

# ロボットと機械学習による楽器演奏パラメータ推定と 練習補助システムへの応用

黒田 迅<sup>†1,a)</sup> 上瀧 剛<sup>†1</sup>

**概要:** 楽器の練習を補助するためのシステムは、これまでに多く開発されている。代表的なものとして、音色などの評価視点において、プロの演奏家と近い評価を行うことのできるシステムや、センサを使用することで実際の演奏動作を取得し、指導を行うシステムが提案されている。しかし、これらのシステムは具体的な改善方法を提示できない、設備が大掛かりになってしまうなどの問題点がある。これに対して、本研究ではリコーダー演奏ロボットと機械学習を使用した、演奏を録音するだけで息の量や唇の張り具合といった具体的な指導を行うことのできる練習補助システムを提案する。学習したモデルは、ロボットの演奏データに対して、音高と流量を高い精度で推定することができた。また、人間の演奏データを入力した場合、人間の演奏時の息の流量と、モデルの推定した流量には相関が見られた。

## 1. はじめに

楽器の学習者が練習するにあたって一番効率的な方法は、指導者に演奏を聞いてもらい、その演奏に対する評価を受け、改善案を提案してもらうことである。しかし、初学者は、指導者の都合や初学者の置かれている環境等の理由により、指導をいつでも受けられるわけではない。そのため、学習者はもっぱら自主練習を行う必要がある。自主練習には様々な方法があるが、学習者が自身の演奏を客観的に評価して、改善を行うことは困難なことであり、練習効率が落ちてしまう。こうした自主練習の質を改善するために、様々な楽器練習補助システムが提案、開発されている [1][2]。しかし、これらのシステムは、具体的な指導を行うことができない、指導のための設備が大掛かりとなるなどの問題がある。これに対して本研究では、楽器演奏ロボットと機械学習を用いた新しい練習補助システムの提案を行う。また、学習した機械学習モデルに人間の演奏を入力したとき、出力される値が人間の演奏と相関があるかどうかを確認した。

## 2. 関連研究

練習補助システムの代表的なものとして、音色などの評価点に対して、プロの演奏家などと同等の評価を下すことのできるシステムがすでに開発されている [1]。これは、演

奏に対しプロの演奏家が評価をしたものをデータセットに使用し、機械学習等でプロと同じような評価を行えるシステムを構築するものである。このシステムを利用すると、学習者は音声を録音してシステムに入力することで、プロと同程度の評価を受けることができる。しかし、このシステムでは、評価を行うことができて、具体的な指導、つまり息の量や腕の動作、唇の動作などの改善案を提示することができないという問題がある。そのため、学習者が自身の演奏の改善を行うには、様々な演奏を試行錯誤的に試すことで評価値を上昇させていくこととなる。実際の指導者による指導では、改善案を提案してくれるため、指導を受けるよりは効率の悪い練習になってしまう。

他にも、センサで身体的な動作や物理的な量を取得することで、具体的な指導を目指すシステムも提案されている [2]。このシステムでは、主に弦楽器等において、腕の動きを検出する加速度センサやモーションセンサを演奏者や楽器に取り付け、演奏時の腕の動作を取得する。この動作をプロの動作などの基準となる値と比較することで、指導を行うことができるというものである。このシステムを利用すると、具体的な指導まで行うことができるため、自主練習の効率を向上させる事ができる。しかし、センサなどの設備や機器が必要となるため、導入に手間がかかるという問題がある。また、弦楽器等の比較的取り出しやすい動作に対しては効果的であるが、管楽器演奏時の口腔内の様子など、取得が難しい動作に対しては対応が困難である。

こうした問題を解決するために、録音するだけで唇の幅や息の量などを具体的に指導できる、管楽器向けの練習補

<sup>†1</sup> 現在, 熊本大学  
Presently with Kumamoto University  
<sup>a)</sup> kuroda@navi.cs.kumamoto-u.ac.jp

助システムの開発を目指している。具体的には、楽器演奏ロボットの演奏を入力に用いて、演奏ロボットの演奏パラメータを推定するモデルを機械学習によって構築する。その後、学習させたモデルを練習補助システムとして使用するというものである。

