

Specmurt における準最適共通調波構造パターンの 反復推定による多声音楽信号の可視化と MIDI 変換

亀岡 弘 和[†] 齊藤 翔 一郎^{††}
西本 卓 也[†] 嵯峨山 茂樹[†]

これまで我々は、多声音楽信号の高調波周波数成分を抑圧低減して基本周波数のみを強調することでピアノロールに似た可視化を可能にする新しい信号処理方法“Specmurt 法”を提案した^{1),2)}。この手法は、多重音の各音が共通の音色(調波構造パターン)をもつと想定することで、観測多重音スペクトルと共通調波構造パターンとの逆畳み込みにより基本周波数成分だけを強調することができる。ただし、得られる可視化結果の精度は共通調波構造パターンの定め方に大きく依存するにも関わらず、これまでは人手で調節しながら設計する必要があった。本報告では、基本周波数分布関数の sigmoid 関数による非線形変換ステップと LU 分解によるパラメータ最適化ステップを繰り返すことで準最適な共通調波構造パターンを推定する方法を提案する。これにより、人手による手間をかけずに鮮明な音高可視化表示が可能となることを実音楽信号を対象とした実験により確認した。また、得られた可視化結果をもとに MIDI フォーマットに変換することができ、70 ~ 80% の音高正解率を得た。

Recursive Estimation of Quasi-Optimal Common Harmonic Structure Pattern for Specmurt Anaylis: Piano-Roll-Display Visualization and MIDI Conversion of Polyphonic Music Signal

HIROKAZU KAMEOKA,[†] SHOICHIRO SAITO,^{††} TAKUYA NISHIMOTO[†]
and SHIGEKI SAGAYAMA[†]

We have previously proposed a new signal processing technique, “specmurt anaylis,” that suppresses harmonic components and provides piano-roll-like visual display of multi-tone signals (e.g., polyphonic music). “Specmurt” enhances fundamental components by the deconvolution of the input multi-pitch spectrum and the common harmonic structure pattern in log-frequency domain. However, how intelligible the output display may be depends highly on the common harmonic structure pattern we determine. This had to be adjusted experimentally by trial and error. In this paper, we propose a method that automatically estimates the quasi-optimal common harmonic structure pattern by the iteration; nonlinear mapping of the fundamental frequency distribution using sigmoid function and the optimization of the parametric impulse train pattern using LU decomposition. We were able to visualize intelligible piano-roll-like displays of real music signal data without spending effort on the pattern adjustment. The resulted display was transformed into a MIDI format and the musical notes were recognized with an accuracy of 70 ~ 80%.

1. はじめに

本報告は、実音楽信号中の複数音高の視覚的理解、および自動推定を支援する技術を扱う。大量に市販されているクラシック音楽などの楽譜は、ある程度の能力や経験や労力を要する採譜作業によって通常作成されているが、音高(基本周波数)情報が可視化され

ば非常に有用である。また、携帯電話の着信メロディやカラオケの伴奏などのように MIDI フォーマットで記述されている媒介は商用目的やそれ以外の目的として広く利用されているが、現状では音響信号からの MIDI 変換作業は多大な人手をかけて行われているため、作業の支援技術や自動変換技術は需要が高い。音楽検索の研究分野においても、音楽信号を MIDI 化できれば移調や別編曲も含めた検索が容易になる。

ただし、短時間周波数解析における低い周波数分解能やスペクトルの広がり、異なる音同士の周波数成分の重複などに起因して多重音信号からの自動音高推定は容易ではなく、これまで打開策としてさまざまなア

[†] 東京大学大学院情報理工学系研究科

Graduate School of Information Science and Technology, The University of Tokyo.

^{††} 東京大学工学部

School of Engineering, The University of Tokyo.

アプローチがなされてきた^{3)~12)}。後藤は、多重音スペクトルを複数の正規分布によって構成される調波構造モデルを多数混合した分布の各重みパラメータをEMアルゴリズムで推定する方法を提案した^{9),10)}。また、我々はこれまで fuzzy クラスタリングの応用により調波スペクトル成分を分離してクラスタ中心として周波数をパラメトリックに推定する, “Harmonic Clustering” を提案した。この手法では, 同時発音数を推定する手段と倍ピッチ/半ピッチ誤りを防ぐ手段をいずれも情報量規準を用いて効果的に定式化した^{11),12)}。これらの手法は, パラメータの最適解を反復計算によって求めているが, 初期値の設定に強く依存して予測不能な誤り解をもたらす可能性がある。

そこで我々は, これまで提案された大多数の手法に代表されるような, 一意の解を与えるいわゆる「推定」に基づくアプローチではなく, スペクトログラムから調波成分を抑圧して基本波成分だけを強調してすべての成分(理想的には基本波成分だけとなり, 結果的にピアノロールに良く似た濃淡画像)を表示する「可視化」に基づくアプローチをとる。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られ, 修正後の結果を利用して他の「推定」技術でより高精度に仕上げた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。

