

頭部動作推定による演奏者と自動伴奏システムの演奏同期*

○齋藤 康之（木更津高専・情報），西本 卓也，嵯峨山 茂樹（東大院・情報理工）

1 はじめに

本研究の目的は、頭部動作により自動伴奏システムの演奏開始タイミングと演奏者のテンポ感を推定し、演奏同期をとることである。

演奏者の演奏を追跡し、対応する伴奏パートを再生する自動伴奏システムは、Dannenbergsの研究以降注目され続けている [1]。その理由として、単にアプリケーションとしての面白さや使って楽しいというだけでなく、合奏曲演奏者の練習補佐としての使用を期待されている点が挙げられる。また、武田らは MIDI 入出力可能な電子ピアノのための自動伴奏システム（以下、Eurydice）を構築した [2]。Eurydice の新たな特徴として、演奏者の演奏誤りや演奏欠落だけではなく、弾き直しやジャンプも考慮したモデル化を行い、これらに対応できることが挙げられる。

上記の先行研究による自動伴奏システムに共通する問題点として、曲の冒頭で演奏者が休みで伴奏が先に演奏する場合に、楽譜で指定されたテンポで演奏し始め、必ずしも演奏者のテンポ感と合致するとは限らない点が挙げられる。また、演奏開始を指示したらずぐに伴奏演奏を開始するため、演奏者が演奏準備を整えにくい場合もある。

そこで本研究では、自動伴奏システムにおけるこれらの問題を解決するために、演奏者の頭部動作により伴奏開始タイミングを制御し、演奏者のテンポ感を推定することで、演奏者と自動伴奏システムの演奏同期をとる。

2 演奏者の頭部動作推定方法

2.1 演奏者の演奏同期

少人数のアンサンブルのように指揮者が不在で演奏者自身が演奏開始合図を出すときは、鼻からのプレスで音を立てる方法や身体の動き（頭部や楽器の振りなど）による方法が挙げられる。

堀内ら [3] は、前者のプレスによる合図による伴奏システムの制御について検討している。しかしながら、余計な音を発しないことが望ましい状況や演奏中では、プレスによる合図は原理的に不向きである。

一方、身体動作であればそのような問題は生じないことから、本研究では身体の動きにより演奏者と自動伴奏システムの演奏同期（伴奏開始タイミングと、テンポ感の合致）をとる。

2.2 演奏者の頭部動作情報の取得

演奏者の頭部動作を得る上で、マーカやセンサ類を身体に付着することなく、かつ、自然な演奏動作ができることが望まれる。そこで、今回は距離情報とカメラ映像を同時に取得できる Microsoft XBOX360 Kinect センサー（以下、Kinect）を用いる。距離情報を用いることで演奏者の背景の影響を抑えられ、ま

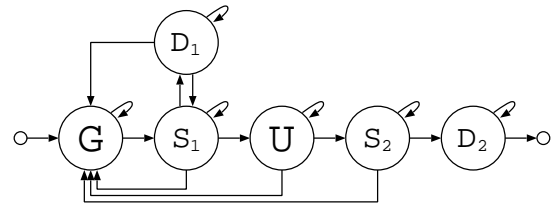


Fig. 1 頭部動作モデル

た、演奏者の身体領域の存在範囲を限定できるので、演奏者の頭部を精度よくかつ高速に抽出できる。

頭部による演奏同期動作により、顔の向きが変化するので、見かけ上、顔部品が移動する。したがって、この顔部品の移動量を頭部動作情報として取得すればよい。このとき、目や口は形状変化が生じるので、今回は鼻領域を追跡することとした。その処理の流れを以下に示す。

1. Kinect から得られる距離情報により、演奏者の身体領域を抽出する。
2. 身体領域内に限定して、カメラ映像から Haar-like 特徴量 [4] により顔領域を検出する。
3. 顔領域の中心付近（鼻領域）を抽出し、テンプレートとする。
4. テンプレートを用いてパターンマッチングを行い、鼻領域を検出（顔追跡）する。
5. 前の映像フレームと現在の映像フレームでの鼻領域の移動量 (dx, dy) を、頭部動作情報とする。

2.3 演奏同期動作のモデル化

演奏同期動作をするとき、演奏者は基本的に「静止→上向き→静止→下向き」の動作をする。また、演奏同期動作をする前に、横を向くなどの余分な動作（以下、一般動作と呼ぶ）も生じうるし、テンポを取りながら若干下に向く可能性もある。これらの動作は確定的ではなく、すなわち、仮に同じテンポ感であっても各動作の持続時間は変化しうる。

これら頭部動作情報は、各動作を最も出力しやすい複数の信号源が次々に切り替わることで観測される非定常時系列データとみなすことができ、隠れマルコフモデル (HMM) によりモデル化するのが適切と考えられる。

まず、HMM の状態数を決定する。演奏同期動作は、基本的に本節冒頭のような動きをすることから、これらを出ししやすい 4 状態と、さらに、一般動作と演奏同期動作をする前の若干下に向く動作を出ししやすい 2 状態を加えた計 6 状態から成るとした。Fig. 1 は、設計した HMM を示す (HMM は本質的に各状態にラベル付けすることはしないが、以下、便宜上、

*Synchronization of musical performance between human player and automatic-accompaniment system using head motion estimation. by Yasuyuki SAITO (Dept. of Info. Eng., KisarazuNCT), Takuya NISHIMOTO and Shigeki SAGAYAMA (Grad. School of Info. Sci. and Tech., the Univ. of Tokyo)

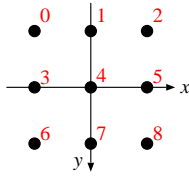


Fig. 2 ガウス関数の平均 μ_k の位置. 赤字の番号は μ_k の k に対応.

