

出力合流並列隠れマルコフモデルと その多声鍵盤音楽の楽譜追跡・両手部分離への応用

中村 栄太¹ 齋藤 康之² 嵯峨山 茂樹¹

概要：鍵盤楽器における両手の演奏は、リズムが左右非同期である場合や独立性を持つ声部が合わさった楽譜の場合などに、互いに独立性を持つ。また、ピアノの運指は多くの場合、両手独立に考えられる。これらの場合、楽譜追跡や自動採譜、ピアノ運指決定等を目的とした演奏モデルは、並列した両手の演奏が依存しながら合わさったものとして記述するのが有効である。この際、MIDI 信号としての演奏は同一チャンネルを通して観測されるため、信号分離を含む逆推定が効率的に行えるモデルが必要である。本稿では、この様な鍵盤演奏を記述するモデルとして、複数の並列した隠れマルコフモデルの出力が同一チャンネルに合流して観測される、出力合流並列隠れマルコフモデルを提案し、その推定アルゴリズムについて議論する。また、このモデルを鍵盤演奏の多声部楽譜追跡アルゴリズムおよび鍵盤楽譜からの両手部分離アルゴリズムに応用し、その有効性を評価する。

1. 序論

ピアノ等の鍵盤楽器の両手による演奏は、互いに独立性を持っている。例えば、リズムが左右独立である楽譜や両手ともに速い楽譜の演奏時などには、両手間の非同期性が大きくなる傾向がある。また、鍵盤楽譜はしばしば2つ以上の独立性を持つ声部が合わさっており、演奏にもその独立性が反映される場合がある。鍵盤演奏における両手の運指は、大部分においてそれぞれ独立に決定される。

楽譜追跡やピアノ運指決定、自動採譜（リズム解析）などは演奏音からその生成過程を求める逆問題として捉えられるため、その生成過程を理解することが鍵となり、演奏のモデルを用いた手法が多く研究されている [1-4]。これらの問題を扱う際、上述の両手間（声部間）の演奏の独立性は精度低下の原因となり得り、性能向上のためには演奏モデルにこれを組み入れることが重要である。

本稿では、この様な演奏を記述するモデルとして、互いに並列に状態遷移を行う複数の隠れマルコフモデル (hidden Markov model, HMM) からの出力が同一チャンネルに合流した結果の信号を記述する、出力合流並列隠れマルコフモデル（以下、合流 HMM）を提案し、その推定アルゴリズムについて議論する。また、このモデルを鍵盤演奏の多声部楽譜追跡アルゴリズムおよび鍵盤楽譜からの両手部分離アルゴリズムに応用し、その有効性を評価する。

2. 出力合流並列隠れマルコフモデル

2.1 両手の独立性を取り入れた鍵盤演奏のモデル

楽譜追跡やピアノ運指決定、自動採譜における演奏は、隠れ状態を持つ確率モデルとして記述できる。その具体的な演奏モデルとして、効率的な逆推定が行える HMM [5] が多く用いられている。

両手の独立性を組み込んだモデルを作る際の方法として、状態遷移の変更や高次 Markov の効果を入れる方法が考えられる。しかし、必要以上にモデルのエントロピーを上げるとは認識の誤り率の増大を引き起こし、また、高次のモデルは逆推定の際の空間・時間計算量の増大を引き起こす。従って、これらの方法は、モデルの拡張が難しい、あるいは効率的でないことから、しばしば実際的でない。もう一つの方法として、まず両手を記述する独立な HMM を考え、それらを合わせる事により互いに依存性を持つモデルを作ることが考えられる。この方法は、両手の独立性を効率的に組み込むことができるため、特に独立性が強い状況において、先の方法に比べて適切なモデル化であることが期待される。以下、この方針に従ったモデルを考える。

鍵盤演奏を MIDI 信号などを通して観測する場合、両手の演奏音は同一チャンネルに合流して観測される。これは、各出力記号に対して片方の HMM のみが遷移および出力を行うというモデルにより記述される。複数の並列した HMM が出力に寄与するモデルとして Factorial HMM [6] があるが、全ての HMM が同時に遷移するモデルのため、この様な信号の記述には適切でない。また、上の様なモデルの逆推定は信号分離問題を含むが、実用上、効率的な推定が可能である必要がある。

¹ 東京大学 大学院情報理工学系研究科
The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo
113-0033, Japan

² 木更津工業高等専門学校 情報工学科
Kisarazu National College of Technology, 2-11-1 Kiyomidai-
Higashi, Kisarazu, Chiba 292-0041, Japan

