

令和 4 年 6 月 17 日現在

機関番号：33803

研究種目：基盤研究(C) (一般)

研究期間：2019～2021

課題番号：19K04300

研究課題名(和文)人とロボットの高度な協調作業に関する研究：純正律協奏システムの工学的実現

研究課題名(英文) Research on advanced collaborative work of human and robot : Engineering realization of the pure temperament concerto system

研究代表者

飛田 和輝 (Tobita, Kazuteru)

静岡理科大学・理工学部・教授

研究者番号：70739094

交付決定額(研究期間全体)：(直接経費) 3,300,000円

研究成果の概要(和文)：高次元の協調作業として、人とロボットの協調演奏に取り組んだ。期間中の研究範囲は、奏者の演奏タイミングに合わせ和音構成音を発音するシステムの実現である。
RGB-Dカメラの画像から奏者の動作をトラッキングし、時系列データを特徴量としたニューラルネットワークにより、演奏開始点の検出を±0.5秒、検出率85%で実現した。また、回転円盤弓による発音装置を設計、弦の振動モデルを考慮したPID制御を実装し1Hzの分解能で演奏可能とした。
これらを統合し、奏者の演奏タイミングに合わせ、和音構成音を純正律で奏でることを実現した。今後、検出速度の向上、ロバスト化、発音装置の応答性向上などの課題に取り組む予定である。

研究成果の学術的意義や社会的意義

少子高齢化に応じて製造現場での人とロボットの協調作業に関する研究、開発、一部実用化が進んでおり、今後さらなる高度な協調作業に発展すると考えられる。高い次元での協調作業として複数演奏者のアンサンブルに着目し、人とロボットの協調による演奏を実現することで、幅広く展開が可能であると考えている。本研究の成果は、産業界において、ティーチング作業の効率化や、遠隔での設備立ち上げなど様々な分野での応用に発展が期待される。

研究成果の概要(英文)：As a high-dimensional collaborative work, we worked on the collaborative performance of humans and robots. The scope of research during the period is the realization of a system that produces chord constituent notes according to the player's performance timing. The movement of the player was tracked from the image of the RGB-D camera, and the detection of the performance start point was realized in ± 0.5 seconds and the detection rate was 85% by the neural network using the time series data as the feature quantity. In addition, we designed a sounding device with a rotating disk bow, implemented PID control considering the vibration model of the strings, and made it possible to play with a resolution of 1 Hz. By integrating these, it was possible to play chord constituent notes in just intonation according to the player's playing timing. In the future, we will solve problems such as improvement of detection speed, robustness, and improvement of responsiveness of the sounding device.

研究分野：知能機械システム, 計測工学

キーワード：協調作業 純正律 演奏ロボット ニューラルネットワーク 骨格検出

様式 C-19、F-19-1、Z-19 (共通)

1. 研究開始当初の背景

少子高齢化に応じて製造現場での人とロボットの協調作業に関する研究、開発、一部実用化が進んできており、今後さらなる高度な協調作業に発展すると考えられる。われわれは高い次元での協調作業として複数演奏者のアンサンブルに着目しており、人とロボットの協調による演奏(協奏)を実現することで、幅広く展開が可能であると考えている。

近年、AIによる人とのアンサンブルに関する研究が進められている。過去の巨匠の演奏を自動演奏ピアノで再現し、時代を超えた幻の共演の実現や、楽器初心者、初級者の演奏の成功体験提供などの研究例がある。

従来研究では、人とロボットの共演の際に、タイミング、テンポ合わせる配慮されているが、音楽のそれ以外の要素についてはいまだ研究対象に取り上げられていないと見受けられる。われわれは、音楽にはハーモニーと音のバランスが不可欠であるという点に着目している。

ピアノは通常平均律で調律されており、1オクターブを便宜上等分している都合上、和音を演奏した際には常に濁った音、つまりうねりが発生する。一方、管楽器、弦楽器などのピッチを動的に調整できる楽器によるアンサンブルでは、一部の現代音楽を除き、純正律による調和が重視され、記譜上は同じ音であっても、和音構成における役割、すなわち基音が3音か、5度かによってピッチを上下に調整することでハーモニーを構成し、うねりをなくし、共鳴により高次の倍音を発生させ、感動を聴衆に与える。

