

# 和音の発音順序交替を許容した動的計画法による多声 MIDI 演奏の楽譜追跡\*

武田晴登, 西本卓也, 嵯峨山茂樹 (東大情報理工)

## 1 はじめに

人間の演奏者による演奏を対象としたリアルタイムで動作する楽譜追跡は、自動伴奏や音楽鑑賞支援に必要であり要求は多い。電子ピアノ等の電子楽器による実演奏のように演奏の MIDI (Musical Instrument Digital Interface) 信号が得られる場合、演奏に誤りが無ければ楽譜追跡は容易であり実際に商品化されている。しかし、人間の演奏には、演奏の誤りによる音の脱落・置換・挿入が含まれ、更に、演奏者の演奏の表情付けによるリズムやテンポの変動、無意識に起こす発音時刻の変動も含まれるため、MIDI 信号を対象とした場合でも実演奏の楽譜追跡は簡単な問題ではない。本稿では、Fig. 1 のように演奏による MIDI 信号が逐次入力されているときに、演奏された音のノートオンメッセージ (以下、発音情報) のみを用い、入力された発音情報に対応する楽譜上の拍位置をリアルタイムで求める楽譜追跡について議論する。

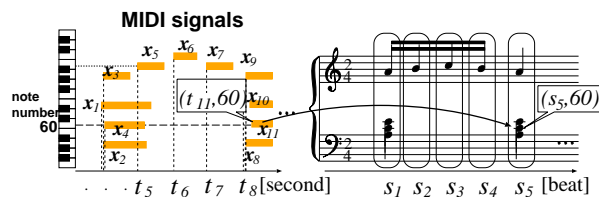


Fig. 1 MIDI 信号に対応する拍位置を求める楽譜追跡

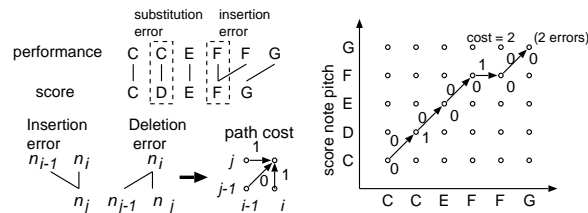


Fig. 2 DP を用いた単旋律の楽譜追跡

## 2 単旋律 MIDI 演奏の楽譜追跡

### 2.1 音高の時系列によるマッチング

まず、MIDI 演奏が単旋律である場合を考える。このとき、楽譜のある拍位置にはひとつの音しか存在しないため、楽譜の拍位置  $s$  の時系列に対応する音高  $n$  の時系列とは一意に決まる。従って、発音時刻が  $t$  秒でノートナンバーが  $n$  である演奏の発音情報  $\vec{x} = (t, n)$  と楽譜の拍位置  $s$  との対応を求めるには、音高をもとに演奏と楽譜の対応付けを行えばよい。

### 2.2 動的計画法の適用

MIDI 演奏と楽譜から得られる音高  $n$  の時系列のマッチングは、通常の文字列の比較と同様に脱落・挿入・置換の誤り数の合計を最小とするマッチングの問題として DP (Dynamic Programming, 動的計画法) を用いて求められる。即ち、音の脱落・挿入による誤りは Fig. 2 左下に示すように DP の経路コストで数え、置換誤りは演奏の  $i$  番目の発音情報  $\vec{x}_i = (t_i, n_i)$  と楽譜の  $j$  番目の拍位置  $s_j$  の音高の  $n_j$  一致を局所コスト

$$c_1(i, j) = \begin{cases} 1 & (n_i = n_j) \\ 0 & (n_i \neq n_j) \end{cases} \quad (1)$$

で数え、誤りの総数はを最小とする対応はコスト最小とする経路にあたる。この経路探索問題は最適性の原理を用いて効率的な計算で解くことができる。Daneberg は DP の原理を用いて [1] 独自に設定したコストを用いリアルタイムでの単旋律 MIDI 演奏の楽譜追跡を提案した。

## 3 多重音を含む MIDI 演奏の楽譜追跡

### 3.1 同時発音における発音の順序の交替

次に、MIDI 演奏が多重音を含む場合を考える。楽譜のある拍位置  $s$  には複数の音高が対応し、同時に発音することを意図された音の集合 (以下、和音) の各音は発音情報は MIDI メッセージとして任意の時間順序で表れる。従って、単旋律の場合と異なり、一般には正しく演奏された場合の音高の時系列が一意には決まらない。また、楽譜の和音と演奏された和音とを比べて演奏誤りを数えるには、和音内では各音の音高の出現を独立に仮定することは適切でない。従って、単旋律で用いた音高の DP をそのまま適用することはできない。

