Vol.2016-MUS-111 No.47
2016/5/22

高速近似連続ウェーブレット変換による振幅スペクトログラムからの逐次位相推定法

中村 友彦[1,a)]　亀岡 弘和[1,2,b)]

概要：我々は以前，高速近似連続ウェーブレット変換で得られた振幅スペクトログラムから高速に位相を推定するアルゴリズムを提案した．しかし，このアルゴリズムでは信号全体の離散 Fourier 変換を必要とするため，計算機のメモリ量により計算可能な信号長が限られたり，実時間動作を要求するアプリケーションに対し原理的に適用できなかった．本報告ではこのアルゴリズムを拡張し，固定長のセグメント毎に分割された信号の振幅スペクトログラムから逐次位相を推定するアルゴリズムを提案する．提案アルゴリズムでは，セグメント間の重複部分の信号が無矛盾であることを考慮しつつ各セグメントで処理を行う．そのため，空間計算量が信号長に依存しない．実験により，提案法が各セグメント対して独立に我々が以前提案した位相推定アルゴリズムを適用したアルゴリズムよりも高音質な信号を再構成できることを確認した．

1. はじめに

連続ウェーブレット変換（continuous wavelet transform; CWT）は定 Q フィルタバンクと本質的に同一であり，時間信号を対数周波数スケールで一様な解像度をもつ時間周波数表現に変換する．この周波数解像度は，平均律での音高の基本周波数の分布だけでなく，特に高周波数帯域で人間の聴覚フィルタバンクとも合致する．これらの性質から，聴覚システムに着想を得た音響信号処理手法の開発には，線形周波数スケールで一様な解像度の時間周波数表現を与える短時間 Fourier 変換（short-time Fourier transform; STFT）よりも，CWT で得られたスペクトログラム領域での処理が望ましいと考えられる．実際に，複素スペクトログラムを用いたマルチチャネル音源分離 [1]，振幅スペクトログラムを用いた多重基本周波数推定 [2], [3], [4] や歌声分離 [5] などで CWT の有用性が確認されている．音源分離や音響信号加工などの時間信号を得ることが目的のアプリケーションにおいては，分離や加工された振幅スペクトログラムを音響信号に変換する必要がある．しかし，振幅スペクトログラムは位相情報を持たないため，本報告では与えられた振幅スペクトログラムに対し適切な位相を求める問題に取り組む．

[1] 東京大学大学院情報理工学系研究科
東京都文京区本郷 7-3-1, 113–0033
[2] 日本電信電話株式会社 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若宮 3-1, 243–0198
[a)] tomohiko.nakamura.jp@ieee.org
[b)] kameoka.hirokazu@lab.ntt.co.jp

CWT で得られた振幅スペクトログラムからの位相推定アルゴリズムは，最初に入野らにより提案された [6]．このアルゴリズムでは以下の 2 つのステップを交互に繰り返す．1 つ目は与えられた振幅スペクトログラムに位相の推定値を割り当てた複素ベクトルに対し逆 CWT の後 CWT を適用するステップであり，2 つ目は適用後の振幅スペクトログラムのみを所望の振幅スペクトログラムに置換するステップである．しかし，CWT は計算量が高いため入野らのアルゴリズムは多大な計算時間を必要する．そのため，実際のアプリケーションに適用するには計算量の削減が必要である．CWT と逆 CWT の様々な高速計算法は近年相次いで提案されており [7], [8], [9]，CWT の代わりにこれらの高速計算法を利用することで計算法を削減できるはずである．しかし，そのようなアルゴリズムでは目的関数の収束性が保証されるかどうかが不明であった．

そこで，我々は補助関数法 [10] と呼ばれる最適化原理に基づき導出したアルゴリズムが入野らのアルゴリズムと一致することを示し，冗長な線形変換であれば目的関数が各反復で非増加であることが保証されることを示した [11]．この結果に基づき，CWT の代わりに高速近似 CWT[7] を用いて，目的関数の収束性を保証しつつ高速な位相推定アルゴリズム（オフライン位相推定アルゴリズム）を提案した．