### 3. 提案手法

今回提案する練習補助システムを図1に示す。提案システムは、演奏録音を用いて機械学習モデルを学習させる部分と、学習させたモデルをシステムとして用いて、実際に演奏指導を行う部分に分かれている。

初めに、演奏録音から機械学習モデルを学習させる。楽器を演奏することのできるロボットの演奏を、息の量や体の動作といった、演奏時のロボット制御パラメータとともに記録し、学習用のデータセットとする。そして、データセットを使用し、演奏から制御パラメータを推定することのできるモデルを構築する。

モデルの学習が終了したら、モデルを用いて演奏指導を行う。演奏指導時には、学習済みのモデルに人間の演奏録音を入力する。このとき、モデルから出力されるのは、人間の演奏を再現することができるようなロボットの制御パラメータである。この出力を人間の演奏時のパラメータとみなすことで、演奏者の演奏時のパラメータを数値として取り出すことが可能となる。そして、学習者と上級者の間でパラメータを比較することで、具体的な指導を行うことができるようになると考えている。

本研究では、この練習補助システムのうち、ロボット演奏データセットの作成と、パラメータ推定モデルの学習を行った。その後、学習したモデルに人間の演奏を入力し、出力結果の観察を行った。

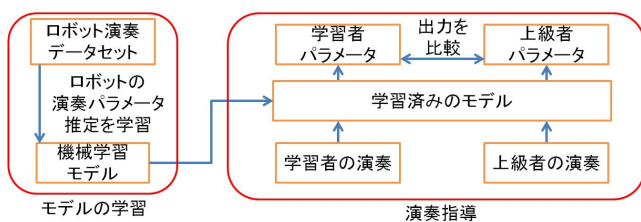


図1 提案システムの全体図。初めにロボットの演奏時のパラメータを推定するモデルを機械学習を用いて学習する。その後、学習したモデルに人間の演奏を入力することで、人間のパラメータをロボットのパラメータに置き換えて出力することができる。この値を学習者と上級者で比較することで、録音のみから具体的な指導が可能になる。

### 4. ロボットパラメータ推定モデルの学習

初めに、リコーダー演奏ロボットを用いてデータセットを作成し、ロボットの制御パラメータを推定するモデルの学習を行った。

#### 4.1 リコーダー演奏ロボット

楽器演奏を行うことのできるロボットは [3] などが存在する。著者らの研究室でも、ヒューマンインタラクティブな楽器ロボットの作成を目指して、ギター演奏ロボット [4] やリコーダー演奏ロボットの作成を行っている。

本研究のシステムを構築するために、著者らの研究室で作成しているリコーダー演奏ロボットを使用した。リコーダーロボットの全体図を図2に示す。このロボットは、リコーダーのトーンホールを抑えることで音高を変える運指部、リコーダーの吹口に空気を送り込む空気回路、演奏を制御するための制御回路からなり、ソプラノリコーダーで演奏することのできる音域すべてを演奏することができる。空気回路に取り付けられている流量弁の開度を変化させることで、楽器に吹き込まれる空気の量を変化させることができる。また、制御回路はMIDI信号によって制御され、簡単に楽曲演奏を行うことができる。使用しているリコーダーはYAHAMAのソプラノリコーダーYRS-27IIIである。

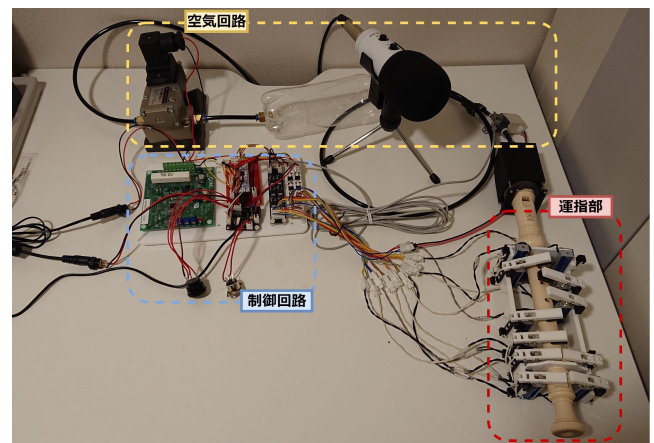


図2 リコーダー演奏ロボットの全体図。リコーダーロボットは運指部、空気回路、制御回路からなる。ソプラノリコーダーの演奏できる音域をすべて演奏することができ、様々な楽曲をMIDI信号で簡単に演奏することが可能である。

