HTC 多重音ピッチ推定と HMM リズム・テンポ推定を統合した音響信号からの自動採譜の検討*

◎宮本賢一, 亀岡弘和, 武田晴登, 西本卓也, 嵯峨山茂樹 (東大情報理工)

1　はじめに

音楽音響信号からの自動採譜は、楽譜作成・音楽検索を始め、今後の音楽産業におけるさまざまな用途が挙げられ、長く高い関心を集めてきた研究課題であるが、多重音からの音高推定や各音の音価認識などの複数の難しい問題設定が内在し、その実現までの道のりは遠いと考えられてきた。そこで、本研究では、これまで我々が開発した HTC(Harmonic Temporal Structured Clustering) 多重音ピッチ推定 [1]、HMM(Hiden Marcov Model) リズム・テンポ推定 [2] を統合した自動採譜手法を検討する。広義の自動採譜には、各音の音程と音価のみならず、調号、テンポ指定、表情記号、アーティキュレーションなどさまざまな楽譜要素の推定が含まれるが、本研究ではまず各音符の音程と音価の推定に限定して議論することとする。

2　多段処理による自動採譜の検討

自動採譜は、演奏された音楽音響信号から元の楽譜を推定する逆問題と捉えることができるが、音楽演奏過程の統合的な逆問題解決は将来の課題として措くことにし、また対象を音高のある楽器音に限定すれば、図 1 に示すように、以下のような多段の変換過程による手法が考えられる。

(1) 音響信号からスペクトル系列を得る
(2) 各音符のピッチ推定と分離により、各音のパワーパターンを得る
(3) 各音符の始終端推定により、ピアノロールデータを得る
(4) 各音符の音価とテンポ推定により、各音符の時刻と音価を得る
(5) 楽譜データを生成する

3　自動採譜のための要素処理

3.1　HTC による多重音分離

入力音響信号 $f(t)$ をステップ (1) でウェーブレット変換して求められたスペクトルにおいては、各構成音の基本周波数およびその倍音周波数に音響エネルギーが分布し、それが各音の音長に対応して時間的に継続して観測される。これを、ステップ (2) では、我々の HTC 多重音ピッチ推定 [1] により、スペクトログラム $W(x,t)$(x は対数周波数、t は時刻) を、調波構造と時間伸縮の拘束を持った 2 次元 GMM でモデル近似する。事前分布として倍音成分の関係や時間エンベロープの減衰を仮定することで、倍音と基本周波数の多義性が改善できる。さらに HTC で推定された番号 k の音響オブジェクトの基本周波数推定値 $\widehat{\mu}_k$(単位：cents) とモデル化分布密度 $q_k(x,t;\widehat{\mu}_k)$ を用いて、$W(x,t)$ を

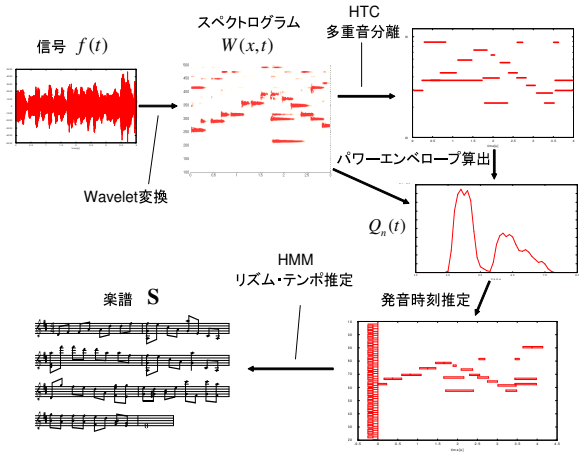

Fig. 1　自動採譜処理の概要

$$\tilde{q}_k(x,t;\widehat{\mu}_k) = \frac{q_k(x,t;\widehat{\mu}_k)}{\sum_k q_k(x,t;\widehat{\mu}_k)} W(x,t) \qquad (1)$$

