

## 事後確率最大化 Specmurt 分析による多重ピッチの 反復推定アルゴリズム

齊藤 翔一郎<sup>†</sup> 亀岡 弘和<sup>†</sup>  
小野 順貴<sup>†</sup> 嵯峨山 茂樹<sup>†</sup>

本稿では、音楽音響信号の多重ピッチ推定を行う手法である Specmurt 分析を MAP 推定の観点から定式化し、事後確率を最大化することによって基本周波数分布と共通調波構造パターンを求めるアルゴリズムについて述べる。我々はこれまでに非線型写像を用いて基本周波数分布に関する先験情報を利用しつつ共通調波構造パターンを推定する手法を提案してきた。この手法は直感的には理解がしやすいが、アルゴリズムが何を目的関数としているかなどの見通しが立ちにくく、アルゴリズムの収束も保証されなかった。そこで我々は Specmurt 分析における多重ピッチ推定を事後確率を最大化する推定として定式化し直し、今までの反復推定に新たな解釈を加えるとともに、そこで明らかになった問題を踏まえ事後確率を最大化するような新たなアルゴリズムを提案する。評価実験によって平均で 64.11% の MIDI 変換正解率を得た。

### Iterative Multipitch Estimation Algorithm for MAP Specmurt Analysis

SHOICHIRO SAITO,<sup>†</sup> HIROKAZU KAMEOKA,<sup>†</sup> NOBUTAKA ONO<sup>†</sup>  
and SHIGEKI SAGAYAMA <sup>†</sup>

This paper describes a iterative algorithm for estimating a fundamental frequency distribution and a harmonic structure pattern by reformulating Specmurt analysis form a standpoint of maximum a posteriori probability. We have proposed specmurt Analysis which calculate the fundamental frequency distribution of multi-pitch music signals by iterative estimation algorithm using non-linear mapping function based on prior information. This iterative algorithm is easy to understand, but it is not obvious what the algorithm considers as an objective function, and the convergence is not ensured. Therefore, we make a explanation of the algorithm from the standpoint of maximum a posteriori probability, and propose a new algorithm based on MAP estimation. Evaluation result shows an avarage 64.11% accuracy of WAV to MIDI conversion.

#### 1. はじめに

本稿では、音楽音響信号から音高情報を推定し可視化するための技術である Specmurt 分析について、事後確率最大の観点から最適な分布の推定を行う手法について報告する。我々の身の回りには音響信号としての音楽が溢れているが、これを音高情報という一種のシンボリックな情報へ自動変換することが出来れば様々な応用が見込める。例えば、携帯電話の着信メロディやカラオケ楽曲などはその多くが人手で作成されているが、経験や労力を必要とすることもあり、これらが元楽曲の音響信号から簡単に生成することができ

ば非常に有用である。また近年のインターネットの急速な発展、ポータブルミュージックプレイヤーの普及に伴い、Web 上、もしくは個人のプレイヤーの中の楽曲から所望の曲を検索、ということがユーザーのニーズとして大きなものになっていくと考えられ、データベース上の膨大な楽曲に対して検索をかけることが出来れば望ましいことである。音高推定技術はそれらの目的を達成するために必要な技術であると言える。

しかし、多重音の音響信号からの音高情報の抽出は、楽音の調波成分の影響や信号の非正常性、短時間解析におけるスペクトルの広がりなどが原因で、困難であることが多い。音楽の信号領域でのモデル化は例えば Godsill<sup>1)</sup>の研究があり、信号をパラメトライズされた様々な周波数の正弦波の重ね合わせとし、Bayes 推定と MCMC 法を用いてピッチパラメータを推定する手

<sup>†</sup> 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.

法を提案している。またパワースペクトルでのモデル化手法としては、たとえば後藤らは多重音の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の重みつき和としてモデル化し、重みパラメータをEM アルゴリズムによって推定する手法を提案している<sup>2),3)</sup>。また我々は、多重音のスペクトログラム全体を調波構造をもった複数の音響オブジェクトの和によってモデル化し、時間周波数平面での分布距離最小の観点から各音響オブジェクトの調波構造およびパワーエンベロープのパラメータを最適推定する手法を提案している<sup>4)</sup>。これらの手法は最適パラメータを反復推定によって求めるものであるが、初期値によっては予測不能な誤り解におちいる可能性も持つ。

そこで我々は、スペクトログラムから楽音の調波成分を抑圧することで基本周波数成分をノンパラメトリックに「強調」するというアプローチを提案してきた。このアプローチの利点はピッチを一意に「推定」しないので、ピッチを確定的に決めない代わりに予測不能な誤りを大幅に低減することが出来ることである。パラメトリックな手法の多くは音源数(同時発音数)をあらかじめ与えることでモデルが計算されるため、音源数の推定を誤るとその後の推定も誤ってしまう傾向があるが、ノンパラメトリックな手法では特定の音源数を仮定する必要がないのでそのような問題は起きない。その上、もちろん多くの手法に見られるような解を一意に「推定」することも閾値処理などによって可能である。この強調された結果は時間周波数平面で濃淡表示が可能であるため、推定が困難な部分であっても濃淡の具合から人間がある程度の候補のうちから推定することが出来るのである。

