

第2回音楽情報科学萌芽・デモ・議論セッション

深山覚¹ 藤田良祐² 河原英紀³ 大矢隼士⁴ 下道雄太⁵ 山本雄也⁶ 須山夢菜⁷ 丸山新世⁸
関晋之介⁹ 澤田恭平¹⁰ 類家怜央¹¹ 渡邊一樹¹² 城田晃希¹¹ 塚本康太¹⁰ 中島凱斗¹¹ 櫻沢繁¹¹
武田郁弥¹¹ 名畑皓正¹¹

概要：本稿では第126回音楽情報科学研究会における萌芽・デモ・議論セッションの発表内容について述べる。このセッションは十分な結果は出ていない萌芽的な研究から、既に国際会議や論文誌投稿まで進んだ完成度の高い研究まで、幅広くポスター発表・デモ・議論できる場である。今回のセッションでは、歌唱、音楽の分析、楽器演奏、楽音合成・音楽生成に関する合計17件の発表が行われる。

1. はじめに

深山覚（産業技術総合研究所）

音楽情報科学は学際的な学術領域であり、音楽学・計算機科学・心理学・認知科学・生理学・図書館情報学・電気電子工学・機械学習等の多様な学術的方法の組合せによって研究が推進される。実際、音楽情報科学研究会の参加者には、複数の学術的方法を駆使する、いわば二足以上の草鞋を履いた方々が多くみられる。しかし先に挙げた学術領域のすべてをマスターしてから調査・研究に取り組むというのでは、準備に時間がかかりすぎてしまう。音楽情報科学の研究をするにあたっては、自分のフィールドに閉じこもらず、自身の知らない学術領域があることを素直に認め、その道の実践者の教えに謙虚に耳を傾けるのが賢明である。

萌芽・デモ・議論セッションは、そのような多様な学術領域の実践者と触れ合う絶好の機会であり、音楽情報科学研究会のハイライトでもある。既に国際会議や論文誌投稿まで進んだ完成度が高い研究紹介は言うまでもなく、十分な結果が出ていない萌芽的な内容の発表は、自身がもし同じような研究テーマに取り組んだ場合に直面するであろう課題を知るのに役立つ。また、デモの実演や体験が含まれ

る発表は、あたかも自分が追実験を行ったかのような実感を持って発表を聞くことができ、一段と研究に対しての理解を深められる。

今回のセッションには大きく4群に分類できる合計17件の発表が集まった。音声信号処理や聴覚心理学との接点の多い歌唱に関する研究が3件、人間が音楽を理解する過程を情報処理により実現する音楽の分析に関する研究が3件、楽器音響の物理から演奏者たる人間について幅広く明らかにする楽器演奏に関する研究が5件、そして機械学習の知見を取り入れて楽音や音楽の自動生成を行う楽音合成・音楽生成の研究が6件発表される。一方で、コンピュータ音楽などの音響芸術の創作や、邦楽を含む民族音楽に関連する研究など、いくつか重要な領域からの発表が含まれなかった。今後、本セッションのような企画を通じて、さらに音楽情報科学の学際性の質と幅広さを磨いていくことが、有意義な学術的知見を生み出す研究会の動力源となると考える。

2. 歌唱に関する研究

2.1 聴覚刺激による歌唱音声パラメタの変調計測 -周波数領域ベルベットノイズに基づく直交系列の応用-

河原英紀（和歌山大）、榊原健一（北海道医療大）

森勢将雅（明治大）、松井淑恵（豊橋技大）

津崎実（京都市立芸大）、入野俊夫（和歌山大）

歌唱などで母音を長時間発声している状況で、聴覚による刺激が発声される音声のパラメタに与える影響を測定する手段を発見・発明した。この方法の鍵は、自由度が高く相互にほぼ直交する時間伸長パルス (TSP: Time Stretched Pulse) の新しいファミリーとなる周波数領域ベルベットノイズ (FVN: Frequency domain variant of Velvet Noise) に

¹ 産業技術総合研究所
² 明星大学
³ 和歌山大学
⁴ 株式会社レコチョク
⁵ 東京大学
⁶ 筑波大学
⁷ 洗足学園音楽大学
⁸ 早稲田大学
⁹ 日本大学
¹⁰ 電気通信大学
¹¹ 公立はこだて未来大学
¹² 中京大学

