Vol.2015-MUS-108 No.16
2015/9/2

階層的確率生成モデルによる装飾音を含む多声MIDI音楽演奏の楽譜追跡

中村 栄太[1,a)]　Philippe Cuvillier[2]　Arshia Cont[2]　小野 順貴[3]　嵯峨山 茂樹[1]　渡邊 健二[4]

概要：本研究報告では，確率モデルによる多声 MIDI 演奏の記述に基づく楽譜追跡手法について議論する．音楽演奏の確率モデル化により楽譜追跡を行う手法は，既に広く研究され現在最も成功しているアプローチの一つである．本研究では，MIDI レベルの音楽演奏を楽譜中の演奏者の進行と個々の演奏音符の生成過程を階層的に統合したモデルにより記述し，それが隠れセミマルコフモデルの拡張モデルにより表されることを議論する．この演奏モデルを従来研究されてきた隠れマルコフモデル (HMM) に基づく演奏モデルと比較し，現モデルが特にトリル，トレモロ，アルペジオを含む演奏の楽譜追跡により有利であることを説明する．続いて，実データを用いた比較評価および誤り解析によりこの主張を経験的・定量的に実証する．また，計算量の問題を論じた上で，このモデルの有利な点を残しつつ計算時間を削減可能な，このモデルと HMM との混成モデルも構成する．考案のモデルにより，装飾音による楽譜追跡の誤りが大幅に低減された．

1. はじめに

音楽演奏と対応する楽譜の実時間マッチング問題（楽譜追跡と呼ぶ）は，自動伴奏などを目的として過去約 30 年間多く研究されてきた [1], [2], [3], [4], [5], [6], [7], [8], [9]．本研究では多声音楽の記号的 (MIDI) 演奏の楽譜追跡を扱う．楽譜追跡の中心的課題は，演奏の多様性を適切に，計算効率の良い方法で捉えることである．近年では，確率モデルによりこの多様性を記述し，柔軟な楽譜追跡アルゴリズムを導出する手法が広く用いられている（2.1 節と文献 [3] を参照）．

確率モデルの一種である隠れマルコフモデル（Hidden Markov Model; HMM）は，記号的演奏の楽譜追跡に応用され，現在最高精度の結果を出している [4], [5], [8], [10]．これらの文献で使われたモデルでは，楽譜中の音符，和音，トリル等の音楽的イベント (以下では単にイベントとも呼ぶ) が状態として表され，演奏音符は背後の状態遷移過程からの出力として記述される．また計算量の効率化のため，出力確率と遷移確率の両方に無記憶性を持つ統計的依存性が仮定される．こうした単純化のため，これらのモデルではイベントごとに演奏される音符の数やトリル全体の音長 (持続時間) などの演奏に重要な側面がうまく記述できない．

現象論的には，演奏は階層的な音符生成の過程と見なせる．即ち，楽譜内での演奏者の進行を音楽的イベント単位で記述した上層のモデルと，個々の演奏音の生成を記述する下層のモデルに分けて考えることができる [10], [11]．本研究では，この過程を隠れセミマルコフモデル（Hidden Semi-Markov Model; HSMM）[12] の自己回帰的な拡張モデル [13] によって記述し，演奏の上記の側面を捉えた多声部音楽演奏のモデルを構成する (2 節)．この演奏モデルは，いくつかの単純化により従来研究されている HMM に基づく演奏モデル [10] に還元される．3 節では，これらのモデルを捉えられる情報およびアルゴリズムの観点から比較し，本モデルが特にトリルやトレモロ，アルペジオを含む演奏の楽譜追跡により有利であることを説明する．また実データを用いた評価と誤り解析を通して，これを定量的に確認する（4 節）．最後に今後の課題と見通しを議論する

[1] 明治大学
164-8525 東京都中野区中野
[2] フランス国立音響音楽研究所
IRCAM, 75004 Paris, France
[3] 国立情報学研究所
101-8430 東京都千代田区一ツ橋
[4] 東京藝術大学
110-8714 東京都台東区上野公園
[a)] eita.nakamura@gmail.com

(5 節).

本研究は，著者の一部らが文献 [14] で発表した研究が発展したものである．本研究とこの研究の間をつなげる関連研究として文献 [8], [10] があるが，紙面の都合上これらの研究結果は本稿では詳しく取り上げることができなかった．特にアルゴリズムの詳細や装飾音の不確定性，モデルパラメータ推定を兼ねた演奏解析結果に関して，本稿では参照にとどめた部分の詳細に関してはこれらの文献を読んで頂きたい．

2. モデル

2.1 確率モデルに基づく音楽演奏の記述

楽譜の記述法に内在する不確定性や奏者や楽器の動きに含まれる不確実性により，楽譜に基づく音楽演奏には多様性がある．これらの不確定性や不確実性は，テンポや発音時刻のノイズ，強弱，アーティキュレーション，装飾音の演奏法や演奏誤りや弾き直し・弾き飛ばしの仕方の中に存在する．正確で柔軟性を持った楽譜追跡を行うには，アルゴリズムにこの演奏の多様性を捉えた規則を (明示的とは限らずとも) 組み込むことが必要である．

