

# 実環境におけるサブバンドピークホールド 音源方向推定法の有効性の検討\*

佐藤耕平, ○金田豊 (東京電機大・工)

## 1 はじめに

音源方向推定技術の問題点として、室内反射音による推定精度の低下が挙げられる。本研究では、反射音耐性がある手法として知られている PHAT 法(または CSP 法)[1]と、筆者らが提案したサブバンドピークホールド (SB-PH)法[2]とを、反射音が強い実環境において性能比較したので報告する。

## 2 時間差に基づく音源方向推定

図 1 に示すように、 $\theta_s$  方向から音波が到来する時、受信信号  $x_1(t), x_2(t)$  に生じる時間差  $\tau_s$  から到来方向  $\theta_s$  を求めることができる。時間差  $\tau_s$  の推定は、2つのマイクロホン受信信号  $x_1(t), x_2(t)$  の相互相関関数  $\phi_{12}(\tau)$

$$\phi_{12}(\tau) = \sum_t x_1(t) \cdot x_2(t + \tau) \quad (1)$$

が最大値をとる  $\tau$  の値として求められる。

## 3 反射音の影響

### 3.1 パルス音の相関関数

図 2(a)に、パルス音が到来した場合の受信信号  $x_1(t), x_2(t)$  のモデルを示す。直接音は  $\tau_s$  の時間差で受信されるが、反射音は音源とは異なる方向から到来するため、 $\tau_s$  とは異なった時間差で受信される。

この 2つの信号の相関関数は図 2(b)のように複数のピークを持つ。具体的には、直接音同士、反射音同士、および直接音と反射音同士の相関である。通常、直接音は反射音に比べてエネルギーが大きいので、直接音同士の相関値  $\phi_{12}(\tau_s)$  が最大となる。

### 3.2 音声信号の相関関数

信号が音声の場合の相互相関関数は、音声の自己相関関数(図 2(c))を図 2(b)に畳み込んだものになる。音声の自己相関関数は時間軸上に広がったものであるため、これを畳み込むと、図 2(b)が平滑化され、複数のピークが融合されるのでパルスの密度の高い時刻が高い値を持つようになり、最大値が  $\tau_s$  とは異なるものとなる(図 2(d))。

## 4 反射音の影響改善

### 4.1 PHAT 法(CSP 法)

反射音の影響を軽減する方法として PHAT 法が知られている[1]。この方法は、平滑化された相関関数を白色化することで、相関関数のピークを鋭くする。その結果、複数の反射音ピークが分離され、直接音の相関値の最大化を図るものである(図 2(e)(f))。

しかし、有限帯域での白色化であるので、狭い時間間隔の反射音ピークの融合は避けることはできず、性能向上には限界がある。

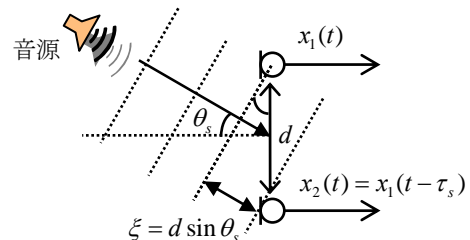


図 1 時間差に基づく音源方向推定の原理

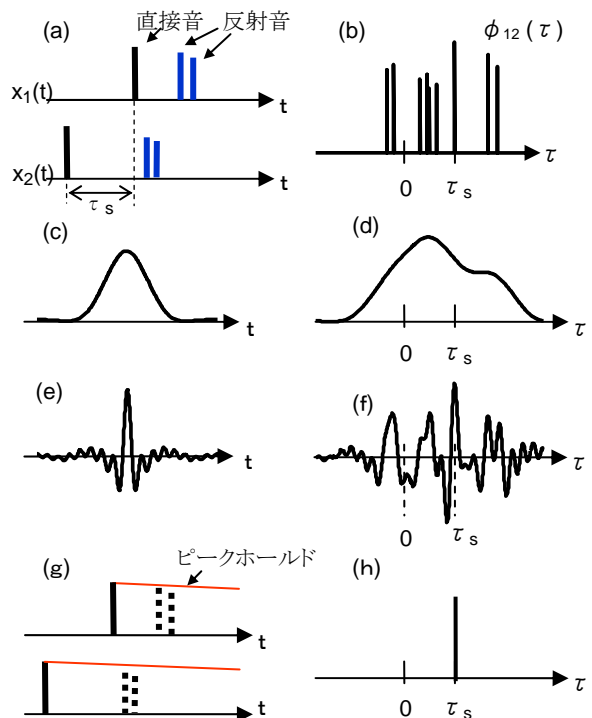


図 2 反射音の影響とピークホールド処理

\* A study of the effectiveness of sub-band peak-hold sound source direction estimation method in a real environment, by SATO, Kohei and KANEDA, Yutaka (Tokyo Denki University).

