3-410-69
2017/3/2

スマートフォンセンサーを用いた即興合奏のための身体動作認識機構の試作

水野創太[†1]　一ノ瀬修吾[†1]　白松俊[†1]　北原鉄朗[†2]

概要：本研究では，演奏経験の少ないユーザでも即興合奏を可能にするシステムの実現を目指している．演奏未経験者にとってコード進行に合う音高を決めるのは困難だが，旋律概形すなわち音高の上下動なら身体動作により表出可能である．我々のこれまでのシステムでは，モーションセンサーカメラを用いてユーザの身体動作を認識していた．本稿では，スマートフォンセンサーを用いて，ユーザの身体動作を認識する手法を開発し，より手軽に即興合奏に参加できるシステムを目指す．既存楽曲に合わせてスマートフォンを動かした際のセンサデータを訓練データとして収集し，背景楽曲のコード進行に合う音高を決定するベイジアンネットワークのモデルを獲得する実験を行った．

Prototyping a Function to Recognize Users' Body Motion for Improvisational Ensemble Using Smartphone Sensors

Souta MIZUNO[†1]　Shugo ICHINOSE[†1]
Shun SHIRAMATSU[†1]　Tetsuro KITAHARA[†2]

Abstract: We aim to develop a system enabling non-experts of musical performance to play improvisational ensemble. Although non-experts of music cannot determine a pitch suitable for a particular chord, they can express pitch contour by their body motion. Our previous system recognizes users' body motion on the basis of a sensor camera. In this paper, we aim to develop a method for recognizing users' body motion on the basis of smartphone sensors in order to enabling multiple users to easily join improvisational ensemble. We collected training data of correspondence between melody and body motion. Moreover, we developed three Bayesian Network models for estimating pitch name, vertical movement of pitch, and attack time.

1. はじめに

本研究ではこれまで，演奏経験の少ないユーザを対象とし，手の上下動によって背景楽曲との即興合奏が可能なシステムを開発してきた[1]．従来は，モーションセンサーモジュール「Intel RealSense 3D」カメラを用いてユーザの手の上下動を入力してきたが，特別なデバイスを必要とする，カメラの認識範囲内でしか行えないなど，多人数が容易に参加するのが困難になるという問題があった．そこで本稿では，より手軽に参加できる即興合奏支援システムを目指し，スマートフォンのセンサーを用いて同様のシステムを実現する手法を検討する．手軽な参加を可能にすることで，例えば初対面の人間が多数集まるワークショップなどにおいて場を和ませるアイスブレイクと呼ばれる用途にも使える可能性がある．

加速度センサーをはじめとした各種センサーによって基準点からの高さを認識することでユーザの動きをトレースするプログラムを作成し，高さに基づいて音高を決定するシステムを用いたところ，人間の小さな動きを判定することが難しいことが課題となったため，本研究では，北原らの手法[2]を参考に，ベイジアンネットワークの確率モデル推定を用いて，トレースした動きのデータを訓練データとしてユーザの動き，その際に出力するべき音高を推測する手法を提案する．また，本稿では予備段階として背景楽曲として流れている曲とのセッションを目指し，流れている曲に合わせてスマートフォンを動かすことで背景楽曲のコード進行，ユーザの動きに適した音が出力されることを目的とし，スマートフォンに搭載されたセンサーからユーザの動きをトレースする手法，トレースしたユーザの動きからベイジアンネットワークによって出力するべき音高を推測する手法について述べる．

2. ポジショントラッキング

図 1 にシステム構成図を表す．本システムは，ユーザが動かすスマートフォンの上下動から旋律概形 (pitch contour) を決定し，コード進行と整合する音高へ補正する．スマートフォンの上下動をトレースするにあたって本研究では，スマートフォンに搭載された加速度センサー，ジャイロセンサー，重力加速度センサーの 3 種類のセンサーを用いたポジショントラッキングプログラムを作成した．

このプログラムでは，スマートフォンに加速度センサー，ジャイロセンサーによって計測した加速度 a と経過時間を用いて，速度 v，移動距離 p を求める．さらに，重力加速度センサーによって重力加速度 g の値を計測する．また，現在速度 v の前回の計測からの変化量を vc とする．

†1 名古屋工業大学　Nagoya Institute of Technology.
†2 日本大学　Nihon University.

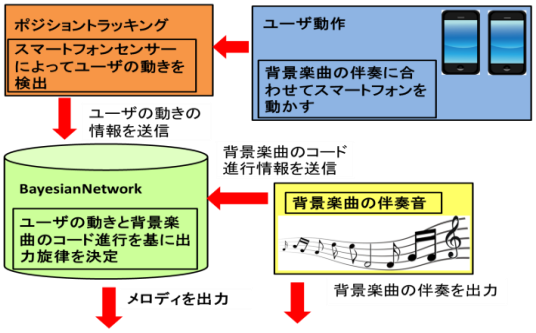

図 1　システム構成図
Figure 1 System Architecture

3. ベイジアンネットワーク

本研究では，センサーによって計測された値を入力値としたベイジアンネットワークを用いて出力する音高を推測する．

3.1 ベイジアンネットワークによる出力音推測

図 2 に本稿において用いているベイジアンネットワークモデルを表す．このモデルは，ポジショントラッキングによって求めた加速度 a，速度 v，速度の変化量 vc，移動距離 p, 重力加速度 g, 加えて背景楽曲のコード進行を表す c, 1 つ前の出力すべき音を表す n_{i-1} を入力値として出力するべき音 n_i を推測するモデルとする．音の種類には，音の出力なしを示す値も含まれているものとする．