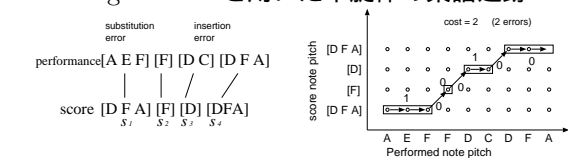


Fig. 3 2段 DP を用いた多旋律の楽譜追跡

### 3.2 2段 DP 法の適用

多重音を含む実演奏の発音情報から DP を用いて楽譜上の拍位置を求める手法の一つとして、和音を発音情報のクラスタと考えると DP を適用することが考えられる。各音が同時に発音することを意図したものを装飾音のように異なるタイミングで聴かせることを意図したものを発音時刻間隔から決定することはできないので、MIDI 演奏の発音情報の和音としてのグルーピングは事前には与えられない。従って、発音情報の和音へのクラスタリングと、各和音と楽譜の拍との最適な対応を求める必要がある。

和音単位で DP を行うので、単旋律の場合の DP の自然な拡張として、局所コストに楽譜と演奏の和音を比較して数えた音の脱落・挿入・置換の誤りを用いる。  $i$  番目の発音情報  $\vec{x}_i$  からその  $L$  個前までの発音情報  $\vec{x}_{i-N+1}$  の音高を和音  $N(i, L) = \{n_{i-L+1}, \dots, n_i\}$  とし、また、楽譜のある拍位置  $s_j$  の和音を構成する音高の集合を  $N(j)$  としたとき、局所コスト

$$c_2(i, j, L) = \min\{|N(i, L)|, |N(j)|\} - |N(i, L) \cap N(j)| \quad (2)$$

と与えられる。Fig. 3 に示すように和音となる音の個数の可能性を考え、誤りの個数を最小に数える対応を求めることができる。これは 2段 DP の原理 [2] を利用したものである。

アルゴリズムを実装し、演奏 MIDI と楽譜との対応付けを正しく与えられることを人間の MIDI 演奏を用いて確認した。この手法は、多声部音楽の演奏と楽譜との間で楽譜に対する演奏誤り数を求めるのに使用可能である。

## 4 リズムを考慮した MIDI 信号の楽譜追跡

### 4.1 和音とリズムの一致を用いた楽譜追跡

これまで述べた MIDI 演奏の楽譜追跡法は、音高情報のみを用いて誤り数最小を数え上げるマッ

\*Score Following of Polyphonic MIDI Performance based on Dynamic Programming. by Haruto Takeda, Takuya Nishimoto, Shigeki Sagayama (University of Tokyo)

グであったが、同じ構成音の和音が長い間連続し音高のみでは演奏位置の特定が難しい場合には、演奏のリズムやテンポに特徴を用いてより頑健に演奏位置を推定できると考えられる。そこで、ここではリズムの特徴を含むように拡張した楽譜追跡を論じる。

#### 4.2 事後確率最大化による演奏位置の推定

発音時刻音の変動の可能性を考えると、一般に演奏された音は、ある拍で音高を誤ったものか、異なる拍の音をずれた発音時刻で演奏したものであるかは、演奏途中で決定することは原理的にできない。我々は原理的に一意に解決できない問題を、確率的モデルによるモデル化を通して楽譜追跡を最も確からしい対応を求める推定問題として扱う。

演奏の発音情報  $X = \{\bar{x}_1, \dots, \bar{x}_i\}$  に対応する拍位置を  $S = \{s(1), \dots, s(i)\}$  とするとき、最も確からしい  $\hat{S}$  は、事後確率最大化問題として、

$$\hat{S} = \operatorname{argmax}_S P(S|X) = \operatorname{argmax}_S P(X|S)P(S) \quad (3)$$

と定式化される。事後確率を最大にするために、今までの誤り数をコストとしてコスト最小となる対応を求める DP の計算原理は、最適性の原理を確率最大の計算に用いた DP に置き換えて使用できる。

#### 4.3 2段 DP 法の適用

リズムは連続する音の時間長の組合せで表現されるので、楽譜と演奏をリズムについてマッチングを行うには、両者の時間的幅を持った区間を比較する必要がある。そこで、楽譜の拍位置  $s_j$  から  $M$  個前の拍位置  $s_{j-M+1}$  の区間と  $N$  個の発音情報  $\bar{x}_{i-L+1}, \dots, \bar{x}_i$  による演奏される確率  $c_3(i, j, L, M)$  を用いて局所的なマッチングを行う。ここでは、

$$\begin{aligned} c_3(i, j, L, M) = & \log P(\text{和音の一致}) \\ & + \log P(\text{拍単位の脱落・挿入誤り}) \\ & + \log P(\text{リズムの一致}) \end{aligned} \quad (4)$$

とおく。Fig. 4 下に示すように、発音情報  $\bar{x}_i$  に楽譜の拍位置  $s_j$  が対応する確率は 2 段 DP により効率的に求められ、 $i-L$  番目までの様々な  $j$  から遷移から最適なものを選べば良い。

#### 4.4 和音とリズムの一致確率の定式化

##### 4.4.1 和音構成音の音階について的一致確率

楽譜の和音の構成音に対して、実演奏がどのような構成音の和音が演奏されるかに関する確率を考える。各音高での音の存在を 0,1 の 2 値で表現した音高の数の成分がある音高ベクトル  $\vec{v}$  は、演奏誤りがなければ一致する。ここでは、楽譜の  $\vec{v}_m$  に対して  $\vec{v}_l$  と演奏される確率  $p(\vec{v}_l|\vec{v}_m)$  を誤り傾向を考慮して定める。

