

## ニューラルネットワークを用いた音楽ゲーム譜面自動生成の検討

柴崎 大地<sup>†</sup>酒井 充<sup>†</sup>  
富山大学<sup>†</sup>丸山 博<sup>†</sup>

## 1. はじめに

音楽ゲームとは、リズムや音楽に合わせて画面に表示される指示に従って、プレイヤーがボタンを押したり、画面をタッチすることで、得点を獲得するゲームである。音楽ゲームにはプレイヤーに指示するアクション等の情報(ノート)を記述したゲーム譜面が使用される。ゲーム譜面を自動生成するプログラムはすでに開発されているが、生成された譜面に対するプレイヤーの評価は良いとは言えず、現状ではほとんど人間が作った譜面がプレイに使用されている。

文献 [1]では Dance Dance Revolution という音楽ゲームの譜面を、音楽データから CNN や LSTM を用いて生成する方法が提案されている。本稿では対象となる音楽ゲームを限定せず、文献 [1]の CNN を用いた手法でゲーム譜面の生成を行い、検討を行う。

## 2. 提案手法

一般にプレイヤーへの操作のタイミングとその時のアクション内容という二つの情報の組をノート、一楽曲分のノート列を譜面と呼ぶ。本研究では自動譜面生成の第一段階としてノートのタイミングのみの生成を行う。なお、BPM(Beat Per Minute)情報や曲のオフセットは既知とする。

## 2.1 音特徴

入力する音響信号はサンプリング周波数 44.1[KHz]でサンプリングしたものである。音響信号が 2 チャンネル以上の場合には平均を取って 1 チャンネルにする。本研究では BPM の値とオフセットを使用しノート入力可能位置を設定する。ノート入力可能位置はオフセットから 32 分音符間隔に設定する。図 1 は入力音響信号に対してノート入力可能位置(赤の縦線)を示した例である。

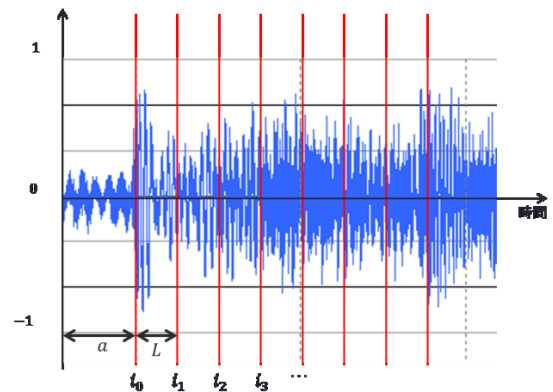


図 1 入力音響信号とノート入力可能位置

次に各ノート入力可能位置周辺の周波数スペクトルを短時間フーリエ変換(STFT)で求める。STFT の窓長は 23ms, 46ms, 93ms を用いる。求めた周波数スペクトルはメルフィルタバンクを用いて次元圧縮を行い 80 次元にする。前後 7 ステップの周波数情報を用い、最終的に各ノート入力可能位置から得られる音特徴は  $15 \times 80 \times 3$  の情報になる。

## 2.2 タイミング推定

タイミング推定には CNN を用いる。使用するモデルは、2 層の畳み込み層と 2 層の全結合層で構成される。一つ目の畳み込み層では 10 個の  $7 \times 3$  のフィルタで畳み込みを行い、二つ目では 20 個の  $3 \times 3$  のフィルタで畳み込みを行う。各畳み込み層の後に周波数軸方向に 1D max-pooling を行う。また、どちらの畳み込み層も活性化関数は rectified linear units(ReLU)を使用する。全結合層はそれぞれ 256, 128 ノードで構成する。最後にシグモイド関数でそのノート入力可能位置での推定値を出力する。図 2 は今回使用する CNN モデルの全体像である。出力する推定値が閾値より高ければそのタイミングにノートありと判定する。

