

単一マイクロホンを用いた音響伝達特性の判別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

音源方向を推定するためにこれまで様々な手法が提案されてきたが [1], マイクロホンアレーを用いた信号の時間差や音圧差による推定方法が多く, 複数のマイクロホンという条件が必要であった.

単一マイクロホンで方向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といった様々な利点があることから, 筆者らはこれまでに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している [2, 3]. 本稿では, これまでの手法を用いた追加実験により得られた結果と知見について述べる.

2 音源方向の推定

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.

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

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事によりケプストラムが得られる.

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

ここで, d はケプストラムの次元を表す. 本稿では, このケプストラム領域にてクリーン音声モデルを作成する.

(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 を抽出する.

(2) 式の音響伝達特性の時系列データを, 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す.

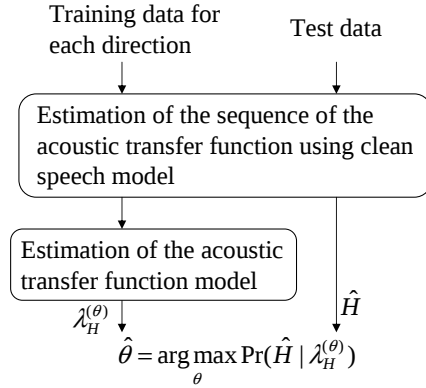


Fig. 1 音源方向推定の概要

Q 関数は, 次式のように導出される [4].

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

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

ここで, j, k はそれぞれクリーン音声モデルの状態系列と混合要素系列を表し, 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事が出来る.

$$\hat{H}_{n,d} = \frac{\sum_j \sum_k \gamma_{j,k,n} \frac{O_{n,d} - \mu_{j,k,d}}{\sigma_{j,k,d}^2}}{\sum_j \sum_k \frac{\gamma_{j,k,n}}{\sigma_{j,k,d}^2}} \quad (6)$$

2.2 モデルによる方向の判別

評価発話毎に, (6) 式により $\hat{H}$ を推定し, 尤度最大基準に基づき方向推定を行う.

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

ここで, $\lambda_H^{(\theta)}$ は, θ 方向に対応する音響伝達特性モデルである.

Fig. 1 に音源方向推定の概要を示す. 各方向に対応する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構築する. 評価発話毎に推定された $\hat{H}$ に対して尤度を計算し, 最大尤度を与えるモデルを正解方向とする.

* A study on single-channel DOA estimation by discrimination of acoustic transfer function. by Yuji Sumida, Tetsuya Takiguchi and Yasuo Ariki (Kobe University)

Table 1 分析パラメータ

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)

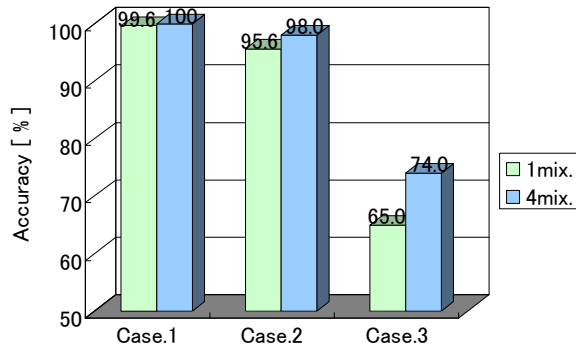


Fig. 2 Case.1 ~ 3 における正解率

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．

Case.1 ~ 3 における実験環境について述べる．評価発話データの到来方向は， 30° ， 90° ， 130° である．

Case.1 では，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 [5] をドライソースに畳み込んでシミュレーションを行った．音声データとして，話者は男性 1 名，クリーン音声の学習データに 40 文，音響伝達特性の学習データに 10 文，評価用のデータに 100 文用いた．クリーン音声モデルには 64 混合の GMM を用い，音響伝達特性のモデルは上記 3 方向の GMM を作成し，1 混合，4 混合で比較を行った．

Case.1 と比較して，Case.2 ではモデルの方向数を 3 方向から 9 方向（ 10° から 20° 間隔で 170° まで）に増やし，Case.3 では残響時間が 0 msec のインパルス応答を用いた．

実験結果を Fig. 2 に示す．Case.1，2 においては正確に方向を推定できており，モデルの方向数が増加しても大きな精度の低下はみられなかった．これらの Case では，複数のマイクロホンで方向推定を行った場合と同程度の精度が出ているといえるが，残響がない Case.3 では精度が低下した．

Case.1，3 における音響伝達特性の各次元のケプストラム平均値を Fig. 3，4 に示す．残響がある場合は方向によってケプストラムの値が異なり，方向毎に特徴がみられる．一方，残響がない場合は値に差がなく，これにより誤識別が起きていると考えられる．

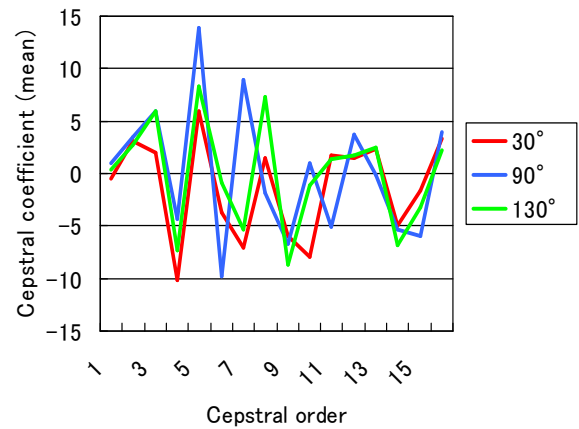


Fig. 3 Case.1 (残響時間: 300 msec) のケプストラム平均値

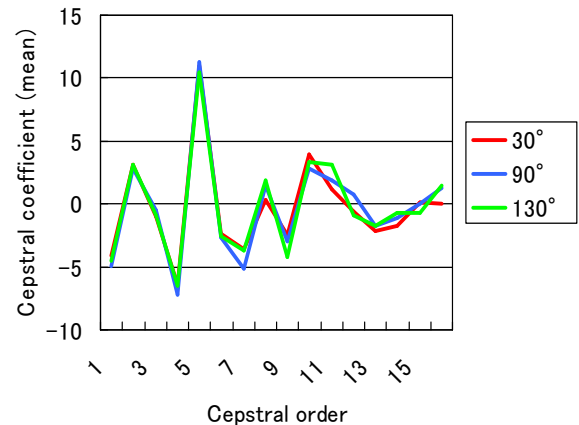


Fig. 4 Case.3 (残響時間: 0 msec) のケプストラム平均値

4 まとめ

単一マイクロホンを用いた音源方向推定の方法として，音響伝達特性の判別による方法を提案した．評価実験より，提案手法は残響が存在する場合に特に有効であることを示した．

今後の課題として，不特定話者への対応や雑音の想定など，実環境での適用を考えて精度を向上させていく必要がある．具体案として，指向性マイクロホンの導入，LDA による特徴成分強調などを検討している．

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] 住田他，音講論（秋），539-540，2006．
- [3] 住田他，信学研報，EA2006-90．
- [4] A. Sankar and C-H. Lee, IEEE Trans. Speech and Audio Processing, 4(3), 190-202, 1996.
- [5] <http://tosa.mri.co.jp/sounddb/>