高速近似 CWT を含め多くの CWT 高速化アルゴリズムでは信号全体の高速 Fourier 変換（FFT）を用いており，信号長が長いほど位相推定アルゴリズムの空間計算量は増加する．そのため，メモリに制約のある環境下では処理可能な音響信号が限られる．また，原理上実時間動作を要求す

るアプリケーションにそのまま適用することはできない．

本報告では，[11] で提案した位相推定アルゴリズムを拡張し，信号長によらない空間計算量をもち音響信号を逐次処理できる位相推定アルゴリズムを提案する．提案アルゴリズムでは，音響信号を重複のある固定長のセグメントに分割し，各セグメントに高速近似 CWT を適用して得られたスペクトログラムを対象とする．オフライン位相推定アルゴリズムの解は一意でないため，重複部分のあるセグメントで独立にオフライン位相推定アルゴリズムを適用するとセグメント間で整合しない位相が推定されてしまい，得られる信号の音質が劣化する原因となる．そこで，重複区間の位相の整合性を考慮した更新則を提案し，提案アルゴリズムを有効性を実験により検証する．

2. 高速近似ウェーブレット変換を用いた位相推定アルゴリズム

2.1 高速近似 CWT

高速近似 CWT は以下の考えに基づき提案された．CWT は，スケーリングされたアナライジング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解釈できる．各サブバンドフィルタの周波数特性の主要部分が局所的に分布していれば，主要な値が存在する周波数区間のみを計算に用いることで CWT を計算量を削減できるはずである．周波数特性に関するこの性質は，Morlet や対数正規分布型ウェーブレット [3] など多くのアナライジングウェーブレットが満たす．

T 次元の時間信号を $\boldsymbol{f}$，T 次元の DFT 行列を F_T とする．高速近似 CWT では，時間信号全体の FFT，$F_T\boldsymbol{f}$ を求めた後に各サブバンドフィルタの主要な部分が存在する領域 $k \in [B, B+D-1]$ に帯域制限を行う．$k = 0, \cdots, T-1$ は角周波数インデックスである．ここで，ある対数周波数インデックス $\omega(= 0, \cdots, \Omega-1)$ に対応するサブバンドフィルタに着目すると，帯域制限を表す行列はの $D \times B$ 次元の零行列 $0_{D\times B}$ と $D \times D$ 次元の単位行列 I_D を用いて，$L := [0_{D\times B}, I_D, 0_{D\times(T-D-B)}]$ と書ける．帯域制限された時間信号の FFT に対して，帯域制限された周波数帯域の当該フィルタの周波数特性 $\boldsymbol{\psi}_\omega \in \mathbb{C}^D$ を乗算し，$\mathrm{diag}(\boldsymbol{\psi}_\omega)LF_T\boldsymbol{f}$ を得る．ここで，$\mathrm{diag}(\boldsymbol{\psi}_\omega)$ は $\boldsymbol{\psi}_\omega$ を対角成分に並べた対角行列を表す．帯域制限により正規化角周波数が $[2\pi B/D, 2\pi B/D + 2\pi]$ に分布するため，$\mathrm{diag}(\boldsymbol{\psi}_\omega)LF_T\boldsymbol{f}$ を巡回的にシフトさせ，先頭の成分の位相が 0 となるようにする．この操作は以下の行列 C で表せる．

$$C := \begin{bmatrix} 0_{(B-(p-1)D)\times(pD-B)} & I_{B-(p-1)D} \\ I_{(pD-B)} & 0_{(pD-B)\times(B-(p-1)D)} \end{bmatrix} \quad (1)$$

