

令和 5 年 6 月 26 日現在

機関番号：32665  
研究種目：基盤研究(C)（一般）  
研究期間：2019～2022  
課題番号：19K12288  
研究課題名（和文）ミクロとマクロの統合によるコミュニケーションの相互予測・適応モデルと音楽への適用  
  
研究課題名（英文）Mutual Prediction and Adaptation Model of Communication by Integrating Micro-and Macro-structures and Its Application to Music  
  
研究代表者  
北原 鉄朗（KITAHARA, Tetsuro）  
  
日本大学・文理学部・教授  
  
研究者番号：00454710  
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）：本研究では、人間とコンピュータが即興演奏によりセッションを行う「ジャムセッション」の実現を目指し、人間の即興演奏の予測およびコンピュータによる即興演奏の生成技術の開発に取り組んだ。特に、次の技術の開発を行った。

畳み込みニューラルネットワーク（CNN）を用いて即興演奏を生成する方法を実現した。CNNは、段階的な畳み込みによりデータの抽象度を上げていく手法であり、音楽が持つ階層的な拍節構造に合わせて畳み込みが行われるようにすることで、ブルーススタイルの旋律の生成を実現した。この技術を用いて即興演奏を支援するシステム「JamSketch」に応用した。

研究成果の学術的意義や社会的意義

音楽の生成は、近年、機械学習を中心としたコンピュータ技術が創造性（クリエイティビティ）を獲得できるかを探求する「計算論的創造性」の研究対象の一つとして注目を浴びている。本研究では、即興演奏の旋律をほぼリアルタイムで生成する技術を実現しており、当該分野に一定の貢献をもたらしたと言える。また、この技術に応用した「JamSketch」は、即興演奏ができない非専門家が簡単に即興演奏もどきを楽しむことができ、これまでにない体験をもたらすと期待できる。

研究成果の概要（英文）：Towards a jam session system that enables a human player and a computer to interact with each other through improvisation, we developed technologies that predict human improvisation and generate computational improvisation as follows:

We developed a method for generating improvisational melodies using a convolutional neural network (CNN). Because a CNN extracts features using hierarchical convolution layers, we achieved to generate Blues-style improvisational melodies by designing convolution layers based on the hierarchical metric structure. We applied this method to the JamSketch system, which allows users to play improvisation using melodic outlines.

研究分野：音楽情報処理

キーワード：即興演奏 旋律生成

## 1. 研究開始当初の背景

コミュニケーションは、人類が生活する上で必要不可欠である。コミュニケーションには「情報の共有」が目的のものと「情動の共有」が目的のものがある。後者では、当事者同士の相手の振る舞いや反応に対する相互の予測が適度に当たり、適度に外れるときに快がもたされると言われる。予測の的中により相手への共感が生まれ、予測の非的中により相手の新たな一面を発見できるからである。コミュニケーションを繰り返すことで予測的中の精度が上がれば（相互適応）、相手への相互理解が深まったことによる快が得られる。

「音楽」は、コミュニケーションの絶好の研究題材である。言語的コミュニケーションでは、会話中の話題の理解など難しい課題があるのに対し、音楽は、音符（ドレミ）の系列として表され、実世界の事物や概念にグラウンディングする必要がない。また、「音楽理論」の形で膨大な専門知識が蓄積されている。そのため、相互予測や相互適応の数理的なモデル化が比較的容易である。

申請者はこれまで、ベイジアンネットワークを用いた四声体和声の生成など、機械学習を用いた音楽生成の研究に従事してきた。この経験を活かし、ジャムセッション（即興演奏による合奏）を題材として音楽の予測・適応モデルを構築することにより、コミュニケーションの工学モデルの実現ができる。また、その応用例として即興演奏やジャムセッションをコンピュータとできるシステムを実現できれば、演奏の楽しみの幅が広がると期待できる。

