

無矛盾性規準に基づく連続ウェーブレット変換 スペクトログラムへの位相推定法と高速化

中村 友彦^{1,a)} 亀岡 弘和^{1,2,b)}

概要：人間の聴覚システムは対数的な周波数サブバンドを持つフィルタバンクとみなせるとされており、連続ウェーブレット変換 (Continuous wavelet transform, CWT) により同様のフィルタバンクを構成することができる。したがって、人間の聴覚システムの重要な機能を模倣した信号処理手法を開発するために、CWT により得られる振幅スペクトログラム (振幅 CWT スペクトログラム) をモデル化、解析、処理するアプローチは有効であると考えられる。このアプローチを実現するためには、位相情報を含まない振幅 CWT スペクトログラムを時間領域の信号に変換する必要がある。これは、特に音響信号を出力するアプリケーションにおいて重要である。そのため、本論文では音響信号を得るために、与えられた振幅 CWT スペクトログラムから位相を推定する高速アルゴリズムを提案する。定量評価実験により、従来のアルゴリズムに比べて提案アルゴリズムは、ほぼ同程度の高品質な音響信号を約 100 倍高速に得られることが示された。

1. はじめに

連続ウェーブレット変換 (Continuous wavelet transform, CWT) は、対数周波数での時間周波数表現が可能な解析法である。人間の聴覚フィルタバンクは、特に高周波帯域に関し対数周波数上で一様な解像度を持つことが知られている [1]。このような人間の聴覚システムの重要な機能を模倣した信号処理を実現するには、CWT によって得られるスペクトログラム (CWT スペクトログラム) を解析する手法が有望である。実際、近年の研究により振幅 CWT スペクトログラム上での多重基本周波数推定などの多重音解析手法が有効であることが示している [2-5]。したがって、振幅 CWT スペクトログラム上での音源分離や音響信号加工も有効であると考えられる。しかし、音源分離や音響信号加工では、処理された振幅 CWT スペクトログラムを音響信号に変換する必要がある。そのためには与えられた振幅 CWT スペクトログラムに対し適切な位相を推定しなければならない。本論文では、この変換のために必要な位相を、与えられた振幅 CWT スペクトログラムから推定する手法を提案する。

振幅 CWT スペクトログラムからの位相推定アルゴリズムは、Iriño らによって提案された [6]。らのアルゴリズムでは、逆 CWT と CWT を行って得られた CWT スペクトログラムの振幅部分 (振幅 CWT スペクトログラム) を所望の振幅 CWT スペクトログラムと置換する処理を繰り返し行う。

ただし、CWT の計算時間は短時間 Fourier 変換 (short-time Fourier transform, STFT) に比べ非常に長いので、このアルゴリズムは長時間の計算を必要とする (5 節参照)。そのため、実際のアプリケーションへの用いるには計算量の削減が欠かせない。

そこで、我々が以前開発した STFT での位相推定法 [7] から計算量削減のヒントを得る。我々は、STFT により得られたスペクトログラム (STFT スペクトログラム) の冗長性に着眼し、振幅 STFT スペクトログラムからの高速位相推定アルゴリズムを導いた。フレーム間にオーバーラップする部分がある場合は、異なるフレーム間で音響信号の要素が共有されるため、得られた STFT スペクトログラムは冗長性を持つ。この冗長性から、実際の音響信号から得られた STFT スペクトログラムが満たす条件を導き、その条件を基に STFT スペクトログラムらしさを測る規準 (無矛盾性規準) を提案し、STFT スペクトログラムに対する位相推定問題を定式化した。この定式化に補助関数法と呼ばれる原理を適用して得られたアルゴリズムは、Griffin らによる位相推定アルゴリズム [8] と一致することも示した。この無矛盾性規準に基づく方法論は、アルゴリズムの高速化や収束性の証明にも有用であった。ここで、CWT スペクトログラムも冗長性を持つことや、Griffin らによる STFT での位相推定法 [8] と Iriño らによる CWT での位相推定法は統一的に記述できること [9] も鑑みると、この方法論は CWT にも有用だと考えられる。

我々はこの方法論に従い、CWT 振幅スペクトログラムからの位相を推定する高速アルゴリズムを提案し、その収束性も示す。まず、実際の音響信号から得られた CWT スペクトログラムが満たす条件 (無矛盾条件) を導出し、その条件から CWT スペクトログラムらしさを測る規準とし