このために本報告では, 人が目で見て修正すべき箇所がある程度分かり易く, 処理が簡単(少ない計算量)な手段を狙いとしている。上で述べたように, これまでの「推定」の考え方によるアプローチでは多くの場合, 出力される推定結果に現れる誤りを直接見ても, どの箇所が誤りあるいは正解であるかを判断するのは難しかったが, 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。オーケストラなどはもちろん, 発音数がある程度多い楽曲を現存の技術を用いて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は今後, これまでこの分野において開発された多くの成果をより広い用途で一般に活用されることを支援できる可能性がある。

本稿の概要は以下の通りである。次章では, これまで我々が提案した音楽信号の音高可視化手法 “Specmurt” を概説し, 3章でこれまで人手で定めていた共通調波構造パターンを反復推定により自動的に決定する方法を提案する。4章では, 提案方法により改良された Specmurt 法を実音楽信号に適用し, 可視化結果を基に MIDI に自動変換して性能評価を行う。

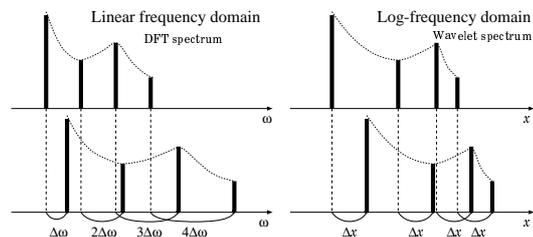


図1 線形周波数および対数周波数領域における基本周波数と高調波周波数の相対的な位置関係

2. Specmurt 法

2.1 対数周波数領域の多重音スペクトル

簡単のため, 以後扱う音響信号は調波性をもつ周期信号であることを前提とする。また, パワースペクトルは加法性が仮定できるとする。

線形周波数スケールでは, n 次高調波周波数は基本周波数の n 倍であるため, 基本周波数が $\Delta\omega$ 変化すると, n 次高調波周波数はそれに伴って $n\Delta\omega$ 変化する。これに対し, 対数周波数スケールでは, 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 Δx 変化するとすべての高調波周波数も Δx 変化する(図1)。ただし, x は対数周波数である。

ここで, すべての音の周波数成分間のパワー比が, 図1のように基本周波数に関わらず普遍的に一定である調波構造パターンを持つ場合を議論する。このようなパターンを共通調波構造パターンと呼び, $h(x)$ と表す。また, 図2のように, このパターンは基本周波数に相当する位置を原点としたものであり, 基本波成分エネルギーを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,

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

により多重音スペクトル $v(x)$ が得られる。

2.2 対数周波数スペクトルの逆畳み込み

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

$$u(x) = h^{-1}(x) * v(x) \quad (2)$$

により, 得ることができる。畳み込みの定理よりフーリエ変換領域では, 逆畳み込みは除算となるので, $u(x), h(x), v(x)$ をフーリエ変換したものを $U(y), H(y), V(y)$ とすると,

$$U(y) = \frac{V(y)}{H(y)} \quad (3)$$

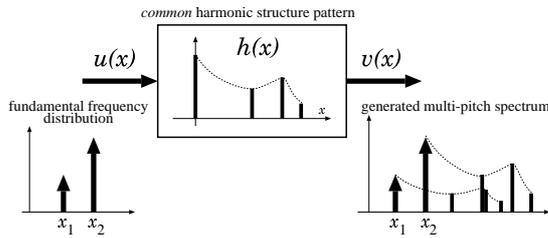


図2 基本周波数分布と共通調波構造パターンの畳み込みによる多重音スペクトルの生成の解釈

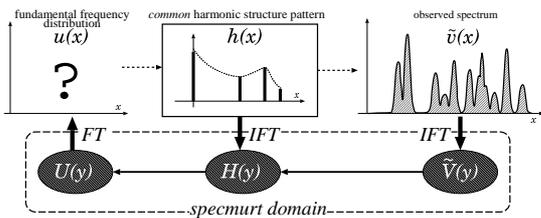


図3 Specmurt法の概観

で $U(y)$ は簡単に求めることができるので、 $u(x)$ は

$$u(x) = \mathcal{F}[U(y)] \quad (4)$$

で求まる。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(命名の由来は文献^{1),2)} 参照)。また、 y 領域は対数周波数スペクトルのフーリエ変換領域であり、これを Specmurt 領域と呼ぶ。

2.3 短時間分析における Specmurt 法

短時間周波数解析においては、窓関数などの影響による各周波数成分の広がりを考慮しなければならない。

式(1)において、 $u(x)$ をインパルス関数とし、 $h(x)$ の各周波数成分が広がりを持つ場合を考えると、 $u(x)$ は $h(x)$ を対数周波数軸上をそのまま平行移動させる働きがあるが、出力される $v(x)$ を短時間フーリエ変換のスペクトルを対数周波数スケール変換したものと見なすには、移動に伴って広がりを変化させなければならず、畳み込み演算から逸脱する。一方、 $h(x)$ をインパルス列の関数とし、 $u(x)$ を任意の広がりをもつ分布とした場合を考えると、式(1)により $v(x)$ はどの周波数成分も同じ大きさの広がりをもち、短時間フーリエ変換のスペクトルとは一致しない。