## 2. 研究の目的

音楽（特に即興演奏やジャムセッション）を題材としてコミュニケーションの工学モデルの実現を目標に、次の研究に取り組む。

- 人間が行う即興演奏に対して、高精度に予測するモデルの実現。
- システムが即興演奏を生成するモデルの実現。
- 人間とシステムが相互に適応（相手に合わせて自分のふるまいを変えること）する方法の実現。
- これらの技術を活用した即興演奏システムの実現。

本研究期間では、主に2つ目と4つ目に着目して研究を進めた。

## 3. 研究の方法

### (1) システムが即興演奏を生成するモデルの実現

#### ① 研究方法の概要

旋律の大まかな形を曲線で表した「旋律概形」を入力とし、そこから旋律を生成するニューラルネットワークモデルを構築する。学習用の旋律には、実際の演奏家が演奏した旋律をデータベース化したものを使用する。その各々の旋律の音高の軌跡を平滑化することで、疑似的にその旋律の旋律概形を得る。つまり、ここで実現するのは、平滑化によって滑らかな曲線になった旋律から、平滑化前のものを推定するモデルである。これは、畳み込みニューラルネットワーク (CNN) を用いて実現する。具体的には、畳み込み層によって旋律概形から特徴抽出した後、逆畳み込み層によって音符レベルの旋律に変換する。

#### ② 使用データ

使用した楽曲は、Weimar Jazz Database (WJazzD) に収録されている楽曲のうち、Tonality Type に BLUES が指定されているもの 96 曲である。これらを 24 小節ごとに切断し、それぞれを別楽曲として扱った。また、データ拡張のために各楽曲の旋律を 1 オクターブ上げたものと下げたものを用意した。その結果、全楽曲数は 539 となった。これらを半数ずつに分け、前半を学習用、後半をテスト用に割り当てた。

#### ③ データ表現

入力データ、出力データともに 16 分音符ごとにベクトルが並んだ行列として表される。入力データと出力データの模式図を Figure 1 に、例を Figure 2 に示す。入力データ、出力データどちらも、時間分解能が 16 分音符のベクトルの時系列である。時刻  $n$  における入力ベクトル  $\mathbf{x}[n]$  は、旋律概形の音高を表す one-hot ベクトルとコード（和音）を表す many-hot ベクトルを結合したものである。一方、時刻  $n$  における出力ベクトル  $\mathbf{y}[n]$  は、旋律の音高を表す one-hot ベクトルである。時刻  $n$  で鳴り始めた音を前の時刻から継続している音を区別するため、これらは別々の要素によって表すこととする。

