

位相を付加したサブバンドピークホールド音源方向推定法の 処理時間短縮法の検討

今西 祐生[†] 金田 豊[†]

[†] 東京電機大学工学研究科 〒120-8551 東京都足立区千住旭町 5 番

E-mail: [†] 14kmc03@ms.dendai.ac.jp, kaneda@c.dendai.ac.jp

あらまし 音源方向推定技術は、指向性集音技術やロボットの聴覚機能などに応用が期待される技術である。実環境で音源方向を推定する場合、反射音の影響で誤推定が発生する。これに対して、筆者らが提案するサブバンドピークホールド(SBPH: Sub-Band Peak Hold)処理や、それに位相を付加した SBPHP(Sub-Band Peak Hold Phase)処理は、2ch マイクロホンアレーにおいて、反射音耐性に優れる音源方向推定法である。しかし、SBPHP 処理は処理時間が長く、実時間処理に適さないことが課題であった。本稿では、SBPHP 処理の処理時間短縮法を提案する。提案法では、サブバンド化の際に窓関数のシフト長を大きくし、計算コストを削減した。提案法の有効性を検討した結果、従来法の約 1/4 の処理時間で推定が可能となった。

キーワード 音源定位, 到来方向推定, SBPH 処理, 計算コスト削減, 反射音耐性

A study of processing time reduction for sound source localization by sub-band peak hold phase processing

Yuki Imanishi[†] and Yutaka Kaneda[†]

[†] Graduate School of Engineering, Tokyo Denki University

5 Senju Asahicho Adachi-ku, Tokyo, 120-8551 Japan

E-mail: [†] 14kmc03@ms.dendai.ac.jp, kaneda@c.dendai.ac.jp

Abstract

Sound source localization is applied to various technologies such as directional sound receiving and robot audition. In a real environment, however, a localization error occurs owing to reflected sounds. To reduce the localization error, we previously proposed sub-band peak hold (SBPH) processing and sub-band peak hold phase (SBPHP) processing. These processing methods work well but require a large processing time, making them unsuitable for real-time processing. In this paper, we propose a method of reducing the time required for SBPHP processing. In this method, we reduce the window shift length for sub-band processing while maintaining the localization performance. Experimental results show that the processing time of the proposed method is about 25% shorter than that for the conventional method.

Keywords Source localization, DOA estimation, SBPH processing, Computational cost reduction, Robustness to reflective sound.

1. はじめに

音源方向推定は、指向性集音技術やロボットの聴覚機能など様々な応用が期待される技術である。これまで、マイクロホンアレーを用いた手法が数多く研究されてきた。2つのマイクロホンを用いて音源方向を推定する代表的な手法としては、相互相関関数(CC: Cross Correlation)法、クロススペクトルを白色化する白色化相互相関(CSP: Cross-power Spectram Phase analysis)法がある[1]。また3つ以上のマイクロホンを使用する場合は、MUSIC (MUltiple SIgnal Classification)法[2]が主に用いられている。

音源方向推定はマイクロホンに録音された信号の

時間差に基づいて推定される。しかし、実環境では壁や天井からの反射音の影響で、誤推定するという問題がある。そのため、音源方向推定法は反射音耐性に優れることが求められる。

筆者らが提案するサブバンドピークホールド(SBPH: Sub-Band Peak Hold)処理 [3]と、SBPH 処理を発展させて位相成分を付加したサブバンドピークホールド(SBPHP: Sub-Band Peak Hold Phase)処理は、録音した信号から反射音の影響を取り除き、直接音のみを利用することで、高い推定性能を持つことが確認されている[4]。

この SBPHP 処理は、CC 法と組み合わせることで、

2ch という小規模なマイクロホンアレーを用いて、反射音耐性の強い方向推定が可能である。しかし、従来の 2ch マイクロホンアレーでの方向推定法と比較すると、処理時間が長いことが課題であった。本稿では、この問題を解決するために、SBPHP 処理の処理時間の短縮法を提案し、従来法と比較することで、その有効性を確認する。