従って、短時間分析に基づいて Specmurt 法を用いるには、 $v(x)$ が対数周波数領域で周波数成分の大きさが同じ大きさであるようなスペクトルであればよいので、ウェーブレット変換や定 Q フィルタなどにより $v(x)$ を求めればよい。

2.4 調波成分抑圧フィルタとしての働き

以上では、調波構造が共通な場合の理論を議論した。しかし、実際には、一般的に調波構造パターン(包絡)の形状は、楽器の種類や、同一種類の楽器であっても基本周波数や演奏の仕方によって異なる。従って、実際得られるスペクトル $\tilde{v}(x)$ がもともと共通の調波構造パターンの重ね合わせではないならば、 $u(x)$ と $h(x)$ の畳み込みでは表現できないので $\tilde{v}(x) \neq v(x)$ である。そこで、図3のような $\tilde{v}(x)$ と任意の $h(x)$ による逆畳み込み演算にはどのような解釈ができ、 $u(x)$ がどのような分布として出力されるかを定性的に論じる。

$h^{-1}(x)$ はパワースペクトル $\tilde{v}(x)$ を入力信号と見立てた場合、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。 $h(x)$ において想定する調波成分は $\tilde{v}(x)$ に含まれる調波成分をどれだけの割合で抑圧するかを表す。例えば、共通調波構造パターン $h(x)$ が基本波成分しかもたない(単一のインパルス)パターンの場合、 $H(y) = 1$ なので $h^{-1}(x)$ はパワースペクトルをそのまま通過させる。 $h(x)$ の包絡が任意の入力音のスペクトル包絡よりも緩やかであれば、過剰に調波成分を減殺することになるので、調波成分を負の値として通過させる。逆の場合は、すべてを除去することはできないが、ある程度抑圧して通過させる。

多重音のパワースペクトルは、対応する周波数成分の位相差によって必ずしも個々の音のパワースペクトルの和に等しくない(期待値は等しい)が、Specmurt 法の本質的な働きは、個々の音を厳密に分離することよりも、調波成分を抑圧するフィルタリング処理と考えれば、2.1 節冒頭におけるパワースペクトルの加法性は近似的に仮定してもよい。

2.5 $h(x)$ の包絡形状による出力結果の差異

図4(b), (c) は、図4(a)のような音高が C4 と E4 のヴァイオリンの音響信号¹³⁾(2音)を合成した多重音信号のパワースペクトルに対して、 $f^{-1.0}$ および $f^{-0.5}$ 包絡(周波数の 1.0 乗および 0.5 乗に反比例した包絡形状)の $h(x)$ を仮定した場合の出力結果の例である。いずれの図にも、C4 と E4 に相当する 2 つの大きなピークが見られるが、抑圧され切れていない調波成分、負の調波成分やその他の誤差成分が見受けられる。

この誤差成分は $\tilde{v}(x)$ や $h(x)$ が実際は離散分布であるなどの理由から発生する。 $h(x)$ が図4(b)のように急な包絡の場合、抑圧の効果は下がり調波成分がいくらか残り、図4(c)のように緩やかな包絡の場合、抑圧の効果は高いので多くの成分が負の値になる反面、誤差成分が増大することが実験的に確認されている。

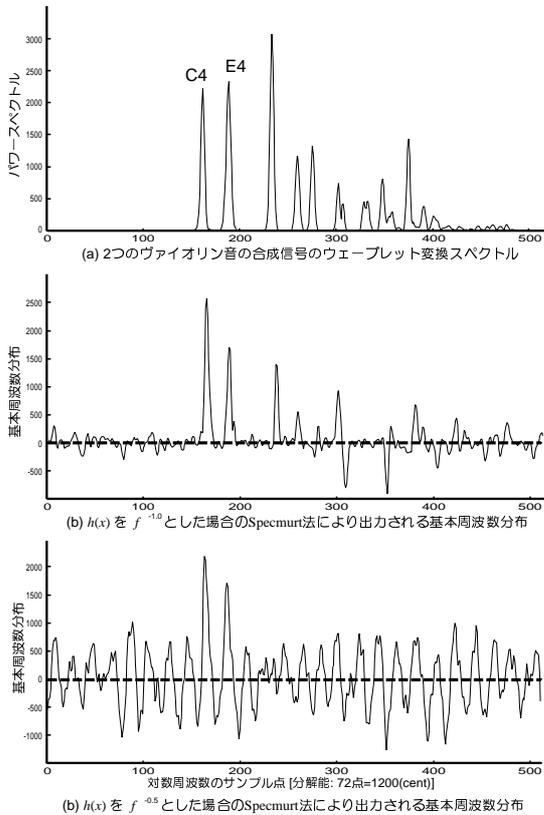


図4 入力パワースペクトル (a) に対して $f^{-1.0}$ 包絡 (b) および $f^{-0.5}$ 包絡 (c) の $h(x)$ を仮定した場合の出力基本周波数分布