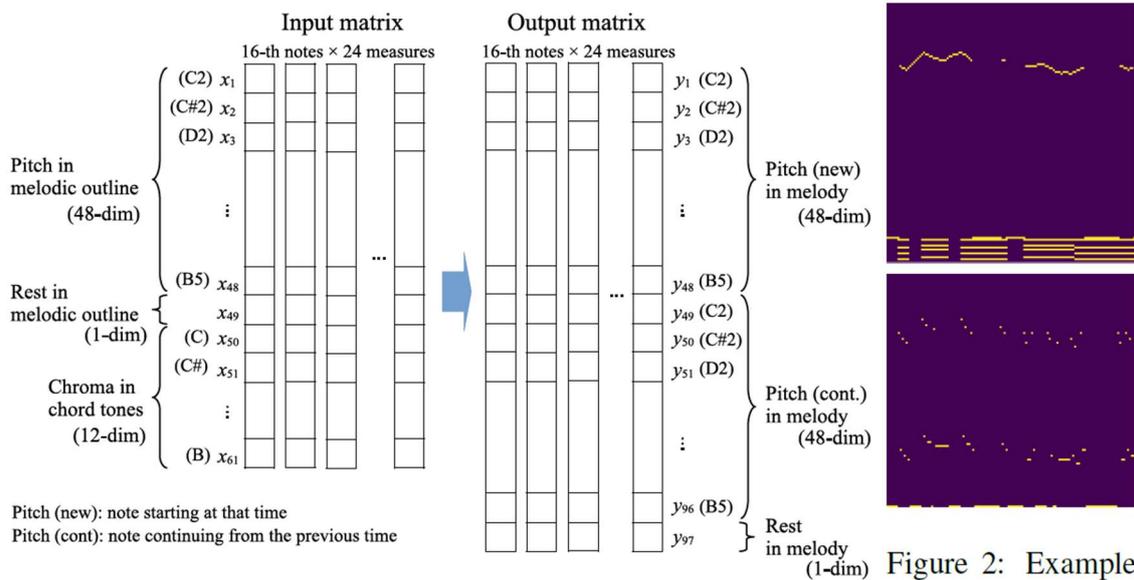


Figure 1: Representation of input and output data

Figure 2: Example of inputs (upper) and outputs (lower)

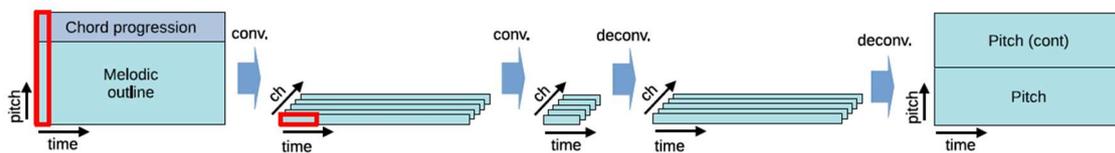


Figure 3: CNN architecture (Red rectangles represent filter sizes)

### ③ 前処理

各 MIDI データ (24 小節ごとに切断してある) が与えられると, C に移調した後, 16 分音符の音高 (および 1 つ前からの継続かどうか) に変換する. 元の MIDI データに 16 分音符よりも細かい音符がある場合, 一部の音符が欠落することになる. その後, 音高列に対して移動平均を用いて平滑化する. 平滑化の際のフィルタ長は 9 (およそ二分音符) とした. 平滑化の際, 短い休符は補間した上でやっている. これにより得られた音高列を旋律概形とみなして入力に用いる. 入力データでは音高は one-hot ベクトルで表されるので, 旋律概形の時刻ごとの値は整数に変換して用いられる.

### ④ CNN の学習

CNN のアーキテクチャを Figure 3 に示す. 入力される行列のサイズを  $M \times N$  とすると ( $N$  を時間軸とする), 1 層目のフィルタのサイズを  $M \times 1$  とする. チャンネル数を  $C$  とすると出力のサイズは  $1 \times N \times C$  である. 2 層目は,  $w$  拍分 ( $w=1,2,4$ ) の情報が 1 つのベクトルに集約されるようにフィルタを定義する.  $w=2$  の場合, 時間分解能が 16 分音符なので, フィルタのサイズは  $1 \times 8$  である. スライドも同じ値に設定する. 3 層目は 2 層目の逆演算, 4 層目は 1 層目の逆演算であるが, 出力される行列に合わせて 4 層目のフィルタサイズが若干異なる. 活性化関数は ReLU (第 4 層のみ softmax) とし, 各層にドロップアウト (ドロップアウト率: 0.2) とバッチ正規化を導入する.

このモデルを TensorFlow を用いて学習する. バッチサイズは 16, エポック数は 1000 とした. 学習には Adam を用い, 各種パラメータは TensorFlow のデフォルト値をそのまま用いる.

### ⑤ 評価方法

評価は, 正解 (平滑化前の音高列) との一致率の他, 音名の度数分布および隣り合う音高の差の度数分布についても, 正解と比較する. 度数分布の正解との比較は, 度数分布を正規化した後 KL ダイバージェンスの計算によって行う.