このための方法の一つは，音楽演奏の確率モデルを構成し，不確定性や不確実性を確率の言葉で記述することである．この方法では，演奏モデルの確率的推論問題として楽譜追跡アルゴリズムを導出できる．確率モデルを用いた手法は多くの先行研究で有用性が確認されており，本研究では以下この手法を議論する．

2.2 演奏者の楽譜内の進行のモデル

以下，モデルの詳細を述べる．音楽演奏を上層と下層の 2 つのモデルを階層的に組合せて記述することにする．上層のモデルは，「音楽的イベント」を単位とした演奏者の楽譜内の進行を記述するものとする．ここで音楽的イベントとは，和音 (アルペジオ奏法を含んでもよいとする) [*1]，トリル/トレモロ，前打音，後打音を表すものとし，これらそれぞれを 1 つの状態 (上層状態と呼ぶ) として表現する．添字 i で上層状態を表す．演奏者の進行は，これらの状態間の一連の遷移過程として表現でき，これを以下 $i_{1:N} = (i_1, \ldots, i_N)$ と記すことにする (ここで N は演奏される MIDI 音符の総数を表す)．なお，発音時刻の順に入力される演奏音符の添字を $n(= 1, \ldots, N)$ とし，対応する音楽的イベントを i_n と記す．

演奏の進行の統計的性質は，確率 $P(i_{1:N})$ により表現される．実際に動作可能な計算量を持つアルゴリズムを得るためには，演奏モデルを単純化する必要がある．一般的には，この確率が遷移確率の積に分解されるという仮定をおく：$P(i_{1:N}) = \Pi_{n=1}^{N} P(i_n|i_{n-1})$ (ここで $P(i_1|i_0) \equiv P(i_1)$ は初期確率を表すものとする)．遷移確率 $P(j|i)$ は，次のイベントへの通常の進行 $(j = i + 1)$ やイベントの挿入誤り $(j = i)$ 及び脱落誤り $(j = i + 2)$ や弾き直し及び弾き飛ばし $(|j - i - 1| > 1)$ の相対的な頻度を表す．これらの確率値は，演奏データから推定することができる．文献 [8] では，$P(i|j)$ が $i - j$ にのみ依存すると仮定した場合の確率値がピアノ演奏データにより求められている (文献 [8] の表 3).

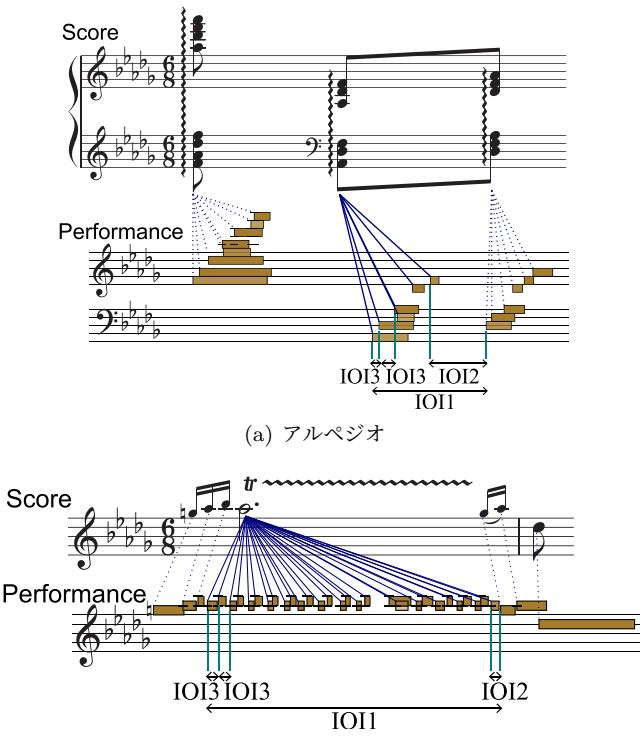


図 1 音楽的イベントとその演奏音符の例．3 種類の発音間時間 IOI1，IOI2，IOI3 は本文中に説明されている．

2.3 個々の演奏音符生成のモデル

下層のモデルは各音楽的イベント内で演奏される音符の生成過程を記述する．強弱とアーティキュレーションは一般的に不確定性が大きく，楽譜追跡への重要性は比較的小さいため，ここでは音高と発音時刻に注目し，それぞれ p_n と t_n と記す．和音やトリルでは複数の音符が演奏される (図 1)．和音は楽譜上では同時発音の音符の集まりであるが，MIDI 信号の演奏音符は順序化され，完全に同期していない．よって p_n は常に一つの音高を表すものとする．

