

多声音楽の楽譜追跡への Factorial HMM の適用に関する検討

五十嵐 巧輝† 山田 昌尚†

概要: 楽譜追跡とは、演奏者が譜面上のどの部分を演奏しているかをリアルタイムに推定するものである。応用例として演奏の進行に合わせて自動で楽譜をめくる自動譜めくりや、演奏に伴奏をつける自動伴奏等の技術があげられる。一般的な楽譜追跡では、一つのパートもしくは音楽全体の楽譜追跡を扱っている。しかし、多声音楽において個別のパートごとに楽譜追跡をすることは行われていない。そこで、多声音楽を対象として Factorial HMM を用い、個別の演奏パートごとに楽譜追跡を行う手法を検討する。

1. はじめに

楽譜追跡とは、演奏者が譜面上のどの部分を演奏しているかをリアルタイムに推定するものである。応用例として演奏の進行に合わせて自動で楽譜をめくる自動譜めくりや、演奏に伴奏をつける自動伴奏等の技術が上げられる [1]。一般的な楽譜追跡では、一つのパートもしくは音楽全体の楽譜追跡を扱っている。しかし、多声音楽において個別のパートごとに楽譜追跡をすることは行われていない。

そこで、本研究では多声音楽を対象として隠れマルコフモデルを用い、個別の演奏パートごとに楽譜追跡を行う手法を提案する。本稿では、音高情報を使った多声音楽の楽譜追跡の手法と、楽譜追跡を行った結果について述べる。

2. 楽譜追跡

隠れマルコフモデルとは、観測された観測系列から直接知ることのできない状態の遷移を推定することができるモデルである。隠れマルコフモデルは、状態系列、状態遷移確率、観測系列、出力確率の4つの要素からなる。本研究では、多声音楽の楽譜追跡を行うために、Factorial HMM (階層隠れマルコフモデル) [2] を用いた。これは、1つの観測系列から複数の状態系列を扱うモデルである。本稿では状態系列を楽器の演奏パート数だけ用意し、1チャンネルのMIDI データから複数の演奏パートの状態系列の遷移を推定する問題を扱う。状態系列数が2のモデルを図1に示す。

$z_n^{(k)}$ は状態系列 k の n 番目の要素、 x_n は n 番目の観測系列の要素、 $p(z_n^{(k)} | z_{n-1}^{(k)})$ は状態 $z_{n-1}^{(k)}$ から $z_n^{(k)}$ への遷移確率、 $p(x_n | z_n^{(k)})$ は状態 $z_n^{(k)}$ に対応する x_n の出力確率である。 $\mathbf{X} = \{x_0, x_1, \dots\}$ は観測系列であり、 x が観測されるたびに楽器のパートごとの状態 $\mathbf{Z}^{(k)} = \{z_0^{(k)}, z_1^{(k)}, \dots\}$ が遷移するモデルとなっている。これを楽譜追跡の問題に適用すると、 k は演奏パートの番号、 $z_n^{(k)}$ は楽譜上の音符、 x_n は観測された音高となる。

今回の観測には、MIDI の音高情報を用いた。式で表すと次式ようになる。

$$p(\mathbf{X}, \mathbf{Z}^{(k)}) = p(z_0^{(k)}) \left[\prod_{n=1}^N p(z_n^{(k)} | z_{n-1}^{(k)}) \right] \prod_{n=1}^N p(x_n | z_n^{(k)}) \quad (1)$$

$$p(x_n | z_n^{(k)}) = \sum_{i=1}^I e_{n_i}^{(k)} x_{n_i} \quad (2)$$

ここで、 I は MIDI で表現できる音高の総数、 $e_{n_i}^{(k)}$ は状態系列 k の n 番目の観測における音高 i の出力確率である。(1) 式は一般的な隠れマルコフモデルの式であり、状態遷移確率の計算部と出力確率の計算部に分かれている。(2) 式は隠れマルコフモデルにおける出力確率であり、音高情報を考慮した式で、本研究で提案するものである。 x_{n_i} は n 番目の観測で音高 i が観測されると 1、観測されていない音高は 0 とする。観測 x_{n_i} と音高 i における出力確率 $e_{n_i}^{(k)}$ の積をとり全音高について足し合わせることで出力確率を表現している。状態系列の推定は Viterbi アルゴリズムにより行った。

