

# HMMによるMIDI演奏の楽譜追跡を用いた自動伴奏\*

武田晴登, 西本卓也, 嵯峨山茂樹 (東大・情報理工)

## 1 はじめに

我々は、電子楽器を演奏する演奏者に合わせて、伴奏楽譜を自動演奏する技術进行研究している。この技術は、楽器の片手ずつの練習、自動連弾、コンチェルトの練習、仮想的アンサンブル演奏など広い応用が見込まれる。さらに演奏者が拍打ちのみを行うならば、仮想指揮システムともなり得る。

演奏者が誤り無く楽譜通りに弾く場合には、実演奏の音を時間順序で楽譜と対応させることによって自動伴奏は容易であり、実際にそのような機能を備えたキーボード製品もある。しかし、現実には、弾き間違い(ミスタッチ)や、(特に練習中は)弾き直しや部分的スキップを含む可能性もあり、そのような演奏に追隨して伴奏するのは単純な問題ではない。

この問題は、演奏者の演奏と楽譜を対応付けて演奏位置を推定する処理、即ち「楽譜追跡 (score following)」と、その結果を用いて伴奏を再生する「自動再生」とに分けて考えられる。局所的な誤りを含む可能性のある演奏の楽譜追跡としては、楽譜と実演奏を音高について DP(Dynamic Programming, 動的計画法) マッチングを行なう手法 [1] が知られている。また、拍に情報に注目した自動伴奏 [2] の研究も行われている。しかし、これらの手法は弾き直しやスキップのように演奏中に演奏箇所が大きく移動する演奏には対応していない。

本稿では、演奏者のキーボード演奏の MIDI (Musical Instrument Digital Interface) 信号を主たる対象とし、第2節で、弾き直しやスキップを含む演奏に対する楽譜追跡を、第3節で、演奏者のテンポ推定結果を用いた伴奏再生方法と具体的なテンポ推定法を論じる。

## 2 弾き直しを含む MIDI 演奏の楽譜追跡

### 2.1 HMM による誤りのある楽譜演奏のモデル

図1左に示すように、演奏している楽譜上の  $i$  番目の音符が存在する拍位置を状態  $s_i$  としてモデル化すると、楽譜を読み進めていく過程は接続する状態を順に遷移していく過程と捉えられる。ここで、ある位置で演奏を誤る(ミスタッチ)確率はそれ以前の音符列履歴の影響を受けないと仮定できるならば、この遷移はマルコフモデルとして扱うことができる。図1右のような、局所的な音符の脱落は状態のスキップ、余分な音符の挿入は同一の拍位置への遷移する過程として扱うことができる。また、さらに、一度弾いた部分の弾き直しは以前の状態への遷移、途中の部分スキップは以後の状態への遷移として扱える。和音の挿入・脱落誤り、弾き直し・スキップは全て状態  $s_i$  から状態  $s_j$  への遷移として記述できるので、これらの演奏に与える確率は、全て遷移確率  $p(s_j|s_i)$  として統一して記述できる。

このような音楽進行の  $i$  番目の拍位置  $s_i$  で、必ずしも楽譜通りの音高と音長が観測されるとは限らず、近隣の音へのミスタッチや臨時記号忘れなどの置換や、音長の変動や誤りの傾向も含めて、音高と音長を表す観測量  $x$  が観測される確率は  $p(x|s_i)$  として表現できる。

演奏者が拍位置の演奏順序  $S = \{s_i\}_{i=1}^I$  を意図して各音を  $X = \{x_i\}_{i=1}^I$  (ただし、 $x_i$  は発音時刻  $t_i$  と音高  $n_i$  からなる) のように演奏する確率を  $P(X|S)$  で表すと、これらの各確率モデルを統合したモデルは、表1に示す対応から HMM (Hidden Markov Models,



Fig. 1 和音内の状態遷移による和音の演奏のモデル化(左)と誤りのある演奏の拍位置の遷移(右)

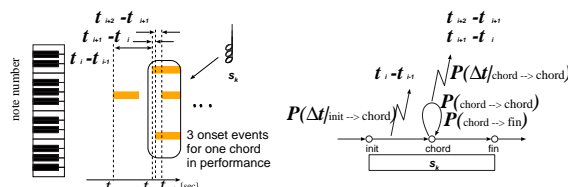


Fig. 2 一和音の演奏の IOI (左) を出力する HMM (右)

Table 1 HMM と演奏生成の確率モデルの対応

HMM	演奏生成モデル
隠れ状態	和音の開始, 和音
状態遷移確率	同時発音数
状態遷移出力	音高, IOI
状態遷移出力確率	和音の一致
HMM	一和音(拍位置)
HMM 遷移	拍位置の演奏順序