Table 1: Participants' musical experience

	Playing	Improvisation	Composition
0 year	1	12	19
1 to 5 years	7	8	8
6 to 10 years	5	1	1
11 years or longer	17	9	2
Total	30	30	30

Table 2: Results of objective evaluation

$w$	1			2			4		
	$C$	64	256	1024	64	256	1024	64	256
Conc.	0.2477	0.2149	0.1927	0.2507	0.2047	0.1931	0.2320	0.1956	0.1979
PCD	0.3414	0.0878	0.0314	0.5012	0.1016	0.0294	2.2294	0.1863	0.0701
PMD	18.6550	0.0618	0.0332	14.3351	0.1022	0.0478	18.4874	0.1463	0.0710

Conc.: note-level concordance rate, PCD: pitch class dissimilarity, PMD: pitch motion dissimilarity

Table 3: Results of subjective evaluation (all participants)

		Overall	Dissonant	Variation	Originality	Blues	Human
Given	Mean	3.88	3.82	4.09	4.82	4.02	3.97
	SD	0.82	0.70	0.66	0.51	0.60	0.82
	Lowest	2.67	2.89	3.00	3.92	3.33	2.78
	Highest	5.15	5.15	5.18	5.64	5.18	5.36
	5 or higher	2	1	1	4	1	2
Generated	Mean	3.95	3.97	4.17	4.52	3.69	3.97
	SD	0.95	0.74	0.77	0.59	0.78	1.01
	Lowest	2.67	3.07	3.11	3.44	2.60	2.89
	Highest	5.86	5.29	5.86	5.57	5.29	6.14
	5 or higher	2	2	1	2	1	2

それに加え、主観評価も行う。Web上のクラウドソーシングサービスを利用し、30名の参加者を募った。参加者の音楽経験はTable 1の通りである。評価対象は、11個の旋律を選択し、各々に対応する旋律概形から生成した旋律を加えた、計22個である。各参加者はweb上でランダムに選ばれた8つの旋律を聴き、次の項目に対して7段階(1~7)で回答してもらった。

- 全体的な質 (Overall)
- 不協和音の少なさ (Dissonant)
- 1巡目と2巡目のメリハリ (Variation)
- オリジナリティ (Originality)
- ブルースらしさ (Blues)
- 人間らしさ (Human)

## (2) 旋律生成技術を活用した即興演奏システムの実現

(1)で実現した旋律概形からの旋律生成技術を用いて、新たな即興演奏システムを実装した。申請者らが以前開発した「JamSketch」は、ユーザが旋律概形を画面に描画すると、それに沿った旋律を即座に生成して演奏するシステムであるが、旋律の生成に遺伝的アルゴリズム(GA)を使用していたため、音楽的に適切な旋律を得るのに時間がかかるという欠点があった。リアルタイム性を重視するために、GAによる最適化を0.5秒で止めていたため、十分な最適化がなされていない(≒音楽的に適切ではない)旋律が出力されることもあった。(1)で実現した旋律生成技術は、モデルの学習こそ時間がかかるものの、学習済みのモデルを使って旋律を生成するのはほとんど時間がかからない。そこで、GAによる旋律を最適化を廃止して(1)でモデルを導入した新バージョン「JamSketch Deep  $\alpha$ 」を実装する。

## 4. 研究成果

### (1) システムが即興演奏を生成するモデルの実現

客観評価の結果をTable 2に示す。Conc.は音符レベルの正解との一致率、PCDは音名の度数分布の正解からの距離(非類似度)、PMDは隣り合う音高差の度数分布の正解からの距離である。このことから、音符レベルでの正解との一致率は20~25%程度と低いものの、隣り合う音高差の度数分布との比較については、フィルタ数が多いほど非類似度が下がることが分かった。

主観評価の結果をTable 3に示す。Given(データベースから取り出した旋律)とGenerated(生成された旋律)との間で、いずれの指標でも大きな差は見られず、人間の演奏に近い旋律を生成できたと言える可能性がある。

### (2) 旋律生成技術を活用した即興演奏システムの実現

(1)で構築したモデルをJamSketchに組み込んだ「JamSketch Deep  $\alpha$ 」を実装した。JamSketchはGroovyで実装されているため、学習済みのモデルをTensorFlow Javaで読み込む形とした。その結果、旋律が生成される速度や生成される旋律の質を改善することができた。