3. 共通調波構造パターンの反復推定

これまでの、共通調波構造パターン $h(x)$ は人手で定める必要があり、得られる基本周波数分布 $u(x)$ はその選び方に大きく依存した。しかし、調波成分をよく低減でき、かつ誤差成分が小さくて済むような出力結果が得られるような $h(x)$ の選定は人手では容易ではない。そこで、より鮮明な可視化と人手の労力削減のために、これを自動的に決定する方法を提案する。

3.1 $u(x)$ と $h(x)$ の 2-Step 反復更新

式 (2) 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。しかし我々は、 $u(x)$ は同時発音数分に相当する数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ことが分かっている。何らかの変換で $u(x)$ にこのような特徴を与えて定め直すことができるならば、 $u(x)$ と $h(x)$ をともに繰り返し更新していくことができる。

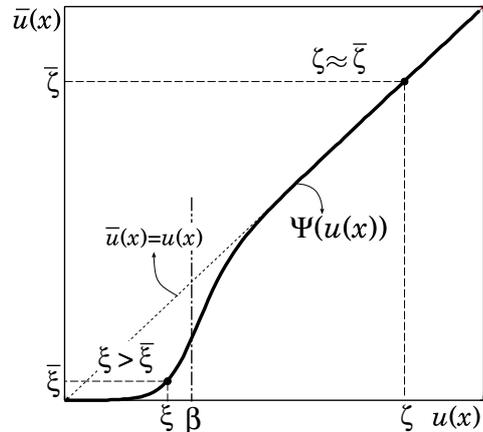


図5 非線形写像関数 $\Psi[u(x)]$

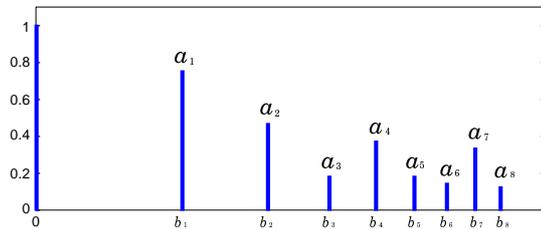


図6 パラメトリック共通調波構造パターン

Step1: 基本波成分パターンの非線形変換

ここでは、図4のような $u(x)$ の出力を $\bar{u}(x)$ に更新する方法を述べる。 $u(x)$ を「良い」 $\bar{u}(x)$ に変換するためには、(1) 顕著なピーク成分は保持し、(2) 負の成分および誤差成分を除去すべきである。そこで、以下のような sigmoid 関数を利用した非線形写像 (図5)

$$\bar{u}(x) = \Psi[u(x)] = \frac{1}{1 + \exp\{-\alpha(u(x) - \beta)\}} u(x) \quad (5)$$

により $u(x)$ を $\bar{u}(x)$ に更新する。この写像は、 β より十分大きい値はほぼ同じ値を返し、 β 近辺または小さい値 (負の値も含まれる) を 0 に近づける働きがある閾値処理に近い変換である。sigmoid 関数は近似的に正規分布の累積分布関数に相当するので正規分布に従って確率的に変動する閾値をどれだけ通過するかを意味する。式 (5) は sigmoid 関数に $u(x)$ を乗じた関数のため、顕著なピークは保持したまま、誤差成分や負の成分を抑圧することができる。このように確率的な閾値処理とするのは、 $u(x)$ (標本値) の確率密度分布に基づいて α, β を学習して誤差成分を効率的に除去する方法を見越したためであるが、本稿では最初の段階として α, β は実験で定める定数とする。

Step2: LU 分解による共通調波構造パターン最適化

共通調波構造パターンを, 各成分 a_n ($n = 1, \dots, N$) をパラメータとする関数

$$\begin{aligned} \bar{h}(x, a) &= \delta(x) + a_1 \delta(x - b_1) + \dots + a_N \delta(x - b_N) \\ &= \sum_{n=0}^N a_n \delta(x - b_n) \end{aligned} \quad (6)$$

とする (図 6). ただし n は調波成分の次数, b_n は定数とする. また, $a_0 = 1, b_0 = 0$ とする. 前節の非線形変換から得た $\bar{u}(x)$ と $\bar{h}(x, a)$ の畳み込みによって得る生成多重音パワースペクトル $v(x, a)$ も $a = \{a_1, \dots, a_N\}$ をパラメータとする関数となる. ここで, 観測パワースペクトル $\tilde{v}(x)$ と $v(x, a)$ との二乗誤差の積分

$$\sum_{i=0}^I \{ \tilde{v}(x_i) - v(x_i, a) \}^2 \quad (7)$$

を最小化するパラメータ $\bar{a} = \{\bar{a}_1, \dots, \bar{a}_N\}$ を求めたい. 以後, 対数周波数 x は実際の計算機上の扱いを踏まえて離散値 x_i ($i = 0, \dots, I-1$) で表すことにする. I はサンプル数を表す. 式 (7) を a について偏微分して 0 と置いて整理すると

$$\begin{aligned} \frac{\partial}{\partial a} \sum_{i=0}^I \left\{ \sum_{n=0}^N a_n u(x_i - b_n) \right\}^2 \\ = 2 \frac{\partial}{\partial a} \sum_{i=0}^I \tilde{v}(x_i) \left\{ \sum_{n=0}^N a_n u(x_i - b_n) \right\} \end{aligned} \quad (8)$$