このようなアプローチは、ユーザー介入型のインタラクティブツールとの親和性が高いとも言える。ユーザーはこのツールを用いてまず分析を行い、大筋で正しい推定結果を得た後、自分の感覚や知識から推定が誤っている部分だけを手直しすることで、正確な音高推定を完成させることが出来る。

我々はこの「強調」ないしは「可視化」のアプローチとして、これまで Specmurt 分析という手法を提案し<sup>5)</sup>、またその分析に用いる共通調波構造パターンの反復推定アルゴリズムを考案した<sup>6)</sup>。この反復推定アルゴリズムは基本周波数分布や調波構造が満たすべき条件(先験情報)を用いて共通調波構造パターンを反復計算によって求めるものである。本稿ではこれらの推定を「事後確率最大化」の問題ととらえ直し、従来の反復推定法に確率論的な解釈を加えると共に、事後確率を最大化する為の新しいアルゴリズムを提案する。

本稿の構成は以下のようなものである。まず、2章で Specmurt 分析の概要について述べ、多重ピッチ推定の問題

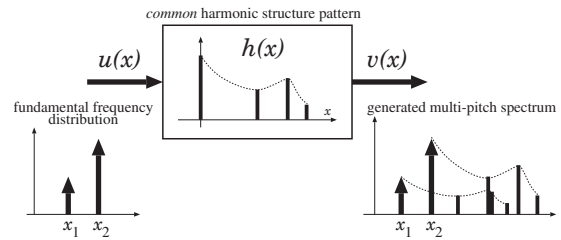


図1 基本周波数分布と共通調波構造の畳み込みによる多重音スペクトルの生成モデル

を具体的に定式化する。また、ここで従来の反復推定アルゴリズムについても述べる。そして3章で反復推定アルゴリズムに確率論的解釈を与え、4章で新しい反復推定アルゴリズムを提案する。その性能を5章で評価し、6章でまとめと展望について述べる。

## 2. 問題の定式化と従来法

### 2.1 Specmurt 分析の概要

一般に楽音は、音高に相当する基本周波数以外に倍音を多く含んでおり、それが音色を構成している。この調波構造の倍音パワー比が基本周波数に依らず共通であると仮定すると、調波構造を持つ音の基本周波数がシフトした場合、線形周波数領域では倍音の位置関係は線形伸縮となる。一方、対数周波数領域で考えれば、倍音の位置関係も一定で調波構造全体がシフトする。ここで、基本周波数分布を  $u(x)$ 、対数周波数軸上の倍音位置に(のみ)倍音強度比の成分を持つ分布(以後、共通調波構造パターンと呼ぶ)を  $h(x)$  とすると、多重音のスペクトル  $v(x)$  を

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

と畳み込みによってモデル化することができる(図1,  $x$  は対数周波数)。  $h(x)$ 、  $v(x)$  の逆フーリエ変換をそれぞれ  $H(y)$ 、  $V(y)$  と定義すると、フーリエ変換の畳み込み定理から、基本周波数分布  $u(x)$  のフーリエ逆変換  $U(y)$  ( $y$  は対数周波数の逆フーリエ変換領域)は

$$U(y) = \frac{V(y)}{H(y)} \quad (2)$$

と表されるため、求めたい基本周波数分布  $u(x)$  は  $U(y)$  をフーリエ変換することによって得られる。このようにして基本周波数分布を求める手法を specmurt 分析<sup>5)</sup> という。

### 2.2 共通調波構造の自動推定問題

前節で述べた Specmurt 分析では、共通調波構造パターンは楽音の倍音構造としてある程度一般的なものをユーザーが事前に決定していたが、楽器の違いや時間の経過による調波構造パターンの変化に応じて逐一

適切なパターンを人手で調べることには限界がある。これを解決する一つの方法は、観測スペクトル  $v(x)$  から共通調波構造パターン  $h(x)$  を自動的に推定することであるが、式 (1) のモデルにおいては基本周波数分布  $u(x)$  も未知であるから、これは不良設定問題であり、何らかの先験情報や制約条件が必要となる。

この問題に対して仮定できる先験情報としては、例えば以下のようなものが考えられる。

- (A)  $u(x)$  もスペクトルを表すから非負であること
- (B)  $u(x)$  は基本周波数分布であるから、ある大きさ以上の鋭いピークから構成されていること
- (C) 共通調波構造  $h(x)$  は各倍音に相当する複数の  $\delta$  関数により以下のように表わされていること

$$h(x) = h(x; \mathbf{a}) = \sum_{n=1}^N a_n \delta(x - \log n) \quad (3)$$

ただし、 $N$  は調波成分の数を表し、 $a_n$  は各調波成分の強度比を表す非負の値である ( $a_1 = 1$ )。

一般にこのような問題を解く方法としては、先験情報によりコスト関数を定義し最小化する方法や、先験情報を満たすような何らかの反復を行なう方法などがある。前者はコスト関数の設計を慎重に行う必要があり、また多くの場合、局所最適解を得るまでに多数のステップが必要となる。一方、反復法はアルゴリズムがわかりやすく、高速な場合が多いが、必ずしも収束性や安定性が保証されるとは限らない。