2.2 モデルの記述

まず、2つの独立な HMM の出力が合流して観測されるモデルを述べる。これは、まずどちらの HMM が遷移を行うかが選ばれ、選ばれたほうの HMM のみがもう片方の HMM の影響を受けずに遷移・出力を行うモデルとして記述できる。HMM を表す添え字を $\eta = 1, 2$ とすると、全体のモデルの状態は、 η および両方の HMM の状態（それぞれの添え字を i と j とする）の対で記述されるため、 $k = (\eta, i, j)$ の添え字を持つ。片方の HMM が選ばれる確率を α_η ($\alpha_1 + \alpha_2 = 1$) とすると、全体のモデルは次の状態遷移確率と出力確率を持つ HMM として記述される：

$$a_{kk'} = P(k'|k) = \begin{cases} \alpha_1 a_{ii'}^{(1)} \delta_{jj'}, & \eta' = 1; \\ \alpha_2 a_{jj'}^{(2)} \delta_{ii'}, & \eta' = 2, \end{cases} \quad (1)$$

$$b_{kk'}(o) = P(o|k, k') = \begin{cases} b_{ii'}^{(1)}(o) \delta_{jj'}, & \eta' = 1; \\ b_{jj'}^{(2)}(o) \delta_{ii'}, & \eta' = 2. \end{cases} \quad (2)$$

ここで、 $a^{(\eta)}$ と $b^{(\eta)}$ はそれぞれの HMM の状態遷移確率と出力確率であり、 o は出力記号を表している。式中、 δ は Kronecker の記号を表している（以下同様）。以下、各 HMM のことをトラック、 η のことをトラック添え字、また、全体の HMM のことを合流 HMM と呼ぶ*1。

実際の演奏では両手が完全に独立ではなく、互いに依存していることが多い。この依存性は、トラックの選択、状態遷移、出力において考えられるが、それぞれ次の式中の相互作用を表す項 $\alpha_\eta(k)$, $\phi_{kk'}^{(\eta)}$, $\psi_{kk'}^{(\eta)}$ により表現できる：

$$a_{kk'} = \begin{cases} \alpha_1(k) a_{ii'}^{(1)} \delta_{jj'} \phi_{kk'}^{(1)}, & \eta' = 1; \\ \alpha_2(k) a_{jj'}^{(2)} \delta_{ii'} \phi_{kk'}^{(2)}, & \eta' = 2, \end{cases} \quad (3)$$

$$b_{kk'}(o) = \begin{cases} b_{ii'}^{(1)}(o) \delta_{jj'} \psi_{kk'}^{(1)}(o), & \eta' = 1; \\ b_{jj'}^{(2)}(o) \delta_{ii'} \psi_{kk'}^{(2)}(o), & \eta' = 2. \end{cases} \quad (4)$$

ここで、 $\alpha_1(k) + \alpha_2(k) = 1$ であり、 $a_{kk'}$ と $b_{kk'}(o)$ は適切な正規化条件を満たすものとする。

2.3 推定アルゴリズムと計算量

前節に述べた通り、出力が合流する複数の HMM の並列モデルは HMM として記述されるため、推定アルゴリズムには Viterbi アルゴリズムや前向きアルゴリズムを用いることができるが、以下、Viterbi アルゴリズムを具体例として議論する。前節の合流 HMM の状態数は $N = 2IJ$ である (I と J はそれぞれの HMM の状態数)。状態数 N の通常の全結合 HMM の場合、出力記号毎の確率の更新には $O(N^2)$ の計算量が必要であり、今の場合これは $O(4I^2J^2)$ である。合流 HMM の場合、状態遷移確率行列が式 (3) の様な特殊な形をしているため、確率更新の計算量は $O(4IJ(I+J))$ に抑えられる。

3.2 節で述べる様に、演奏モデルでは、発音間時間 (inter-onset interval, IOI) の出力の様に出力分布が一つ前の出力記号に依存している場合がある。この様な HMM を前出力

*1 合流 HMM の状態の添え字 k の成分 η は、トラックが自己遷移する場合にどちらのトラックかを区別するのに必要である。

依存型と呼ぶことにする*2。複数の全出力依存型 HMM から成る合流 HMM の場合には、各トラックに関して最後の出力記号を記憶しておく必要がある。このため、この合流 HMM の最尤状態列は、Viterbi アルゴリズムにより全探索することができない。

実際の最尤推定では、遠く遡った過去の状態に関する情報に関する依存性は小さいことが多い。このことから、各時点で各状態に関してその数個前までの最尤状態の出力記号を正解として用いることにより、前出力依存型の場合でも近似探索を効率的に行える。

2.4 モデルの拡張

2.2 節と同様に、 $\eta = 1, \dots, N_{tr}$ で表される複数の HMM があり、その出力が合流して観測される場合も HMM により記述できる。全体のモデルは添え字 $k = (\eta, i_1, \dots, i_{N_{tr}})$ (ただし、 i_η は η 番目の HMM の状態の添え字) で表される状態を持ち、その状態遷移確率と出力確率は、各 HMM が独立な場合、

$$a_{kk'} = \alpha_{\eta'} a_{i_\eta i'_\eta}^{(\eta')} \prod_{\rho \neq \eta'} \delta_{i_\rho i'_\rho}, \quad (5)$$

$$b_{kk'}(o) = b_{i_\eta i'_\eta}^{(\eta')}(o) \prod_{\rho \neq \eta'} \delta_{i_\rho i'_\rho}. \quad (6)$$

