Vol.2020-ICS-197 No.8
2020/2/17

CycleGANを用いたゲーム音楽のシーン別変換

星 雄輝[†1,a)]　清 雄一[†1,b)]　田原 康之[†1,c)]　大須賀 昭彦[†1,d)]

概要：昨今，PCやスマートフォン，インターネットの普及に伴い個人でもゲームを制作することが身近な存在となっている．しかし，作曲・演奏経験の乏しいゲームアプリ開発者にとって，オリジナルのBGMを用意するのはコストが必要となる場合がほとんどである．また，ゲーム音楽といっても，ゲーム音楽ならではの音楽的特徴は幅広く，ゲーム音楽そのものの理論は確立されていないのが現状である．そこで，ゲームのシーンに着目し，音楽をゲームのシーンに合った曲調に変換してくれるシステムによって，音楽理論を理解していなくてもゲームのBGMを用意できるのではないかと考えた．本研究では，CycleGANを用いたドメイン変換に加え，テンポと音色を考慮することによって，クラシック音楽からシーン別のゲーム音楽を生成することを提案する．評価としては，分類器を用いた変換精度の客観評価とアンケート形式による生成楽曲に対するイメージ及び変換精度を答えてもらう主観評価の2つの評価方法を用いて評価を行った．客観評価と主観評価の総合的な評価の結果，「フィールド」と「戦闘」のシーンに変換した楽曲において高い評価が得られた．

キーワード：ドメイン変換，GAN，CycleGAN，ゲーム音楽

1. はじめに

1.1 背景

PCやスマートフォン，インターネットの普及に伴いサービスを利用するのはもちろん，サービスを提供することも容易になってきている．ゲームアプリもそのうちの1つであり，仕事や趣味でゲームを制作することが情報通信機器の普及前と比べ，身近なものになっている．以前と比べ身近な存在となったゲームの制作だが，個人レベルで制作する際には，BGMを必要とすることがある．ゲーム制作と同様に作曲環境も進歩しているとはいえ，作曲する場合には音楽理論の知識等が必要となり，そういった知識の無い初学者にとって，ゲーム音楽の作曲は敷居が高くなっている．

ゲーム音楽に関しては，ゲーム音楽の黎明期から発展途上にかけては，ゲーム機の制約上，限られた音でしか構成されないという特徴はあったが，技術の進歩に伴い，近年では映画の音楽に近いクオリティのゲーム音楽も存在する．ゲーム音楽に関する研究も焦点が当てられてきているが，Karen Collinsによると，ゲーム音楽に関する研究は散在しており，ゲーム音楽理論は未だ確立されていない状況である [1].

1.2 目的

ゲーム音楽という広い枠組みではなく，ゲームのシーンに着目し，更に音楽理論の知識を有せずとも作曲できるよう，ニューラルネットワークを用いたドメイン変換によってゲーム音楽を生成することを考えた．ドメイン変換とは特定のセマンティックな情報を維持したまま，画像であれば別のスタイルの画像に，音声であれば別のスタイルの音声に変換することである．本研究では，CycleGANを用いたテンポと音色を加味したドメイン変換によってゲーム音楽を生成し，その精度の評価を行うことを目的とする，また，ドメイン変換前のデータには，著作権の保護期間が切れている曲が少なくなく（ただし著作隣接権には配慮が必要），実際にクラシック音楽をモチーフとしたり類似しているゲーム音楽が存在するということを考慮して，クラシック音楽を用いた．

1.3 論文の構成

本論文の構成は以下のとおりである．以降，本論文は2章では，関連研究について，3章では提案手法の説明，4章では提案手法を用いた実験について，5章では実験結果の評価について，6章では考察，最後に7章で本論文の結論

[†1] 現在，電気通信大学 〒182-8585 東京都調布市調布ケ丘 1-5-1 , Chofu, Tokyo, 182-8585, Japan
[a)] hoshi.yuki@ohsuga.lab.uec.ac.jp
[b)] seiuny@uec.ac.jp
[c)] tahara@uec.ac.jp
[d)] ohsuga@uec.ac.jp