ここで，p は $pD \le B+D < (p+1)D$ となる最大の整数である．この巡回シフトを行ったものに対し D 次元の逆 FFT を行うことで，当該サブバンドでのスペクトログラムが得られる．D, B はサブバンドによって異なってもよいが，簡単のため以下では全サブバンドで D, B は同一とする．

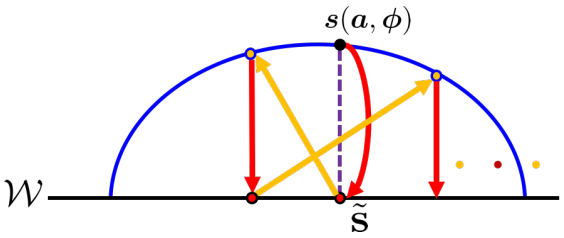


図 1 オフライン位相推定アルゴリズム [11] の概略図．青の曲線は，振幅が $\boldsymbol{a}$ である複素ベクトルの集合を表す．

これらをまとめると，対数周波数インデックス ω に対応する高速近似 CWT は $W_\omega = F_D^{\mathsf{H}} C\,\mathrm{diag}(\boldsymbol{\psi}_\omega)L$，高速近似 CWT は $W := [W_0^\top, \cdots, W_{\Omega-1}]^\top$ と表せ，複素スペクトログラムは $W\boldsymbol{f}$ で表される複素ベクトルとして得られる．ここで，$^{\mathsf{H}}$ は当該変数のエルミート共役を表す．高速近似 CWT は線形変換なので，その逆変換（逆高速近似 CWT）は W の擬似逆行列 W^+ で表される．

2.2 オフライン位相推定アルゴリズム

$T < D\Omega$ とすれば高速近似 CWT で得られるスペクトログラムは時間信号の冗長な表現であり，この場合にはスペクトログラムの集合 $\mathcal{W}$ は $\mathbb{C}^{D\Omega}$ の部分空間となる．逆高速近似 CWT の後高速近似 CWT を適用する操作 WW^+ は $\mathcal{W}$ への直行射影であるため，$D\Omega$ 次元の複素ベクトルとそれに対し WW^+ を適用して得られた複素ベクトルの距離が小さければ小さいほど，当該複素ベクトルは「複素スペクトログラムらしい」とみなせる．したがって，与えられた振幅スペクトログラムに対して，この距離を最小化する問題として位相推定問題は定式化できる．

振幅スペクトログラム $\boldsymbol{a} \in \mathbb{R}_{\ge 0}^{\Omega D}$ が与えられたとする．このとき，位相推定問題は位相の推定値 $\boldsymbol{\phi} \in [-\pi, \pi)^{\Omega D}$ を用いて以下のように定式化できる．

$$\min_{\boldsymbol{\phi}\in[-\pi,\pi)^{\Omega D}} \mathcal{I}(\boldsymbol{\phi}), \quad \mathcal{I}(\boldsymbol{\phi}) = \boldsymbol{s}^{\mathsf{H}}(\boldsymbol{a},\boldsymbol{\phi})(I_{\Omega T} - WW^+)\boldsymbol{s}(\boldsymbol{a},\boldsymbol{\phi}). \quad (2)$$

ここで，$\boldsymbol{s}(\boldsymbol{a},\boldsymbol{\phi})$ は振幅 $\boldsymbol{a}$，位相 $\boldsymbol{\phi}$ の複素ベクトルを表す．

$\boldsymbol{s}(\boldsymbol{a},\boldsymbol{\phi})$ 自体をパラメータとみなし $\boldsymbol{s} = [s_{0,0}, s_{0,1}, \cdots, s_{0,D}, s_{1,0}, \cdots, s_{\Omega-1,D}]^\top$ と置くと，各時間周波数成分 $s_{\omega,t}$ に関し $|s_{\omega,t}|^2 = a_{\omega,t}^2$ という 2 次制約を持つ 2 次計画問題に書き換えられる．これに対し，Lagrange 未定乗数法により直接解を求めることは難しい．

