

頷き動作による自動譜めくりシステムでの 合図とリズムノリの判別

地曳 はるか^{1,a)} 齋藤 康之^{1,b)} 中村 栄太² 嵯峨山 茂樹³

概要: 本報告では、自動譜めくりシステムにおいて、演奏者の譜めくり合図の頷きとリズムノリ動作を判別する方法について議論する。実際のピアノ演奏において、演奏者は譜めくり係に頷いて譜めくりの合図を送っていることから、我々は頷き動作をキューとした自動譜めくりを開発してきた。しかし、頷き動作は譜めくりの合図だけではなくリズムのノリにも現れる。そのためシステムは譜めくり合図の頷き動作のみを検出する必要がある。そこで、頷き動作とリズムノリ動作のサンプルを集めた結果、頷き動作よりもリズムノリ動作は浅く頷いていることが明らかとなった。このことから、2つの動作の各々における鼻の最下点の平均を求め、さらに両者の平均を閾値として設定し、それによって判別可能かどうかを実験により調べた。その結果、両者の判別は可能であったが、譜めくり合図の頷きがリズムノリとして誤認識されることがあった。これは閾値算出に用いた動作のサンプル数が足りず、妥当な閾値でないことが原因と考えられる。今後の課題として、閾値の個人適応も必要であろう。

キーワード: 頷きによる自動譜めくり, 楽譜, 隠れマルコフモデル, 頷きとリズムノリの判別

JIBIKI HARUKA^{1,a)} SAITO YASUYUKI^{1,b)} NAKAMURA EITA² SAGAYAMA SHIGEKI³

Abstract: This study discusses a method for discriminating players' page turning cues against head shaking motions for the purpose of developing an automatic page turning system for music scores. Nodding is commonly used as a page turning cue in actual piano playing and has been used in our proposed system. Since a similar motion is made when pianists shake head in time to the music, it is necessary to discriminate page turning cues from this head shaking motion. Through the experimental results, we noticed that displacement of player's nose position in head shaking motion is shallower than that in page turning cue. Thus we substituted an average between average of the deepest nose position in page turning cue and in head shaking motion into a threshold to distinguish two types of nodding, and confirmed this effectiveness. Although the system could perform page turning properly, sometimes page turning cues were recognized as head shaking motions incorrectly. As a future work, the threshold should be properly set by personalization process to avoid such error caused by personal differences of nodding deepness.

Keywords: Automatic page turning by nodding, musical scores, Hidden Markov model, Discrimination of nodding between page turning cue against head shaking motions

1. はじめに

楽器の演奏において楽譜をめくることを譜めくりという。譜めくりは基本的に演奏者本人が行うが、特にピアノ

の演奏では譜めくり係をつけることがある。この場合、事前に演奏者と譜めくり係で譜めくりのタイミングを練習しておかねばならず、譜めくり係の技量によっては楽譜をしっかりとめくれない、楽譜を落としてしまうなど演奏の妨げになってしまうことがある。そのため、人の手に代わり、機械やソフトウェアによってめくる自動譜めくりが開発されてきた。たとえば、図1はドイツのforma:Bene（フォルマ・ベーネ）社が開発した紙楽譜をめくる譜めくり機「フメクール」[1]である。フメクールにはフットスイッチが付属しており、それを踏むとアームが旋回し、その先端にあ

¹ 木更津工業高等専門学校
NIT, Kisarazu College, Chiba 292-0041, Japan
² 京都大学 大学院情報学研究科 知能情報学専攻
Kyoto Univ., Sakyo-ku, Kyoto 606-8501, Japan
³ 明治大学 総合数理学部 先端メディアサイエンス学科
Meiji Univ., Nakano-ku, Tokyo 164-8525, Japan
a) sdj15b08@j.kisarazu.ac.jp
b) saito@j.kisarazu.ac.jp



図 1 フメクール.



図 2 フェアリー.

る磁石が予めページ隅に貼っておいた鉄片付きシール（クリップ）を吸着して、そのページをめくる。ただし、安定な動作のためには外部電源が必要であることや、ピアノのペダルを両足を使って踏んでいる場合、フットスイッチを踏めずに楽譜をめくれないという欠点がある。また、図2の iPhone・iPad 向け電子楽譜サービス「フェアリー」[2]では、演奏音から演奏位置推定をし自動譜めくり機能を実装している。持ち運びが容易であるのは利点だが、専用の楽譜を1曲400円で購入しなければならないため、手軽に利用できるとは言い難い。

現在、発表会やコンクールなどの実際の演奏では、演奏者は譜めくり係に頼いて譜めくりの合図を送っている。そこで、我々は頷きのジェスチャ認識による自動譜めくりについてソフトウェアの開発を行ってきた [3]。本研究ではさらに、頷き動作とリズムノリ動作を判別し、譜めくりの合図を出したときのみ楽譜がめくられるような方法を検討する。