をまとめ，今後の展望を記す．

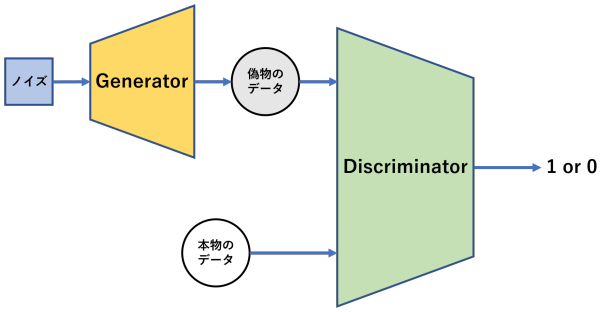

図 1 GAN のモデル

2. 関連研究

2.1 GAN

Goodfellow ら [2] が提案した GAN（Generative Adversarial Network）は，2 つのニューラルネットワーク（Generator と Discriminator）で構成される生成モデルの一種であり，敵対的生成ネットワークとも呼ばれる．Generator は，ランダムノイズを入力として受け取りデータの生成を担い，もう 1 つのネットワークである Discriminator は，Generator が生成した偽物のデータと訓練データから取り出された本物のデータを入力として受け取り，データの正否を判別する．Generator は生成データを Disciminator に本物として判別されるように精度を向上させ，Discriminator は Generator が生成したデータを偽物と判別できるるよう精度を向上させて，互いに競合させながら学習することによって，より精度の高いデータを生成できる仕組みになっている．GAN のモデルを図 1 に示す．また，GAN の損失関数はミニマックス最適化として以下のように定式化できる．

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{data}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1 - D(G(z)))] \quad (1)$$

近年では，この GAN のネットワークに CNN を用い，更にネットワークの層を深くした DCGAN（Deep Convolutional GAN）が用いられることが多くなっている．

2.2 CycleGAN

Zhu ら [3] が提案した CycleGAN は GAN のネットワークを 2 つ組み合わせたようなネットワークの構造を持ち，学習の際には 2 つのデータセットを対にして，教師なしで学習が行われる．2 つのドメインをそれぞれドメイン X とドメイン Y，Generator によるドメイン X からドメイン Y への変換を $G_{X\to Y}$ とする．$x \in X$ を $G_{X\to Y}$ に入力することで $y' = G_{X\to Y}(x)$ が生成され，y' を $G_{Y\to X}$ に入力することで $x'' = G_{Y\to X}(y')$ が生成される．このときの x''

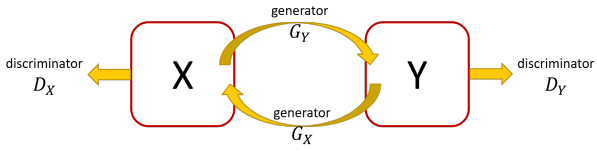

図 2 CycleGAN のモデル

が x と一致しているかを確認しながら学習が行われ，y' はドメイン Y の Discriminator，D_Y によって判別が行われる．この一連の変換が逆方向に対しても行われることにより，循環による一貫性が確認され，双方向のドメイン変換が可能となっている．CycleGAN のモデルを図 2 に示す．

また，CycleGAN の損失関数は，Adversarial Loss と Cycle-Consistency Loss によって以下のように定式化できる．

$$\mathcal{L}_{full} = \mathcal{L}_{adv}(G_{X\to Y}, D_Y) + \mathcal{L}_{adv}(G_{Y\to X}, D_X) + \lambda_{cyc}\mathcal{L}_{cyc}(G_{X\to Y}, G_{Y\to X}) \quad (2)$$

Adversarial Loss は以下のように表される．

$$\mathcal{L}_{adv}(G_{X\to Y}, D_Y) = \mathbb{E}_{y\sim p_{data}(y)}[\log D_Y(y)] + \mathbb{E}_{x\sim p_{data}(x)}[\log(1 - D_Y(G_{X\to Y}(x)))] \quad (3)$$

$$\mathcal{L}_{adv}(G_{Y\to X}, D_X) = \mathbb{E}_{x\sim p_{data}(x)}[\log D_X(x)] + \mathbb{E}_{y\sim p_{data}(y)}[\log(1 - D_X(G_{Y\to X}(y)))] \quad (4)$$