5. 主な発表論文等

〔雑誌論文〕 計5件（うち査読付論文 5件／うち国際共著 1件／うちオープンアクセス 3件）

1. 著者名 Yeh Yin-Cheng, Hsiao Wen-Yi, Fukayama Satoru, Kitahara Tetsuro, Genchel Benjamin, Liu Hao-Min, Dong Hao-Wen, Chen Yian, Leong Terence, Yang Yi-Hsuan	4. 巻 50
2. 論文標題 Automatic melody harmonization with triad chords: A comparative study	5. 発行年 2021年
3. 雑誌名 Journal of New Music Research	6. 最初と最後の頁 37 51
掲載論文のDOI（デジタルオブジェクト識別子） 10.1080/09298215.2021.1873392	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 該当する
1. 著者名 Ayumi Shiga, Tetsuro Kitahara	4. 巻 12631
2. 論文標題 Generating Walking Bass Lines with HMM	5. 発行年 2021年
3. 雑誌名 Perception, Representations, Image, Sound, Music. CMMR 2019. Lecture Notes in Computer Science	6. 最初と最後の頁 248-256
掲載論文のDOI（デジタルオブジェクト識別子） 10.1007/978-3-030-70210-6_17	査読の有無 有
オープンアクセス オープンアクセスではない、又はオープンアクセスが困難	国際共著 -
1. 著者名 Mina Shiraishi, Kozue Ogasawara, and Tetsuro Kitahara	4. 巻 Vol.27
2. 論文標題 HamoKara: A System that Enables Amateur Singers to Practice Backing Vocals for Karaoke	5. 発行年 2019年
3. 雑誌名 Journal of Information Processing	6. 最初と最後の頁 683-692
掲載論文のDOI（デジタルオブジェクト識別子） 10.2197/ipsjjip.27.683	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -
1. 著者名 Yusuke Tsuchiya, Tetsuro Kitahara	4. 巻 Vol.3, No.1
2. 論文標題 A Non-notewise Melody Editing Method for Supporting Musically Untrained People's Music Composition	5. 発行年 2019年
3. 雑誌名 Journal of Creative Music Systems	6. 最初と最後の頁 1-25
掲載論文のDOI（デジタルオブジェクト識別子） 10.5920/jcms.624	査読の有無 有
オープンアクセス オープンアクセスとしている（また、その予定である）	国際共著 -

1. 著者名 Tetsuro Kitahara	4. 巻 3
2. 論文標題 Generating Melodies from Melodic Outlines Towards an Improvisation Support System for Non-musicians	5. 発行年 2023年
3. 雑誌名 IIAI Letters on Informatics and Interdisciplinary Research	6. 最初と最後の頁 1-8
掲載論文のDOI (デジタルオブジェクト識別子) 10.52731/liir.v003.081	査読の有無 有
オープンアクセス オープンアクセスとしている (また、その予定である)	国際共著 -

〔学会発表〕 計38件 (うち招待講演 2件 / うち国際学会 9件)

1. 発表者名 関晋之介, 北原鉄朗
2. 発表標題 ユーザの演奏のペロシティ変化を考慮するドラム演奏表情付けシステム
3. 学会等名 情報処理学会研究報告 音楽情報科学 (MUS)
4. 発表年 2021年

1. 発表者名 北原鉄朗
2. 発表標題 誰もが創作を通じて音楽を楽しめる世界を目指して
3. 学会等名 音学シンポジウム (招待講演)
4. 発表年 2021年

1. 発表者名 北原鉄朗
2. 発表標題 北原研究室の研究事例紹介: ベーシストの旋律分析とイコライザーの印象分析
3. 学会等名 Music×Analytics Meetup #5 (招待講演)
4. 発表年 2021年

1. 発表者名 島村 美羽, 阿久井 愛, 北原 鉄朗
2. 発表標題 多人数演奏楽譜から連弾譜への自動編曲
3. 学会等名 日本音響学会 2022年 春季研究発表会
4. 発表年 2022年

