# 楽譜からの音楽音響信号生成モデルに基づく 楽譜と音響信号の詳細なアラインメント*

◎松本 恭輔, 西本 卓也, 小野 順貴, 嵯峨山茂樹 (東大院・情報理工)

## 1 はじめに

本稿では、楽譜と音響信号の詳細なアラインメントについて議論する。これは、演奏のテンポ、楽譜上の各音符に対応する各音の発音時刻・音長・音高等を詳細に求めるもので、楽譜を用いた音楽音響信号の解析手法として、デビエーションデータベース [4] の作製支援や、楽譜情報を用いた音楽音響信号加工の準備等の様々な応用が見込まれる。

従来のアラインメント手法 [1, 2] が扱う問題は、録音へのランダムアクセス・自動伴奏システム等への応用を想定した、音響信号上の時刻と楽譜上の拍との対応付け問題であり、実演奏にテンポ変動のみを仮定した楽譜と音響信号のマッチング問題と捉えられる。しかし実演奏は、テンポ変動のみならず、一音毎の音長・発音時刻・音高に微小変動を含むため、従来手法単独での詳細なアラインメントには限界がある。

これに対して我々は、楽譜から音楽音響信号が生成される過程に着目し、詳細なアラインメントを行う手法の検討を行ったのでこれを報告する。

## 2 楽譜からの音楽音響信号生成モデル

### 2.1 楽譜からの音楽音響信号の生成過程

従来手法で考慮された、テンポ変動に加え、どのような変動を考慮すればよいだろうか. 実際の演奏場面を考えてみよう。まず、楽譜 (Score; $\mathcal{S}$) を見て、指揮者 (Tempo; $\mathcal{T}$) はテンポを設計し、楽譜に明示的に記されていないテンポを演奏全体に与える (テンポ変動)。次に、そのテンポに合わせて演奏が行われるが、各演奏音 (Model; $\mathcal{M}$) には、演奏者によって意識的/無意識的に揺らぎが生じる (発音時刻・音長・音高のデビエーション、以下これを単にデビエーションと呼ぶ)。そして楽器から音が発せられ、音響信号 (Audio; $\mathcal{A}$) が生成される (音色変動)。指揮者がいない場合は、演奏者自身が頭に思い描く演奏のテンポを、指揮者のテンポに対応付け、同様の説明ができる。

上述の各種変動は、(楽曲、指揮者、演奏者が同じであったとしても) 毎回の演奏で厳密に一定であることはないので、確率的に扱うことが必要である。また、各種変動は原因が異なり、独立に発生すると近似できる。以上から、「楽譜に上記三種の変動が順次加わることで、音楽音響信号が生成される」とする、「楽譜からの音楽音響信号生成モデル」を考えることができる (Fig. 1 参照)。

次項以降、このモデルに基づき、詳細なアラインメントを確率的逆問題として定式化するために各種変動に対して数理モデルを与える。本来、各種変動は、音楽的な解釈や意図に従って生じるものである。将来は高度な楽譜の自動解析と理解からの変動の予測が可能になるかも知れないが、当面はそのような知識は仮定せず、単純な確率変動のモデルによるアラインメントを検討することは妥当であろう。

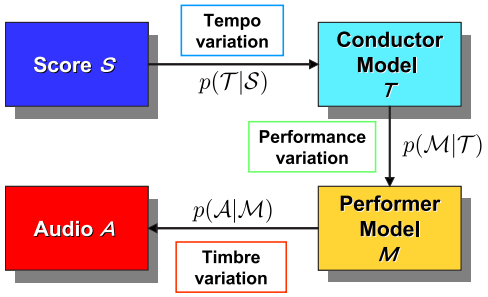


Fig. 1 楽譜からの音楽音響信号生成モデル: 楽譜$\mathcal{S}$に三つの変動が加わり音響信号$\mathcal{A}$が生成される。

特に本稿では、よりシンプルな問題として、ピアノ曲のように、音高の揺らぎが十分無視でき、音長には十分大きい変動を許すアラインメントを想定したモデル化を行うことで、テンポ変動と密接に関係するであろう、発音時刻のデビエーションをモデルとして導入した効果を確認する。

### 2.2 楽譜から指揮者へ – テンポ変動のモデル化

局所的なテンポ揺らぎを演奏者の各音のデビエーションとして扱えば、指揮者テンポは楽曲のほとんどの場所で区分的に一定であると近似できる。 そこで、楽譜上の拍位置 $b$ と音響信号上時刻 $T(b)$ の対応関係を、区分線形モデル (区間の数 $J$, 間隔 $L$ は固定)

$$T^{(j)}(b) = \frac{\beta_{j+1} - \beta j}{L} b + ((j+1)\beta_j - j\beta_{j+1}) \quad (1)$$