¹ 東京大学 大学院情報理工学研究所,
東京都文京区本郷 7-3-1, 113-0033

² NTT コミュニケーション科学基礎研究所,
神奈川県厚木市森の里宮 3-1, 243-0198

^{a)} nakamura@hil.t.u-tokyo.ac.jp

^{b)} kameoka@hil.t.u-tokyo.ac.jp

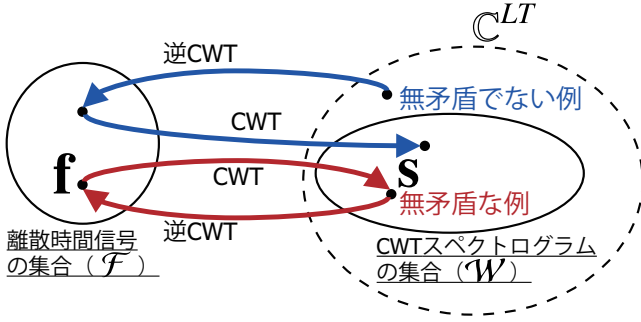


図1 連続ウェーブレット変換によるスペクトログラムに対する無矛盾条件。赤矢印と青矢印はそれぞれ無矛盾条件を満たす場合と満たさない場合を示している。

て無矛盾性規準を導き位相推定を最適化問題として定式化する。次に、補助関数法と呼ばれる原理を用いて最適化アルゴリズムを導出し、従来法との関連を明らかにする。その後、CWTの高速近似アルゴリズム [10] を用いて高速なアルゴリズムを与える。また、位相推定のためのスケールとマザーウェーブレットに関する要請も明らかにする。以下では、計算機への実装のため離散時間及び離散スケールドメインでのCWTと逆CWTを考える。

2. 連続ウェーブレット変換スペクトログラムの無矛盾性

2.1 無矛盾条件

CWTのスケールはウェーブレットの局所的な周期に対応しており、本論文では対数周波数軸上で解析周波数の間隔が一樣となるようなスケールの場合を考える。スケールと時間シフトのインデックスをそれぞれ $l \in [0, L-1]$ と $t \in [0, T-1]$ とし、 L 個の複素ベクトル $s_l := [s_{l,0}, s_{l,1}, \dots, s_{l,T-1}]^T \in \mathbb{C}^T$ を結合した複素ベクトル $s := [s_0^T, s_1^T, \dots, s_{L-1}^T]^T \in \mathbb{C}^{LT}$ を導入する。このとき、 s が直流成分のない複素離散信号 $f = [f_0, f_1, \dots, f_{T-1}]^T \in \mathcal{F}$, ($\mathcal{F} := \{f' \in \mathbb{C}^T; \sum_l f'_l = 0\}$) のCWTスペクトログラムとなるためには、CWTの基底波形を各行に並べた行列を $W \in \mathbb{C}^{LT \times T}$ として

$$s = Wf \quad (1)$$

$$W := \begin{bmatrix} W_0 \\ W_1 \\ \vdots \\ W_{L-1} \end{bmatrix}, \quad W_l := \begin{bmatrix} \psi_{l,0} & \psi_{l,1} & \cdots & \psi_{l,T-1} \\ \psi_{l,T-1} & \psi_{l,0} & \cdots & \psi_{l,T-2} \\ \vdots & & \ddots & \vdots \\ \psi_{l,1} & \psi_{l,2} & \cdots & \psi_{l,0} \end{bmatrix}. \quad (2)$$

を満たす必要がある。ここで、 $W_l \in \mathbb{C}^{T \times T}$ はCWTスペクトログラムの l 番目の成分に対応するCWTを表す行列であり、 $\psi_{l,t} := \psi(t\Delta/a_l)/a_l$ はスケール a_l でスケールされた時刻 $t\Delta$ のマザーウェーブレットを表す。 Δ はサンプリング周期、 $\psi(\cdot) \in \mathbb{C}$ は連続時間領域でのマザーウェーブレットであり、アドミッシブル条件を満たす。一方で、逆CWTは W の擬似逆行列 W^+ によって

$$f = W^+s, \quad W^+ := (W^H W)^{-1} W^H \quad (3)$$

と表現される。ここで、 H は複素共役転置を表す。また、 W^+ は擬似逆行列であるため、

$$W^+s = \underset{f \in \mathcal{F}}{\operatorname{argmin}} \|s - Wf\|_2^2 \quad (4)$$

をみだす。ここで、 $\|\cdot\|_2$ は l^2 ノルムである。

式(1) から分かるように、CWTは T 次元複素ベクトルから LT 次元の複素ベクトルへの変換であるため、CWTスペクトログラムは冗長な表現である。したがって、図1に示すようにCWTスペクトログラムの集合 W は LT 次元の複素ベクトルの集合 \mathbb{C}^{LT} の部分空間である。実際の信号のCWTスペクトログラムは逆CWTとCWTを行った後にも元のCWTスペクトログラムと同一であるので、実際の信号のCWTスペクトログラムが満たす条件である無矛盾条件は、

$$\mathbf{0}_{LT} = s - WW^+s \quad (5)$$

と定義できる。ここで、 $\mathbf{0}_{LT}$ は LT 次元の零ベクトルである。すなわち、 W は(5) を満たす複素ベクトル s の集合である。

また、 W の各行をSTFTの基底波形に置換すると、我々が以前提案したSTFTスペクトログラムに対する無矛盾条件と一致する [7]。

2.2 無矛盾性条件と位相推定問題の関連

振幅CWTスペクトログラムからの位相推定問題は、与えられた振幅CWTスペクトログラムを持つような信号を再構成するための位相を推定する問題である。ここで、振幅CWTスペクトログラムにランダムな位相を付与した複素ベクトルによって得られた信号は、図1の無矛盾でない例のように、元の振幅CWTスペクトログラムとは異なる可能性がある。したがって、再構成された信号の振幅CWTスペクトログラムが与えられた振幅CWTスペクトログラムと一致するためには、再構成された信号のCWTスペクトログラムが無矛盾条件を満たさなくてはならない。そのため、位相推定に無矛盾条件が利用できる。

