

研究報告

ミクスト音楽のためのフルート奏法自動判別手法の検討 Machine Learning Based Classification of Flute Playing Techniques for Mixed Music

ブロシエック・ニコラ, 田中翼

Nicolas BROCHEC, Tsubasa TANAKA

東京藝術大学音楽環境創造科

Department of Musical Creativity and the Environment,
Tokyo University of the Arts

概要

ミクスト音楽 (musique mixte) においては、生楽器演奏に対してリアルタイムにデジタル音響エフェクトが施されるケースが多い。通常、こうしたエフェクトは、セクションごとに固定され、コンピュータの手動操作や演奏者によるフットペダル操作でエフェクトの切り替えがなされるが、操作者への負担が発生する。本研究が目指すのは、機械学習に基づき、奏法に応じたエフェクトの自動的な切り替えを行えるインタラクティブ・システムを構築し、こうした操作者の負担を軽減することである。また、このようなシステムが実現すれば、よりリアルタイム性を活かしたミクスト音楽の創作的可能性が広がるだろう。機械学習に基づく奏法判別の既存研究においては、未だ実用化にとって十分な精度には達していない。特に管楽器については研究例が見当たらなかった。そこで、本稿ではフルートの奏法判別を対象とし、複数の機械学習手法を適用した実験結果を報告する。

In mixed music (musique mixte), digital sound processing is often applied to live instrumental performances in real time. However, in most cases, such processing techniques are fixed for each section and switched by manual computer operation or the performer's foot pedal operation, which burdens the operator. This research aims to build an interactive system that can automatically switch the processing techniques and their parameters according to the type of playing techniques of the instrument, thereby reducing the burden on the operator. Also, realizing such a system will expand the creative possibilities of mixed music to those that further take advantage of its realtimeness. In existing studies, sufficient accuracy has not been achieved yet for practical use. In particular, we have

not found research examples of wind instruments for this purpose. Therefore, in this paper, we tackle the problem of automatic classification of flute playing techniques and report the experimental results of applying several machine learning methods.

1. はじめに

生楽器演奏と電子音を組み合わせた現代音楽としてのミクスト音楽 (musique mixte) (Boulez 1981-1988; Manoury 1993) の領域において、コンピュータの計算速度の向上の恩恵を受け、1980年頃から人間による生楽器演奏に対してリアルタイムにデジタル音響エフェクトをかけるような楽曲が制作されるようになった。しかし通常、音響エフェクトはセクションごとに固定されており、エフェクトの切り替えはコンピュータの操作や演奏者のフットペダルなどで行われることが多く、操作者への負担となっている。これを自動化するために、演奏中の楽譜上の位置を追跡し、自動的にエフェクトの切り替えを行う Antescofo (Etchevest 2015) のような自動伴奏システムも開発されてきたが、演奏位置を推定するために楽曲ごとに楽譜情報の入力が必要になるなど、別のコストがかかる弱点がある。

そこで本研究が提案するのは、楽器奏法の違いの自動判別システムを構築し、奏法に応じたリアルタイムな音響エフェクトの自動的な切り替えを行えるようにし、操作者への負担や楽譜入力のコストを低減させることである。また、それが実用化できれば、ミクスト音楽の創作的可能性が、音響エフェクトのリアルタイム性をより活かしたものと広がるだろう。

関連する既存研究としては、音響特徴量を用いて奏法イベントの検出を行うもの (プログラムは人間が行う) (Malt 2009) や、主成分分析で認識に有効な特徴量

を自動選択するもの (Malt 2015) がある。こうした奏法の検出目的で機械学習を用いた手法が研究されるようになったのは比較的最近のことである。Ducher らは畳み込みニューラルネットワーク (CNN) を用いたチェロの奏法判別を試みたが、未だ実用化に十分な精度には達していない (Ducher 2019a; Ducher 2019b)。

管楽器については関心の近い研究例が見当たらなかったが、第一著者のプロシエックはフルートの演奏者でもあり、フルートとエレクトロニクスのための楽曲の作曲を計画している。そのため本稿では、研究の最初のステップとして、機械学習に基づくフルートの自動奏法判別を試み、精度の検証を行う。