1. 発表者名 阿久井 愛, 島村 美羽, 北原 鉄朗
2. 発表標題 Jpopのラテン風ピアノ編曲システムの試作
3. 学会等名 日本音響学会 2022年 春季研究発表会
4. 発表年 2022年

1. 発表者名 山川 峻, 北原 鉄朗
2. 発表標題 GTTM分析のオープンソース実装
3. 学会等名 日本音響学会 2022年 春季研究発表会
4. 発表年 2022年

1. 発表者名 Aiko Uemura and Tetsuro Kitahara
2. 発表標題 Morphing-Based Reharmonization using LSTM-VAE
3. 学会等名 The 2020 Joint Conference on AI Music Creativity (国際学会)
4. 発表年 2020年

1. 発表者名 Mio Kusachi, Aiko Uemura and Tetsuro Kitahara
2. 発表標題 A Piano Ballad Arrangement System
3. 学会等名 The 2020 Joint Conference on AI Music Creativity ( 国際学会 )
4. 発表年 2020年

1. 発表者名 Ayumi Shiga and Tetsuro Kitahara
2. 発表標題 Generating Walking Bass Lines with HMM
3. 学会等名 The 14th International Symposium on Computer Music Multidisciplinary Research (CMMR 2019) ( 国際学会 )
4. 発表年 2019年

1. 発表者名 Seiya Masuda, Eriko Aiba, and Tetsuro Kitahara
2. 発表標題 An Investigation towards Verbally Controllable Equalizer for Singing Voices
3. 学会等名 The 5th Workshop on Intelligent Music Production (WIMP 2019) ( 国際学会 )
4. 発表年 2019年

1. 発表者名 関 晋之介, 井上 湧哉, 北原 鉄朗
2. 発表標題 ドラム演奏表情付けに向けたドラム演奏のペロシティの分析
3. 学会等名 情報処理学会 第126回音楽情報科学研究会 萌芽・デモ・議論セッション
4. 発表年 2020年

1. 発表者名 塚本 康太, 饗庭 絵里子, 北原 鉄朗
2. 発表標題 クラシック曲の楽譜データに対する自動和声付与システムの構築に向けて
3. 学会等名 情報処理学会 第126回音楽情報科学研究会 萌芽・デモ・議論セッション
4. 発表年 2020年

1. 発表者名 Yuki Iwamoto and Tetsuro Kitahara
2. 発表標題 A Music Loop Sequencer with User-Adaptive Music Loop Selection
3. 学会等名 Proceedings of the 4th ACM International Conference on Multimedia in Asia (国際学会)
4. 発表年 2022年

1. 発表者名 Tetsuro Kitahara and Akio Yonamine
2. 発表標題 JamSketch Deep : A CNN-Based Improvisation System in Accordance with User's Melodic Outline Drawing
3. 学会等名 Proceedings of the 4th ACM International Conference on Multimedia in Asia (国際学会)
4. 発表年 2022年

1. 発表者名 Sai Oshita and Tetsuro Kitahara
2. 発表標題 Automatic classification of blowing properness in flute sounds
3. 学会等名 Proceedings of the 24th International Congress on Acoustics (ICA 2022) (国際学会)
4. 発表年 2022年

1. 発表者名 Aiko Uemura and Tetsuro Kitahara
2. 発表標題 Morphing-based Reharmonization with VAE: Reducing Dissonance with Consonance-based Loss Function
3. 学会等名 Proceedings of the 3rd Conference on AI Music Creativity (AIMC 2022) (国際学会)
4. 発表年 2022年

1. 発表者名 北原 鉄朗
2. 発表標題 即興演奏支援に向けた旋律生成の一試行
3. 学会等名 第36回人工知能学会全国大会論文集
4. 発表年 2022年

〔図書〕 計0件

〔産業財産権〕

〔その他〕

日本大学 文理学部 情報科学科 北原研究室 <a href="http://www.kthrlab.jp/">http://www.kthrlab.jp/</a>
--

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
---------	---------