

平成 30 年 6 月 11 日現在

機関番号：13903

研究種目：国際共同研究加速基金（国際共同研究強化）

研究期間：2016～2017

課題番号：15KK0008

研究課題名（和文）演奏者の個人性を転写する演奏生成と協調演奏システムの研究（国際共同研究強化）

研究課題名（英文）A research on automatic music performance rendering system and automatic accompaniment system(Fostering Joint International Research)

研究代表者

酒向 慎司（SAKO, SHINJI）

名古屋工業大学・工学（系）研究科（研究院）・准教授

研究者番号：30396791

交付決定額（研究期間全体）：（直接経費） 9,300,000円

渡航期間： 8ヶ月

研究成果の概要（和文）：音響信号による楽譜追跡技術の高度化に取り組み、基本的な音符列の情報だけでなく、打楽器やメロディなど楽譜情報を活用する新たな演奏追跡手法を開発した。RWC音楽データセットを用いたシミュレーション実験により、リアルタイム性を損なうことなく楽譜追跡精度の改善が可能であることを示した。また、演奏動作の指形状変化を取得する画像処理手法として、演奏中のマルチモーダルデータの構築を行うとともに、畳み込みニューラルネットワークによる手形状認識手法において、精密3次元手形状モデルによって疑似的な画像を多数生成することによって学習データセットを拡張し、実写画像に対する認識精度が大きく向上することを確認した。

研究成果の概要（英文）：I worked on improving music score following method using acoustic signals.

We developed a new score following method that utilizes musical score information such as part information of percussion instrument and melodies in addition to basic note sequence that are used in the conventional method. I confirmed that it is possible to improve the accuracy of score following without impairing the real time performance by conducting experiment using the RWC music data set. I also worked on an image processing method to acquire finger shape movements of music performance. I developed the multi-modal data set of music performance. In the hand shape recognition method using the Convolution Neural Network, it has been confirmed that the expanding training data set using precise 3-dimensional hand model could improve the recognition accuracy for real images.

研究分野：音楽情報科学

キーワード：マルチモーダル実演奏データ 楽譜追跡 セグメンタル条件付き確率場 畳み込みニューラルネットワーク 運指推定

### 1. 研究開始当初の背景

これまでに音楽情報処理分野において特に演奏支援に関わる研究を進めてきた。特定の個人性を備えた演奏を機械によって生成する演奏表情生成では、実演奏データから楽譜上のパターン（音符列）と演奏表情の関係を分類し、未知楽譜に対して演奏者の特徴を再現した演奏を生成することができる。人間の演奏に対して音響信号から楽譜上の演奏位置をリアルタイムで追跡する技術では、セグメンタル条件付き確率場(SCRF)と線形動的システム(LDS)を組み合わせることで、位置推定と局所的なテンポ変化をリアルタイムで推定する手法を提案し、これを応用した人間の演奏に合わせて機械が同期した伴奏を再生するシステムなども構築されている。さらに、バイオリンの運指のように形式知として記述が難しい演奏時の身体動作をモデル化することで、初級者のだけでなく様々な技量をもった演奏者に応じた適切な運指を推定する自動運指推定の研究などを進めてきた。これらの研究を元に、次元信号による音響情報だけでなく画像情報や三次元的な位置情報によって表される演奏の身体動作を同時に扱うことで、個々の技術の高度化を目指すとともに新たなマルチモーダル演奏情報処理技術を開拓する。

### 2. 研究の目的

本研究では、複数モダリティ統合による演奏解析技術の高度化および、楽器演奏に関する情報処理技術とヒューマンインタフェース技術との融合システムの探求を目的とする。演奏に対して非接触で計測される音情報と視覚情報などを対象とした演奏データ収集方法について検討し、マルチモーダルな演奏データベース構築方法について検討する。さらにそれらのデータを活用することでより高度な演奏追跡手法、演奏動作認識手法などの新たな技術開発に取り組むため、ドイツミュンヘン工科大学ヒューマンマシンコミュニケーション研究所に滞在し、Gerhard Rigoll 教授のもとで研究を実施した。演奏中の動作認識が必要となる人体の形状や動きを抽出するための要素技術は、動画像中の身体動作のパターン認識の分野では様々な研究がなされていると考えられる。また、近年高度化が進んでいる深度センサを用いた人体動作の計測・認識技術も盛んに研究が進められており、このような研究分野において、Gerhard 教授の研究グループでは先進的な研究に取り組んでいる。共同研究期間では、Gerhard 教授および研究グループに所属するポスドク研究員と連携して個々の研究課題を進める。