とする。$T^{(j)}$ は $j$ 番めの区間の直線、$\beta_j (_{j=1,2,\ldots,J-1})$ は $j$ 番めの区間の開始時刻である。

音楽研究で広く利用されるテンポの特徴「テンポ変動は滑らか」を利用して、「隣り合う区間のテンポの差は平均が 0 で、分散の小さい正規分布に従う」と仮定すると、テンポ変動の確率を以下のようになる。

$$p(\mathcal{T}|\mathcal{S}) = \prod_{j=0}^{J-1} \frac{1}{\sqrt{2\pi}\sigma_t} e^{-\frac{(\beta_j - \beta_{j+1}/2 - \beta_{j-1}/2)^2}{2\sigma_t^2 L^2}} \quad (2)$$

### 2.3 指揮者から演奏へ – デビエーションのモデル化

「発音時刻 $\tau_k$ は、テンポに従う理想時刻 $T(b_k)$ を中心とする分散の小さな正規分布に従う」と仮定すると各音デビエーションの確率は

$$p(\mathcal{M}|\mathcal{T}) = \prod_{j=0}^{J-1} \prod_{k_j=1}^{K_j} \frac{1}{\sqrt{2\pi\sigma_{on}^2}} e^{-\frac{(\tau_{k_j} - T^{(j)}(b_{k_j}))^2}{2\sigma_{on}^2}} \quad (3)$$

で与えられる。$k_j, K_j$ は、区間 $j$ に含まれる、単音のインデックスと、個数である。

### 2.4 演奏音から音響信号へ – 音色変動のモデル化

Kameoka らは、多重音解析を時間-対数周波数平面上での単音モデルの配置問題として捉えた、時間調波

---

*"Detailed Alignment of Score to Audio Signal Based on Generative Model of Musical Acoustic Signal from Score" by MATSUMOTO, Kyosuke,(The University of Tokyo).

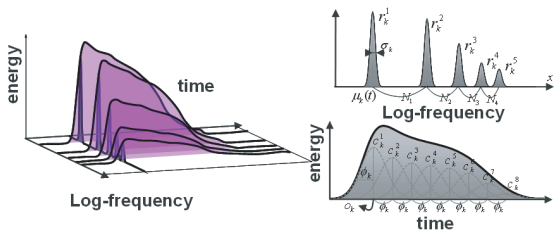


Fig. 2　パラメトリック HTC における単音のモデル:ガウシアンを基底関数として時間・周波数方向に連ねる

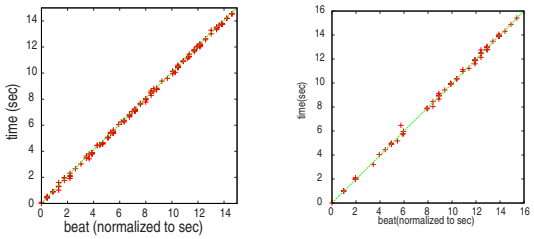


Fig. 3　推定テンポ (緑点線の区分直線) と、各音発音時刻 (赤点): (左)No.39 (右)No.29

構造クラスタリング (Harmonic Temporal Clustering; HTC) を提案している [3]。この捉え方は、時間周波数各方向のデビエーションを統一的・直感的に扱うのに適しており、HTC は演奏音のモデルからの音響信号生成のモデルに基づく手法として解釈可能でもある。そこで本稿では、HTC でが用いる演奏音モデル (Fig. 2 参照) と音色変動の確率を用いる。

## 3　詳細なアラインメントの定式化と解法

### 3.1　詳細なアラインメント問題の定式化

前節で与えられたモデルに基づき、ベイズの定理、各種変動の独立性によって、詳細なアラインメント問題は、「楽譜$\mathcal{S}$, 音響信号$\mathcal{A}$が既知の下での、テンポと演奏のパラメータ $\Theta \triangleq (\Theta_{\mathcal{M}}, \Theta_{\mathcal{T}})$ の事後確率最大化推定」として以下のように定式化される。

$$\begin{aligned}\hat{\Theta} &= \operatorname*{argmax}_{\Theta} p(\Theta|\mathcal{S},\mathcal{A}) \\ &= \operatorname*{argmax}_{\Theta} \Big( \log p(\mathcal{A}|\mathcal{M}) + \log p(\mathcal{M}|\mathcal{T}) + \log p(\mathcal{T}|\mathcal{S}) \Big) \end{aligned} \tag{4}$$