まずイベント毎に演奏される音符数を考えよう．和音や

[*1] 本研究では，以下「和音」を楽譜上で同時に発音される全ての音符の集まりと定義する．この定義により，本研究では和音は単音を表すこともある．

前打音，後打音の場合は演奏されるべき音符数は決まっているが，演奏誤りにより音符の追加や削除が起こった場合には音符数は変わり得る．トリルや (不確定音価の) トレモロの場合は，装飾音の演奏速度によって演奏音符数も変化する．これらを踏まえ，イベント毎の演奏音符数を確率分布 $d_i(s)$ により記述する (s は演奏音符を表す確率変数であり，$\sum_{s=1}^{\infty} d_i(s) = 1$ を満たす)．例えばイベント i が和音の場合，$d_i(s)$ は楽譜に記された音符数に s のピークを持つ．イベント i が 1 音からなるトリルの場合は，ピークの位置は $s_i^{\mathrm{peak}} \simeq \nu_i v/\delta t_{\mathrm{trill}}$ と書ける (ここで $\delta t_{\mathrm{trill}}$，$\nu_i$，$v$ はそれぞれ連続したトリル音の平均発音間時間 (Inter-Onset Interval; IOI)，イベント i の音価，単位音価当たりの秒数を単位とするテンポ (の逆数) を表す)．現在のところ $d_i(s)$ の分布形を定めるのに十分な経験知識がないため，以下では標準偏差を調整可能なパラメータとする正規分布を仮定する：$d_i(s) = N(s_i; s_i^{\mathrm{peak}}, \sigma_i)$.

次に演奏音符の音高の確率を考える．イベント i の演奏音符の音高を確率分布 $P_i^{\mathrm{pitch}}(p)$ として，計算効率のため各演奏音符に関して分布は独立と仮定する．楽譜上と一致しない音高に対する小さな確率値によって，音高誤りの頻度が表現できる．文献 [8] (式 (30)) では，全ての楽譜音符について演奏誤り確率が同じであり，演奏誤りはいくつかの典型的な場合に分類されるという仮定の下，分布 $P_i^{\mathrm{pitch}}(p)$ の近似型がピアノ演奏データにより推定されている．

最後に演奏音符の発音時刻の記述を考える．演奏の確率は時刻全体をずらしても不変であるという自然な仮定により，モデルは時間間隔のみに依存することが要求される．音楽演奏の発音時刻を局所的に記述するのに関連する時間間隔には (少なくとも) 次の 3 種類がある：(IOI1) 連続したイベントの各々の最初の演奏音符間の IOI，(IOI2) あるイベントの最初の演奏音符と直前のイベントの最後の演奏音符との間の IOI，(IOI3) イベント内での連続した演奏音符間の IOI (図 1)．これらの IOI の確率が現在と直前の状態のみに依存するという単純化の仮定により，確率は $P_\kappa(\delta t|i_{n-1}, i_n, v)$ ($\kappa = \mathrm{IOI1}, \mathrm{IOI2}, \mathrm{IOI3}$) の関数形を持つ (ここで δt と v は対応する IOI とテンポを示す)．IOI3 の時間間隔はおおよそ関連するイベントのみに依存し，テンポや他の文脈にはほぼ独立であるという経験的知識により，対応する分布の関数形はさらに $P_{\mathrm{IOI3}}(\delta t|i_n)$ と簡単化できる．全ての時刻の履歴が保たれている場合，IOI1 と IOI2 は互いに独立な量ではないことは注意が必要である．しかし，下述するマルコフ的な記述では，これらは異なる重要性を持つ．

2.4 自己回帰隠れセミマルコフモデル

2.2 節と 2.3 節のモデルの統合は隠れセミマルコフモデルの拡張モデルにより記述できる．いくつかの等価な定式化の 1 つでは [15] (また文献 [12] の 3.3 節)，セミマルコフモデルは拡張された状態空間をもつマルコフモデルとして表現される．この拡張状態空間は上層状態 i とイベント内の演奏音符数を表す変数 $s = 1, 2, \ldots$ [*2]のペア (i, s) の全体により表され，遷移確率は次のように書かれる：

$$P(i_n, s_n|i_{n-1}, s_{n-1}) = \delta_{s_n,1} P(i_n|i_{n-1}) P_{i_{n-1}}^{\mathrm{exit}}(s_{n-1}) + \delta_{s_n, s_{n-1}+1} \delta_{i_n, i_{n-1}} \left(1 - P_{i_{n-1}}^{\mathrm{exit}}(s_{n-1})\right) \quad (1)$$

ただし，

$$P_i^{\mathrm{exit}}(s) = d_i(s)/\Sigma_{s'=s}^{\infty} d_i(s'). \quad (2)$$

ここで式 1 の δ は Kronecker のデルタである．式 (2) の退出確率は，演奏者がイベント i で既に s 音符演奏している時に，次の演奏音符で他のイベントに移る確率である．式 (1) の右辺の第 1 項はイベント i_{n-1} で s_{n-1} 音符演奏した後に，イベント i_n に移る確率を表し，第 2 項はイベント i_n で s_{n-1} 音符演奏した後にさらに同じイベントに留まる確率を表している．このように，このモデルは楽譜中の演奏者の進行と各イベントでの演奏音符生成を統合的に記述する．