ある。FVN を二値の直交系列で変調し周期的に配置することにより、完全に直交する系列を構成することができる。この系列を聴覚刺激として与える音響信号のパラメタの変調に用いることにより、発声された音声のパラメタへの影響をインパルス応答として、定量的に誤差の推定値とともに求めることができる。会場では、この現象を測定することができるデモシステムを用意し、参加者に体験してもらう。

関連 Web ページ:

<https://github.com/HidekiKawahara/FVN>

<https://github.com/HidekiKawahara/voiceRTFB>

2.2 深層弱ラベル学習を用いたポピュラー音楽の歌唱テクニック検出の基礎検討

山本雄也 (筑波大学)

中野倫靖, 後藤真孝 (産業技術総合研究所)

寺澤洋子, 平賀譲 (筑波大学)

本研究では歌声に含まれる歌唱テクニックの時刻と種類を自動検出することで、歌手の個性を定量化することを目標とする。ビブラート、しゃくり、エッジボイス等の歌唱テクニックの特性や出現箇所・頻度は、歌唱力や個性を分析する上で重要である。従来、ビブラート検出や声区（地声や裏声等）の判定、単一のテクニックが含まれる短い歌唱ファイルのからのテクニック識別など、特定の歌唱テクニックの検出を目的とした研究はあったが、一つの歌唱に含まれる複数のテクニックを総合的に扱える研究はなかった。明らかになっている歌唱テクニックの特性は一部分・一側面のみであるため、検出のための音響特徴量は深層学習によって自動獲得したい。しかし同様の理由から、歌声合成によって全ての学習データを自動生成することは難しく、歌唱テクニックの時刻付きラベルデータ（強ラベル）も存在しない。そこで本稿では、時刻なしラベルデータ（弱ラベル）を伴う歌声データセットとして VocalSet を学習データとした、畳み込みニューラルネットワーク（CNN）による歌唱テクニック検出手法を提案する。10種類のテクニックを対象とした評価実験では、10種類中4種類のみ、70%以上の精度で検出できた。本セッションではモデルの改良、検出するテクニック、今後の応用などを議論したい。

2.3 自己音声を用いた深層学習によるお手本歌唱生成システム

渡邊一樹, 久野文菜, 谷口航平, 濱川礼 (中京大学)

本稿では、ユーザが目標とする歌唱から、深層学習を用いてユーザ自身の声でのお手本歌唱を生成する手法について述べる。歌唱練習を行うには、カラオケマシンでの採点結果や分析画面から自分に足りないテクニックを知るといった方法や、自分の歌唱を録音し、お手本となる歌唱と聞き比べを行い、ピッチのズレなどを認識するといった方

法がとられる。しかし、それらの方法では、ユーザはカラオケマシンで分析されても、具体的にどのように歌えばいいのかをイメージしづらかったり、お手本となる声とユーザの声は違うため、二つの認識の齟齬を埋められるようになる必要があるといった問題がある。そこで、本システムではユーザ自身の声での歌のお手本を生成するシステムを提案する。手法として、深層学習を用いてユーザが目標とする歌唱の声質に相当するスペクトル包絡をユーザのスペクトル包絡に変換を行う。また、ユーザが目標とする歌唱の声の高さに相当する基本周波数をユーザの声の高さに調整を行う。これらの方法により、生成されたお手本を聴き練習を行うことで、ユーザは具体的な歌い方をイメージしやすくなり、歌唱練習がより効率的に行えるようになるのではないかと期待できる。

3. 音楽の分析に関する研究

3.1 コンサート映像コンテンツのメタ情報付与のためのユーザインターフェース

大矢隼士 (株式会社レコチョク), 森島繁生 (早稲田大学)

