

Transformerモデルによる感情を基にした動画BGMの生成と評価

原田楓* 小高知宏** 黒岩丈介** 諏訪いずみ*** 白井治彦****

Generation and Evaluation of Emotion-based Video BGM Using Transformer Model

Kaede HARADA* Tomohiro ODAKA** Jousuke KUROIWA**
Izumi SUWA*** Haruhiko SHIRAI****

(Received September 30, 2022)

In this paper, we developed a system that automatically generates appropriate background music for a video using the emotions contained in the video as input. A dataset of piano music was created for each emotion and trained using Music Transformer models. Music was generated by inputting emotion words to the 4 models that had been trained. Two evaluation experiments were conducted on the generated songs. In the evaluation experiment of the music piece alone, we asked the subjects to listen to and evaluate 16 generated music. In the evaluation experiment of the videos as background music, we asked the subjects to watch 4 videos without sound, and determined the emotions of the videos. Then, 16 videos with background music were created by combining the 4 videos and 4 songs that contained each emotion. Then, we asked the subjects to watch the created videos and evaluate which background music was most suitable for them. For the experimental results, some of the models produced suitable music for the videos with expected emotions, but not all of the models were able to produce suitable music for the videos. In order to complete the system, we need to improve the training dataset and revise the experimental method.

Key words : music genelation, Transformer, BGM

1. 緒言

近年、動画サイトの使用率、使用人数は増加しており、加えて動画投稿者の数も増えている。また、ショート動画を投稿する機能に力を入れた SNS アプリが普及し、誰でも手軽に動画の制作や投稿ができる環境にあると

言える。加えて、多種多様なジャンルの動画、動画投稿者の増加によって日本を含め、世界の様々な分野に影響を与えている。このことから、動画というコンテンツの需要が増加傾向にあると考えられる。

動画制作の重要な要素の一つとして、適切な BGM の選択が挙げられる。動画の内容にマッチした BGM を付与することで、動画のクオリティ向上に繋がると考えられる。しかし、個人利用、商業利用が可能な BGM に絞ったとしても存在する楽曲数は膨大である。そのため、個人が気に入る BGM の選択、動画に合わせる編集などの作業には時間を要する。

本研究では、動画に含まれる感情を入力として、動画に適した BGM を自動生成するシステムの開発を目的とする。BGM を生成する手法には、深層学習モデルである Transformer を応用した音楽生成モデル Music Trans-

*大学院工学研究科 知識社会基礎工学専攻

*Fundamental Engineering for Knowledge-Based Society, Graduate School of Engineering

**知能システム工学講座

**Department of Human and Artificial Intelligent Systems

***仁愛女子短期大学 生活科学学科

***Jin-ai Women's College

****工学部 技術部

****Technical Division

former を用いる。Music Transformer には楽曲の感情を基に作成したデータセットを学習させる。学習させたモデルに感情を入力して生成された BGM を評価することで、動画に適した BGM が生成されているか判定する。

本論文では、2 章で現在の動画サイト等の動向、BGM 生成に関する研究について述べる。3 章では Music Transformer を用いた楽曲の生成について述べる。4 章で生成された楽曲の評価実験について説明し、5 章で評価実験の結果を示す。6 章で実験結果に対する考察を述べ、7 章で本論文についてのまとめを述べる。

2. 動画サイトの現状と BGM の自動生成

近年では、動画投稿を主軸とした SNS の普及によって、動画サイトや SNS の利用者数は増加傾向にある。加えて、操作が簡易化された動画作成アプリの登場などにより、動画の投稿者数も増加している。動画投稿者の増加に伴い、以前より多種多様なジャンルの動画が確立され、投稿される動画の本数も増加している。また、視覚と聴覚に情報を伝達できる動画というコンテンツの制作において、適切な BGM の選択は重要な要素の一つである。このことから、BGM の選択にかかる時間の簡略化、個人の気に入る BGM の生成という点に需要が存在すると考える。2.1 節で現在の動画サイト等の動向、2.2 節で動画と BGM の自動生成に関する過去の研究について述べる。