そこで，我々は [11] で補助関数法と呼ばれる最適化原理に基づきオフライン位相推定アルゴリズムを提案した．このアルゴリズムの概略図を図 1 に示す．導出の詳細は省くが，下記の 2 つの更新式を交互に更新することにより，各反復で $\mathcal{I}(\boldsymbol{\phi})$ が非増加となる．

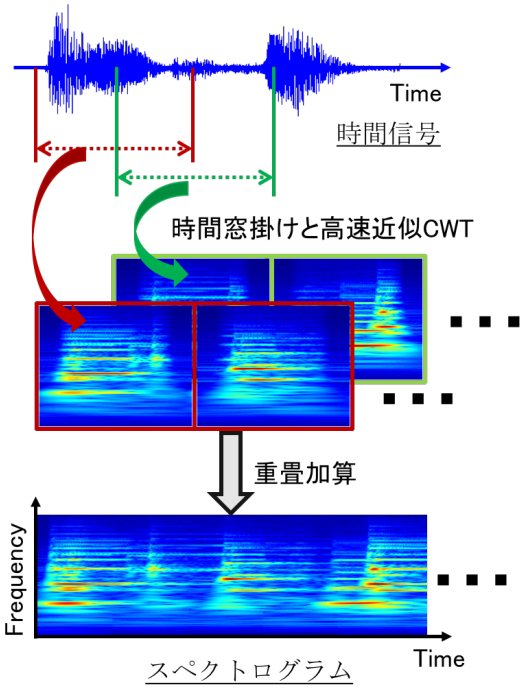


図 2 逐次高速近似 CWT の処理フロー.

$$\tilde{\boldsymbol{s}} \leftarrow WW^{+}\boldsymbol{s}(\boldsymbol{a},\boldsymbol{\phi}) \tag{3}$$

$$\boldsymbol{\phi} \leftarrow \angle\tilde{\boldsymbol{s}} \tag{4}$$

ここで，$\angle\tilde{\boldsymbol{s}}$ は複素ベクトル $\tilde{\boldsymbol{s}}$ の各要素の偏角を $[-\pi,\pi)^{\Omega T}$ のベクトルとして返す演算子である．(3) 式は，現在のスペクトログラム推定値に対し逆高速近似 CWT を行ったのちに高速近似 CWT を適用することを表す（図 1 の赤矢印）．(4) 式は，(3) 式の操作で得られたスペクトログラム $\tilde{\boldsymbol{s}}$ の位相を新たな位相の推定値とすることに対応する（図 1 の橙矢印）．

図 1 からも分かる通り，位相推定問題の解は一意ではない．例えば，位相全体に定数を加えても振幅スペクトログラムは変化しないため目的関数値は変わらない．

3. 位相推定アルゴリズムの実時間化

3.1 逐次高速近似 CWT

前節までの処理では高速近似 CWT を計算するために時間信号全体の FFT が必要であるため，音響信号が逐次入力される場合には適用できない．このように場合には逐次処理が可能な CWT の高速計算法 [8] を用いることができ，本節では [8] と同一の方法で高速近似 CWT を拡張する．この手法を逐次高速近似 CWT と呼ぶ．

逐次高速近似 CWT は以下の 3 ステップの処理からなる（図 2）．

(i) セグメント長を $2N$，セグメントのシフトを N として，入力音響信号の当該時刻でのセグメントを切り出す．

(ii) 当該セグメントの信号成分に分析窓 $\boldsymbol{h} = [h_0,\cdots,h_{2N-1}]^{\top}$ を掛け，高速近似 CWT を適用し当該セグメントの信号の複素スペクトログラムを得る．