2. SBPHP-CC 法

2.1. 2ch マイクロホンによる音源方向推定

音源方向は、2 つのマイクロホンに入力された信号の時間差によって推定される。Fig.1 に示すように、角度 θ の方向から音が到来した場合、音源から 2 つのマイクロホンまでに距離差 $d \sin \theta$ が生じるため、受信された信号には、時間差 τ が生まれる。この時、マイクロホン間隔を d 、音速を c とすると、次式によって角度 θ を求めることが出来る。

$$\theta = \sin^{-1} \frac{c\tau}{d} \quad (1)$$

この時、 c 、 d は固定値であるため、時間差 τ を求めることが出来れば、音の到来方向 θ を求めることが出来る。

この時間差 τ は、録音した信号の相互相関関数の最大値をとる時間より求められる。しかし、Fig.2 のように実環境で録音した場合、反射音が混入する。反射音は、音源と異なる方向から到来し、直接音と相関を持つため誤推定の主要因となる。特に、音源の位置が遠距離の場合や、マイクロホンが壁際に設置された場合は、反射音の影響が大きくなる。そこで、相互相関を計算する前に、反射音の影響を軽減する SBPH 処理を加えることで、推定率の向上を図る。SBPH 処理に関して以下に述べる。

2.2. ピークホールド(PH)処理

PH 処理の概要を Fig.3 に従って述べる。先に述べたように、実環境で録音した音には直接音と反射音が混在する。直接音は音源から最短距離で到来するため反射音に先行し、またその振幅も反射音より大きい。この事を利用して、直接音の振幅値を保持(PH)することによって反射音をマスクする。

このとき、直接音の振幅値を一定値で保持した場合、後続する直接音までマスクしてしまい、後続の直接音が利用できなくなってしまうため、ホールド値に減衰をかける。この減衰は、測定室の残響時間から得た指数減衰とする。最後に時間差分(diff)をとることで、反射音を取り除き、直接音のみを得ることが出来る。

ただし、複数の反射音が同時に入力されると、その振幅は PH によるマスクよりも大きくなり、反射音成分が十分に抑圧できない。例えば Fig.4 では、実線で描かれた直接音の振幅が 10、破線で描かれた後続の反

