

6ZH-04

スマートフォンセンサを用いた即興合奏支援のための LSTM による音高決定手法

高瀬 遥矢[†] 水野 創太[†] 一ノ瀬 修吾[†] 白松 俊[†]
名古屋工業大学 工学部 情報工学科[†]

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た、身体動作だけで即興合奏が出来るシステムを開発してきた。まず「RealSense 3D カメラ」で手の上下動を認識し、背景楽曲のコード進行と協和する音高を出力するシステム[1]を開発した。次に、スマートフォンセンサで手の上下動を認識するシステム[2]を開発した。

スマートフォンセンサを用いた研究では、上下動から音高を推測する手法にベイジアンネットワークが用いられ、その訓練データには童謡「エーデルワイス」1曲のみが用いられた。そのため、生成された音高がエーデルワイスのメロディの上下動に強く影響され、スマホの上下動と逆に動く問題があった。そこで本稿では、(1)複数楽曲を用いて、スマホの上下動に合わせてメロディが動くように改善する。また、(2)特定のジャンルやアーティストの楽曲を訓練データとすることで、その性質を持つメロディの生成を実現する手法を検討する。

図.1 に先行研究のシステム構成図[3]を表す。親機はコード進行や伴奏パターン決定を行い、子機に時間調整データと子機演奏情報変更データを送信する。子機は身体動作と親機から送信されたデータに基づいて旋律線を決める。本稿のシステムは子機のシステムを実装する。

2. 実装方法

図.2 に本システムの構成図を表す。青の四角はユーザデバイス、緑の円はデータ、赤の四角はモジュールである。ユーザデバイスが RealSense の場合、RealSense で手の高さを絶対値で認識したものを相対値変換し、その値とあらかじめ用意した背景楽曲の伴奏音を音高推測 LSTM の入力とすることでメロディを生成する。ユーザデバイスがスマートフォンの場合、まず、スマートフォンを上下に動かし、そのときのスマートフォンセンサの計測値を手の高さ推測 LSTM に入力すること

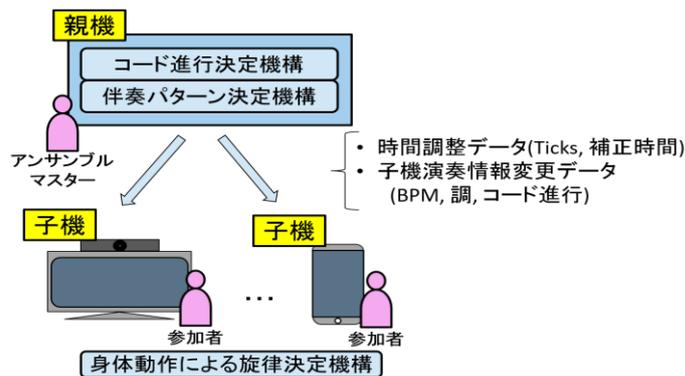
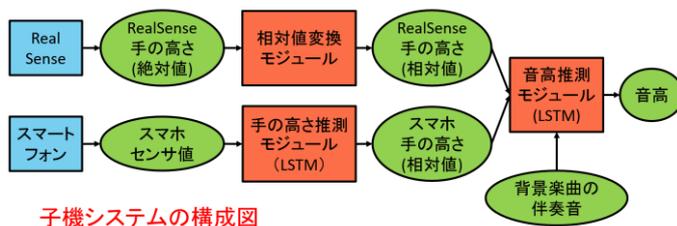


図.1. 先行研究のシステム構成図



子機システムの構成図

図.2. 本研究のシステム構成図

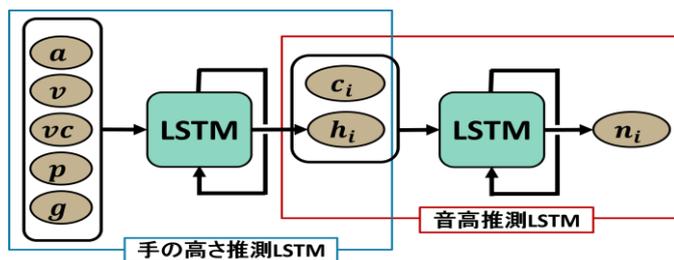


図.3. 手の高さ推測 LSTM と音高推測 LSTM

で手の高さの相対値を推測する。推測した手の高さとコード進行を音高推測 LSTM に入力することでメロディを生成する。

図.3 にスマートフォンをユーザデバイスとした時の LSTM の入力値と出力値を表す。手の高さ推測 LSTM では、入力はスマートフォンの各種センサから得た値のベクトルであり、手の高さの推測値を出力する。訓練データは、各種センサ値と手の高さの正解値(Kinectで計測)を用いる。また、音高推測 LSTM では、入力は

An LSTM-based Pitch Determination Method for Supporting Improvisational Ensemble Using Smartphone Sensors

[†]Haruya Takase, Souta Mizuno, Shugo Ichinose, Shun shiramatsu
[†]Department of Computer Science, School of Engineering, Nagoya Institute of Technology

