

研究報告

機械学習を用いた楽器演奏音響データからのフレージング要素の検出

小長谷 恵美, 塩田 さやか, 貴家 仁志, 安藤 大地
Megumi KOHASE, Sayaka SHIOTA, Hitoshi KIYA, Daichi ANDO
東京都立大学
Tokyo Metropolitan University

概要

フレージングとは楽譜に記載されている音符、休符に演奏表情をつけることであり、構成要素には音量、ピッチの微細な変化、速度、アーテキュレーション（スタッカート等）等がある。楽器演奏ではこれらに捉えづらいう微細な変化を付けることが求められている。これらには良し悪しが存在するが、構成要素が複数あり変化が微細であることから、良し悪しを具体的に説明することは困難である。そこで本研究では、演奏を要素ごとに分析することで部分的な説明を可能にし、演奏者に具体的なフィードバックを与えることを目標とした。まず構成要素の1つである発音タイミングに着眼した。分析方法は、演奏家から楽器音響データを録音、スペクトログラム化した学習データを作成、畳み込みニューラルネットワークを用いて音源の良し悪しを分類した。その結果、画像判別は可能だが、発音タイミング以外に着眼して判別が行われたことが分かった。

1. 導入

1.1. 教育現場でのフレージングの指導と問題点

音楽教育現場ではフレージング指導に抽象的な言葉（例：池を飛び越えるように、馬が駆け上がるように演奏しなさい等）が多用されている。なぜなら、フレージングには複数の要素（リズム、速度、アーテキュレーション等）があり、それらに微細な変化を付けるため、具体的に説明することが困難であるからだ。これらを理解するには、指導者の言葉に対して生徒が演奏し、それに対して指導者よりフィードバックをもらう、という工程が長期間必要となる。しかし楽器練習の大半は個人練習が占めており、自身の練習が指導者の言葉に添えているのか練習者（特に初心者）自身が判定することは難しく、効率的な個人練習が行えない場合がある。また長期間レッスンを受講していても、未経験のことや、イメージし辛いことなど指導者と生徒の人生経験の差より理解することが困難な場合もある。そのため多くの楽器演奏者にとってフレージングを習得

することは大きなハードルとなっている。そこで本研究ではフレージングを構成要素ごとに分析することで、部分的な説明を可能にし、演奏者に具体的なフィードバックを与えることを目標とする。

1.2. 本研究全体のアプローチ

本研究では、フレージングの構成要素の1つである「発音タイミング」に着眼して研究を行う。なぜなら発音タイミングはグルーブ感に大きく影響するからである。楽器は音を発音する操作を行ってから、実際に発音するまでに時間がかかる。この発音までに要する時間は楽器によって異なり、打楽器、管楽器、擦弦楽器の順で遅くなる。本研究では管楽器と打楽器での発音タイミングに着眼して研究を行う。管楽器は自身の息や唇、舌を用いて発音するため、発音操作やタイミングを視覚的に見ることが難しく、他楽器と比べより楽器演奏者の技量が求められ、発音操作が演奏表情にそのまま結びついている。そのため演奏者へのフィードバックの重要性が高いと考え、管楽器に着眼した。プロ演奏家は音が立ち上がり切るまでの時差を感覚的に考慮し、実際に音を出す瞬間よりも以前に発音操作に入ることが出来る。しかし多くの初心者はその時差を感覚的に掴むことが出来ず、立ち上がりきれっていない音で発音をしてしまう。そのためグルーブ感が崩れ、演奏が遅れているように聞こえる。そこで発音タイミングの良し悪しを分析し、どの程度時差を考慮して発音操作に入ると良い発音タイミングになるのか考察を行った。

1.3. 先行研究と本研究全体の位置付け

これまでにメトロノームなどリズムの指標から演奏が微細に前後することがグルーブ感に与える影響について Keil(Keil, 1995) や宮丸ら(宮丸, 2017) によって研究されている。Keil はメトロノームと演奏が微細に前後することがグルーブ感の表現において重要だとし、PD 理論 (Participatory Discrepancies theory) を提唱し