2.1 動画サイト等の動向

1. YouTube

「YouTube」は Google 社が運営する動画プラットフォームである。2021 年には、全世界で 23 億人以上（ソーシャルメディアユーザーの 50% 以上）が月に 1 回利用している。全世界ユーザーのうち、日本のユーザー数は 6000 万にも上る。また、投稿者として活動するために新しく開設されたチャンネルの数は 2019 年～2020 年にかけて約 2 倍に増加している。クリエイターの増加により、日本では音楽産業への貢献、学習ツールとしての需要増加、国を超えたカルチャーの発信などに影響を与えている^{[1][2]}。

2. TikTok

「TikTok」は 2016 年にサービスを開始し、2021 年には ios, Google Play を合わせると世界で最もダウンロードされたモバイル向けの動画プラットフォームである。2022 年には、月間アクティブユーザー数が 10 億人を超え、世界第 6 位のソーシャルネットワークアプリとなっ

た。全世界ユーザーのうち、約 50% が 34 歳以下、32.5% は 10 歳～19 歳の利用者であることから、若年層を中心に成長している SNS アプリだと言える。TikTok の特徴として、エンゲージメント率（視聴者の反応率）の高さが挙げられる。TikTok におけるマイクロインフルエンサーのエンゲージメント率は約 18%、メガインフルエンサーのエンゲージメント率は約 5% と他の動画プラットフォームの約 5 倍である。このことから、動画メディアによるマーケティング効果が大きく期待できるアプリだと言える^[3]。

これら以外にも、様々な動画プラットフォームがアクティブユーザー数や動画投稿数の面で成長を見せており、動画という情報メディアの需要が増加していることが伺える。

2.2 BGM の自動生成に関する研究

BGM の自動生成に関する研究は数多く存在する。特に近年では、自然言語処理の分野に登場した深層学習モデル Transformer を応用した音楽生成モデル Music Transformer が注目されている。2018 年に Curtis Hawthorne らによって発表された研究では、ピアノ演奏の midi データで構成された大規模データセット「MAESTRO」を用いて、オリジナルのピアノ演奏を生成するモデルの開発とテストを行った^[4]。この研究では、モデルの一部に Music Transformer を使用しており、生成されたピアノ演奏楽曲はデータセットに含まれるピアノ演奏と同様の音楽的特性を持っているという結果が得られた。加えて、長期的な依存関係を含んだ約 1 分程度のピアノ演奏楽曲の生成が可能になった。

本研究の目的は、感情を基にして動画に適した BGM を自動生成するシステムの開発である。そのため、感情を基に収集した楽曲のデータセットを作成し、Music Transformer に学習させることで、ある感情の音楽的特徴を持った楽曲が生成できると考えた。この理由から、本研究では楽曲の学習、生成モデルに Music Transformer を使用することとした。

3. 楽曲の生成

本章では、Music Transformer を用いた感情語を入力とした楽曲の生成方法について述べる。まず、感情語の取得方法や設定について説明する。次に、使用する Music Transformer モデルについて説明し、モデルに学習させるデータセットの作成方法を述べる。最後に、Music Transformer を用いた楽曲生成の方法を説明する。

3.1 感情語の取得と設定

本研究では、動画の内容から楽しい、悲しいなどの感情語を取得し、取得した感情語を Music Transformer に入力することで感情に適した BGM の生成を行う。目標は、自動で動画内容から感情語を取得できるようにすることだが、今回は手動で動画に含まれる感情語を取得した。対象者に動画を視聴してもらい、どのような感情を含んでいるか判定してもらうことで、動画の持つ感情語を決定した。得られた感情語を Music Transformer への入力に使用した。また、感情には様々な種類の表現が存在するが、今回は動画から取得する感情を楽しい、穏やか、悲しい、恐ろしいの 4 種類に設定した。

