

サブバンドピークホールド処理を用いた音源方向推定の検討*

©鈴木敬, 金田豊 (東京電機大・工)

1 はじめに

一般の室内で音源方向推定を行う際、反射音の影響により推定精度が低下する。この対策として、我々は、ピークホールド処理を用いた手法を検討し、拍手音に対する有効性を確認した [1,2]。しかし、音声信号に対しては十分な効果が得られなかった。そこで本報告では、その原因を検討し、周波数サブバンド毎にピークホールド処理を行うことで、大幅な精度向上が確認できたので報告する。

2 2chマイクとピークホールド処理による音源方向推定

2つのマイクロホンで受信した信号 $x_1(t)$, $x_2(t)$ の時間差 τ_s が分かれば音波の到来方向 θ_s は次式のように求められる。

$$\theta_s = \sin^{-1}(c \cdot \tau_s / d) \quad (1)$$

ただし、 c は音速、 d はマイク間距離である。この時間差 τ_s は、例えば、 $x_1(t)$ と $x_2(t)$ との相互相関関数 $\phi_{12}(\tau)$ の最大値を与える τ の値として推定されるが、反射音の多い環境下においては良好な推定結果を得ることができない。

図1(a)に、パルス音に単一反射音が付加された受信波形 $x_1(t)$, $x_2(t)$ の例を示す。これらから相互相関関数を計算すると、直接音の時間差 τ_s 以外に直接音と反射音、反射音同士の時間差に起因する複数のピークを持ち、誤推定の原因になる。

図1(b)に、受信波形にピークホールドをかけた波形を示す。このように、ピークホールドは直接音の大きさを保持するので、後続する低振幅の反射音はマスクされ、直接音の時間差 τ_s が明確となる。

3 サブバンドピークホールド処理

従来のピークホールド法は、反射音耐性に優れているが [1]、音源が拍手音である場合と比べて、音源が音声である場合は、満足できる性能が得られなかった [2]。この原因は、拍手音と音声の波形の違いにある。拍手音と

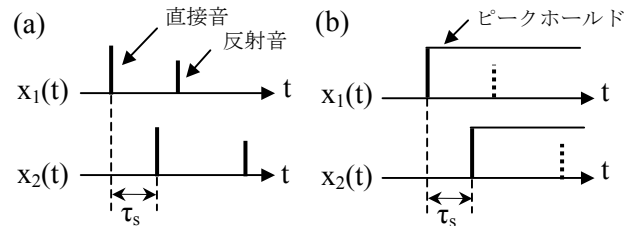
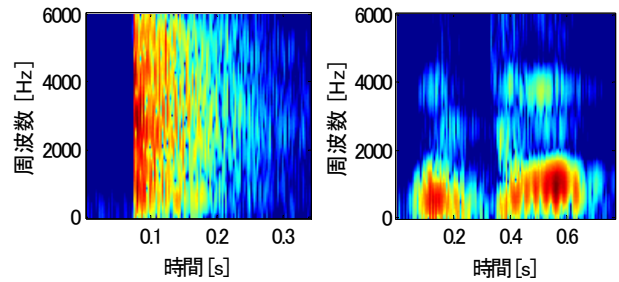


Fig. 1 ピークホールド処理の原理図



(a) 拍手音

(b) 音声

Fig. 2 拍手音と音声のスペクトログラム

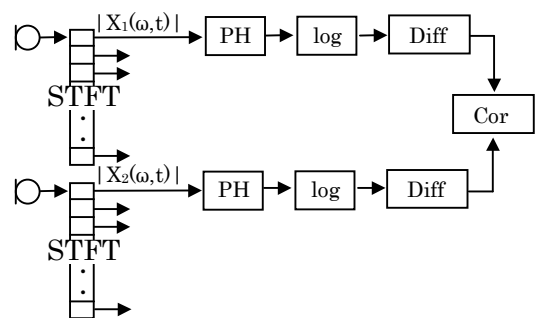


Fig. 3 提案手法のブロック図

音声のスペクトログラムを図2に示す。一般に、音の方向性情報は信号の立ち上がり部分の直接音において明確であり、図2(a)の拍手音は、全ての帯域で信号の立ち上がり時刻が同一である。一方、図2(b)の音声では、帯域ごとに立ち上がり時刻が異なり、帯域によっては成分を持たない。よって、波形上では各帯域の直接音成分が時間的に分散するため、不明確となる。

そこで、音声信号を帯域分割し、帯域毎にピークホールドをかけて直接音のみの観測機会を増やすことを狙い、サブバンドピークホ

* A study of DOA estimation based on Sub-band Peak Hold processing, by SUZUKI, Takashi and KANEDA, Yutaka (Tokyo Denki University).