2. フルートの奏法判別

個々の種類の楽器においては、さまざまな奏法がある。フルートの場合、レガートやスタッカートといったアーティキュレーション、ビブラートやフラッターなどの奏法、スラップ、キー・パーカッションなどの特殊奏法、といったさまざまな演奏法の差異を作り出すことができ、場面に応じて使い分けられる。

フルートの音は主にマウスピースに吹き込まれた空気の摩擦から生じる。吹き込まれた空気の数と特性の違いは奏法の変化をもたらす (Fletcher 1998)。例えば、空気の数や音量の変化やオクターブの変化を引きおこし、舌による空気の流れの調整はスタッカート、スラップ、フラッターなどの違いを生み出す。

フルートの出せる音高は基本的には平均律であり、指で押さえるキーの組み合わせ (運指) により半音階を実現することができる。マルチフォニック (重音奏法) やクォータートーンのような特殊奏法も通常とは異なる運指によって実現することができる (Levine 2002)。こうした様々な奏法を、オーディオデータとしてリアルタイムに自動判別することが本研究の目的である。

3. 学習データ

3.1. 音源の選択

さまざまな奏法の判別をシステムに学習させるためには、奏法ごとに録音したデータセットが必要となる。本研究では楽器音の既存のサンプリング音源 FullSOL (Cella et al. 2020) をデータセットとして用いる。FullSOL は多数の奏法を含む音源であり、フルートにおいては 27 種類の異なる奏法が収録されている。

3.2. 奏法の選別

FullSOL の音源では、それぞれの奏法の録音だけでなく、一つの奏法から別の奏法へ移行する中間状態も

1	aeolian	11	pizzicato
2	aeolian-and-ordinario	12	play-and-sing
3	discolored-fingering	13	play-and-sing-unison
4	flutterzunge	14	sforzando
5	fortepiano	15	staccato
6	harmonic-fingering	16	tongue-ram
7	key-click	17	trill-major-second-up
8	multiphonics	18	trill-minor-second-up
9	note-lasting	19	whistle-tones
10	ordinario	20	whistle-tones-sweeping

表 1: 対象とする全 20 カテゴリーのフルート奏法

収録されている。例えば、エオリアン・トーン (通常音と同時に歌口や歯の間から漏れるノイズを発する奏法) からオーディナリオ (通常の音) までの移行などがある。本稿においては、まずは奏法カテゴリーの判別に集中するため、中間状態は用いないこととする。それにより、27 種類の録音音源のうち 20 種類のみを用いる。したがって、解くべき問題は 20 クラスの分類問題である。20 種類の分類ラベルのリストを表 1 に示す。

3.3. データの形式

データセットの形式は、先行研究 (Ducher 2019a) に倣い、サンプリングレートを 24000Hz、FFT 時間窓長を 2048、ホップサイズを 512 とし、60 フレーム分のメルスペクトログラムをデータ単位とする。各オーディオファイルにおける発音前の無音区間を取り除いた後、60 フレーム毎に分割する。ピチカートやスタッカートのように短い音源においてはゼロパディングを行い長さを調整した。

4. 奏法判別の手法

4.1. 機械学習アルゴリズム

奏法判別の手法としては、sklearn ライブラリ (Pedregosa et al. 2011) から複数の機械学習アルゴリズムを用いた。基本的にアルゴリズムのパラメータは初期設定のままを用いた (ただし MLP の *max_iter* のみ初期設定の 100 ではなく 1000 とした)。使用したアルゴリズムは、サポートベクターマシン (SVM、カーネルとしては RBF、Poly、Linear の三種類)、決定木、k 近傍法 (k-NN)、ランダムフォレスト (RF)、決定木ベースの AdaBoost、LightGBM、多層パーセプトロン (MLP) である。

学習データとテスト用データの分割のため、データセットから各奏法ごとに 25 パーセントをテスト用に

ランダムに振り分けた。ただし、ランダムな振り分けによる偶然的な偏りを是正するため、精度は10回の結果の平均をとる。

4.2. データ拡張

学習データ量の不足を補うためには、手持ちのデータに何らかの変形を加えたものをも用いてデータ量を増やすデータ拡張 (data augmentation) が有効であることが多い。本実験では、チェロに関する奏法分類の先行研究 (Ducher 2019a; Ducher 2019b) と同様の三種類のデータ拡張の手法を用いる。具体的には、ピッチシフト、リバーブ付加、ノイズ付加の三種類である。