5G 実用化が目前に迫りいつでも場所を選ばずに映像コンテンツを楽しめる時代が到来しつつある。株式会社レコチョクでは、AKB48 グループ映像倉庫や RecTV といった映像サービスを展開している。配信される映像コンテンツは今後も増加の一途をたどるが、個々のユーザにより良くコンテンツを届けるサービスにしていくためには様々なメタ情報が不足している。また、映像サービスのメタ情報は事業部の社員が人力で入力しており、人的・時間的コストがかかっているのが実情である。そこで我々はこれらの問題を解決するため、楽曲解析を用いたコンサート映像の楽曲推定システムと推定誤りを最小限の労力で修正するユーザインターフェースの開発をおこなっている。コンサート楽曲推定では、まず、コンサート音源の各サンプルごとに RMS・MFCC・テンポを抽出し楽曲区間を推定する。その後、楽曲区間における MFCC の平均・分散とデータベース中の楽曲の MFCC の平均・分散をもとに楽曲を推定する。本発表では、現在開発中である AKB48 グループ映像倉庫のためのコンサート映像の楽曲推定システムの全体の構想とサービスの将来的なビジョンを紹介するとともに、推定誤り修正ユーザインターフェースを展示する。

関連 Web ページ:

<https://akb48group-eizo.jp/>

3.2 楽曲を表現する形容詞と音響特徴量をそれぞれ学習した 2 つの VAE 同士の対応づけについて

類家怜央, 竹川佳成, 平田圭二 (公立はこだて未来大学)

近年、ストリーミングサービス等の普及により、楽曲の推薦手法に注目が集まっている。しかし、推薦手法は主に協調フィルタリングが使用されるか用意されたプレイリ

ストを提示するだけとなっている。また、推薦楽曲のアーティストが同一であるなどの問題がある。本研究では、楽曲情報に関する VAE と、その楽曲の曲調を表現する形容詞に関する VAE を同時に学習させ、得られた両 VAE の潜在空間どうしを対応付ける。これより、一方の VAE に形容詞を入力すると、他方の VAE から楽曲情報が出力される。実験では提案システムに 240 曲を学習させ、推薦楽曲の多様性と推移の妥当性を確認した。

3.3 スパイラルアレイモデルとベイズ推論を応用した楽曲の局所調性階層を検出する試みについて

名畑皓正, 竹川佳成, 平田圭二 (公立はこだて未来大学)

本研究は、任意の楽曲の局所調性階層を検出することにより、調性が局所的に階層構造を持っており、それが楽曲の特徴量として計算可能であることを検証することが目的である。調性音楽における音どうしの関係を三次元にまとめ幾何学計算により表せるようにしたモデルにスパイラルアレイモデル (E. Chew, 2000) (以下 SAM) がある。SAM では、ピッチ・コード・キーという次元の異なる概念を同一の三次元座標空間内で座標値として表すことができる。SAM に対して Center of Effect Generator のアルゴリズムを適用することで楽曲の調性同定を行い、Boundary Search Algorithm (以下 BSA) によって転調境界を求めることができるが、距離の計算を用いるため結果は離散的である。本研究では、最小二乗法を用いた最適化とベイズ推論により、SAM のキーの点群を確率分布として表した B-SAM を提案、B-SAM を用いた転調境界を検出するアルゴリズム B-BSA を考案し、局所調性階層を検出するアルゴリズムとして提案する。

4. 楽器演奏に関する分析

4.1 クラリネットリードについて—ブライントテストによる EXT 処理リードと EXT 未処理リードの比較—

須山夢菜 (洗足学園音楽大学大学院)

平野剛 (電気通信大学), 辻功 (洗足学園音楽大学)

クラリネットは、マウスピースという歌口リードを取り付け、息を吹き込むことでリードを振動させて音を鳴らしている。クラリネットのリードは植物体の葦から作られるため、一定の品質を確保することは非常に難しく、良い状態のリードを確保することは、奏者にとって重大なことであると同時に困難なことでもある。近年、低温処理を施したリード (EXT リード) が開発されこれを利用する奏者も多いが、EXT リードが EXT 未処理のリードと比べて奏者によってどの程度良いものなのかを調べた研究はない。そこで本研究では、音楽大学生 14 名を対象に EXT リードと EXT 未処理のリードの良し悪しを比較する主観評価を行った。実験の結果、1 番演奏しやすいリードに EXT リードを選んだのは、被験者 14 名のうち 9 名で、EXT 処

理の効果が見られる結果となった。しかし EXT 未処理のリードの中には、EXT リードよりも明らかに評価の高いものも存在した。これは主観評価には、吹き心地の良さの他に音色など他の要因も順位に関係しているためと考えられる。今後、リードの個体差を考慮した研究や、サンプル数を増やした研究などを行う必要があると考えられる。

