1P-03

音響信号からのベース奏者の認識に関する研究

西村　奈那子†
津田塾大学大学院理学研究科†

1 はじめに

1.1 ジャズについて

ジャズバンドは、ピアノ(ギター)、ベース、ドラムの 3、4 種類の楽器からなる「リズムセクション」と、サックスやトランペットといった管楽器を演奏する「フロント」というパートで編成されることが多い。バンドや曲によっては、リズムセクションの3名のみで演奏する場合や、ギターがフロントとして参加する場合などがある。また、バイオリンやビブラフォンなどの楽器が編成に加わることもある。このように、編成に多少の違いがあるが、ジャズは合計 3〜6 名ほどで演奏することが多い。

ジャズの楽曲は、バンドメンバーが1人ずつ順に行うアドリブ演奏が曲の大部分を占め、決まったメロディを演奏する部分は少ない。上記のような編成で演奏を行う際、例えば、バンドがサックス、ピアノ、ベース、ドラムからなる場合、アドリブ演奏はフロントパートであるサックスから始まり、ピアノ、ベース、ドラムと続くことが多い。ただし、フロントパートの楽器がどのような曲でもほぼ必ずアドリブ演奏を行うのに対し、リズムセクションの楽器、特にベースは、どの曲でも必ずアドリブ演奏を行うわけではない。そのため、CD や動画などで公開されているジャズの音源の中でも、アドリブ演奏音源の数は楽器ごとに異なる。

楽器の演奏技術を向上させるための訓練として、音程やリズムに関する基礎的な練習や理論の学習に加え、有名な演奏者が過去に行った演奏の模倣練習が行われる。音楽には様々なジャンルがあり、どのジャンルにおいても模倣練習が行われるが、特に頻繁にアドリブ演奏を行うジャズにおいては、演奏技術習得のために模倣練習を多く行う必要がある。

Study on acoustic signal recognition of the bass players
†Nanako Nishimura, Tsuda University

1.2 ジャズベースについて

ある曲の模倣練習を行う際、アドリブ演奏音源が多ければ、模倣する演奏者の選択肢が広がる。すると、自身の演奏レベルや好みにあった演奏を選択することができ、練習の幅が広がる。しかし、前述の通りアドリブ演奏の音源の数には楽器によって差があり、特にベースのアドリブ演奏音源は少ない。そのため、演奏レベルや好みに合わせた選択を行えない場合がある。よって、ベースのアドリブ練習用の音源を生成することがベース演奏者の練習支援につながると考えられる。

ベーシストにはそれぞれの演奏特徴があり、その特徴を模倣することで、より高度な練習を行うことができるようになる。したがって、本研究では、アドリブ音源の少ないベースに注目し、模倣練習に使用できるような音源を作成するために、あるベース演奏者のもつ音の選び方、音の強弱、リズムといった演奏特徴を模したアドリブ演奏音源を自動生成することを最終目標とする。

2 検討内容

2.1 データ形式

音声データを扱うにあたり、データを音響信号としてそのまま扱うか、MIDI のようなシンボリック表現に変換して扱うかを決定する必要がある。データを処理しやすいのはシンボリック表現であるが、繊細な音の強弱やなめらかさなどの演奏特徴が失われる可能性がある。本研究の目的は、演奏特徴を再現することであるため、詳細な情報まで保持するために、主に音響特徴を扱うこととする。ただし、条件付けなどの目的で音源をシンボリック表現に変換して使用することも検討する。

2.2 生成手法

次に、自動生成に用いる手法について検討する。音声の自動生成に用いられるモデルには、RNN(LSTM) 、 CNN 、 Generative Adversarial Network[4](GAN)などがある。RNN は時系列データを

Copyright ©2021 Information Processing Society of Japan.
All Rights Reserved.