である。ただし、 α_η は η 番目の HMM の遷移が選ばれる確率を表している ($\sum_\eta \alpha_\eta = 1$)。2.2 節と同様の方法により、相互作用項を導入できる。この場合の確率更新の計算量は $O(N_{tr}^2 I_1 \dots I_{N_{tr}} (I_1 + \dots + I_{N_{tr}}))$ となる。

合流 HMM のその他の拡張としては、2.2 節のモデルにおいてそれぞれ 0 次と 1 次の Markov モデルであった η と i_η の確率モデルとして高次 Markov モデルを用いることが考えられる。

3. 鍵盤演奏の多声楽譜追跡

3.1 HMM による演奏モデルを用いた楽譜追跡

楽譜追跡は、演奏音と楽譜中の音符とのマッチング問題であるが [7,8]、これは演奏音をもとに奏者の意図した楽譜位置を推定するという逆問題と捉えられる。この逆問題を解く際には、演奏音の生成過程を知ることが鍵となるが、演奏にはテンポや発音時刻の揺らぎ、演奏誤り、弾き直しや弾き飛ばし（併せて以下、弾き飛ばし）、裝飾音など多くの不確実性があるため、これらの不確実性を記述する演奏モデルを作ることが重要である。演奏過程は、奏者による楽譜位置の意図と実際の発音の 2 段階からなる過程として捉えられるが、両過程ともに上記不確実性を含むため、それぞれを確率的に記述することが考えられる。

HMM は 2 段階の確率過程を記述するモデルであり、効率的な逆推定が可能であるため [5]、演奏過程を記述するモデルとして用いられてきた [1,9]。本章で議論するピアノなどの多声鍵盤演奏の楽譜追跡 [10] においても、MIDI 信号に基づく HMM が提案されている [11-14]。

*2 例えば、状態列 $\{i_m\}$ 、出力記号列 $\{o_m\}$ の場合の前出力依存型 HMM の出力確率は、 $b_{i_{m-1}i_m}(o_{m-1}, o_m)$ の様に書ける。この様なモデルは、状態が $x_m = (i_m, o_m)$ で表されるような HMM を考えることと等価である (3.2 節参照)。

3.2 単一の HMM による演奏モデル

以下、文献 [12–14] に基づき単一 HMM による演奏モデルについて簡潔に記す。楽譜上での音符の位置は通常、音価を基に記述されるが、この方法による楽譜位置（の座標）のことを以下、楽譜時刻と呼び、変数 τ で表す。奏者による楽譜位置の意図は、各時刻 t における楽譜時刻の対応づけであるので、関数 $\tau(t)$ として表現される。演奏は演奏音として観測されるが、鍵盤演奏の MIDI 信号の場合、これは発音時刻と音高の情報として観測される*3。対応する楽譜位置は離散個のペア $\{(\tau_m, t_m)\}$, $m = 1, 2, \dots$ により表される。ここで、各 τ_m は楽譜上の音符の発音楽譜時刻であるが、各発音楽譜時刻に対して HMM の一つの状態を対応させると、楽譜位置の意図は $\{(i_m, t_m)\}$, $m = 1, 2, \dots$ と表される。ただし、 i は HMM の状態を示す添え字とする。

多声鍵盤楽譜の場合、楽譜には和音や複雑旋律が含まれるが、単一 HMM による演奏モデルでは、同楽譜時刻のものはまとめて一つの状態として表される。またトリルやアルペジオなどの装飾音は発音時刻や音数に不確実性があるが、これらも一つの状態として表現できる [14]。観測時刻における意図情報 (i_m, t_m) は、それ以前の演奏に依存していると考えられるが、この依存性が Markov 的であると仮定することにより、近似的な Markov モデルが得られる。演奏は時間並進に対して不変であると仮定すると、状態遷移確率は

$$P(i_m, t_m | i_{m-1}, t_{m-1}) = a_{i_{m-1}i_m}(t_m - t_{m-1}) \quad (7)$$

と表される。

演奏音の発音は、演奏意図の結果生じると考えられるが、この過程も確率的であり、直前の意図までに依存しているとする、出力確率は

$$P(p_m | i_{m-1}, t_{m-1}; i_m, t_m) = b_{i_{m-1}i_m}(p_m; t_m - t_{m-1}) \quad (8)$$

により表される。ただし、 p_m は m 番目の発音の音高を表す。演奏は意図と発音の結果として表現されるので、確率は

$$P(\{p_m, i_m, t_m\}_m) = \prod_m c_{i_{m-1}i_m}(p_m; t_m - t_{m-1}) \quad (9)$$

$$= \prod_m a_{i_{m-1}i_m}(t_m - t_{m-1}) b_{i_{m-1}i_m}(p_m; t_m - t_{m-1}) \quad (10)$$

$$= \prod_m a_{i_{m-1}i_m} \tilde{b}_{i_{m-1}i_m}(p_m, t_m - t_{m-1}) \quad (11)$$