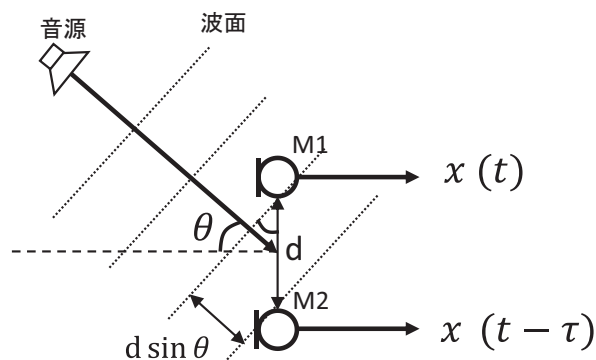


Fig.1 2ch マイクロホンによる音源方向推定の原理

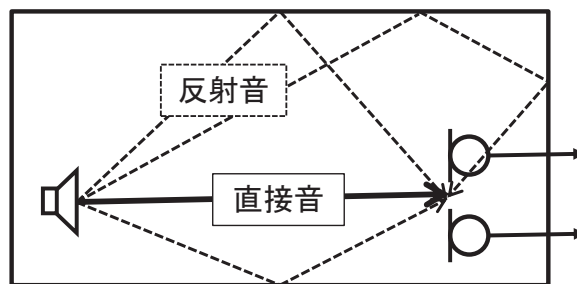


Fig.2 実環境における反射音の影響

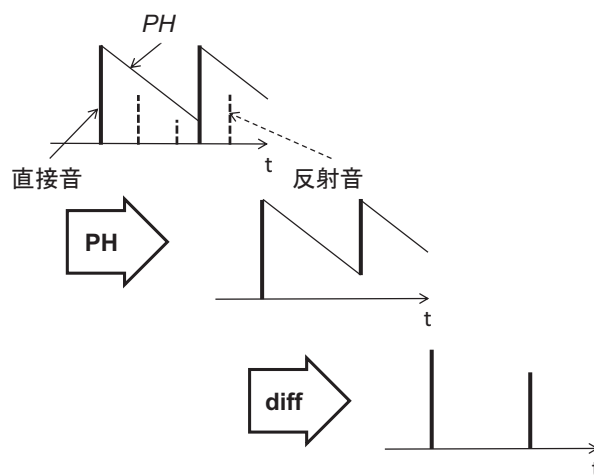


Fig.3 ピークホールド(PH)処理

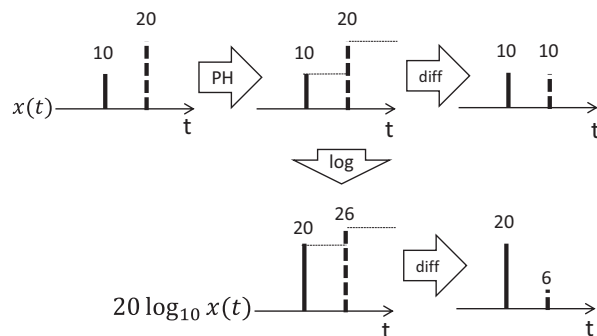


Fig.4 ピークホールド振幅値の対数操作の効果

射音の振幅が 20 であったとする．この場合，PH 及び差分処理後の振幅値が直接音，反射音ともに 10 の値となる．

そこで，振幅に対して対数値をとる(log)．その場合，直接音の振幅が 20，反射音の振幅が 6 となり，直接音と比較すると，反射音の振幅が小さくなるのが分かる．以上より，Fig.4 のように対数値で計算することで，直接音より大きな反射音成分の値を小さくし，その影響を軽減させる．

以降では，以上に述べた PH, diff, log を合わせて PH 処理と呼ぶ．

2.3. サブバンド(SB)処理

PH 処理は，直接音成分と考えられる信号の立ち上がり部分に注目して，2ch 間の時間差を求めるものである．一方，音声信号などは，Fig.5 のスペクトログラムに示されるように，周波数帯によって立ち上がり時刻が異なっている．このような信号の各周波数成分の立ち上がりを有効に活用するために，周波数帯域分割をする SB 処理を行う．

SB 処理には短時間フーリエ変換(STFT)を使用する．STFT は，窓関数を $w(t)$ ，窓関数の時間的位置を t_s とすると，以下のような式で定義される．

$$X(\omega, t_s) = \int_{-\infty}^{\infty} w(t - t_s)x(t)e^{-j\omega t} dt \quad (2)$$

SB 処理は，窓関数を一定の時間幅(シフト長)で移動させて STFT を行う．STFT 後は，時間と周波数の 2 変数を持つ関数となる．この関数の周波数変数を固定して考えると，Fig.6 のように，周波数帯域ごとの時間信号(サブバンド信号)が得られる．SB 処理後の信号は，フーリエ変換によって複素数となるので，PH 処理を行うためには絶対値を取る必要がある．

SB 処理において分析窓長(フレーム長)を増加させると，周波数分解能が向上する．一方，窓のシフト長を増加させると時間分解能が劣化する．そのため，時間分解能が角度分解能に対応する音源方向推定においては，シフト長を最小値である 1 サンプルとして計算を行なうと最良の結果が得られる．

2.4. SBPHP 処理

前述した SB 処理と PH 処理を組み合わせ，反射音耐性を得る手順をサブバンドピークホールド(SBPH)処理と呼ぶ．Fig.7 に従って処理手順を説明する．最初に，受信した信号は短時間フーリエ変換(STFT)によってサブバンド化を行う．次に振幅値を取り出すため，絶対値(abs)を取った後，各周波数帯域の時間信号に対して PH 処理をかける．その後，対数(log)をとり，時間差分(Diff)を行うことで，反射音の影響を低減した信号を得ることが出来る．SBPH 処理を受音信号にかけた後，