2.3 スケールとマザーウェーブレットへの要請

位相推定に無矛盾条件を用いる際に必要な条件の有無を調べるために、無矛盾条件について考察する。CWTは、スケールされたマザーウェーブレットが各フィルタのインパルス応答に対応するフィルタバンクとして解釈できる。時間インデックスに対して(5)の両辺に離散Fourier変換 (discrete Fourier transform, DFT) を適用すると、

$$0 = \hat{s} - \hat{W}\hat{W}^+\hat{s}, \quad (6)$$

$$\hat{W} = \begin{bmatrix} \hat{W}_0 \\ \hat{W}_1 \\ \vdots \\ \hat{W}_{L-1} \end{bmatrix}, \quad \hat{W}_l = F_T W_l F_T^{-1}, \quad \hat{W}^+ = (\hat{W}^H \hat{W})^{-1} \hat{W}^H \quad (7)$$

が得られる。ここで、 $F_T \in \mathbb{C}^{T \times T}$ はDFT行列、 $\hat{\cdot}$ は当該変数のDFTを表す。 W_l は F_T と F_T^{-1} によって対角化され、対角化された行列 \hat{W}_l の対角要素が l 番目のスケールに対応するフィルタの周波数特性を表す。この対角要素は、角周波数インデックスを $k \in [0, T-1]$ として

$$0 = \hat{s}_{l,k} - \frac{1}{C_k} \sum_{l'} \hat{\psi}_{l,k} \hat{\psi}_{l',k}^* \hat{s}_{l',k} \quad (8)$$

と書ける．ここで， C_k は正規化係数， $*$ は当該変数の複素共役を表す．

異なるフィルターのサブバンドが重複していれば ($\exists l \neq l', \hat{\psi}_{l,k} \hat{\psi}_{l',k} \neq 0$)，CWT スペクトログラムにとって (5) は非自明な制約である．しかし重複がなければ，振幅 CWT スペクトログラムに任意の位相を付与しても (5) が満たされるため，位相推定には無矛盾条件を用いることができない．したがって，位相推定のためにはスケールとマザーウェーブレットが適切に選択される必要がある．スケールが適切に選択されていれば，Morlet や対数正規分布型ウェーブレット [3]，auditory wavelet transform で用いられるウェーブレット [6] などが，適切なマザーウェーブレットとして使用できる．以下では，異なるフィルター同士の重複があるものと仮定する．

この CWT のサブバンドの重複に対する要請は，STFT でのフレームの重複に対する要請に類似している．STFT スペクトログラムに対する無矛盾条件は，重複したフレームから得られたスペクトルが無矛盾でなくてはならないと言い換えることができる [7]．CWT の場合には，CWT スペクトログラムの無矛盾条件が，重複したサブバンドを持つフィルターの出力が無矛盾でなくてはならないと解釈できる．

3. 無矛盾条件に基づく位相推定

3.1 位相推定問題の定式化

振幅 CWT スペクトログラムからの位相推定問題では，まず振幅 CWT スペクトログラム $\mathbf{a} \in [0, \infty)^{LT}$ が与えられる．しかし，この与えられた振幅 CWT スペクトログラムはいかなる位相を付与しても無矛盾条件を満たすことができない可能性があるため，無矛盾条件を厳密に満たす位相を推定することは難しい場合もある．そこで我々は，位相推定問題を振幅 CWT スペクトログラムと推定された位相 $\phi \in [-\pi, \pi)^{LT}$ によって得られた CWT スペクトログラムができるだけ無矛盾となるような位相を推定する問題として定式化する．この定式化では $I(\phi) \geq 0$ を ϕ について最小化する問題として，

$$I(\phi) := \|\mathbf{s}(\mathbf{a}, \phi) - WW^+ \mathbf{s}(\mathbf{a}, \phi)\|_2^2, \quad (9)$$

$$\mathbf{s}(\mathbf{a}, \phi) := \mathbf{a} \odot \begin{bmatrix} e^{j\phi_0} \\ e^{j\phi_1} \\ \vdots \\ e^{j\phi_{LT-1}} \end{bmatrix} \quad (10)$$

と表す．ここで， $\mathbf{s}(\mathbf{a}, \phi)$ は推定 CWT スペクトログラムであり， \odot は要素積を表す．導入した目的関数 $I(\phi)$ は，どの程度 $\mathbf{s}(\mathbf{a}, \phi)$ が無矛盾条件を満たすかを表現しており， $I(\phi)$ が 0 に近づくほど $\mathbf{s}(\mathbf{a}, \phi)$ は近似的に無矛盾条件を満たす．