Cycle-Consistency Loss は以下のように表される．

$$\mathcal{L}_{cyc} = \mathbb{E}_{x\sim p_{data}}[\|G_{Y\to X}(G_{X\to Y}(x)) - x\|_1] + \mathbb{E}_{y\sim p_{data}(y)}[\|G_{X\to Y}(G_{Y\to X}(y)) - y\|_1] \quad (5)$$

2.3 音楽を対象としたドメイン変換について

ニューラルネットワークによるドメイン変換といえば，ほとんどが画像に焦点を当てて行われていたが，昨今では，自然言語や音声といった，画像以外の非構造化データを対象としたドメイン変換も盛んに行われている．音楽に対するドメイン変換の手法は，主に Recurrent Neural Networks（RNN）([4, 5]) や Long-short term model（LSTM）[6]([7–11]) によって行われていたが，近年では畳み込みニューラルネットワーク（CNN）を用いた手法も注目されている．

CNN を用いた音楽のドメイン変換の手法の 1 つに，Brunner ら [12] が提案した MIDI を対象として CycleGAN

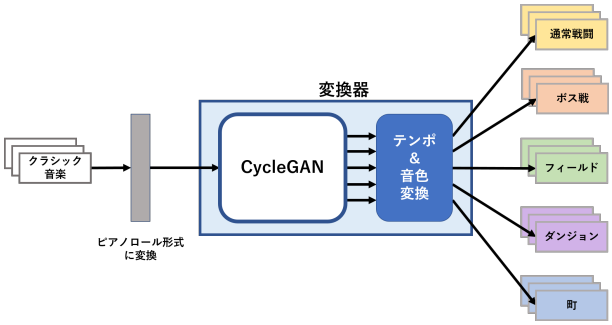

図 3 提案システムの全体像

を用いたものがある．この CycleGAN のモデルでは，Discriminator を追加し，偽物のデータとターゲットのドメインの比較だけでなく，複数のドメインとも比較することによって，音楽としての構造を保持するよう制約がなされている．更に GAN の学習をより安定させるために，両方の Discriminator にガウシアンノイズ $\mathcal{N}(0, \sigma_D^2)$ を入力するモデルとなっている．本研究で導入した CycleGAN では同様のモデルを使用した．

3. 提案手法

3.1 提案手法の全体像

本研究では，多くの RPG ゲームで共通して存在する「通常の戦闘」，「ボス戦」，「フィールド」，「ダンジョン」，「町」の 5 つのシーン別にゲーム音楽の生成を行った．本システムでは，まず，入力データとして MIDI 形式のクラシック音楽を与え，ピアノロール形式に変換する．そのデータをゲームのシーン別に学習させた CycleGAN に渡し，元のメロディをセマンティックな情報として維持されるよう変換が行われる．更に楽曲のテンポと音色を変換することによって，ゲーム音楽をゲームのシーン別に生成するシステムとなっている．以上のゲーム音楽のシーン別変換システムの全体像を図 3 に示す．

3.2 テンポと音色の変換

テンポおよび音色の変換では，CycleGAN で出力されたデータを入力として受け取り，目的の楽曲を出力する．テンポと音色に関しては，Livingstone ら [13] らが提案した，音楽における感情の想起を決定づける 8 つの要素を参考にそのうち 2 つを選定した．テンポの変換では，各ゲームシーンの学習データからテンポの平均値を求め，出力時にそのテンポに変換するよう設定した．また，音色の変換では，ゲーム音楽黎明期にメロディ部分で主に使用されていた矩形波に音色を変換させるよう設定した．このテンポと音色の変換器は Python の pretty_midi パッケージを用いて作成した．

4. 実験

4.1 ネットワークアーキテクチャとハイパーパラメータ

システムに使用した CycleGAN のアルゴリズムは先行研究 [12] を参考にしたものである．使用した CycleGAN の Discriminator のアーキテクチャを表 1 に，Generator のアーキテクチャを表 2 に示す．また，Discriminator に入力するガウシアンノイズ $\mathcal{N}(0, \sigma_D^2)$ のハイパーパラメータ σ_D については，各シーンの学習時に σ_D を 0.01，0.1，1.0 の 3 パターンで入力し，$\mathcal{L}_{cyc}$ の値が最も低かったものを選定した．ゲームのシーンごとの σ_D は表 3 の通りである．

