

連続ウェーブレット変換の高速近似アルゴリズムに基づく 振幅スケログラムへの無矛盾位相付加法の検討*

☆中村友彦（東大院・情報理工），亀岡弘和（東大院・情報理工/NTT CS 研）

1 背景

音源分離や加工のための音響信号処理手法において、短時間フーリエ変換（short-time Fourier transform, STFT）の絶対値として得られる振幅スペクトログラム、連続ウェーブレット変換（continuous wavelet transform, CWT）の絶対値として得られる振幅スケログラムの領域で処理を行うアプローチが有効である [1-5]。処理後の振幅スペクトログラムや振幅スケログラムから音響信号を生成するためには、適切な位相を付加する（位相付加）必要がある。

振幅スペクトログラムへの位相付加は、Griffinら [2] によって提案され、LeRouxら [3] により高速化が行われた。振幅スケログラムに対しては、振幅スケログラムを保存しつつ CWT と逆 CWT を交互に繰り返す位相付加（replace and iterative wavelet transform-modified wavelet transform, RIWT-MWT）が提案されている [4]。しかし、CWT は STFT に比べ計算コストが高い。

そこで我々は、LeRoux らの方法に倣い、実際の音響信号から得られるスケログラムが満たす条件（無矛盾条件）を導出し、RIWT-MWT を高速化したアルゴリズムを提案する。CWT と逆 CWT は元来連続関数に対する変換であるが、計算機への実装のため離散化した CWT と逆 CWT を考える。本稿では、単に CWT、逆 CWT と書いた場合、離散化した CWT と逆 CWT を表すこととする。

2 スケログラムの無矛盾条件

2.1 位相付加法と無矛盾性の関連

実際の信号を CWT して得られるスケログラムは、逆 CWT により信号に変換した後 CWT を用いてスケログラムに変換する操作 \mathcal{W} を行っても、不変である。このようなスケログラムを「無矛盾なスケログラム」と呼ぶことにする。位相付加法の目的は、与えられた振幅スケログラムを持つような信号を構成するための位相を推定することである。したがって、与えられた振幅スケログラムをもち、無矛盾なスケログラムになるような位相を求めればよい。本節では、操作 \mathcal{W} によって無矛盾なスケログラムが満たすべき条件を導出する。

2.2 無矛盾条件の導出

ウェーブレット変換においてスケールパラメータは基底波形（ウェーブレット）の局所的な周期に対応する。スケールパラメータを対数周波数軸上で等間隔で離散化する場合を考え、 $l \in [0, L-1], t \in [0, T-1]$ をスケール値と時間シフト値のインデックスとする。スケール $a_l > 0$ に対するスケログラムの要素を $\mathbf{S}_l = [S_{l,0}, S_{l,1}, \dots, S_{l,T-1}]^T \in \mathbb{C}^T$ として、スケログラム $\mathbf{S} = [\mathbf{S}_0^T, \mathbf{S}_1^T, \dots, \mathbf{S}_{L-1}^T]^T \in \mathbb{C}^{LT}$ は、CWT により離散信号 $\mathbf{f} = [f_1, f_2, \dots, f_{T-1}]^T \in \mathbb{C}^T$ から

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

により得られる。ここで、連続領域でのマザーウェーブレットを $\psi(\cdot) \in \mathbb{C}$ 、 Δ をサンプリング周期として、 $\psi_{l,t} := \psi(t\Delta/a_l)/a_l$ と定義した。 $\psi_{l,t}$ は、スケール a_l でスケールされたマザーウェーブレット（ベビーウェーブレット）を離散化した値を表す。一方、逆

CWT は $W = [W_0^T, W_1^T, \dots, W_{L-1}^T]^T \in \mathbb{C}^{LT \times T}$ を用いて、

$$\mathbf{f} = (W^H W)^{-1} W^H \mathbf{S} \quad (2)$$

と W の擬似逆行列を用いて書ける。 H は複素共役転置を表す。したがって、無矛盾なスケログラムは定義より、

$$\mathbf{0} = \mathbf{S} - \mathcal{W}[\mathbf{S}] = \mathbf{S} - W(W^H W)^{-1} W^H \mathbf{S} \quad (3)$$