(iii) 得られた各セグメントでのスペクトログラム同士を対応する時刻で加算することにより所望のスペクトログラムが得られる．

逆変換では，ステップ（ii）で得られた当該セグメントのスペクトログラムと，当該セグメントと直前のセグメントの重複部分の時間信号があればよい．当該セグメントのスペクトログラムに逆高速近似 CWT を適用し得られた時間信号に合成窓 $\boldsymbol{v} = [v_0,\cdots,v_{2N-1}]^{\top}$ を掛け，適切に加算することで入力された時間信号が得られる．ここで，合成窓は時刻インデックス $n = 0,\cdots,N-1$ において $h_n v_n + h_{n+N} v_{n+N} = 1$ を満たす．そのため，空間計算量の主な増大要因である各セグメントのスペクトログラムの要素数は信号長に依存せず，信号長に比べ小さな N を用いれば空間計算量が削減可能である．

ここで，h_n として

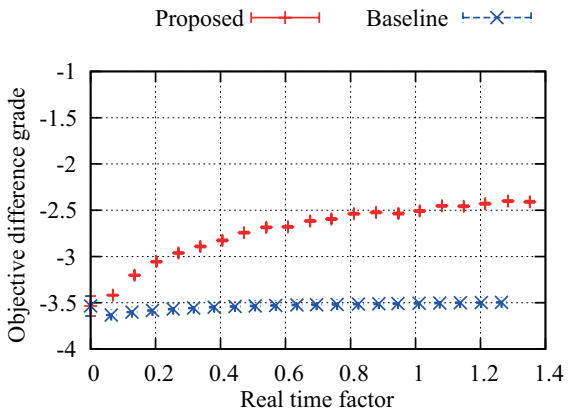
$$h_n = \begin{cases} 0 & (0 \le n < \frac{N-M}{2}) \\ 0.5 - 0.5\cos\left(\pi\dfrac{n-\frac{N-M}{2}}{M-1}\right) & (\frac{N-M}{2} \le n < \frac{N+M}{2}) \\ 1 & (\frac{N+M}{2} \le n < \frac{3N-M}{2}) \\ 0.5 + 0.5\cos\left(\pi\dfrac{n-\frac{3N-M}{2}}{M-1}\right) & (\frac{3N-M}{2} \le n < \frac{3N+M}{2}) \\ 0 & (\frac{3N+M}{2} \le n < 2N) \end{cases} \tag{5}$$

で定義される Tukey 窓を用いると，$n = 0,1,\cdots,N-1$ において $h_n + h_{n+N} = 1$ が成り立つため各 $n = 0,\cdots,2N-1$ で $v_n = 1$ となる．M $(0 < M < N)$ はオーバーラップ量を調節するパラメータであり，M が大きいほどセグメント同士のオーバーラップ量が多い．以下では (5) 式で定義される $\boldsymbol{h}$ を用いる．

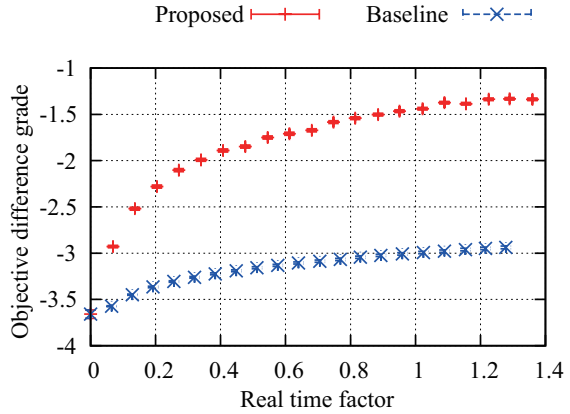
3.2 逐次位相推定アルゴリズム

3.1 節で述べたように，オフライン位相推定問題の解は一意でないため，互いに重複部分のあるセグメントに対して独立にオフライン位相推定アルゴリズムを適用すると，各セグメントで推定された位相が重複区間で整合しないことがありうる．次節で示す通り，位相の不整合性は振幅スペクトログラムから再構成された信号の音質の劣化要因となる．そこで，重複部分で整合するような位相を推定するアルゴリズムを提案する．