3.2 使用モデルの説明

本研究では、データセットの学習、楽曲の生成に Music Transformer を使用する。Music Transformer は 2018 年に Google LLC が発表した Transformer を用いた自動作曲 AI である^[5]。回帰型ニューラルネットワーク (RNN) や長短期記憶ネットワーク (LSTM) の登場により、楽曲 (時系列データ) の過去情報を長期的に記憶することが可能になった。Music Transformer では、情報の長期記憶性能が強化されている。これによって、従来のモデルでは困難だった繰り返しフレーズ (楽曲中に複数回登場する似たメロディ) を持った楽曲の生成が可能になった。また、音楽には文脈性が存在する。音楽では、前の時系列データ情報から、次に使用できる音階 (スケール) や曲の調 (キー) などが決定される。しかし、長い時系列になるほど文脈性の表現は難しくなる。Music Transformer の特徴として、音楽の文脈性を従来のモデルより上手に表現できることも挙げられる。

3.3 データセットの作成方法

今回使用する Music Transformer モデルでは、学習データ、楽曲生成時の入力データにピアノ楽曲の midi データが必要となる。そのため、以下の配布サイトでピアノ BGM を収集し、データ形式を変換することで、データセットの作成を行った。

- 音楽の卵 (<https://ontama-m.com/>)
- 甘茶の音楽工房 (<https://amachamusic.chagasi.com/>)
- ポケットサウンド (<https://pocket-se.info/>)
- 魔王魂 (<https://maou.audio/>)

はじめに、各配布サイトにて、4 種類の感情 (楽しい、穏やか、悲しい、恐ろしい) をキーワードとして検索を行い、ピアノ BGM を wav データ形式でダウンロードした。そして、私自身がダウンロードした BGM を聴き、感

情に沿った楽曲であるか判断を行った。楽曲は各感情で 20 曲、計 80 曲収集した。その後、オンラインファイルコンバーター AnyConv^[6] を用いて、収集した楽曲を wav データから midi データに変換した (図 1)。これらの作業を行い、各感情 (楽しい、穏やか、悲しい、恐ろしい) のデータセットを作成した。



図 1: データセットの作成工程

3.4 学習方法と生成方法

入力する感情語に適した楽曲を生成するためには、各感情を持つ楽曲の特徴を学習したモデルが必要になる。そこで、3.2 節で説明した各感情のピアノ楽曲データセットをそれぞれ Music Transformer に学習させる。楽しいピアノ楽曲 20 曲分のデータセットを学習させた FUN モデル、穏やかなピアノ楽曲 20 曲分のデータセットを学習させた RELAX モデル、悲しいピアノ楽曲 20 曲分のデータセットを学習させた SAD モデル、恐ろしいピアノ楽曲 20 曲分のデータセットを学習させた FEAR モデルの計 4 種類の楽曲生成モデルを作成する。作成したモデルを用いて既存の BGM 数秒を始めとしたオリジナル楽曲を生成する。

作成した 4 種類のモデル (FUN モデル、RELAX モデル、SAD モデル、FEAR モデル) を用いて BGM の生成を行う。学習させたモデルに感情語を入力すると、まずモデルは入力された感情語と関連性のある既存の BGM を選択する。そして、選択された BGM の冒頭数秒を始めとして、選択された BGM の持つ音楽的情報 (使用されている音、音の長さなど) を基に続きとなるメロディーを生成する。最後に、生成された BGM を wav 形式で出力する (図 2)。

今回は評価実験に使用するために、4 種類のモデルでそれぞれ楽曲を 20 曲 (合計 80 曲) 生成を行った。生成された BGM の長さは約 30 秒~1 分 30 秒だった。

4. 楽曲の評価実験

本章では、生成した楽曲が動画 BGM として適しているのか判定するために 2 種類の評価実験を行う。4.1 節では生成楽曲の評価実験について、4.2 節では生成楽曲を動画の BGM として使用した際の評価実験について説明する。

