室内残響を考慮した大規模マイクロホンアレイによる発話方向の推定*

◎醍醐徹[1]，菊池慶子[1]，中島弘史[2]，中臺一博[2]，長谷川雄二[2]，金田豊[1]
（1 東京電機大・工，2 (株)ホンダ・リサーチ・インスティチュート・ジャパン）

1　はじめに

本稿では，室内に設置されたマイクロホンアレイを用いて人の発話方向を推定する方法について検討する。

マンマシーンインタフェースを考えるとき，発話者位置を特定するとともに，発話者の向いている方向を知ることも重要である。例えば，ロボットとの対話を考えたとき，ロボットは「誰が（どこから）」話しかけたのか，という情報とともに，発話者が，「誰（どこ）に向かって」話しかけたのかという発話方向情報は，その場の情景理解において大変重要である。

発話方向検出には近年進歩の著しい顔画像認識技術を利用することも有効かと考えられる。しかし，検出できる角度の精度に課題が予想されることと，画像だけでなく，音と画像による複数の検出結果を統合することで，人間同様に，より高機能，高精度の状況把握が期待される。

これまで，マイクロホンアレイを用いて音源位置を推定する試みは多数なされてきた[1, 2]。しかし，発話方向検出の例は数少ない[3, 4]。従来の方法では，周波数平坦化ビームフォーマーの考え方で行ったため，低ゲイン方向での悪影響が発生していた。また，処理の元となる伝達関数を自由音場に仮定して行っていたため，残響の多い室内では実際の伝達関数とのモデル化誤差によって，十分良好な結果は得られなかった。

そこで，本報告ではこの問題をパターンマッチングの問題と捉えるとともに，室内で実測した伝達関数を利用することで，残響のある室内においても適用可能な発話方向推定法を検討した。

2　発話方向推定原理

2.1 推定原理

Fig. 1 に発話とマイクロホンアレイによる受音のモデルを示す。図においてSは発話者，$M_1 \sim M_N$はN個のマイクロホン，$H_1 \sim H_N$はそれらの間の空間伝達関数を表す。話者は，室内の平面座標(x, y)の位置において，θ方向を向いて音声$S(\omega)$を発話している。伝達関数 $H_i(\omega, \mathbf{p})$ $(i = 1, \cdots, N)$ $(\mathbf{p} = (x, y, \theta))$ は発話方向θにも依存し，室内反射音も含んでいる。

Fig. 2 は，提案する処理系の構成を示す。最初に，マイクロホンで受音した信号にビームフォーマーなどを適用して，話者位置(x, y)を検出する。この検出は，良好に行えることがすでに報告されている[5]。そこで，本検討では，話者位置が既知・固定であるとして，話者方向θのみの検出を行うことを考える。

ここで，伝達関数ベクトル$\mathbf{h}(\theta)$ (ωは簡単化のため省略)を，次式のように定義する。

$$\mathbf{h}(\theta) = [H_1(\omega, \theta), \cdots, H_N(\omega, \theta)]^T \qquad (1)$$

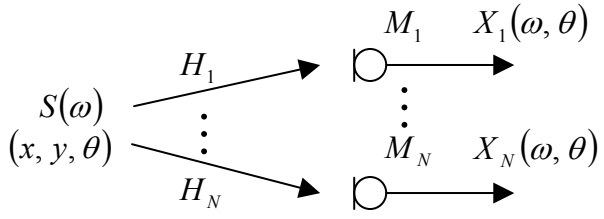

Fig. 1 発話と受音のモデル

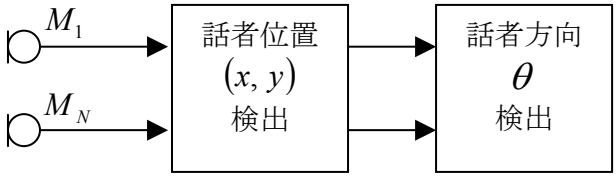

Fig. 2 処理の流れ

*Estimation of Talker-Facing-Direction Using Large Scale Microphone Array in Consideration of Room Reverberation, by Tôru DAIGO[1], Keiko KIKUCHI[1], Hirofumi NAKAJIMA[2], Kazuhiro NAKADAI[2], Yuji HASEGAWA[2] and Yutaka KANEDA[1] (1 Tokyo Denki University, 2 Honda Research Institute Japan Co., Ltd.)