### 2.3 従来の反復推定アルゴリズム

我々は前節の後者の解法、つまり先験情報を満たすような反復を繰り返す手法として、非線形写像を用いた反復推定法を以前に提案した<sup>6)</sup>。この手法は大きく分けて3つのステップで構成されており、これらを繰り返すことで  $u(x)$ 、 $h(x)$  が準最適に求まるというものである。

- (I) 逆畳み込み  $u(x) = v(x) * h^{-1}(x)$  により  $u(x)$  を求める。

- (II) 非線形写像

$$\bar{u}(x) = \frac{u(x)}{1 + \exp\{-\alpha(u(x) - \beta)\}} \quad (4)$$

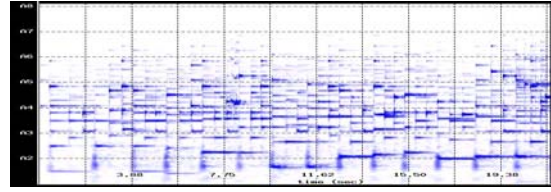
により  $u(x)$  を更新。

- (III)  $\|v(x) - u(x) * h(x; \mathbf{a})\|^2$  を最小にするような  $\mathbf{a}$  を

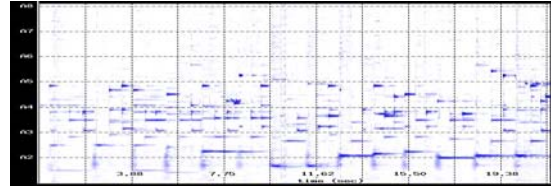
$$\frac{\partial}{\partial \mathbf{a}} \sum_x (v(x) - u(x) * h(x; \mathbf{a}))^2 = 0 \quad (5)$$

によって求め、 $h(x)$  を更新。

式 (4) は  $u(x)$  の大きな正の成分をそのまま残しつつ、負の成分および小さな正の成分を0へ近づける働きがあり、 $\alpha$  と  $\beta$  はそれぞれ抑圧のファジーネスと閾値を規定するパラメータとなっている。また式 (5) の



(a) スペクトログラム



(b) specmurt 分析結果

図 2 specmurt 分析による倍音抑圧

求解においては  $a_n < 0$  となることも許容している。<sup>\*</sup>

ここで2.2の先験情報と照らしあわせると、A, Bの先験情報が(II)において、Cの先験情報が(III)において反映されていることがわかる。この反復推定によって基本周波数の強調性能が向上するとともに出力結果の共通調波構造パターンの初期値に対する依存性が大幅に減少した。RWC 研究用音楽データベース<sup>10)</sup>のクラシック No.30(Nocturne Op. 9, No.2, ショパン)について specmurt 分析を行った結果を図2に示す。

## 3. 非線形写像による反復推定の確率論的解釈

### 3.1 事後確率最大化としての Specmurt 分析

前節では、Specmurt 分析において未知数  $u(x)$  と  $h(x)$  の満たすべき条件を順に適用していくことで所望の分布を得るアルゴリズムを述べた(以降このアルゴリズムを「従来の反復推定法」と表現することにする)。この方法は直感的には何をしているのか理解がしやすいが、アルゴリズムとしての終着を何に設定しているかの見通しはいまひとつ明確でない。そこでこの節では、Specmurt 分析を「事後確率を最大化する」という目的をもって行うことを考えることにする。

今考えている問題は、観測スペクトル  $v(x)$  が得られている状態で、基本周波数分布  $u(x)$  と共通調波構造パターン  $h(x)$  を求めるというものである。ここで、観測スペクトル  $v(x)$  を

$$v(x) = u(x) * h(x) + n(x) \quad (6)$$

とする。 $n(x)$  は誤差項であり、雑音成分および調波構造のピッチによる違いから現れる畳み込みモデルから

<sup>\*</sup> これは主に調波成分の数を実際より多く見積もった場合に高調波において微少な負の値をとるという場合において発生するが、実際にはほとんど影響がないことを実験において確かめている。

の逸脱量を含む。解くべき問題は

$$\operatorname{argmax}_{h(x), u(x)} \prod_x p(h(x), u(x)|v(x)) \quad (7)$$

である。これは

$$\begin{aligned} & \operatorname{argmax}_{h(x), u(x)} \prod_x p(h(x), u(x)|v(x)) \\ &= \operatorname{argmax}_{h(x), u(x)} \prod_x p(v(x)|h(x), u(x)) \cdot p(u(x), h(x)) \\ &= \operatorname{argmax}_{h(x), u(x)} \prod_x p(v(x)|h(x), u(x)) \cdot p(u(x)) \cdot p(h(x)) \\ &= \operatorname{argmax}_{h(x), u(x)} \sum_x \log p(v(x)|h(x), u(x)) \\ & \quad + \log p(u(x)) + \log p(h(x)) \quad (8) \end{aligned}$$

