

SBPH-MUSIC 法による複数音声方向推定の検討*

☆川田 希望, 金田 豊 (東京電機大)

1 はじめに

音源方向推定法は、多チャンネル信号を用いる MUSIC 法をはじめ多数の手法が提案されている[1]。しかし、一般的な室内で音源方向推定を行う際、反射音が誤推定要因となる。加藤らは、反射音への対策法として、SBPH(Sub Band Peak Hold)処理を MUSIC 法に適用した SBPH-MUSIC 法を提案し、1 方向の音声の到来方向推定について反射音耐性が向上することを確認した[2]。

しかし、複数音声の到来方向推定の場合、SBPH 処理において一方の音声他方の必要な音声の成分をマスクしてしまう可能性があった。そこで本報告では、SBPH-MUSIC 法による音声方向推定を、音声複数方向から到来する場合に条件を拡張し、実環境実験により検討を行った。

2 MUSIC 法による音源方向推定

複数マイクロホンの受信信号を用いて方向推定を行う。この時、各マイクの受信信号には音源方向 θ に依存した時間差 τ が生じる。マイクロホンの第 1 チャンネルを基準として、 θ 方向から到来する多チャンネル受信信号を周波数領域で表すと

$$\begin{aligned} \mathbf{X}(\omega, \theta) &= X_1(\omega) [1, e^{-j\omega\tau_2(\theta)}, e^{-j\omega\tau_3(\theta)}, \dots]^T \\ &= X_1(\omega) \mathbf{d}(\omega, \theta) \end{aligned} \quad (1)$$

となる。ここでは、 $X_1(\omega)$ は第 1 チャンネルの信号スペクトル、 ω は角周波数、 $\mathbf{d}(\omega, \theta)$ は方向制御ベクトル(steering vector)を表している。

MUSIC 法では、受信信号から算出される雑音部分空間の相関行列 $\mathbf{R}_n(\omega)$ と方向制御ベクトル $\mathbf{d}(\omega, \theta)$ を用いて、次式で方向スペクトル $P_{MUSIC}(\omega, \theta)$ を求めることができる。

$$P_{MUSIC}(\omega, \theta) = \frac{1}{\mathbf{d}^H(\omega, \theta) \mathbf{R}_n(\omega) \mathbf{d}(\omega, \theta)} \quad (2)$$

3 サブバンドピークホールド(SBPH)処理

SBPH 処理は、受信信号を短時間フーリエ変換するサブバンド処理と、周波数帯域毎に先行する直接音を利用し、後続の反射音をマスクするピークホールド処理で成り立つ。

SBPH 処理の流れを説明する(Fig. 1)。最初

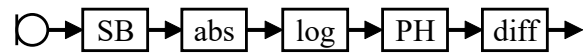


Fig. 1 SBPH 処理のブロック図

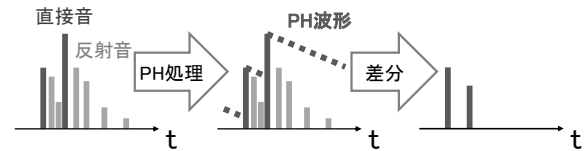


Fig. 2 PH 処理のモデル図

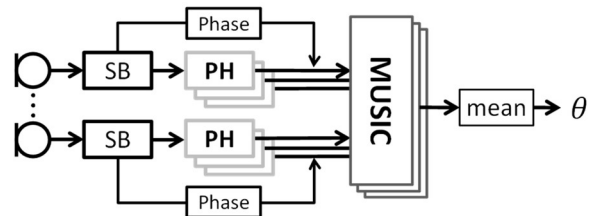


Fig. 3 SBPH-MUSIC 法のブロック図

に短時間フーリエ変換によって信号をサブバンド化(SB)する。次に各サブバンド信号の振幅成分(abs)の対数値(log)を取り出し、ピークホールド処理(PH:Fig. 2)を行う。ピークホールド処理では、まずは先行する音の振幅成分を保持し、減衰をかけていく。また、現在の値よりも大きいデータが来た場合はそちらの値を保持し、減衰をかけていくという操作をする。最後に時間差分(diff)をとる (以下、この abs から diff までの 4 つの処理をまとめて PH 処理と呼ぶ)。その結果、周波数帯域毎に直接音成分のみが抽出される。

Fig. 3 に SBPH-MUSIC 法のブロック図を示す。一連の操作の中で信号の振幅値を取るため、サブバンド信号の位相成分は失われてしまう。そこで、あらかじめ位相成分を保存しておき、SBPH 処理後に位相成分の情報を付加する。

Fig. 2 のモデル図からわかるように、PH 処理では先行する強い成分により後続の成分をマスクしている。複数の音声の到来方向推定の場合、一方の音声到来すると、他方の音声の成分をマスクしてしまい、正しく推定が行えない可能性が考えられた。

* A study of multi sound source direction estimation using SBPH-MUSIC, by KAWADA Nozomi, and KANEDA Yutaka (Tokyo Denki University).

