

# Specmurt 法による音楽信号の音高可視化における 共通調波構造パターンの自動決定\*

亀岡弘和<sup>†</sup> 齊藤翔一郎<sup>††</sup> 西本卓也<sup>†</sup> 嵯峨山茂樹<sup>†</sup>  
<sup>†</sup> 東京大学大学院 情報理工学系研究科 <sup>††</sup> 東京大学 工学部計数工学科

## 1 はじめに

我々は、音高を可視化してそれをユーザが容易に手で修正が入れられ、修正後の結果を利用してより高精度に仕上げていけるようなインタラクティブツールの完成を目指しており、本稿はこの構想における前半の可視化技術に位置づけられる。オーケストラのように発音数がある程度多い楽曲を全自動で完全に採譜あるいは分離するのは難しく、このようなインタラクティブツールが今後有用となることが期待される。

これまで我々は、人が目で出力を見て修正すべき箇所が分かり易い音高「可視化」手法 (Specmurt 法) を提案した [1]。一意の解を出力する従来の音高「推定」手法では多くの場合、脱落誤りのように推察しづらい誤りを伴うことがあったが、我々のアプローチではスペクトログラムに一種のフィルタリングをかけたものを濃淡表示出力する。本報告では、この手法において想定する共通調波構造パターンを自動で決定する方法を提案する。

## 2 Specmurt 法 [1]

### 2.1 対数周波数スペクトルの逆畳み込み

簡単のため、以後扱う音響信号は調波性をもち、パワースペクトルは加法性が仮定できるとする。

対数周波数スケール (対数周波数を  $x$  とする) では、 $n$  次高調波周波数は基本周波数の値に関わらず基本周波数から  $\log n$  だけ離れて位置し、基本周波数が  $\Delta x$  変化すると高調波周波数は揃って  $\Delta x$  変化する。

ここで、周波数成分間のパワー比が基本周波数に関わらず共通である調波構造を想定し、基本周波数に相当する位置を原点とした図 1 のようなパターン  $h(x)$  を共通調波構造パターンと呼ぶ。また、基本波成分を 1 とする。次に、基本周波数がどの値でどれだけ成分をもつかを表した基本周波数分布と呼ぶ関数  $u(x)$  を定義すると、多重音スペクトルは

$$v(x) = h(x) * u(x) \quad (1)$$

により表される。

逆に、 $v(x)$  を観測情報として  $h(x)$  が既知ならば、 $u(x)$  は  $v(x)$  と  $h(x)$  の逆畳み込みにより得ることができる。畳み込みの定理より、 $u(x)$ 、 $h(x)$ 、 $v(x)$  を逆フーリエ変換したものを  $U(y)$ 、 $H(y)$ 、 $V(y)$  とすると、

$$u(x) = \mathcal{F} \left[ \frac{V(y)}{H(y)} \right] \quad (2)$$

で  $u(x)$  は求まる。このように対数周波数領域における逆畳み込みにより基本周波数分布関数を求める方法を Specmurt 法と呼ぶ。

\*“Automatic Determination of the Common Harmonic Structure Pattern in Specmurt Method for Pitch Visualization of Music Signals” by Hirokazu KAMEOKA, Shoichiro Saito, Takuya NISHIMOTO and Shigeki SAGAYAMA (The University of Tokyo).

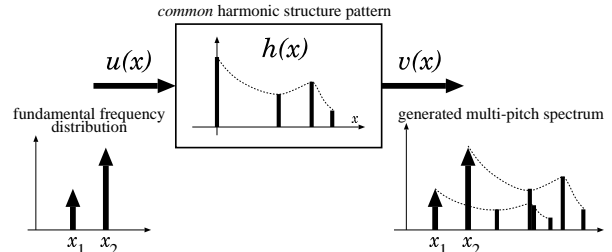


図 1: 基本周波数分布と共通調波構造パターンの畳み込み

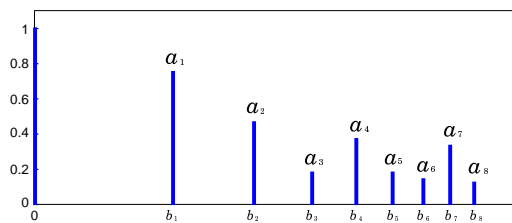


図 2: パラメトリック共通調波構造パターン

短時間周波数解析においては、窓関数などの影響により各周波数成分は広がって分布するが、Specmurt 法を実際利用する上では、ウェーブレット変換や定  $Q$  フィルタなどのように対数周波数領域で周波数成分の広がりの大きさが均一であるようなスペクトルを出力する短時間分析方法で  $v(x)$  を得るのが適切である。その理由は、高調波周波数と基本周波数の相対位置関係だけでなく、分布の形状も基本周波数の違いに関わらずに一定である方が良いからである。対数周波数スケールの短時間フーリエ変換のスペクトルでは、高周波になるにつれ周波数成分の広がりは小さくなるため、畳み込み演算から逸脱してしまう。