ハーモニーには音量の制御も欠かせない。基音や5度は十分な音量が必要だが、三音は控えめの音量とすることでバランスが取れる。さらに、オーケストレーションによっては、それぞれの構成音を発音する楽器の数が異なるので、同時発音数に応じて音量を調節することも欠かせない。また、音量については、クレッシェンドの仕方、楽譜に書かれていないフレーズの歌い方についても、先行する奏者の表現方法に、後から演奏する奏者は合わせる必要がある。また、主旋律を担当するなら大きく、しかし明らかな伴奏なら落とす。如何にこれらを人とロボットの協奏で実現するかを、本研究の出発点としている。

2. 研究の目的

本研究の最終的な目標は、あらかじめ入力された楽譜情報をもとに、リアルタイムに人の演奏者に対して純正律のハーモニーを演奏するシステムを実現することにある。これまでの楽器演奏ロボットや自動演奏システムでは考慮されることのなかった、純正律のハーモニーを実現するところに独自性、新規性がある。本研究の該当期間における研究の範囲は、リコーダー演奏者が発音するタイミングに合わせて、和音構成音を発生するシステムの実現にある。

3. 研究の方法

(1) 奏者動作検出システムの研究開発

演奏家の演奏開始タイミング(アインザッツ)を検出するシステムである。開発ステップは以下の通り進めた。

1stステップ: カメラ, 3D TOF カメラ, 加速度センサ, ジャイロ, マイク等のセンサを用いた検出システムの構成と、演奏者のデータ取得。

2ndステップ: 演奏データによる学習結果からタイミング予測信号を抽出。

(2) リアルタイム音量ピッチ制御システムの研究開発

発音すべき音は与えられているものとして、基準音が発生すると、目標の周波数で発音を開始し、基準音と発音した音との関係から、ピッチと音量を制御するシステムである。

1stステップ: 張力を制御可能とした弦と、押し下げ力を制御する押し下げ機構、円盤を回転させる回転弓を用いた発音装置を構築。さらに基準音を入力するマイク、発音装置の音を入力するマイクなどからなる、リアルタイム音量ピッチ制御システムを構成。

2ndステップ: 基準音と同じ音を発音し、張力制御により音のうねりを無くす制御アルゴリズムを構築

3rdステップ: 基準音に対する和音構成音を発音し、張力制御による和音の制御。

(3) システム統合

(1)(2)の開発結果を統合し、演奏者の基準音に合わせて和音構成音を発音可能とする。

4. 研究成果

(1) 奏者動作検出システムの研究開発

奏者動作検出システムのシステムブック図を図1に示す。リコーダーに取り付けたマイクをミキサ、ADコンバータを介して取り込みPC上でFFTによる基音の分析を行う。また演奏開始前の合図となるアインザッツの動作をカメラを介して取り込み、動作のトラッキングをPC上で解析し、機械学習により打点の学習、推定を行う。

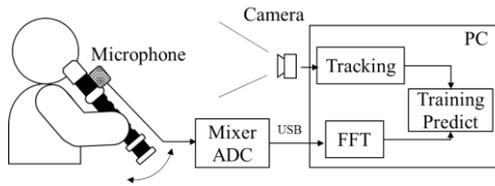
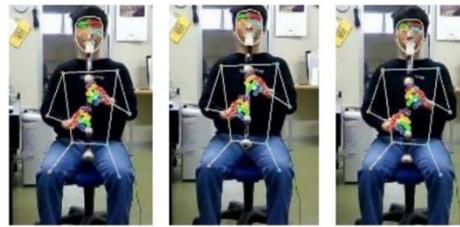
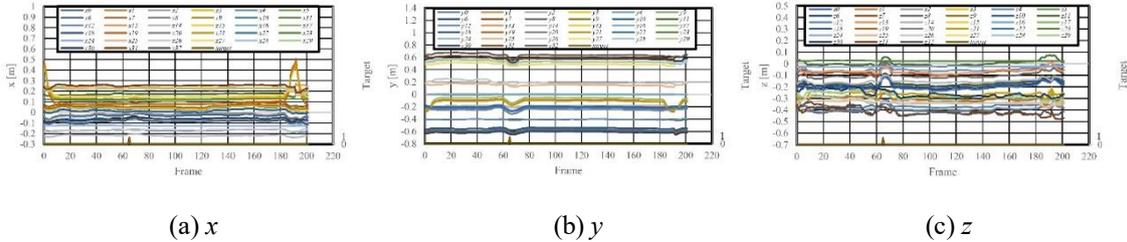