演奏音符の音高と発音時刻の確率は．このセミマルコフ過程に付与された出力確率により記述できる．計算効率を目的とした単純化のため，音高と発音時刻の確率は互いに独立であると仮定する．音高の出力確率は次の様に与えられる：$P(p_n|i_n, s_n) = P_{i_n}^{\mathrm{pitch}}(p_n)$.

n 番目の演奏音符の発音時刻の出力確率は，次で与えられる：

$$P(t_n|i_n, s_n, i_{n-1}, s_{n-1}, v, t_{1:n-1}) = \begin{cases} w_1 P_{\mathrm{IOI1}} + w_2 P_{\mathrm{IOI2}}, & s_n = 1; \\ P_{\mathrm{IOI3}}, & s_n \neq 1. \end{cases} \quad (3)$$

ここで，

$$P_{\mathrm{IOI1}} = P_{\mathrm{IOI1}}(t_n - t_{n-s[n-1]}|i_n, i_{n-1}, v), \quad (4)$$

$$P_{\mathrm{IOI2}} = P_{\mathrm{IOI2}}(t_n - t_{n-1}|i_n, i_{n-1}, v), \quad (5)$$

$$P_{\mathrm{IOI3}} = P_{\mathrm{IOI3}}(t_n - t_{n-1}|i_n)\delta_{i_n i_{n-1}} \quad (6)$$

であり，表記上 $s[n-1] = s_{n-1}$ と表した．この 3 つの場合分けは，2.3 節で説明した 3 種類の IOI にそれぞれ対応する．IOI1 と IOI2 に対する確率両方が楽譜追跡に重要であるた

[*2] 注意：ここのモデルでは，s は音楽的イベント内で演奏される音符の数を表しており，これはそのイベントの (例えば秒単位の) 継続時間ではない．

め，これらの確率の混合分布を用いている ($w_1 + w_2 = 1$)．上の様に出力確率が過去の出力にも依存するモデルは音声処理の分野でも研究されており，そこでの慣習に基づき以下では以上のモデルを自己回帰 (Autoregressive; AR) 隠れセミマルコフモデル (HSMM) と呼ぶ．

分布 P_{IOI1}，P_{IOI2} と P_{IOI3} は演奏データの解析により推定できる．P_{IOI2} と P_{IOI3} はピアノ演奏データを用いてこれまでに推定されている [10]．この文献では，最も重要な場合である $i_n = i_{n-1}+1$ (誤りなしの次のイベントへの遷移) の時，$P_{\mathrm{IOI2}}(\delta t|i{+}1, i, v)$ は次の形の Cauchy 分布により良く近似されることが示されている．

$$\mathrm{Cauchy}(\delta t; v(\tau_i^{\mathrm{end}} - \tau_i) - \mathrm{dev}_i, 0.4\ \mathrm{s}). \qquad (7)$$

ここで $\mathrm{Cauchy}(x; \mu, \Gamma)$ は中央値 μ と半値幅 Γ を持つ Cauchy 分布関数を表し，τ_i はイベント i の発音楽譜時刻，τ_i^{end} はイベント i 内で演奏音の発音が継続される楽譜時刻の上限を表す．また dev_i はイベント i での装飾音による IOI のずれを表し，その期待値は前打音やアルペジオされる音符数とこれらの音符の平均 IOI の積により与えられる．この結果により $i_n = i_{n-1}+1$ の場合の P_{IOI1} は

$$P_{\mathrm{IOI1}}(\delta t|i{+}1, i, v) = \mathrm{Cauchy}(\delta t; v\nu_i, 0.4\ \mathrm{s}) \qquad (8)$$

で近似的に与えられる (ここで $\nu_i = \tau_{i+1} - \tau_i$ はイベント i の音価)．分布 P_{IOI3} は和音内の音符や装飾音の IOI の測定により推定されている (文献 [10] の 3.3 節と 4.2 節を参照)．

最後にテンポ v_n は，スイッチングカルマンフィルター ([10]，3.4 節) により逐次的に推定できる．以上をまとめると，完全データ確率 $P(i_{1:n}, s_{1:n}, t_{1:n}, p_{1:n})$ は次の積の反復式により与えられる：

$$\prod_{m=1}^{n} \Big[P(t_m|i_m, s_m, i_{m-1}, s_{m-1}, v_{m-1}, t_{1:m-1}) \cdot P(i_m, s_m|i_{m-1}, s_{m-1}) P_{i_m}^{\mathrm{pitch}}(p_m) \Big]. \qquad (9)$$