3.2 補助関数法による反復アルゴリズムの導出

$I(\phi)$ を ϕ について直接最小化することは，(9) の第 2

所望のCWTスペクトログラムをもつ
複素ベクトルの集合

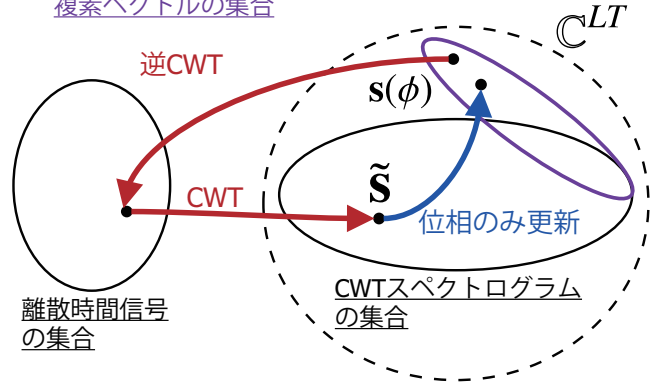


図 2 位相推定の反復アルゴリズム．赤矢印と青矢印はそれぞれ (14) と (15) の更新ステップに対応している．

項に ϕ が含まれるため難しい．そこで，補助関数法 [11] と呼ばれる原理を用いて $I(\phi)$ を最小化する．補助関数法は，コスト関数 $I(\phi)$ が与えられたときに，補助変数 $\tilde{\mathbf{s}}$ と補助関数 $I^+(\phi, \tilde{\mathbf{s}}) \geq I(\phi)$ を導入する．次に， $I^+(\phi, \tilde{\mathbf{s}})$ を ϕ と $\tilde{\mathbf{s}}$ について交互に最小化することによって，間接的に $I(\phi)$ を単調減少させる方法である．特に， $I(\phi)$ は凸関数であるため局所最適解が大域最適化と同一であり，補助関数法に従えば反復的に $I(\phi)$ を最小とする解に漸近する．

擬似逆行列は (4) を満たすので，補助変数 $\tilde{\mathbf{s}} \in \mathcal{W}$ を導入して補助関数 $I^+(\phi, \tilde{\mathbf{s}})$ は

$$I(\phi) = \min_{\tilde{\mathbf{s}} \in \mathcal{W}} \|\mathbf{s}(\mathbf{a}, \phi) - W\tilde{\mathbf{s}}\|_2^2 \quad (11)$$

$$= \min_{\tilde{\mathbf{s}} \in \mathcal{W}} \|\mathbf{s}(\mathbf{a}, \phi) - \mathbf{s}'\|_2^2 \quad (12)$$

$$\leq \|\mathbf{s}(\mathbf{a}, \phi) - \tilde{\mathbf{s}}\|_2^2 =: I^+(\phi, \tilde{\mathbf{s}}) \quad (13)$$

と定義できる．(13) の等号成立条件は $\tilde{\mathbf{s}} = WW^+ \mathbf{s}(\mathbf{a}, \phi)$ である．この等号成立条件と $I^+(\phi, \tilde{\mathbf{s}})/\partial \phi = \mathbf{0}_{LT}$ の解から，更新式は

$$\tilde{\mathbf{s}} \leftarrow WW^+ \mathbf{s}(\mathbf{a}, \phi), \quad (14)$$

$$\phi \leftarrow \angle \tilde{\mathbf{s}} \quad (15)$$

と導出できる．ここで， \angle は複素ベクトルの各要素の偏角を $[-\pi, \pi)^{LT}$ 上のベクトルとして与える．

3.3 従来法との関連

図 2 に示すように，(14) と (15) は $\mathbf{s}(\mathbf{a}, \phi)$ に逆 CWT と CWT を行うステップと， $\mathbf{s}(\mathbf{a}, \phi)$ の位相を $\tilde{\mathbf{s}}$ と置換するステップを表している．このアルゴリズムは，Iriño らによるアルゴリズム [6] と同一である．さらに，もし W の各行を STFT の基底波形と置換すれば我々が以前導出した振幅 STFT スペクトログラムへの位相推定アルゴリズム [7] が得られる．したがって，STFT と CWT での位相推定アルゴリズムとその導出は統一的に記述できる．これは，Lopes らによる考察 [9] と一致している．

4. 高速位相推定アルゴリズム

4.1 高速近似連続ウェーブレット変換

3 節で無矛盾条件に基づく反復アルゴリズムを導出した

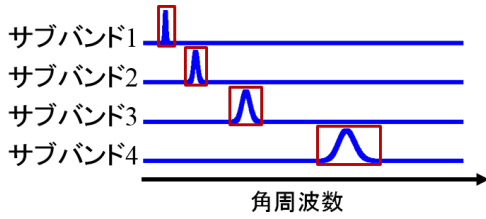


図3 スケーリングされたマザーウェーブレットの離散 Fourier 変換が、角周波数軸上で局所的にしか値を持たない例。マザーウェーブレットとして対数正規分布型ウェーブレット [3] を用いた。