##### 4.4.2 リズムの一致確率

演奏者のテンポは楽譜追跡過程では未知であるので、ここではテンポに依存しない特徴量を用いて楽譜と演奏の発音タイミングを比較する。Fig. 5 に示すように時間情報である発音時刻「秒」とそれに対応する楽譜の「拍」をリスケールした時間情報  $r_l, r_m$  は、演奏が音価と一致するリズムでかつテンポが一定である厳密な発音時刻である場合は厳密に一致する。実演奏にはテンポや音長の変動が含まれるので、楽譜の  $r_m$  に対応する発音情報が  $r_l$  として観測される確率  $p(r_l|r_m)$  を 0 を平均とした正規分布で与える。

##### 4.4.3 2段 DP による和音とリズムの一致確率

$c_3(i, j, L, M)$  の計算における和音のグルーピングは、Fig. 4 上に示すように、3 節における局所コストを  $c'_2(l_1, l_2, m) = \sum_{l=l_1}^{l_2} \log p(r_l|r_m) + \log p(\vec{v}_l|\vec{v}_m)$  に置き換えた 2 段 DP により、同様の方法で求められる。

#### performance

onset message  $\mathbf{x}_{i-L} \quad \mathbf{x}_{i-L+1} \quad \dots \quad \mathbf{x}_i$   
 $[[\dots] \dots [\dots]] \quad [[\text{A F D}] [\text{F}] [\text{E D}] [\text{F}] [\text{F A D}]]$

#### score

$[[\dots] \dots [\dots]] \quad [[\text{D F A}] [\text{F}] [\text{D}] [\text{A D F}]]$   
 beat positions  $S_{j-M} \quad S_{j-M+1} \quad \dots \quad S_j$

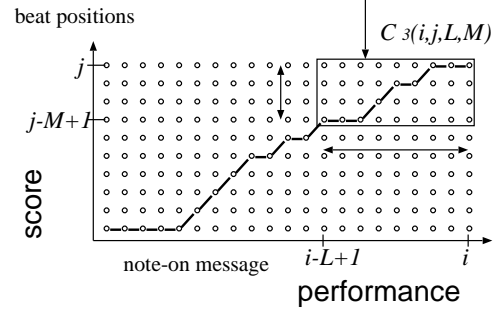
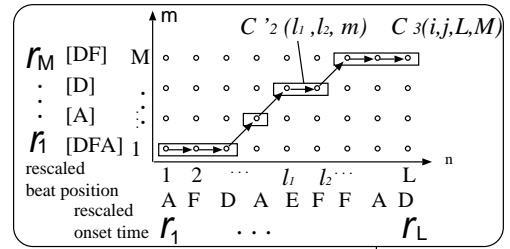


Fig. 4 2 段 DP を用いたリズムを考慮した楽譜追跡

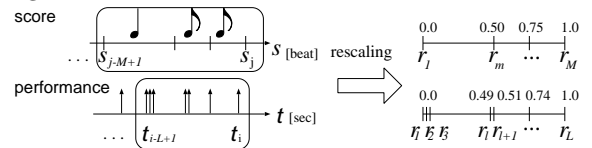


Fig. 5 リスケールした時間と拍によるリズムの比較

#### 4.5 動作検証

現実を使用する場合には実時間で少ない負荷での動作が望まれるので、拍位置の字と確率を与える DP の漸化式の計算は MIDI ノートオンの受信時刻間隔が 0.05 秒以上のときのみに行い、拍位置  $s$  の個数の幅を 32 とする傾斜制限を加え、局所距離の計算に用いるノートオンメッセージの個数を  $N=10$  に固定した。

評価には Burgmüller 作曲の練習曲 Op.100 の全 25 曲の電子ピアノ演奏を用いた。この曲集は初心者用の練習曲であるが、テンポの変化や裝飾音の使用等リズムを含む。演奏誤りのない演奏と、その演奏に人工的に音階の誤りを 5 箇所加えた MIDI データを用意した。楽譜追跡は正解率は、全推定回数に対する正しく推定した回数の割合で計算し、演奏に誤りがある・ないの両方の場合とも平均 99.8% の正解率を得た。24 曲では完全に追跡できた。誤推定は楽譜にフェルマータの指示がある箇所で見られ、テンポ一定してリズムの一致を計算したことが原因であった。実行中には誤りが起きても、その後は直ちに正しい拍位置を推定し正確な楽譜追跡が続けられた。

#### 5 おわりに

本稿では、和音とリズムの確率的なマッチングを与える 2 段 DP を用いた実演奏の MIDI 信号を対象とした楽譜追跡アルゴリズムを議論した。今後は、提案アルゴリズムを用いた自動伴奏システムの設計と開発を行いたい。

#### 参考文献

- [1] R. B. Dannenberg, Proc. ICMC, pp. 193–198, 1984.
- [2] H. Sakoe, IEEE Trans. ASSP-27, pp. 585–595, 1979.