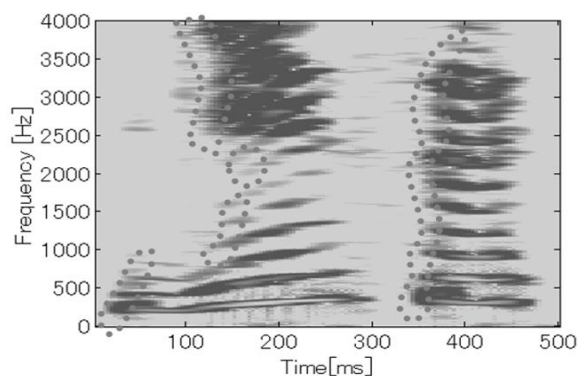


Fig.5 周波数による立ち上がり時刻の違いを示す音声のスペクトログラム(破線が立ち上がり部分)

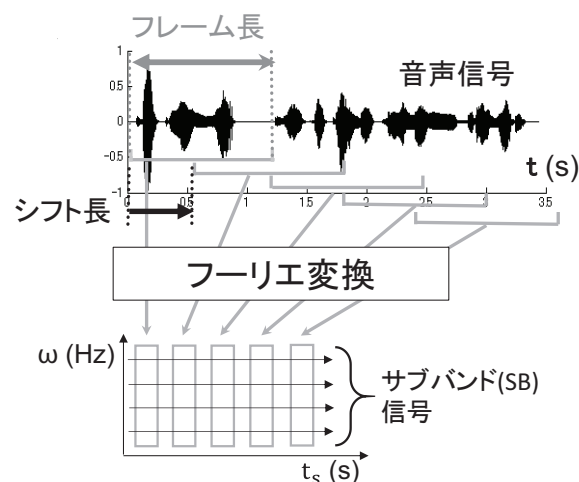


Fig.6 サブバンド処理

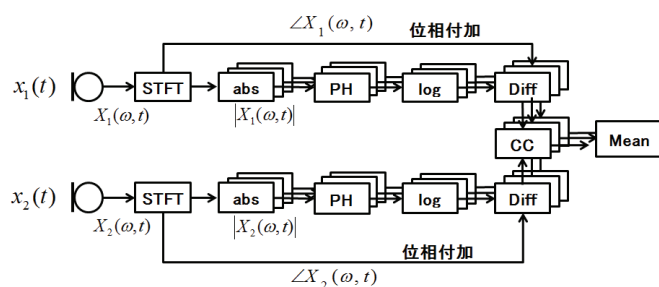


Fig.7 SBPHP-CC 法のブロック図

各周波数帯域の正規化相互相関関数(CC)を計算し，それらを平均して得られた波形の最大値から時間差 τ を計算する．このように，SBPH 処理と CC 法を組み合わせた音源推定法を SBPH-CC 法と呼ぶ．

しかし，SBPH 処理ではサブバンド信号の絶対値を取ったことで，位相成分を失ってしまう．位相成分を失い，振幅成分のみとなった場合，Fig.8 のように，サブバンド信号は全ての周波数帯において，低い周波数成分のみを持つようになる．低い周波数成分のみの相関波形はピークが鈍化するため，誤差が発生しやすく

なる。そこで、PH 処理で失われた位相成分を保存しておき、PH 処理後に付加するという手順を加える。

この手順を加えた SBPH 処理を、位相(Phase)を付加したサブバンドピークホールド(SBPHP)処理と呼ぶ。位相を付加することで、Fig.9 に示すように、高い周波数帯では、高い周波数成分のみを持つようになる。

SBPHP 処理の後に CC 法を適用した手法を SBPHP-CC 法と呼ぶ。各サブバンドの相関関数が低周波数成分しか含まない SBPH 処理よりも、様々な周波数成分の相関波形の平均をとる SBPHP 処理は相関波形が先鋭化され、誤推定が減少するため、SBPHP-CC 法は SBPH-CC 法と比較すると、誤推定率が減少する。