4 複数音源に対する評価実験

4.1 実験条件

実験条件を Fig. 4、Table 1 に示す。Fig. 4 と Table 1 の条件で最初に、音声を単一話者ごとに録音する。次に、録音した2つのデータを足し合わせ、同時に2方向から到来する音声のデータを作成する。作成したデータに対し、MUSIC 法および SBPH-MUSIC 法で方向推定実験を行った。

得られた方向スペクトル $P_{MUSIC}(\omega, \theta)$ を使用帯域で平均した $P_{MUSIC}(\theta)$ からピークの値を大きい順に2点取り、2方向の推定結果とする。その時の角度 θ が片方でも許容誤差 $\pm 10^\circ$ から外れていたら誤推定とした。MUSIC 処理における想定音源数は1および2の両方で行った。サブバンド化パラメータ(短時間フーリエ変換を行うときのフレーム長とシフト長)は、フレーム長を64~2048、シフト長を2~1024の範囲で変化させ、各推定法で一番良い推定結果を最終結果とした。

4.2 実験結果

MUSIC 法と SBPH-MUSIC 法のそれぞれについて、合計の誤推定数を Table 2 に示す。Table 2 より、MUSIC 法と SBPH-MUSIC 法の誤推定数を比較すると、想定音源数が1の場合では誤推定数は99個から47個に減少した。また、想定音源数が2の場合では、誤推定数は110個から35個に減少した。

Fig. 5 に MUSIC 法による推定で反射音成分が誤差要因となってしまう結果の例を示す。また、Fig. 6 に、同一音声に対して SBPH 処理により反射音成分が低減できている結果の例を示す。

想定音源数を1にした場合、周波数ごとにパワーが最も強い音声の方向を推定する。音声のスパース性より、各音声が高い成分を持つ周波数は異なることが多いので、周波数平均を取ることで、反射音成分を抑圧し、かつ2方向検出することが可能となる。

想定音源数を2にした場合、片方の音声のパワーが強い周波数帯において、MUSIC 法では強い音声の直接音成分と反射音成分の2方向を検出してしまうことがあった。それにより、誤推定数が増加した。一方、SBPH-MUSIC 法では反射音の影響が低減できるため、誤推定数は減少した。

5 むすび

本報告では、SBPH-MUSIC 法の複数の音声の方向推定においての効果を実環境実験により検討をした。その結果、SBPH 処理を付加することで推定性能が向上することが確認できた。

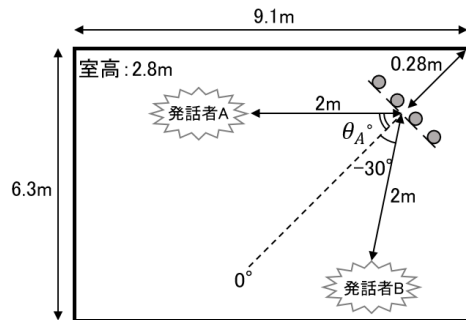


Fig. 4 実験配置図

Table 1 実験条件

残響時間	1.3 s (500 Hz)
マイクロホンアレイ	4 ch, 0.04 m 間隔
サンプリング周波数	48000 Hz
発話角度 θ_A, θ_B	発話者 A: $\theta_A = 45, 30^\circ$ 発話者 B: $\theta_B = -30^\circ$
使用帯域	3000~18000 Hz
音声データ	発話者 A: 男性・10 単語 発話者 B: 男性・14 単語

Table 2 各推定法の誤推定数

推定法 \ 想定音源数	MUSIC	SBPH-MUSIC
1	99	47
2	110	35

総推定数 280(10×14×2)

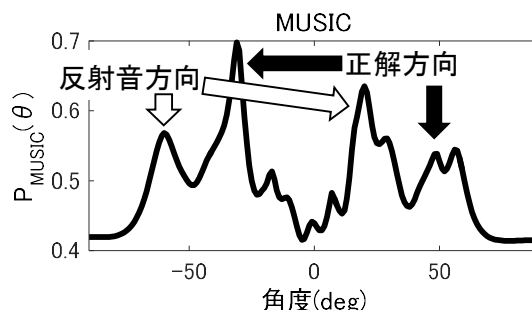


Fig. 5 MUSIC 法による推定結果

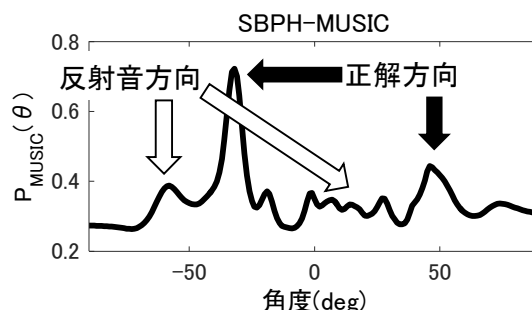


Fig. 6 SBPH-MUSIC 法による推定結果

参考文献

- [1] 浅野, “音のアレイ信号処理”, 日本音響学会, pp.107-124, 2011.
- [2] 加藤、金田、音講論(秋), pp.731-732, 2012.