#### 4.2 データセット

リコーダーロボットを用いて機械学習モデルの学習に使用するデータセットの作成を行った。

##### 使用する演奏制御パラメータ

使用するリコーダー演奏ロボットは、電圧値によって流量弁の開度を調節し、演奏中の息の量を変化させることができる。そのため、演奏制御パラメータとして、流量弁の制御電圧値を用いた。

##### 使用する音高

リコーダーロボットに使用されているソプラノリコーダーは、主に小学校教育などで使用されているものである。そのため、実際に使用する場面として小学校が挙げられる。

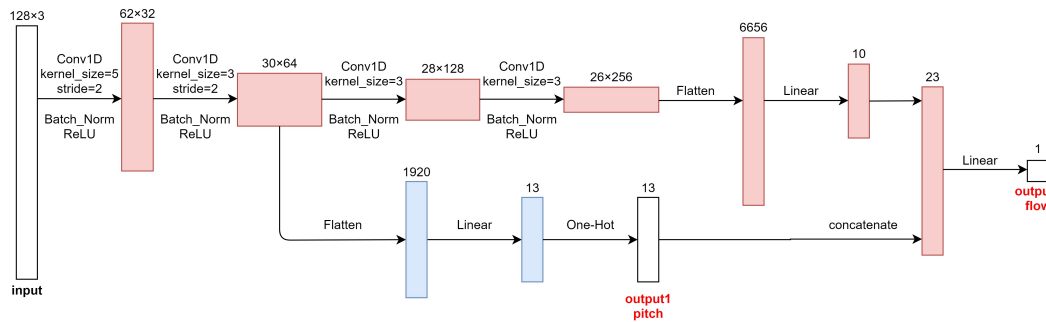


図 3 モデルの構成. 演奏されている音高とその時の制御パラメータを同時推定することができる. 入力には 1 次元のスペクトル画像が使用されており, 各層は 1 次元の CNN を中心に構成されている.

よって, 今回作成するデータセットに使用する音高は, 小学校教育で使用される音域 C5~A6 を使用した.

#### 録音環境

リコーダー演奏ロボットの録音は, 防音室内で自動で行った. 録音には TASCAM MC-VT1 と JTS MA-500 の 2 本のマイクを使用した

#### 使用する流量

以下の表 1 に, 音高ごとの制御電圧値と総データ数を示す. 1 データは 0.5 秒間の録音である. 音高ごとに電圧値の幅やデータ数が異なるのは, 音高に寄って音を鳴らすことのできる流量が異なるためである.

表 1 音高ごとの制御値とデータ数

音高	最小電圧値 [mV]	最大電圧値 [mV]	データ数
C5	1683	1762	2600
D5	1703	1802	3200
E5	1703	1828	4000
F5	1703	1848	4600
G5	1729	1848	3800
A5	1703	1868	5200
B5	1716	1848	4200
C6	1716	1881	5200
D6	1749	1881	4200
E6	1769	1861	3000
F6	1769	1894	4000
G6	1782	1901	3800
A6	1802	1914	3600

#### 4.3 モデル構造

パラメータ推定モデルとして, 一次元の CNN ベースのモデルを作成した. 以下の図 3 にモデルの構成を示す. 今回作成したモデルは, 13 クラスの分類による音高推定と, 回帰による制御電圧値推定を同時に行う. 音高推定は単純な CNN で構成されている. 流量推定の全結合層には, 推定した音高から作成された one-hot 配列を連結しており, これによって音高ごとのバイアスをかけている. 入力は 128 次元 3 チャンネルのものを使用する. 入力の各チャンネルにはそれぞれデータセットの音声データから作成した

スペクトル画像を用いており, メルフィルタ数 128 の対数メルスペクトル, メルフィルタ数 256 のメルスペクトルを半分に分割したものを使用している.