と表される。ここで、

$$a_{i_{m-1}i_m} = \int dt_m a_{i_{m-1}i_m}(t_m - t_{m-1}), \quad (12)$$

$$\tilde{b}_{i_{m-1}i_m}(p_m, t_m - t_{m-1}) = \frac{c_{i_{m-1}i_m}(p_m; t_m - t_{m-1})}{a_{i_{m-1}i_m}} \quad (13)$$

とした。発音間時間 (IOI, inter-onset interval) $\Delta t_m = t_m - t_{m-1}$ は観測量であるので、式 (11) により、このモデ

*3 MIDI 信号としては他に発音終了時刻と鍵盤速度の情報が得られるが、これらは楽譜との関連がより小さく不確実性が大きいいため、本稿では観測の対象から外すことにする。

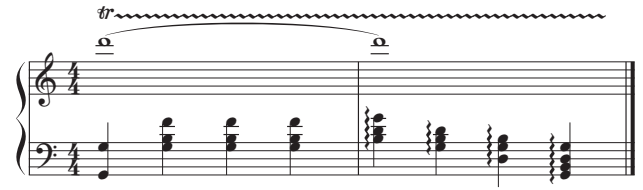


図 1 単一 HMM による演奏モデルによる表現が難しい楽譜の例。

ルは楽譜位置 i を状態とし、音高と IOI を出力する HMM とも解釈できる。これらの確率分布は、経験または実際の演奏データをもとに与えることができる。

楽譜追跡は、観測された時刻と音高が与えられた時の演奏意図の系列 $\{i_m\}$ に関する最尤推定により行え、これには、(前方) Viterbi アルゴリズムを用いることができる。

2.1 節で述べたように両手の独立性が大きい演奏の場合には、IOI や音の順序が楽譜による理想の演奏と異なる。これらは、演奏誤りとしてモデルに組み込むことができるが、一般的には、演奏誤りの確率が高いほど、認識の誤り率は大きくなる。例えば、図 1 に示した様なトリルと和音の楽譜は、HMM において和音とトリルの交互列として表現可能であるが、ここでの IOI はほとんど重要な情報となり得ず、楽譜位置推定を誤る可能性が高くなる。特に、左手の和音の脱落がある場合にそれを認識することは困難である。また、多声の複雑な楽譜においてどの様に状態を構成すればいいかを楽譜から決定するかが明らかでない場合も少なくない。さらに、このように誤りや IOI の大きな揺らぎを含んだ楽譜位置推定の結果からのテンポ推定の誤りも大きくなる。

3.3 合流 HMM による演奏モデルを用いた多声楽譜追跡

前節で述べた単一の HMM による演奏のモデルにおいて問題となる両手の独立性は、合流 HMM を用いることにより有効的に記述できると期待される。例えば、図 1 の楽譜の演奏は、右手 HMM は一つのトリル状態、左手 HMM は複数の和音状態を持つ合流 HMM により記述される。この場合、両手 HMM の遷移はそれぞれの手のみによる発音間での IOI に依存するため、左手の和音の脱落も認識することができる。また、モデルの記述も楽譜の記法に沿った単純なものになる。

2.3 節で述べた通り、合流 HMM を用いた演奏モデルの状態数は両手の HMM の状態数の積程度になるため、音符数の多い楽曲の場合、空間計算量が大きくなる。また、弾き飛ばしを含む演奏を記述するためには、楽譜上で遠く離れた音符間を状態遷移で結ぶ必要があるため、時間計算量も大きくなる。実時間での動作が必要な楽譜追跡の場合、これらの計算量を大きく削減する必要があるが、これは以下のモデルの単純化により可能である。まず、弾き飛ばしの確率は一定と近似すると計算時間を大きく削減でき、この近似による追従性能の低下は小さいことが知られている [12, 13]。次に、演奏中に両手の楽譜位置が大きく離れないと考えられるため、状態 (η, i, j) のうち楽譜上で大き

表 1 単一 HMM および合流 HMM を用いた多声楽譜追跡の結果. 数値は人手によりラベル付けした結果を正解とした時の誤り率を示す. 楽曲 2 ではトリル音以外の音符数を示している.

楽曲	演奏音符数	単一 HMM	合流 HMM
1	1068	0.219	0.127
2	1226	0.233	0.113

く離れた i と j を持つものへの遷移は無視できる*4. この際、楽譜上で大きく離れた状態同士を一定の確率遷移で結ぶことにより、弾き飛ばしをモデル化できる*5. 以上の単純化のもと、空間計算量と時間計算量はともに $O(I+J)$ 程度となる.

