

HMM を用いた指揮者による表現指示動作の識別

古川絵梨^{*1}, 泉正夫^{*2}, 真嶋由貴恵^{*3}

^{*1}大阪府立大学工学域

^{*2}大阪府立大学大学院 工学研究科

^{*3}大阪府立大学大学院 人間社会システム科学研究科

Identification of Action to Indicate Expression by the Conductor using HMM

Furukawa Eri^{*1}, Izumi Masao^{*2}, Majima Yukie^{*3}

^{*1}Faculty of Engineering, Osaka Prefecture University

^{*2}Graduate School of Engineering, Osaka Prefecture University

^{*3}Graduate School of Humanities and Sustainable System Science,
Osaka Prefecture University

現在, 市民オーケストラや市民合唱団など音楽を楽しむ多くの人たちが幅広く活動している. しかし, 初学者は指揮者の表現の指示が即座に理解することが困難であったり, また市民楽団の指揮者は経験者でない場合もあり, 指揮初心者の多くは指揮における指示動作を独学で習得することが難しい. そこで本研究では, 実践的な指揮の学習を行うことを支援するために, 指揮者の動作から表現指示動作を自動的に抽出し, 指揮中の表現を識別するシステムを提案する. 提案手法では, 直接機器を身体に装着する必要がないため, 身体への負担がない. また, より自然な指揮を扱うことができ, 実際の指揮での識別を可能にする. 実験では, 指揮映像から指揮表現を識別し, 手法の有効性を確認した.

キーワード: 指揮動作, 動作識別, 自己学習サポート, 隠れマルコフモデル (HMM), Kinect

1. はじめに

現在, 市民オーケストラや市民合唱団など音楽を楽しむ多くの人たちが幅広く活動している. しかし, 初学者は指揮者の表現の指示が即座に理解することが困難であったり, また市民楽団の指揮者は経験者でない場合もあり, 指揮初心者の多くは指揮における指示動作を独学で習得することが難しい. そこで本研究では, 実践的な指揮の学習を行うことを支援するために, 指揮者の動作から表現指示動作を自動的に抽出し, 指揮中の表現を識別するシステムを提案する.

提案手法では, 直接機器を身体に装着する必要がないため, 身体への負担がない. また, より自然な指揮を扱うことができ, 実際の指揮での識別を可能にする.

実験では, 指揮映像から指揮表現を識別し, 手法の有効性を確認した.

2. 提案手法

2.1 提案手法の概要

Kinect センサを用いて指揮者の動作から表現指示動作を識別する手法を提案する. 提案手法の概要を下図に示す.

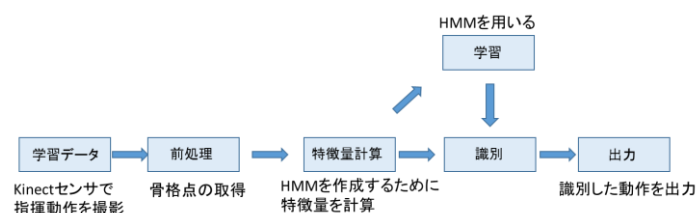


図1 提案手法の概要

2.2 学習データ

本研究で扱う識別データは, 学習者が演奏を行う上で必ず理解しなければならない指揮中で頻繁に扱われる表現かつ, 多くの指揮者が類似した動作を行う 4 種類

の表現指示動作と表現を含まない通常の指揮動作の計 5 種類を扱う。

- 1) スタッカートを示す動作
手を開いた状態で、手腕を振り下ろす動作を繰り返すものである。
- 2) 音の維持を示す動作
手のひらを真上に向けた状態で、手腕を体の前方に伸ばし、同じ高さで保つ動作。
- 3) 音の停止を示す動作
開いた手を円状に動かしながら徐々に閉じていく動作。
- 4) 音の大きさを示す動作
手を軽く握ったような形にし、体の前方に伸ばした位置から真っ直ぐ戻す動作。
- 5) 表現が含まれていない通常指揮動作

