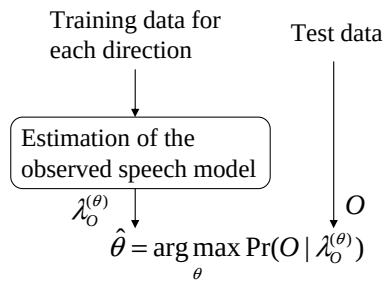


Fig. 2 観測信号の尤度判定による音源位置推定

Table 1 分析パラメータ

サンプリング周波数	12 kHz
窓関数	Hamming
フレーム長	32 msec
フレームシフト	8 msec
特徴量	MFCC (order 16)

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。分析パラメータを Table 1 に示す。比較のために、音響伝達特性ではなく観測信号の特徴量を用いて音源位置を推定する実験も行った。概要を Fig. 2 に示す。以降、本稿ではこれを従来手法と呼ぶ。

実験環境について述べる。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た。クリーン音声のモデルは 64 混合の GMM を用い、2620 単語を使用して学習を行った。実環境をシミュレーションするために、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み、評価データと音響伝達特性・観測信号の学習データを作成した。評価データには 1000 単語を使用し、30°、90°、130° の位置より到来するものとした。音響伝達特性・観測信号の学習データには 10 単語もしくは 50 単語を使用し、上記 3 方向のモデルを 1 混合もしくは 4 混合の GMM で作成した。

実験結果を Fig. 3 に示す。学習データ数やモデルの混合数にかかわらず、提案手法の方が従来手法より良い結果となった。最も正解率が良かったのは 50 単語で 4 混合の音響伝達特性 GMM を作成して判別した場合で 83.0%，最も差が出たのは 10 単語で 4 混合の GMM を作成して判別した場合で、その差は 11.7% であった。

50 単語で学習した 30°位置モデルの分散値を Fig. 4 に示す。観測信号には音響伝達特性に加えてクリーン音声の成分が含まれているため、観測信号の分散値は音響伝達特性より大きくなっている。従って観測信号 GMM の山は音響伝達特性 GMM よりなだら

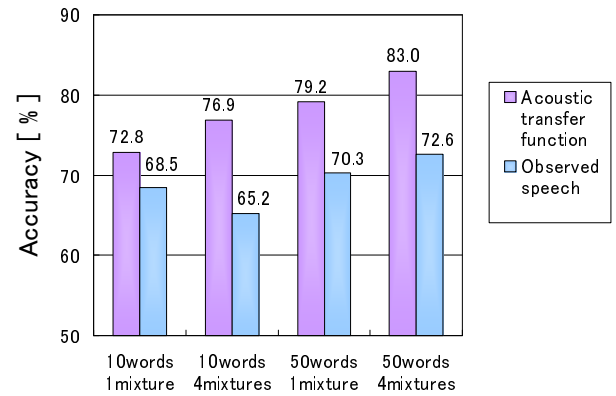


Fig. 3 音響伝達特性と観測信号における尤度判定の正解率

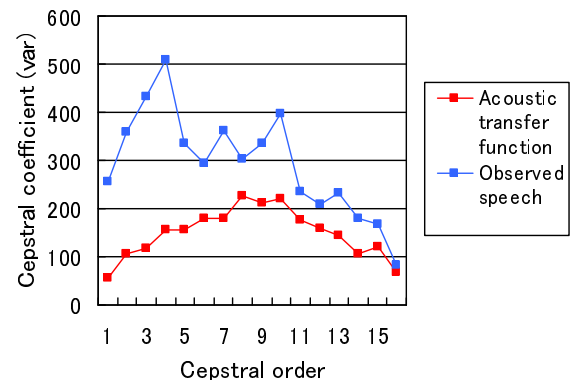


Fig. 4 音響伝達特性 MFCC と観測信号 MFCC の分散値

となり、各モデル間の尤度差の減少から誤識別が起きているものと考えられる。

4 まとめ

本稿では、単一マイクロホンのみを用いた音源位置推定の方法として、音響伝達特性の尤度判定による方法を提案した。実験結果から、提案手法は従来手法と比較して 10% ほど正解率に改善がみられた。今後の課題として、話者が不特定になった場合や雑音が混入した場合、未知の方向から音源が到来する場合や学習と評価で部屋が違う場合などに対応していくことが挙げられる。

参考文献

- [1] C. H. Knapp and G. C. Carter, IEEE Trans. On Acoust., Speech and Signal Proc., ASSP-24(4), 320-327, 1976.
- [2] 西野他, 音学誌, 3-12, 2007.
- [3] 住田他, 音講論 (秋), 749-750, 2007.
- [4] Tetsuya Takiguchi, Yuji Sumida and Yasuo Ariki, IEEE SSP Workshop, 336-340, 2007.
- [5] A. Sankar and C-H. Lee, IEEE Trans. On Speech and Audio Proc., 4(3), 190-202, 1996.