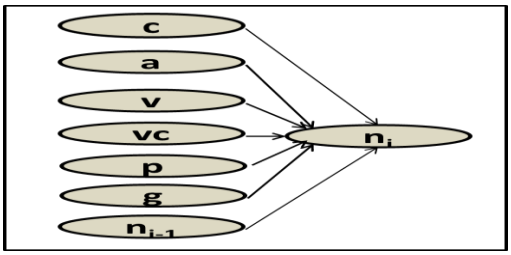

図 2　音名予測のためのベイジアンネットワーク(1)
Figure2 BayesianNetwork for Predicting Pitch Notation(1)

本稿では予備実験として，実際に背景楽曲に合わせてユーザがスマートフォンを動かした際，約 5ms ごとの加速度 a，速度 v，速度の変化量 vc，移動距離 p 重力加速度 g を取得する実験を行った．実験を行う際，発音タイミングの指定方法として以下の 3 つの方法を仮定した．

表 1　音名予測の精度
Table 1 Accuracy of Predicting Pitch Notation

	原曲と同一の音が出力されるノート数/全体のノート数
シェイク動作	0.49(131/270)
手拍子動作	0.49(133/270)
タッチ動作	0.55(148/270)

1. **シェイク動作**: スマートフォンを振るシェイク動作によって発音タイミングを指定（加速度，ジャイロセンサー等）
2. **手拍子動作**: スマートフォンを持ちながら手拍子をして発音タイミングを指定（照度センサー）
3. **タッチ動作**: スマートフォンの画面内に配置されたボタンにタッチして発音タイミングを指定

これらの実験をそれぞれ被検者 5 人に対して行い約 65000 サンプルを用意し，モデル作成のための訓練データを作成した．機械学習ソフトウェア Weka を用い，作成した訓練データを学習し，同じデータを用いて，発音タイミングにおいてデータと同じ種類の音を予測できているかどうかを表した結果を表 1 に示す．

表 1 では原曲と全く同一の音を予測した精度を示したが，本研究の目的は原曲と同一の音を予測することではなく，コード進行と不協和にならない音高を出力することである．よって，本来はコードと不協和にならない音高を正解とすべきであり， 精度も表 1 の値より高いはずであるが，これについては今後の課題とする．

この実験の結果から，精度の向上を目的として音の出力を以下の 3 つの要素に分け，3 つのベイジアンネットワークモデルを用いる手法を新しく提案する．

- 発音タイミング
- 出力する音の種類
- 音高の上下動

3.2 発音タイミングの推測

図 3 に発音タイミングを推測するベイジアンネットワークモデルを表す．このモデルでは，ポジショントラッキングによって求めた加速度 a，速度 v，速度の変化量 vc，重力加速度 g を入力値として音を出力するタイミングを表す t を{1，0}によって推測する．正しい発音タイミングから±30ms のサンプルを発音タイミングとして扱うものとする．なお，ボタンにタッチして発音タイミングを指定する方法については，センサーを用いておらず確実にタイミングを推定できるため，発音タイミングを予測する必要がないものとする．

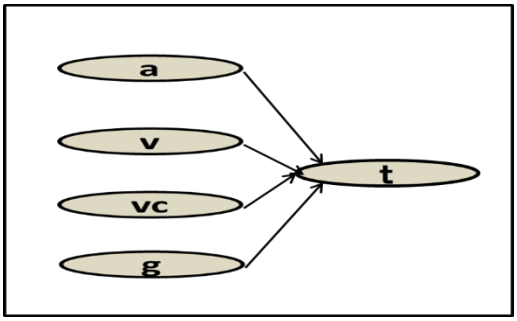

図 3　発音タイミングのためのベイジアンネットワーク
Figure 3 BayesianNetwork for Timing of Note Attack

3.3 出力する音の種類の推測

図 4 に，出力する音の種類を推測するベイジアンネットワークモデルを表す．このモデルでは，図 2 に表されているモデルを基にさらに，発音タイミングモデルの予測結果となる t と直前 n サンプルの予測結果から一番多く予測された結果を表す r_n を入力値として用いている．本稿では，n=10 と仮定して用いている．

スマートフォンの画面内に配置されたボタンを押す演奏方法の場合，発音タイミングを予測するモデルを用いていないため，発音タイミングの予測結果を表す t は用いないとする．