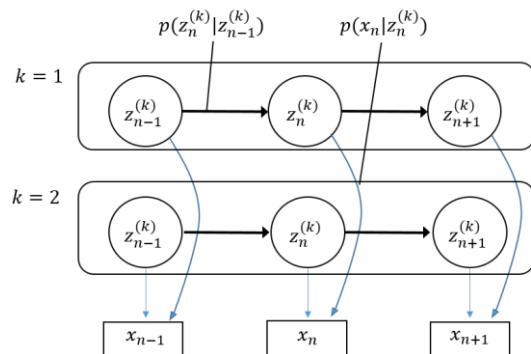


図1 Factorial HMM の例

3. 実験方法と結果

演奏者が楽譜上と異なった音高を演奏してしまう音高間違いが発生したことを想定した楽譜追跡の実験を行った。音高間違いによる追跡結果の変化のみを確認するために遷移確率は固定し、観測確率を変化させて実験を行った。初めに楽譜と遷移確率について示す。使用する曲はモーツァルトのピアノ・ソナタ K.545 の冒頭 2 小節を使用した。楽譜を図2に示す。状態系列は、右手のメロディ部の状態系

†National Institute of Technology, Kushiro College

列 1 と、左手の伴奏部の状態系列 2 を用意した。遷移確率は正しい状態への遷移確率を 0.8 とし、0.2 を他の全ての状態への遷移確率に振り分けた。音が発音されている途中で、もう一方の音に変化があった場合、自己遷移して状態を維持しなければならない。状態系列 1 の最初の音符のように自己遷移が 3 回必要な場合、自己遷移確率を 0.6、次の音符の状態への遷移確率を 0.2 とし実験を行った。



図 2 実験に使用した楽譜

次に演奏と出力確率について示す。MIDI ファイルから音高のみのデータを抽出したものを観測とした。まず、楽譜どおりに演奏した MIDI ファイルで楽譜追跡を動作させたところ、適切な状態遷移が得られた。次に音高を 1 か所だけ半音ずらしたデータを、全ての音符について作成した。正しい音高 i から半音ずれた音高 $i \pm 1$ の出力確率 $e_{n_{i \pm 1}}^{(k)}$ を「音高間違いの出力確率」と呼ぶことにする。弾き間違いのデータ数は音符数と同じ 23 個である。正しい音高の出力確率と半音ずらした音高の出力確率の合計を 0.9 とし、0.1 を他の全ての出力確率に振り分けた。楽譜上のある時点での音高が C3 のときの出力確率を図 3 に示す。図 3 は正しい音高の出力確率を 0.5 とし、半音ずれた音高の出力確率をそれぞれ 0.2 としたものである。

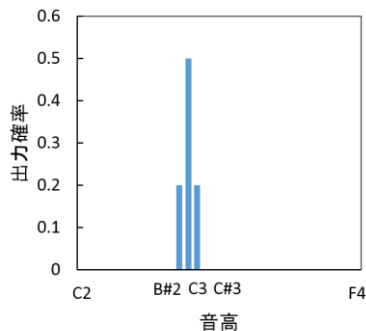


図 3 出力確率の分布

音高間違いの出力確率 $e_{n_{i \pm 1}}^{(k)}$ を 0 から 0.5 まで 0.05 刻みで変化させたときの、楽譜追跡成功数の変化を図 4 に示す。成功とは、音高間違いがない入力と同じく状態が遷移することである。

図 4 から音高間違いの出力確率が高いほど弾き間違いを追跡しやすくなることがわかる。音高間違いの出力確率が 0.3 付近で 23 個全ての状態遷移が成功した。図 3 から音高間違いの確率が 0.45 付近で正しい出力確率が 0 になるので、楽譜追跡の成功数が 0 になることが確認できる。また、状態遷移系列を確認したところ、左手の伴奏部は音高間違

いの出力確率が低い場合でも、適切な結果が得られた。右手のメロディ部は音高間違いの出力確率が 0 付近では成功しなかったが、確率が上がるにつれて成功数が増加した。

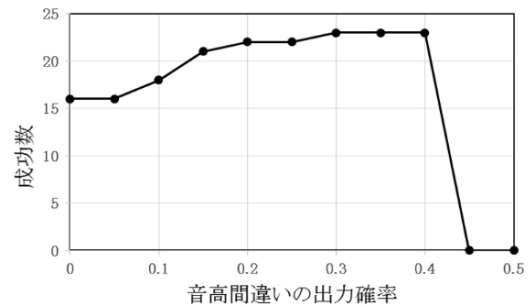


図 4 楽譜追跡の成功数

4. まとめ

多声音楽における楽譜追跡を行うモデルを提案し動作の確認を行った。今後は音高間違いや、弾き飛ばしに対応できる出力確率、状態遷移確率を設定していくことが必要である。

参考文献

- [1] 鈴木孝輔, 上田雄, 齋藤康之, 小野順貴, 嵯峨山茂樹, 「HMM を用いた音響演奏の楽譜追跡による弾き直しに追従可能な自動伴奏」, 情報処理学会研究報告, Vol.2011-MUS-89, No.29, pp. 1-6, 2011.
- [2] Z. Ghahramani, M. I. Jordan, "Factorial Hidden Markov Models", Machine Learning, Vol 29, Issue 2-3, pp.245-273, 1997.