

DNN を用いた音源のギター分離・評価

Study on the sound source separation and sound source evaluation of guitars using DNN

萬谷 亮太[†] 土居 茂雄[‡] 中村 嘉彦[‡] 山本 椋太[‡]
 Ryota Many Shigeo Doi Yoshihiko Nakamura Ryota Yamamoto

1. 緒論

音源分離は活発な研究が行われている分野である。音源分離の研究は以前から行われており、機械学習の発展に伴い従来手法と機械学習の融合によりさらに活発な研究が進められている。代表的な音源分離の事例としてはノイズ除去や複数話者の分離、楽器の分離などがある。楽器の分離については、一般的なものとして伴奏、ボーカル、およびいくつかの決まったタイプの楽器の分離の研究が行われている[1]。しかし、同じ種類の楽器に関する音源分離、例えばアコースティックギターとエレキギターの音源分離に関しては、ほとんど研究事例が見られないのが現状である。

LALAI.AI[2]のような音源分離ツールは広く商業的に利用されており、音源分離ツールにはアコースティックギターとエレキギターの音源分離機能が備わっているものもある。しかし、これらのツールに関する論文などは公開されておらず、その手法を再現することは困難である。そこで、本研究ではアコースティックギターとエレキギターの分離を、音源分離ライブラリの Spleeter[3]を用いて学習モデルを作成することで実現を目指す。

最終的には、学習モデルを作成と分離性能の評価を目指す、その前段階としてデータセットに用いる MIDI 音源の波形比較を行うことで、MIDI 音源を学習に使用できるかを判断する。

2. Spleeter について

Spleeter は、音源分離タスクの支援を目的に、Deezer 社が開発したオープンソース音源分離ツールである。このツールは、目的の音源を分離するために自分のデータを用いてモデルを学習させることが可能であり、そのためにはクラスに設定する楽器ごとのデータとそれを合体させたミックスデータが必要となる。

Spleeter は、研究者や音源分離のタスクを行う人にわかりやすく設計されており、パラメータや設定の変更が可能である。また、このツールは学習アルゴリズムが複数搭載されており、CNN や Bi-LSTM[4] などユーザは必要に応じて選択することができる。

3. 研究方法

今回はアコースティックギターとエレキギターの混合音源を分離することを目標にモデルを作成する。

今回学習に用いる音源データは、MIDI ベースの合成音源が有用な代替手段であることを実証した Manilow らの研究結果[5]から、実際の楽器音がサンプリングされ、仮想環境で演奏できるようになっている MIDI 音源を採用する。

[†] 苫小牧工業高等専門学校 創造工学専攻 情報エレクトロニクス系 Advanced Course, Department of Engineering for Innovation, National Institute of Technology, Tomakomai College

[‡] 苫小牧工業高等専門学校 創造工学系 情報科学・工学系 Department of Engineering for Innovation, National Institute of Technology, Tomakomai College

まず、データは楽器の音階である C から B までのコード音を MIDI 音源で作成し、いくつかの音源加工を施してデータ数を増加させた後、それを Spleeter に入力してモデルを学習させる。その後学習させたモデルを用いて音源を分離し、その結果を評価する。

Manilow らの研究結果から MIDI ベースの合成音源が有用な代替手段であることが実証されているが、MIDI 音源には様々な種類があり、研究で使用される MIDI 音源が学習に用いることのできる特徴があるかは不明である。そのため実際の楽器の波形と、MIDI 音源の波形を比較・考察した上で、MIDI 音源を学習に使用できるかを判断することとした。

3.1 データセットの収集方法

音源データは、エレキギターとアコースティックギターそれぞれにおいて、DAW と呼ばれる音楽編集ソフトウェアの FL Studio [6]で、電子楽器音生成プラグイン(以降、MIDI プラグインと呼ぶ)にて作成した単音コードの音源(以降、MIDI 音源と呼ぶ)を C から B までとその半音を含む 12 音階用意した。