表 1 Discriminator のアーキテクチャ

Input: ($batchsize$ × 64 × 84 × 1)					
layer	filter	stride	channel	instance norm	activation
conv	4 × 4	2 × 2	64	False	LReLu
conv	4 × 4	2 × 2	256	True	LReLu
conv	1 × 1	1 × 1	1	False	None
Output: ($batchsize$ × 16 × 21 × 1)					

表 2 Generator のアーキテクチャ

Input: ($batchsize$ × 64 × 84 × 1)					
layer	filter	stride	channel	instance norm	activation
conv	7 × 7	1 × 1	64	True	ReLu
conv	3 × 3	2 × 2	128	True	ReLu
conv	3 × 3	2 × 2	256	True	ReLu
conv	3 × 3	1 × 1	256	True	ReLu
conv	3 × 3	1 × 1	256	True	ReLu
conv	3 × 3	2 × 2	128	True	ReLu
conv	3 × 3	2 × 2	64	True	ReLu
conv	7 × 7	1 × 1	1	False	Sigmoid
Output: ($batchsize$ × 64 × 84 × 1)					

表 3 各シーンの学習で設定した σ_D について

	通常戦闘	ボス戦	フィールド	町	ダンジョン
σ_D	0.1	0.01	0.1	1.0	1.0

4.2 データセット

実験では MIDI 形式のデータを扱うものとし，VGMusic [14] から世界的に有名な RPG ゲームである FINAL FANTASY の曲をゲーム音楽のデータとして収集した．クラシックの MIDI データに関しては，先行研究で公開されているものを使用した [15]．収集した MIDI を扱いやすいデータにするために，以下のような前処理を行った．

- ゲーム音楽の MIDI データの水増し（Data augmentation）
- 4/4 拍子かつ拍子が変化しないものに絞る
- ドラムのトラックを除去

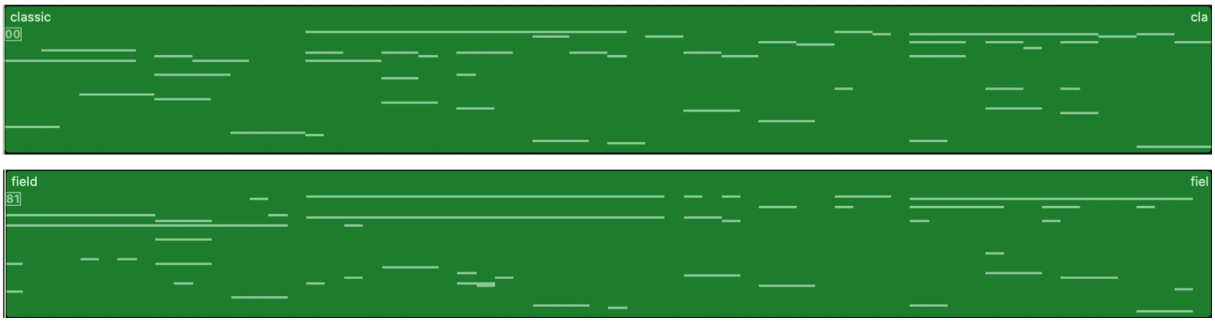

図 4 変換の例（上段がクラシック音楽，下段がフィールドのシーンに変換した楽曲）

- C0〜C8 以外の音域を切り捨てる
- 全ての音を同じ大きさに揃える
- 複数のトラックを 1 つのトラックに統合する
- 1 つの曲から 4 小節ごとのフレーズに分割
- ピアノロール形式に変換

ゲーム音楽の Data augmentation については，ゲーム音楽の MIDI データをゲームのシーン別に分類した場合，明らかにデータ数が足りなくなってしまうため，適切な音域を超えない範囲で移調させたデータを加えた．

前処理後のデータのサイズは，16 歩音符を最小単位とした 4 小節分を 1 フレーズとし，C0〜C8 内の 84 音域から成り立っていることから，[タイムステップ，音域，出力チャネル] = [64，84，1] となる．

前処理後のデータセットについて，表 4 に示す．

表 4 前処理後のデータ数（フレーズ数）

	通常戦闘	ボス戦	フィールド	町	ダンジョン
学習	1240	6650	3020	1366	3628
テスト	132	640	360	159	444

4.3 生成実験

前処理を行ったデータを使用して，クラシック音楽とシーン別のゲーム音楽を対にして学習を行った．学習の際には，学習データの少ない方と同じ数になるように，一方のデータに対しダウンサンプリングを行った．また，学習回数は 30epoch とした．提案システムによって生成したフィールドのシーンの楽曲を例として図 4 に示す．