ールド法を提案する。サブバンドピークホールド処理のブロック図を図3に示す。まず、受信信号を短時間フーリエ変換 (STFT) し、振幅成分の時系列 $|X_i(\omega, t)|$ を出力する。そして、ピークホールド処理 (PH), 対数操作 (log), 時間差分 (Diff) を行い[1], 2つの受信信号の対応する周波数成分時系列毎の正規化相関関数 (Cor) を求めて、これを全ての周波数に対して平均化する。

4 実験条件

提案法の反射音耐性を確認するため、室内において音源方向推定実験を行った。表1に実験条件を、図4に実験配置図を示す。この実験では、反射音の影響が大きい条件を想定し、マイクロホンの近くに2つの反射板を配置した。比較対象の従来法としては、従来のピークホールド法と反射音耐性を持つとされている PHAT 法(または CSP 法)[3]を用いた。

5 実験結果

実験結果の正解率を図5に示す。(a)が許容誤差を 5° とした時、(b)が許容誤差を 10° とした時の正解率である。図5(a)より、従来のピークホールド (PH) 法の正解率が46%, PHAT 法が57%であるのに対し、提案法では82%と高い正解率が得られている。また、許容誤差を 10° とすると (図5(b)), 提案法の正解率が97%と大幅に向上していることが分かる。

次に、推定方向誤差の大きさのヒストグラムを図6に示す。(a)が PHAT 法, (b)が提案法である。図6(a)より、PHAT 法は推定方向の大きな誤りがいくつも見られるのに対して、図6(b)の提案法では、誤差の大きさがほぼ 10° 未満に収まっていることが分かる。よって提案法は、従来法に比べて反射音耐性があり、また、大きな誤りが少ないことが分かる。

6 むすび

従来のピークホールド法を改良したサブバンドピークホールド法を提案した。提案法は、反射音の影響が比較的大きいと思われる条件下で、音声に対して従来のピークホールド法と PHAT 法よりも反射音の影響を受けにくいことを示した。さらに、推定方向を大幅に誤推定してしまう PHAT 法に比べ、提案法は誤差の大きさがほぼ 10° 未満に収まっており、その有効性が見てとれる。

Table 1 実験条件

標本化周波数	12000 [Hz]	部屋の寸法	5.0×9.0×2.4 [m]
マイク間距離	0.6 [m]	音源	人間 (男性 2人)
残響時間	0.4 [s]	音声	“爆音”, “音響”
騒音レベル	45 [dBA]	音源距離 r	1,2,3 [m]
		音源方向 θ	-60,-30,0,30,60 [deg.]

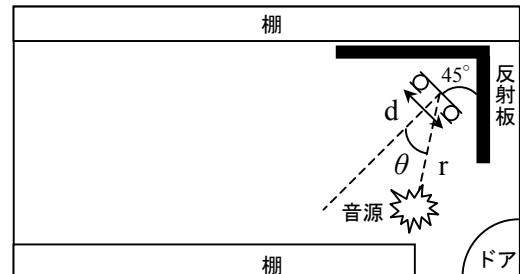


Fig. 4 実験配置図

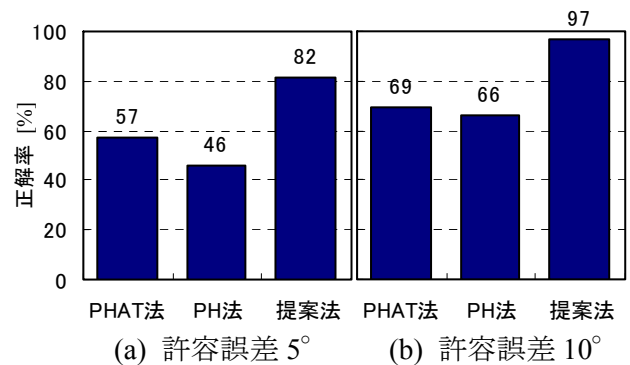


Fig. 5 実験結果

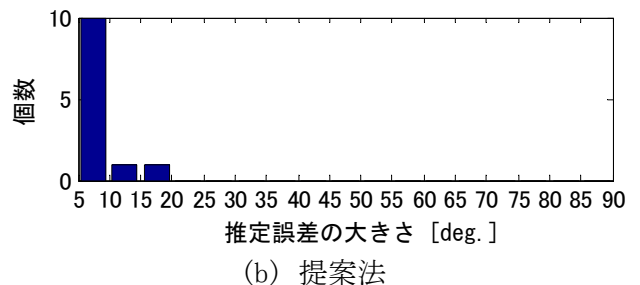
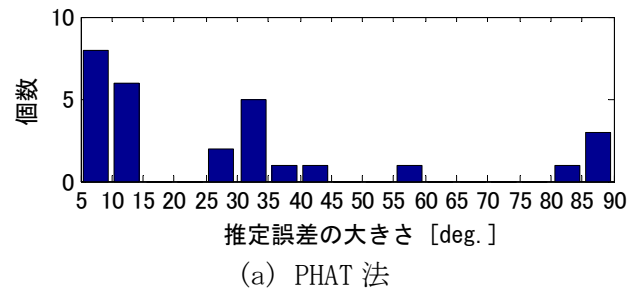


Fig. 6 推定方向誤差の大きさ

参考文献

- [1] 木皿, 金田, 音講論(秋), 711-712, 2005.
- [2] 木皿, 金田, 音講論(春), 631-632, 2006.
- [3] C. H. Knapp *et al.*, IEEE Trans. ASSP, 24(4), 320-327, 1976.