のように音響オブジェクトごとに分離できる。

3.2　時間エンベロープ分布からの発音時刻の推定

続くステップ (3) は、各ピッチについて得られた音響エネルギーの時間関数形状から、音符の始終端を推定するもので、新たな検討を行った部分であるので、以下に論じる。前節の HTC による音響オブジェクト分離においては、一音を複数のモデルで近似してしまったり、その逆もしばしば起こり得る。そこで、式 (1) で分離した $\tilde{q}_k(x,t;\widehat{\mu}_k)$ は必ずしも実際の一音の分布に対応しない。そこで HTC によって得た $\tilde{q}_k(x,t;\widehat{\mu}_k)$ を音高ごとに統合し、

$$Q_n(t) = \sum_{k\in C_n} \int \tilde{q}_k(x,t;\widehat{\mu}_k)\mathrm{d}x \qquad (2)$$

$$C_n = \left\{k \middle| A\left(n-\tfrac{1}{2}\right) \leqq \widehat{\mu}_k < A\left(n+\tfrac{1}{2}\right),\ k,n\in\mathbb{N}\right\}$$

(ただし、A は 100cent(半音間隔)) のように音高 n のパワーエンベロープ $Q_n(t)$ を得て、これから再度、各音へ分離することを検討する。

この $Q_n(t)$ から発音時刻を推定することを考えたいが、ウェーブレット変換ではサブバンドごとに異なる度合の時間方向のエネルギー拡散が生じるため、このエネルギー拡散を考慮した発音時刻推定を行うべきである。ピアノのように発音の急峻な音響信号を想定し、発音時刻 τ、周波数 ω_0、減衰率 α で指数減衰する解析信号

$$g(t) = cu(t-\tau)e^{j\omega_0(t-\tau)}e^{-\alpha(t-\tau)}$$

に対する Gabor ウェーブレット変換の中心周波数 ω_0 のサブバンドパワーエンベロープは解析的に求められる。ただし、$u(t)$ は $t \geqq 0$ で 1、$t < 0$ で 0 をとる

* "Automatic Music Transcription Combining HTC Multipitch Analysis and HMM-based Rhythm and Tempo Estimation," by Ken-ichi Miyamoto, Hirokazu Kameoka, Haruto Takeda, Takuya Nishimoto and Shigeki Sagayama, Graduate School of Information Schience and Technology, The University of Tokyo.

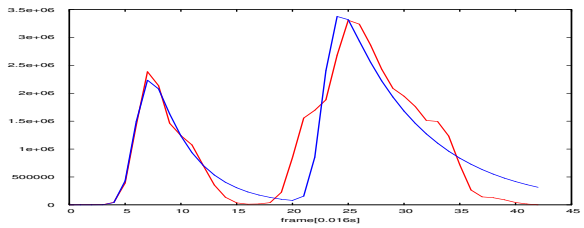

Fig. 2　パワーエンベロープのモデルフィッティングの様子

ステップ関数、c は振幅を表す。導出は省略するが、上記パワーエンベロープは

$$\Psi(t;\tau,\alpha,C) = Ce^{-2\alpha(t-\tau)}\left(\int_{-\infty}^{t-\tau-\frac{d\alpha}{2\omega_0^2}} e^{-\frac{\omega_0^2}{d}s^2}\mathrm{d}s\right)^2$$

と求まる。$Q_n(t)$ はこの関数の L 音分の重ね合わせの結果であると考えれば、

$$\forall t:\ Q_n(t) \approx \sum_{l=1}^{L} \Psi(t;\tau_l,\alpha_l,C_l) \qquad (3)$$

のようにモデル化できる。連続発音を考慮するため、本研究では音量変化の探索によりモデル数を決定する。

ここで、最適パラメータを以下の目的関数

$$J = \int_{-\infty}^{\infty} \left|Q_n(t) - \sum_{l=1}^{L} \Psi(\alpha_l,\tau_l,C_l,t)\right|^2 \mathrm{dt} \qquad (4)$$

の最小化により推定する。具体的には

1. α_l、τ_l を固定して、$\mathbf{C} = (C_1,\cdots,C_L)^T$ を更新
2. $\mathbf{C}$ を固定して、最急降下法より α_l、τ_l を更新 (直線探索によりステップサイズ取得)

上記各ステップで目的関数は減少するので、局所最適解への収束が保証される。図 2 において、HTC 分析と式 (2) によって得た $Q_n(t)$ を提案手法によってモデルフィッティングした例を示す。