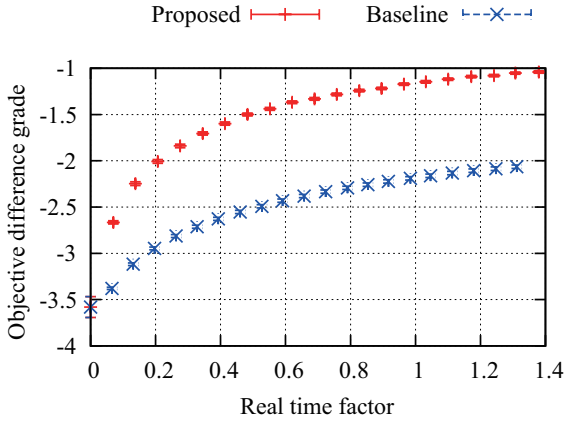
直前のセグメントと当該セグメントの重複区間において，直前のセグメントで推定された時間信号を N 次元の複素ベクトル $\boldsymbol{g}$ で表す．$\boldsymbol{s}(\boldsymbol{a},\boldsymbol{\phi})$ を当該セグメントの複素スペクトログラムの推定値として再定義すると，セグメントの重複部分の信号は同一でなければならないので，$W^{+}\boldsymbol{s}(\boldsymbol{a},\boldsymbol{\phi})$ の直前のセグメントとの重複部分について再度窓を掛けなおしても同一の波形となるはずである．したがって，



(a) セグメント長 170 ms



(b) セグメント長 340 ms



(c) セグメント長 680 ms

図 3 提案法（Proposed）とベースライン法（Baseline）による様々なセグメント長での RTF に対する ODG の平均値と標準誤差．各点は，RTF が小さい方からそれぞれ各セグメントで反復回数を $0, 10, \cdots, 200$ としてアルゴリズムを実行した場合の結果である．

$$W^{+}\boldsymbol{s}(\boldsymbol{a},\boldsymbol{\phi}) = \mathrm{diag}(\boldsymbol{h})\left(\begin{bmatrix}\boldsymbol{g}\\ 0_N\end{bmatrix} + \begin{bmatrix}I_N & 0_N\\ 0_N & 0_N\end{bmatrix} W^{+}\boldsymbol{s}(\boldsymbol{a},\boldsymbol{\phi})\right) + \begin{bmatrix}0_N & 0_N\\ 0_N & I_N\end{bmatrix} W^{+}\boldsymbol{s}(\boldsymbol{a},\boldsymbol{\phi}) \tag{6}$$

が成り立つ．(6) 式が常に成り立つと仮定できれば，(6) 式の右辺を逐次高速近似 CWT での逆変換とみなせる．そのため，(3) 式の代わりに

$$\tilde{\boldsymbol{s}} \leftarrow W\left\{\mathrm{diag}(\boldsymbol{h})\left(\begin{bmatrix}\boldsymbol{g}\\ 0_N\end{bmatrix} + \begin{bmatrix}I_N & 0_N\\ 0_N & 0_N\end{bmatrix} W^{+}\boldsymbol{s}(\boldsymbol{a},\boldsymbol{\phi})\right) + \begin{bmatrix}0_N & 0_N\\ 0_N & I_N\end{bmatrix} W^{+}\boldsymbol{s}(\boldsymbol{a},\boldsymbol{\phi})\right\} \tag{7}$$

を用いることで，セグメントの重複部分で整合する位相となるように誘導できる可能性がある．本稿では，このアルゴリズムを逐次位相推定アルゴリズムと呼ぶ．このアルゴリズムでは，当該セグメントのスペクトログラムと $\boldsymbol{g}$ のみを保存すればよく，空間計算量は $O(N+\Omega D)$ である．そのため，空間計算量が信号長に依存せず，セグメント長に比べ信号長が大きい場合にオフライン位相推定アルゴリズムに比べて空間計算量を格段に削減できる．