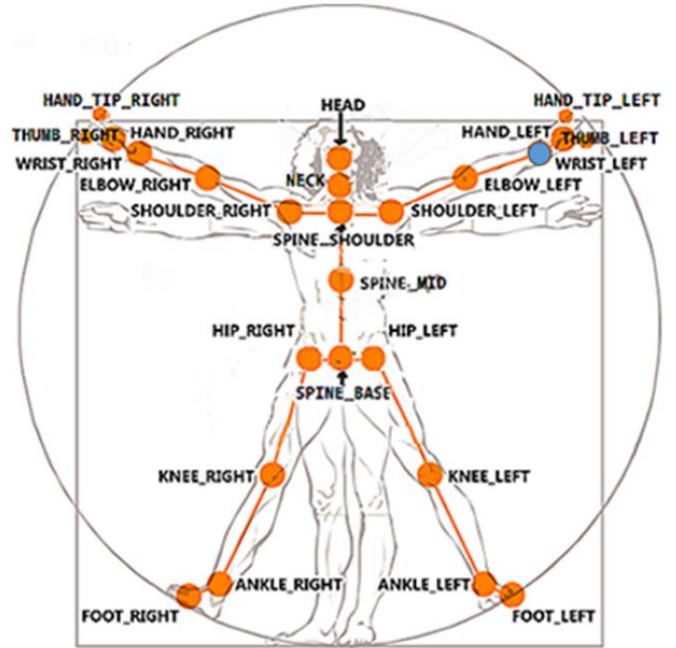


図 2 Kinect で取得することができる骨格点

2.3 骨格点の取得

識別に使用する座標情報を Kinect によって得られる骨格点から取得する。Kinect for Windows SDK 2.0(以下 SDK)⁽¹⁾とは、2014 年 7 月に Microsoft が公開したソフトウェア開発キットである。SDK では Kinect に搭載された各種センサから得られるデータを取得可能であり、RGB 画像、深度画像、骨格線情報を取得することができる。ここで骨格線情報は深度情報から Kinect 独自のアルゴリズムにより算出される。また、骨格線中の空間座標も取得可能である。三次元座標が取得できる人体の骨格点 (25 点) の位置を図 2 に示す。

2.4 特徴量

特徴量には Kinect センサから取得される骨格点の三次元座標を用いる。指揮における表現指示動作とは、「主導的ではない方の手による身振りを表現に関する身振り」(Braem&Bräm, 2000)⁽²⁾としている。そこで本研究では、左手で表現指示動作を行うと仮定した。また、指揮動作を識別する上で、手の動きを取得する必要がある。しかし、指や手のひらの座標は安定的に取得できないことがある。そこで本研究では、左手首の三次元座標を特徴量として用いる。図 2 における青い点が今回使用した骨格点の位置である。

2.5 学習

本研究では、隠れマルコフモデル(以下 HMM)によって動作の識別を行う。HMM とは、確率モデルの一つであり、観測されない状態をもつマルコフ過程である。時系列パターンに対して統計的、確率的なパターン認識の手法として一般的に用いられている。実験では、Python で実装された hmmlearn⁽³⁾を用いた。

表現指示動作 4 種類に対し、各 1 つずつ HMM を作成し、表現指示動作の含まれていない通常の指揮動作に対し、1 つの HMM を作成した。以上 5 つの HMM を用いて識別を行った。

HMM を作成する際には、指揮動作の映像を目視で 30 フレームずつに分割し、学習を行った

3. 実験

3.1 実験環境

図 3 にデータ収集のための実験環境を示す。取得した座標の位置が大きくずれることを防ぐため、Kinect から被験者までの距離は、常に全身の 25 点の骨格点が計測できるように設定した。この位置から動くことなく指揮を行い、撮影した。

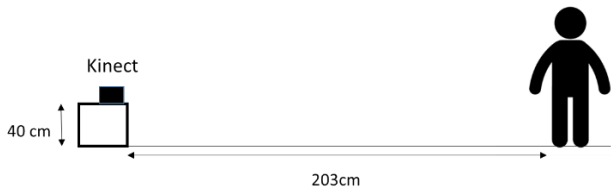


図 3 撮影環境

3.2 データ収集

表現指示動作の登録用データとして 1 名のデータを収集した。データの種類は 2 種類収集した。

1 つ目は 4 つの表現指示動作のうち 1 つの表現指示動作が含まれた指揮を 10 セットずつ、計 40 セット行った。1 セットには、表現指示動作が 2 回分と通常指揮動作が含まれている。

2 つ目は 4 つの表現指示動作が一回ずつと通常指揮動作が含まれた指揮を 5 セット行った。

3.3 評価方法

本研究では、指揮中で表現指示動作が行われているタイミングの検出を行うことと、識別することを目指している。そこで、各表現指示動作の識別率と指揮中での表現指示動作が行われているタイミングの検出率で評価する。

各表現指示動作の識別率の評価は一個抜き交差検証により行う。この評価法では、標本となるデータから 1 つを抜き出しテストデータとし、残りのすべてのデータで学習を行う。データの解析及び導出されたモデルがどれだけ母集団に有効であるか良い近似で検証できる。実験では少数のデータでも精度を評価できるよう交差検証を用いた。今回は撮影したデータセットのうち指揮動作 20 回分のデータで一個抜き交差検証を行った。本実験での一個抜き交差検証は、以下の手順で行った。