#### 4.4 モデルの学習

作成したデータセットを用いて, モデルの学習を行った. 学習は 2 段階に分けて行われる. 初めに, 制御電圧値を正解ラベルとして, 図 3 の赤色の部分の重みを更新し, 学習を行う. その後, 音高を正解ラベルとして, 初めに学習した赤色の部分の重みを固定したまま, 青色の部分の重みのみを更新し, 学習を行う. どちらの学習もバッチサイズ 256, epoch 数 50 で行った. 損失関数は, 回帰には平均二乗誤差 (MSE) を使用し, クラス分類には交差エントロピー誤差を使用した. また, 最適化関数には Adam を使用した. データセットは学習用データと評価用データに 4:1 の比率で分けて使用した.

制御電圧学習の結果を図 4 に, 音高学習の結果を図 5 に示す. 流量推定, 音高推定ともに学習が正しく進んでいる. また, 学習済みモデルの評価データに対する音高推定精度は 1, 電圧値の絶対平均誤差 (MAE) は 0.812 mV となり, 高い精度で推定できていることが確認できた.

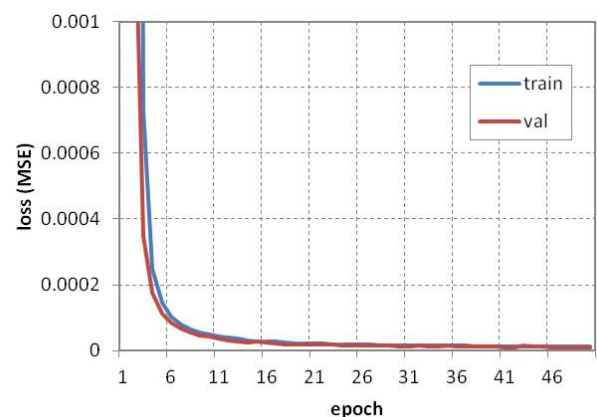


図 4 流量制御値の学習による loss の変化

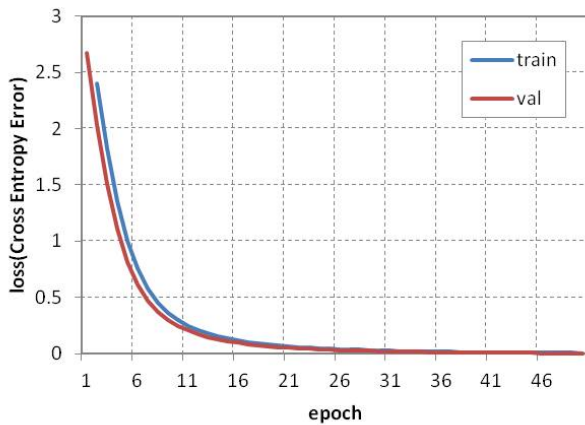


図 5 音高推定の学習による loss の変化

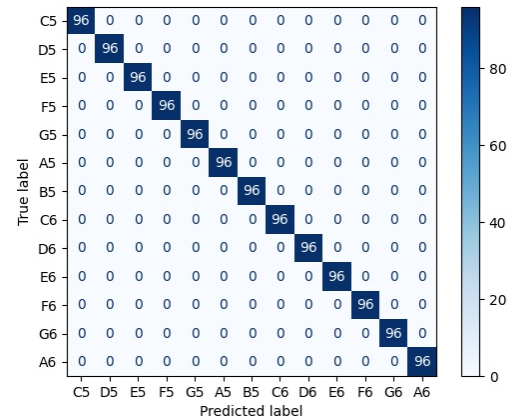


図 6 音高推定の混同行列

## 5. 人間の演奏に対するモデルの出力

学習済みのモデルが人間の演奏に応用できるかを確認するために、人間の演奏を入力したときの音高推定精度と、演奏時の流量と制御値の推定結果との相関を観察した。

モデルに入力する人間の演奏データとして、著者によるリコーダー演奏の録音を行った。録音に使用したリコーダーとマイクは、リコーダーロボットのデータセットを作成したときと同じものを使用した。演奏時の息の流量に差を作るために、フォルテ、メゾフォルテ、ピアノの三段階の音量を録音した。それぞれの音高、音量に対して 60BPM で 8 拍の録音を行い、0.5 秒間で分割して 1 データとした。そのため、今回作成したデータ数は合計で 1248 データとなった。

