事後確率最大化 Specmurt 分析による音楽音響信号の多重ピッチ推定*

◎齊藤翔一郎, 亀岡弘和, 小野順貴, 嵯峨山茂樹 (東大情報理工)

1 はじめに

我々はこれまで, 多重音の音楽音響信号の音高情報を可視化する手法である specmurt 分析 [1] を提案してきた. 本稿では従来提案してきた反復推定 [2] を事後確率最大化を用いて達成する枠組を提案し, その改善点と性能比較実験の結果について報告する.

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. この倍音と基本周波数の位置関係は対数周波数領域では線形シフトの関係になっていることに注目すると, 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 $h(x)$ として, 多重音のスペクトル $v(x)$ を

$$v(x) = u(x) * h(x) \tag{1}$$

とモデル化することができる (x は対数周波数). これに基づき, 基本周波数分布 $u(x)$ を多重音スペクトル $v(x)$ に対する対数周波数での逆畳み込み

$$u(x) = \mathcal{F}\left[\frac{\mathcal{F}^{-1}[v(x)]}{\mathcal{F}^{-1}[h(x)]}\right] \tag{2}$$

で求める手法を specmurt 分析 [1] と呼ぶ.

2.2 共通調波構造の自動推定

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. ここでは, $u(x)$ は基本周波数のパワースペクトルのため非負でありかつ少数の支配的なピークを持つということ, $h(x)$ は倍音位置にのみインパルスを持つような分布であることが先験情報として仮定できる.

我々は一昨年, これらの条件をできるだけ満たすよう, 以下の 3 つのステップを繰り返すことで, $u_n(x)$, $h_n(x)$ を更新していく反復推定法を提案した [2]. ただし n はステップ数である.

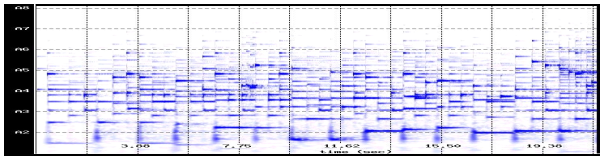
I. $v(x)$ と $h_n(x)$ に対し, 式 (2) により $u_n(x)$ を得る

II. u の先見情報に基づいて非線形写像

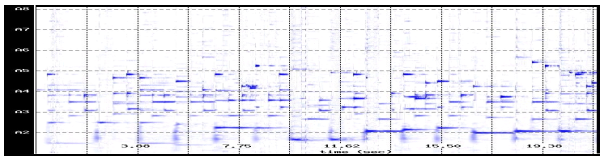
$$\bar{u}_n(x) = u_n(x)/\big(1+\exp\{-\alpha(u_n(x)-\beta)\}\big) \tag{3}$$

を用い, $u_n(x)$ の負値をとる部分を除き, 不要成分と思われる部分を抑圧する.

III. h の先験情報に基づき, $\int |v(x) - \bar{u}_n(x) * h(x;\boldsymbol{a})|^2 dx$ を最小化するインパルス列のパラメータ $\boldsymbol{a} = (a_1\ a_2\ \cdots\ a_N)^T$ を求め, $h_{n+1}(x) = h(x;\boldsymbol{a})$ を得る.



(a) スペクトログラム



(b)specmurt 分析結果

Fig. 1 specmurt 分析による倍音抑圧

この反復推定によって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のクラシック No.30 (F. ショパン: ノクターン Op.9, No.2) について specmurt 分析を行った結果を Fig. 1 に示す.

3 事後確率最大化 Specmurt 分析

3.1 最大事後確率推定問題としての定式化

前節のアルゴリズムは, 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ものであったが (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.

ここで, 観測スペクトル $v(x)$ を

$$v(x) = u(x) * h(x) + n(x) \tag{4}$$

とモデル化する. $n(x)$ は確率的な振る舞いを持つ誤差項と考え, 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の逸脱量を含む. 解くべき問題は $v(x)$ が与えられたときの事後確率 $\prod_x P(h(x), u(x)|v(x))$ を最大にする $h(x)$ と $u(x)$ を求めることであり, これは, その対数をベイズの定理により展開した

$$\begin{aligned} J \triangleq \sum_x & \log P(v(x)|h(x),u(x)) \\ & + \log P(u(x)) + \log P(h(x)) \end{aligned} \tag{5}$$

を最大化することに等しい.

3.2 従来手法の解釈

式 (5) の各項を $J_1 \triangleq \sum_x \log P(v(x)|h(x),u(x))$, $J_2 \triangleq \sum_x \log P(u(x))$, $J_3 \triangleq \sum_x \log P(h(x))$ と定義する. ここで J_3 は共通調波構造の事前確率であり, 本稿では問題の単純化のためにこれを一様分布と見なすことにする.

