2ZA-04

即興合奏支援システムのための スマートフォンセンサーを用いた身体動作認識手法

水野創太† 一ノ瀬修吾† 白松俊† 北原鉄朗‡

名古屋工業大学 工学部 情報工学科† 日本大学 文理学部 情報科学科‡

1. はじめに

旋律歌唱や旋律聴取における 3 つの処理側面として，リズム，旋律線，調性がある[1]. これら 3 要素のうち，音楽未経験者の即興合奏を難しくするのは調性判断である．そのため本稿では，音楽未経験者が即興合奏するうえで困難となる調性判断をシステムが行い，ユーザの身体動作からスマートフォンに搭載されたセンサーを用いてリズムと旋律概形（pitch contour）の入力を行うことで，スマートフォンを用いた音楽未経験者による即興合奏を支援するシステムの開発を目指す．なお，菅[2]は旋律線の手なぞりなどの身体動作が音楽理解を促進する方法として有効であるとしていることから，旋律概形の入力にユーザの身体動作を用いる手法が有効である．

本システムを実装するために，まずはスマートフォンに搭載されたセンサー（加速度センサー，ジャイロセンサー等）で計測した値からスマートフォンの上下動をトレースする単純なポジショントラッキング手法を実装した．しかし，この単純な手法では，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．そのため本稿では，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，ポジショントラッキングによって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. 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やタイミングを決定する．その推定にあたり(1)発音タイミング推定，(2)半音毎の音名推定，(3)音高の上下動推定の 3 要素に分ける．本稿では，その中でも特に重要な(1)，(2)のベイジアンネットについて述べる．

2.1 発音タイミングの推測

図 1 に発音タイミングを推測するベイジアンネットを示す．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 のインスタンスを発音タイミングとして扱うものとする．発音タイミングの推測の場合，加速度を x 軸方向の加速度 a_x，y 軸方向の加速度 a_y の 2 種類を用いる．

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．

2.2 半音毎の音名推定

図 2 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．ポジショントラッキングによって加速度 a，速度 v，速度の変化量 vc，重力加速度 g に加えて移動距離 p，発音タイミングモデルの予測結果となる t と直前 m インスタンスの予測結果から一番多く予測された結果を表す r_m，1 つ前のノートの音名を表す n_{i-1} を入力値として用いて出力するノートの音名 n_i を予測する．

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 は用いないものとする．

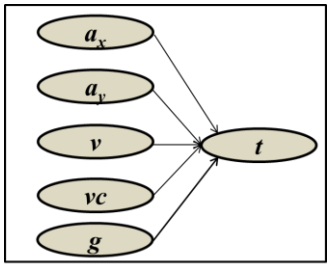

図 1 発音タイミング推定のためのベイジアンネット

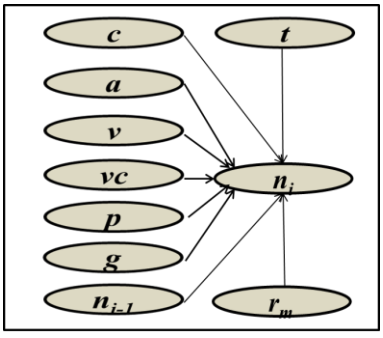

図 2 半音毎の音名推定のためのベイジアンネット

A Method for Recognizing User's Body Motion using Smartphone Sensor for Improvisational Ensemble Support System

†Department of Computer Science, School of Engineering, Nagoya Institute of Technology

‡Department of Information Science, College of Humanity and Sciences, Nihon University

3. 評価実験

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．

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 インスタンスを用意し，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，作成した図 1，2 の 2 種類のベイジアンネットワークについて，それぞれの予測精度について確認した．

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 1 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．

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．

また，発音タイミングの判定において，1 時点のデータだけでなく直前の状態との前後関係を考慮した手法への改良を検討する．具体的には，発音タイミングを推測するベイジアンネットの入力変数として，x 軸方向の加速度 a_x，y 軸方向の加速度 a_y，速度 v，速度の変化量 vc，重力加速度 g それぞれについて直前 m インスタンスのデータを入力値として加えた場合について同様に実験を行った．実験の結果を表 3,4 に示す．

表 3,4 の結果から再現率には向上が見られるが，適合率が低下していることからシステムが予測する発音タイミングの数が大きく上昇しているがその発音タイミングが間違っている場合も多いことがわかる．再現率を維持したまま，適合率を向上させることを今後の課題とする．

半音毎の音名推定実験では，m=10 に設定し，直前 10 インスタンスの予測結果のうち一番多く予測された結果を r_m の値とした．表 2 に，原曲と全く同一の音が推定されたノート数の割合を示す．これは，発音タイミングの最初のインスタンスが原曲と同一の音名だった場合を正解とした精度である．ここでは原曲と全く同一の音名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名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名を正解とすべきであり，精度も表 2 の値より高くなるはずであるが，これについては今後の課題とする．

なお，表 2 の結果は発音タイミングの瞬間の 1 インスタンスだけの音名推定結果を用いたものだが，直後の数インスタンスを考慮することで精度向上の可能性があると考えて手法の改善を試みたが，実験の結果，表 2 からの顕著な精度向上は認められなかった．これは，発音タイミングの最初のインスタンスから 9 インスタンスまでの r_{10} は，1 つ前のノートの予測結果を考慮してしまっているために間違った予測情報を用いているためであると考えられる．

表 1　発音タイミング推定の再現率と適合率

	再現率	適合率
シェイク動作	0.63 (171/270)	0.26 (171/661)
手拍子動作	0.14 (39/270)	0.31 (39/151)

表 2　半音毎の音名推定の精度

	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	0.45 (121/270)
手拍子動作	0.47 (127/270)
タッチ動作	0.56 (152/270)

表 3　発音タイミング推定の再現率と適合率（前後関係を考慮，n=1）

	再現率	適合率
シェイク動作	0.80 (217/270)	0.20 (217/1087)
手拍子動作	0.60 (164/270)	0.16 (164/1036)

表 4　発音タイミングについての再現率と適合率（前後関係を考慮，n=2）

	再現率	適合率
シェイク動作	0.86 (231/270)	0.19 (231/1241)
手拍子動作	0.70 (189/270)	0.15 (189/1296)

4. 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．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予備実験により，出力する音高や発音タイミングを推定できる可能性を示した．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，より目的に合致した評価を行う．また，時系列データの前後関係をより考慮するために，HMM 等の手法も検討する予定である．

謝辞　本研究は，JSPS 科研費（25870321， 16K16180, 16H01744, 26280089, 26240025）の助成を受けた．

参考文献

[1]　波多野誼余夫(編)：音楽と認知．Vol.8,東京大学出版会，2007.
[2]　菅道子：身体表現を取り入れた参加型音楽コンサートの可能性:カノンの理解を目指した「追いかけっこをしよう」の事例から．　和歌山大学教育学部教育実践総合センター紀要 18，pp. 121-129，2008.
[3]　北原鉄朗，戸谷直之，徳網亮輔，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.