図1 奏者動作検出システムブロック図



0フレーム 5フレーム 10フレーム

図2 骨格検出の例



(a) x

(b) y

(c) z

図3 骨格座標時間推移と打点の一例

種々のトラッカ (Cascade, MIL, KCF, Posenet, MediaPipe) について基礎実験を行った結果, フレームレートと追従安定性が良好で, かつトラックポイント数が 33 点と多い MediaPipe を採用した. MediaPipe により骨格検出した例を図 2 に示す. また, 時々刻々の関節の座標の変化の一例を図 3 に示す. x 軸はカメラ画像水平方向, y 軸はカメラ画像垂直方向, z 軸はカメラの奥行方向をそれぞれ意味する. 33 点の特徴点と発音タイミングを target として示している. 65 フレーム付近がザッツによる演奏開始指示点を意味しており, その地点のみ, target を 1 としている. ザッツにより, x 座標はあまり変化しないが, y 座標, z 座標は変化がみられる.

アインザッツで身体や楽器を揺らす大きさや速度は, 奏者の癖や楽器の種類によっても異なることから, 一律の閾値を設けてそれに合致するもののみを判別するような手法は好ましくない. そこで, 機械学習を用いて奏者のアインザッツの出し方を学習することで, 柔軟かつ高精度な判別を行えるよう検討を行った.

ザッツ判別原理を図 4 に示す. 使用する機械学習の種類については, ザッツであるかないかを判別 (分類) することから, 分類器型を採用した. また, リアルタイムで遅延なく学習予測できる処理速度も重要である. 映像単体では 30 fps 以上が望ましいが, 映像解析と動作判別の両機能の組合せ時における処理速度は 10 fps 程度を想定する.

本研究では, scikit-learn の MLP Classifier (多層パーセプトロン分類器) を用いたアインザッツ判別を行うこととした.

ターゲットは演奏開始タイミング (無音状態から有音状態に移行した瞬間) であり, これをラベル 1, それ以外をラベル 0 として学習させる.

入力層に特徴量を入力すると, 隠れ層で多数のニューロンによって処理され, 出力層にて予測判別結果が生成される (ザッツが判別されたら 1, ザッツではないと判別された場合は 0 が出力される) 構成である.

グリッドサーチにより, パラメータを選定する.

MLP 分類器のパラメータとして, 最適化手法 (adam, lbfgs, sgd) および活性化関数 (identity, logistic, relu, tanh) を選定する.

また, 分類器に学習させる特徴量として, 奏者をトラッキングして得られた三次元座標の履歴から使用するフレーム数 (10, 20, 30 フレーム), 特徴量を相対座標化 (最初のフレームにおける座標値を原点とみなす処理), 正規化, 標準化などの前処理の有無についてグリッドサーチを実施する.

なお, 特徴量の種類は MediaPipe Pose の pose_world_landmarks から得られる奏者の全身の三次元座標値 (腰の midpoint を原点) を使い, ターゲットは奏者が楽曲演奏を開始した (初めて演奏音が入力された) 瞬間をラベル 1, それ以外をラベル 0 としたものを用いる.

ザッツの判別については, ターゲットのフレーム前後 10 フレーム (± 0.5 s に相当) 以内を判別に成功したと定義する.

判別成功回数の多い組合せの上位 10 位までを表 1 に示す.