### 3. 研究の方法

[3-1] マルチモーダルな実演奏データを記録する際、演奏音の他に演奏中の身体動作を記録する場合には、モーションキャプチャ

などの併用によって正解データを取得することが望ましい。ただし、本格的なモーションキャプチャは高価であることやマーカーが演奏の妨げになる恐れがあることから、より身体的拘束がなく、簡便な方法による身体動作の計測手法を検討した。演奏者の妨げになることが少ない手法としてカメラを用いる手法が考えられ、近年小型化や高精度が進む深度センサを活用する手法について検討する。

[3-2] 演奏追跡技術の高度化のため、演奏者が参照する楽譜情報をより積極的に活用することが望ましい。類似研究を含めた従来手法では基本的な音符の列と演奏とのマッチング問題として捉えられていたが、本研究では音符列以外に楽譜に記載される高次な情報を活用することで演奏追跡の性能を改善する手法について検討する。

[3-3] 実演奏データをもとに音情報と視覚情報による複数モーダル情報を用いることで、画像処理による運指の自動推定技術について取り組む。運指推定においては、複雑な指の動作を認識する必要があるが演奏中の指の動きは複雑で演奏中の画像を精密に記録することも容易ではない。そこで本研究では精密な3次元手形状モデルによってさまざまな指形状画像を生成することで画像認識で有効な手法とされている畳み込みニューラルネットワークの学習データを拡張する手法について検討する。

### 4. 研究成果

[3-1] 表情を伴った実演奏に対して聴覚情報・視覚情報・身体動作など多元的な演奏情報を含んだ表情付きマルチモーダル演奏情報データベースの仕様と構築方法について検討し、電子ピアノを用いて演奏された音響信号、MIDI 信号による打鍵タイミング・打鍵強度などの演奏情報、演奏中の指動作などの視覚情報として深度情報を含んだ RGB-D 情報を同時に記録する基本的なシステムを開発した。さらに、視覚情報を用いない手法についても検討した。筋電センサによって五指の動作を認識する手法が報告されていることから、演奏中の指動作に影響のない筋電センサを用いた上腕動作および指動作の取得方法について検討し、有効な手法となり得ることを確認した。

[3-2] 音響信号による楽譜追跡技術の高度化に取り組み、これまでに提案してきた SCRF と LDS に基づいたリアルタイム楽譜追跡技術を拡張し、従来手法のように基本的な音符列の情報だけでなく、楽譜に記載されている高次の情報に着目し、演奏モデル内で打楽器やメロディなど楽譜情報を活用する新たな仕組みを提案した。RWC 音楽データベースを用いた計算機実験では、リアルタイム性を損なうことなく楽譜追跡精度が改善することを確認した。

[3-3] 視覚情報を利用した運指推定手法

について検討した。まず演奏動作における特に指形状変化を取得するための画像処理手法として、畳み込みニューラルネットワーク(CNN)による手形状認識手法において、モデルの学習用データセットを拡張するために精密な3次元手形状モデルによって生成した合成画像を用いる手法を検討した。実写画像に加えて典型的な手形状の合成画像を多数生成することでデータセットを拡張し、実写画像に対する認識精度が大きく向上することを確認した。

[3-4] その他、これまでに取り組んでいた、演奏者の技量に応じたバイオリン運指推定の研究を進展させ、一つのメロディに対して多義性のあるギター TAB譜を音響信号から直接推定する手法について検討を行い、単独のコードを演奏した音響信号から押弦パターンそのものを再現したTAB譜を推定できることを示した。