3. 処理時間短縮法の提案

従来、SBPHP-CC 法では方向推定に必要な時間分解能を確保するために、STFT の分析窓のシフト長を最小値の 1 サンプルとして計算を行っており、その結果、処理時間が長いという問題があった。そこで、STFT のシフト長を大きくし、フーリエ変換の回数を減らすことで、処理時間を短縮させる方法を検討した。

Fig.10(a)(b)(c)は、各サブバンド信号の相互相関関数の大きさを示した図で、横軸に 2ch 間の遅れ時間に対応する角度を表し、縦軸にサブバンド周波数を表している。白くなるにつれて相関値が高くなる。以降この図のことを角度スペクトログラムと呼ぶ。図において、サブバンド周波数が高くなるに従って、相関関数の周波数成分も上昇するので、白と黒の幅が狭くなっている。

従来、SBPHP-CC 法で使用されていたシフト長 1 の時の角度スペクトログラムの例を Fig.10(a) に示した。この時、相関値の大きい白い部分が音源方向(60°)に直線状に存在していることが分かる。SBPHP-CC 法では、最後に全ての周波数帯域の相関波形の平均(縦方向の平均)を取るのので、Fig.10(a)のような特性を持っていると、音源方向の相関値が強調されて、正しい方向を推定できる。

さて、処理時間短縮のためにシフト長を増加させることは、サブバンド信号のサンプリング周波数が小さくなる事に相当し、サンプリング定理を満たさなくなる。Fig.10(b)は、Fig.10(a)と同じ受信信号に対して、2 倍のシフト長でサブバンド処理を行った結果である。図におけるサンプリング周波数 f_s は 48 kHz であるため、 $f_s/2 = 24$ kHz まで表示されている。

シフト長を 2 倍にし、サンプリング周波数を $f_s/2$ とした結果、 $f_s/4 = 12$ kHz 以上のサブバンドに折り返しが発生する。折り返しによって高いサブバンド周波数の相関関数に含まれる高い周波数成分が、低い周波数成分になるため、相関波形が鈍化し、誤差の原因となる。また、折り返しが発生した帯域では、位相が共役

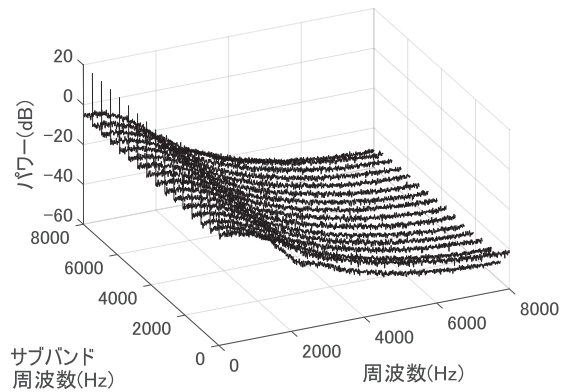


Fig.8 各サブバンド信号を振幅成分のみとした場合周波数特性

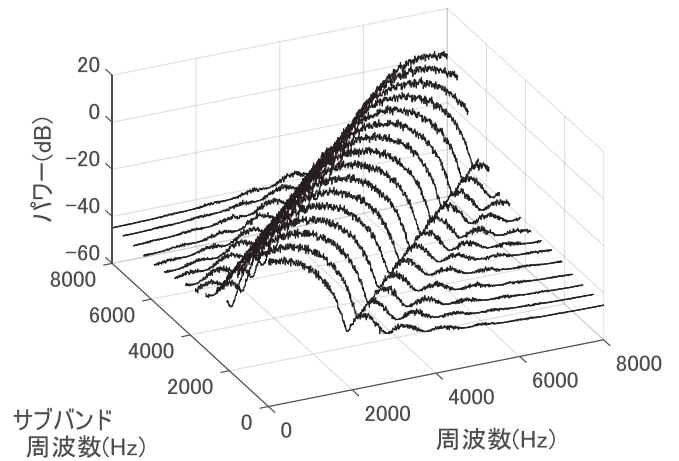
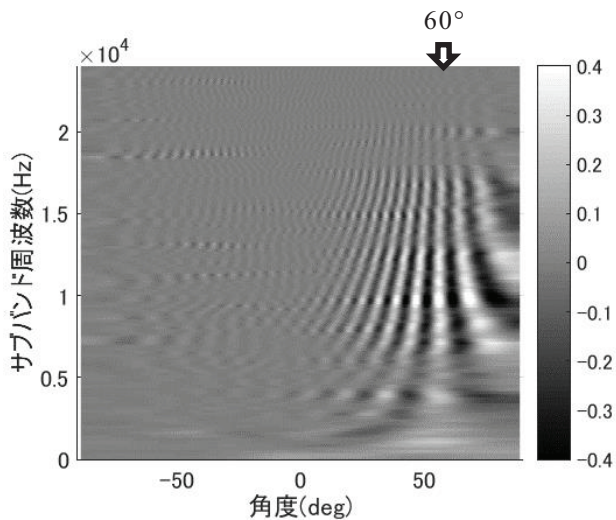