また推定結果では、倍音成分と基本周波数との多義性やエンベロープモデル数の決定手法の性質より、音の脱落や不要音が生じ得る。そのため本研究では、推定したエンベロープの大きさを閾値にして不要音の除去を行なう。

以上の手法により、各音高ごとに発音時刻推定値 τ_l を取得できる。

3.3　発音時刻推定値を用いたリズム・テンポ同時推定

前節の提案手法により、多重音に関して音程と発音時刻の情報を取得できる。ステップ (4) は、音程と発音時刻情報から、和音を考慮したリズムパターンの推定問題である。我々の HMM を用いたリズム・テンポ同時推定 ([2] 参照) では、得られた発音時刻ごとの間隔から、多項式近似したテンポ曲線とリズム音価列を反復的に推定できる。また和音を考慮した HMM を利用しているため、和音のリズム推定を可能にしている。この結果、テンポが滑らかに変動する曲に対応したリズムパターン推定が実現できる。

4　自動採譜システムの動作検証

本システムを実装し、実際の音楽音響信号から自動採譜を行う動作実験を行った。Bürgmuller のピアノ練習曲を用い、提案システムを用いて発音時刻のリズム譜を自動生成した。図 3 は実際の演奏 MIDI から [2] の手法を用いて推定した楽譜であり、正しい発音リズムが推定されている。同じ演奏 MIDI を WAV 形式に変換し、提案システムを用いて自動採譜を行った結果を図 4 に示す。なお、リズム・テンポ推定でのリズム曲線の次数は 3 次で固定し、楽譜の調号はあらかじめ人為的に与えている。

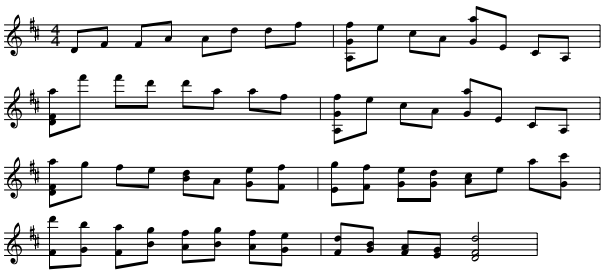

Fig. 3　演奏 MIDI からリズム推定した楽譜

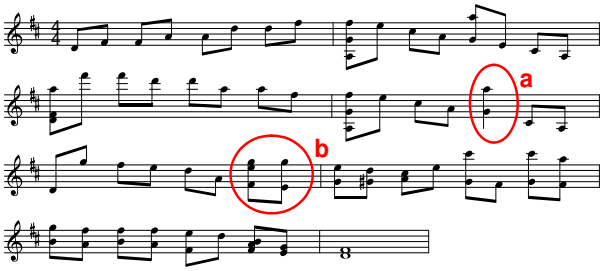

Fig. 4　提案手法による自動採譜結果

2 図の比較結果、2 段目一小節目までは HTC による分離、発音推定が成功し、リズム推定も正解した。図 4 の a 部では音の脱落が起こったが、リズムは 8 分音符 2 つ分の音価を 4 分音符ひとつとして正しく推定された。しかし b においてパワーエンベロープにより正解の音を誤って除去した後では、リズムを誤推定が見られた。これは、HMM リズム・テンポ推定手法は入力情報の脱落やノイズを想定していないためである。リズムや和音情報を用いた不要音除去や、ノイズを想定したリズム推定を行なうことで、このような誤りは改善できると考えられる。

5　おわりに

本研究では、HTC 多重音ピッチ推定と HMM リズム・テンポ推定の手法を発音時刻推定手法を用いて統合した、音響信号からの自動採譜を議論した。さらに提案する自動採譜システムを実装し動作を確認した. 今後は、リズムや和音の情報を利用した HTC での音響オブジェクトの推定について検討したい。

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なわれた。

参考文献

[1] H. Kameoka *et al.* , "A Multipitch Analyzer Based on Harmonic Temporal Structured Clustering," *IEEE Trans. on Audio, Speech and Language Processing*, in Press.

[2] 武田ら, 音講論 (春), pp. 721-722, 2006.