4.2 ピアニストのサステインペダル制御と音響特徴量の関係

澤田恭平, 饗庭絵里子, 中鹿亘 (電気通信大学大学院)

ピアノは鍵盤を押すとハンマーが弦を打つことによって音が出る構造になっている。弦には、ダンパと呼ばれる弦を押さえつけ振動を止める装置が取り付けられている。ピアノのサステインペダルは、その操作によってダンパの押し付け具合を調整し、弦振動の持続や停止を制御することができる。プロピアニストは、これを細かに調整し、音の濁り具合などを調整しているとされている。そこで、レーザー測位計を用いてピアニストのペダリングを測定し、様々な音響特徴量との関係性を観察することにより、どのような音響特徴量に基づいて制御されているのかを検証する。その第一段階として、本研究では音の協和度に着目することとし、協和度と相関がある指標としてスパースネスを用いて、プロピアニストのサステインペダル制御との関係を観察したところ、特に打鍵のタイミングにおいて、スパースネスの低下を抑制するようなペダル制御を行っていることが明らかになった。しかしながら、ペダル制御とスパースネスに完全な相関があるわけではないことから、他の指標についても検討が必要である。この点について議論し、将来的に自動ペダル付与システムの開発に貢献したい。

4.3 鍵盤ピッチ配列を動的変更することで不協和な音を補正する即興演奏支援システムのデモンストレーション

武田郁弥, 竹川佳成, 平田圭二 (公立はこだて未来大学)

本研究の目的は、未熟な楽器演奏者の即興演奏への参加を支援するシステムの提案である。対象ユーザは音楽知識がなく、不協和音を意図して避けることができない演奏者を想定している。そこで本研究ではユーザの演奏した旋律中の不協和音を補正する鍵盤楽器型の支援システムを作成した。また、ユーザは実際に他の演奏者と即興演奏をすることを想定して、リアルタイムに補正処理が行われることをシステム要件とした。本システムでは不協和音となるピッチに対応する鍵盤が打鍵された時、対応するピッチとは異なる、不協和音ではないピッチが出力されるように鍵盤上のピッチ配列を変更する。演奏者が打鍵するたびに、鍵盤のピッチ配列が動的に変更される。現在時刻に旋律に対する後続音候補は、旋律のデータベースに基づいた N-gram によって求める。後続音候補としての生起確率が 0 であるピッチを不協和音とする。不協和音と判断した

ピッチは、鍵盤上の距離が近く生起確率が高いピッチへ変更される。デモでは旋律が補正される様子を実演し、本システムによって即興演奏支援が実現されていることを示す。

4.4 ドラム演奏者のスティック操作における筋電位波形と熟達度の関係

櫻沢繁, 郡司貴明 (公立はこだて未来大学)

生体信号を用いた打楽器演奏時の身体動作解析に関する研究が報告されている。特に、筋電位から抽出した打点時刻前後の和太鼓のバチさばきやドラムのスティックコントロールに関する情報と、打楽器の鳴音との関連が多く報告されている。しかしそのほとんどが、筋電位のパワーとタイミングのみから推定される随意的な腕の制御に着目されている。一方、打楽器の演奏時には楽器をたたいた瞬間の反作用により、身体は反射的な活動を示すことが想定される。そこで本研究では、ドラム演奏時のスネアドラムをたたいた時の右前腕の筋電位を計測し、打点時刻前後の筋電位波形に着目し、楽器の反作用の影響で生じたと考えられる不随意的な身体活動がないか調べた。その結果、熟達者は打点時刻において完全な脱力状態であり、ドラムからの反作用を受けたときの脊髄反射と考えられる特徴的な波形の筋電位が観測された。ドラム演奏の熟達者は、楽器の反作用に対する精髄反射を不随意で利用しながらスティックコントロールをしていることが推定される。

4.5 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

関晋之介, 井上湧哉, 北原鉄朗 (日本大学)