2. ソフトウェアの要求仕様

本研究で開発したソフトウェアの仕様は以下の通りである。

- 頭部の動きを web カメラで認識し、それを譜めくりのキューとして使用する。
- 楽譜データは PDF 形式とする。

本研究では手軽に利用できることを目的としたため、安価に入手できる USB カメラを用いることとした。また、すでに手元にある紙の楽譜を利用する場合、スキャナにより楽譜を取り込む必要がある。そのとき PDF 形式として保存することが多い。さらに、PDF 形式では1つのファイル内に複数ページを格納できるため、楽譜を保存する形式として最適である。以上より本研究では楽譜データは PDF 形式とした。

3. ソフトウェアに用いた手法

本研究で開発したソフトウェアは顔追跡部とジェスチャ

表 1 演奏中の動作の状態.

状態	説明
S_1	動かない
S_2	下を向く
S_3	動かない
S_4	上を向く

認識部に分けられる。本章ではそれぞれの部分で用いた手法について述べる。

3.1 顔追跡部

顔追跡部は二つの段階で構成されている。第1段階では、USB カメラによって取得した画像から Haar-like 特徴を用いて顔検出を行う [4]。そして、検出した顔画像から鼻の部分の画像を切り取りテンプレート画像とする。また、鼻の位置を取得しておく。この第1段階はソフトウェアが起動されたとき一度だけ実行される。第2段階では、再度 USB カメラによって取得した画像と、テンプレート画像でテンプレートマッチングを行う。そして、マッチングされた鼻の位置を取得し、前回位置と今回位置の差分をとることで顔の移動距離・移動方向を得ることができる。この二段構成によって、画像フレーム全体から毎回顔検出を行うよりも高速に顔追跡を行うことができる。

3.2 ジェスチャ認識部

3.2.1 隠れマルコフモデル

頷き動作の深さや速度などは、演奏している楽曲や演奏者によって異なる。そのため、加速度や変位量、頷き向きなどに閾値を設定することでは全ての演奏者の頷き動作を認識するのは不可能である。そこで、演奏中の動作を状態とし、確率モデルとして考える。ただし、観測できる情報は USB カメラから得られる画像のみであるため、頷き動作は直接観測できない隠れ状態である。そのため、隠れマルコフモデル (HMM, hidden Markov model) を用いる。

本研究では演奏中の動作を表1に示す4つの状態に分け、

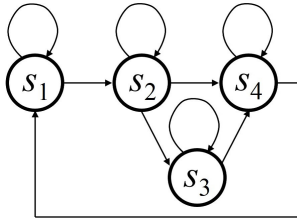


図 3 演奏中の動作のモデルの状態遷移図.

表 2 式中で使われている記号.

記号	説明
c	状態の個数
n	時系列の個数
$S_j^{(t)}$	時刻 t での状態 ($j = 1, 2, \dots, c$)
$o^{(t)}$	出力
r_i	状態 i が初期選択される確率 ($i = 1, 2, \dots, c$)
a_{ij}	状態 i から j への遷移確率
$b(s_j^{(t)}, o^{(t)})$	出力 $o^{(t)}$ が状態 $S_j^{(t)}$ から 観測される確率
$p_j^{(t)}$	時刻 t における状態 j の最大確率
$q_j^{(t)}$	最適状態系列の配列
$z^{(t)}$	最適状態系列の状態番号

隠れマルコフモデル化する. このモデルの状態遷移を図 3 に示す. たとえば, $S_1 S_2 S_3 S_4 S_1$ や $S_1 S_2 S_4 S_1$ のような状態系列の場合, 1 回領いたと考えられる. また, S_3 の状態が長く続く場合は, 領きの動作ではなく手元を見ていると考えられる.

3.2.2 Viterbi アルゴリズム

前項で述べたように, 領き動作は直接観測できず, 観測できる情報は USB カメラから得られる画像のみである. そこで, USB カメラから得られる画像より顔の移動距離・移動方向を算出し, それらを状態の出力として考え状態遷移の推定を行う. この状態遷移の推定には Viterbi アルゴリズムを用いる.

Viterbi アルゴリズムは時系列の状態遷移の最適経路を推定することができ, 以下の 4 つのステップにより構成されている [5]. 式 (1)~(6) で使われている記号を表 2 に示す.