隠れマルコフモデル) と捉えることができる。この HMM により、楽譜に従って演奏  $X$  が行われる確率は  $P(X|S)P(S)$  で与えられる。

### 2.2 HMM における和音演奏のモデル

和音の構成音は、現実の演奏では必ずしも同時に発音されるとは限らず、和音を構成する各音は拍位置の付近に、図2左に示すように順不同で発音される可能性がある。あらゆる順序を考慮して拍位置に対応づけるには組合せが多いが、拍位置の状態の内部に設けた状態の自己遷移として和音構成音が発音するようにモデル化すれば、単純化できる。この場合は、楽譜にない音の挿入誤りによる自己遷移と同時発音の自己遷移は同一のモデルで表現されるが、同時発音の自己遷移確率は和音の同時発音数に依存した値を設定することで改善できる。

和音構成音の2つの音の発音時刻  $t_{i-1}, t_i$  の間の IOI (inter-onset interval, 発音時刻間隔)  $t_i - t_{i-1}$  は小さい正值を取ることが多いと考えられる。ここで、同時発音判定閾値を設定し、同時発音の IOI 値はそれ以内のみ分布し、ミスタッチによる挿入はそれより大きい値のみ分布すると仮定すれば、やはり HMM の枠組の中で扱いつつ、楽譜位置の推定の為の確率計算を必要とする仮説数を減らして計算量の削減ができ、実時間処理に有利になる。但し、このように同時発音であるか否かを IOI に設けた閾値で判定すると、閾値以上に時間的にずれた複数の音(アルペジオ奏法など)は和音と見なされないため誤推定を起こす場合もあり得る。閾値は実験的に設定する必要がある。

### 2.3 HMM を用いた MAP 推定による楽譜追跡

演奏者の演奏の入力  $X$  に対して、HMM の中で最も尤もらしい拍位置の系列  $S$  を求める問題は、Bayes の定理を用いて以下のように定式化される。

$$\hat{S} = \underset{S}{\operatorname{argmax}} P(S|X) = \underset{S}{\operatorname{argmax}} P(X|S)P(S)$$

\* Automatic accompaniment using score following of MIDI performance using HMM. by Haruto Takeda, Takuya Nishimoto, Shigeki Sagayama (University of Tokyo)

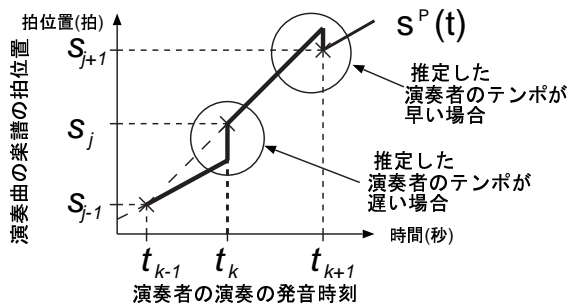


Fig. 3 演奏者した拍位置  $\hat{s}_i$  と過去の演奏者の演奏から推定した  $s^P(t)$  との誤差

従って、楽譜追跡は HMM における事後確率を最大化する状態遷移系列を求める問題となる。この最適解は、DP の原理を応用した時間同期 Viterbi 探索を用いて効率的に求められる。

#### 2.4 性能評価実験

クラシックの室内楽曲である A. Dvořák 作曲のピアノ五重奏曲 Op. 81 第 3 楽章のピアノパートの開始から 57 小節までの部分を、弾き直し・スキップを含む 2 種類の拍位置の順序で各 4 回演奏し、演奏誤りが含まれた演奏を MIDI ファイルとして記録して楽譜追跡の評価を行なった。同時発音の判定は 0.045 秒以下とし、自己遷移確率は演奏曲の楽譜から定め、音高のミスタッチには全音違いの範囲まで 0.1 の確率を与えた。その結果、評価データに対しては全て弾き直しやスキップを正しく追跡でき、推定した拍位置の正解率は各々 95.2%, 97.2% であった。

### 3 楽譜追跡に基づく伴奏の自動演奏

#### 3.1 伴奏再生のための演奏者の拍位置の推定

次に、演奏者に合わせた伴奏の演奏方法を考える。楽譜追跡が正しく行われているならば、伴奏 MIDI の再生器は各伴奏音の発音を任意の時刻に行えるので<sup>1</sup>、演奏者が打鍵した音符に同期した伴奏音符は、ほとんど遅れなく同期させて発音できる。

しかし、演奏者のパートが休符である場合や伴奏の方が細かい音符である場合など、演奏者の打鍵に同期して伴奏発音するわけにはいかない場合も多い。演奏者の打鍵は時間的に離散的であるので、任意の時刻  $t$  の演奏者の拍位置  $s^P(t)$  は楽譜追跡の結果だけからは得られない。即ち、演奏者の演奏の発音イベントがあった時刻  $t_i$  における演奏位置は楽譜追跡  $s^P(t_i) = \hat{s}_i$  により与えられるが、この直後から次の演奏の発音があるまでの間は、演奏者の拍位置は陽には分からない。

