

# SBPH-MUSIC 法による複数音声方向推定の検討 A study of multi-talkers direction estimation using SBPH-MUSIC

川田 希望<sup>†</sup>      金田 豊<sup>†</sup>

<sup>†</sup> 東京電機大学工学部

Nozomi KAWADA<sup>†</sup>

Yutaka KANEDA<sup>†</sup>

<sup>†</sup> Tokyo Denki University

## 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は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、音声は複数方向から到来する場合に条件を拡張し、シミュレーション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  $\theta$  に依存した時間差  $\tau$  が生じる。マイクロホンの第 1 チャンネルを基準として、 $\theta$  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、 $\omega$  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。

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

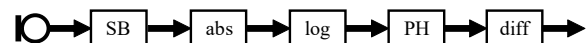


図 1 SBPH 処理のブロック図

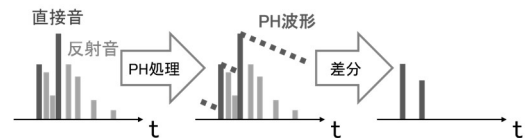


図 2 PH 処理のモデル図

## 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(図 1)。最初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)の対数值(log)を取り出し、ピークホールド処理(PH:図 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。

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。図 3 に SBPH-MUSIC 法のブロック図を示す。

図 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、一方の音声は到来すると、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。

## 4 シミュレーションによる評価実験

### 4.1 実験条件

実験条件を図4、表1に示す。図4における発話者A、Bから各マイクロホンまでのインパルス応答を、鏡像法により作成する。このインパルス応答と、異なる二人の音声を畳み込み、 $45^\circ$ 方向から到来する音声と $-30^\circ$ 方向から発された音声のデータを作成する。この時、各音声は1つの単語を繰り返している。こ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 $\theta$ 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・2のそれぞれの場合で行った。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、合計の誤推定数を表2に示す。表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50個から9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176個から118個に減少した。

想定音源数を1にした場合、周波数ごとにパワーが強い音声の方向を推定する。使用する周波数帯でそれぞれの音声が高いパワーを持つ成分が異なる場合、2方向を検出することが可能となる。よって、想定音源数を1とすることで、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。

一方、想定音源数を2にした場合、片方の音声のパワーが高い周波数帯において、強い音声の直接音成分と反射音成分の2方向を誤検出したり、雑音源が存在する $0^\circ$ 方向を誤検出したりし、これが誤推定数を上昇させる原因となった。

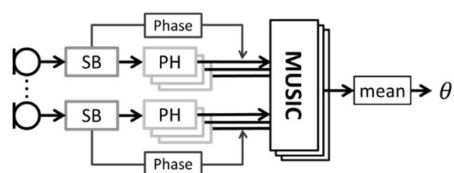


図3 SBPH-MUSIC法のブロック図

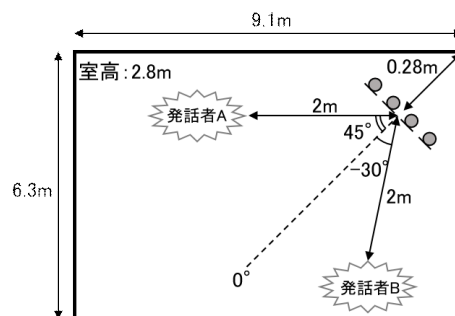


図4 シミュレーション配置図

表1 シミュレーション条件

残響時間	1.0 s (500 Hz)
マイクロホンアレー	4 ch, 0.04 m 間隔
サンプリング周波数	48000 Hz
発話角度 $\theta_A, \theta_B$	発話者 A: $\theta_A = 45^\circ$ 発話者 B: $\theta_B = -30^\circ$
使用帯域	3000~15000 Hz
音声データ	発話者 A: 女性・15 単語 発話者 B: 男性・15 単語

表2 各推定法の誤推定数

想定音源数 \ 推定法	MUSIC	SBPH-MUSIC
1	50	9
2	176	118

総推定数 255(15×15)

## 5 むすび

本報告では、SBPH-MUSIC法が複数の音声の方向推定においての効果を実験により検討をした。その結果、想定音源数は1とすることが適当であり、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。

### 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.