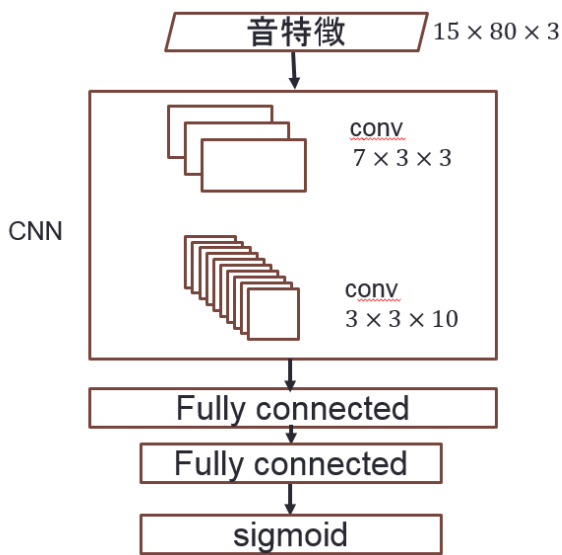


図2 CNNを用いたタイミング推定

### 3. 実験

実験に使用する各楽曲は BPM が途中変化しない 4 分の 4 拍子のものであり、それぞれの BPM や曲の長さはばらばらである。一般に音楽ゲームでは 1 楽曲に複数の難易度のゲーム譜面が用意されているが、今回はその中で難易度の高いもののみを使用した。140 曲の譜面を学習用に、25 曲の譜面をテスト用に使用した。

使用する言語は Python3.5.2 であり、ニューラルネットワークのライブラリには Keras を使用した。

### 4. 結果

表 1 に実験結果を示す。F-Score<sup>c</sup> は楽曲ごとに閾値を変動させ、最も F-Score が高かった時の平均をとったものである。F-Score<sup>m</sup> は全楽曲での micro F-score である。

表 1 実験結果

F-Score <sup>c</sup>	F-Score <sup>m</sup>
0.551	0.568

文献[1]では CNN での F-Score<sup>c</sup> が 0.678, F-Score<sup>m</sup> が 0.750 であったが、本実験ではそれよりも低い結果になった。学習データやノート入力可能位置の設定など、実験方法に差があるため、詳細については今後検討していきたい。

一方、学習データを再代入して結果を求めたところ、F-Score<sup>c</sup> が 0.826, F-Score<sup>m</sup> が 0.868 であった。このことから、学習データでは認識に必要な特徴抽出ができていていると考えられる。テストデータの前半を学習用データとして使用し、後半をテストデータとする半自動生成を行った。その結果を転移学習なしとして表 2 に示す。さらに、特徴抽出部と見なせる畳み込み層の重みを 140 曲の学習データで得た重みを使用することで、特徴抽出の精度が向上すると考え、追加実験を行った。重みの初期値を 140 曲の学習データで得た重みにした場合を転移学習 1、畳み込み層を固定した場合を転移学習 2 として表 2 に示す。

表 2 半自動生成の結果

	F-Score <sup>c</sup>	F-Score <sup>m</sup>
転移学習なし	0.666	0.833
転移学習 1	0.704	0.836
転移学習 2	0.693	0.832

140 曲の学習データで得た重みを使用する転移学習により、F-Score<sup>c</sup> が改善できた。

### 5. おわりに

文献[1]とほぼ同じ手法で音楽ゲームの譜面生成を試み、CNN を用いたタイミング推定に効果があることがわかったが、対象とする音楽ゲームを限定しなかったことにより、文献[1]より推定精度が低くなったと考えられる。また、半自動生成の場合では、他の楽曲で学習した重みを使用する転移学習により結果の向上が見られた。

#### 参考文献

- [1] Chris Donahue, Zachary C. Lipton, Julian McAuley. "Dance Dance Convolution", <https://arxiv.org/pdf/1703.06891.pdf>, 2017/1/21.
- [2] 柴崎他, "NTF を用いた音楽ゲーム譜面自動生成の検討", 電気関係学会北陸支部連合大会, 2017