FN は特徴量となる三次元座標取得値を使用するフレーム数 (履歴) を, 前処理については RC は相対座標化, NO は正規化, ST は標準化を示す. また, activation は活性化関数, solver は最適化手法を, それぞれ示す. 7 回のザッツのうち, 判別できた回数を OK に, ザッツではない動作をザッツだと誤判別した回数を NG に示す.

この結果から, フレーム数 10, 前処理が相対座標化のみ, 活性化関数が relu, 最適化手法が lbfgs の場合に判別成績が良好であった.

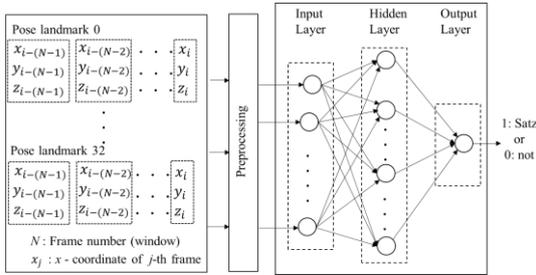
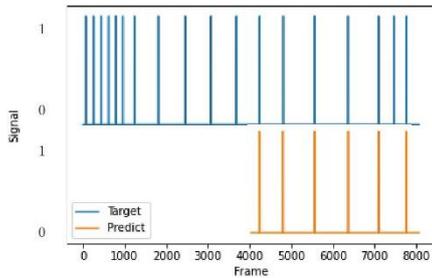


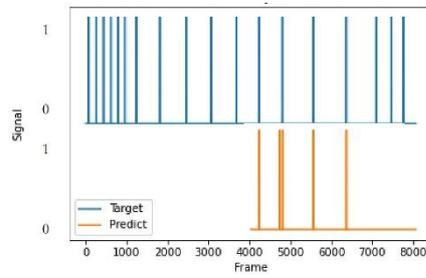
図4 ザッツ判別原理

表1 グリッドサーチランキン

Rank	FN	Preprocessing	Activation	Solver	OK	NG
1	10	-/RC	relu	lbfgs	6	0
2	20	-/-	tanh	lbfgs	4	0
3	30	-/-	relu	lbfgs	4	0
4	10	-/-	identity	lbfgs	3	0
5	10	-/RC	identity	adam	4	1
6	10	-/RC	relu	adam	3	0
7	20	-/-	relu	lbfgs	3	0
8	20	-/RC	relu	adam	3	0
9	20	-/RC	relu	lbfgs	3	0
10	10	NO/RC	tanh	lbfgs	2	0



(a)ランキン 1



(b)ランキン 5

図5 学習による打点推定結果一例

ランキン 1 位におけるターゲットと予測値との推移を示したグラフを図 5(a) に示す。図中、前半 11 回が学習用のデータであり、4,000 フレーム以降の後半の 7 回が予想用のデータである。ターゲットが 1 になる回数、つまりザッツの回数は全 7 回あるが、正確に判別できたのは 6 回であり、判別確率は約 85% と言える。

また、ランキン 5 位の場合のターゲットと予測値との推移を示したグラフを図 5(b) に示す。こちらはザッツ判別成功数は 4 回であるが、ザッツではない部分を 1 か所誤判別している様子が見受けられる。

(2) リアルタイム音量ピッチ制御システムの研究開発

リアルタイム音量ピッチ制御システムの外観を図 6 に、演奏時のピッチ f [Hz] を制御するための張力調整機構の構成を図 7 に示す。目標ピッチ f_{ref} [Hz] に対する張力制御用モータの理論トルク T_{th} [Nm] は次式で求められる。

$$T_{th} = 2\mu D l^2 f_{ref}^2 \quad (1)$$

なお、 μ [kg/m] は弦の線密度である。ここで、摩擦等の外乱の影響を補正するため、図 8 に示すような PID コントローラによってトルク目標値 T_{ref} [Nm] を補正する仕組みを導入した。ピッチ制御実験として、C4、E4、G4 についてそれぞれ 3 回実施した際の結果を図 9 に示す。