この$\mathbf{h}(\theta)$は，既知であると仮定する。次に，話者方向が$\hat{\theta}$であったときに受音された信号を要素とするベクトル$\mathbf{x}$を次式で定義する。

$$\mathbf{x} = \left[X_1(\omega, \hat{\theta}), \cdots, X_N(\omega, \hat{\theta})\right]^T \quad (2)$$

この時，

$$X_1(\omega, \hat{\theta}) = H_1(\omega, \hat{\theta})S(\omega) \quad (3)$$

であるので，

$$\mathbf{x} = \left[H_1(\omega, \hat{\theta})S(\omega), \cdots, H_N(\omega, \hat{\theta})S(\omega)\right]^T = S(\omega)\mathbf{h}(\hat{\theta}) \quad (4)$$

と表すことができる。（注：音声$S(\omega)$は未知なので，受音信号 $\mathbf{x}$から直接的に$\mathbf{h}(\hat{\theta})$を求めることはできない）

ここで，ベクトル$\mathbf{h}(\theta)$の長さを正規化して単位ベクトル

$$\mathbf{h}_0(\theta) = \mathbf{h}(\theta) / \|\mathbf{h}(\theta)\| \quad (5)$$

とした時， $\mathbf{x}$との内積

$$\mathbf{h}_0(\theta)^* \mathbf{x} = S(\omega)\mathbf{h}(\theta)^* \mathbf{h}(\hat{\theta}) / \|\mathbf{h}(\theta)\| \quad (6)$$

（*は共役転置）は，ベクトル$\mathbf{h}(\hat{\theta})$とベクトル$\mathbf{h}(\theta)$との方向類似度に対応する。従って，θを変化させながら式(6)の内積を計算し，最大値をとるとき 2 つのベクトルは一致する。すなわち，式(6)の内積の最大値を与えるθが話者方向$\hat{\theta}$となる。この操作は周波数ごとに行われるので，適宜周波数平均を行い，最終判断結果とする。

本研究では，室内の壁面に話者を取り囲んでマイクロホンが設置されており，伝達関数ベクトル$\mathbf{h}(\theta)$の各要素$H_1 \sim H_N$の大きさは話者方向に応じた指向性に対応している。よって上記の内積は，室内残響を含めた指向性パターンのマッチングを行っていることに相当する。

2.2 伝達関数ベクトル$\mathbf{h}(\theta)$の推定

伝達関数ベクトル$\mathbf{h}(\theta)$を求める方法としては，スピーカなどを用いて事前に測定しておく方法が考えられる。ただし，この方法では人間の指向性ではなく，スピーカの指向性が反映され，誤推定が生じる可能性がある。

第 2 の方法としては，人間を用いて実測した音声の利用が考えられる。人間がθ方向を向いて音声$S_0(\omega)$を発生した場合の受音信号ベクトル $\mathbf{x}_0(\theta)$は，式(4)と同様に

$$\mathbf{x}_0(\theta) = S_0(\omega)\mathbf{h}(\theta) \quad (7)$$

となって，伝達関数ベクトルが定数倍されたものになっている。ただし，音声$S_0(\omega)$は，短時間では全ての周波数成分を持たないので，リファレンスマイクロホンの出力$X_{10}(\omega, \theta)$を乗じて次式のような平均化操作を行なう。

$$\mathbf{h}_x(\theta) = \overline{X_{10}(\omega, \theta)\mathbf{x}_0(\theta)} = \overline{S_0{}^2(\omega)} \cdot H_1(\omega, \hat{\theta}) \cdot \mathbf{h}(\theta) \quad (8)$$

このようにして得られたベクトル$\mathbf{h}_x(\theta)$は，周波数成分の欠落が無く，$\mathbf{h}(\theta)$の定数倍になっている。これより，θのみを変数とする場合には，この$\mathbf{h}_x(\theta)$を伝達関数ベクトル$\mathbf{h}(\theta)$の代わりに用いて式(5)(6)の内積計算を行っても，同じ話者方向推定結果が得られる。

ただし音声は，周波数によってはエネルギーが小さく SN 比が悪い場合があり，全周波数測定の容易なスピーカ利用の伝達関数で精度の良い推定ができれば，それが望ましい。