### 3.2　演奏とテンポの反復推定アルゴリズム

上述の問題に対する最適パラメータを一挙に解析的に求めることはできないが、一方を固定し、もう一方に関して目的関数を単調増加させることは、$\Theta_{\mathcal{T}}$に関して解析的に、$\Theta_{\mathcal{M}}$に関して安定性のある反復法により行える ([5] 参照)。これを利用した詳細なアラインメント手法のアルゴリズムを以下に示す。目的関数は上に有界、各ステップで関数値は単調非減少するので、本手法は局所最適解への収束性が保証される。

1. $\Theta_{\mathcal{M}}, \Theta_{\mathcal{T}}$に適当な初期値を与える
2. $\Theta_{\mathcal{T}}$を固定、$\Theta_{\mathcal{M}}$を更新する
3. $\Theta_{\mathcal{M}}$を固定　$\Theta_{\mathcal{T}}$を更新する
4. 目的関数が収束したら終了, それまで 2,3 を繰り返す

## 4　実験・今後の課題

提案法は、特に各音のデビエーションに対処するための反復推定手法である。将来的には大域的最適化手法の併用を予定しているが、提案法もテンポ変動を考慮に入れた枠組であり、単独でどの程度のテンポ変動に対処し得るかは興味深い。そこで今回は、提案法単独での実演奏への適用例を示す。

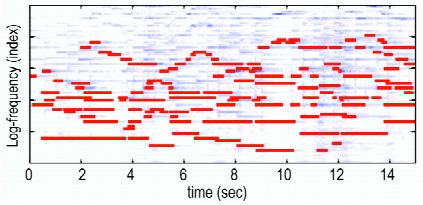


Fig. 4　No.39 推定結果 (赤実線) と入力スケーログラム

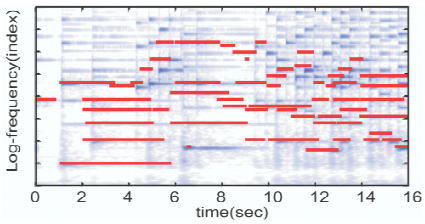


Fig. 5　No.29 推定結果 (実線) と入力スケーログラム

入力音響信号には、RWC 研究用音楽データベース RWC-MDB-C-2001 より、No.39(C. フランクのヴァイオリンソナタ)・No.29(R. シューマンのトロイメライ: ピアノ独奏曲) を抜粋して利用した。楽譜情報には、対応する RWC-MDB の MIDI ファイルを楽譜通りの発音拍と音価に修正し利用した。演奏音モデルの初期値は、音響信号長に合わせて線形伸縮をした楽譜に従い配置し、テンポの初期値は、平均のテンポ ($\alpha$ とする) とした。また、分析条件は、テンポ一定の区間を約 3 秒、発音時刻・テンポの標準偏差をそれぞれ $\sigma_{on} = 0.25$ 秒, $\sigma_t = \alpha/20$ とした。

演奏のテンポがほぼ一定である No.39 に対しては、推定テンポは一定 (Fig. 3 左参照)、各音毎にデビエーションが適切に推定されている (Fig. 4 参照)。一方、No.29 はテンポ変動が豊富な演奏で、後半部分はテンポ変動に追従しながら、各音デビエーションを推定しているが、前半部分では、ロングトーンの間に変化したテンポに追従することができずに、推定を失敗している (Fig. 5 参照)。このような部分に対処するには、音長の変動を考慮したデビエーションモデルの構築や大域的最適化手法の併用が必要である。

今後は、今回の検討を参考に、より適切なモデル構築、大域的最適化手法との併用、手法の定量的評価実験を行う予定である。

**謝辞**　本研究の一部は、科学技術振興機構 CREST 研究課題「時系列メディアのデザイン転写技術の開発」として行われた。また、亀岡弘和氏 (NTT CS 研) には、有益なコメントを頂いた。

## 参考文献

[1] N. Orio *et al.*, "Alignment of Monophonic and Polyphonic Music to a Score," *ICMC 2001,* pp. 155–158, 2001.

[2] C. Raphael, "A Hybrid Graphical Model for Aligning Polyphonic Audio with Musical Scores," *the 5th ISMIR,* pp.387–394, 2004.

[3] H. Kameoka *et al.*, "A Multipitch Analyzer Based on Harmonic Temporal Structured Clustering," *IEEE Trans. ASLP.,* Vol. 15, No. 3, pp.982–994, Mar, 2007.

[4] 橋田光代 他, "音楽演奏表情データベース構築に向けて," 人工知能学会全国大会 2007.

[5] 松本恭輔 他, "パート除去を目的とした楽譜と音響信号のアラインメント手法の検討," 情処学会研究報告, 2007-MUS-71.