を満たす。ここで、 $\mathbf{0}$ は LT 次元の零ベクトルである。すなわち、(3) が無矛盾なスケログラムが満たす無矛盾条件である。 W のかわりに STFT における各基底波形を各行に並べた行列を用いれば、 \mathcal{W} は逆 STFT を行った後 STFT を行う操作に対応する [3]。

2.3 マザーウェーブレットと無矛盾条件の関係

無矛盾条件を陽に書き下すために、(3) の両辺を t に関して離散フーリエ変換（DFT）すると、 $\forall k \in [0, T-1], \forall l$ に対し、

$$0 = \hat{S}_{l,k} - \frac{1}{C_k} \sum_{l'} \hat{\psi}_{l,k} \hat{\psi}_{l',k}^* \hat{S}_{l',k}, \quad (4)$$

が導出できる。ここで、 C_k は k 番目の周波数に対応する正規化項、 $\hat{\cdot}$ は当該変数の DFT 後の値を表し、 k は周波数のインデックス、 $*$ は複素共役を表す。 $\hat{\psi}_{l,k}$ は、スケール a_l に対応するベビーウェーブレットを DFT したときの k 番目の周波数成分である。

異なるベビーウェーブレット同士が周波数軸上で完全に分離している（ $\forall l \neq l', \hat{\psi}_{l,k} \hat{\psi}_{l',k} = 0$ ）ときは、任意の \mathbb{C}^{LT} の元 \mathbf{S} について (4) は成立する。一方で、異なるベビーウェーブレット同士が周波数軸上でオーバーラップする部分があれば（ $\exists l \neq l', \hat{\psi}_{l,k} \hat{\psi}_{l',k} \neq 0$ ）、任意の \mathbb{C}^{LT} の元 \mathbf{S} に対して (4) は必ずしも成り立たない。これは、ベビーウェーブレット同士に周波数軸上でオーバーラップ部分があれば、異なるスケールの無矛盾なスケログラムの要素間に制約があることを示している。したがって、オーバーラップがある場合にはこの制約を手がかりとして、位相付加法に用いることができる。以下では、異なるベビーウェーブレットに周波数軸上でオーバーラップが存在する場合に限り議論する。

3 無矛盾条件を用いた振幅スケログラムからの位相付加法

振幅スケログラム $\mathbf{A} \in [0, \infty)^{LT}$ が与えられたとする。このとき、無矛盾条件を満たすような位相 $\phi \in [-\pi, \pi]^{LT}$ を求めたい。しかし、任意の \mathbf{A} に対し、無矛盾条件を満たす位相が存在するとは限らない。そこで、無矛盾条件を緩和した

$$I(\phi) := \|\mathbf{A} \odot e^{i\phi} - \mathcal{W}[\mathbf{A} \odot e^{i\phi}]\|_2^2 \quad (5)$$

を最小化するような ϕ を考える。ここで、 $\|\cdot\|_2$ は L2 ノルム、 \odot は要素積を表す。

RIWT-MWT は、振幅スケログラムが所与のベクトル \mathbf{A} となるような信号を得るために提案されたアルゴリズムである。このアルゴリズムは、振幅スケログラムを固定しながら逆 CWT と CWT を行って得られたスケログラムの位相を反復的に更新する。

実は、RIWT-MWT は、補助関数法 [6] を用いた $I(\phi)$ の最適化アルゴリズムと同一であることが示せる。補助関数法は、補助変数を導入し $I(\phi)$ の上界（補助関数）を導出した後、補助関数を補助変数と ϕ につい

* Consistent Phase Estimation from Amplitude Scalogram using Fast Approximate Continuous Wavelet Transform by Tomohiko Nakamura (The University of Tokyo) and Hirokazu Kameoka (The University of Tokyo/NTT CS Lab.)