Fig.9 位相を付加した場合の各サブバンド信号の周波数特性

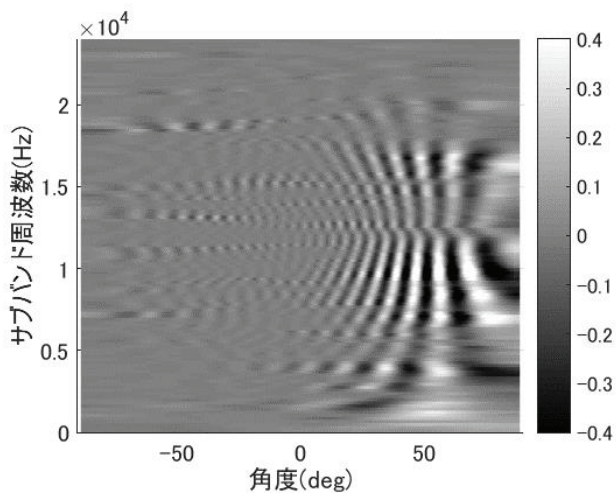
になる。そのため、相関値の正負が反転し、平均を取るとピーク値が小さくなり、誤推定の要因となる。そこで、正しい推定を行うためには、サンプリング周波数を元の値に戻し、折り返しが発生した帯域を元に戻す必要がある。

シフト長の増大に伴うサンプリング周波数の減少を元に戻すため、各サブバンドの相関波形に対して補間(アップサンプリング)を行う。しかし、単純に補間を行っても、折り返しによって低い周波数となってしまった帯域は、低い周波数成分のまま補間される。そこで、補間を行ってサンプリング周波数を増加させた後、折り返しが発生した帯域に、ナイキスト周波数の正弦波をかけることで、低くなってしまった周波数成分を、元の高い周波数成分に戻すという手順を加える。

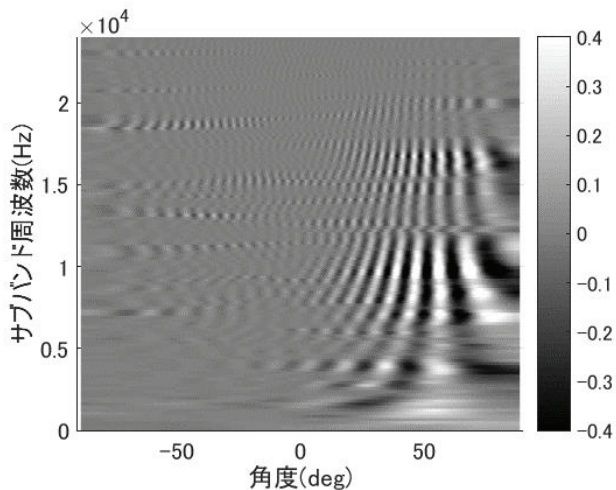
Fig.11 に折り返しを補正する方法の概念図を示す。シフト長を 2 倍にして相互相関関数のサンプリング周波数が $1/2$ (f_s から $f_s/2$) になったと仮定する。この時、 $(f_s/2)/2$ を越える成分は、低周波数成分に折り返