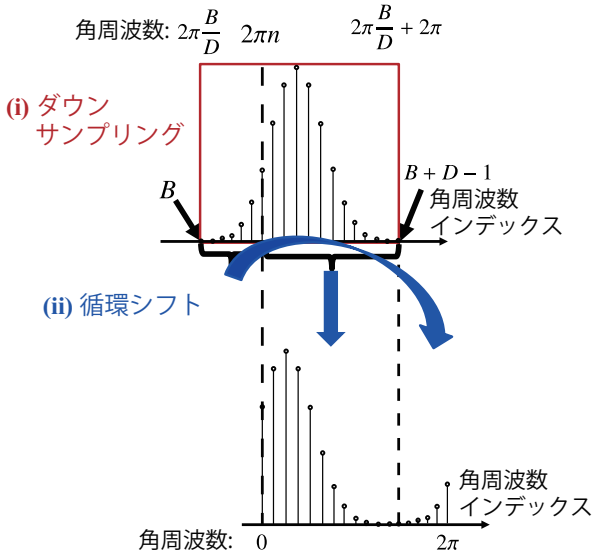


図4 高速近似連続ウェーブレット変換 [10] の概要。

が、CWT の計算量は STFT の計算量よりも非常に大きい
ため、実際のアプリケーションで使用するためには、計算
量を削減し計算速度を向上する必要がある。

計算量を削減するためには、我々が以前提案した高速
近似 CWT を利用できる [10]。2.3 節のフィルタバンク
としての CWT の解釈から、図 3 に示すように高速近似
CWT はフィルタの周波数特性が角周波数軸上で局所的
にしか値を持たないことを利用する。このような局所化
は、Morlet や対数正規分布型ウェーブレット [3] をマザー
ウェーブレットとして用いたときに現れる。 l 番目のフィル
タに対し、周波数特性 $\{\hat{\psi}_{l,k}\}_k$ が角周波数インデックス
 $k \in [B, B+D-1]$ ($0 \leq B, 0 < D \leq T$) に主要な値を持つと
き、高速近似 CWT は $k \in [B, B+D-1]$ ($0 \leq B, 0 < D \leq T$)
の各周波数領域の $\{\hat{\psi}_{l,k}\}_k$ のみを用いてダウンサンプルされ
た CWT スペクトログラムの l 番目の成分を計算する。た
だし、 D は 2 の冪乗でなくてはならない。また、高速近似
CWT で得られる CWT スペクトログラムはダウンサンプ
ルによって従来の CWT によって得られる CWT スペクト
ログラムに比べ時間解像度が低下しているが、適切に近似
精度を調整すれば実用に十分な時間解像度で高速に計算で
きる。

高速近似 CWT は、音響信号 f とスケーリングされたウェ
ーブレットの高速 Fourier 変換 (fast Fourier transform, FFT) を
計算後、各フィルターについて 2 ステップの処理を行う。(i) l

番目のフィルターの出力の FFT $\hat{W}_l f$ を $k \in [B, B+D-1]$ の成分
に帯域制限し、 T 点から D 点にダウンサンプルする。このダ
ウンサンプルされた要素の角周波数は $[2\pi B/D, 2\pi(B+D)/D]$
に分布しているので、(ii) ダウンサンプルされた要素を並
び替えて角周波数を $[0, 2\pi]$ に正規化する。信号の DFT は
周期的であるため、角周波数の正規化を行うためにはダウ
ンサンプルされた要素を巡回的にシフトさせ、零位相に対
応する部分が先頭となるようにすればよい。ここで、床関
数 $\lfloor \cdot \rfloor$ を用いて $n := \lfloor (B+D)/D \rfloor$ となる整数 n を定義する。
(ii) の操作は、図 4 に示すように $[2\pi B/D, 2\pi n]$ の要素をダ
ウンサンプルされた要素の後ろに連結すればよい。この処
理の後、並び替えられた要素を逆 FFT することによって
 D 点にダウンサンプルされた CWT スペクトログラムの l
番目の成分を得られる。ここで、 B, D は各フィルターに異
なってもよい。

ダウンサンプルによって削除された角周波数要素は、従
来の CWT で 0 とみなして計算していることに相当する。
したがって、高速近似 CWT の逆変換 (逆高速近似 CWT)
は、高速近似 CWT によって得られた CWT スペクトログ
ラムの各スケールに対応する成分に FFT を適用後、(ii) とは
逆の操作の後ダウンサンプルによって削除された角周波数
要素を 0 と詰めし、従来の逆 CWT と同様に計算すればよい。

4.2 反復アルゴリズムの高速化

(i) と (ii) のステップは、

$$K := \underbrace{\begin{bmatrix} 0_{B_0 \times (D-B_0)} & I_{B_0} \\ I_{D-B_0} & 0_{(D-B_0) \times B_0} \end{bmatrix}}_{(ii) \text{ 循環シフト}} \underbrace{\begin{bmatrix} 0_{D \times B} & I_D & 0_{D \times (T-D-B)} \end{bmatrix}}_{(i) \text{ ダウンサンプリング}} \quad (16)$$