3. 他の手法との比較

3.1 HMM に基づく手法との比較

これまで提案されてきた MIDI 演奏の楽譜追跡手法で最も高精度なものの一つは，通常の HMM による演奏モデルに基づいて開発されている [10]．本研究のモデルはこの HMM の演奏モデルを 2 つの点で拡張したものと見なせる．まず，HMM の遷移確率は式 (1) の遷移確率で，退出確率 $P_i^{\mathrm{exit}}(s)$ が s に関して定数である特別な場合に対応する．特にこの確率はイベント i 内の演奏音符数の期待値の逆数で与えられる．よく知られている通り，この制約は $d_i(s)$ が幾何分布で $s = 1$ にピークを持つことを示しており，大きな和音や長いトリル/トレモロの場合には特に悪い近似となる．

2 つめの違いは，発音時刻の出力確率の構造である．通常の HMM では，出力確率に関してもマルコフ性が仮定される．これにより，このモデルでは IOI2 と IOI3 のみが記述可能であり，式 (3) の IOI1 の確率分布は省かれている．言い換えれば，HMM では IOI の出力確率において $w_1 = 0$，$w_2 = 1$ とおいた場合に相当する．よって HMM では，トリル/トレモロの全体の音長やアルペジオの音長がうまく捉えられない．

これらのモデルの記述の差異は，楽譜追跡への応用で重要な効果を持つ．楽譜追跡では一般的には音高の情報が最も重要であるが，似た様な音高が含まれる音楽的イベントが連続している場合は，発音時刻やイベントごとの演奏音符数が正しく音符をマッチする上で重要な情報となる．例えば，トリル/トレモロが連なった部分の楽譜音符間マッチングを正しく行うには，イベント毎の演奏音符数や各トリル/トレモロの音長が重要な視点となる．HMM ではこれらがうまくモデルに組み込まれていないため，この場合は自己回帰 HSMM の方がより効果的であると期待される．同様のことはアルペジオが連なっており，IOI2 と IOI3 が演奏ごとで大きな変動をもつ場合にも当てはまる．これに対して，通常の和音の連なりの場合は，IOI1 と IOI2 はほとんど等しく，これらは和音のクラスタリングに必要な情報の大部分を担っている．よって，装飾音を含まない楽節の場合は，両方のモデルは同様の効果を持つと期待される．

3.2 前処理の手法との比較

楽譜追跡における装飾音の取り扱いの問題の解決法として，前処理を用いる手法が提案されている [16]．この手法の考え方は，前処理により装飾音符が楽譜と演奏のマッチングをするモジュールに直接送られない様にするというものである．提案者自体が分析している通り，この方法は装飾音があまり複雑に多声的に絡み合っていない楽譜や演奏誤りが少ない場合にはうまく機能する一方で，装飾音の周りで演奏誤りがあった場合や弾き直し・弾き飛ばしが行われた場合には前処理が失敗することがある．先行研究では，前処理の手法と HMM の手法の直接比較評価が行われ，誤りや弾き直し・弾き飛ばしを含むピアノ演奏に対して HMM の優位性が示されている [10]．そこで，本研究では 4 節で提案したモデルと HMM との比較をする．

3.3 計算コスト

楽譜追跡では，入力演奏に対して最大確率を与える隠れ状態系列を探索する．実時間処理を実現するためには，こ

の推定アルゴリズムの計算コストは十分小さい必要がある．本節では以下，提案したモデルと 3.1 節で述べた HMM とを計算コストに関して比較する．

HMM の状態推定問題には Viterbi アルゴリズムが適用できる．遷移確率と出力確率の積を $a_{ij}(o) = P(j|i) \cdot P(o|i,j)$ と記す (o は音高と発音時刻を表す)．Viterbi アルゴリズムの更新式は次の漸化式で表現される：

$$\hat{p}_N(i_N) \equiv \max_{i_1,\ldots,i_{N-1}} \left[\prod_{n=1}^{N} a_{i_{n-1}i_n}(o_n) \right] \tag{10}$$

$$= \max_{i_{N-1}} \left[\hat{p}_{N-1}(i_{N-1}) a_{i_{N-1}i_N}(o_N) \right]. \tag{11}$$

状態は音楽的イベントと対応するため，状態数は N である．弾き直し・弾き飛ばしを含む楽譜内の任意の進行を許す場合，Viterbi アルゴリズムを直接適用すると各更新の差異に $O(N^2)$ 回の確率計算が必要となる．ところが，もし $a_{ij}(o)$ が幅 D の帯行列 α_{ij} と 2 つのベクトル S_i，r_j の直積の和で書ける場合には，アルゴリズムの組換え手法により計算量を $O(DN)$ に削減できることが知られている [8]．直感的には，α_{ij} は比較的大きな確率値を持つ近傍の状態間の遷移に対応し，S_i と r_j は通常非常に小さな確率値を持つ遠くへの弾き直しや弾き飛ばしに対応する．$a_{ij}(o) = \alpha_{ij} + S_i r_j$ を式 (11) に代入すると，α_{ij} は $O(DN)$ の計算量を誘発し，$S_i r_j$ は組換えにより $O(N)$ の計算量を誘発することが分かる．この単純化した遷移確率行列は，長い楽譜での実時間処理を可能とするために使われてきた．