扱うことに長けており、CNN は時系列データの扱いには不向きとされている。しかし、WaveNet[1]が PixelCNN をベースとして時間的依存関係を学習し、自然な音声の生成を可能としたことから、音声領域への CNN の利用も有効であると考えられる。GAN は主に画像生成の領域で用いられ、生成を行う生成器と、生成器によって生成されたデータが本物のデータか偽物のデータかを識別する識別器を持つ。GAN や、GAN を発展させた DCGAN[2]を用いることで違和感のない画像を生成することができる。

GAN の音楽領域への応用例として、MuseGAN[3]が挙げられる。MuseGAN は、シンボリックなマルチトラック(ベース、ドラム、ギター、ピアノ、ストリングス)の音楽を生成することができ、生成器、識別器の両方が CNN で構成されている。MuseGAN の持つ 3 つの生成モデルのうち 1 つは各トラックを個別に生成するというモデルであるため、今回の研究への応用が可能と考えられる。よって、今回は CNN をベースとした GAN による演奏特徴の学習と、新たな音源の生成に注目する。

3 実験準備

3.1 音源生成にあたって

本研究では、CNN を基とした GAN を用いて、ベース演奏者の特徴を模したアドリブ演奏音源を自動生成する。それにあたり、まずは CNN やその応用技術を用いた既存の機械学習モデルを用いた場合、どの程度の精度で演奏者を識別できるかを調査する。

調査を行うための前準備として、ジャズベース演奏者の演奏音源を用いてデータセットを作成した。

3.2 データセット

今回、識別を行うジャズのベース演奏者として、模倣練習を行うにあたっての需要が高い Jaco Pastorious、Stanley Clarke、Anthony Jackson の 3 名を選択した。各アーティストがベース演奏者として参加している楽曲を 1 時間から 2 時間分ずつ用意し、以下の方法でデータセットを作成した。

まず、Spleeter[5]という音源分離ライブラリを使用し、すべての曲についてベースの演奏音源のみを抽出した。Spleeter には音源をボーカル/伴奏に分離する 2 分離、ボーカル/ベース/ドラム/その他に分離する 4 分離、ボーカル/ベース/ドラム/ピアノ/その他に分離する 5 分離という 3 つのモードがあるが、今回はベースのみ分離できれば良いため、4 分離モードを使用した。次に、抽出した音源をそれぞれ 5 秒ずつに区切った後にメル周波数スペクトルを求め、RGB 画像を作成した。

スペクトル画像作成後、無音区間や Spleeter での音声抽出がうまくいかなかった区間については音声ファイル、スペクトル画像ともに削除した。その結果、各アーティストのデータ数はそれぞれ 627 個、527 個、430 個となった。

5 まとめ

今回は、ジャズベース演奏者の演奏特徴を模したアドリブ演奏音源の自動生成を行うための生成手法を検討した。また、生成手法として選択した GAN での生成と識別に CNN を用いることを検討した。最後に、ベースのみを抽出した音源を CNN の学習に用いた場合、どの程度の精度で演奏者を特定することができるのかを調査するために、前準備として 3 名のベーシストによる演奏音源を用いてデータセットのサンプルを作成した。

今後は、データの増量を行うとともに、実際に CNN を用いて学習を行い、他手法による学習との精度比較を行う。

参考文献

[1] Aaron van den Oord, Sander Dieleman, Heiga Zen, Karen Simonyan, Oriol Vinyals, Alex Graves, Nal Kalchbrenner, Andrew Senior, Koray Kavukcuoglu “WaveNet: A Generative Model for Raw Audio”, arXiv:1609.03499, 2016.
[2] Alec Radford, Luke Metz, Soumith Chintala , “Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks”, arXiv:1511.06434, 2015.
[3] Hao-Wen Dong, Wen-Yi Hsiao, Li-Chia Yang, Yi-Hsuan Yang “MuseGAN: Multi-track Sequential Generative Adversarial Networks for Symbolic Music Generation and Accompaniment”, In AAAI, 2018.
[4] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio “Generative adversarial nets”, NIPS, 2014.
[5] Romain Hennequin, Anis Khlif, Félix Voituret, Manuel Moussallam “Spleeter: A Fast And State-of-the Art Music Source Separation Tool With Pre-trained Models”, In ISMIR, 2019.

Copyright ©2021 Information Processing Society of Japan.
All Rights Reserved.