

JamGesture:スマートフォンを用いた身体動作による 即興演奏支援システム

水野創太^{†1}, 白松俊^{†1}, 北原鉄朗^{†2}, 一ノ瀬修吾^{†1}

^{†1}名古屋工業大学 ^{†2}日本大学

概要: 演奏者の身体動作は、その動きの視覚的効果によって音楽理解を促進する。特に、旋律の上下動（旋律概形）は直感的な身体動作と親和性が高い。我々はこれまで、モーションセンサーカメラ、スマートフォンセンサーによってユーザの身体動作を認識する手法を提案し、身体動作による演奏行為を支援するシステムの開発を行ってきた。本稿では、これまで開発してきたスマートフォンを用いた身体動作認識手法と、北原らによる旋律概形からメロディ生成するシステム JamSketch を統合することで、即興演奏支援システム JamGesture を開発した。JamGesture は、スマートフォンを用いて認識したユーザの手の上下動から描画した旋律概形を基に、JamSketch の機構によってメロディを生成することで、ユーザの直感的な身体動作を入力とした即興演奏を可能とする。

キーワード: 即興演奏, 身体動作, スマートフォン, インタラクション

JamGesture: Gesture-based Improvisation Support System using Smartphone Sensors

Souta MIZUNO^{†1}, Shun SHIRAMATSU^{†1}, Tetsuro KITAHARA^{†2},
Shugo ICHINOSE^{†1}

^{†1}Nagoya Institute of Technology ^{†2}Nihon University

Abstract: The Physical gesture promotes musical comprehension because performer's gesture can provide the visual information of musical performance to others. Especially, pitch contours (melodic outlines) have a high affinity the intuitive physical gesture. We have proposed a method to recognize user's physical gesture using a motion sensor camera and smartphone sensors, and have developed a system for supporting the musical performance by physical gestures. In this paper, we develop an improvisation support system "JamGesture" by integrating a method of recognizing physical gesture using smartphone that we have developed and "JamSketch", a system for melody generation based on melodic outline developed by Kitahara et al. The JamGesture enables users to play improvisation using input as user's intuitive physical gestures by the melody generation function of the JamSketch based on melodic outline, which is drawn by a vertical motion of user's hands recognized using smartphone sensors.

Keywords: Improvisation, physical gesture, smartphone sensor, interaction

1. はじめに

演奏者の身体動作は、その動きの視覚的効果によって音楽理解を促進する。特に、旋律の上下動（旋律概形）は直感的な身体動作と親和性が高い。身体動作を用いた音楽演奏システムを Serge らは Meta-Instrument と定義した[1]。実際に、身体動作を含んだ様々な Meta-Instrument が開発されてきた。例えば、Leap Motion センサーを用いた Marcella らによって開発された "Hand Composer" [2], Intel RealSense Camera を用いた "KAGURA" [3], センサーカメラ Kinect を用いて Sertan らによって開発された "Crossole" [4]や、センサーグローブである CyberGlove と 3 次元位置計測センサーである Polhemus FASTRAK を用いて Horace らによって開発された "Cyber Composer" [5]などが存在する。これらの研究は、ユーザの身体動作を認識するために特別なデバイスを用いて行われている。本研究では、我々はスマートフォンに搭載されたセンサーを代表とした幅広く一般に普及したデバイスを用いた Meta-Instrument の開発を目的としている。

身体動作の代表的な利用法として、ユーザの上下位置を

音高に変換する手法が考えられる。我々はこの手法を用いて、ユーザの手の上下動を手を持ったスマートフォンでトラッキングし入力として用いることで、単純な手の上下動から背景楽曲に合わせた即興合奏を支援するシステムを開発してきた [6]。

本稿では、これまでに開発してきたスマートフォンを用いたユーザの身体動作認識手法と、北原らによる旋律概形からメロディ生成するシステム JamSketch [7]を統合することにより、即興演奏支援システム JamGesture を開発した。以下では、スマートフォンを搭載されたセンサーやカメラを利用したポジショントラッキング手法や、JamSketch システムとの統合について述べる。