ピッチシフトにおいては、オリジナルの録音のピッチの440ヘルツに対し、±20ヘルツの範囲でランダムにずらしたものをを用いる。これは、調律の微妙なずれに対応するためである。リバーブ付加においては、オリジナルの音源に対して一定の深さのリバーブをかけたものをを用いる。これは、演奏環境によって残響が異なることを想定したものである。ノイズ付加においては、オリジナルの録音に対してホワイトノイズを付加したものをを用いる。これは、演奏音をマイクで把握する際などのノイズを想定したものである。ノイズの大きさはファイルごとにランダムに設定する。

このデータ拡張により、学習データの数はオリジナルのデータの4倍となる。データ拡張後の学習データのファイル数は一万強である。

5. 実験結果

5.1. 精度

前章で述べた、それぞれの手法に対して奏法判別を行い、データ拡張なしの場合とありの場合でそれぞれ10回の平均正解率を算出した (表2)。

20カテゴリーであることから、ランダムな分類では5%がベースラインとなることから、データ拡張なしの場合でも、おおむね70%台以上であり、最高で92.3% (SVM-Poly) と、機械学習が有効に働いていることが確認できる。データ拡張ありの場合はおおむね80%台以上、最高で92.4% (SVM-Poly) と、データ拡張なしの場合に比べて高い分類精度となる傾向が確認できる。

ただし例外として、Decision Tree と AdaBoost の場合は精度が50%台にとどまっている。また、SVM-Linear と MLP の場合ではデータ拡張によって精度が下がっている。これは、手法の分類能力の限界や、パラメータの調整不足が原因として考えられる。

手法	データ拡張なし	データ拡張あり
SVM-Rbf	81.1%	92.2%
SVM-Poly	92.3%	92.4%
SVM-Linear	89.0%	84.9%
Decision Tree	53.2%	56.1%
k-NN	76.7%	90.2%
Random Forest	79.2%	85.0%
AdaBoost	53.2%	56.1%
LightGBM	84.5%	87.8%
MLP	86.5%	85.1%

表2: 各手法の正解率 (10回の平均)

5.2. 本実験の限界と今後の課題

本実験をうけて、今後は、誤分類の原因の分析、さらなるデータの整備やパラメータ調整、深層学習を用いた分類器の改良などによって、完璧に近い精度への改善を試みたい。

ただし、本実験においては、学習用とテスト用のコーパスが同一のサンプリング音源をデータ分割したものであり、別のサンプリング音源や生演奏をテストデータにした場合の検証はまだ行っていない。別の音源をテスト用データとした場合、先行研究 (Ducher 2019a; Ducher 2019b) に示されたように、精度が落ちることが予想される。本研究を実用化するには、別の音源による実験の方がより実情に近い可能性もあるため、今後はそちらについても実験を行いたい。

さらに、今回の実験ではメルスペクトログラムのみを入力データとしたが、その他の特徴量を用いることで精度が改善する可能性も検討したい。

6. 結論

本研究では、ミクスト音楽のための新しいシステムの開発の一環としてフルート奏法の自動判別の実験を行った。そこでは奏法判別を20カテゴリーの分類問題としてとらえ、複数の機械学習手法で分類器を構築し、精度を比較した。チューニングをほとんど行わなくても高い分類精度が実現し、最も良い分類器の精度は90%を超えた。また、データ拡張によって、多くの機械学習手法で精度の改善が見込めることが確認できた。

今後の課題としては、誤分類の原因分析、さらなるデータの整備やパラメータ調整、深層学習を用いた分類器の改良、メルスペクトログラム以外の特徴量の検討、複数のサンプリング音源の使用などがある。

分類精度の問題の他に、奏法判別システムの実用化に向けた他の課題もある。例えば、実演時のシステムの使用においては、時々おこる分類ミスの影響を排除するため、移動平均で頑健性を高めるなどの工夫が必

要になると思われる。また、リアルタイムの奏法判別にかかる時間遅延がどこまで許容できるかを検証する必要があり、アルゴリズムごとの時間遅延も評価すべき項目となるであろう。