## 5. 主な発表論文等 (研究代表者は下線)

### [雑誌論文](計 6件)

H. Hosoe, S. Sako, B. Kwolek  
Recognition of JSL Finger Spelling  
Using Convolutional Neural Networks,  
IAPR International Conference on  
Machine Vision Applications, pp.72-75,  
2017. DOI:10.23919/MVA.2017.7986796  
B. Kwolek, S. Sako, Learning Siamese  
Features for Finger Spelling  
Recognition, Advanced Concepts for  
Intelligent Vision Systems (ACIVIS  
2017), LNCS 10617, pp. 225-236, 2017.  
DOI: 10.1007/978-3-319-70353-4\_20  
S. Awata, S. Sako, T. Kitamura,  
Japanese Sign Language Recognition  
Based on Three Elements of Sign Using  
Kinect v2 Sensor, International  
Conference on Human-Computer  
Interaction 2017, pp.95-102, 2017.  
DOI: 10.1007/978-3-319-58750-9\_13  
S. Sako, M. Hatano, T. Kitamura,  
Real-time Japanese Sign Recognition  
Based on Three Phonological Elements  
of Sign, International Conference on  
Human-Computer Interaction (HCI  
2016), CCIS 618, pp.130-136, 2016.  
DOI:10.1007/978-3-319-40542-1\_21  
奥村 健太, 酒向 慎司, 北村 正, 統計  
モデルと楽譜情報に基づく使用を用い  
た演奏表情の比較分析, 感性工学会論  
文誌, Vol.15, pp.255-263, 2016.  
DOI:10.5057/jjske.TJSKE-D-15-00061  
奥村 健太, 酒向 慎司, 北村 正, 楽譜  
と表情を関連付けた統計モデルに基づ  
く鍵盤楽器の自動生成手法, 日本知能  
情報学会誌, Vol.28, pp.557-569, 2016.  
DOI:10.3156/jsoft.28.557

### [学会発表](計 4件)

酒向 慎司, 細江 花, Bogdan Kwolek, "3  
次元モデルとCNNを用いた指文字認識の  
検討", 電子情報通信学会技術報告 福  
祉情報工学研究会, Vol. 117, No. 66,  
WIT2017-10, pp. 45-49, 2017年5月27  
~28日, リオン株式会社(国分寺)  
野口 綾子, 酒向 慎司, 北村 正, 楽譜  
情報を考慮した SCRF に基づく音響信号  
と楽譜のアライメント, 日本音響学会  
2017年春季研究発表会, 3-4-7, 2017年  
3月15~17日, 明治大学(川崎)  
S. Awata, S. Sako, T. Kitamura, Vowel  
duration dependent hidden Markov  
model for automatic lyrics  
recognition, 5<sup>th</sup> Joint Meeting of the  
Acoustical Society of America and  
Acoustical Society of Japan, 2016年  
11月28日~12月2日, ホノルル(アメ  
リカ合衆国)  
A. Noguchi, S. Sako, T. Kitamura,  
Segmental Conditional Random Fields  
based Audio-to-Score Alignment  
Distinguishing Percussion Sounds from  
other Instruments, 5<sup>th</sup> Joint Meeting of  
the Acoustical Society of America and  
Acoustical Society of Japan, 2016年  
11月28日~12月2日, ホノルル(アメ  
リカ合衆国)

### [図書](計 0件)

### [産業財産権]

#### 出願状況(計 0件)

名称:  
発明者:  
権利者:  
種類:  
番号:  
出願年月日:  
国内外の別:

#### 取得状況(計 0件)

名称:  
発明者:  
権利者:  
種類:  
番号:  
出願年月日:  
取得年月日:  
国内外の別:

### [その他] ホームページ等

<http://www.mmsp.nitech.ac.jp/>  
<http://sakoweb.net/>

## 6. 研究組織

### (1) 研究代表者

酒向 慎司 (SAKO, Shinji)  
名古屋工業大学大学院工学研究科・助教  
研究者番号：30396791

### (2) 研究協力者

〔主たる渡航先の主たる海外共同研究者〕

Gerhart Rigoll  
ドイツ・ミュンヘン工科大学ヒューマンマシン  
コミュニケーション研究所・教授・

〔その他の研究協力者〕

Bogdan Kwolek  
ポーランド・AGH 科学技術大学・准教授  
Rafael Ramirez Merendez  
スペイン・ポンペイファブラ大学・准教授