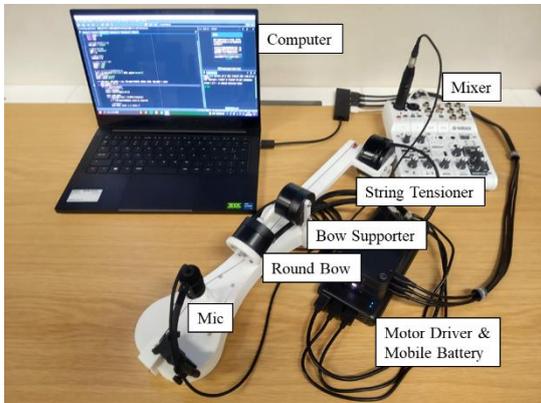


図6 バイオリン型の擦弦楽器および演奏用の各種機構の外観

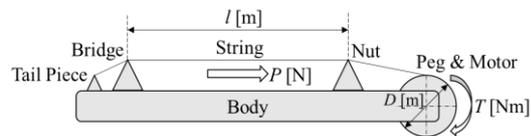


図7 張力調整機構の構成

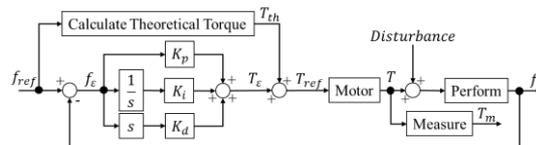


図8 ピッチ制御用 PID コントローラ

C4, E4 については、収束後の 6 秒以降のピッチの標準偏差は約 1.5Hz であった。G4 では標準偏差が約 10Hz にまで悪化することがあったが、張力制御モータの最大トルク 0.5 Nm に近づいているため、安定した制御が行いにくいためと考えられる。なお、収束に時間を要するがこれは今後の課題とする。

(3) システム統合

(1) (2) のシステムを統合し純正律協奏実験として、楽器奏者の演奏音の周波数 f_p [Hz] に対して、3 度の和音構成音となる周波数 ($f_p \times 1.25$) を目標値 f_{ref} [Hz] としてリアルタイム音量ピッチ制御システムに追従演奏させ、その周波数 f [Hz] が目標値に収束しているかを評価した。

奏者が連続して C4, D4, E4, 休符, E4, D4, C4 と演奏したときの結果を図 10 に示す。C4, D4 の三度の E4, F#4 については約 ± 5 Hz に収束しており、奏者に追従して和音を演奏する機能の実装が確認できた。なお、E4 の三度の G#4 が安定しないが、これは G4 の単音発音実験結果同様、張力制御用モータのトルク上限付近となるためであると考えられる。

人と機械とのより高度な協調作業を実現するための基礎研究として、楽器奏者の動作を学習して伴奏する純正律協奏システムの研究開発を行い、両者の演奏開始タイミングを合致させ、また、ピッチを和音構成音に自動調節させることで、人と機械とが協調して合奏するための基礎的な研究成果を得た。

5. 主な発表論文等

[学会発表] (計 4 件)

国際学会 (2 件) (すべて査読あり)

- ① K. Tobita, R. Ishikawa, K. Mima, Study on human-robot ensemble : Experiments on Performance Timing Detection and Pure Chord Playing, 2022 JSME-IIP/ASME-ISPS Joint Int. Conf. on Micromechatronics for Information and Precision Equipment (MIPE), August, 2022, Nagoya, Japan [Accepted]

- ② R. Ishikawa, K. Tobita, Concept and Prototype of the Pure Intonation Ensemble System , The 11th TSME International Conference on Mechanical Engineering, Ubon Ratchathani, Thailand, DRC001 (2020.12).

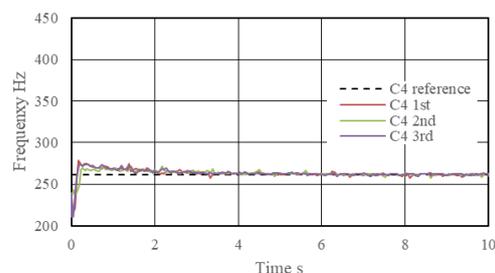
国内学会 (2 件) (査読なし)

- ① 飛田和輝, 石川隆介, 美馬一博, 人と機械の協奏に関する研究-アインザッツ検出手法の考察-, 日本機械学会ロボティクス・メカトロニクス講演会 2022 講演論文集, 1P1-J02, 2022.
- ② 石川隆介, 飛田和輝, 人と機械の協調に関する研究-純正律協奏システムの構想と試作-, 日本機械学会ロボティクス・メカトロニクス講演会 2020 講演論文集, 2A1-D02, 2020.