ている。宮丸らはドラムを用いて、リズム指標と演奏のズレについて印象評価実験を行った。時間のズレが与えるグルーブ感への印象が最も良いのは 20ms 後に逸脱させることであり、30ms 前後に逸脱する場合はグルーブ感が損なわれると結論づけた。どちらの検討もリズムの指標と演奏のズレに着眼している。

既存研究は発音タイミングのズレに関して議論しているが対象は打楽器のみである。一方、本研究は、最終目標を管楽器で行うことが既存研究とは異なる。そのために、単純な発音タイミングだけではなく、管楽器特有の「演奏家がコントロールする音の立ち上がり」を演奏表情として用いる」ことに対応しなければならない。そこで本研究では、スペクトログラムを用いて、画像処理的に演奏を分類するという手法を取る。スペクトログラムは楽器の同定などに使われており楽器音響での実績がある。もし楽器音響で発音タイミングの分類を行えるならば、教育現場での実装としてはスマートフォン 1 台があればよく、特殊な改造を施した管楽器も必要ない。基礎研究と教育現場への応用の 2 つの視点において、本研究が取り扱うスペクトログラムで発音タイミングを検討するという手法は、ユニークである。

1.4. 本稿の位置づけ

本稿では、本研究全体の目標である「スペクトログラムによる画像処理的な発音タイミングの分類」の基礎研究として、ピアノを対象として行った実験について述べる。管楽器の分類は難しいことが予想されるため、まず既存研究でも取り扱われている打楽器の発音タイミングのズレが演奏に与える影響をスペクトログラムでも分類できることを確かめるためである。

2. 実験環境

2.1. 対象としたデータ

普段の練習環境を再現するため、メトロノームをピアノの上に置き実験を行なった。また演奏家の耳元の高さにマイクを設置し、観客ではなく演奏家が聞いている音を録音した。(サンプリング周波数:44100Hz,16ビット)今回はメトロノームを 1 小節に 1 回鳴らし、正しい速度の指標を作った。対象とした曲はバロック時代など 1 小節内で拍と拍の間隔に変化が少ないもの、特に小節の頭拍はメトロノームに合うことが良いとされる曲を選曲した。

2.2. 分析手法

被験者の発音タイミングがメトロノーム音とどの程度ズレることが良いフレージングに聞こえるかに着眼

し、録音を行った。実験では練習過程のデータを収集することが目的であったため、被験者には譜読みの時間は与えなかった。被験者が曲を演奏できるようになったと感じたら、実験監督者に合図を送り、実験監督者も同様に感じたら次の曲に移行した。計 20 曲の演奏データを収録した。それらのデータに演奏の良し悪しラベリングを行った後、メトロノームから前後 1 秒を切り出した。その後、短時間フーリエ変換を用いて音源をスペクトログラムに変換、画像データを 3000 枚作成した。今回は良し悪しの差が顕著な音源を選択したため、3000 枚中 800 枚を学習データとして用いた。短時間フーリエ変換は窓関数：ハン窓、窓サイズ：2048、ゼロ埋め倍数：2 で実装した。また画像解像度は画像サイズ：173×173 とし、機械学習を用いて画像分類を行った。画像分類のモデルとして、2 次元畳み込みネットワーク (2DCNN, 2-dimensional Convolution Neural Network) を用いた。これは画像に特化したニューラルネットワークであり、今回はスペクトログラムの形状とメトロノームの位置関係を探るために使用し、発音タイミングとメトロノーム音の最適なズレを探った。