4. 実験

4.1 実験条件

セグメントの重複部分の位相の整合性による効果と計算速度を評価するため，振幅スペクトログラムからの時間信号の再構成実験を行った．提案法と，各セグメントに対してそれぞれオフライン位相推定アルゴリズムを独立に適用する方法（ベースライン法）を比較した．実験データとして，RWC 音楽ジャンルデータベース [12] の 10 曲を用いた．サンプリング周波数は 48 kHz とし，各曲の冒頭 30 s に逐次高速 CWT を適用し得られた振幅スペクトログラムを入力とした．セグメント長は $170, 340, 680$ ms（$N = 2^{12}, 2^{13}, 2^{14}$）とし，(5) 式で定義される分析窓を用いた（$M = N/4$）．アナライジングウェーブレットとして対数正規分布型のウェーブレット [3] を用いた．このウェーブレットの Fourier 変換は対数周波数領域で正規分布と同形であり，正規分布の標準偏差に対応するパラメータを 0.02 とした．フィルタの中心周波数が 25 cent 毎（各オクターブ 48 ビン）に 27.5 から 23679.5 Hz となるようスケールを設計し，高速近似 CWT での帯域制限の範囲は中心周波数から対数周波数上で $[-2\sigma, 2\sigma]$ とした．位相はランダムに初期化し，各セグメントでの反復回数をそれぞれ $0, 10, \cdots, 200$ 回として変えて比較を行った．

計算速度の評価指標として，セグメントのシフトに対する処理時間の比で定義される real time factor（RTF）を用いた．RTF が 1 以下であれば実時間で実行可能であり，低ければ低いほど高速である．再構成信号の音質の評価指標として，perceptual evaluation of audio quality（PEAQ）[13] による objective difference grade（ODG）を用いた．ODG は −4 から 0 までの値をとり，ODG が大きいほど音質が高い．

4.2 結果

図 3 に，提案法とベースライン法による再構成信号の ODG と計算時間の結果を示す．提案法はベースライン法に比べ再構成信号の ODG が高く，重複部分の位相の整合

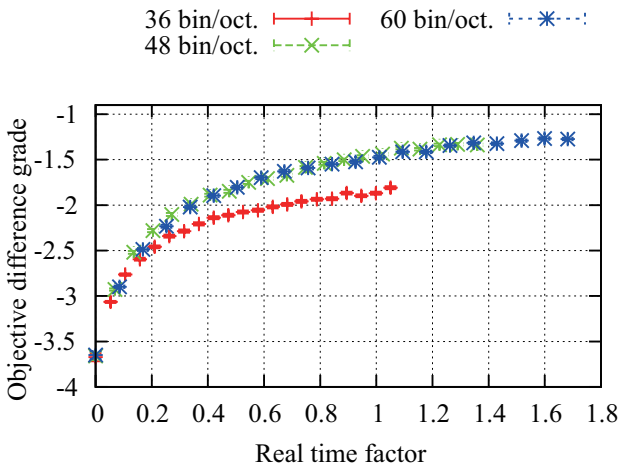


図 4 様々なオクターブ毎の周波数ビン数での提案法による RTF に対する ODG の平均値と標準誤差．“36 bin/oct.” はオクターブ毎の周波数ビン数が 36 であることを表す．各点は，RTF が小さい方から反復回数を $0, 10, \cdots, 200$ としてアルゴリズムを実行した場合の結果である．

性が位相推定に有効であることが確認できる．実際に筆者がベースライン法の再構成信号を聴取したところ，セグメントの重複部分で音量が下がったように聴こえた．これは，重複部分で互いに打ち消し合うような信号成分が得られたためであると推察できる．短いセグメント長ほどベースライン法で得られた再構成信号の ODG が低い傾向があるが，これはセグメント長が短くなるほどセグメント数が増え重複する信号成分が増えるために重複部分の位相の整合性の重要度が増すからである．また，セグメント長 340, 680 ms とした提案法では実時間で ODG が −2 以上の再構成信号が得られており，実時間制約下でもある程度の音質の音響信号が再構成可能であった．