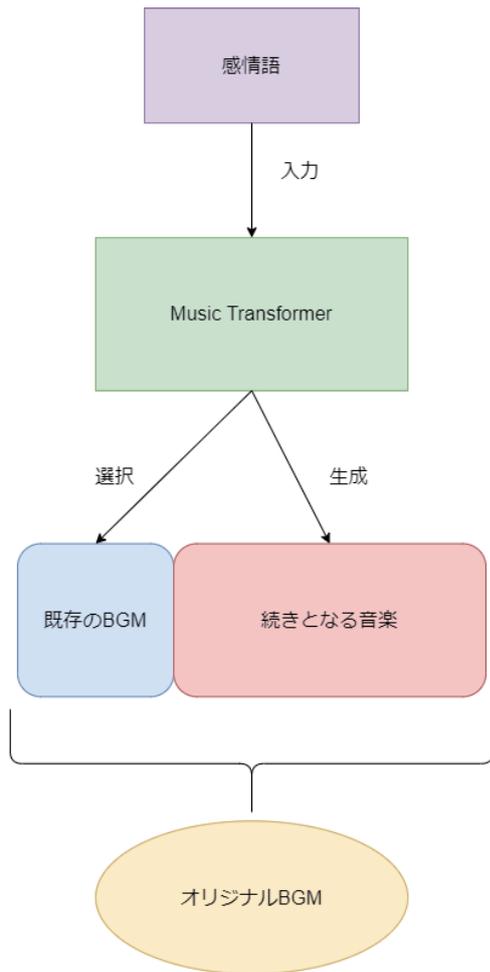


図 2: BGM 生成の工程図

4.1 生成楽曲の評価実験

各モデルから生成された楽曲が期待する感情を含んでいるか確認するために評価実験を行った。まず、各モデルで生成した楽曲 80 曲 (モデルごとに 20 曲) の中から、私自身が対象者に聴いてもらう楽曲を 16 曲 (モデルごとに 4 曲) 選定した。その後、対象者 4 名に選定した 16 曲を聴いてもらい、楽曲が持つ感情を評価してもらった。また、対象者 4 名は全員 20 代の男子大学生である。

楽曲の評価実験には、対象者が楽曲の持つ感情を定量的に評価する指標が必要である。今回は評価指標として、曲に含まれる各感情 (楽しい, 穏やか, 悲しい, 恐ろしい) に 1~4 の 4 段階の値を設定した。設定した値が 4 に近いほど楽曲は対象の感情を含んでおり (強く感じる), 1 に近いほど対象の感情を含んでいない (あまり感じない) ものとした。この評価指標を基に対象者に楽曲を聴かせ、各感情の値を決定してもらった。このようにして、楽曲にどの感情が多く含まれているか、各感情がどの程度含まれているのかを評価した。

4.2 動画 BGM としての評価実験

生成された楽曲が BGM として動画に適しているか確認するために、動画と生成楽曲を組み合わせて評価実験を行った。こちらの実験も、4.1 節の評価実験と同様の人物に対象者とした。

まず、対象者 4 名に音が付いていない動画 4 本を視聴してもらい、動画から感じた感情を評価してもらった。今回の評価実験では、フリー素材配布サイト Pixabay^[7] の動画を使用した。動画の評価指標は 4.2 節で説明したのと同様に、各感情 (楽しい, 穏やか, 悲しい, 恐ろしい) に 1~4 の 4 段階の値を設定した。対象者に値を決定してもらうことで、どの感情を多く含む動画か評価した。

その後、4.1 節の評価実験で感情値が高かった楽曲を感情ごとに 1 曲ずつ (計 4 曲) を選定し、評価してもらった動画 4 本に BGM として組み合わせた。つまり、動画 4 本と楽曲 4 曲を組み合わせることで、BGM 付きの動画を合計 16 本分作成した (図 3)。