によって表現できる。ここで、 $B_0 := B - (n-1)D$ 、 I_D は
 $D \times D$ の単位行列、 $0_{D \times B}$ は $D \times B$ の零行列である。高速近
似 CWT によって得られる CWT スペクトログラムの l 番
目の成分 $\check{s}_l \in \mathbb{C}^D$ は、行列 K を用いて

$$\check{s}_l = F_D^{-1} K \hat{W}_l F_T f \quad (17)$$

と書ける。逆高速近似 CWT も高速近似 CWT の擬似逆行
列によって表現できるため、3 節で得られた反復アルゴ
リズムの CWT と逆 CWT の代わりに高速近似 CWT と逆高
速近似 CWT を用いれば、収束性を保証したまま反復アル
ゴリズムを高速化できる。

4.3 時間及び空間計算量の削減

信号長 T が十分大きければ、CWT と高速近似 CWT の時間
計算量は、FFT の計算量に依存する。FFT は $O(T \log_2 T)$ の
計算量をもつため、CWT の計算量は $O(T \log_2 T + LT \log_2 T)$
である。一方、高速近似 CWT で l 番目のフィルターで
 $D_l \leq T$ にダウンサンプルした場合、計算量は $O(T \log_2 T +$
 $\sum_{i=0}^{l-1} D_i \log_2 D_i)$ に削減される。フィルタの周波数特性が局
所化していればいるほど、計算量の削減は大きい。

空間計算量についても、Iriano らのアルゴリズム [6] に比
べ提案法では削減されている。信号が十分長い場合には、
空間計算量の主な増大要因は CWT スペクトログラムの要

素数である。Iriano らのアルゴリズムでは LT 個の要素が存在するが、提案法では $\sum_l D_l$ 個にまで削減されている。

5. 定量評価実験

5.1 実験 1: 計算時間と音質

5.1.1 実験条件

計算時間と位相推定により得られた信号の音質を評価するため、Iriano らのアルゴリズム [6] と提案法の比較評価実験を行った。与えられる振幅 CWT スペクトログラムとして、ATR 日本語音声データベースの A セット [12] の女性話者 (faf) 115 ファイルと男性話者 (mht) の 113 ファイルの音響信号から得られた振幅 CWT スペクトログラムを用いた。FFT は 2 の冪乗の長さの信号に対して非常に高速に動作するので、使用した信号に対し信号長が最も近い 2 の冪乗となるまで零詰めを行った。位相はランダムに初期化し、両アルゴリズムとも 1000 イテレーションまで計算した。マザーウェーブレットとして、対数正規分布型ウェーブレット [3] を用いた。このウェーブレットは Fourier 変換されたドメインで定義されており、

$$\hat{\psi}(\omega) := \begin{cases} \exp\left(-\frac{(\log \omega)^2}{4\sigma^2}\right) & (\omega > 0) \\ 0 & (\omega \geq 0) \end{cases} \quad (18)$$

と表される。ここで、 ω は角周波数、 σ は標準偏差を表す。本実験では、 σ を 0.02 とし、各フィルタの中心周波数が 20 cent 毎に 27.5 から 7040 Hz となるようにスケールを設定した。提案法では、対数周波数軸で中心周波数から $\pm 3\sigma$ の領域を用いて計算を行った。Intel Xeon CPU E31245 (3.3 GHz)、32 GB RAM をもつ計算機を用いた。

音質の評価尺度として、音声の世界標準客観評価尺度である perceptual evaluation of speech quality (PESQ) [13] を用いた。PESQ は -0.5 から 4.5 までの値を取り、大きければ大きいほど音質が高いことを示している。計算時間の評価尺度としては、イテレーション毎の計算時間を用いた。

5.1.2 結果

PESQ の平均値と標準誤差は、Iriano らのアルゴリズムは 4.20 ± 0.08 、提案法は 4.1 ± 0.1 であり、得られた信号の音声品質は実用上十分高いことが示された。また、両アルゴリズムの差異は実用上大差ないものであった^{*1}。

計算量は主に信号長に依存するため、図 5 に信号長に対するイテレーション毎の平均計算時間と標準誤差を示した。提案法は Iriano らのアルゴリズムに比べ 100 倍の速度で計算できており、例えば Iriano らのアルゴリズムでは 15 s の信号に対し約 10 s/iteration で計算が行われるのに対し、提案法では 0.1 s/iteration であった。

5.2 実験 2: 近似精度と音質の関連

5.2.1 実験条件

提案法は近似を含むため、近似精度に対する再構成された信号の音質を評価した。RWC 音楽ジャンルデータベース [14] の 102 個の 16 kHz サンプリングの音楽音響信号