を得る. 各成分 a_n ($n = 1, \dots, N$) で偏微分すると,

$$\begin{pmatrix} \gamma_{1,1} & \dots & \gamma_{1,n} & \dots & \gamma_{1,N} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \gamma_{n,1} & \dots & \gamma_{n,n} & \dots & \gamma_{n,N} \\ \vdots & \ddots & \vdots & \ddots & \vdots \\ \gamma_{N,1} & \dots & \gamma_{N,n} & \dots & \gamma_{N,N} \end{pmatrix} \begin{pmatrix} a_1 \\ \vdots \\ a_n \\ \vdots \\ a_N \end{pmatrix} = \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ \vdots \\ \lambda_N \end{pmatrix} \quad (9)$$

$$\gamma_{j,k} = \sum_{i=0}^I u(x_i - b_j) u(x_i - b_k) \quad (10)$$

$$\lambda_j = \sum_{i=0}^I \{ \tilde{v}(x_i) - u(x_i) \} u(x_i - b_j) \quad (11)$$

のような連立 1 次方程式の形となるので, LU 分解で左辺の第一項の逆行列を求めて, 最適パラメータ \bar{a} を得ることができる. 新たに更新された共通調波構造パターン $\bar{h}(x, \bar{a})$ を $h(x)$ とし, 観測スペクトル $\tilde{v}(x)$ との逆畳み込み (Specmurt 法) により基本周波数パターン $u(x)$ を再度計算し, Step1 の非線形写像に戻る. この反復操作手順の模式図を図 7 に示す.

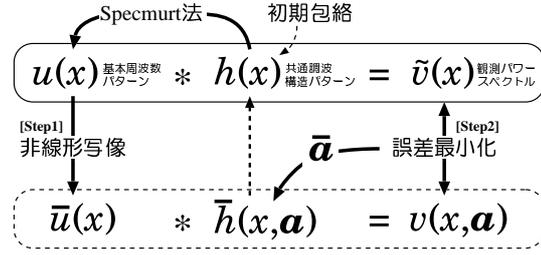


図 7 $h(x), u(x)$ の反復更新の模式図

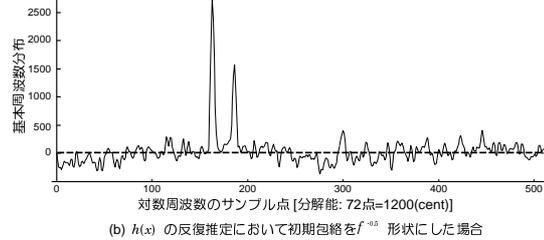
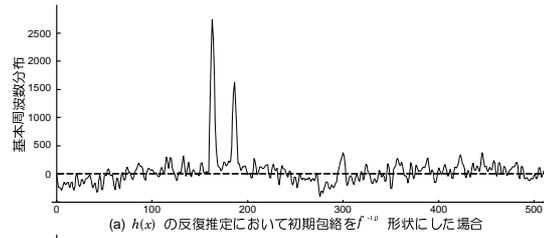


図 8 共通調波構造パターンの初期包絡を $f^{-1.0}$ (a), $f^{-0.5}$ (b) として 5 回の反復計算後に出力される基本周波数分布

提案方法を用いて図 4(b), (c) に示した $u(x)$ の出力結果から 5 回の反復計算を経て更新した $u(x)$ を図 8(a), (b) に示す. 初期 $u(x)$ に関係なく最終的にはほぼ同一の $u(x)$ が出力されており, いずれも C4 と E4 に対応する 2 つの成分が適切に検出されている. 特に図 4(b) では C4 の 2 次調波成分が大きく残っていたが, 反復推定後にはほとんど除去されている.

この方法により決定される共通調波構造パターンは多重音中の各音の中間音色としての意味をもつため, 提案方法は, 異なるスペクトル包絡をもつ音源による多重音に対して鮮明な基本周波数分布を出力する上で, バランスの良い $h(x)$ を見つけ出すための手段である.

3.2 MIDI への自動変換

得られた $u(x)$ の時系列から, 以下のような手順により MIDI データに自動変換することができる.

Step1: 一定閾値を超える $u(x)$ のピークを抽出する.

Step2: 各ピークの基本周波数を最も近い音階の基本周波数 (A4: 440Hz) に量子化する.