7. 参考文献

- Cella, Carmine Emanuele, Ghisi, Daniele, Lostanlen, Vincent, Lévy, Fabien, Fineberg, Joshua, Maresz, Yan. 2020. “OrchideaSOL: a dataset of extended instrumental techniques for computer-aided orchestration“. in *ICMC 2020*.
- Ducher, Jean-François, Elsing, Philippe. 2019a. “Apprentissage profond pour la reconnaissance en temps réel des modes de jeu instrumentaux“, Journées d’Informatique Musicale, May 2019, Bayonne, France.
- Ducher, Jean-François, Elsing, Philippe. 2019b. “Folded CQT RCNN for Reals-Time Recognition of Instrument Playing Techniques“ 20th International Society for Music Information Retrieval Conference, Delft, The Netherlands.
- Echeveste, J.M. 2015. “Un langage de programmation pour composer l’interaction musicale : la gestion du temps et des événements dans Antescofo“ Thèse de doctorat de l’université Pierre et Marie Curie, Paris.
- Fletcher, Neville H., Rossing, Thomas D. 1998. “The Physics of Musical Instruments“ Springer, 2nd edition, 511–512.
- Levine, Carin, Mitropoulos-Bott, Christina. 2002. “The Technique of Flute Playing“. Bärenreiter-Verlag.
- Malt, Mikhail, Gentilucci, Martha. 2015. “Real Time Vowel Tremolo Detection Using Low Level Audio Descriptors“ arXiv:1511.07008 [cs], November 2015. arXiv: 1511.07008.
- Malt, Mikhail, Jourdan, Emmanuel. 2009 “Real-Time Uses of Low Level Sound Descriptors as Event Detection Functions Using the Max/MSP Zsa.Descriptors Library“ in *SBCM 2009*, Recife, Brazil.
- McFee, Brian, Raffel, Colin, Liang, Dawen, Ellis, Daniel, McVicar, Matt, Battenberg, Eric, Nieto, Oriol. 2015. “Librosa: Audio and Music Signal Analysis in Python“ in *the Proceedings of the 14th Python in Science Conference. (SCIPY)*

Pedegrosa, Fabian, Varoquaux, Gael, Gramfort, Alexandre, Michel, Vincent, Thirion, Bertrand, Grisel, Olivier, Blondel, Mathieu, Prettenhofer, Peter, Weiss, Ron, Dubourg, Vincent, Vanderplas, Jake, Passos, Alexandre, Cournapeau, David. 2011. “Scikit-learn: Machine Learning in Python“. *Journal of Machine Learning Research* 12. 2825–2830.

8. 参考作品

- Boulez, Pierre. 1981-1988. *Répons*. Universal Edition, 1988.
- Manoury, Philippe. 1993. *En écho*. Édition Durand, 1994.

9. 著者プロフィール

ブロシェック・ニコラ

Broche Nicolas はフランスの作曲家、ソフトウェア開発者である。パリ第八大学修士課程（音楽学専攻）、ストラスブール大学大学院および音楽院修士課程修了（作曲専攻）。現在、東京藝術大学音楽環境創造科（後藤英研究室）に研究生として在籍中。作品はフランスを始めスペイン、オーストリア、ルクセンブルクなどヨーロッパ各地で演奏されている。画像から着想を得ることが多く、ノイズと和声を組み合わせる作品作りを目指している。現在の研究テーマは機械学習のミクスト音楽への応用。

田中翼

東京藝術大学音楽環境創造科非常勤講師。JST ACT-X 研究員（AI 活用学問革新創生領域）。AI 技芸研究会代表。家庭教師（科研費、学振、国費留学、音楽プログラミングなど）。専門は音楽情報科学、アルゴリズム作曲、数理的音楽理論。京都大学理学部で数学、東京大学情報理工学系研究科で情報科学を学び、東京藝術大学先端芸術表現科で博士号を取得。2014年より博士研究員として渡仏し、IRCAM、パリ左岸＝ジュシュー数学研究所、ソルボンヌ大学で研究する傍ら音楽院で作曲を学ぶ。2021年に帰国し現職。



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/> をご覧頂るか、Creative Commons, PO Box 1866, Mountain View, CA 94042, USA までお手紙をお送りください。