さて，2.4 節での自己回帰 HSMM の定式化により明らかな通り，このモデルにも通常の Viterbi アルゴリズムを適用できる．実際上は，各イベント i に対して演奏音符数の上限値 $s_i^{\max}$ を設ける．すると HSMM の状態数は $\sum_i s_i^{\max} \equiv SN$ となる (S は $s_i^{\max}$ の平均)．式 (1) の遷移確率は特殊な形のため，Viterbi アルゴリズムの計算量は一般には $O(SN^2)$ となる．上記と同様の組換え法を適用すると，直積型の遷移確率の場合には計算量は $O(DSN)$ に削減できる．結果的に，自己回帰 HSMM の計算量は縮小モデルである HMM に比べて約 S 倍大きい．例えば，$s_i^{\max}$ をイベント i の音符数の期待値の 2 倍とすると，中程度の度合いの多声部の楽譜の場合 $S \simeq 3$–10 であり，大きな和音や長いトリル/トレモロが多くある楽譜では S はより大きい値となる．

3.4 HMM・HSMM 混成モデル

3.1 節で議論した通り，現モデルは HMM に比べて楽譜追跡でより良い結果をもたらす論理的理由があるが，その反面で計算量は増大し，これは長い楽譜に対しては好ましくない．一方で，大部分の音楽的イベントは (単音を含む) 普通の和音であり，HMM をそのまま適用しても良い結果が得られることが分かっている．そこでもし普通の和音に対しては HMM の状態表現を適用し，装飾音を含むイベントに関しては自己回帰 HSMM の状態表現を適用し，これらを結合できれば，最小限の計算量増大により楽譜追跡の改良法が得られるであろう．この様な HMM と HSMM の結合は HMM・HSMM 混成モデル (英語では Hidden Hybrid Markov/Semi-Markov Model; 以下混成モデルとも呼ぶ) [6], [17] により行える．混成モデルでは，通常の和音は HMM 状態で表され，トリル，トレモロ，アルペジオ，前打音と後打音は HSMM 状態で表される．このモデルの Viterbi アルゴリズムの計算量は，自己回帰 HSMM に対する表記 $S = \sum_i s_i^{\max}/N$ において i が HMM 状態の場合に $s_i^{\max} = 1$ を代入した値となる．

表 1 楽譜追跡の誤り率 (%)．自己回帰隠れセミマルコフモデル ("HSMM")，混成モデル ("Hybrid")，HMM [10] の 3 種類のモデルによる楽譜追跡の結果を示す．上側の 4 種類の楽曲は，Couperin の "Allemande à deux clavecins", Beethoven のピアノ協奏曲第 1 番，Beethoven のピアノ協奏曲第 2 番，Chopin のピアノ協奏曲第 2 番 [10] を表し，下側の 2 つの楽曲は本文中に説明がある．

楽曲	音符数	HSMM	Hybrid	HMM
Couperin	1763	5.50	6.02	6.66
Beethoven 1	17587	3.16	3.13	3.16
Beethoven 2	5861	2.01	2.20	2.35
Chopin	16241	9.22	9.22	11.1
Debussy	3294	3.64	3.58	4.66
Tchaikovsky	2245	0.40	0.40	4.55

表 2 タイプ別のマッチング誤りの回数．タイプの分類は本文中を参照．モデルの略語は表 1 と同様．

タイプ	音符数	HSMM	Hybrid	HMM
トリル	8159	282	281	508
トレモロ	2603	115	115	151
アルペジオ	1081	36	33	127
その他の装飾音	2401	340	339	362
その他	32030	1580	1599	1673

4. 楽譜追跡の精度比較

これまで議論したモデルを楽譜追跡の精度の点で評価・比較するために，自己回帰 HSMM (2.4 節)，混成モデル (3.4 節)，と HMM [10] に基づく 3 つの楽譜追跡アルゴリズムを実装し，様々な装飾音を含む音楽演奏に対してそれらを走らせた．文献 [10] で用いられた演奏誤りや，弾き直し・弾き飛ばしを含むピアノ演奏データに加え，連続したトレモロを含む Debussy の「白と黒で」(第 2 楽章の第 1 ピアノパート) と大きなアルペジオの連続を含む Tchaikovsky

表 3 Viterbi アルゴリズムの更新 1 回に要した平均処理時間 (ms). モデルの略語は表 1 と同様.

楽曲	HSMM	Hybrid	HMM
Couperin	1.6	1.1	0.3
Beethoven 1	5.9	2.9	1.1
Beethoven 2	7.0	3.0	1.6
Chopin	7.1	3.5	1.2
Debussy	0.9	0.8	0.1
Tchaikovsky	1.2	1.0	0.1

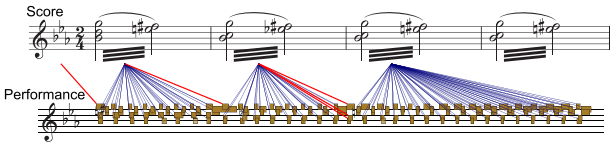
のピアノ協奏曲第 1 番のソロピアノパート (第 2 楽章の終わりのセクション) の演奏を録音したものをデータとして用いた.