我々は、人間がキーボードを演奏し、システムがドラムパートなどを演奏するジャムセッションシステムの構築を目指している。このようなシステムを実現する上での課題の1つは、ドラム演奏に適切な演奏表情を付与することである。本研究では、演奏表情の中でも音の強さ(ペロシティ)に焦点を当てて研究を進める予定である。ペロシティに着目した演奏表情の付与で重要なのは、大局的な演奏の盛り上がりを考慮することであると考えている。演奏の盛り上がりに合わせて人間によるキーボード演奏のペロシティが大きくなると予想されるため、システムによるドラム演奏も同様に、演奏の盛り上がりに合わせてペロシティを大きくすることで、息の合ったセッションになると予想される。我々は、この仮説を検証するため、人間同士のセッションの記録・分析を進めている。大学のモダンジャズ研究会に所属する2名に電子ピアノと電子ドラムでジャズのスタンダードナンバーを演奏してもらい、MIDI形式で記録した。両者のペロシティを図示して比較したところ、不明瞭ではあるものの、両者のペロシティの大局的な変化には、若干の関連がありそうだという結果が得られた。

5. 楽音合成・音楽生成の研究

5.1 一般化FM合成における微分項の導入による近似精度改善の検討

中島凱斗, 竹川佳成, 平田圭二 (公立はこだて未来大学)

周波数変調 (FM) 合成を用いて作られた FM シンセサイザは、構成が単純で制御すべきパラメータ数が少ないにもかかわらず、非整数次の複雑な倍音成分を含む楽音を合成することができる。一方、パラメータ値と出力音との関係や、FM 発振器の接続パターンと出力音との関係を直感的に理解するのが困難であると言われている。そこで我々は一般化 FM 合成方式を提案する。まず、FM 発振器を活性化関数として持つようなニューラルネットワークを構成し、FM 発振器のパラメータ、下位層の FM 発振器出力の重み付き和のパラメータをバックプロパゲーションによって最適化を行う。すべての活性化関数が同じ時間 t を共有しており、一般化 FM 合成は従来の FM 合成の一般形と考えることができる。一般化 FM 合成は、周期信号に限らず任意の信号波形を学習できる。本デモ発表では、より長時間で複雑かつ非定常の信号波形を学習するため、教師信号の微分とネットワークの出力信号の微分の差を損失関数に加える方式を提案する。教師信号と出力信号の微分を考慮することで、波形の急峻な部分の近似精度が向上することが期待される。デモでは、損失関数に微分項を加えた効果を可視化し、実際に音として聴取していただく予定である。

5.2 和音データの学習によるバッハ風楽曲の自動生成

藤田良祐, 村杉亮, 相澤俊, 武藤良, 横山真勇, 植木一也 (明星大学)

本研究では、ニューラルネットワークを用いてバッハの楽曲を学習し、対位法的な楽曲を自動生成するシステムを提案する。和声理論を考慮して楽曲を学習させる手法により、生成された曲がバッハの作曲したものにより近いものになることを目標としている。バッハの楽曲の MIDI ファイルから音の情報を2次元的なテキストデータに表すことで和音を学習させるためのデータを作成し、それを機械学習の入力に使用した。機械学習により、生成された曲の検証を行った結果、和音の生成が可能であることが確認されたが、学習用データの不足により、生成される音のバリエーションが少なくなってしまう、学習させた曲のメロディしか生成されないなどの問題も確認された。今後の課題として、学習用データを増加させ、生成される音のバリエーションを増やすことや、和声法や対位法など、他の音楽理論を学習に組み込む手法を検討するなどが挙げられる。

5.3 身体の動きを用いたインタラクティブな音楽生成に関する研究

下道雄太, 入江英嗣 (東京大学), 坂井修一 (東京大学)

本研究では, 身体の動きを用いてインタラクティブに音楽を生成する方法について検討する. 身体の動きを音に結びつけたサウンドインタラクションについてはこれまでも様々な提案がなされてきたが, その多くは生成される音が抽象的であったり, 演奏が難しい, 自由な身体の動きが制限されるなどの課題があった. そこで本研究では, 音楽経験のない初心者であっても, 自由に身体を動かすことで音が鳴り, それが音楽として自然に構築されていくようなシステムの実現を目標とする. 一つのアプローチとして, プレイヤーが入力した単純なフレーズをループ, 変形, 重っていくことで音楽として形作っていくことはできないかと考えた. すなわち, 入力されたフレーズを素材として音楽を構築しながら, その過程に身体の動きによる相互作用をいれた, 音楽の生成と演奏を同時に行えるようなシステムである. 本デモ発表ではその前段階として, まずは単音を一つのフレーズとみなし, 身体を動かすと音楽的な音がインタラクティブに生成されるという体験を実現した. 実装には骨格推定ライブラリの openpose と SuperCollider を用いた.