て交互に最小化することにより、反復的に $I(\phi)$ を減少させる。逆 CWT は (2) の通り CWT を表す行列 W の擬似逆行列として表現できるため、離散信号の集合 \mathcal{F} と無矛盾なスケログラムの集合 $\text{Image}(W)$ を用いて

$$I(\phi) = \min_{\tilde{f} \in \mathcal{F}} \|\mathbf{A} \odot e^{j\phi} - W\tilde{f}\|_2^2 = \min_{\tilde{S} \in \text{Image}(W)} \|\mathbf{A} \odot e^{j\phi} - \tilde{S}\|_2^2 \quad (6)$$

と書ける。したがって、補助変数 $\tilde{S} \in \text{Image}(W)$ を用いて (7) の右辺の通り補助関数 $I^+(\phi, \tilde{S})$ を定義できる。

$$I(\phi) \leq \|\mathbf{A} \odot e^{j\phi} - \tilde{S}\|_2^2 =: I^+(\phi, \tilde{S}). \quad (7)$$

後は、 $I^+(\phi, \tilde{S})$ を ϕ, \tilde{S} について交互に反復的に更新すれば、 $I(\phi)$ を単調減少させることができる。更新則は、 $\partial I^+(\phi, \tilde{S}) / \partial \phi = \mathbf{0}$ の解と (7) の等号成立条件から、

$$\phi \leftarrow \arg \tilde{S}, \quad \tilde{S} \leftarrow W[\mathbf{A} \odot e^{j\phi}] \quad (8)$$

と導ける。ここで、 \arg は各要素の偏角を $[-\pi, \pi]$ の範囲で表す演算子である。(8) は RIWT-MWT の更新則にほかならない。

ここで、CWT を STFT と置き換えると、Griffin らの振幅スペクトログラムに対する位相付加法 [2] と一致することが LeRoux ら [3] によって示されている。したがって、振幅スペクトログラムと振幅スケログラムに対する位相付加法は、補助関数法を用いて統一的に記述できる。

4 連続ウェーブレット変換の高速近似アルゴリズムによる高速化

4.1 連続ウェーブレット変換の高速近似アルゴリズム

5 節で示すとおり、RIWT-MWT は従来からよく使われる高速フーリエ変換 (FFT) による CWT と逆 CWT の高速化を用いても、各イテレーションでの計算コストが高く、実用的な計算速度にする必要がある。本研究では、亀岡らによって提案された CWT の高速近似アルゴリズム [7] (高速 CWT) を用いて、CWT と逆 CWT を高速化し \tilde{S} の更新を高速化する。

高速 CWT は、フーリエ変換後のベビーウェーブレットが周波数軸上で局在することを利用する。これを満たすマザーウェーブレットは、Morlet や対数正規分布型ウェーブレット [5] などである。 $\hat{\psi}_{l,k}$ が $k \in [B, B+D-1]$, ($0 \leq B, 0 < D \leq T$) にしか主要な値をもたないとすると、 $[B, B+D-1]$ の範囲の $\hat{\psi}_{l,k}$ の値のみを保存しつつ CWT が行えれば、十分な精度でスケログラムが近似的に得られるはずである。

$T \times T$ の離散フーリエ変換行列 F_T を用いて $\hat{W}_l := F_T W_l F_T^{-1}$ を導入すると、主要な区間 $[B, B+D-1]$ を $\hat{W}_l \hat{f}$ から切り出し、 T 点から D 点へのダウンサンプルする。ここで、ダウンサンプル後に得られた区間の位相は $[2\pi B/D, 2\pi(B+D)/D + 2\pi]$ である。ダウンサンプルされて得られた区間の要素を逆 FFT を用いてスケログラムに変換するため、位相を $[0, 2\pi]$ の区間に整える必要がある。そこで、FFT で変換された信号は周波数軸上で 2π の周期をもつことを利用すると、位相を $[0, 2\pi]$ となるよう要素を循環シフトすればよいことがわかる。

高速 CWT では、ダウンサンプルで除外された区間は 0 として計算することに相当する。したがって、高速 CWT に対応する逆 CWT (高速逆 CWT) は、スケログラムを FFT したのちダウンサンプルで失われた周波数の要素を 0 詰めで復元すれば、FFT による逆 CWT の高速化と同様に計算できる。