## 4.2 サブバンドピークホールド(SB-PH)法

筆者らは直接音の振幅を保持することで反射音をマスクする(図 2(g)) ピークホールド処理を提案した[2]. ピークホールド後の信号の時間差分をとれば, 直接音の立ち上がり部分が検出され, 反射音の影響のない相関関数を得ることができる(図 2(h)).

音声は周波数によって立ち上がり時間が異なるのでサブバンド信号に分割しピークホールド処理を行う. 処理の流れを図 3 に示すブロック図に従って説明する. まず①短時間フーリエ変換(STFT)より信号をサブバンド化し, ②時変振幅スペクトル $|X_1(\omega, t)|$ を出力する. 次に各バンドに対して③ピークホールド処理(PH)を行い. 大きな複合反射音の対策のための④対数操作(log)を行い, ⑤時間差分(Diff)を取ることでサブバンド信号の立ち上がりを取り出す. これらの⑥相互相関関数(Cor)の⑦総和をとり, 時間差を推定する.

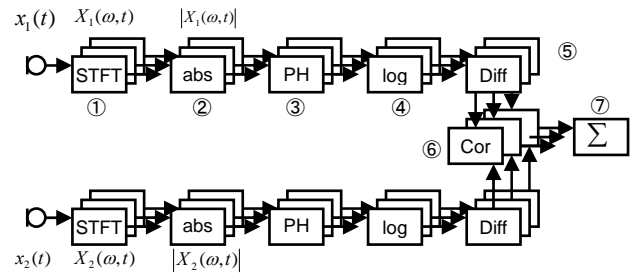


図 3 SB-PH 処理のブロック図

表 1 実験条件

部屋の寸法	3.3[W]×5.8[D]×2.4 [H] [m]
残響時間	0.5 s
SN 比	30 dB
マイクロホン間隔 $d$	0.3 m
音源距離 $r$	1m, 2m
音源方向 $\theta_s$	-60, -30, 30, 60 deg.
音声	音響、信号、野球など 男性話者による 30 発声

## 5 評価実験

先の報告[2]では  $30\text{m}^2$  程度の会議室で評価を行ったが, 今回は  $20\text{m}^2$  と狭くて壁面がコンクリートまたはガラス窓の強い反射音が多く存在する環境で, PHAT 法, と SB-PH およびサブバンド信号に位相成分を付加した SB-PHP 法[3]の評価実験を行った. 実験条件は, 表 1 と図 4 に示す.

実験結果を図 5 に示す. 推定方向が正解方向に対して $\pm 5^\circ$  となったものを正答とした. 図 5 (a)は音源距離が 1m の場合の結果で, いずれの方法もほぼ 100%の正答率となっている. 一方, 図 5(b)は音源距離が 2m の場合である. 音源距離が大きくなると直接音のエネルギーが低下するため反射音の影響が大きくなり, PHAT 法の正答率は大幅に低下した. しかし, 反射音をマスクする効果を持った SB-PH ならびに SB-PHP 法では, 1m の場合と同様にほぼ 100%の正答率が得られた.

## 6 むすび

本報告では, 反射音の強いコンクリート壁付近で方向推定実験を行い, SB-PH 法が従来法である PHAT 法と比べて有効であることを示した. これまで複数の環境で評価実験を行ってきたが, 今回のような環境は従来法との差が大きく, 反射音の強い小室は SB-PH 法の有効性が顕著な環境の一つと考えられる.

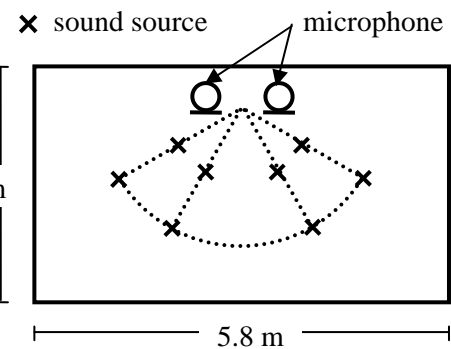


図 4 本実験の配置図

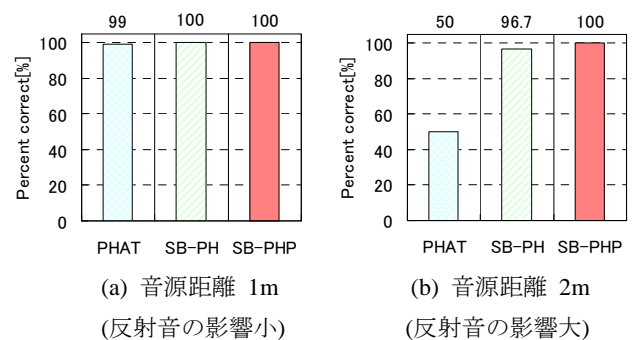


図 5 実験結果

## 参考文献

- [1] M.Brandstein ほか, Microphone Arrays, Springer, New York (2001) pp.157-180.
- [2] 鈴木ほか, 日本音響学会誌, 65, 10, (2009).
- [3] 佐藤ほか, 音講論, 3-Q-19, (2009.9).