と変形できる。よって

$$J \triangleq \sum_x \log p(v(x)|h(x), u(x)) + \log p(u(x)) + \log p(h(x)) \quad (9)$$

とすると、 $J$  を最大化することが事後確率最大化になっている。

### 3.2 従来手法の解釈

記号を

$$J_1 \triangleq \sum_x \log p(v(x)|h(x), u(x)) \quad (10)$$

$$J_2 \triangleq \sum_x \log p(u(x)) \quad (11)$$

$$J_3 \triangleq \sum_x \log p(h(x)) \quad (12)$$

と定義する。

ただし、 $J_3$  は共通調波構造の事前確率のため、定数と見なしてもよい。

今、ノイズ項にガウス分布を仮定すると、モデル  $u * h$  からゆらぎを持って  $v$  が観測される確率としての  $p(v|u, h)$  は

$$p(v|u, h) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(v - u * h)^2}{2\sigma^2}\right) \quad (13)$$

となり、 $J_1$  の項は以下のように書ける：

$$J_1 = -\frac{(v - u * h)^2}{2\sigma^2} - \log(\sqrt{2\pi}\sigma) \quad (14)$$

従来の反復推定法の (I) では  $u = v * h^{-1}$  によって  $u$  を更新するので、この  $u$  は  $J_1$  の最大化において明らかに大域最適解である。そして、この計算の段階において  $h^{-1}(x)$  のフィルタとしての安定性が保証されず、 $u(x)$  の推定に支障をきたすが稀にあったことも留意しておく。

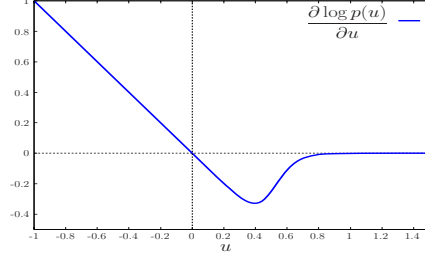


図 3  $\frac{\partial \log p(u)}{\partial u}$  の概形

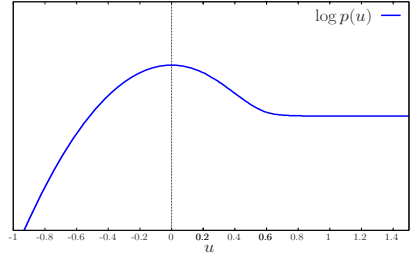


図 4  $\log p(u)$  の概形

次に、(II) における更新を考える。

式 (4) は

$$u^{(t+1)} = u_{(1)}^{(t)} + \frac{-u_{(1)}^{(t)} \exp\{-\alpha(u_{(1)}^{(t)} - \beta)\}}{1 + \exp\{-\alpha(u_{(1)}^{(t)} - \beta)\}} \quad (15)$$

とも書ける。 $u_{(1)}^{(t)}$  は  $t$  回目の反復の (I) で更新された  $u$  とする。この式を、学習係数 1 の最急降下法を用いて  $J_2$  を大きくする更新であると解釈すると

$$u^{(t+1)} = u_{(1)}^{(t)} + \frac{\partial J_2}{\partial u_{(1)}^{(t)}} \quad (16)$$

となる。この式は  $J_1$  部分の最大化を終えてから  $J_2$  に関する最急降下法を行うもので、現在考えている目的関数  $J$  を最大化するものにはなっていないが、このように表記すると (II) は

$$\frac{\partial \log p(u)}{\partial u} = \frac{-u \cdot \exp\{-\alpha(u - \beta)\}}{1 + \exp\{-\alpha(u - \beta)\}} \quad (17)$$

となるような事前分布  $p(u)$  を仮定して最急降下法を用いたと解釈できる。

$\alpha = 15.0, \beta = 0.5$  の場合の  $\frac{\partial \log p(u)}{\partial u}$  と  $\log p(u)$  の例を図 3, 4 に示す。ただし  $\frac{\partial \log p(u)}{\partial u}$  は解析的に積分が不可能なため、あくまでおおまかな概形を描いている。 $\frac{\partial \log p(u)}{\partial u}$  は  $u < 0$  で正、 $u > 0$  で負であり原点を通るため、 $\log p(u)$  は  $u = 0$  で最大値をとる。

最後に、(III) の更新について考える。ここでは前の 2step での  $u$  の更新を受けて、 $h$  に関する  $J_1$  の最大化を行っていることに相当する。具体的には偏微分の式が 0 であるという連立方程式解くことで極値を求めて

いる。

以上をまとめると、従来の非線型写像を用いた反復推定法は、モデルの誤差の確率分布にガウス分布を仮定した上で、

- (I)  $u$  に関する  $J_1$  の最大化
  - (II)  $u$  に関する  $J_2$  における最急降下法
  - (III)  $h$  に関する  $J_1$  の最大化
- ということを行っていたと解釈できる。