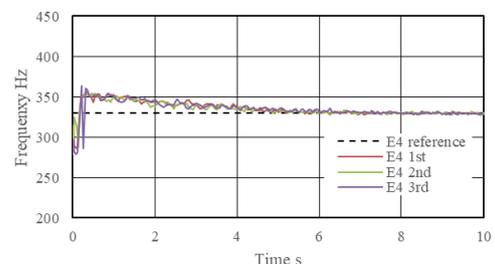
[その他]

○受賞

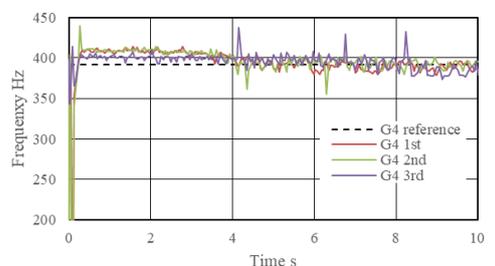
- ① Best Paper Runner-Up Award in The 11th Thai Society of Mechanical Engineers Int. Conf. on Mechanical Engineering 2020.12



(a) C4



(b) E4



(c) G4

図 9 ピッチ制御結果一例

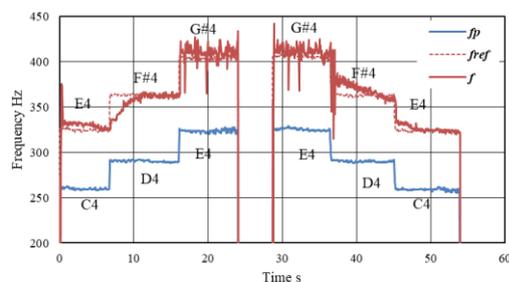


図 10 奏者と演奏システムのピッチの時間推移

5. 主な発表論文等

〔雑誌論文〕 計0件

〔学会発表〕 計4件（うち招待講演 0件 / うち国際学会 2件）

1. 発表者名 Kazuteru Tobita, Ryusuke Ishikawa, Kazuhiro Mima
2. 発表標題 Study on human-robot ensemble : Experiments on Performance Timing Detection and Pure Chord Playing
3. 学会等名 2022 JSME-IIP/ASME-ISPS Joint International Conference on Micromechatronics for Information and Precision Equipment (MIPE) (国際学会)
4. 発表年 2022年

1. 発表者名 飛田和輝, 石川隆介, 美馬一博
2. 発表標題 人と機械の協奏に関する研究-アインザッツ検出手法の考察-
3. 学会等名 日本機械学会ロボティクス・メカトロニクス講演会2022
4. 発表年 2022年

1. 発表者名 Ryusuke Ishikawa, Kazuteru Tobita
2. 発表標題 Concept and Prototype of the Pure Intonation Ensemble System
3. 学会等名 The 11th TSME International Conference on Mechanical Engineering (国際学会)
4. 発表年 2020年

1. 発表者名 石川隆介, 飛田和輝
2. 発表標題 人と機械の協調に関する研究-純正律協奏システムの構想と試作
3. 学会等名 日本機械学会ロボティクス・メカトロニクス講演会2020
4. 発表年 2020年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

静岡理科大学 機械工学科 メカトロニクス研究室
<https://www.sist.ac.jp/lab/me/tobita/>

6. 研究組織

	氏名 (ローマ字氏名) (研究者番号)	所属研究機関・部局・職 (機関番号)	備考
--	---------------------------	-----------------------	----

7. 科研費を使用して開催した国際研究集会

〔国際研究集会〕 計0件

8. 本研究に関連して実施した国際共同研究の実施状況

共同研究相手国	相手方研究機関
---------	---------