### 2.2 調波成分抑圧フィルタの働き

以上では、調波構造が共通な場合の理論を議論したが、一般的に調波構造パターン (包絡) の形状は、楽器の種類などによって異なる。 $h^{-1}(x)$  は観測スペクトル  $\tilde{v}(x)$  を入力信号 ( $x$  が時間) と見立てた場合、線形系のフィルタと解釈でき、 $\tilde{v}(x)$  の中から調波構造を見つけ出し、調波成分だけを抑圧低減しようとする働きがある。すなわち Specmurt 法は、共通の逆フィルタで  $\tilde{v}(x)$  の調波成分を抑圧する手法であり、多重音中の各音色が異なる場合も有効である。

## 3 共通調波構造パターン自動決定

これまででは、共通調波構造パターン  $h(x)$  は実験的に定める必要があった [1]。本章では、適切な  $h(x)$  を自動的に決定する方法を提案する。

Specmurt 法は、2 つの未知で独立な関数の演算の出力値だけが既知であるという不良設定問題であり、 $h(x)$  の数学的な良し悪しの評価基準はないため  $u(x)$  と  $h(x)$  の組合せは無数通り考えられる。しかし、 $u(x)$  はいくつかの鋭いピークをもち、それ以外は 0 であ

るような分布であるはずである．そこで，このような分布に誘導していくように  $u(x)$  と  $h(x)$  を次の2つのステップにより反復更新していく方法を考える．

### Step1: 基本周波数分布の非線形変換

$u(x)$  の中の不要成分が正規分布に従うと仮定し， $u(x)$  を正規分布に従って確率変動する値による閾値処理を施す．正規分布の累積分布関数を sigmoid 関数で近似し，以下のような非線形写像

$$\bar{u}(x) = \Psi[u(x)] = \frac{1}{1 + \exp\{-\alpha(u(x) - \beta)\}} u(x) \quad (3)$$

により  $u(x)$  を  $\bar{u}(x)$  に更新する．この写像は， $\beta$  より十分大きい値は  $u(x)$  とほぼ同じ値を返し， $\beta$  周辺または小さい値は 0 に近づける働きがある．

### Step2: 共通調波構造パターン最適化

共通調波構造パターンを，図 2 のような成分  $a = \{a_1, \dots, a_N\}$  をパラメータとするデルタ関数列

$$\tilde{h}(x, \mathbf{a}) = \sum_{n=0}^N a_n \delta(x - b_n) \quad (4)$$

で表す．ただし  $n$  は調波成分の次数， $N$  は調波成分の数， $b_n$  は  $\log(n+1)$  (定数) である．また， $a_0 = 1$  とする． $\bar{u}(x)$  と  $\tilde{h}(x, \mathbf{a})$  の畳み込み  $v(x, \mathbf{a})$  も  $\mathbf{a}$  の関数となるので，観測パワースペクトル  $\tilde{v}(x)$  と  $v(x, \mathbf{a})$  との二乗誤差の積分を  $\mathbf{a}$  について偏微分して 0 と置くと  $N$  次元の連立 1 次方程式の形となり，LU 分解により最適パラメータ  $\bar{\mathbf{a}}$  を得ることができる．新たに更新された  $\tilde{h}(x, \bar{\mathbf{a}})$  を  $h(x)$  とし，Specmurt 法により  $u(x)$  を再度計算し，Step1 の非線形写像に戻る．