Step3: 55 音 (A1 ~ C6) に対応する 55 個の隠れマルコフモデル (Hidden Markov Model; HMM) を用意する. 各 HMM は発音状態と無音状態の 2 状態か

らなり、基本周波数が量子化されたピークの集合を観測系列とした確率モデルである。独立に 55 個の HMM の状態遷移系列を Viterbi アルゴリズムで求め、各音階の発音開始時刻（ノートオン）と発音終了時刻（ノートオフ）を推定する。

Step4: Step3 の結果を MIDI 形式でデータ出力する。

以下の実験評価では、この手順により得られた MIDI データを基に性能評価した。

4. 評価実験

4.1 実験条件

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¹⁴⁾ に収録されている合計 7 トラック（ジャズ曲 6 トラック、クラシック曲 1 トラック）の単一種類の楽器による多重奏（ピアノソロ、ギターソロ）音響信号をそれぞれ 16kHz にリサンプリングし、モノラルに変換したものとした。周波数解析などの詳細な実験条件を表 1 に示す。 $h(x)$, $u(x)$ の反復更新の回数は毎フレーム 5 回とした。

表 1 実験条件

周波数解析	サンプリング周波数	16 kHz
	フレーム周期	32 msec
	ウェーブレット関数	Gabor 関数
	分散	30 cent
	周波数分解能	≈16.7 cent
	周波数帯域	55–7604 Hz
$h(x)$	初期包絡	$f^{-1.5}$ 形状
	調波成分の数 N	10

4.2 音高可視化結果

基本周波数分布 $u(x)$ の時系列を濃淡表示することで得られる音高可視化結果の例を、スペクトログラムの濃淡表示および手入力による MIDI 信号のピアノロールウィンドウ表示 (XGworks) とともに図 10, 11, 12 に示す。それぞれの図において中段の音高可視化結果を見てみると、上段に示されている入力スペクトログラムから調波成分が効果的に抑圧されていることが確認でき、下段の参照用ピアノロールウィンドウ表示に類似した表示となっていることが分かる。

また、図 9 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可視化結果の比較を示す。左図では、誤差成分や除去されずに残った調波成分により、実際の演奏とかけ離れた可視化結果となったが、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。

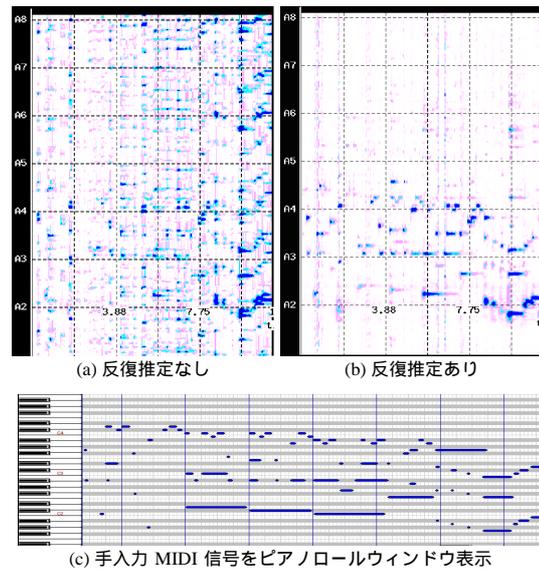


図 9 $h(x)$ を固定 (包絡形状が $f^{-0.5}$) とした場合 (a) と反復推定を行った場合 (b) との比較。両者ともに、実験データ “Jive (Piano solo)” 同一部分の可視化結果。

4.3 MIDI 自動変換結果

可視化結果がどの程度正確に得られているかを確認するために、得られた $u(x)$ の時系列から 3.2 節で述べた手順により MIDI に自動変換を行い、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 2 に示す。

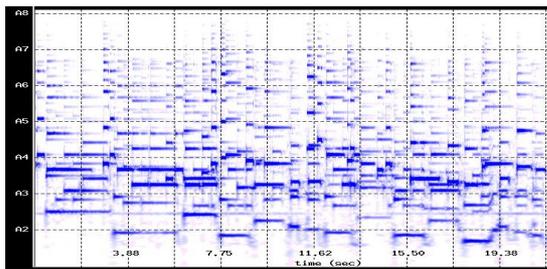
全 7 トラックの実演奏音響信号に対し、全体として 70 ~ 80% の音高正解率を得た。主な誤りは、Specmurt 法の原理的な性質上生じる誤りである調波成分を除去しきれないことによる挿入誤り、オクターブ重音を 1 音と見なしてしまうことによる脱落誤りなどであった。また、ピアノは非調和性楽器であることが知られており、調和性を仮定した共通調波構造パターンを用いていることが誤差成分の発生に影響し、上記以外の誤りに反映されている可能性がある。

5. まとめと展望

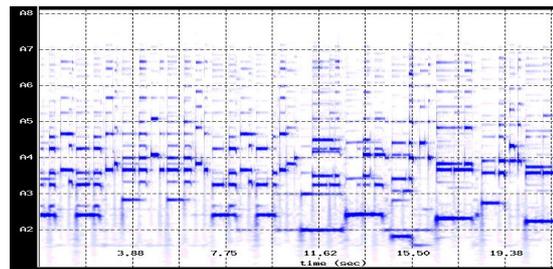
本報告では、基本周波数分布を sigmoid 関数を用いた非線形変換により更新するステップと変換された基本周波数分布をもとにパラメトリックに表現された共通調波構造パターンを最適化するステップからなる共通調波構造パターンの反復推定方法を提案し、我々がこれまでに提案した Specmurt 法の発展を試みた。ま