多声楽譜追跡の両手部に用いる楽譜は、通常実際に両手で弾かれる楽譜をそれぞれ用いることが想定されるが、これ以外の楽譜の分割をすることも可能である. 例えば、右手 HMM には両手の楽譜を、左手 HMM には空の楽譜を入力とすると、多声楽譜追跡は前節に述べた単一 HMM による楽譜追跡と一致する. 楽譜の一部のみ両手の楽譜を統合して用いることもできる. また、上と同様の方法により 3 つ以上の声部を入力として多声楽譜追跡も行える.

合流 HMM の逆推定結果として各音がどちらの HMM からの出力であるかが得られるので、この結果より各時点での局所的なテンポが求まる. この局所テンポに対して移動平均または Kalman フィルタを用いることにより、演奏テンポの推定を行える.

3.4 多声楽譜追跡の評価

合流 HMM を用いた多声楽譜追跡の有効性を確認するため、実際の人間による演奏を用いた評価を行った. 使用した楽曲は、両手でリズムが 4 対 3 と異なる楽譜である Chopin の幻想即興曲の allegro 部 (楽曲 1) および、図 1 の様に保続したトリルと同時に演奏される和音やアルペジオを多く含む練習曲 (楽曲 2) である. 演奏は、計 3 人の奏者による練習時の演奏を MIDI 録音したものをを用いた.

表 1 に楽譜追跡の結果を示す. 表中、誤り率は人手によりラベル付した結果を正解として計算した. 楽曲 2 の演奏では多くのトリル音があったが、トリル音の楽譜位置は不確定であるため、トリル以外の和音やアルペジオの楽譜位置をもとに評価を行った. 楽曲 1 と 2 両方において、合流 HMM を用いた時の誤り率は、単一 HMM を用いた場合に比べ約半分程度に収まっている. 特に、楽曲 1 では両手の演奏音の順序が大きく入れ替わった場合に、また、楽曲 2 ではトリルと同時に同一和音の連打やアルペジオがある場合に合流 HMM を用いた方が誤りが少なかった. 一方、演奏中の弾き飛ばしに関する追従は単一 HMM の方が速い傾向が見られたが、これは両手の同時関係の情報をより多く用いているためであると考えられる.

*4 これは、式 (3) で $|i-j|$ が大きい時に $\phi_{kk'}^{(\eta)} = 0$ とすることに対応している.

*5 厳密には、これは式 (1) で定義される合流 HMM では現れない遷移である. これを含むモデルは合流 HMM の拡張と見なせるが、ここではこのモデルも合流 HMM と呼ぶことにする.

3.5 実時間動作に関する議論

前節の評価に用いた音符数が約 1000 程度以下の楽譜の場合、合流 HMM を用いた多声楽譜追跡アルゴリズムは普及型パソコンの計算性能程度でも、実時間で動作することを確認した. 数千音符以上の楽曲の場合、同様の計算性能下での実時間動作は難しいと考えられ、この点において約 10000 音符まで実動作可能な単一 HMM によるアルゴリズム [12,13] に比べ劣っている. これは、3.3 節で述べた状態数削減を行った後でも、合流 HMM の状態数は単一 HMM の状態数に比べて大きくなるからである. しかし、3.3 節で述べた様に、楽譜中で単一 HMM でも適切に楽譜追跡ができる箇所については両手の楽譜を統合することにより状態数増加を避けることが可能である.

4. 鍵盤楽譜の両手部分離

4.1 楽譜追跡およびその他の分野における両手部分離問題

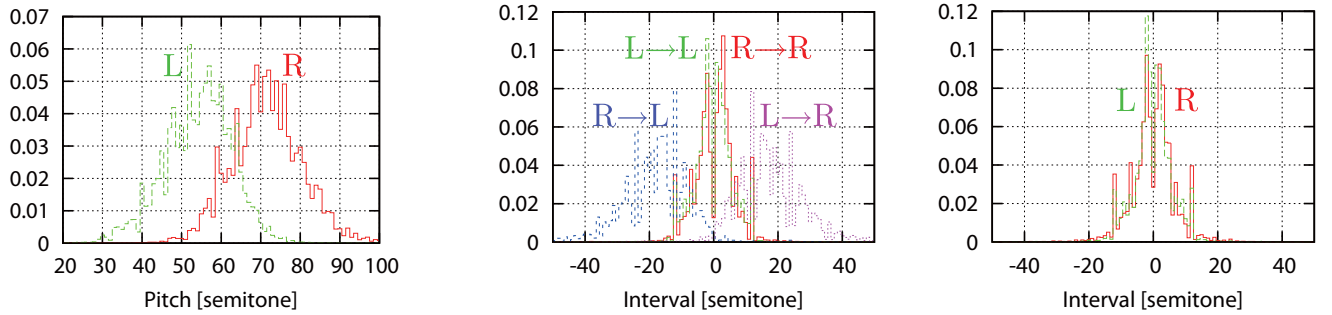
3 章で議論した方法により多声鍵盤演奏の楽譜追跡を行うためには、予め両手パートが分離している楽譜情報が必要である. ピアノ譜 (大譜表) に対応する楽譜情報が手に入る場合は、それを直接用いることができるが、楽譜が 1トラックにまとめられた MIDI ファイルで与えられる場合は、それを両手パートに分離するという処理が必要となる. 本章では、両手パートがまとめられた楽譜から両手パートに分離された楽譜を得る問題 (以下、両手部分離問題という) を扱う.