^{*1} サンプル音は <http://hil.t.u-tokyo.ac.jp/~nakamura/demo/fastCWT.html> で試聴できる。

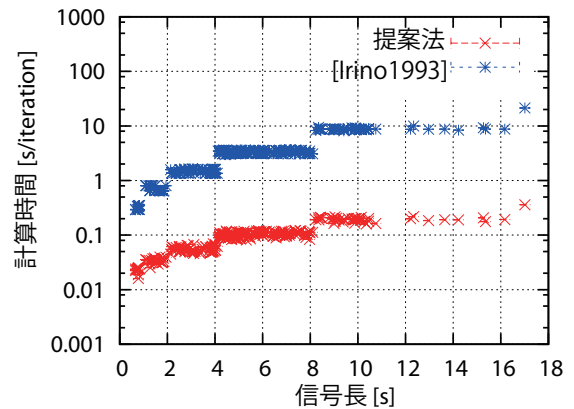


図 5 信号長に対するイテレーション毎の平均計算時間と標準誤差。

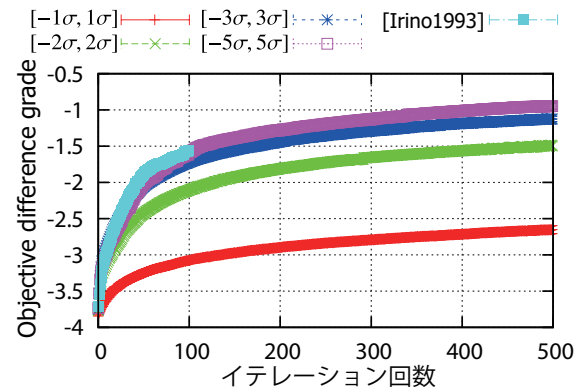


図 6 様々な近似精度 ($[-P\sigma, P\sigma]$ ($P = 1, 2, 3, 5$)) の提案法と Iriano らのアルゴリズム [6] による、Perceptual evaluation of audio quality で得られた objective difference grades のイテレーションに対する平均と標準誤差の時間発展。

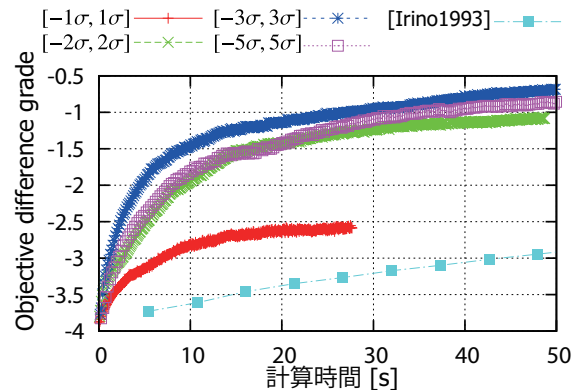


図 7 様々な近似精度 ($[-P\sigma, P\sigma]$ ($P = 1, 2, 3, 5$)) の提案法と Iriano らのアルゴリズム [6] による、Perceptual evaluation of audio quality で得られた objective difference grades の計算時間に対する平均と標準誤差の時間発展。

ファイルの 30 s から 35 s の部分の振幅 CWT スペクトログラムを、データとして用いた。マザーウェーブレットは $\sigma = 0.02$ の対数正規分布型ウェーブレットとした。提案法の近似精度はステップ (i) での帯域制限の範囲に依存する。この範囲を、中心周波数から対数周波数軸上で $\pm P\sigma$ ($P = 1, 2, 3, 5$) と変化させて計算した。提案法に関するイテレーション数は 500、Iriano らのアルゴリズムでは

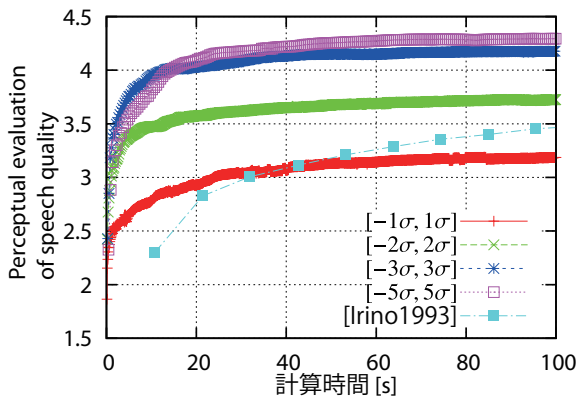


図8 様々な近似精度 $[-P\sigma, P\sigma]$ ($P = 1, 2, 3, 5$) の提案法と Irino らのアルゴリズム [6] による、Perceptual evaluation of speech quality スコアの計算時間に対する平均と標準誤差の時間発展。

100 とした。計算は、Intel Core i3-2120 CPU (3.30 GHz) と 8GB RAM を持つ計算機で行った。他の条件は、5.1.1 節と同一である。

音質評価尺度として、音質評価の世界的な標準尺度である perceptual evaluation of audio quality (PEAQ) [15] の objective differential grade (ODG) を用いた。この値は -4 から 0 まで値をとり、ODG が大きくなればなるほど音質が高いことを示している。

5.2.2 結果

図 6 にイテレーションに対する平均 ODG と標準誤差を示した。 $P = 3, 5$ での提案法の ODG は、100 イテレーション後には -2.0 より大きく、十分高音質である^{*2}。音質について、これらの結果は Irino らのアルゴリズムによる結果と大きく異ならず、 $P \geq 3$ とすれば提案法は Irino らのアルゴリズムとほぼ同等の音質をもつ音響信号を得られると言える。