自己回帰 HSMM および混成モデルにおけるパラメータ σ_i は，トリル/トレモロに対して $\sigma_i = 0.4 s_i^{\mathrm{peak}}$，それ以外の場合は $\sigma_i = 1$ と設定した．また発音時刻の出力確率のパラメータは，$w_1 = w_2 = 1/2$ とした．これらの値は基準として設定されたが，さらなる最適化の余地は残されている．

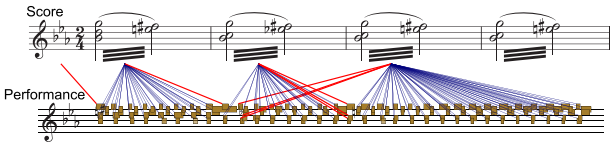
評価基準には誤り率を用いた．即ち，ミスマッチの音符数を総演奏音符数で割った比率を計算した．ただし，演奏音符の中には，専門家にとってもどの楽譜音符にも関係づけにくいものが少数あり，これらは入力データには含めたが，誤り率の計算の際には除外した．表 1 に示した結果から，自己回帰 HSMM と混成モデルは同様の精度であり，HMM は総じて精度が低くなったことが分かる．詳細な誤り解析のために，分類ごとのマッチング誤りの頻度を表 2 に示す．ここで数字はデータ全体の中でのそれぞれのタイプの誤りの個数を示している．装飾音符は，表中の上方の 4 つのタイプに分類し，その他の音符は最後のタイプにまとめられている．これにより，最初の 3 つのタイプ (トリル，トレモロ，アルペジオ) のマッチング誤りは有意に低減されており，その他の誤りは減ってはいるが減少率は比較的わずかであることが確認できる．

図 2 には，データ中で自己回帰 HSMM が HMM よりも良い結果を出した典型的なケースが 2 例示されている．最初の例では，楽節は類似した音高を持ったトレモロが連なったものである．HMM ではマッチング誤りだったいくつかの音符が自己回帰 HSMM では正しくマッチしていることが分かる．同様に 2 つめの例では，大きく両手にまたがったアルペジオの連続の楽節において，HMM では起きたマッチング誤りが自己回帰 HSMM では全てなくなっていることが分かる．これらは，3.1 節での議論と整合性を持つ結果である．

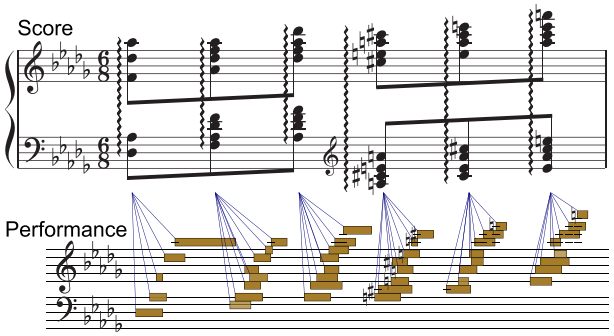
各モデルによるアルゴリズムが要した処理時間も測定した (表 3)．全ての場合において Viterbi アルゴリズムの計算時間は，どの演奏音符に対しても定数であるため，その



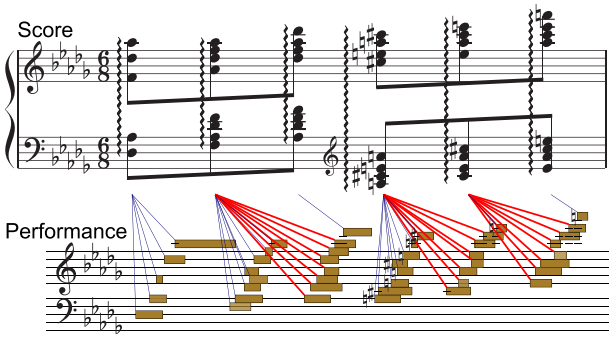
(a) Debussy: 「白と黒で」の一部に対する自己回帰 HSMM での結果.



(b) Debussy: 「白と黒で」の一部に対する HMM での結果.



(c) Tchaikovsky: ピアノ協奏曲第 1 番の一部に対する自己回帰 HSMM での結果.



(d) Tchaikovsky: ピアノ協奏曲第 1 番の一部に対する HMM での結果.

図 2 自己回帰 HSMM および HMM [10] による楽譜追跡の結果例. マッチング誤りは赤い太線で示されている.