#### 4. 事後確率最大化 Specmurt 分析

##### 4.1 提案する反復推定法

以上で見てきたように、従来の反復推定法は MAP 推定の観点から解釈することができ、同時に事後確率の意味での目的関数を必ずしも最大化しないことも分かった。よって我々は観点からこのアルゴリズムを再考することによって、勾配法の意味で収束をさせつつ高い性能を持つアルゴリズムを提案することを目指す。

式 (7) で表される事後確率を最大化する目的において、目的関数を

$$J = J_1 + J_2 \quad (18)$$

とする。ただし

$$J_1 = -\frac{1}{2\sigma^2} \sum_x (v(x) - u(x) * h(x))^2$$

$$= -\frac{1}{2\sigma^2} \sum_x \left( v(x) - \sum_{\tau} u(x - \tau) h(\tau) \right)^2 \quad (19)$$

$$J_2 = \int \frac{-u \exp(-\alpha(u - \beta))}{1 + \exp(-\alpha(u - \beta))} du \quad (20)$$

であり、尤度確率の分散  $\sigma^2$  は尤度と事前分布の重みを決めるパラメータとなっている。

目的関数を最大にする  $u(x)$  は陽には求まらないが、その微分係数が既知であるため、最急降下法によって目的関数の最大化を図る。また、目的関数の  $J_2$  の項に関しては  $h$  に依存しないので、 $h$  に関して目的関数を最大化するには従来と同じく式 (5) を解くことで最適解を求めればよい。

以上から、目的関数を最大化するには、

- (1) 最急降下法によって  $u$  に関して  $J$  を最大化
  - (2) 式 (5) を解いて  $h$  に関して  $J$  を最大化
- の 2 つの step を交互に繰り返せばよい。更新による目的関数の最大化を概念的に表したものを図 5 に示す。ただし、図の  $u(x)$  と  $h(x)$  の軸はそれぞれの  $x$  に対するスカラーの値を示すわけではなく、分布の bin 数である  $I$  次元のベクトルとしての  $u$  や  $h$  の組合せが敷き詰められているようなイメージである。(正確に描くと  $2I$  次元平面が必要になるため、あくまで概念図である)

なお、 $J_1$  は明らかに上に有界な関数であり、 $J_2$  の導

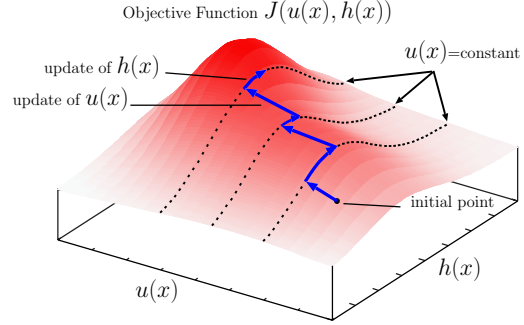


図 5 提案法による目的関数最大化の概念図。  $u(x)$  に関しては勾配の方向へ更新し、次に  $u(x)$  を固定した断面について最大になる  $h(x)$  を求める。

関数は正から負へ一度だけ入れ替わるためやはり明らかに最大値を持ち、上に有界である。

$u$  の具体的な更新則は、

$$u^{(t+1)} = u^{(t)} + A \left( \frac{\partial J_1}{\partial u} + \frac{\partial J_2}{\partial u} \right) \quad (21)$$

となる。  $A(> 0)$  は学習係数である。

第二項の微分は  $i$  番目の要素について

$$\frac{\partial J_1}{\partial u(i)} = 2 \sum_x h(x - i) \left( v(x) - \sum_{\tau} u(x - \tau) h(\tau) \right) \quad (22)$$

と書き下せ、共通調波構造が

$$h(x) = \sum_n a_n \delta(x - b_n) \quad (23)$$

というデルタ列で定義されるとすると ( $b_n$  は倍音位置)

$$\frac{\partial J_1}{\partial u(i)} = 2 \sum_m a_m \left( v(i - b_m) - \sum_n a_n u(i - b_m - b_n) \right) \quad (24)$$

となる。一方第三項の微分は

$$\frac{\partial J_2}{\partial u} = \frac{-u \exp(-\alpha(u - \beta))}{1 + \exp(-\alpha(u - \beta))} \quad (25)$$

である。

$h$  の更新は  $J_2$  が  $h$  に依存しないことから従来の (III) と同じであるため、式 (5) を解いて更新を行う。

以上 2 つの step を繰り返すことで、事後確率を最大にする  $u(x)$  と  $h(x)$  が求まることになる。

反復の初期値  $u(x)$  は従来の反復推定法の (I) の逆畳み込みによって与えるとする。逆畳み込みを行うには逆フィルタとしての  $h^{-1}(x)$  が安定に働くことを保証する必要があるが、初期値として一度だけ行う分には問題ない。その際ユーザーは適当な共通調波構造を選択することが出来るが、安定性の観点から避けるべき  $h(x)$  の形状については文献<sup>7)</sup> に記述がある。また、