2. JamGesture のシステム構成

図 1 に JamGesture のシステム構成図を示す。JamGesture ではスマートフォンに搭載された各種センサー、カメラを利用したポジショントラッキングによってユーザの手の高さを認識し、その認識結果を Bluetooth 通信によって PC 上で動作する JamSketch システムに送信する。JamSketch システムでは、送られたユーザの手の上下動を基に旋律概形の

描画と描画された旋律概形を基にメロディ生成を行っている。図 2 に、実際の旋律概形描画時のユーザインタフェースを示す。この描画された旋律概形を基にした旋律の生成手法については、JamSketch [7]で用いられている手法をそのまま用いている。JamSketch システムでは、図 2 のように描画された旋律概形を基に、リズムの決定、音高の決定による旋律の生成と演奏の表情付けを行っている。しかし、今回 JamGesture の実装にあたっては、JamSketch システムによる演奏の表情付け処理は用いていない。

旋律の生成は 1 小節単位で行われ、旋律の生成後に演奏の表情づけが行われている。これらの処理には約 1.0 秒前後の処理時間を必要とするため、システムによる旋律概形の描画は各小節が演奏される 1.0 秒程度前に行われている。

リズムの決定においては、仮のリズム R' を生成する。 R' は 1 小節ごとにリズムの分解能を 3 連 8 音符とした各要素を 1 (オンセット) か 0 (直前の音符の継続) の値をとる 12 次元ベクトルを持つ。例: $R'=(1, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0)$ は 2 分音符が 2 つ続くことを意味する。あらかじめ定められたリズムパターン集合 R のうち R' に最も類似したものをリズムの生成結果としている。リズムの生成結果を基に遺伝的アルゴリズム(GA)を用いて、生成されるノートそれぞれにたいして音高の決定を行っている。

3. 身体動作入力による旋律概形描画

本研究では、即興演奏における旋律生成に用いる旋律概形の描画にユーザの身体動作を入力として用いる。JamSketch では、図 3 のピアノロール上にマウスなどのポインティングデバイスを用いて直接旋律概形を描画するのに対し、本システムでは、手の上下動のみをユーザからの入力として使用する。そこで、横軸(時間軸)の座標は、現在時刻のちょうど 1 小節先とし、伴奏の進行に合わせて自動的に変化するように JamSketch を修正した。