5.4 機械学習生成モデルによる楽曲の分類と生成

丸山新世, 柳尾朋洋 (早稲田大学大学院)

近年, 機械学習の発達により画像分野において大きな研究成果が見られ, 音楽分野においても人間と同じように楽曲生成を実現することが期待されている. 教師あり学習を用いた楽曲生成の研究例が存在するが, 教師あり学習による出力結果は学習データのパターンを模倣したものになるためコンピュータ独自のオリジナリティを生み出すのが難しい. したがって, 教師なし学習によってコンピュータが見出した特徴量から楽曲生成を行うことが有効であると考えられる. 本研究では学習データの模倣にならないオリジナリティのある楽曲生成を行うことを目的とし, 変分オートエンコーダを用いて楽曲データの特徴量を抽出し楽曲分類を行う. そして, 敵対的生成ネットワークを楽曲データに適用し楽曲生成を試みる. 結果, 楽曲データのパワースペクトルの時間変化を学習データにすることで, 楽曲が持つ特徴量を用いてジャンルごとの分布を低次元空間で可視化し特徴づけることができた. また, ピアノロールとして表された楽曲データを学習データにすることで, 画像データと同様の扱いで学習データの模倣にならない楽曲生成を行った.

5.5 深層学習を用いたモーフィング出力に含まれる非和声の抑制方法の検討

城田晃希, 竹川佳成, 平田圭二 (公立はこだて未来大学)

近年, 深層学習による音楽自動生成の研究が隆盛である. 自己符号化器に旋律を学習させる時, 従来の方式では, その旋律に含まれるスケール音や和音を区別せずその他の音と同じ重要度で損失関数を計算していた. その結果, モーフィングによって中間の旋律を生成する時に, 出力旋律に非和声音が現れ, 音楽的一貫性が損なわれる場合がある. モーフィングとは, 2つの楽曲 A, B に対してその中間的な楽曲を生成する方法である. 本研究では, 自己符号化器には VAE を用い, 潜在空間上でにおける楽曲間の領域を移動する操作によって, 楽曲のモーフィングを実現する. 我々は, 学習用旋律に含まれるスケール音や和音を考慮し, 非和声音に対する誤差の重みが大きいような損失関数を提案する. それより, 非和声音の出現確率の低減が期待される. 現在は, 教師データの旋律に, 手動で和声か非和声かというアノテーションを付与し, 生成モデルの実装と予備実験を進めている. デモでは, 学習用と生成用に RWC ミュージックデータベースの楽曲を用いる. 提案手法を用いて, 出力に現れる非和声音の数が変化することを示す.

5.6 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

塚本康太, 饗庭絵里子 (電気通信大学大学院)

北原鉄朗 (日本大学)

現在私たちが耳にする多くの楽曲は機能和声と呼ばれる概念に基づいて作られている. 和声とは, 和音を構成すること, または和音を継続的に連結していくことであり, 特に調性音楽の和声を, 音階各音度上の和音の機能の観点から説明したものを機能和声と呼ぶ. 本研究は, クラシック曲に対して和声付与を予め行った教師データを基に, 機械学習によって, クラシック曲に対する機能和声を考慮した自動和声付与を行うシステムを構築することを目的としている. 自動和声付与システムの構築には, 複数の課題が存在する. 例えば, 和音区間の認識や非和声音の扱いなどが挙げられる. 特に, クラシック曲は和音区間やその区間で響いている和音の判断が難しく, さらに, 転調が行われたか, それとも借用和音が使われているかの判断なども必要となる. オーディオデータを扱うのか, シンボリックデータ (MIDI や楽譜など) を扱うのかで, これらの課題に対する対処方法は変化する. 今回は, 上述したような自動和声付与システムを構築する際の課題について, 過去の文献を参照しながら, オーディオデータを用いた和声付与, シンボリックデータを用いた和声付与の両面から, 議論を深めたいと考えている.