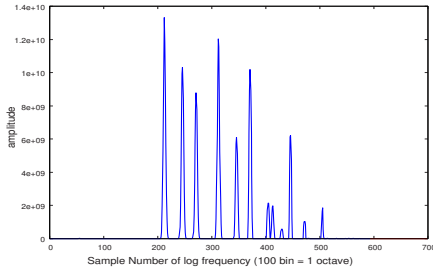


図6 ピアノの三和音 (C4,E4,G4) の観測スペクトル

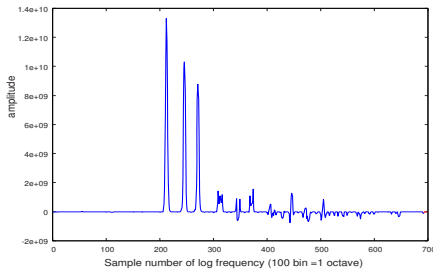


図7 提案手法によって基本周波数推定を行った結果 ( $\alpha = 15, \beta = 0.2, A = 0.9, \sigma^2 = 1.5$  で 10 回反復)

事前分布の  $\beta$  の値がデータ依存にならないよう、 $v(x)$  の値は最大値が 1 になるようにあらかじめ正規化しておき、分析後に元のスケールに戻す。

アルゴリズムの適用例を図 6,7 に示す。図 6 は RWC 研究用音楽データベースの楽器音データベース<sup>8)</sup> を用いて合成したピアノの三和音 (C4,E4,G4) のスペクトルであり、図 7 はそれに対し提案手法によって推定された基本周波数分布である。倍音が適切に抑圧されている様子が見られる。

#### 4.2 事前分布の役割

従来の非線型写像による反復推定と同じく、この提案手法の推定において主要な役割を担っているのは、観測スペクトル  $v(x)$  と Specmurt 分析におけるスペクトルのモデル  $u(x) * h(x)$  との二乗誤差を最小にする部分である。確率の表現を用いれば「ガウス分布の対数尤度を最大にする」推定の部分である。しかし、本手法ではそれに加えて、事前分布として  $x$  によらず全ての  $u(x)$  について 0 を最大値に持つような同一の分布を導入している。これにより、「対数尤度の意味で最適な値が 0 からそれほど離れていない場合、対数尤度を多少偽性にしても事前分布のペナルティを軽くするよう値を 0 へ持っていく」という効果が得られる。突き詰めれば、「出来るだけ少数の大きなピークで、観測スペクトルを最もうまく説明できるモデルを求める」ということが事前分布の導入によって可能になったことになる。つまり、2.2 の先験情報 B が従来とは別角度から達成されたとも言える。

表 1 実験条件

周波数解析	サンプリングレート	16kHz
	フレームシフト	16ms
	周波数分解能	12cent
	分析周波数	60Hz~7680Hz
Specmurt 分析	調波構造の初期値 $a_n$	$\frac{1}{n}$
	調波成分の数 $N$	8
	学習係数 $A$	0.9
	分散パラメータ $\sigma^2$	1.5

## 5. 評価実験

### 5.1 実験条件

前章で述べたアルゴリズムの有効性を示すための評価実験を行った。実験に用いたデータは、RWC 研究用音楽データベースのジャズとクラシック<sup>10)</sup> の一部を 20 秒強切り出した 8 つのデータである (曲の一覧を表 2 に示す)。これらのデータに対し、共通調波構造パターンを固定して反復推定を行わない Specmurt 分析、非線型写像を用いた従来の反復推定法、今回提案する反復推定法の 3 つについて基本周波数分布推定を行う。次に出力された基本周波数分布をある閾値で処理したあと、正解に相当するハンドラベリングされた MIDI データと比較し、フレームごとに ON/OFF の正誤判定を行い、正解率を計算する。正解率は  $X$  を発音があった全フレーム数、 $D$  を脱落誤り個数、 $I$  を挿入誤り個数、 $S$  を置換誤り個数として

$$\frac{X - D - I - S}{X} \times 100(\%) \quad (26)$$

のように計算した。

基本周波数分布の閾値処理ではデータや手法ごとに最適な閾値が異なるため、閾値を統一して性能を比較することは難しいと判断した。そこで、各データで様々な閾値を試行し、それぞれにおいてもっとも正解率の高かった値を理論上の最大性能という意味で比較することとした。また、同等の条件で、後藤の提案する preFEst、亀岡の提案する HTC とも性能を比較した。これらの実験においては、亀岡が製作した“MIDI Reference Editor”<sup>9)</sup> (図 8) を使用して性能を比較した。このシステムには HTC と preFEst\* の MIDI 変換正解率を計算する機能があるため、新たに提案する Specmurt 分析の手法とあわせて容易に性能比較を行うことが出来る。preFEst, HTC の実験条件および実験結果のデータについては文献<sup>9)</sup> を参照していただきたい。

実験に用いたパラメータの値を表 1 に示す。周波数解析にはウェーブレット変換を用いている。事前分布

\* ここで実装されている preFEst は「preFEst-core」というピッチ推定の部分のみであり、マルチエージェントによりピッチトラッキングを行う部分は実装されていない。