また、DAW では音源編集用プラグイン(以降、編集プラグインと呼ぶ)と呼ばれる拡張機能を用いることで、波形として表示することができる。

楽器のエフェクトについては、今回は考慮しないこととし、エフェクトをかけないクリーントーンで作成した。

MIDI 音源のうち、アコースティックギターの MIDI プラグインは Ample Sound 社が提供している、Martin 社の D-41 をモデルにした Ample Guitar M Lite [7]を、エレキギターの MIDI プラグインは Impact Soundworks 社が提供している、Fender 社の Stratocaster をサンプリングした Shreddage 3 Stratus Free[8]を使用して、データセットとなる音源を作成した。各音源のサンプルレートは 44.1kHz に統一し、仮にデータのサンプルレートが 44.1kHz ではない場合には、作成したリサンプリングプログラムを用いて統一した。データセットとなる音源は、すべての音源をモノラルに統一した。

3.2 MIDI 音源の比較方法

実際の楽器音と、実際の楽器からサンプルされた MIDI 音源の比較は DAW の編集プラグインによる音源波形の記録と、各音源およびその波形の音量正規化、波形解析ツール Sonic Visualiser[9]を用いた。音量の正規化は、Edison という編集プラグインを用いて、最大音量を 0db に設定することで行った。波形は、時間あたりの周波数ごとの音の強さをプロットしたスペクトログラムで比較した。

実際の楽器音は、秋田大学の田中研究室が公開している Guitar chord sound dataset[10]より、アコースティックギターとエレキギターの C コードの録音音源を使用した。

MIDI 音源は、3.1 で用意したアコースティックギターとエレキギターのプラグインから Cmajor コードの音を用いた。

3.3 学習結果の評価方法

音源分離の研究においてモデルの評価方法には様々なものがある。例としては信号対雑音比(SNR)、信号対干渉比(SIR)、信号対歪み比(SDR)、等々存在する。本研究では、学習したモデルによる分離結果が、入力音源と異なり、音源に分離モデルが影響を与えられているかを確認するためこれらの評価指標は用いる前に、主観的な経験則から出力音源に変化があるかを確認する。具体的には、アコースティックギターは、弦の振動がボディ内部での共鳴による自然な音色、エレキギターは、電気信号変換による長いサステイン(音の持続時間)を持つという特徴を持つ。これらを筆者の経験則から得た各楽器の特徴とし、分離された各システムがそれらを正しく分離しているか、という観点で比較する。また、時間あたりの振幅をプロットした波形や、時間あたりの周波数ごとの音の強さをプロットしたスペクトログラムから出力されたファイル同士を比較評価する。

4. 実際の楽器と MIDI 音源の波形比較結果と考察

DAW にて音量の正規化を行ったうえで、実際の楽器と MIDI 音源をスペクトログラムで表示したときの比較結果を図 1、図 2 に示す。図の縦軸は周波数、横軸は時間を表している。

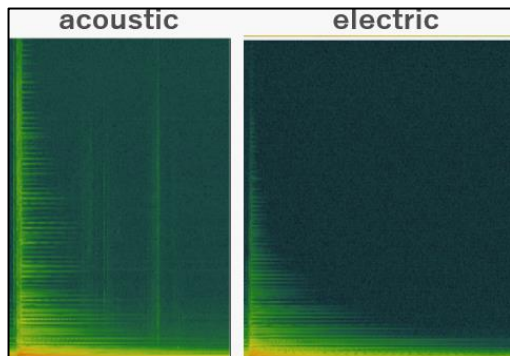


図 1 実際の楽器音源のスペクトログラム
(左:アコースティックギター, 右:エレキギター)

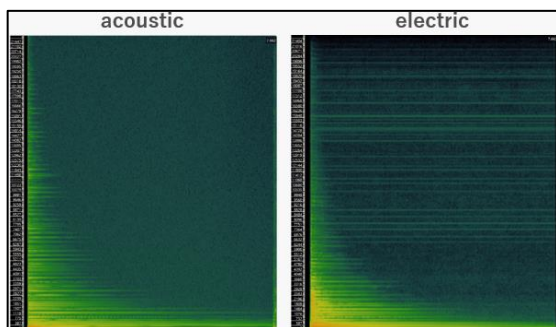


図 2 MIDI 音源のスペクトログラム
(左:アコースティックギター, 右:エレキギター)