計算時間の削減のためには、 P が小さければ小さいほど削減できる。図 7 は RWC-MDB-G-2001 No. 1 の音源に対する、計算時間と ODG の時間発展であり、提案法により得られた信号の ODG が従来法よりも急速に高くなったことが確認できる。同様の結果が、ATR 日本語音声データベース [12] の A セットの fafsc110 (7 s の音源) でも見られた (図 8)。我々は結果から、 $P = 3$ 程度の近似精度で提案法を用いれば、実用的な時間で高音質な信号が得られると結論づけた。

6. 結論

本論文では、収束性のある高速な位相推定アルゴリズムを提案した。位相推定問題を無矛盾条件をベースにして定式化し、補助関数法によって反復アルゴリズムを導出した。また、このアルゴリズムが Irino らによって示されていたアルゴリズム [6] と同一であることも確認した。さらに、無矛盾条件から位相推定のためにスケールとマザーウェーブレットが満たすべき条件を導出した。高速近似 CWT [10] を用いて、反復アルゴリズムの時間及び空間計算量が削減

できることを示した。定量評価実験により、提案法が Irino らのアルゴリズムに比べ約 100 倍高速であり、音楽と音声音響信号に対して実用上十分な音質で信号を得られることを示した。

今後は、和音や調、音階の変換のための音楽音響信号の加工手法のポストプロセッシングとして使用する予定である。また、実時間での処理のためにオンラインアルゴリズムの構築も行う予定である。

謝辞 本研究の一部は、JSPS 科研費 26730100 の助成を受けたものである。

参考文献

- [1] 入野 俊夫: [レクチャー講演] 聴覚フィルタの測定と定式化について, 日本音響学会聴覚研究会資料, Vol. 39, No. 6, pp. 413–418 (2009).
- [2] Schmidt, M. N. and Mørup, M.: Nonnegative matrix factor 2-D deconvolution for blind single channel source separation, *Independent Component Analysis and Blind Signal Separation*, Springer, pp. 700–707 (2006).
- [3] Kameoka, H.: Statistical Approach to Multipitch Analysis, PhD Thesis, The University of Tokyo (2007).
- [4] Muller, M., Ellis, D. P. W., Klapuri, A. and Richard, G.: Signal processing for music analysis, *IEEE J. Sel. Topics. Signal Process.*, Vol. 5, No. 6, pp. 1088–1110 (2011).
- [5] de León, J. P., Beltrán, F. and Beltrán, J. R.: A complex wavelet based fundamental frequency estimator in single-channel polyphonic signals, *Proc. Digital Audio Effects* (2013).
- [6] Irino, T. and Kawahara, H.: Signal reconstruction from modified auditory wavelet transform, *IEEE Trans. Signal Process.*, Vol. 41, No. 12, pp. 3549–3554 (1993).
- [7] Le Roux, J., Kameoka, H., Ono, N. and Sagayama, S.: Fast Signal Reconstruction from Magnitude STFT Spectrogram Based on Spectrogram Consistency, *Proc. Int. Conf. Digital Audio Effects*, pp. 397–403 (2010).
- [8] Griffin, D. and Lim, J.: Signal estimation from modified short-time Fourier transform, *IEEE Trans. Acoust., Speech, Signal Process.*, Vol. 32, No. 2, pp. 236–243 (1984).
- [9] Lopes, D. M. and White, P. R.: Signal reconstruction from the magnitude or phase of a generalised wavelet transform, *Proc. Eur. Signal Process. Conf.*, pp. 2029–2032 (2000).
- [10] 亀岡 弘和, 田原 鉄也, 西本 卓也, 嵯峨山 茂樹: 信号処理方法及び装置 (2008). 特開 2008-281898.
- [11] Ortega, J. M. and Rheinboldt, W. C.: *Iterative solution of nonlinear equations in several variables*, No. 30 (2000).
- [12] Kurematsu, A., Takeda, K., Sagisaka, Y., Katagiri, S., Kuwabara, H. and Shikano, K.: ATR Japanese Speech Database as a Tool of Speech Recognition and Synthesis, *Speech Commun.*, Vol. 9, No. 4, pp. 357–363 (1990).
- [13] ITU-T: Recommendation P.862, Perceptual Evaluation of Speech Quality (PESQ): An Objective Method for End-To-End Speech Quality Assessment of Narrow-Band Telephone Networks and Speech Codecs (2001).
- [14] Goto, M.: Development of the RWC Music Database, *Proc. Int. Congress Acoust.*, pp. 1–553–556 (2004).
- [15] ITU-T: Recommendation BS.1387-1, Perceptual Evaluation of Audio Quality (PEAQ): Method for Objective measurements of perceived audio quality (2001).

^{*2} c.f.) 使用した音楽音響信号を MPEG-3 に 160 kbps でエンコードした場合は、平均 ODG と標準誤差は -3.68 ± 0.03 であった。