表 2 実験に使用した音楽データ (RWC 研究用音楽データベース<sup>10)</sup> より抜粋)

Symbol	タイトル (ジャンル)	作曲家/演奏者	Catalog number	楽器	平均音源数
data(1)	Crescent Serenade (Jazz)	S. Yamamoto	RWC-MDB-J-2001 No. 9	Guitar	2.13
data(2)	For Two (Jazz)	H. Chubachi	RWC-MDB-J-2001 No. 7	Guitar	2.67
data(3)	Jive (Jazz)	M. Nakamura	RWC-MDB-J-2001 No. 1	Piano	1.86
data(4)	Lounge Away (Jazz)	S. Yamamoto	RWC-MDB-J-2001 No. 8	Guitar	4.04
data(5)	For Two (Jazz)	M. Nakamura	RWC-MDB-J-2001 No. 2	Piano	2.34
data(6)	Jive (Jazz)	H. Chubachi	RWC-MDB-J-2001 No. 6	Guitar	1.78
data(7)	Three Gimnopedies no. 1 (Classic)	E. Satie	RWC-MDB-C-2001 No. 35	Piano	2.96
data(8)	Nocturne Op.9, No.2 (Classic)	F. F. Chopin	RWC-MDB-C-2001 No. 30	Piano	1.55

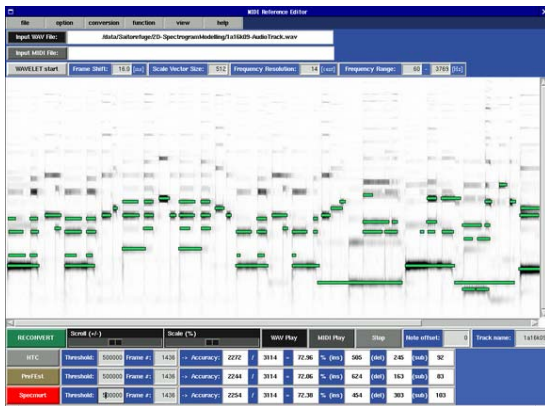
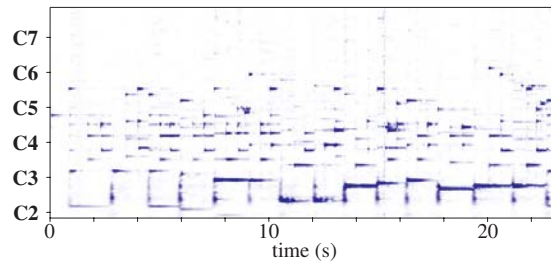


図 8 "MIDI reference editor." 閾値に応じた正解率を計算し表示することが出来る。

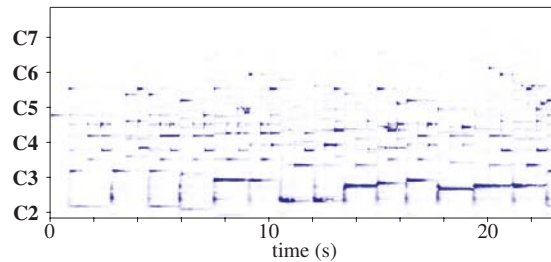
のパラメータ (従来法では非線型写像のパラメータ) は  $\alpha = 15.0$  とし,  $\beta$  を 0.2 と 0.5 の 2 通りについて実験を行った. 学習係数  $A$  と分散  $\sigma^2$  は予備的な実験のうちうまく動作するものを適当に選んで固定したが, この値の選択および反復中の更新にはなお検討の余地がある. また反復回数は目的関数値が解析的に得られないなどの理由から固定の値にしているが, こちらも今後の課題である.

## 5.2 実験結果

提案手法と従来の Specmurt 分析の MIDI 変換精度を比較した結果を表 3 に示す. 提案手法は以前から高い精度が得られることが確認できていた従来の反復法と同程度の性能を示し, 反復計算を行わない場合よりも大幅な改善が見られた. また推定結果が稀に発散してしまう従来法に対し提案法ではそのような現象は見られなかった (図 9). また, 我々は以前に凸射影法を用いてアルゴリズムの収束性を保証する枠組を提案した<sup>11)</sup> が, その際には収束性を優先させたため倍音抑圧性能が低下してしまったことを考えると, そのアルゴリズムより高い倍音抑圧性能が得られていると言える. 提案法において  $\beta = 0.5$  の場合に正解率が落ちているのは, 事前分布の傾斜を広い範囲にかけたため, 中程度のエネルギーのピッチまでもが不要成分と見なされ抑



(a) 従来の反復推定による可視化結果



(b) 提案する反復推定による可視化結果

図 9 data(8) における従来手法と提案手法の可視化結果の比較. 2 つはほぼ同じ推定結果となっているが, 提案法では従来法に見られる縦線 (推定誤り) がない.

圧されてしまったと考えられる.

また表 4 では preFEst と HTC の正解率との比較を行っている. 提案手法はこれら 2 手法に勝るとまではいかないまでも, ほぼ同程度までの性能は得ることが出来た.