ステップ 1: 初期化

$$p_i^{(1)} = r_i b(S_i^{(1)}, o^{(1)}) \quad (1)$$

$$q_i^{(1)} = 0 \quad (2)$$

$$(i = 1, 2, \dots, c)$$

ステップ 2: 再帰的導出

$$p_j^{(t)} = \max_i \{p_j^{(t-1)} a_{ij}\} b(S_j^{(t)}, o^{(t)}) \quad (3)$$

$$q_j^{(t)} = \operatorname{argmax}_i \{p_j^{(t-1)} a_{ij}\} \quad (4)$$

$$(t = 2, 3, \dots, n) (i, j = 1, 2, \dots, c)$$

ステップ 3: 終端の状態の決定

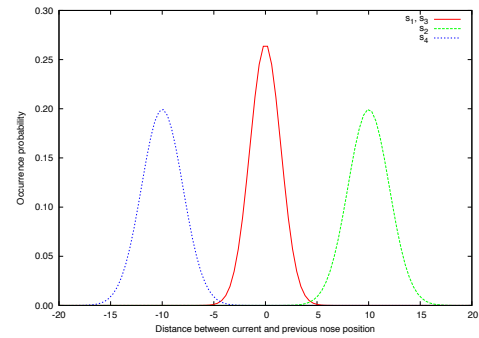


図 4 $b(s_j, d_p)$ の出力確率.

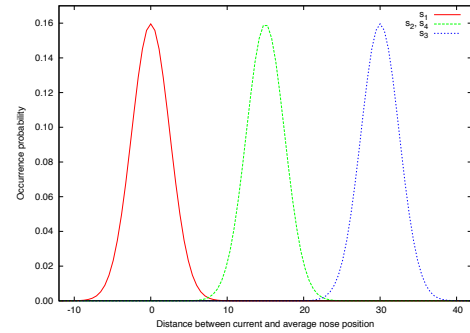


図 5 $b(s_j, d_a)$ の出力確率.

$$z^{(n)} = \operatorname{argmax}_i \{p_i^{(n)}\} \quad (5)$$

$$(i = 1, 2, \dots, c)$$

ステップ 4: 状態系列の確定

$$z^{(t)} = q_{z^{(t+1)}}^{(t+1)} \quad (6)$$

$$(t = n-1, \dots, 2, 1)$$

なお, 求められる状態系列は時刻に対して逆順であることに注意が必要である.

また, 本研究では $b(s_j^{(t)}, o^{(t)})$ は式 (7) で求めている. d_p は現在と前回の鼻の位置の距離であり, d_a は現在と平均の鼻の位置の距離である. $b(s_j, d_p)$ の値は図 4 に, $b(s_j, d_a)$ の値は図 5 に示す.

$$b(s_j^{(t)}, o^{(t)}) = b(s_j, d_p) b(s_j, d_a) \quad (7)$$

4. 実装

本研究ではソフトウェア開発に Java[6] を用いた. Java は汎用性に優れており, プラットフォームに依存しないアプリケーションソフトウェアの開発, および展開が容易である. そのため現在はノート PC 用のソフトウェアであるが, 今後タブレット端末など, モバイル端末向けへのソフトウェアとして展開していきたいと考えている. また, 顔検出には OpenCV[4] を用いている. そして, PDF 楽譜の表示には PDFRenderer[7] を用いた.

5. 領き動作とリズムノリ動作のサンプリング調査

ピアノの演奏中にテンポを感じて領き動作をすることが

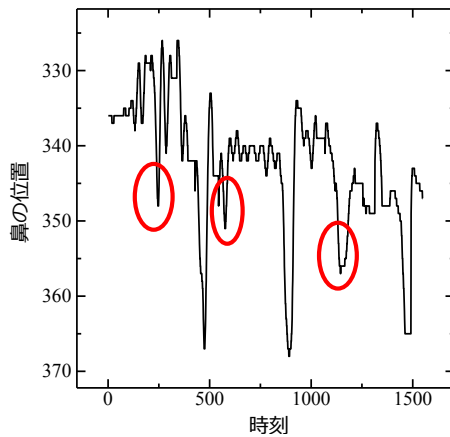


図 6 リズムノリ動作を含む鼻の位置.

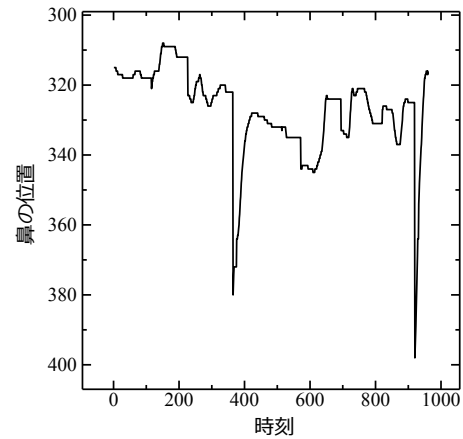


図 7 リズムノリ動作を含まない鼻の位置.