両手部分離は、楽譜追跡のみならず他の分野に関しても応用可能である. 1 章で述べた様に、多声鍵盤演奏の自動採譜 (リズム解析) をする際にも両手の非同期性・独立性が問題となり得る. また、ピアノ運指の自動決定においても、両手の運指を決定することは従来の方法では難しい [2,15]. これらの際、両手パートが分離された演奏や楽譜が与えられれば問題を軽減することができるため、両手部分離は、これらの前処理としても用いることができる.

鍵盤楽譜中のある音が両手のどちらで弾かれるかは、その音高や周辺の音との間の音程、音価、運指や演奏表情付け、音楽的な voice leading や演奏効果等による影響を受けて決定されるであろう. 通常の演奏では、両手はある程度の距離をおいて演奏されるため、音高や周辺音との音程のみによってもある程度高い精度での両手部分離ができると期待される. 本章では、これらの情報による両手部分離の方法を考える. 以下で議論する様に、これらの情報による両手部分離は、演奏者の身体における物理的な制約が大きく影響している. よって、両手部分が与えられている実際の鍵盤楽譜に対して、ここで議論する両手部分離の結果を比較することは、こうした単純な情報では説明ができないより音楽的な制約や表現が大きく影響している両手部分離の箇所を検出する手段として用いることができ、鍵盤楽曲の音楽学的な研究をする上で有効な手段の一つとなり得る. これが本章で両手部分離を議論するもう一つの目的である.

4.2 鍵盤分割および単純な HMM による方法

両手部分離の方法として最も単純なものと考えられるのは、各音符の音高のみにより行う方法である. 図 2(a) は、



(a) 両手部の音高分布. 横軸は MIDI ノート番号 (60 = C4; 一点ハ音) を表わし, 緑と赤のヒストグラムは, それぞれ左手と右手により弾かれる音高の分布を示している.
 (b) 両手の bigram に対する音程分布. 横軸は半音単位での音程を表わしている. 緑, 赤, 紫, 青のヒストグラムは, それぞれ左 → 左, 右 → 右, 左 → 右, 右 → 左における音程の分布を示している.
 (c) 片手の bigram における音程分布. 横軸は半音単位での音程を表わし, 緑と赤のヒストグラムは, それぞれ左手と右手に関する音程の分布を示している.

図 2 ピアノ曲における両手部の音高分布および音程分布.

ピアノ曲における両手部の音高分布を示している. 用いた楽曲は, 4.4 節で評価に用いるものと同じである (以下同様). この分布をもとに, 各音高が両手どちらによって弾かれやすいかを推定できるが, これは両手を表す隠れ変数 $\eta = L, R$ を持った 0 次の HMM を用いた Viterbi 推定と見なせる. 一般には, 両手の使われる頻度の分布を状態出現確率として用いられ, 出力確率は図 2(a) に示した分布から得られる. 図より, この方法は閾値 $p_{th} \simeq 62.5$ により両手部分離をすること (鍵盤分割による方法) とほぼ等しいことが分かる. この 0 次 HMM によるモデル固有の誤り率^{*6}は, 約 16% である.

図 2(a) が示す様に, 両手部の音高分布はかなり重なり合っているため, 各音の音高情報のみによる両手部分離には限界があり, 誤り率をより下げるためには, 周辺の音との間の音程を用いることが必要である. 連続した 2 音間の音程を用いる方法として, 次数を増やした HMM を用いることが考えられる. 例えば, 1 次の HMM は, 状態遷移確率を $a_{\eta\eta'}$, 出力確率を $b_{\eta\eta'}(\delta p)$ (δp は連続した 2 音間の音程) として記述される. このモデルの出力分布に対応するピアノ曲の音程の分布を図 2(b) に示す. 同様に, 状態遷移確率も学習可能である. 1 次 HMM におけるモデル固有の誤り率は約 10% である.

同様の方法は, $n(\geq 2)$ 次の HMM を用いても作ることが可能である. しかし, 演奏における両手の順序に関する物理的制約は緩いため, 高次の状態遷移確率の偏りは少ないと考えられる. また, 出力分布に関しても同様に, δp に関する分布の高次において現れる偏りは少ないと予想され, 音高に関する高次の効果を取り入れない限り, 単純に高次の HMM を考えても誤り率の減少は少ないと期待される.