表 2 各実験データに対する音高正解率

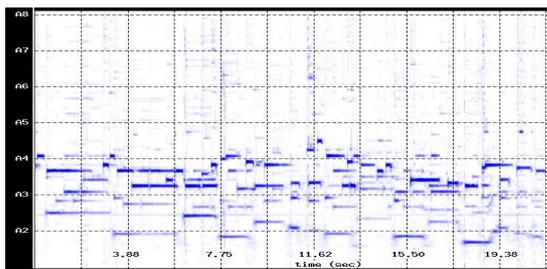
実験データ				MIDI 変換 正解率 (%)
タイトル	作曲者	ジャンル	楽器	
Jive (Piano Solo)	中村真	ジャズ	ピアノ	77.8
Lounge Away (Piano Solo)	永井隆雄	ジャズ	ピアノ	78.4
Jive (Guitar Solo)	中鉢洋夫	ジャズ	ギター	77.6
For Two (Guitar Solo)	中鉢洋夫	ジャズ	ギター	76.9
Crescent Serenade (Guitar Solo)	山本聡	ジャズ	ギター	74.5
Abyss (Guitar Solo)	中鉢洋夫	ジャズ	ギター	72.0
ノクターン第 2 番変ホ長調 op.9-2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



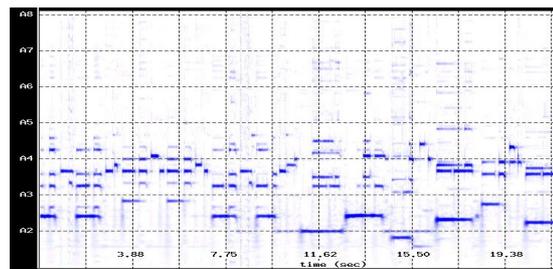
(a) 実音楽信号から切り出したスペクトログラム



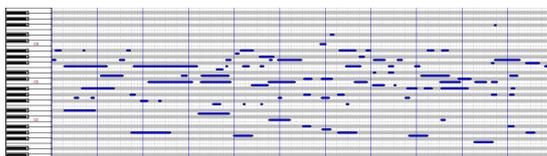
(a) 実音楽信号から切り出したスペクトログラム



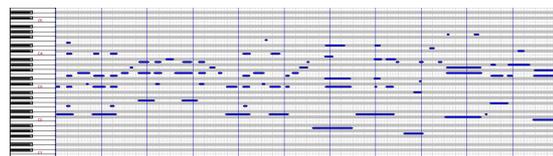
(b) 反復推定後の Specmurt 法による基本周波数パターンの時系列



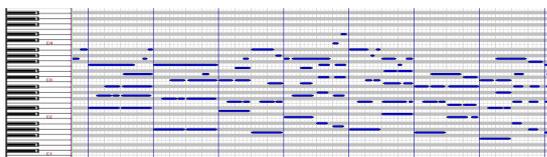
(b) 反復推定後の Specmurt 法による基本周波数パターンの時系列



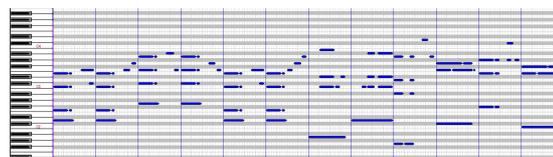
(c) MIDI 自動変換結果のピアノロールウィンドウ表示 (XGworks)



(c) MIDI 自動変換結果のピアノロールウィンドウ表示



(d) 手入力 MIDI 信号のピアノロールウィンドウ表示



(d) 手入力 MIDI 信号のピアノロールウィンドウ表示

図 10 “For Two (Guitar Solo)¹⁴⁾” の実演奏音響信号を入力とした音高可視化結果と参照データとの比較

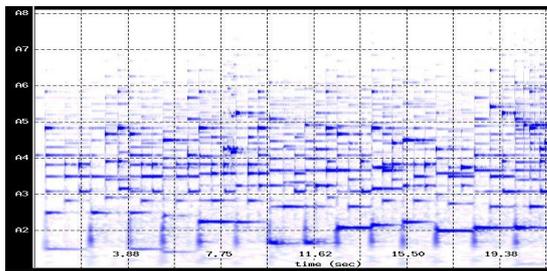
図 11 “Crescent Serenade (Guitar Solo)¹⁴⁾” の実演奏音響信号を入力とした音高可視化結果と参照データとの比較

た、単一種類の楽器による実演奏データを対象に可視化実験と MIDI 変換実験を行い、その効果と性能を確認した。評価実験により Specmurt 法では、単一種類の楽器による多重奏に対しては 70 ~ 80% 前後の音高正解率が得られることが確認できた。

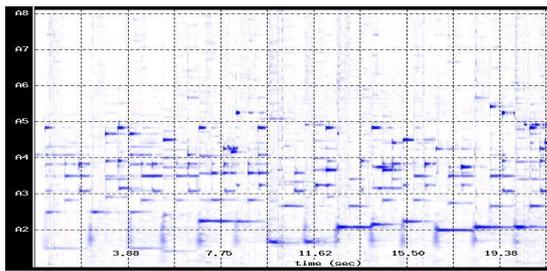
今後は、実用性の高い MIDI 自動変換や音源分離に向けて、タブレット PC などで可視化結果を簡単に