そして、作成した動画 16 本を対象者 4 名に視聴してもらい、各動画に対してどの感情を多く含んだ楽曲が一番適していたか評価してもらった。

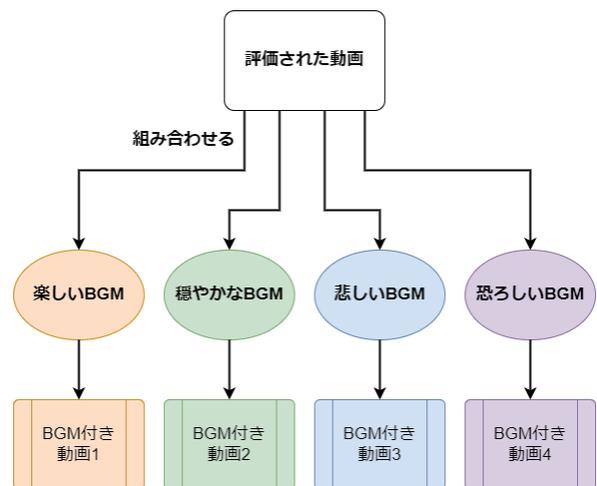


図 3: ある評価された動画に対する BGM の組み合わせ

5. 評価結果

本章では、4 章で説明した 2 種類の評価実験の結果を示す。5.1 節では、生成楽曲の評価実験結果、5.2 節では、生成楽曲を動画 BGM として使用した際の評価実験結果を示す。

5.1 生成楽曲の評価結果

各モデル (FUN モデル, RELAX モデル, SAD モデル, FEAR モデル) から生成した楽曲の評価平均を表 1~4 に示す。また、表の数値は楽曲の評価実験で対象者 4 名

が評価した各感情の数値(1~4)の平均を取ったものである。

RELAX モデルからは穏やかな楽曲が生成されることを期待しており、期待通り生成された楽曲の多くが穏やかに感じると評価された(表2)。SAD モデルからは悲しい楽曲が生成されることを期待しており、期待通り生成された楽曲の多くが悲しく感じると評価された(表1)。FEAR モデルからは恐ろしい楽曲が生成されることを期待しており、期待通り生成された楽曲の多くが恐ろしく感じると評価された(表4)。これらの結果から、ほとんどのモデルからそれぞれに期待する感情を多く持っている楽曲が生成されていたと言える。しかし、FUN モデルからは楽しい楽曲が生成されることを期待していたが、生成された楽曲の多くが穏やかに感じると評価された(表1)。

表1: FUN モデル生成楽曲の評価平均

曲	楽しい	穏やか	悲しい	恐ろしい
Acc1	2.75	3.5	1.25	1
Acc2	2.25	3.25	1.75	1
Loss1	3.5	2	1	1
Loss2	1.75	3.5	1.75	1

表2: RELAX モデル生成楽曲の評価平均

曲	楽しい	穏やか	悲しい	恐ろしい
Acc1	3.5	1.25	1	1.75
Acc2	1	3	2.5	1.25
Loss1	1.25	2.5	2.25	1.5
Loss2	2.5	3.25	1.75	1

表3: SAD モデル生成楽曲の評価平均

曲	楽しい	穏やか	悲しい	恐ろしい
Acc1	1.25	2.75	3.5	1
Acc2	1	1.5	3	1.25
Loss1	1	2.5	3.5	1
Loss2	1.25	3	2.25	1

表4: FEAR モデル生成楽曲の評価平均

曲	楽しい	穏やか	悲しい	恐ろしい
Acc1	1	1.25	2.25	3.25
Acc2	1	1.75	2.75	2.25
Loss1	1.75	2.75	1.5	1.75
Loss2	1	1	1.5	2.75

5.2 動画 BGM としての評価結果

対象者に映像のみで動画を評価してもらった結果、動画1,2は明るい(楽しい, 穏やかな)感情, 動画3,4は暗い(悲しい, 恐ろしい)感情を持っているという評価が得られた。評価してもらった動画と合わせた際の各BGMの適合率を表5に示す。また、表の数値は対象者4名が各動画に対して、最も適していると評価したBGMの割合を示している。

暗い感情を持った動画に対しては、恐ろしいBGMが最も高い適合率が得られた。しかし、悲しいBGMは適していないと評価された。反対に、明るい感情を持った動画に対しては、悲しいBGMが最も適していると評価された。また、全体的に楽しいBGMや穏やかなBGMは適合率が低い結果となった。これらの結果から、動画に対して適切なBGMを付与できるように多くの改善が必要である。