人間の演奏の流量と、モデルの出力する制御値の相関を調べるために、本研究では相関比を用いた。相関比とは、量的変数と名義尺度感の相関の程度を表す統計量である。 $a$  種類のカテゴリの中に  $n_a$  個の名義尺度変数  $x$  があるとき、変数全体の平均  $\bar{x}$  をとすると、相関比  $\eta$  の 2 乗は下式で表される。

$$\eta^2 = \frac{\sum_{i=1}^a n_i (\bar{x}_i - \bar{x})^2}{\sum_{i=1}^a \sum_{j=1}^{n_i} (x_{ij} - \bar{x})^2} \quad (1)$$

人間のデータを入力した際の音高推定結果の混同行列を図 6 に、B5 における制御電圧値の推定結果のストリップチャートを図 7 に示す。また、人間の演奏の流量と推定された制御値の相関比を表 2 にそれぞれ示す。図 6 より、今回学習したモデルは、人間の演奏に対して、正しく音高を推定することができた。また、制御電圧値に関して、相関比が 0.85~0.98 程度であるため、流量制御電圧と人間の演奏の流量との間には高い相関があることが確認できた。

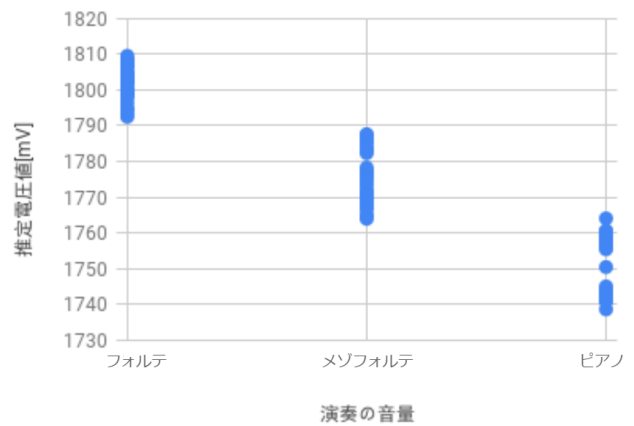


図 7 B5 における流量制御値の推定結果

表 2 音高ごとの相関比

音高	相関比	音高	相関比
C5	0.856	C6	0.873
D5	0.939	D6	0.986
E5	0.878	E6	0.881
F5	0.964	F6	0.908
G5	0.885	G6	0.932
A5	0.944	A6	0.973
B5	0.903		

## 6. 結論

本研究では、リコーダー演奏ロボットの演奏音高及び電圧値を推定する機械学習モデルを構築するために、データセットを作成し、モデルの学習を行った。学習したモデルは、音高、電圧値ともに高い精度で推定することができた。また、学習したモデルに人間の演奏データを入力した際の出力を観察した。その結果、音高を正しく推定することができ、推定電圧値と人間による演奏の流量との相関比

も 0.85 以上と高い相関があることを確認できた。

今後の展望として、リコーダーやマイクなどが異なる状況への対応を行うことが挙げられる。また、実用に向けてスマートフォンアプリや web アプリとしての実装も行っていきたい。

#### 参考文献

- [1] Picas, O. R., Rodriguez, H. P., Dabiri, D., Tokuda, H., Hariya, W., Oishi, K. and Serra, X.: A real-time system for measuring sound goodness in instrumental sounds, *Journal of The Audio Engineering Society* (2015).
- [2] 洗司宮里, 聖也大城, 健太郎野口, 志穂子神里: K-030 慣性センサによるバイオリンの運弓動作指導の検討 (教育工学 (4), K 分野: 教育工学・福祉工学・マルチメディア応用), 情報科学技術フォーラム講演論文集, Vol. 10, No. 3, pp. 783–784 (オンライン), 入手先 (<https://ci.nii.ac.jp/naid/110009623006/>) (2011).
- [3] : 早稲田大学 高西研究室 WAS-5, [http://www.takanishi.mech.waseda.ac.jp/top/research/music/saxophone/was\\_5/](http://www.takanishi.mech.waseda.ac.jp/top/research/music/saxophone/was_5/).
- [4] Kodama, K. and Koutaki, G.: Development of guitar playing robot by PWM control of solenoid, *2019 IEEE 8th Global Conference on Consumer Electronics (GCCE)*, pp. 291–293 (2019).