提案方法は，多重音中の各音の中間音色を探し出す働きがあり，各音が異なる音色をもつとしてもできるだけ鮮明な基本周波数分布を出力しようとする．

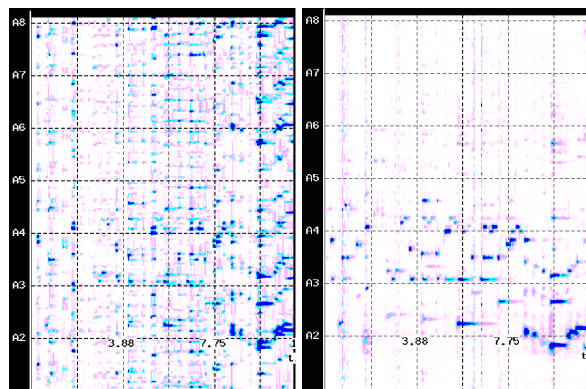
## 4 評価実験

提案方法と Specmurt 法を用いて実音楽信号に対して解析実験を行った．実験データとして採用した音楽信号は，RWC 研究用音楽データベースに収録されている合計 7トラックの単一種類の楽器による多重奏(ピアノ，ギター)音響信号(サンプリング周波数 16kHz，モノラル)とした．周波数分析はウェーブレット変換(ウェーブレット関数: Gabor 関数，分散: 30cent，周波数分解能: 16.7cent，周波数帯域: 55–7604Hz)を用い，提案方法における初期  $h(x)$  形状は  $f^{-1.5}$  ( $f$  は周波数)，反復更新回数は 5 回とした．

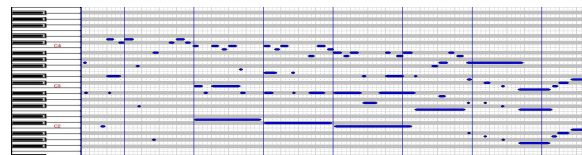
### 4.1 実験結果

図 3 に， $h(x)$  の包絡形状を  $f^{-0.5}$  とした場合と， $f^{-0.5}$  を初期値として  $h(x)$  の反復推定を行った場合の出力基本周波数分布の比較を示す．左図では，除去されずに残った調波成分などの不要成分により，実際の演奏とかけ離れた分布となったが，提案方法によって出力された右図では誤差成分や調波成分が抑圧された可視化結果を得ることができた．

可視化結果がどの程度正確に得られているかを評価する 1 つの目安として，得られた  $u(x)$  の時系列から MIDI に自動変換(手順は省略)を行い，データベースに付属されている各曲に対応する手入力標準 MIDI ファイルを参照して，XGworks のピアノロールウィ



(a) 反復推定なし (b) 反復推定あり



(c) 手入力 MIDI 信号をピアノロールウィンドウ表示

図 3:  $h(x)$  を固定した場合 (a) と反復推定を行った場合 (b) の実験データ “Jive (Piano solo)” 同一部分の可視化結果．

表 1: 各実験データに対する音高正解率

実験データ		MIDI 変換正解率 (%)
タイトル	作曲者	
Jive (Piano Solo)	中村真	77.8
Lounge Away (Piano Solo)	永井隆雄	78.4
Jive (Guitar Solo)	中鉢洋夫	77.6
For Two (Guitar Solo)	中鉢洋夫	76.9
Crescent Serenade (Guitar Solo)	山本聡	74.5
Abyss (Guitar Solo)	中鉢洋夫	72.0
ノクターン第 2 番変ホ長調 op.9-2	ショパン	80.4

ンドウ上で比較しながら目視で音高正解率を計算した．正解率は，総音符数から挿入誤り，置換誤り，脱落誤りの数を引いたものを総音符数で割ったものとした．各実験データに対する音高正解率を表 1 に示す．

全 7トラックの実演奏音響信号に対し，全体として 70 ~ 80% の音高正解率を得た．主な誤りは，Specmurt 法の原理的な性質上生じる誤りである調波成分を除去しきれないことによる挿入誤り，オクターブ重音を 1 音と見なしてしまうことによる脱落誤りなどであった．また，ピアノは非調和性楽器であることが知られており，調和性を仮定した共通調波構造パターンを用いていることが不要成分の発生に影響し，上記以外の誤りに反映されている可能性がある．

## 5 まとめと展望

本報告では，基本周波数分布を非線形変換により更新するステップとパラメトリックに表現された共通調波構造パターンを最適化するステップからなる共通調波構造パターンの反復推定方法を提案し，我々がこれまでに提案した Specmurt 法の発展を試みた．また，単一種類の楽器による実演奏データを対象に評価実験を行い，70 ~ 80% 前後の音高正解率を得た．

今後は，冒頭で述べたように，ユーザおよび別の多重音解析手法が積極的に介入できる実用性の高いツールを開発していきたい．

### 参考文献

- [1] 高橋佳吾, 西本卓也, 嵯峨山茂樹, “対数周波数逆畳み込みによる多重音の基本周波数解析,” 情報処理学会研究報告, 2003-MUS-53-13, pp. 61–66, 2003.