そこで，以下本稿では，

1）音声から得た式(8)の伝達関数を用いて，提案した話者方向推定法の性能を評価

2）スピーカで測定した伝達関数を用いて，話者方向推定の可能性を評価

の２点について検討した結果を報告する。

3 実験結果

3.1 音声から得た伝達関数を用いた推定

実験は，広さが 7m×4m 高さが 3.5m，残響時間が 約 100ms の吸音性の実験室で行なった。室内の壁面に 96ch のマイクロホンを Fig. 3 の○印に示すように配置させた。女性および男性が x=3m，y=2m (部屋中央)の位置で，図の 0° 方向から反時計回りで 15°刻みに 345°までに向いて音声を発声し，録音した。音声は「あ」の音韻の高さを変えながら発声した。録音したデータから，式(8)に従って伝達関数ベクトル$\mathbf{h}_x(\theta)$を計算した（リファレンスマイクロホンの位置は Fig. 3 に示す）。

評価用音声として，女性および男性が，複数の方向を向いて「あ，い，う，え，お」と発話したものを用いて方向推定を行なった。

方向推定結果を Fig. 4～Fig. 6 に示す。図は，横軸が時間，縦軸が推定方向を表している。伝達関数ベクトル$\mathbf{h}_x(\theta)$は評価用音声と同一人物のものを使用した。式(6)の内積の対数値を周波数で平均した結果[dB]をカラー表現した。濃青で表されている値の小さな領域は無音声区間である。色が赤いほど内積値が大きく発話方向の可能性が大きいことを示している。各時刻における推定方向（内積の最大値）を黒線で示した。

Fig.4，Fig.5 はほぼ安定な推定結果が得ら

れたことを示している。FIg.6 では，第 5 発声音「お」において，少し推定のばらつきが見られる。今回は発声者自身の声から求めた伝達関数を使用して推定を行なったが，許容誤差を±15° とすると，ほぼ 80%の正答率を得た。誤推定の分析・改善は今後の課題であるが，発話者本人の音声を用いて得た伝達関数を利用した場合，提案法はほぼ良好に動作することがわかった。

3.2 スピーカの伝達関数を用いた場合

スピーカの位置を x=3m，y=2m (部屋中央) 高さを床から約 1.5m とし，角度を 0°~355°まで 5° 刻み回転させて，TSP 信号を用いて，各マイクロホンまでの伝達関数 $H_i(\omega, \theta)$ の測定を行った。そして，式(6)に従い前項と同様に，発話方向推定を行った。

結果を Fig. 7～Fig. 8 に示す。Fig. 7 の 0° 方向を推定した結果はおおむね推定できている。しかし，Fig. 8 の 90° 方向を推定した結果は不良であった。

そこで，周波数帯域別に発話方向推定結果を見てみた。推定のできていない Fig. 8 の発話区間の一部分を切り出して，発話方向推定角度の統計をとった結果を Fig. 9～Fig. 11 に示す。

Fig. 9 は，各周波数ごとの推定結果を全帯域にわたって頻度分布として表したものである。最終的推定結果は，これらのほぼ平均であるので，正解の 90° ではなく，30° 方向に推定されてしまう。しかし，周波数ごとに見てみると, Fig. 11 の 4000-8000Hz の帯域では，誤推定が多いのに対して，0-1000Hz の帯域 (Fig.10)では正解が多いことがわかる。

この原因は，高周波では人間とスピーカの伝達特性の違いが，特に位相成分において大きいためと考えられる。そこで，今後の改善策として，1) 位相を用いず，振幅値のみを利用する, 2) 周波数に重みをつけて推定を行なう，などの対策が有効と考えられた。

4 まとめ

本稿では，室内壁面に設置されたマイクロホンアレイを使用して，人の発話方向推定を検討した。室内残響を考慮するために，該当室内で発声したときの伝達関数を事前測定し，観測音声とマッチングをとることで，発話方向を推定する方法を提案した。実音場で実験の結果，発話者本人の音声で推定した伝達関数を用いた場合には，80%程度の推定正解率が得られ，提案方法の原理的有効性が確認できた。

しかし，スピーカで測定した伝達関数を用いた場合では, 良好な結果が得られなかった。この原因を検討した結果，高周波域での推定誤差が大きいことがわかった。

提案法に汎用性を持たせるためには，個人の周波数特性に依存することなく，スピーカなどで測定した結果が利用できることが望ましい。今回の推定は，高周波域も含めた全周波数帯域で均一な評価を行なったことが問題と考えられるので，今後は適切な周波数加重を考えることで性能の改善が図れるものと考える。

参考文献

[1] H.F.Silverman, W.R.Patterson, “The Huge Microphone Array,”Techical report, LEMS, Brown University, 1996.