(a)シフト長 1 の時の角度スペクトログラム例



(b)折り返しが発生した角度スペクトログラム (シフト長 2 倍)



(c)折り返し補正後の角度スペクトログラム

Fig.10 折り返し補正の効果

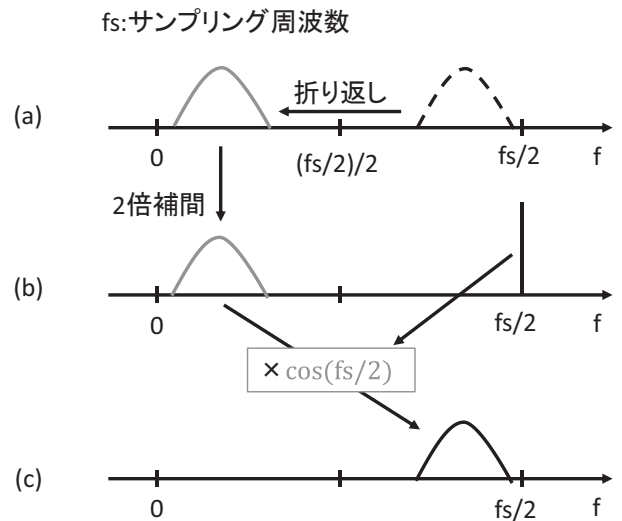


Fig.11 折り返し補正法の概念図

される(Fig.11(a)). これを元の高周波数成分に戻すには 2 倍補間をして元のサンプリング周波数とし(Fig11(b)), ナイquist周波数成分を持つ正弦波 $\cos(fs/2)$ との積を取れば良い(Fig.11(c)).

折り返し補正をかけた後の角度スペクトログラムを Fig.10(c)に示した. Fig.10(b)と比較すると, 高いサブバンド周波数の相関関数に高い周波数成分(白と黒の変化が密な部分)が含まれていることがわかる. 従って, 全帯域の平均を取ることで, 一つの鋭い相関のピークを得ることができるため, シフト長増加による誤推定を防ぐことができる.

4. 提案法の有効性の確認

提案法の有効性を確認するため, シフト長を 1 から 32 まで変化させながら方向推定実験を行った. 実験条件は Table.1 に示す通りである. 録音は反射音の強い壁際で行った. ただし, フレーム長 32, シフト長 32 の時は, 誤推定数が大幅に上昇したため, フレーム長 512, シフト長 1 の時はメモリ不足で実験できなかったため, 除外している.

Fig.12 に各フレーム長, シフト長毎の処理時間を, 処理対象とした音声信号長を基準に正規化して表した. また, それぞれの場合の誤推定数を Table.2 に示した. Fig.12 から, 処理時間はシフト長の増加(=フーリエ変換回数の減少)にほぼ反比例して短くなっていることが分かる. このことから, 折り返し補正処理はシフト長が大きいくほど処理量が増加するが, 全体の処理時間への影響が小さいことがわかる.

Table.2 はシフト長, フレーム長毎の誤推定数を示した. 結果から, シフト長が増加するほど誤推定数が増加する傾向にあり, シフト長がフレーム長の $1/4$ 以上になると, 誤推定数が大幅に増加することが分かった.

Table.2 において、シフト長がフレーム長の 1/4 以上となつて、誤推定が増加した部分を灰色で示す。

Fig.12, Table.2 において、従来使用されていたフレーム長 32, シフト長 1 の場合での結果を破線の○印で示している。従来のパラメータでは、処理対象の音声 1 秒当たりの処理時間は約 1.91 秒, 誤推定数は 6 個であった。提案法によってシフト長を変更した際、フレーム長 128, シフト長 16(実線の○印)のとき, 1 秒当たりの処理時間約 0.47 秒となり, 誤推定数は従来と同じ 6 個であることが確認された。よつて, 推定性能を維持しつつ, 従来の SBPHP-CC 法と比べて 1/4 の時間で処理が可能となつた。