最も出力しやすい観測を用いて各状態を指し示すこととする).

次に, 各状態から出力する動作の観測確率を設定する. 各状態とも, Fig. 2 のような静止を含む 9 方向の頭部動作情報を出力するが, 状態の特性を考慮してその確率を高くあるいは低く設定する. 状態 G は, 一般動作を吸収するためのガーベッジモデルに対応させ, 横方向や斜め方向の動きが観測されやすいように観測確率を設定した. S_i は静止状態が, U は上向きの動作が, D_j は下向きの動作が, 各々の観測されやすいように確率を高くした.

状態遷移確率は, 経験的に, 自己遷移確率を 0.8 とし, 他の状態への遷移確率を 0.1 とした.

そして, 頭部動作情報を観測値としてそのまま確率計算するために, 確率密度関数をガウス関数でモデル化した. 確率密度関数は, 今回は大小関係が判別できれば十分なので, 計算の簡単化のためにガウス関数の係数と分散を 1 とした. また, 平均 $\mu_k = (\mu_{k_x}, \mu_{k_y})$, ($k = 0, 1, \dots, 8$) の位置を Fig.2 の 9 方向に変化させながら, (dx, dy) が各方向の動きである確率を式 (1) で求める.

$$N(dx, dy, \mu_k) = \exp\left\{-\frac{(dx - \mu_{k_x})^2}{2} - \frac{(dy - \mu_{k_y})^2}{2}\right\} \quad (1)$$

これより, ある状態 m での観測確率 P_{o_m} を式 (2) で求める.

$$P_{o_m} = \max_k \{P_{b_{m,k}} N(dx, dy, \mu_k)\} \quad (2)$$

ここで, $P_{b_{m,k}}$ は, 状態 m で平均 μ_k のガウス関数が観測される確率を示す.

2.4 演奏同期動作の推定

前節で述べた HMM に対し, Viterbi アルゴリズムにより累積状態確率 P_a を算出する. そして, トレースバックによりどの状態を経由したのかを調べる. 状態 S_1 から状態 U に遷移したときから, 状態 S_3 に到達するときまでの映像フレーム数 F を用いて式 (3) によりテンポ T (M.M.) を求める. ここで, 映像フレームレートは 30 fps (0.033 sec/frame) である.

$$T = \frac{60}{F \times 0.033} \quad (3)$$

ただし, Viterbi パスをトレースバックするとき, 仮に全て一般動作であっても, また, 演奏同期動作中であっても, 初期状態まで到達する. そこで, 演奏同期動作をしていたか, それが終了したかを判定する必要がある.

演奏同期動作中であれば, 最新の時刻で $P_{a_{D_2}}$ が他の P_{a_m} よりも大きくなる. また, 演奏同期動作が終了すれば, 最新の時刻で $P_{a_{D_2}}$ よりも, 他の P_{a_m} のうちいずれかが大きくなる. これらの特性を用いて, 演奏同期動作の開始時点および終了時点を推定する.

3 実験結果

まず, 人工的に作成した動き情報のデータ系列を用いて検証したところ, 演奏同期動作の終了時点もテンポ推定も正確に行えた.

次に, M.M.=60 としてメトロノームを鳴らし, 演奏同期動作を行ったところ, 56~67 の範囲に, M.M.=88 の場合は, 72~96 の範囲に収まった.

誤差の原因として, 人の演奏同期動作の正確性が挙げられる. その対策として, 演奏者特有の癖 (遅れ, 進み) を事前に学習しておいて補正することや, 演奏に必要なテンポを何段階かあらかじめ設定したルックアップテーブルを構築しておき, 演奏同期動作から推定したテンポに近いものを動的に選択できるようにすることなどが考えられる.

そして, Eurydice に組み込んで自由に演奏同期動作をしたところ, あまり違和感のないテンポおよび伴奏開始タイミングで伴奏が演奏された. 一方, メトロノームに合わせる場合は, 伴奏開始タイミングにずれが生じた. これは, 本来ならばメトロノームが鳴った瞬間に伴奏が始まらなければならないが, 演奏同期動作の終了時点がメトロノームの鳴った瞬間に合わせやすいためである. すなわち, 現在の方法では原理的にメトロノームの鳴りよりも若干遅れて伴奏が開始されることになり, メトロノームと完全に一致させることは困難である (メトロノームの鳴りよりもわずかに早く演奏同期動作を終了させる必要がある). そのため, 半拍到相当する時点 (最新の時刻で状態 S_2 の確率 $P_{a_{S_2}}$ が他の P_{a_m} よりも高いとき) で演奏同期動作の終了時点を予測するなどの工夫を要する.

4 まとめ

本研究は, 演奏者と自動伴奏システムの演奏同期をとるべく, 演奏者の頭部動作により伴奏開始タイミングの制御と演奏者のテンポ感を推定した.

今後の課題として, 伴奏開始タイミングの予測や, 曲中に演奏者が休みまたは長い音符を演奏している場合および ritardando から a tempo になるときの演奏同期などが挙げられる.

参考文献

- [1] Dannenberg, R. B., "An On-line Algorithm for Real-time Accompaniment," Proc. Int. Comp. Music Conf., pp.193-198, 1984.
- [2] 武田, 西本, 嵯峨山, "HMM による MIDI 演奏の楽譜追跡と自動伴奏," 情処研, 2006-MUS-66, pp.109-116, Aug. 2006.
- [3] 堀内, 西田, 市川, "プレスによる合図を検出する伴奏システム," 情処論, Vol.50, No.3, pp.1079-1089, Mar. 2009.
- [4] 永田, "実践 OpenCV 映像処理&解析," カットシステム, 2009.