ユーザの手の上下動は、スマートフォンのセンサーまた

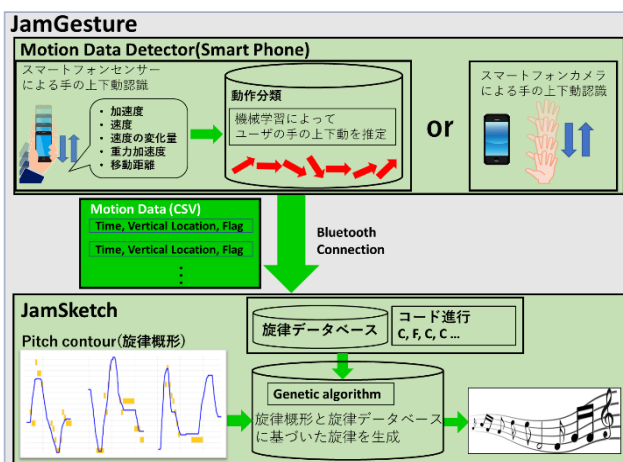


図 1 JamGesture のシステム構成図

Figure 1 System Configuration of JamGesture

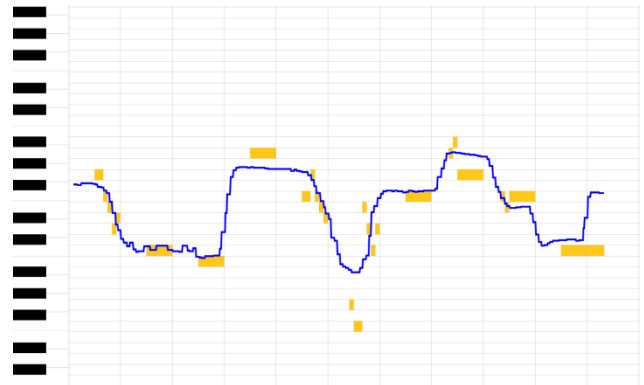


図 2 システムによる旋律概形描画時の画面例

Figure 2 example when drawing pitch contour by the system

はカメラで認識した後、JamSketch における縦軸(音高軸)の座標として用いられる。演奏経過時間 t におけるユーザの手の上下動を表す入力 H_t を生成し、Bluetooth 通信によって PC 上で動作する JamSketch に送信することによって描画を行う。また、JamSketch システムによる旋律概形描画では、マウスのクリックによって描画タイミングを指定していたため、本システムではスマートフォンを使った操作によって描画タイミングの指定も行う。

本稿では、スマートフォンを用いた身体動作の入力への変換手法として以下の 2 手法を紹介する。

1. スマートフォンセンサーの利用: スマートフォンに搭載された各種センサー(加速度センサー, ジャイロセンサー等)を利用しポジショントラッキングを行う。
2. スマートフォンカメラの利用: OpenCV ライブラリを利用したスマートフォンカメラによるポジショントラッキング

3.1 スマートフォンセンサーによるポジショントラッキング

スマートフォンセンサーによるポジショントラッキング手法ではベイジアンネットワークを用いて、スマートフォンセンサーから得た値を入力として用いることでスマートフォンを持つユーザの手の上下動推定を行う。

(1) トレーニングデータの作成

ベイジアンネットワークのモデル構築に用いるトレーニングデータの作成する際、被験者に背景楽曲に合わせてスマートフォンを持った手を上下に動かした際のスマートフォンに搭載されたセンサーから求めた加速度 a , 速度 v , 速度の変化量 vc , 移動距離 p , 重力加速度 g を約 5ms 秒おきに取得する予備実験を行った。

実験を行う際、被験者の手をセンサーカメラ Kinect によって認識し、記録した手の座標値をセンサーから求めた値に対する教師データとして利用した。教師データとして用いる際、図 3 のようにスマートフォンセンサー値を取得した時間と一致した時間における手の座標値の変化量の値に

応じて、5段階に量子化した瞬間のユーザの上下動作を表す変数 $h = \{-2, -1, 0, 1, 2\}$ を用いている。

図4に作成したトレーニングデータを用いて構築したベイジアンネットワークモデルを示す。スマートフォンセンサーから得られる加速度 a 、速度 v 、速度の変化量 vc 、移動距離 p 、重力加速度 g を入力することでユーザの上下動作を表す h の推定を行う。

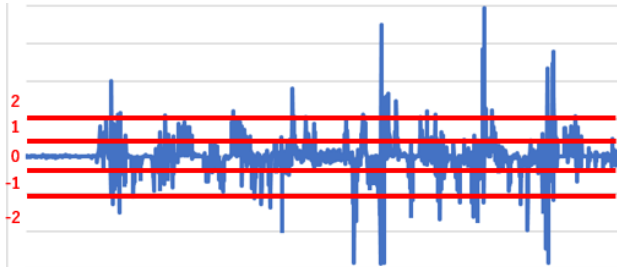


図3 教師データの量子化

Figure 3 Quantization of training data.

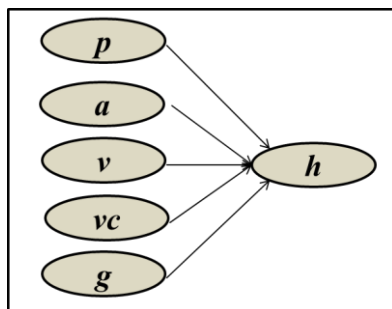


図4 ユーザの動作の上下動を推定するベイジアンネットワーク

Figure 4 Bayesian Network estimates vertical movement of user's motion

(2) 旋律概形描画座標への変換

JamSketchによって旋律概形の描画を行う際には、(1)のモデルによって推定される瞬間のユーザの上下動作 h_t からその時点のユーザの手の高さ H_t を求める必要がある。