3.3.1 一個抜き交差検証の手順

- 1) 撮影した 20 データのうち、1 つのデータをテスト用データとし、残りの 19 データを用いて学習する。
- 2) 1) を 20 回繰り返す。
- 3) 20 回の識別結果の平均を全体の結果とする。

指揮中での表現指示動作が行われているタイミングの検出率の評価は、以下の手順で行った。

3.3.2 表現指示動作が行われているタイミング検出率評価手順

- 1) 1 つの表現指示動作が含まれた指揮 40 セットすべてを用いて学習を行い、4 つの表現指示動作が含まれた指揮 5 セットのうち 1 セットをテスト用データとして識別を行う。
- 2) 1) を 5 回繰り返す。
- 3) 5 回の識別結果をテスト用データと目視で確認し、検出率とする。

4. 結果

表 1 に各表現指示動作の識別結果を示す。今回は通常指揮動作についての識別は行わず、表現指示動作の識別のみを行った。

表 1 各表現指示動作の識別結果

		正解				
		1)	2)	3)	4)	5)
識別結果	1)	15	0	3	2	0
	2)	0	18	1	0	1
	3)	0	1	19	0	0
	4)	1	0	3	16	0

1) スタッカートを示す動作、2) 音の維持を示す動作、3) 音の停止を示す動作、4) 音の大きさを指示する動作、5) 表現が含まれない通常指揮動作を示している。

以上の結果より、スタッカートを示す動作は 75%、音の維持を示す動作は 90%、音の停止を示す動作は 95%、音の大きさを指示する動作は 80% という結果となった。

表 2 に表現指示動作が行われているタイミングの検出結果を示す。

表 2 指揮中の表現指示動作のタイミング検出結果

	表現指示動作の種類			
	1)	2)	3)	4)
検出数	5	5	4	3

1) スタッカートを示す動作、2) 音の維持を示す動作、3) 音の停止を示す動作、4) 音の大きさを指示する動作を示している。

以上の結果より、スタッカートを示す動作と音の維持を示す動作の検出率は 100%、音の停止を示す動作は 80%、音の大きさを指示する動作は 60% という結果となった。

5. 考察

表現指示動作の識別率に関して、音の維持を指示する動作と音の停止を指示する動作は比較的高い識別率であるといえる。スタッカートを指示する動作については、動作が速いため、データが連続で取ることができていない可能性が考えられる。音の大きさを指示する動作については、左手首の前後の動きに重点をおいた HMM を作成したため、深度情報のずれが結果に大きく関係したのではないかと考えられる。

表現指示動作のタイミング検出に関しては、動作の長さが違うため、検出できなかったものがあると考えられる。

Kinect で撮影することによって、身体に装置を付ける必要がないため、自然な指揮を行うことができたので、より実践的な指揮学習に用いることができると思われる。

6. おわりに

6.1 まとめ

HMM を用いて指揮者の動作から表現指示動作を識別する手法を提案した。特徴量には Kinect センサから得られる骨格点の三次元座標から、左手首の三次元座標を用いた。また、学習には HMM を用いた。提案手法を実証するために、実験を行い、有効性を確認した。

6.2 今後の課題

演奏学習者に対しては、指揮者が行っている指揮動作が何を示すのかをリアルタイムで伝えることができるシステムの構築を目標としている。また、指揮学習者に対しては、指揮熟達者の表現指示動作の振り方や、タイミングを解析したもので自己学習できるシステムの構築、そして自分自身の動作が正しく演奏者に伝えられるものであるかを確認できるシステムの構築を目標としている。そのため、まず識別率を向上させるために撮影データを増やすこと、それぞれの表現指示動作の長さが異なる場合に対して頑強なシステムとすること、また、汎用的なシステムとするために被験者を増やすことも課題として挙げられる。

謝辞

本研究の一部は JSPS 科研費 JP25330412, JP26293451 の助成を受けたものです。

参考文献

- (1) Developing with Kinect for Windows.
<https://developer.microsoft.com/en-us/windows/kinect/develop> (2017年2月3日確認)
- (2) Bream, P. B. & Bräm, T: A Pilot Study of the Expressive Gestures Used by Classical Orchestral Conductors. In Emmorey, K. & Lane, H. (Eds.), *The Signs of Language Revisited: An Anthology to Honor Ursula Bellugi and Edward Klima*, 143-167. Lawrence Erlbaum Associates.(2000)
- (3) Hmmlern0.2.0:PythonPackageIndex.
<https://pypi.python.org/pypi/hmmlern>(2017年2月3日確認)