4.2 計算量の比較

位相付加法は逆 CWT と CWT を繰り返すため、各イテレーションでの逆 CWT と CWT の計算コストが重要である。計算量は主に FFT と逆 FFT に依存し、RIWT-MWT は、FFT による高速化を行ってもイテレーション毎に $O((L+1)T \ln T)$ である。一方、高速 CWT と高

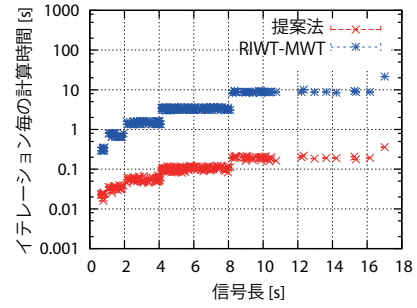


Fig. 1 信号長に際するイテレーション毎の平均計算時間と標準偏差。

速逆 CWT を用いた提案法では、すべてのスケールで D 点にダウンサンプルすれば、 $O(LD \ln D + T \ln T)$ まで計算量が削減できる。ベビーウェーブレットが周波数軸上で十分局在化していれば $D \ll T$ であり、十分な高速化が見込める。また、提案法では D をスケール毎に別々に設定できるため、低周波部分ではさらなる計算量削減が期待できる。

5 計算時間と音質に関する評価実験

計算時間と音質について定量的に、RIWT-MWT と 4 節の提案法を評価した。振幅スケログラムとして、ATR の A セット [8] の男性と女性音声ファイル 230 個から得られた振幅スケログラムを用いた。この振幅スケログラムを計算する際、高速フーリエ変換は 2 のべき乗の長さの信号に対してより高速に実行できるため、信号長を 2 のべき乗になるまで 0 詰めした。位相の初期値はランダムに設定し、1000 反復で終了とした。スケールは、解析する周波数が 27.5 Hz から 1/5 半音毎に 7040 Hz となるように設定した。マザーウェーブレットは対数正規分布型ウェーブレットを用い、分散 σ は 0.2 とした。提案法では、対数周波数軸上で中心周波数から $\pm 3\sigma$ の部分を $[B, B+D-1]$ として計算した。音質の評価指標として、 -0.5 から 4.5 の値をとる音声品質客観評価法 Perceptual evaluation of speech quality (PESQ) [9] を用いた。PESQ の値が大きいほど高音質である。

計算時間に関する結果は図 1 に示す通り、計算時間は約 1/100 まで減少した。また、15 s の音響信号に対しても各イテレーション約 0.2 s で計算でき、数十秒のデータを扱うには十分実用的である。また、1000 イテレーション後の PESQ は厳密法が 4.20 ± 0.08 、高速法が 4.1 ± 0.1 とどちらも高音質であり、提案法は実用上 RIWT-MWT と同程度の音質である。

6 まとめ

本稿では、スケログラムに対する無矛盾条件を導入し、無矛盾条件に基づく位相付加法と RIWT-MWT の関係を示した。また、亀岡らにより提案された CWT の高速近似アルゴリズムを用いて位相付加法の計算量を削減し、約 100 倍の高速に実用的な音質の音響信号が得られることを定量評価実験により示した。今後は、実際に加工したスケログラムに対して適用し、加工への位相付加法の有効性を検証することが課題である。

参考文献

- [1] M. Müller et al., *IEEE J. Sel. Topics Signal Process.*, 5(6), pp. 1088–1110, 2011.
- [2] D. Griffin et al., *IEEE Trans. ASLP*, 32(2), pp. 236–243, 1984.
- [3] J. Le Roux et al., In *Proc. of DAFx*, pp. 397–403, 2010.
- [4] T. Irino et al., *IEEE Trans. SP*, 41(12), pp. 3549–3554, 1993.
- [5] H. Kameoka, *Statistical Approach to Multipitch Analysis*, Ph.D. thesis, The University of Tokyo, Mar., 2007.
- [6] 亀岡 弘和 他, 情報研報, 2006-MUS-66-13, pp. 139–144, 8, 2006.
- [7] 亀岡 弘和 他, “信号処理方法及び装置,” 特開 2008-281898, 11, 2008.
- [8] A. Kurematsu et al., *Speech Commun.*, 27, pp. 187–207, 1990.
- [9] ITU-T Recomm. P.862.2.