約5ms秒おきにスマートフォンセンサーの値を入力として推定される値 h_t を直前までのユーザの手の上下動値を表す値 H_{t-1} に加算していくことで現在のユーザの手の位置を決定する。

$$H_t = H_{t-1} + h_t$$

スマートフォンセンサーによるポジショントラッキング手法を用いた旋律概形描画では、スマートフォンの画面をタップ時のみ旋律概形の描画を行う。

3.2 スマートフォンカメラによるポジショントラッキング

スマートフォンに搭載されたカメラを用いた手法では、カメラによってユーザの手を認識することで、手の座標の上下動を入力に用いる。カメラによる画像処理には、OpenCVライブラリを用いて行う。

(1) 手の認識

カメラによる手の認識方法には様々な手法が考えられるが本稿では、色抽出による手法を用いる。スマートフォンに搭載されたインカメラから得たカメラ画像からあらかじめ指定した範囲の画素値に該当する領域を抽出する。抽出された領域のうち最大範囲を持つ領域をユーザの手を表す領域 f として扱い、領域 f のうち x, y 両軸における中心座標 p_f を入力として用いる。

図5に実際に色抽出による手の認識時の様子を示す。本稿では、あらかじめ色を塗った軍手を手にはめることで色抽出による手の座標の認識を行っている。図5より、青色で塗られた部分が手を表す領域 f として認識していることがわかる。また、領域内の中心点を表す点 p_f のカメラ認識範囲における座標を0~1の範囲で正規化し、旋律概形描画における y 座標として用いる。

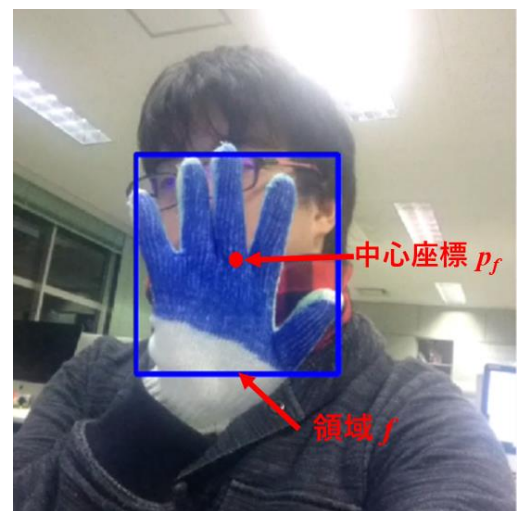


図5 色抽出手法による手の認識

Figure 5 hand recognized by method of color extraction

(2) カメラ手法における描画 On,Off 切り替え

スマートフォンセンサーを用いた手法と同様に旋律概形描画のタイミングをユーザが指定するための操作が必要となる。本稿では、切り替えのための操作として色の認識範囲の比較を利用する。

図6に旋律概形描画の切り替えにおける描画の変化の様子を示す。セクション(1)における手の認識において用いた色の抽出手法によって、手の認識に用いる色とは別の色を表す画素値を指定し、指定の画素値における最大領域 F を