平均値を示している．なお，測定には中程度の計算能力を持つノートパソコンを用いた．結果により，混成モデルは自己回帰 HSMM に比べ計算時間が小さく，HMM に比べ精度が高く，楽譜追跡に実用上有利であるという予想が正しいことを確認できた．

5. 結論

本研究報告では，多声 MIDI 演奏の楽譜追跡を行うために，音楽演奏を階層的な確率過程として記述する手法について論じた．提案した自己回帰 HSMM に基づく記号的音楽演奏のモデルが従来研究されてきた HMM に基づくモデルに比べて楽譜追跡において有利である論理的な理由を説明し，これを実データを用いた比較評価および誤り解析により経験的にも確認した．本文で説明した通り，セミマルコフモデルは拡張された状態空間を持つマルコフモデルと見なせるため，HMM による楽譜追跡手法のために開発されている改良法 [8], [18] は現モデルにも適用可能である．特に，これにより弾き直し・弾き飛ばし直後のマッチング誤りや声部間非同期性などにより局所的に音符の順序入れ換えが含まれる演奏でのマッチング誤りのさらなる低減が期待される．

この先の進展として，現モデルを採譜や関連した問題へ応用することが考えられる．現モデルでは装飾音の音長やイベント内部の時間構造を捉えられるため，楽譜が与えられていない演奏における装飾音の検出やその結果を用いた採譜ができる可能性がある．

謝辞 有益な議論をして頂いた齋藤康之氏と中村友彦氏に感謝する．また評価のためのピアノ演奏データ収集に協力頂いた山中歩夢氏と早坂忠明氏にも感謝する．本研究の一部は，国立情報学研究所の 2014 年度 MOU 予算及び，JSPS 科研費 26240025，25880029，15K16054 の助成を受けて行った．

参考文献

[1] R. Dannenberg, “An on-line algorithm for real-time accompaniment,” *Proc. ICMC*, pp. 193–198, 1984.
[2] B. Vercoe, “The synthetic performer in the context of live performance,” *Proc. ICMC*, pp. 199–200, 1984.
[3] N. Orio, S. Lemouton and D. Schwarz, “Score following: State of the art and new developments,” *Proc. NIME*, pp. 36–41, 2003.
[4] B. Pardo and W. Birmingham, “Modeling form for online following of musical performances,” *Proc. of the 20th National Conf. on Artificial Intelligence*, 2005.
[5] 武田晴登, 西本卓也, 嵯峨山茂樹, “HMM による MIDI 演奏の楽譜追跡と自動伴奏,” 情報処理学会研究報告, MUS, pp. 109–116, 2006.
[6] A. Cont, “A coupled duration-focused architecture for realtime music to score alignment,” *IEEE Trans. PAMI*, **32(6)**, pp. 974–987, 2010.
[7] A. Arzt, G. Widmer and S. Dixon, “Adaptive distance normalization for real-time music tracking,” *Proc. EUSIPCO*, pp. 2689–2693, 2012.
[8] E. Nakamura, T. Nakamura, Y. Saito, N. Ono and S. Sagayama, “Outer-product hidden Markov model and polyphonic MIDI score following,” *JNMR*, **43(2)**, pp. 183–201, 2014.
[9] P. Cuvillier and A. Cont, “Coherent time modeling of semi-Markov models with application to real-time audio-to-score alignment,” *Proc. IEEE MLSP*, 6 pages, 2014.
[10] E. Nakamura, N. Ono, S. Sagayama and K. Watanabe, “A stochastic temporal model of polyphonic MIDI performance with ornaments,” to appear in *JNMR*.
[11] N. Orio and F. Déchelle, “Score following using spectral analysis and hidden Markov models,” *Proc. ICMC*, pp. 1708–1710, 2001.
[12] S.-Z. Yu, “Hidden semi-Markov models,” *Artificial Intelligence*, **174**, pp. 215–243, 2010.
[13] J. Bilmes, “Graphical models and automatic speech recognition,” in *Mathematical foundations of speech and language processing* (Springer New York), pp. 191–245, 2004.
[14] 中村 栄太, 山本 龍一, 酒向 慎司, 齋藤 康之, 嵯峨山 茂樹, “多声 MIDI 演奏の楽譜追跡における演奏の不確定性のモデル化と自動伴奏への応用,” 情報処理学会研究報告, MUS96, No. 14, pp. 1–6, 2012.
[15] M. Russel and A. Cook, “Experimental evaluation of duration modelling techniques for automatic speech recognition,” *Proc. ICASSP*, pp. 2376–2379, 1987.
[16] R. Dannenberg and H. Mukaino, “New techniques for enhanced quality of computer accompaniment,” *Proc. ICMC*, pp. 243–249, 1988.
[17] Y. Guédon, “Hidden Hybrid Markov/Semi-Markov Chains,” *Computational Statistics and Data Analysis*, **49**, pp. 663–688, 2005.
[18] E. Nakamura, Y. Saito, N. Ono and S. Sagayama, “Merged-output hidden Markov model for score following of MIDI performance with ornaments, desynchronized voices, repeats and skips,” *Proc. Joint ICMC|SMC 2014*, pp.1185-1192, 2014.