逆に、伴奏音の発音時刻  $t$  で演奏者の拍位置  $s^P(t)$  が与えられれば、伴奏の拍位置  $s^A(t)$  をそれに一致させることができる。テンポは時間的には緩やかに変動する傾向があることを考慮して、この短い区間ではテンポがほとんど一定値  $r(t_i)$  (拍/秒) であると仮定すると、その直後の時刻  $t = t_i + \Delta t$  の拍位置は

$$s^P(t_i + \Delta t) = s^P(t_i) + r(t_i) \cdot \Delta t \quad (2)$$

と近似できる。ただし、この場合、次の演奏者の演奏の入力があったときに、そのときの楽譜追跡により求められた拍位置と伴奏の拍位置との間には図 3 に示すように、誤差が存在し得る。これらの誤差は、伴奏再生において不自然な音飛びや音のくり返し、もしくは拍の時間長の伸縮に対応する。本節ではこの誤差を最小化するテンポ推定を議論する。

#### 3.2 楽譜追跡結果からの演奏者のテンポの推定

演奏の入力直後から次の入力があるまでの間の演奏者の拍位置  $s^P(t)$  を推定するために、図 3 にお

<sup>1</sup> 正確には、実装しているプログラムの時間制御には数ミリ～数十ミリ秒程度の遅れが存在するが、和音における発音時刻のずれと同程度であるので、ほぼ「同時」と見做せる。

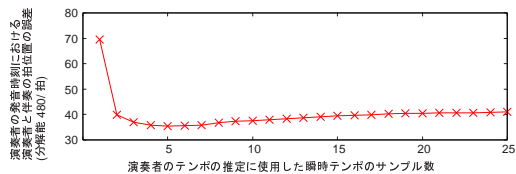


Fig. 4 演奏者の発音時刻における演奏者と伴奏の拍位置の誤差

る誤差が少なくなるような式 (2) における一定テンポ  $r^P(t_i)$  の推定を考える。

テンポは局所的に一定であると見なすと、 $t_i$  より前の演奏された音の各 IOI のテンポ  $r_j^I = \frac{s^P(t_j) - s^P(t_{j-1})}{t_j - t_{j-1}}$

は、テンポ  $r^P(t_i)$  と近い値を取ると考えられる。厳密に一致しないのは、演奏者が無意識にテンポを変動させたり、発音時刻のタイミングがずれたり、あるいは、テンポがこの観測中に変化しているからである。仮に、この誤差が正規分布に従って確率的に変動するもの  $r_j^I - r^P(t_i) \sim N(0, \sigma^2)$  として扱えるなら

ば、最尤推定値  $\hat{a} = \sum_{m=1}^M r_{i-m}^I / M$  が得られる。

#### 3.3 演奏者の発音時刻における伴奏と演奏者の拍位置のずれの評価

この一定テンポの推定は、サンプル数が少なければ局所的なテンポの変動を受け、また、サンプル数が多いと長時間のテンポにわたるテンポの変化の影響を受け、拍位置の誤差が大きくなると考えられる。今回は拍位置の予測に適切なサンプル数に求める為に、テンポ推定に用いるサンプル数と拍位置の誤差との関係を求めた。

楽譜追跡結果に基づいて推定したテンポから  $s^P(t)$  を推定し、演奏者の次の入力音があったときの演奏者と伴奏との間の拍位置の 2 乗誤差  $\sqrt{\sum_i (\hat{s}_i - s^P(t_i))^2}$  を計算した。評価曲には、Burgmüller 作曲の「練習曲」から選んだ 19 曲の電子ピアノによる演奏データを用いた。その結果、図 4 に示すように、拍位置のずれは  $M = 4$  で最小値  $35.4/480 = \text{約 } 0.07$  (拍) となり、それ以上のサンプル数を用いると逆に予測精度が僅かに落ちる傾向が見られた。

#### 3.4 システムの実装

演奏者が伴奏に合わせるのではなく伴奏が演奏者に合わせて演奏する場合を想定し、楽譜追跡と伴奏テンポの設定手法を直列に接続した自動伴奏システムを実装した。楽譜追跡に必要なとされる拍位置と音高は、標準 MIDI ファイルにより与える仕様として実装した。複数の楽曲の伴奏 MIDI ファイルと演奏者の楽譜情報に対応する MIDI ファイルを用意し、これらを用いてシステムの動作を確認した。

### 4 おわりに

本稿では人間の演奏者の電子楽器演奏の MIDI 信号を対象とした楽譜追跡手法、及び、演奏者のテンポ推定を行なう伴奏再生手法について議論した。今後は、具体的な伴奏者の数理モデルや伴奏システムの性能評価方法を検討したい。

#### 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ました。また、自動伴奏のソフトウェアの実装については、戦略ソフトウェア創造人材養成教育コースにおいて、平木敬教授、金子勇元助手をはじめとする教育コースのスタッフの方々から有益な助言を頂きました。

#### 参考文献

- [1] R. B. Dannenberg, Proc. ICMC, pp. 193–198, 1984.
- [2] C. Raphael, Proc. ICMC, 2001.