

# パワースペクトログラムの伸縮と無矛盾位相付加に基づく 音楽音響信号の実時間テンポ / ピッチ変換\*

水野優 (東大・工), ルルージョナトン, 小野順貴, 嵯峨山茂樹 (東大・情報理工)

## 1 はじめに

音楽音響信号を自由にテンポ / ピッチ変換させる技術は、音楽の楽しみ方を広げるだけでなく、カラオケや自動伴奏システム [1] の高音質化などに応用できる基礎技術としても重要である。すでに多くのフリーソフトや商用ソフトで試みられてはいるものの、音楽信号のような多重音信号に対して高品質な加工を行うことは必ずしも容易ではない。本稿では、与えられたパワースペクトログラムに対して矛盾の無い位相を付与する観点から、音楽音響信号加工の手法に関して、そのアルゴリズムと実時間実装について論じる。

## 2 スペクトログラムへの無矛盾な位相付加

### 2.1 パワースペクトログラム領域での信号加工

人間の聴覚系はパワースペクトログラムに相当する特徴を抽出し音響信号を知覚していると考えられる。したがって、音楽信号のテンポ変換は原信号のパワースペクトログラムを時間軸方向に伸縮することにより、ピッチ変換は周波数軸方向に伸縮することにより、達成できると考えられる。これを一般化すれば、原信号のパワースペクトログラムに対して所望の加工を (位相を考えないで) 施した後、そのようなパワースペクトログラムを持つ信号波形を生成することができれば、様々な信号加工の可能性が広がる。一つのパワースペクトログラムには多数の位相の可能性があるが、一意決定はできないが、フーリエ解析をすれば所望のパワースペクトログラムになるように位相が決定できれば、目的が達成できるであろう。

一般に、短時間フーリエ変換によって得られる複素スペクトログラムは、フレームのオーバーラップに起因した冗長な表現であり、そのパワーと位相は独立ではない。この性質を利用し、与えられたパワースペクトルとつじつまがあう、いわば無矛盾な位相を付与する手法が Griffin and Lim[3] によって提案されているが、この手法は多大な反復演算を必要とし、計算量が大きな制約となっていた [4]。

効率的な計算法としてスライディングブロック分析を用いることにより、パワースペクトログラムの伸縮に基づくテンポ / ピッチ変換技術の実時間化が考えられる。同様の検討は、近年、Zhu らからも独立に報告されているが [5]、以下に述べる我々の手法は、

1) FFT を用いた高速なりサンプリングに基づくピッチ変換、2) スペクトル包絡伸縮との併用によるフォルマントとピッチの独立な操作、などに特徴があり、反復演算自体の高速化 [2] もあわせて検討を進めている。

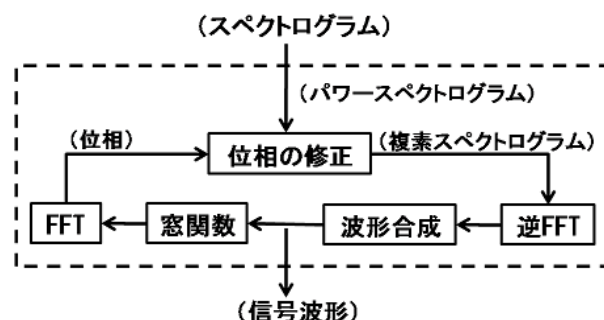


Fig. 1 無矛盾位相付加のアルゴリズム。反復推定により任意のスペクトログラムに対し信号波形を合成。

### 2.2 スペクトログラムからの位相推定

任意のパワースペクトログラムに対し無矛盾な位相を推定し波形を合成する手法としては、最初に、目的とするパワースペクトログラムに対し適当な位相を与えて複素スペクトログラムにした後、以下の処理を反復して行う (Fig. 1) ことにより実現できる。

1. 逆フーリエ変換し、時間領域で波形を合成する。
2. 窓関数を用いて波形を切り出し、フーリエ変換して推定位相を求める。
3. 推定位相とパワースペクトログラムで複素スペクトログラムを再構成する。

実時間化するためには、数フレーム分のパワースペクトログラムに対し反復推定を行いつつ、最も古いフレームから順次推定を終了し、その分だけ次のフレームを取得していけばよい。この手法では、フレームのオーバーラップにより、既に推定された位相情報を新たなフレームでの推定に利用している。そのため、全体の反復回数を少なくすることができ、実時間での位相推定が可能である。

## 3 スペクトログラムの加工

次に、目的とする変換に対し、それを実現するスペクトログラムを得るためには、原信号にどのような加工を施せばよいかについて述べる。

### 3.1 テンポ変換のアルゴリズム

先にも述べたように、テンポ変換はパワースペクトログラムを時間軸方向に伸縮することで得られる。これは、原信号からフーリエ変換により新規フレームのスペクトルを取得するとき、前のフレームからのフレームの移動幅を変化させることにより、計算量を変化させずに実現できる。移動幅を  $\alpha$  倍すると、同じ長

\*"Real-time Time-scale/Pitch Modification of Music Signal by Stretching Power Spectrogram and Consistent Phase Reconstruction" by Mizuno Yuu, Jonathan Le Roux, Ono Nobutaka, Sagayama Shigeki (The University of Tokyo).