5. 評価実験

5.1 客観評価

5.1.1 実験目的

分類器を用いることによって，変換前，変換後，変換後から変換前のドメインに変換し直した楽曲を分類し，変換の精度の評価を行う．

5.1.2 実験内容

客観評価では，クラシック音楽とシーン別のゲーム音楽のデータをペアとして分類器に学習（合計で 5 つのペアで学習）させ，変換の精度を求めた．分類器は先行研究 [12] で使用されていたもを参考にして使用し，同様の方法で精度を算出した．変換前のドメイン（クラシック音楽）を A，変換後のドメイン（シーン別のゲーム音楽）を B とする．このとき，楽曲 x が A である確率を $P_A(x)$ とすると，ドメイン A の変換前のデータ x_A とドメイン B に変換後のデータ $\hat{x}_B$ について，$P_A(x_A) > 0.5$ かつ $P_A(\hat{x}_B) < 0.5$ のときに変換が成功したと見なせる．更に，$\hat{x}_B$ をドメイン A に変換し直したデータを $\tilde{x}_A$ とすると，CycleGAN を用いて A から B の変換を行ったときのドメイン変換の強度を以下のように定義できる．

$$S^D_{A\to B} = \frac{P_A(x_A) - P_A(\hat{x}_B) + P_A(\tilde{x}_A) - P_A(\hat{x}_B)}{2} \quad (6)$$

$P_A(x_A) = 1$，$P_A(\hat{x}_B) = 0$，$P_A(\hat{x}_B) = 1$ のときに，$S^D_{A\to B}$ は最も高い強度となるが，セマンティックな情報をどれだけ維持できているかは判断できない．使用した分類器のアーキテクチャを表 5 に示す．

表 5 分類器のアーキテクチャ

Input: ($batchsize \times 64 \times 84 \times 1$)					
layer	filter	stride	channel	instance norm	activation
conv	1×12	1×12	64	False	LReLu
conv	4×1	4×1	128	True	LReLu
conv	2×1	2×1	256	True	LReLu
conv	8×1	8×1	512	True	LReLu
conv	1×7	1×7	2	False	Softmax
Output: ($batchsize \times 16 \times 21 \times 1$)					

5.1.3 実験結果

実験結果を表 6 に示す．いずれのシーンにおいて，$P_A(x_A) > 0.5$ かつ $P_A(\hat{x}_B) < 0.5$ を満たす結果となった．$S^D_{A\to B}$ については，「フィールド」，「町」，「ダンジョン」のシーンにおいて，50%以上となっており，特に「フィールド」と「ダンジョン」のシーンにおいて，60%を越える結果となった．

表 6　各シーンの変換の精度

	通常戦闘	ボス戦	フィールド	町	ダンジョン
$P_A(x_A)$	88.13%	90.04%	91.64%	93.27%	93.17%
$P_A(\hat{x}_B)$	42.45%	49.92%	4.28%	36.10%	32.53%
$P_A(\tilde{x}_A)$	80.54%	48.12%	54.55%	82.31%	92.84%
$S^D_{A\to B}$	41.88%	21.16%	68.81%	51.69%	60.47%

5.2　主観評価

5.2.1　実験目的

アンケート形式によって，生成した楽曲を聴いてもらい，生成楽曲の精度と変換精度の評価を行う．

5.2.2　実験内容

20〜30 代の男女 26 人の被験者に生成した楽曲を聞いてもらい，評価を行った．評価に用いた楽曲は，変換前のクラシック音楽と 5 つのゲームのシーンに変換したゲーム音楽であるが，被験者にはどのシーンに変換したものかは伝えずに聞いてもらった．被験者には 5 つの変換後の楽曲を 3 セット，計 15 曲のゲーム音楽の評価を行ってもらった．また，被験者には「演奏経験，作曲経験の程度」と「RPG ゲームのプレイの経験の程度」をアンケートで答えてもらい，生成した楽曲ごとに以下のような項目について 5 段階で評価（1 が最低，5 が最高）を行ってもらった．