今, モデル誤差の確率分布 $P(v|u,h)$ に独立なガウス分布を仮定すると, J_1 の項は以下のように書ける:

*Multipitch Estimation of Music Audio Signals through MAP Specmurt Analysis. by SAITO, Shoichiro, KAMEOKA, Hirokazu, ONO, Nobutaka, SAGAYAMA, Shigeki (Graduate School of Information Science and Technology, The University of Tokyo)

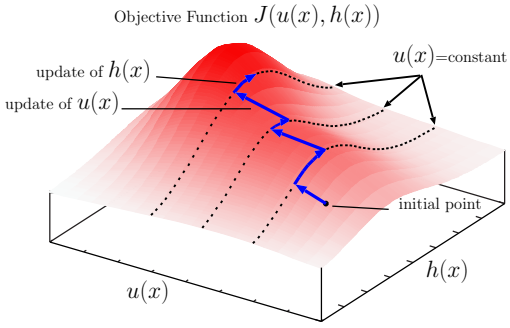


Fig. 2　提案法による目的関数最大化の概念図. $u(x)$ に関しては勾配の方向へ更新し, 次に $u(x)$ を固定した断面について最大になる $h(x)$ を求める.

$$J_1 = -\frac{(v - u * h)^2}{2\sigma^2} - \log(\sqrt{2\pi}\sigma) \tag{6}$$

ここで σ^2 は誤差分散であり, 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.

従来の反復推定法をこの枠組で解釈すると, まず 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 の大域最適解を求めていることに相当する. また, II の更新は, 学習係数 1 の最急降下法による更新と考えることができ, J_2 が

$$J_2 = \int \frac{-u \exp\{-\alpha(u-\beta)\}}{1 + \exp\{-\alpha(u-\beta)\}} \mathrm{d}u \tag{7}$$

となるような事前分布 $P(x)$ を設定して

$$u^{(t+1)} = u_{\mathrm{I}}^{(t)} + \frac{\partial J_2}{\partial u_{\mathrm{I}}^{(t)}} \tag{8}$$

ということを行っていると解釈できる. ただし $u_{\mathrm{I}}^{(t)}$ は t 回目の反復の I で更新された u とする. 最後に, III) は連立方程式を解くことで h に関する J_1 の最大化を行っていることに相当する.

つまり, 従来の反復推定法は順に, u に関する J_1 の最大化, u に関する J_2 における最急降下法, h に関する J_1 の最大化, を行っていたと解釈できる.

3.3　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るが, 各ステップで別々の評価関数を増加させていたため, 事後確率が最大化されるとは必ずしも限らない. よってここでは提案法として, 式 (5) を統一的な目的関数と考え, これを最大化することを考える.

目的関数を最大にする $u(x)$ は陽には求まらないが, その微分係数は導出できるため,

$$\frac{\partial J_1}{\partial u(i)} = 2 \sum_x h(x-i) \Big(v(x) - \sum_\tau u(x-\tau) h(\tau) \Big) \tag{9}$$

$$\frac{\partial J_2}{\partial u(i)} = \frac{-u(i) \exp\big[-\alpha\{u(i) - \beta\}\big]}{1 + \exp\big[-\alpha\{u(i) - \beta\}\big]} \tag{10}$$

を用いて最急降下法

$$u^{(t+1)} = u^{(t)} + A \left(\frac{\partial J_1}{\partial u} + \frac{\partial J_2}{\partial u} \right) \tag{11}$$

により, 目的関数を増加させる u の更新ステップを得ることができる. ただし $A(>0)$ は学習係数である. J_3 はここでは定数と考えているので u に依存しないことに注意する. 一方, 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 III を行うことで最適解を求めればよい. このようにすることで u と h について交互に事後確率を単調増加させることが出来, u と h の収束性が保証される. この最大化を概念的に表したものを Fig.2 に示す.

Table 1　従来法と提案法の MIDI 変換精度の比較

	no iteration	conventional	proposed
平均	17.67%	64.11%	64.05%

Table 2　先行研究と提案法の MIDI 変換精度の比較

	preFEst	HTC	proposed
平均	64.26%	70.37%	64.05%

4　評価実験

4.1　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の一部を 20 秒強切り出した 8 つのデータである. 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. 実験に用いたパラメータは今回は $\alpha = 15.0$, $\beta = 0.2$, $A = 0.9$, $\sigma^2 = 1.5$ に固定した. $u(x)$ の初期値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. 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 $100 \times (X - D - I - S)/X$ のように計算した. また, 同様の実験を用いて後藤が提案する preFEst[3][1] と亀岡が提案する HTC[4] の 2 つの手法とも性能を比較した.

結果を Table 1, 2 に示す. 提案法は従来の反復推定法よりピッチ推定性能を向上させたとは言えないが, ほぼ同程度の性能を示しており, さらに従来法で稀に起きていた推定の大幅な逸脱も見られなくなった. また, 先行研究との比較においても同程度と考えられる性能を得た.

5　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の性能を得られたが, 共通調波構造の事前分布導入など, なお性能向上の余地があると考えている.

謝辞　本研究の一部は科学研究費補助金・基盤研究 B(課題番号 17300054) および科学技術振興機構 CREST プロジェクトの補助を受けて行われた.

参考文献

[1] 高橋他, 情報処理学会研究報告, 2003-MUS-53, pp.61-66, Dec. 2003.
[2] 亀岡他, 音講論 (秋), pp. 803–804, 2004.
[3] M. Goto, ICASSP2001, pp. V-3365–3368, 2001.
[4] 亀岡他, 音講論 (秋), pp. 769-770, 2005

[1] 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含まない.