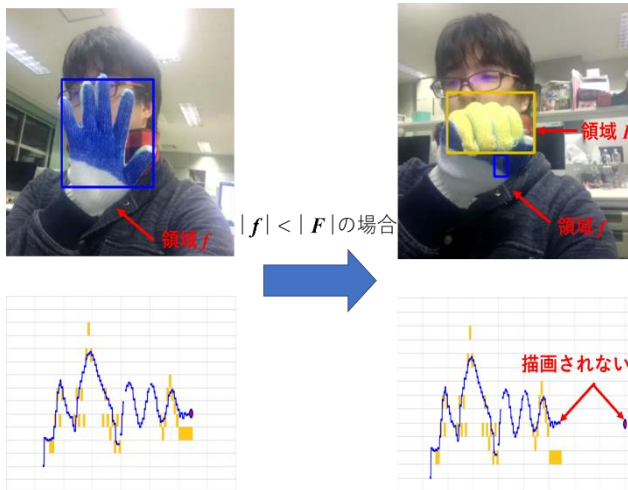


図 6 描画タイミングの切り替え操作

Figure 6 Switching operation of timing of drawing

抽出する。二つの領域 f , F の領域範囲の大きさ $|f|$, $|F|$ を比較し以下のように操作を切り替える。

- $|f| > |F|$ の場合: 旋律概形の描画を行う
- $|F| < |f|$ の場合: 旋律概形の描画を行わない

4. 実行例

本稿で紹介したスマートフォンセンサーを用いた場合、スマートフォンカメラを用いた場合、それぞれの手法における身体動作認識を用いて JamGesture システムを利用し、簡単な上下動の認識を行った際のユーザインタフェースの旋律描画の実行例を図 7 に示す。

図 7 より、両手法において上げた、下げたという大まかな認識については正しく行えているが、スマートフォンセンサーによる認識については図 7 左より、動きの不安定さや動きの大きさの認識という点では改善の必要があるように感じられた。カメラを用いた手法については図 7 右より、ある程度精密な認識が行えていることがわかるが、あらかじめ指定した色を認識に用いなければならない点やカメラで認識する実験環境の色に影響されやすい点などが改善点として感じられた。これらの問題点の改善については今後の課題としていきたい。

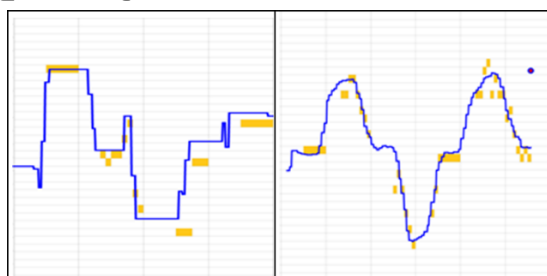


図 7 JamGesture 実行例

(左: センサーを用いた場合, 右: カメラを用いた場合)

Figure 7 Example of JamGesture execution

(left: case of using sensors, right: case of using camera)

5. おわりに

本稿では、我々がこれまで開発してきたスマートフォンによるユーザの手の上下動認識手法と北原らの開発した JamSketch による旋律概形に基づいた旋律生成を統合した即興演奏システムを開発した。今後は、スマートフォンによる手の上下動の認識精度の向上と、演奏支援システムとしての機能追加を図っていきたい。

謝辞 本研究の一部は、JSPS 科研費 (No. 25870321, 17K00461, 17H00749, 16K16180, 16H01744, 26280089, 26240025) の支援を受けた。

参考文献

- [1] Serge de Laubier, "The Meta-Instrument," Computer Music Journal, 22(1), pp. 25-29, 1998.
- [2] M. Mandanici, S. Canazza, "The "Hand Composer": Gesture-driven Music Composition Machines," in Proceedings of Post-Conference Workshop NRF-IAS-2014, 2014.
- [3] P. Rothenberg (2016), "Kagura shows what the future of music could be like", Tech in Asia, <https://www.techinasia.com/kagura-kickstarterfuture-music>, (参照 2017-06-02).
- [4] Sertan Sentürk, Sang Won Lee, Avinash Sastry, Anosh Daruwalla, Gil Weinberg, "Crossole: A Gestural Interface for Composition. Improvisation and Performance using Kinect," in Proceedings of the 2012 International Conference on New Interfaces for Musical Expression, pp. 449-502, 2012.
- [5] Horace H S Ip, Ken C K Law, Belton Kwong, "Cyber Composer: Hand Gesture-Driven Intelligent Music Composition and Generation", in Proceedings of 11th International Multimedia Modelling Conference, 2005.
- [6] Souta Mizuno, Shugo Ichinose, Shun Shiramatsu, Tetsuro Kitahara, "Support System of Improvisational Ensemble Based on User's Motion Using Smartphone Sensors" in Proceedings of The Twelfth 2017 International Conference on Knowledge, Information and Creativity Support Systems (in press).
- [7] Tetsuro Kitahara, Sergio Giraldo, Rafael Ramirez: "JamSketch: Improvisation Support System with GA-based Melody Creation from User's Drawing", Proceedings of the 13th International Symposium on Computer Music Multidisciplinary Research, pp.352--363, September 2017.