図 2(b) において, b_{LR} と b_{RL} の分布の分散は, 図 2(a) の分布の実現により得られるものよりも有意に小さく, これら分布では Markov 連鎖の効果が大きいことを示している. また, これらの分布と, 図 2(c) に示した片手の bigram における音程分布との差異は小さいことが確かめられる.

これより, 片手内での連続音間の音程の分布では, Markov 連鎖の効果が高次まで大きいことが分かる. 一方, 図 2(b) における b_{LR} と b_{RL} の分布は, 両手の交差に対応する領域を除いて概形が, 図 2(a) の分布の実現により得られるものに近く, 両手にまたがる連続音間の音程の分布では Markov 連鎖の効果が比較的小さいことを示している. この理由として, 片手内での音程分布は手の大きさや指および腕の動きによる制約を受けやすい一方, 両手にまたがる音程分布は, 両手の交差を除いて, その様な制約を受けにくいことが考えられる.

4.3 合流 HMM を用いた方法

前節の議論より, 両手の鍵盤演奏による音高列は合流 HMM により効率的かつ有効的に記述できると期待される. 具体的には, これは, 隠れ状態として $x = (\eta, p_L, p_R)$, $\eta = L, R$ を持つ合流 HMM として表される. ここで, p_L と p_R は, それぞれ左手 (右手) の弾く音高を表す. 観測される楽譜上の音高列を $\{y_i\}$ として表すと, 状態遷移確率と出力確率は,

$$a_{xx'} = \begin{cases} \alpha_L a_{p_L, p'_L}^L \delta_{p_R, p'_R}, & \eta' = L; \\ \alpha_R a_{p_R, p'_R}^R \delta_{p_L, p'_L}, & \eta' = R, \end{cases} \quad (14)$$

$$b_x(y) = \delta_{y, p_\eta} \quad (15)$$

で与えられる. a_{p_L, p'_L}^L および a_{p_R, p'_R}^R は図 2(c) の分布により得られる. 音域の制約として図 2(a) の音高分布を用いた場合, 合流 HMM におけるモデル固有の誤り率は約 7% である.

両手部分離は, この合流 HMM を用いて得られた最尤推定状態列 $\{\hat{x}_i\}$ のトラック添え字 $\{\hat{\eta}_i\}$ により推定できる. 上述の通り, この推定は観測される音高列に関する任意の高次の効果を含んでいるが, 実際上は十分大きい次数まで考慮すれば, 有限次数でも等価な推定が可能である. この次数を N_h とすると, この推定は, 各時点での変数 $\{(\eta_i, h_i)\}$, $h_i = 0, 1, \dots, N_h$ に対する次のポテンシャルの最小化の問題と等価であり, N_h が小さい時, 計算量的により効率的な推定が可能である:

*6 ここで, モデル固有の誤り率とは, モデル自体を用いて生成したデータに対する誤り率を意味しており, これはモデルパラメータにより決まる誤り率の目安を示している. 以下, この値は図 2 より学習したモデルパラメータの値に関するものを示す.

表 2 実際の楽譜に対する両手部分離の誤り率. 0-HMM, 1-HMM, 2-HMM は, 0 次, 1 次, 2 次の HMM を表している. 括弧内の数値は「10 度ルール」を導入した場合, 括弧内の数値は導入しない場合の誤り率を示している.

楽曲	音符数	0-HMM [%]	1-HMM [%]	2-HMM [%]	合流 HMM [%]
Bach (15 曲)	9638	5.1 (5.9)	5.3 (5.7)	6.1 (6.4)	1.9 (2.1)
Beethoven (2 曲)	18144	13.0 (19.4)	11.1 (14.6)	11.5 (14.8)	9.28 (11.1)
Chopin (5 曲)	8508	5.7 (16.3)	4.0 (7.3)	4.29 (8.1)	3.8 (10.7)
Debussy (3 曲)	3360	17.8 (24.5)	14.8 (17.8)	14.8 (18.2)	18.7 (24.0)
Total	39650	9.9 (15.9)	8.5 (11.2)	8.9 (11.6)	7.1 (9.9)

$$V(\eta, h) = - \sum_i \ln Q_{(\eta_{i-1}, h_{i-1}; \eta_i, h_i)} \quad (16)$$

$$Q_{(\eta_{i-1}, h_{i-1}; \eta_i, h_i)} = \begin{cases} \alpha_{\eta_i} a_{y_{i-1}, y_i} \delta_{h_i, h_{i-1}+1}, & \eta_i = \eta_{i-1}; \\ \alpha_{\eta_i} a_{y_{i-2-h_{i-1}}, y_i} \delta_{h_i, 0}, & \eta_i \neq \eta_{i-1}. \end{cases} \quad (17)$$

我々は, 次節で行う評価をする際に, N_h が 50 以上の場合に十分良い近似となっていることを確認した.