最後に，オクターブ毎の周波数ビン数を 36，48，60 とした場合を比較した．セグメント長は 340 ms とし，他のパラメータや位相の初期値は上述の実験と同一とした．周波数ビン数が少なければ少ないほど計算量が小さくなるが，CWT のサブバンドフィルタ同士の重複部分が減少するため再構成信号の音質が低くなる可能性がある．図 4 に示す通り，オクターブ毎の周波数ビン数を 48 以上とすれば実時間で動作しつつある程度の音質の再構成信号が得られることを確認した．

5. まとめ

本報告では，振幅スペクトログラムからの逐次位相推定アルゴリズムを提案した．提案アルゴリズムでは，固定長のセグメント毎に得られた振幅スペクトログラムに対して位相を推定するため，空間計算量が信号長に依らず一定である．また，セグメントの重複部分の位相の整合性を考慮した更新式を用いることで，各セグメント対して独立にオフライン位相推定を適用した場合よりも高音質な信号を振幅スペクトログラムから再構成できることを実験により確認した．また，音楽音響信号に対して実時間制約下でもある程度の音質の信号が再構成できることも確認した．今後は，音声に関する性能評価実験や提案アルゴリズムの収束性に関する理論的な保証が課題である．

謝辞 本研究は JSPS 科研費 26730100，15J0992 の助成を受けたものである．

参考文献

[1] Burred, J. J. and Sikora, T.: Comparison of frequency-warped representations for source separation of stereo mixtures (2006).
[2] Schmidt, M. N. and Mørup, M.: Nonnegative matrix factor 2-D deconvolution for blind single channel source separation, *Proc. Int. Conf. Independent Component Analysis and Blind Signal Separation*, pp. 700–707 (2006).
[3] Kameoka, H.: Statistical Approach to Multipitch Analysis, PhD Thesis, The University of Tokyo (2007).
[4] de León, J. P., Beltrán, F. and Beltrán, J. R.: A complex wavelet based fundamental frequency estimator in single-channel polyphonic signals, *Proc. Int. Conf. Digital Audio Effects*, pp. 47–54 (2013).
[5] Ikemiya, Y., Yoshii, K. and Itoyama, K.: Singing Voice Analysis and Editing based on Mutually Dependent F0 Estimation and Source Separation, *Proc. Int. Conf. Acoust. Speech Signal Process.*, pp. 574–578 (2015).
[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] 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.
[8] Holighaus, N., Dorfler, M., Velasco, G. and Grill, T.: A Framework for Invertible, Real-Time Constant-Q Transforms, *IEEE Trans. Acoust., Speech, and Language Process.*, Vol. 21, No. 4, pp. 775–785 (2013).
[9] Schörkhuber, C., Klapuri, A., Holighaus, N. and Dörlfer, M.: A Matlab Toolbox for Efficient Perfect Reconstruction Time-Frequency Transforms with Log-Frequency Resolution, *Proceedings of AES International Conference on Semantic Audio* (2014).
[10] Ortega, J. M. and Rheinboldt, W. C.: *Iterative solution of nonlinear equations in several variables*, No. 30, SIAM (1970).
[11] Nakamura, T. and Kameoka, H.: Fast Signal Reconstruction from Magnitude Spectroggram of Continuous Wavelet Transform based on Spectrogram Consistency, *Proc. Int. Conf. Digital Audio Effects*, pp. 129–135 (2014).
[12] Goto, M.: Development of the RWC Music Database, *Proc. Int. Congress Acoust.*, Vol. 1, pp. 553–556 (2004).
[13] : ITU-T Recommendation BS.1387-1, Perceptual Evaluation of Audio Quality (PEAQ): Method for Objective measurements of perceived audio quality (2001).