- どのようなシーンの印象をどの程度感じたか
  - 戦闘のシーン
  - ボス戦のシーン
  - フィールドのシーン
  - 町のシーン
  - ダンジョンのシーン
- 生成楽曲の変換精度について
  - 音楽として聞き苦しくはなかったか
  - ゲーム音楽らしさを感じるか
  - 元の曲の名残はどの程度あったか

「どのようなシーンの印象をどの程度感じたか」という項目については，5 段階の評価に加え「そもそもシーンに対するイメージがわからない」という評価項目を加えた．

5.2.3　実験結果

3 回以上 RPG ゲームを最後までクリアしたことがあった人数は 16 人（61.5%），4 年以上演奏・作曲経験があった人数は 17 人（65.4%）であった．5 段階評価の数値（1〜5）の平均値を算出し，評価を行った．「そもそもシーンに対するイメージがわからない」という評価項目に関しては，評価値を 0 として計算した．

生成したシーン別のゲーム音楽に対するゲームのシーンのイメージについては，図 5 に示す．「通常戦闘」と「フィールド」のシーンに変換した曲については，その曲のシーンと同じシーンをイメージした平均値が 3.6 を上回り，「ボス戦」と「ダンジョン」に関しては，3 を下回る結果となった．また，シーンに関わらず，RPG のゲームのプレイ数，演奏・作曲経験年数による特徴的な差はなかった．

生成楽曲の変換精度については，RPG のゲームのプレイ数，演奏・作曲経験年数による特徴的な差がなかったため，被験者全員を対象としてまとめた結果を図 6 に示す．「音楽として聞き苦しくはなかったか」という項目については，町以外のシーンに変換した楽曲において，評価の平均値が 3 を上回る結果となり，特に「フィールド」と「ダンジョン」のシーンに変換した楽曲に関しては，4 以上の値となった．「ゲーム音楽らしさを感じるか」という項目については，いずれのシーンに変換した楽曲において，平均値が 3.5 を上回り，特に「通常戦闘」,「ボス戦」,「フィールド」のシーンに変換した楽曲に関しては，4 を上回る結果となった．「元の曲の名残はどの程度あったか」という項目については，「フィールド」と「ダンジョン」のシーンに変換した楽曲において，平均値が 3 を上回り，特に「ダンジョン」のシーンに変換した楽曲に関しては 4 を上回る結果となった．

6.　考察

主観評価では，「通常戦闘」と「フィールド」のシーンに変換した楽曲について，変換したシーンと同じシーンに対する評価値が高い結果となっており（図 5），他の 3 つのシーンと比べて高い精度で変換ができたのではないかと考えられる．特に「フィールド」のシーンにに関しては，客観評価において $S^D_{A\to B}$ の精度が 68%と，5 つのシーンの中でも最も高く，また，変換精度の主観評価（図 6）の全ての項目においても評価値が 3 を越えており，総合的に判断すると最も高い精度で変換できたのではないかと考えられる．

「ボス戦」と「ダンジョン」のシーンに変換した楽曲については，そのシーンのイメージの評価値が低い結果となった（図 5）．「ボス戦」のシーンに関しては，そもそもボス戦の BGM には，本研究で提案したシステム（主に midi を扱う Python のパッケージ）では扱いきれないような特殊な音が多いことや，16 分音符以上に細かい音符が使用されていたことが多かったこと，中ボスとラスボスの BGM に音楽的特徴の違いがあったこと，時代が進むにつれ演奏の表現の幅が広がったことにより作品ごとにモチーフを反映させるような BGM になっていたことなどが，変換の精度が良くなかった原因ではないかと考えられる．「ダンジョン」のシーンに関しては，主観評価において「元の曲の名残はどの程度あったか」という評価値が最も高い結果であったことから（図 6），変換前のクラシック音楽の音楽的特徴が強く残り，その結果，シーンのイメージがしにくかったことが原因ではないかと考えられる．