ここまで, 音高あるいは継時的な音程に関する分布の偏りに基づく両手部分離の方法を述べてきた. 一方で, 例えば 1 オクターブを大きく超えた音程を片手で弾くことは困難である, といった物理的な制約に基づく同時的な音程 (和音) に関する制約も, 両手部分離において重要であろう. 鍵盤演奏において両手交差は稀であるため, 同時に弾かれる多重音中の最低音 (最高音) から 1 オクターブを大きく超えて高い (低い) 音が左手 (右手) で弾かれる頻度分布は, 上の制約を強く反映していると予想される. 多重音の情報は, 五線譜および MIDI データから読み取れるため, この情報をもとに上の制約を確率的重みとして導入することが可能である.

4.4 実際の楽譜を用いた評価

上述した両手部分離アルゴリズムの性能を比較および評価するため, 実際の楽曲を用いて両手部分離を行った. 使用した楽曲は, Bach の 2 声の Inventions 全曲, Beethoven の Piano sonata Nos. 1-2, Chopin の Etudes Op. 10 Nos. 1-5, Debussy の Préludes 第 1 集 Nos. 1-3 である. 評価は, 4.2 節で述べた単純な (0 次から 2 次の) HMM を用いた方法および 4.3 節で述べた合流 HMM を用いた方法に対して, 両手部分離の誤り率を求めて行った (表 2). 多重音の制約に関しては, 現実の楽譜において片手で 10 度を超える多重音は稀であったので, 10 度を超える多重音の上下の音に関して確率的な重みを与えた (以下, 「10 度ルール」).

表 2 に「10 度ルール」を導入した場合とそうでない場合両方における各モデルを用いた両手部分離の結果を示す. 全体としては, 合流 HMM を用いた方法が最も誤り率が低く, 次いで 1 次の HMM と 2 次の HMM が低く, 0 次の HMM の誤り率が最も高かった. 2 次の HMM による方法は, 1 次のものに比べ全体的に誤り率が高いが, 大きな違いは見られなかった. また, 全ての曲において「10 度ルール」の導入により誤り率が低下した. 曲によりばらつきがあるものの, 提案したアルゴリズムにより両手部分離が比較的高い精度で行えることが確認された.

この結果により, 合流 HMM の有効性が確認された. また, 単純な HMM による方法では, 2 次の Markov モデルを用いても, 性能は大きく変わらないことが確認できた. Debussy の楽曲では, 全体的に誤り率が高かったが, これは演奏に大きな手の動きを含む複雑な運指が必要であることによる. 特に合流 HMM での誤り率が高いことは, 正解の両手部が手指の動き以外に音楽的な要因により決められた可能性を示唆しているが, 詳しい解析は今後の課題とする.

5. 結論

本稿では, 鍵盤演奏のモデルにおいて両手の独立性を効率的に組み入れられるモデルとして, 出力合流並列隠れマルコフモデルを提案し, その推定アルゴリズムについて議論した. また, 具体的な応用例として, 提案モデルを用いた鍵盤演奏の多声楽譜追跡アルゴリズムおよび両手部分離アルゴリズムを構築し, 性能を評価することにより有効性を確認した.

今後は, 各トラック間の依存性の効率的な学習やその推定の方法を議論したいと考えている. また, 他の具体例にも応用することを予定している.

謝辞 モデルに関して議論していただいた田中裕彬氏と齋藤大輔氏に感謝する. また, 楽譜追跡の評価のためピアノを演奏して下さった橘秀幸氏と中村友彦氏に感謝する.

参考文献

- [1] N. Orio *et al.*, in *NIME*, pp. 36-41, 2003.
- [2] Y. Yonebayashi *et al.*, *IJCAI*, pp. 2915-2921, 2007.
- [3] A. T. Cemgil *et al.*, *CMJ* **24**(2), pp.60-76, 2000.
- [4] 大槻知史他, 情報処理学会論文誌 **43**(2), pp. 245-255, 2002.
- [5] L. Rabiner, *Proc. IEEE*, **77** No. 2, pp. 257-286, 1989.
- [6] Z. Ghahramani and M. I. Jordan, *Machine Learning* **29**, pp. 245-273, 1997.
- [7] R. Dannenberg, *Proc. ICMC*, pp. 193-198, 1984.
- [8] B. Vercoe, *Proc. ICMC*, pp. 199-200, 1984.
- [9] P. Cano *et al.*, *Proc. ICMC*, pp. 441-444, 1999.
- [10] J. Bloch and R. Dannenberg, *Proc. ICMC*, pp. 279-289, 1985.
- [11] D. Schwarz *et al.*, *Proc. ICMC*, 2004.
- [12] 武田晴登他, 情報処理学会研究報告, MUS66, No. 18, pp. 109-116, 2006.
- [13] 中村栄太他, to appear in 情報処理学会論文誌, Apr. 2013.
- [14] 中村栄太他, 情報処理学会研究報告, MUS96, No. 14, pp. 1-6, 2012.
- [15] A. Al Kasimi *et al.*, *ISMIR*, pp. 355-356, 2007.