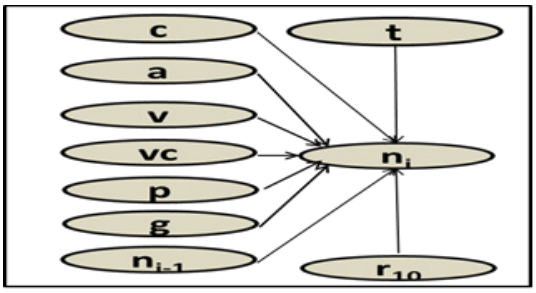

図 4　音名予測のためのベイジアンネットワーク(2)
Figure4 BayesianNetwork for Predicting Pitch Notation(2)

3.4 音高の上下動の推測

図 5 に音高の上下動を推測するベイジアンネットワークモデルを表す．このモデルでは，発音タイミングに関するモデルと同様に，加速度 a，速度 v，速度の変化量 vc，重力加速度 g，加えて移動距離 p を入力値として用いて，前に出力するべき音よりも音高が上がっているか，下がっているかを表す h を推測する．この音高の上下動の推測結果と出力するべき音の推測結果から出力する音のオクターブを決定する．

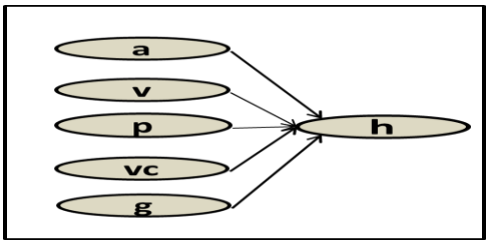

図 5　音高の上下動予測のためのベイジアンネットワーク
Figure5 BayesianNetwork for Predicting Pitch Contour

4. 評価

作成した図 3，4，5 の 3 種類のベイジアンネットワークについて図 2 のベイジアンネットワークと同様に同じ訓練データを用いてそれぞれの予測精度について確認した．

表 2　発音タイミングについての再現率と適合率
Table2 Recall and Compliance Rate about Timing of Note Attack

	再現率	適合率
シェイク動作	0.07(18/270)	0.26(18/70)
手拍子動作	0.03(8/270)	0.31(8/26)

表 3　音名予測の精度
Table3 Accuracy of Predicting Pitch Notation

	原曲と同一の音が出力されるノート数/全体のノート数
シェイク動作	0.49(133/270)
手拍子動作	0.54(147/270)
タッチ動作	0.66(178/270)

表 4　音高の上下動予測の精度
Table4 Accuracy of Predicting Pitch Contour

	正しい上下動予測がされるノート数/全体のノート数
シェイク動作	0.56(152/270)
手拍子動作	0.65(175/270)
タッチ動作	0.68(184/270)

発音タイミングについてモデルが判定したノートの数の再現率と適合率を表 2 に示す．本稿における再現率とは，発音タイミングから±30ms を 1 つのノートの発音タイミングとして，システムが予測した発音タイミングと本来の発音タイミングが一致している数と本来の発音タイミングの割合とする．

表 2 における再現率の低さには，瞬間ごとのデータが用いられていることが関係していることが予測され，これを解消するには，連続性をもつデータによる予測によってユーザの動きを推測する必要があると考えられる．

また，適合率とは，システムが予測した発音タイミングから 60ms までを 1 つの発音タイミングとして，システムが予測した発音タイミングと本来の発音タイミングが一致している数とシステムが予測した発音タイミングの数の割合とする．出力する音の種類についてモデルが判定した原曲と同一の音が出力されるノート数の割合を表 3 に示す．

本稿では，発音タイミングの最初のサンプルが原曲と同一の音を判定している場合，そのノートは原曲と同一の音が判定されているものとして扱う．

表 3 では表 1 と同様に原曲と全く同一の音を予測した精度を示したが，こちらも表 1 の場合と同様に原曲と同一の音を予測することではないため，表 3 の値より高い精度となるはずである．

音高の上下動についてモデルが本来の音高の上下動と

同じ上下動の判定を行うノート数の割合を表 4 に示す.

本稿では，発音タイミングの最初のサンプルが正しい上下動を判定している場合，そのノートは正しい上下動が判定されているものとして扱う.

5. おわりに

本稿では，スマートフォンのセンサーとベイジアンネットワークを利用した直感動作による演奏支援システムについて提案した．実際に背景楽曲に合わせたユーザの動きから得たセンサーの値を利用した訓練データを用いたベイジアンネットワークによって出力する音高をどの程度正しく推定できるかわかった.

今後は，出力する音高が完全に一致しているかどうかではなく，コード進行にあった音高が出力されているかどうかに着目し，さらなる精度の向上を目指す．また，瞬間ごとの単一データを用いた手法だけでなく HMM などの連続性を持つデータによる推測を用いた手法を利用していく予定である.

謝辞 本研究の一部は，科研費若手(B)（25870321，16K16180），JST CREST の支援を受けた.

参考文献

1) 一ノ瀬修吾，白松俊: 調性判断の不要な身体動作入力による即興合奏支援システムの試作．情報処理学会第 78 回全国大会，1Y-04 ，2016.
2) 北原鉄朗，戸谷直之，徳綱亮輔，片寄晴弘: BayesianBand:ユーザとシステムが相互に予測し合うジャムセッションシステム．情報処理学会論文誌，50(12) ，pp. 2949-2953, 2009.