また, フレーム長 512, シフト長 16 の場合は, 誤推定数が従来法よりも少ない 4 個となることが確認された。従つて, 処理時間を考えなければ, より誤推定数を減らすことも可能である。

Table.3 に他の代表的音源方向推定法である MUSIC 法, CSP 法との比較結果を示した。実験条件は, Table.1 と同様とした。誤推定数は, SBPHP-CC 法が最も少なく, MUSIC 法と比べると誤推定数が 40%減少した。また処理時間は, MUSIC 法と比較すると, 約 1.27 倍程度に抑えられるようになった。

5. まとめ

本稿では, 処理時間が課題であつた SBPHP 音源方向推定法の時間短縮法を提案した。サブバンド処理のシフト長を増加し, フーリエ変換回数を減少させることで, 信号データ量を間引いて処理時間の短縮を図つたが, サブバンド信号に折り返しが発生して性能が低下することがわかり, その対策として折り返し補正を提案した。提案法の効果を実験によつて検討した結果, 推定性能を維持しつつ, これまでの SBPHP-CC 法よりも処理時間を 1/4 にすることができた。また, 結果から, シフト長がフレーム長の 1/4 未満であれば, 推定率を維持できることが確認された。

謝 辞

本研究の一部は JSPS 科研費 15H02728 の助成を受けたものです。

文 献

- [1] G.C.Carter, C.H.Knapp, "The generalized correlation method for estimation of time delay," IEEE Trans. on Acoust. Speech and Signal Proc., ASSP-24, pp.320-327, 1976.
- [2] 浅野 太, "音のアレイ信号処理—音源の定位・追跡と分離—," コロナ社, pp.115-124, 2011.
- [3] 鈴木 敬, 金田 豊, "サブバンドピークホールド処理を用いた音源方向推定法," 日本音響学会誌, 65 巻, 10 号, pp. 513-522, 2009.
- [4] 佐藤 耕平, 金田 豊, "位相情報を用いたサブバンドピークホールド処理による音源方向推定の検討," 音響学会秋季講演論文集, 3-Q-19, pp.783-784, 2009.

Table.1 実験条件

室寸法	6.3×9.1×2.8(WDH) m
残響時間	0.4 s
マイクロホンアレー	2 ch, 35 cm 間隔
窓関数	ハニング窓
フレーム長	32, 64, 128, 256, 512
シフト長	1, 2, 4, 8, 16, 32
サンプリング周波数	48000 Hz
音声データ	男性 1 名, 12 単語, 総発話回数 324 回
発話距離	1, 2, 3 m
発話角度	0, 15, 30, 45, 60°
許容誤差	±10°

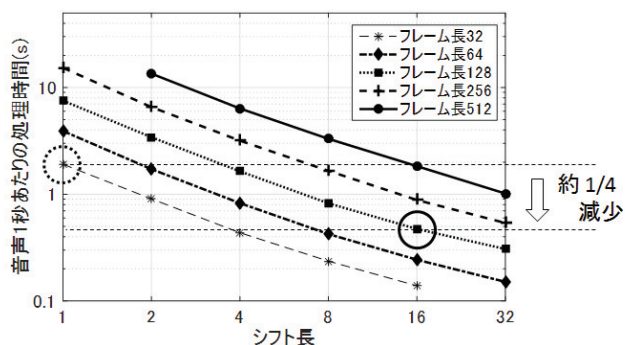


Fig.12 SBPHP-CC 法のフレーム長シフト長別の処理時間

Table.2 フレーム長シフト長別の誤推定数

フレーム長 \ シフト長	32	64	128	256	512
1	6	4	5	5	—
2	8	5	5	6	5
4	11	6	5	6	5
8	8	7	5	7	5
16	49	12	6	8	4
32	—	55	12	10	7

(総データ 324 個の誤推定数)

Table.3 他の代表的方向推定法との比較

	SBPHP-CC	MUSIC(2ch)	CSP(PHAT)
誤推定数	6	10	12
処理時間(s)	0.47	0.37	0.09

(処理時間は音声信号 1 秒あたりの処理時間)