[2] P.C.Meuse, H.F.Silverman, “Characterization of Talker Radiation Pattern Using a Microphone Array,” ICASSP-94, IEEE, vol.2, 257-260, 1994.

[3] K.Nakadai, et. al., “Sound Source Tracking with Directivity Pattern Estimation Using a 64ch Microphone Array,”IROS2005, IEEE, RSJ, 196-202, 2005.

[4] 中島弘史, 他,“拡張ビームフォーミングを用いた音源の指向特性推定,” 音講論（秋）,621-622 , 2005.

[5] 中島弘史, “音源の方向を推定可能な拡張ビームフォーミング,”音講論（秋）, 619-620, 2005.

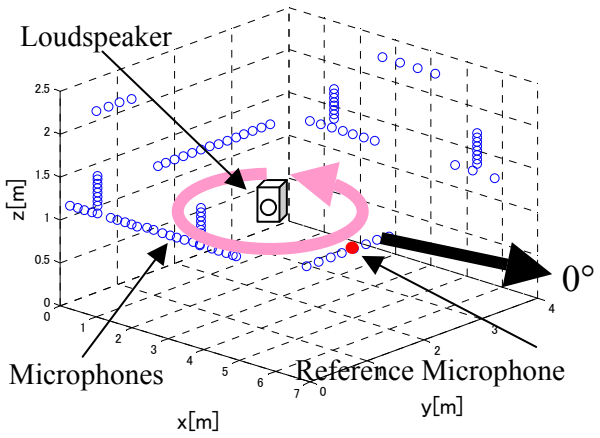

Fig. 3 実験室のマイクロホン配置

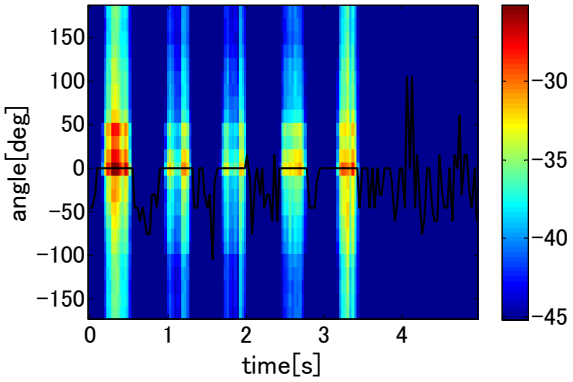

Fig. 4　発話方向推定結果(女声 0°)

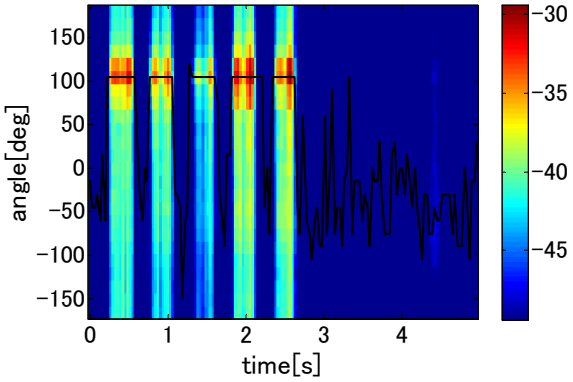

Fig. 5　発話方向推定結果(女声 90°)

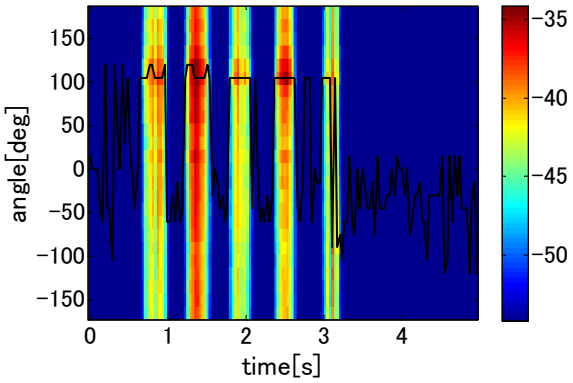

Fig. 6　発話方向推定結果 (男声 90°)