手の高さの推測値と背景楽曲の伴奏音である。

3. 手の高さ推測 LSTM の訓練データ収集

手の高さ推測 LSTM の学習に必要な訓練データを収集した。計測方法は、下準備として midi 形式の RWC 音楽データベースからメロディが覚えやすく身体動作が簡単な楽曲を選別し、被験者に計測する楽曲のメロディを 8 小節単位で何度か聞いて覚えてもらう。その後 Kinect とスマホを同期させ、被験者がスマホから流れるメロディに合わせてスマホを上下動させた時の各種センサ値と Kinect カメラによる手の高さを同時に取得する。またこのとき被験者にはメロディの上下動の遷移が分かるように、紙に印刷したメロディのピアノロールを見ながらスマホを動かしてもらう。

その結果、合計7曲(約38,000フレーム)分のデータを収集した。またデータ収集において、「被験者のリズム感の有無によってメロディの再現度が大きく変わる」といった問題点が上がったため、今後のデータ収集においては、「被験者にとって合わせやすいメロディとなるように楽曲を編集する」、「実験に使用した楽曲より有名な楽曲を用いる」が必要である。

4. 手の高さ推測 LSTM の精度検証

収集した訓練データを用いて手の高さ推測 LSTM の学習モデルを作成し、その精度の検証と訓練データの違いによる精度比較を行った。

図.4 に全楽曲を訓練データとした時に予測された手の高さの絶対値と、正解の手の高さの絶対値(記録開始位置からの距離)の線グラフを表す。青色の線が予測された値で、オレンジ色の線が正解値である。また、横軸はフレーム番号、縦軸は手の高さの絶対値(単位は cm)である。予測線と正解線との間に大きな差は見られず、また時間が経つにつれて予測線がずれてくることもないことから、手の高さ推測に LSTM を用いることは有効であると考えられる。

表.1 に訓練データに用いる楽曲数の違いによる手の高さ推測 LSTM の精度を比較する。楽曲数 2, 4, 7 曲を用い交差検証によってそれぞれ 5 回ずつ学習させた時の予測値と正解値との誤差 0.5cm 以下の割合と平均絶対誤差 (MAE) の平均値を求めた。これら指標の妥当性のために全フレームにおける手の高さの最大値と最小値も示している。表より楽曲数が増えるにつれて精度が上がっていることが分かる。これは楽曲数を増やすことによって、特定楽曲のメロディの上下動パターンによる予測値への影響力が小さくなるためと考えられる。このことから、より様々な上下動パターンを持つ訓練データの作成を行う必要がある。

表.2 に訓練データのステップ数 (予測に使う過去のフレーム数) の違いによる手の高さ推測 LSTM の精度を比較する。表より誤差 0.5cm 以下の割合、平均絶対誤差ともにステップ数 10 が最も高かった。このことより予測に過去フレームを用いることは有効であることが分かった。また、学習時には過学習判定を行っており、ス

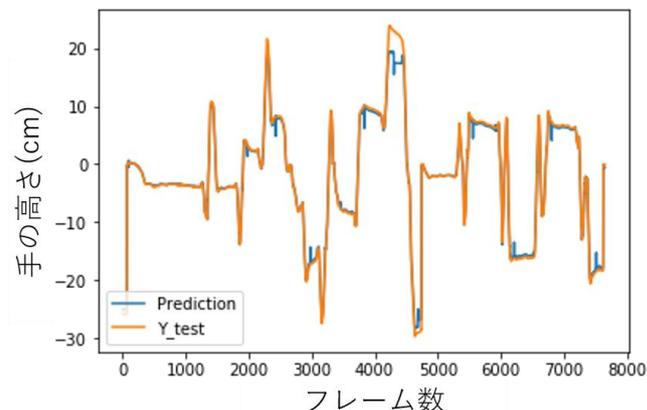


図.4. 手の高さ推測 LSTM による予測値と正解値 (手の高さは開始地点からの距離)

表.1. 楽曲数による手の高さ推測 LSTM の精度

評価指標	訓練データに用いる 楽曲数		
	2	4	7
誤差 0.5cm 以下の割合	74.7%	90.4%	91.6%
MAE	0.33	0.20	0.20
手の高さの最大値	23.9cm		
手の高さの最小値	-38.5 cm		

表.2. ステップ数による手の高さ推測 LSTM の精度

評価指標	訓練データに用いる ステップ数			
	1	5	10	15
誤差 0.5cm 以下の割合	88.8%	86.9%	91.6%	90.7%
MAE	0.26	0.21	0.20	0.26

テップ数 15 では他のステップ数に比べ早い段階で過学習と判定されてしまった。このことから予測に用いる過去フレームは大きすぎると過学習になるため、適切なステップ数を選択する必要があると考えられる..

5. おわりに

本論文では、LSTM を用いた手の高さ推測、音高推測 LSTM による合奏支援システムについて提案し、手の高さ推測 LSTM のデータ収集と精度検証を行った。その結果、手の高さ推測 LSTM の有用性と訓練データに用いる楽曲数、予測に用いる過去フレーム数による精度への影響力の違いが分かった。今後、手の高さ推測 LSTM のさらなる精度向上を行い、その後、特定ジャンルの楽曲集合を複数用意し、それらを訓練データとした音高推測 LSTM を実装、被験者実験によって訓練データに用いた楽曲の性質を持つメロディが生成可能かを検証、比較する予定である。

文献

- [1] S.Ichinose, et al. Proc. of IIAI-AAI 2017, pp.794-798, 2017.
- [2] S.Mizuno, et al. Proc. of KICSS 2017, pp.185-188, 2017.
- [3] 石川亮太他. 第 80 回情報処理学会全国大会, 1N-06, 2018.