実際の楽器同士でスペクトログラムを比較したところ、図の縦軸の約中央に位置する 10kHz 付近では、エレキギターよりもアコースティックギターの方が音の持続時間が約 2 倍以上長く、また低周波数帯の信号では、アコースティ

ックギターが約 700Hz 以下の信号が強いものに対して、エレキギターは約 820Hz 以下の信号が強く、特にエレキギターの立ち上がりの部分においてその特徴がアコースティックギターよりも強く現れていることがわかった。しかし実際の楽器の録音では、奏者によって弾く時の力が異なり、弦が振動する時間はその影響を受けるためあくまで今回の奏者ではこのような特徴であったと言える。

一方、MIDI 音源同士のアコースティックギターとエレキギターを比較したところ、エレキギターのサステインが不自然に長く、音源の終わりまで一定の強さで続いていることがわかった。このような特徴は、実際に録音されたエレキギターの特徴には見られないものである。また、同じ MIDI 音源である acoustic のスペクトログラムにはこのような特徴は存在しないため、MIDI プラグインの変更や音の消え入り方(ベロシティ)を設定しなおすことで改善する必要があると言える。

また、エレキギターの低周波数帯の時刻 0 付近を見ると、アコースティックギターに比べて広範囲に強い信号が存在していることがわかる。この特徴は実際の楽器音のスペクトログラム(図 1)でも見られるため、分離モデルが 2 クラスにおいて異なるパラメータを学習できる可能性を表していると言える。

5. 総括

本研究では、アコースティックギターとエレキギターの音源分離を目的として、実際の楽器音をサンプリングして作成された MIDI 音源の生成および音源分離ライブラリの Spleeter を用いたモデル学習を行う。そのために、Spleeter の学習の際に用いられる音源波形としての MIDI 音源が、実際の楽器の波形の特徴とどのような共通点や違いがあるかを調査した。

その結果、MIDI 音源のエレキギターには、実際の楽器音では見られない、音源の終わりまで一定の強度で出力される特徴が確認されたが、各スペクトログラムの低周波数帯においてエレキギターの方が広く強い信号が特に時刻 0 付近に存在することが確認された。

これらのことから今後は、エレキギターの MIDI プラグインの適切な設定や選択、そして効果的な学習アルゴリズムの検討を行い、学習モデルの作成とそのモデルの分離性能の評価を行う必要がある。

参考文献

- [1] Stöter, F.R., Liutkus, A., Ito, N. "The 2018 signal separation evaluation campaign", (2018)
- [2] <https://www.lalal.ai/>
- [3] Hennequin, R., Khlif, A., Voituret, F., and Moussallam, M. "Spleeter: a fast and efficient music source separation tool with pretrained models", (2020)
- [4] S. Uhlich, M. Porcu, F. Giron, M. Enekl, T. Kemp, N. Takahashi and Y. Mitsufuji "Improving music source separation based on deep neural networks through data augmentation and network blending", Proc. ICASSP, (2017)
- [5] E. Manilow, G. Wichem, P. Seetharaman, and J. Le Roux, "Cutting Music Source Separation Some Slack: A Dataset to Study the Impact of Training Data Quality and Quantity", (2019).
- [6] <https://www.image-line.com/fl-studio/>
- [7] <https://www.amplesound.net/en/pro-pd.asp?id=7>
- [8] <https://impactsoundworks.com/product/shreddage-3-stratus-free-kp/>
- [9] <https://sonicvisualiser.org/>
- [10] 田中研究室, 秋田大学:"Guitar chord sound dataset", <http://www.ee.akita-u.ac.jp/~tm-lab/guitar/dataset/index.html>,(2023)