対象としたデータの典型的なスペクトログラムを図 1 に示す。

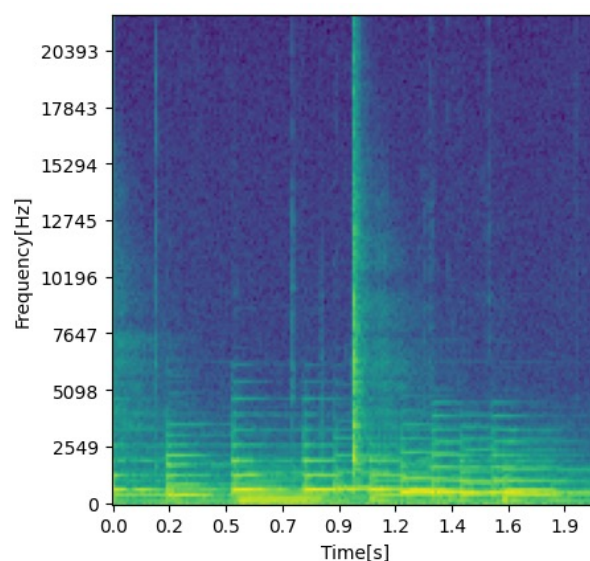


図 1: ピアノの演奏のスペクトログラム。中央部に立っているのがメトロノームの音。

3. 実験結果

結果としては、loss0.22・accuracy0.93 という高確率で画像を分類できるモデルを作成できた (図 2)。しかし後述の判断根拠の分析から、この分類は発音タイミング以外により分類が行われたことが判明した。

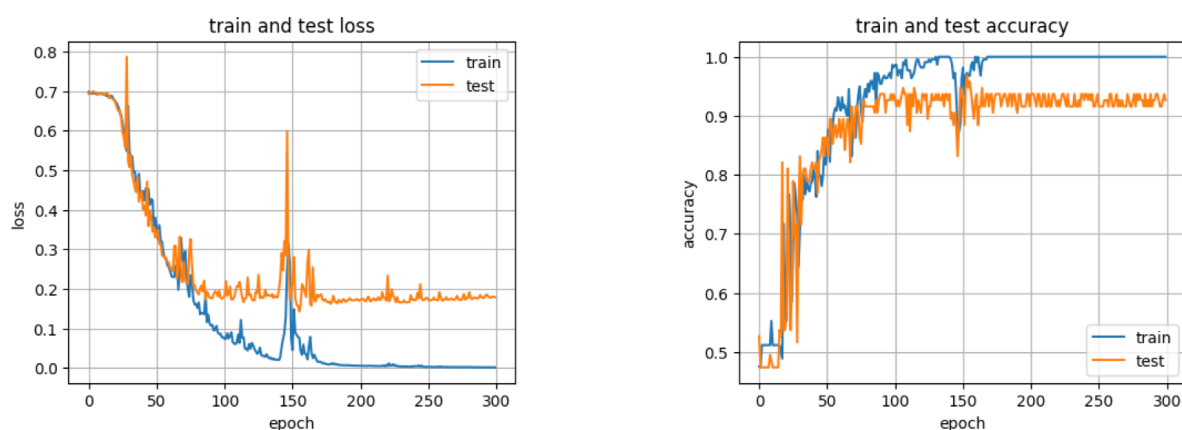


図 2: CNN による分類の結果。分類の学習が進んでいることは確認できる。

4. 考察

5. 学習の結果と分類の根拠となる情報

今回の機械学習モデルでは、画像よりメトロノームと実際の演奏のズレを具体的な位置、数値を指定することなく検知することができた。分類そのものは高精度でできていることから、スペクトログラム当手法には有効性があると著者らは考えた。

モデルの評価を行うため、Grad-Cam(学習結果の判断根拠を局所的に示す手法)を用いて判断根拠を可視化した。図3、4に、それぞれ良い発音タイミングと悪い発音タイミングの、演奏のスペクトログラムとヒートマップを示す。結果として、発音タイミングのズレに着眼したた判別が行われていないことが分かった。

著者らの当初の予想では、演奏家がある程度共通した認識として持っている「発音開始後、振幅が一番大きくなった場所を拍頭に合わせる」ことがスペクトログラムから観察でき、CNNも同じように判別するものと考えていたが、実際はそのような「発音開始タイミングの微細な前ノリ」とも呼ぶべき状態を基準にCNNが判断をしているとは言えなかった。

原因として考えられることは、実験対象であるピアノは打楽器であり発音タイミングが短いため、CNNの畳み込みノードの入力画像サイズである173x173では、時間的な解像度が足りなかったことが予想される。