ある。これをリズムノリ動作と呼ぶ。本研究で開発したソフトウェアではこのリズムノリ動作を譜めくり合図の頷き動作と誤認識してしまう可能性がある。ただし、リズムノリ動作と譜めくり合図の頷き動作に異なる特徴があるならば判別が可能であると考えられる。そこで、開発したソフトウェアを用いて頭部の動きのサンプリング調査を行った。以下の条件でそれぞれ1回ずつ、計2回演奏してもらいそれぞれの頭部の動きのサンプルを集めた。

1. ソフトウェアの最低限の利用法を説明し自由に演奏してもらおう。
2. ページ送り以外ではなるべく動かずに演奏してもらおう。

自由に演奏してもらおうときにリズムノリ動作が含まれると考えられる。

サンプリング調査は19歳から21歳のピアノの演奏経験がある3名によって行われた。楽曲は「エリーゼのために」を用い、テンポは110 BPMを目安に演奏した。自由に演奏してもらったときの鼻の位置のグラフを図6に示す。また、なるべく動かずに演奏してもらったときの鼻の位置のグラフを図7に示す。図6中の丸で囲っている時刻でリズムノリ動作を行っている。時刻1200のときのリズムノリ動作で楽譜がめくられたこと以外、誤動作はなかった。リズムノリ動作と頷き動作を比べると、リズムノリ動作ではあまり深く頷いていないことが明らかとなった。

6. 閾値を用いた頷きとリズムノリの判別

サンプリング調査の結果より、頷きの深度に対して閾値を設定することで頷きとリズムノリの判別ができると考えられる。そこで、3名分のサンプルデータから頷き動作とリズムノリ動作の各々について鼻の最下点の平均を求め、さらに両者の平均値を閾値とした。この閾値よりも深く頷いた場合、譜めくり合図の頷きとしページをめくる。

閾値を設定したソフトウェアで使用実験を行った。ソフ

トウェアの最低限の両方を説明し、自由に演奏してもらい、演奏中の誤動作の有無を確認した。

実験では、15歳から21歳のピアノ演奏経験がある10名を被験者とし、楽曲は指定せず、自由に選曲してもらった。結果としてリズムノリ動作では譜めくりはされず、きちんと譜めくり合図の頷きのときのみめくられることを確認した。しかし、譜めくり合図の頷き10回に対して2回の割合でリズムノリ動作として誤認識してしまい楽譜がめくられないことがあった。これは閾値算出に用いた動作のサンプル数が足りず、妥当な閾値でないことが原因であると考えられる。この問題は、閾値の個人化を行うことで改善できると考えられる。その判別方法として、演奏音からの楽譜追跡技術 [8], [9] を用いることで、楽譜の右端付近以外で行われる動作を無視することができると考えられる。

7. まとめ

本研究では演奏の妨げにならず、手軽に利用できることを目的として頷きのジェスチャ認識による自動譜めくりについてソフトウェアの開発を行ってきた。開発したソフトウェアを実際に用い、頭部の動きのサンプルを集めた。その結果、テンポを感じて動くリズムノリ動作は譜めくり合図の頷き動作と比べて深く頷かないことが判明した。そこで頷き動作とリズムノリ動作を区別する閾値を設定し、使用実験を行った。結果として、きちんと譜めくり合図の頷き動作のときのみ楽譜がめくられることを確認した。しかし、譜めくり合図の頷き10回に対して2回の割合でリズムノリ動作として誤認識してしまった。

今後の課題として、閾値の個人化を行い誤認識が改善されるか調査する必要がある。その他、システムの改善のためにインターフェースの改善や、ページ戻り機能を実装していきたい。

謝辞 本研究の一部は日本学術振興会の科学研究費補助金 基盤研究 (A) 26240025 による。

参考文献

- [1] 自動譜めくり機「フメクール」(online), 入手先 <http://inforent.dreamblog.jp/blog/78.html> (2016.04.19).
- [2] 電子楽譜フェアリー (online), 入手先 <http://www.fairy-score.com/> (2016.04.19).
- [3] 地曳はるか, 清水寿一, 齋藤康之: ジェスチャ認識による自動譜めくりに関する研究, 映像情報メディア学会技術報告, Vol. 39, No. 8, pp. 61-64, 2015.
- [4] OpenCV (online), 入手先 <http://opencv.org/> (2016.04.19).
- [5] 石井健一郎, 上田修功: 続・わかりやすいパターン認識-教師なし学習入門-, オーム社, pp. 138-143, 2014.
- [6] Java (online), 入手先 <https://www.java.com/jp/> (2016.04.19).
- [7] PDFRenderer (online), 入手先 <https://developer.android.com/reference/> (2016.04.19).
- [8] A. Arzt, G. Widmer, and S. Dixon: Automatic page turning for musicians via real-time machine listening, *Proc. of ECAI2008*, pp.241-245 (2008).
- [9] E. Nakamura, T. Nakamura, Y. Saito, N. Ono, and S. Sagayama: Outer-Product Hidden Markov Model and Polyphonic MIDI Score Following, *Journal of New Music Research*, Vol. 43, No. 2, pp. 183-201, 2014.