マイクロホンアレーを用いた発話方向推定における時間 - 周波数選択の検討*

○菊池慶子[1]，醍醐徹[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

ヒューマンマシンコミュニケーションでの音声認識において，発話者がマシンに対して話しかけたかどうかを判断する必要があり，そのためには，発話者の正対方向（発話方向）を知ることは重要である。この発話方向検出の問題に対して，醍醐らはビームフォーミングに基づいた手法を提案した[1]。しかし，この方法では騒音や室内反響などに対する対策がなされていなかったので，実環境性能が不十分であった。本報告では，この問題を解決するため，周波数成分選択および時間区間選択処理の検討を行った。

2 発話方向推定原理

図 1 にマイクロホンアレーと受音信号の関係を示す。図において$S(\omega)$は発話音声，$M_1 \sim M_N$はN個のマイクロホン，$H_1(\omega,\theta) \sim H_N(\omega,\theta)$は話者が$\theta$方向を向いている時の話者 - マイクロホン間の伝達関数を表す。また$X_1(\omega,\theta) \sim X_N(\omega,\theta)$は，各マイクロホンでの受音信号を表しており，$X_i(\omega,\theta) = S(\omega)H_i(\omega,\theta)$と表される。ここで発話方向を$\theta = \hat{\theta}$と固定し，各変数をベクトル化すると，下記のように表現できる。

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

$$\begin{aligned}\mathbf{X}(\omega,\hat{\theta}) &= \left[\left|X_1(\omega,\hat{\theta})\right|,\cdots,\left|X_N(\omega,\hat{\theta})\right|\right]^T \\ &= |S(\omega)|\left[\left|H_1(\omega,\hat{\theta})\right|,\cdots,\left|H_N(\omega,\hat{\theta})\right|\right]^T \\ &= |S(\omega)|\,\mathbf{h}(\omega,\hat{\theta}) \end{aligned} \quad (2)$$

但し，Tは転置である。

この$\mathbf{h}(\omega,\theta)$は，音源が$\theta$方向を向いている時の各マイクロホンでの受音信号の大きさのパターンを表している。従って，各ベクトルを正規化して，$\mathbf{X}(\omega,\hat{\theta})$と各$\theta$方向の$\mathbf{h}(\omega,\theta)$との内積を取り，周波数平均した次式（3）が最大値となるθを求めることで，発話方向$\hat{\theta}$を推定することができる。

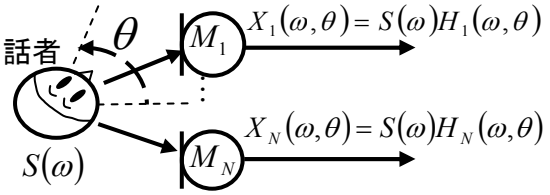

図 1 マイクロホンアレーと受音信号

$$\mathbf{C}(\theta) = \sum_{\omega} w(\omega)\,\mathbf{h}(\omega,\theta)^T\,\mathbf{X}(\omega,\hat{\theta}) \quad (3)$$

ただし，$w(\omega)$は周波数重みを表す。

3 周波数成分の選択 （周波数マスク付ヒストグラムの導入）

式（3）の周波数平均化において，内積の単純平均では内積値の絶対的大きさの影響を受けるので，ヒストグラムを導入した。具体的には，短時間区間ごとに DFT を行い，各周波数での推定結果のヒストグラムをとり，最大頻度の方向をその時刻の発話方向と推定した。しかし騒音環境下では，音声の周波数成分が小さい帯域も含めてヒストグラムを計算すると推定誤差が発生する。

そこで筆者らは，音声の周波数成分が小さい帯域では内積値が低下することに着目して，内積値の低い周波数成分をヒストグラムの計算から除去する周波数マスクを導入した。周波数マスクは次式の周波数重み$w(\omega)$として定義される。

$$w(\omega) = \begin{cases} 1 & (p(\omega)/p_{mean}(\omega) \geq \gamma) \\ 0 & (p(\omega)/p_{mean}(\omega) \leq \gamma) \end{cases} \quad (4)$$

ただし，$p(\omega)$は内積値を，$p_{mean}(\omega)$はその時間平均値を表す。またγは閾値を表し，今回はγ=1.5 とした。

4 時間区間の選択 （発話区間検出）

発話方向推定を行うべき時刻を特定するために，まず，音声が存在する時間区間の検出をおこなった。今回は，周期性の有無に基づ

* A study of time - frequency component selection for estimation of sound source orientation using a microphone array, by Keiko KIKUCHI[1], Tohru DAIGO[1], Hirofumi NAKAJIMA[2], Kazuhiro NAKADAI[2], Yuji HASEGAWA[2] and Yutaka KANEDA[1] (1 Tokyo Denki University, 2 Honda Research Institute Japan).