また、また画像内のメトロノーム音は常に同じ位置にあり、特徴を学ぶ機械学習において、メトロノーム周辺ピクセルの学習が不十分である可能性もある。

画像サイズについては、本研究の最終目標は管楽器であり発音開始後、振幅が一番大きい場所に達するまでの時間が、ピアノのような打楽器と比較し遅いため、この問題は管楽器に適用した時には問題にならないと考えているが、引き続き検討が必要である。また画像内のメトロノームの位置については、演奏とメトロノームを分離し録音し、様々な形状の演奏をスペクト

ログラムとし学習データとすることで、解決できると考えている。

5.1. 今後の展開

今後の展開としては、再実験を行い、楽器演奏とメトロノームを分離して録音し、学習データとすることを考えている。これにより、より高度なフレー징ングの良し悪し判定が行えると考えている。

また今回の研究で機械学習を用いて、画像よりメトロノームと実際の演奏のズレを具体的な位置、数値を指定することなく検知することが出来たため、今後はフレーズの頭拍と中身(拍と拍の間の位置関係)の両方に着眼して研究を進める。そのため、メトロノームを頭拍のみを鳴らしながら演奏する音源と、メトロノームを細かく鳴らしながら演奏する音源の2種類を録音した。この2つを見比べながら、より音楽的な発音タイミングのズレを分析する。またラベリングについても専門技能を持つ複数の音楽家に依頼し、より精度の高い良し悪しのラベリング付けを行う。

6. まとめ

フレー징ングとは楽譜に記載されている音符、休符に演奏表情をつけることである。また構成要素が複数あり変化が微細であることから、良し悪しを具体的に説明することは困難である。そこで本稿では、演奏を要素ごとに分析することで部分的な説明を可能にし、演奏者に具体的なフィードバックを与えることを目標とした。今回は構成要素の1つである発音タイミングに着眼した。分析方法は、演奏家から楽器音響データを録音、スペクトログラム化した学習データを作成、畳み込みニューラルネットワークを用いて音源の良し悪しを分類、特徴を考察した。学習結果としてはlossが0.22, accuracyが0.93という、高確率で発音タイミング