## 6. まとめと展望

本稿では事後確率最大化による Specmurt 分析によって音楽音響信号から多重ピッチを推定する手法について述べた. 本研究の意義は, 直感的手法によって高い性能を示していた従来の反復推定法に確率論的アプローチからの解釈を与えたことに加え, そのような理論的枠組を与えたことで反復推定法を改善するための見通しが立ちやすくなったということにある. 本稿では最急降下法による目的関数最大化の方法を提案し, 推定結果が発散することなく従来の反復推定と同程度

表 3 提案手法と従来の Specmurt 分析の MIDI 変換精度の比較

	フレーム数 $X$	反復推定なし	従来の反復推定法		提案する反復推定法	
			$\beta = 0.2$	$\beta = 0.5$	$\beta = 0.2$	$\beta = 0.5$
data(1)	3063	17.3	<b>77.7</b>	76.8	72.9	69.9
data(2)	3828	29.0	<b>67.6</b>	63.7	65.5	59.5
data(3)	2671	12.8	59.1	59.3	<b>61.1</b>	56.2
data(4)	5798	14.3	<b>66.7</b>	62.6	66.5	62.8
data(5)	3366	20.1	<b>63.6</b>	61.1	62.8	58.7
data(6)	2563	17.2	51.1	49.9	<b>52.8</b>	46.4
data(7)	4244	12.2	63.0	64.3	<b>65.4</b>	63.3
data(8)	2227	20.6	50.7	55.9	<b>58.7</b>	53.1
平均		17.67	64.11	62.37	64.05	59.79

表 4 PreFEst および HTC との正解率の比較

	$X$	PreFEst				HTC				提案手法 ( $\beta = 0.2$ )			
		正解率 (%)	$I$	$D$	$S$	正解率 (%)	$I$	$D$	$S$	正解率 (%)	$I$	$D$	$S$
data(1)	3063	74.2	383	327	81	<b>81.2</b>	210	312	55	72.9	509	239	96
data(2)	3828	71.8	455	397	228	<b>77.9</b>	241	397	208	65.5	428	592	302
data(3)	2671	55.9	553	500	126	<b>64.2</b>	313	524	120	61.1	324	586	128
data(4)	5798	<b>76.2</b>	476	650	254	75.2	361	769	310	66.5	786	757	392
data(5)	3366	62.3	565	515	190	62.2	465	627	178	<b>62.8</b>	505	569	178
data(6)	2563	48.8	531	597	185	<b>63.8</b>	304	476	147	52.8	386	541	282
data(7)	4244	53.6	801	830	337	63.2	427	734	403	<b>65.4</b>	273	399	70
data(8)	2227	57.6	367	482	96	<b>70.9</b>	278	291	79	58.7	616	703	414
平均		64.26				70.37				64.05			

の性能を得られたが、これは事後確率最大化の一例であり、なお性能向上の余地があると考えている。

### 参考文献

- 1) S. Godsill and M. Davy, "Bayesian Harmonic Models for Musical Pitch Estimation and Analysis," Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP2002), pp. 1769–1772, 2002.
- 2) 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.
- 3) 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.
- 4) 亀岡弘和, 西本卓也, 嵯峨山茂樹, "調波時間構造化クラスターリング (HTC) による音楽の音響特徴量同時推定," 情報処理学会研究報告, 2005-MUS-61-12, pp. 71-78, 2005.
- 5) 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.
- 6) 亀岡弘和, 齊藤翔一郎, 西本卓也, 嵯峨山茂樹, "Specmurt における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," 情報処理学会研究報告, 2004-MUS-56-7, pp. 41–48, 2004.
- 7) 小野順貴, 齊藤翔一郎, 亀岡弘和, 嵯峨山茂樹, "Specmurt 分析における共通調波構造の Riemann の  $\zeta$  関数を用いた逆フィルタ解析," 日本音響学会 2006 年春季研究発表会講演論文集, 1-5-25, pp. 555–556, 2006.
- 8) 後藤真孝, 橋口博樹, 西村拓一, 岡隆一, "RWC 研究用音楽データベース: 音楽ジャンルデータベースと楽器音データベース," 情報処理学会研究報告, 2002-MUS-45-4, pp. 19–26, 2002.
- 9) H. Kameoka, T. Nishimoto and S. Sagayama, "A Multipitch Analyzer Based on Harmonic Temporal Structured Clustering," IEEE Transactions on Speech and Audio Processing, *in Press*.
- 10) 後藤真孝, 橋口博樹, 西村拓一, 岡隆一, "RWC 研究用音楽データベース: クラシック音楽データベースとジャズ音楽データベース," 情報処理学会研究報告, 2002-MUS-44-5, pp. 25–32, 2002.
- 11) 齊藤翔一郎, 亀岡弘和, 小野順貴, 嵯峨山茂樹, "凸射影法に基づく Specmurt 分析の共通調波構造推定," 情報処理学会研究報告, 2006-MUS-65-3, pp.13-18, May, 2006.