いた音声区間検出法[2]を利用した。この手法では，周期雑音が存在しない（または抑圧可能な）場合，エネルギーの大きい母音区間を検出することができる。しかし，この手法で検出した音声区間では，発話方向誤検出が発生した。

誤検出の発生部分を調べた典型的な結果を図 2 に示す。図において横軸は時間で縦軸は方向，色は内積値を表し，赤色の強い方向が推定方向となる。図に示すように，誤推定は音声区間の後半に多いことが判明した。これより，音声区間として検出された時間区間の後部は，発話区間ではなく，音声が残響として残っている区間であり，発話方向検出に利用するのは不適切な区間であると判断した。この結果に基づき，今回は音声区間の後部の数フレームを削除した区間を発話区間と定義することとした。

5　評価実験

5.1 測定条件

実験は広さ 7m×4m，高さ 3.5m，残響時間が約 230ms の実験室で行った。実験室の暗騒音レベルは約 40dB であった。マイクロホン数は 96 で，図 3 に○印で示すように室内の壁面に配置されている。

最初に，スピーカを部屋の中央に配置し，図 3 の 0° 方向から反時計回りに 15° 刻みで 345° まで回転させて（計 24 方向），式(1)の伝達関数ベクトルを測定した。

発話方向の推定対象は男性 1 名で，部屋の中央において 0°，90°，180°，270° の 4 方向に向き，「あ，い，う，え，お」と発話した。音声はサンプリング周波数 16kHz で収音した。収音信号は 64ms ごとに 1024 点 DFT をして発話方向推定を行い，誤差評価を行った。推定は①基本手法，②今回提案した周波数成分選択を追加した手法，③さらに時間区間選択を付加した手法の 3 つを比較した。

5.2 実験結果

実験結果を図 4 に示す。棒グラフは左から，上記①②③の条件に対応する。縦軸は推定方向の平均絶対誤差とその標準偏差を表す。

図より，周波数選択・時間選択という騒音・残響対策を行っていない①の手法では平均誤差が 35° 標準偏差が 21° であった。これに周波数成分選択を行うことで平均誤差は 18° となり，さらに時間区間選択を加えることで平均誤差 7°，標準偏差 13° 程度に減少させることができた。今回の伝達関数ベクトルの測定間隔，すなわち推定の分解能は 15° であり，標準偏差も含めて測定誤差を分解能以下とすることができた。

6　まとめ

本稿では，マイクロホンアレーを用いた発話方向推定法において，騒音や残響の存在する実環境での性能向上を目指して，時間 - 周波数選択についての検討を行った。そして，周波数マスク付ヒストグラム，および，残響を考慮した発話区間検出の導入を行うことで，従来 35° であった平均誤差を 7° に低下させることができた。

参考文献

[1] 醍醐他，音講論（春），pp.627-630. 2007.

[2] 伊藤，水島，信学技法，EA95-59，pp.17-25，1995.

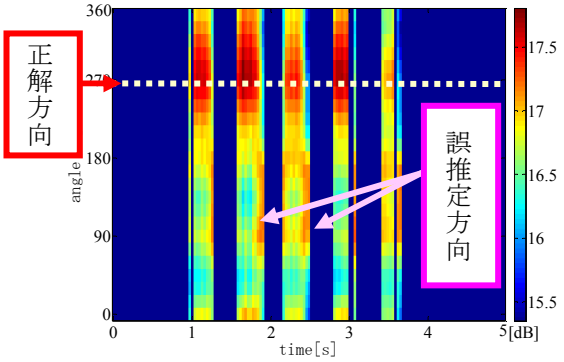

図 2 誤推定部分を含んだ推定結果

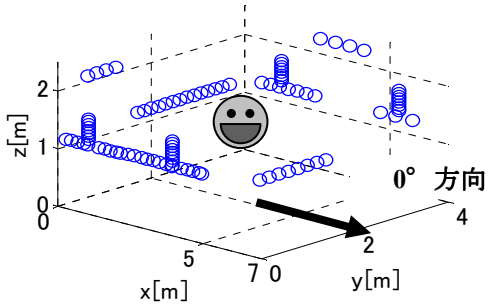

図 3 マイクロホンの配置図

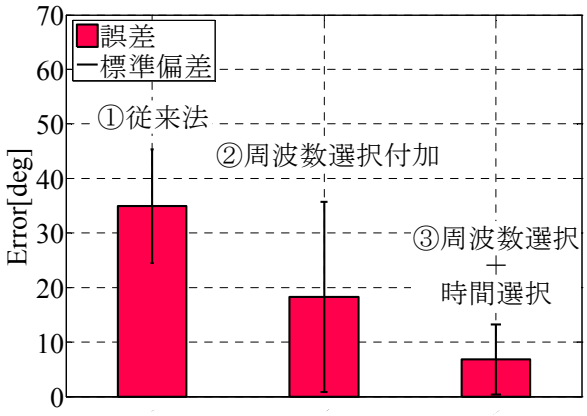

図 4 従来法と提案法の処理結果