「ダンジョン」,「フィールド」,「町」のシーンに変換した楽曲のシーンのイメージの評価値については，「町」と

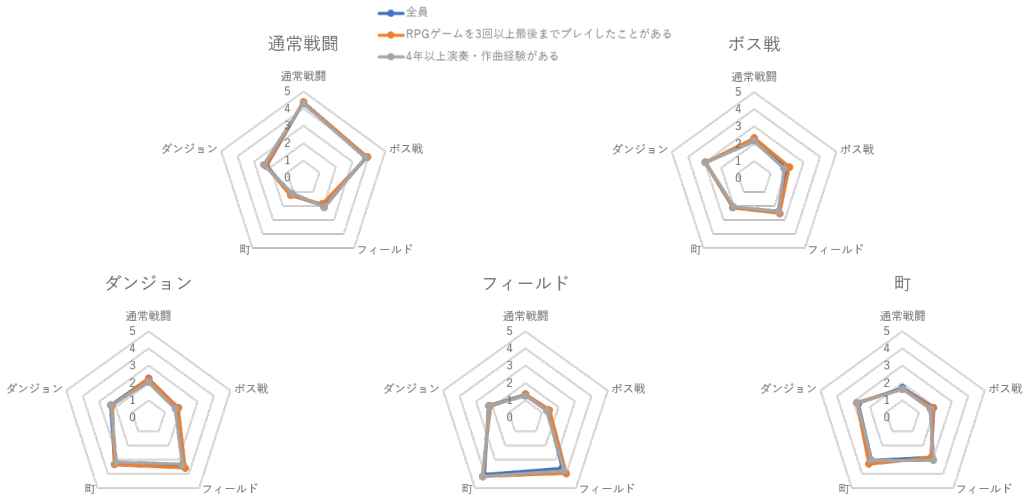

図 5　シーン別に生成した楽曲に対してイメージしたシーンの評価の平均値

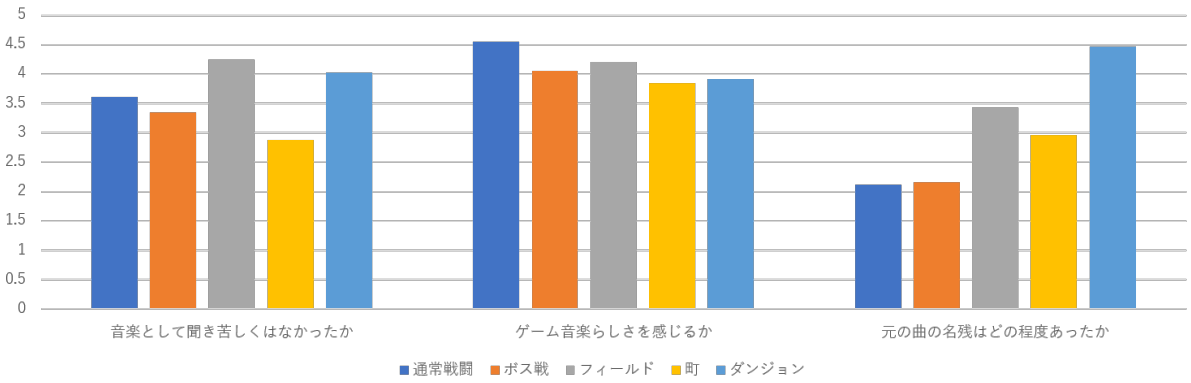

図 6　変換精度の評価の平均値

「フィールド」の評価値が同程度であった（図 5）．このことから，被験者にとって「町」と「フィールド」の曲に対するイメージの差がほとんどない，あるいは，その 2 つの違いを判断することが難しいものなのではないかと考えられる．

主観評価の「音楽として聞き苦しくはなかったか」という項目では，「フィールド」と「ダンジョン」のシーンに変換した楽曲の評価値が高い結果となったが，同様に「元の曲の名残はどの程度あったか」という項目でもその 2 つのシーンの評価値が最も高かった（図 6）．このことから，「フィールド」と「ダンジョン」のシーンについては，クラシック音楽の音楽的特徴がある程度維持されたまま変換されており，その結果音楽としてのクオリティも保たれたのではないかと考えられる．

7.　おわりに

7.1　まとめ

本論文では，ゲーム音楽の音楽としてのカテゴリの幅広さを考慮して，ゲーム音楽をシーン別に着目し，また，音楽理論の知識を有していないものでも，ゲーム音楽の作曲ができるよう，CycleGAN を用いたテンポと音色を加味したドメイン変換によってゲーム音楽を生成するシステムを提案した．分類器による客観評価とアンケートによる主観評価によって，生成した楽曲の評価を行った結果，「通常戦闘」と「フィールド」のシーンに変換した楽曲は，比較的高い精度で変換が行えたことが分かり，特に「フィールド」のシーンに変換した楽曲については，音楽としてのクオリティについても高い評価が得られた．