手で修正でき、その結果と別の多重音解析手法^{11),12)}を用いてより高精度な処理や解析ができるインタラクティブツールを開発していきたい。

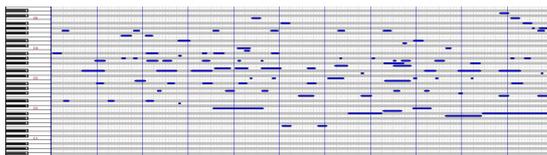
謝辞 本研究は、同研究室に所属した高橋佳吾氏(現警察庁勤務)の修士課程における研究を発展させたものである。本研究を進めるにあたり、同氏が作成したソースコードの一部を利用した。



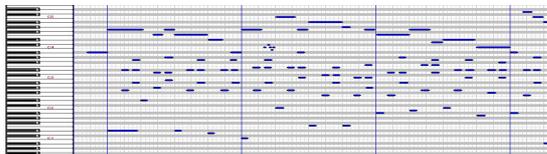
(a) 実音楽信号から切り出したスペクトログラム



(b) 反復推定後の Specmurt 法による基本周波数パターンの時系列



(c) MIDI 自動変換結果のピアノロールウィンドウ表示



(d) 手入力 MIDI 信号のピアノロールウィンドウ表示

図 12 “ノクターン第 2 番ホ長調 op.9-2¹⁴⁾” の実演奏音響信号を入力とした音高可視化結果と参照データとの比較

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.
- 2) S. Sagayama, K. Takahashi, H. Kameoka, T. Nishimoto, “Specmurt Anasylis: A Piano-Roll-Visualization of Polyphonic Music Signal by Deconvolution of Log-Frequency Spectrum,” Proc. ISCA Tutorial and Research Workshop on Statistical and Perceptual Audio Processing (SAPA2004), to appear, 2004.
- 3) 三輪多恵子, 田所嘉昭, 斎藤努: “くし形フィルタを利用した採譜のための異楽器音中のピッチ推定,” 電子情報通信学会論文誌, Vol. J81-D-II, No. 9, pp. 1965–1974, 1998.
- 4) 柏野邦夫, 木下智義, 中臺一博, 田中英彦: “音楽情景分析の処理モデル OPTIMA における和音の認

識,” 電子情報通信学会論文誌, Vol. J79-D-II, No. 11, pp. 1762–1770, 1996.

- 5) K. Kashino, K. Nakadai, T. Kinoshita and H. Tanaka, “Organization of Hierarchical Perceptual Sounds: Music Scene Analysis with Autonomous Processing Modules and a Quantitive Information Integration Mechanism,” Proc. International Joint Conference on Artificial Intelligence (IJCAI), Vol. 1, pp. 158–164, 1995.
- 6) S. Godsill and M. Davy, “Bayesian Harmonic Models for Musical Pitch Estimation and Analysis,” Proc. IEEE 27th International Conference on Acoustics, Speech and Signal Processing (ICASSP2002), Vol. 2, pp. 1769–1772, 2002.
- 7) A. Klapuri, T. Virtanen and J. Holm, “Robust Multipitch Estimation for the Analysis and Manipulation of Polyphonic Musical Signals,” Proc. COST-G6 Conference on Digital Audio Effects, pp. 233–236, 2000.
- 8) T. Virtanen and A. Klapuri, “Separation of Harmonic Sounds Using Linear Models for the Overtone Series,” Proc. IEEE 27th International Conference on Acoustics, Speech and Signal Processing (ICASSP2002), Vol. 2, pp. 1757–1760, 2002.
- 9) 後藤真孝: “音楽音響信号を対象としたメロディーとベースの音高推定,” 電子情報通信学会論文誌, D-II, Vol. J84-D-II, No. 1, pp. 12–22, 2001.
- 10) M. Goto: “A Predominant-F0 Estimation Method for CD Recordings: MAP Estimation Using EM Algorithm for Adaptive Tone Models,” Proc. IEEE 26th International Conference on Acoustics, Speech and Signal Processing (ICASSP2001), pp. V-3365–3368, 2001.
- 11) H. Kameoka, T. Nishimoto and S. Sagayama, “Extraction of Multiple Fundamental Frequencies from Polyphonic Music,” Proc. 18th International Congress on Acoustics (ICA2004), Mo2.C1.3, 2004.
- 12) H. Kameoka, T. Nishimoto and S. Sagayama, “Separation of Harmonic Structures Based on Tied Gaussian Mixture Model and Information Criterion for Concurrent Sounds,” Proc. IEEE 29th International Conference on Acoustics, Speech and Signal Processing (ICASSP2004), AE-P5.9, 2004.
- 13) M. Goto, H. Hashiguchi, T. Nishimura and R. Oka, “RWC Music Database: Music Genre Database and Musical Instrument Sound Database,” Proc. International Conference on Music Information Retrieval (ISMIR2003), pp. 229–230, 2003.
- 14) M. Goto, H. Hashiguchi, T. Nishimura and R. Oka, “RWC Music Database: Popular, Classical, and Jazz Music Database,” Proc. International Conference on Music Information Retrieval (ISMIR2002), pp. 287–288, 2002.