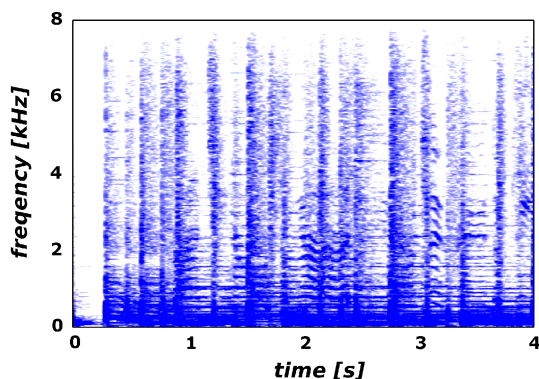


Fig. 2 原信号のパワースペクトログラム例

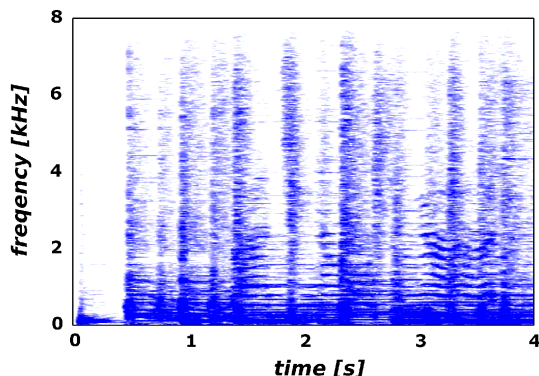


Fig. 3 2/3倍テンポ変換信号のスペクトログラム例

さの原信号から得られるフレームは $1/a$ 倍となり、スペクトログラムは時間方向に $1/a$ 倍、すなわちテンポを $a$ 倍した形となる (Fig. 2,3).

### 3.2 ピッチ変換のアルゴリズム

ピッチ変換はパワースペクトログラムを周波数軸方向に伸縮することで得られ、これは信号をリサンプリングすることに相当する。リサンプリングは通常多くの計算量を必要とするが、新規フレームのスペクトルを取得するとき、フーリエ変換のフレーム長を変化させることで高速リサンプリングが実現できる。通常の $b$ 倍のフレーム長のデータをフーリエ変換した場合、得られるスペクトルは $b$ 倍の長さとなるが、その最大周波数は変化していない。そのため、得られるスペクトルは周波数方向に $b$ 倍、すなわちピッチを $b$ 倍に変化させた形となる (Fig. 2,4)。位相情報を付加する際には、高周波成分を0詰めまたは間引きすることで通常のフレーム長にしてからパワースペクトログラムを取得すればよい。

### 3.3 スペクトル包絡を保ったピッチ変換

音声の認識や合成においてスペクトル包絡が重要な役割を果たすことは広く知られている。そのため、音声加工においては、スペクトル包絡を保存したピッチ変換は有用である。ここでは線形予測分析 (LPC) を利用した手順を以下に示す。

1. 1フレーム分の原信号をLPCによりスペクトル包絡と残差信号に分離する。
2. 残差信号を3.2節の手法により高速にピッチ変換した後、時間領域に戻す。

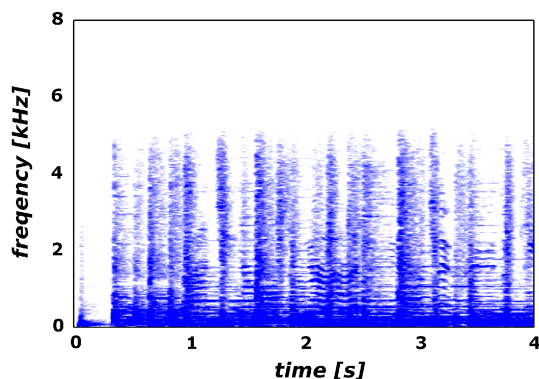


Fig. 4 2/3倍ピッチ変換信号のスペクトログラム例

3. ステップ2の出力信号がステップ1のスペクトル包絡を持つようにフィルタを掛ける。
4. ステップ3の出力信号をフーリエ変換しスペクトログラムの1フレーム分を得る。

以上の3.1節~3.3節の変換は、すべて計算量が小さく抑えられているため、通常のパソコンでの実時間動作が可能である。また、テンポ変換とピッチ変換 (スペクトル包絡を維持した変換も含む) を組み合わせた処理についても通常のパソコンにおいて実時間で動作できる。

## 4 結論

本稿では、任意のパワースペクトログラムへの無矛盾な位相付加の観点から、実時間動作可能な音楽音響信号の高音質なテンポ/ピッチ変換手法を提案した。

今後の発展として、スペクトル包絡保存したピッチ変換でより良い音質を実現する手法や、自動伴奏システムへの組み込みについて検討する。

## 参考文献

- [1] 武田他, 情報処理学会研究報告, “HMMによるMIDI演奏の楽譜追跡と自動伴奏,” 2006-MUS-66, pp.109-116, Aug. 2006.
- [2] J. L. Roux *et al.*, “Explicit consistency constraints for STFT spectrograms and their application to phase reconstruction,” *Proc. SAPA*, Sep. 2008.
- [3] D. W. Griffin *et al.*, “Signal Estimation from Modified Short-Time Fourier Transform,” *IEEE Trans. ASSP.*, 32(2), 236-243, Apr. 1984.
- [4] J. Laroche *et al.*, “Improved Phase Vocoder Time-Scale Modification of Audio,” *IEEE Trans. SAP.*, 7(3), 323-332, May 1999.
- [5] X. Zhu *et al.*, “Real-Time Signal Estimation from Modified Short-Time Fourier Transform Magnitude Spectra,” *IEEE Trans. ASLP*, 15(5), 1645-1653, Jul. 2007.