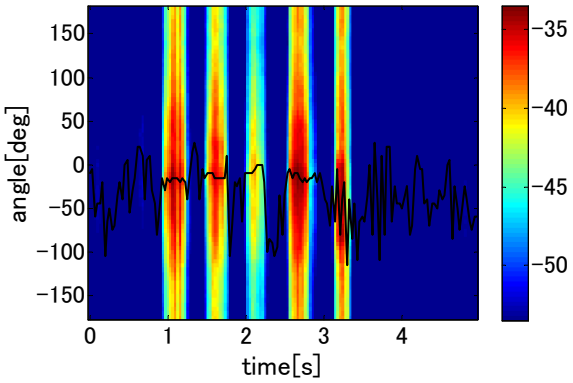

Fig. 7　スピーカの伝達特性を用いた発話方向推定結果(男声 0°)

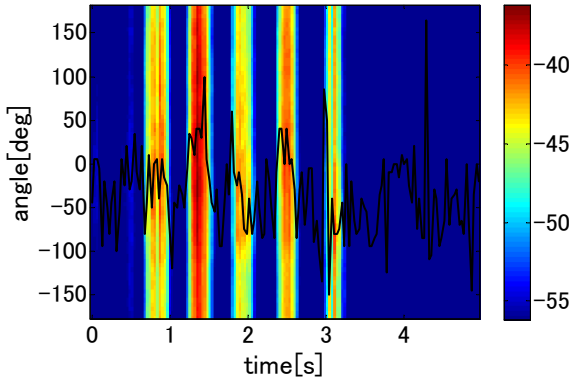

Fig. 8　スピーカの伝達特性を用いた発話方向推定結果(男声 90°)

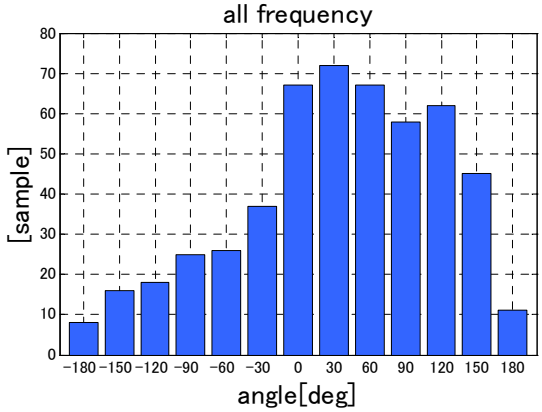

Fig. 9　発話方向推定統計(全帯域)

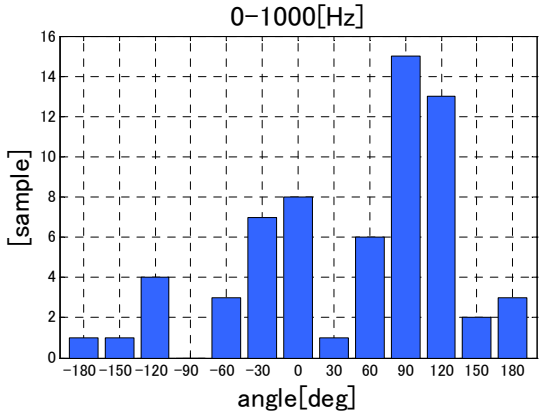

Fig. 10　発話方向推定統計(低周波域)

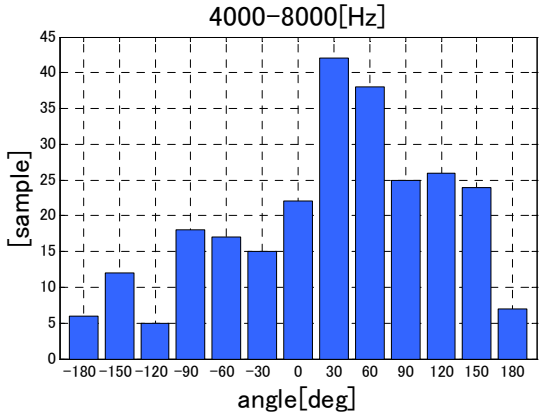

Fig. 11　発話方向推定統計(高周波域)