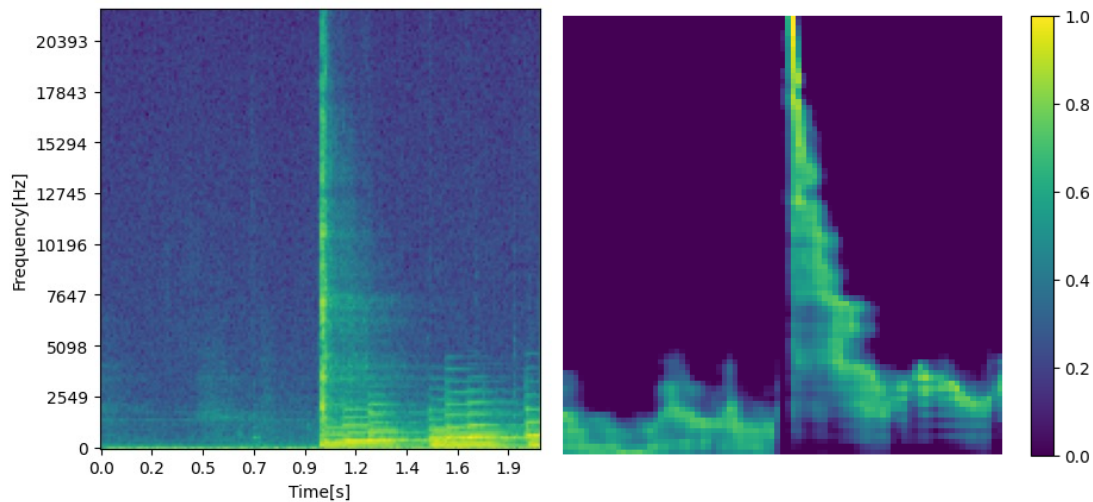


図 3: 人間が良い発音タイミングだと判断した演奏のスペクトログラム（左）と Grad-Cam のヒートマップ（右）

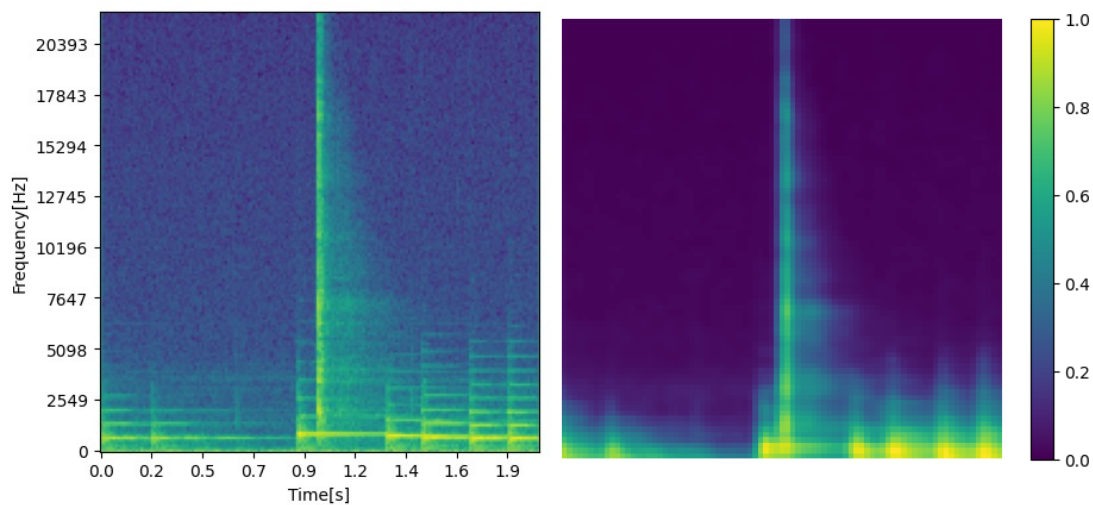


図 4: 人間が悪い発音タイミングだと判断した演奏のスペクトログラム（左）と Grad-Cam のヒートマップ（右）

の良し悪しを判別することが出来た。またメトロノーム（テンポの指標）と演奏のズレが具体的な数値を決めることなく、画像から判別することが出来た。しかし Grad-Cam(学習結果の判断根拠を局所的に示す手法)を用いて分析すると、発音タイミング以外に着眼して画像分類が行われていることが判明した。原因として考えられることは、今回対象としたピアノは音の立ち上がりが比較的早く、画像サイズ 173x173 では発音タイミングを確認することが出来なかったこと、また画像内のメトロノームが常に同じ位置にあるため、画像の特徴を学ぶ機械学習において、メトロノーム周辺ピクセルの学習が上手く進んでいない可能性があることが挙げられる。画像サイズについては、本研究の最終目標はクラリネット（ピアノと比べ音の立ち上がりが遅い）を対象とした分析であるため、当手法は有効性があると考えられる。また画像内のメトロノームについて

は、演奏とメトロノームを分離して録音することで解決できると考える。（録音済み）今回はフレーズの頭拍のみに着眼し、ズレを機械学習において検出することができたため、今後はフレーズ内の拍と拍の間の位置関係に着眼し、より高度なフレージングの要素解析を行っていく。

7. 参考文献

- C.Keil (1995) “The theory of participatory discrepancies: A progress report”, *Ethnomusicology*, vol.39, pp.1—19.
- 宮丸友輔, 江村伯夫, 山田真司 (2017) 「ポピュラ音楽のドラムス演奏におけるグルーブ感の研究」『日本音響学会誌』73 巻 10 号, pp.625-637.

8. 著者プロフィール

小長谷 恵美 (Megumi KOHASE)

東京都立大学 システムデザイン研究科修士課程に所属。音響データを用いた演奏分析、初学者への演奏フィードバックについて研究をおこなっている。2019年に東京音楽大学 器楽専攻に入学。クラリネット演奏を四戸世紀に師事。2023年東京都立大学 システムデザイン・インダストリアルアート研究科に入学、安藤大地に師事。



この作品は、クリエイティブ・コモンズの表示 - 非営利 - 改変禁止 4.0 国際 ライセンスで提供されています。ライセンスの写しをご覧になるには、<http://creativecommons.org/licenses/by-nc-nd/4.0/> をご覧ください、Creative Commons, PO Box 1866, Mountain View, CA 94042, USA までお手紙をお送りください。