7.2　今後の展望

今後の課題としては，音色とテンポ以外の音楽的特徴を考慮する必要があると考えられる．また，本研究では，ドラムのトラックを削除し，複数のトラックを同一のトラックに併合してから変換を行うシステムとなっており，メロディとベースを考慮することによって，精度が向上するのではないかと考えられる．更に，ゲームのシーンによって変換の精度の差があることから，変換前と変換後の音楽的特徴の相性についても考慮し改善する必要があるのではないかと考えられる．

謝辞 本研究は JSPS 科研費 JP17H04705, JP18H03229, JP18H03340, JP18K19835, JP19H04113, JP19K12107 の助成を受けたものです.

参考文献

[1] Karen Collins: Game Sound: An Introduction to the History, Theory and Practice of Video Game Music and Sound Design, MIT Press(2008)
[2] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in Advances in neural information processing systems, 2014, pp. 2672– 2680.
[3] J. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in IEEE International Conference on Computer Vision, ICCV 2017, Venice, Italy, October 22-29, 2017, 2017, pp. 2242–2251. [Online]. Available: https://doi.org/10.1109/ICCV.2017.244
[4] P. M. Todd, "A connectionist approach to algorithmic composition," Computer Music Journal, vol. 13, no. 4, pp. 27–43, 1989.
[5] M. C. Mozer, "Neural network music composition by prediction: Exploring the benefits of psychoacoustic constraints and multi-scale processing," Connect. Sci., vol. 6, no. 2-3, pp. 247–280, 1994. [Online]. Available: https://doi.org/10.1080/09540099408915726
[6] S. Hochreiter and J. Schmidhuber, "Long short-term memory," Neural Computation, vol. 9, no. 8, pp. 1735–1780, 1997. [Online]. Available: https://doi.org/10.1162/neco.1997.9.8.1735
[7] D. Eck and J. Schmidhuber, "A first look at music composition using lstm recurrent neural networks," Istituto Dalle Molle Di Studi Sull Intelligenza Artificiale, vol. 103, 2002.
[8] N. Boulanger-Lewandowski, Y. Bengio, and P. Vincent, "Modeling temporal dependencies in high-dimensional sequences: Application to polyphonic music generation and transcription," in Proceedings of the 29th International Conference on Machine Learning, ICML 2012, Edinburgh, Scotland, UK, June 26 - July 1, 2012, 2012. [Online]. Available: http://icml.cc/2012/papers/590.pdf
[9] G. Brunner, Y. Wang, R. Wattenhofer, and J. Wiesendanger, "JamBot: Music theory aware chord based generation of polyphonic music with LSTMs," in 29th International Conference on Tools with Artificial Intelligence (ICTAI), 2017.
[10] G. Hadjeres, F. Pachet, and F. Nielsen, "DeepBach: a steerable model for bach chorales generation," in Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017, 2017, pp. 1362–1371. [Online]. Available: http://proceedings.mlr.press/v70/hadjeres17a.html
[11] H. Chu, R. Urtasun, and S. Fidler, "Song from PI: A musically plausible network for pop music generation," CoRR, vol. abs/1611.03477, 2016. [Online]. Available: http://arxiv.org/abs/1611.03477
[12] Brunner, G., Wang, Y., Wattenhofer, R., Zhao, S.: Symbolic music genre transfer with cyclegan. In: IEEE 30th International Conference on Tools with Artificial Intelligence, ICTAI 2018, 5-7 November 2018, Volos, Greece. pp. 786–793 (2018). https://doi.org/10.1109/ICTAI.2018.00123
[13] Steven R. Livingstone, Ralf Muhlberger, Andrew R. Brown, William F. Thompson: Changing Musical Emotion : A Computational Rule System for Modifying Score and Performance, Computer Music Journal, Spring 2010, 34:1, pp. 41-64
[14] https://www.vgmusic.com
[15] https://drive.google.com/file/d/1B1ocZ9WMOE2cN7CUVPaAFE9RNgTR0R1f/view