表5: 動画に対する楽曲の平均適合率

	動画1	動画2	動画3	動画4
BGM1(楽しい)	25%	50%	0%	25%
BGM2(穏やか)	25%	0%	25%	25%
BGM3(悲しい)	50%	50%	0%	0%
BGM4(恐ろしい)	0%	0%	75%	50%

6. 考察

本章ではTransformerモデルを用いて生成した楽曲の評価実験の考察を述べる。6.1節では生成楽曲単体の評価実験の考察、6.2節では動画BGMとして使用した際の評価実験の考察を述べる。

6.1 生成楽曲の評価実験考察

表2,3,4より、RELAXモデル、SADモデル、FEARモデルではそれぞれのモデルに期待する感情を含んだ楽曲が多く生成されているという結果が得られた。このことから、作成した4種類のモデルのうち、3種類(RELAXモデル、SADモデル、FEARモデル)は各感情を持つ楽曲生成に適したモデルであったと言える。それと同時にこれらのモデルに学習させたデータセット(楽曲)は、それぞれの感情(穏やか、悲しい、恐ろしい)を持つ楽曲生成に適していたと言えるだろう。

しかし表1より、FUNモデルからは期待した感情と異なる感情を多く含んだ楽曲が生成された。また、RELAXモデル、SADモデル、FEARモデルからも少数だが期待する感情以外の感情評価が高い楽曲が生成された。これらの結果から各学習データセットの内容の見直しが必要であると考えられる。今回は私自身の判断でそれぞれ

のデータセットに使用する楽曲を選定したが、自分以外の人間にもデータセットに使用する楽曲を評価してもらうべきだと考える。多くの人間に楽曲の持つ感情を評価してもらうことで学習に用いる楽曲の一般的な評価を得ることができ、より大勢の感情とマッチする学習データセットの作成に繋がるだろう。学習データセットの改良を行うことで、学習させたモデルから、より期待する感情を多く含んだ楽曲が生成されるようになることを考える。

また、生成した楽曲の中には、終盤のメロディーが音楽として聴きづらいものがいくつか存在し、楽曲の完成度に差が見られた。このような楽曲が生成された原因として、学習する楽曲が少なかったことが考えられる。今回は一つのモデルに学習させた楽曲数は20曲だったが、学習させる楽曲数を増やすことで、より完成度の高いオリジナルBGMの生成が可能になると考える。

6.2 動画BGMとしての評価実験考察

表5より、恐ろしいBGMが動画3では平均適合率75%、動画4では50%と、暗い感情を持った動画に対して最も高い適合率が得られた。また、楽しいBGMは動画2において平均適合率50%という数値が得られた。これらの結果から、FEARモデルからは動画に適した恐ろしいBGMが生成されていると言える。FUNモデルに関しても、今回の実験では動画に適した楽しいBGMが生成できたと考える。

しかし、悲しいBGMは動画3、4ともに平均適合率0%と、暗い感情を持った動画に対して適合率が低いという結果になった。反対に動画1、2では平均適合率50%と、明るい感情を持った動画に対して適合率が高い結果が得られた。このような結果となった理由として、楽曲が悲しい以外の感情を多く含んでいた可能性が考えられる。楽曲の評価実験では悲しい感情を最も多く含んでいるという結果だったが、悲しい以外の感情も多く含んでいたことによって、動画と組み合わせた際の評価実験では対象者の評価が変化したと考えられる。また、楽しいBGM、穏やかなBGMは多くの動画に対して適合率が低い結果となった。この結果から、これらの楽曲には感情の音楽的特徴があまり含まれていなかった可能性がある。音楽的特徴があまり含まれていないことにより、対象者の印象に残りにくく、他の楽曲の方が適していると判断されたと考えられる。ただし、これらの問題は6.1節で述べた学習データセットの改良によって解決できる可能性が高いだろう。

また、今回は20代の男子大学生4名に対象者となってもらい、評価実験を行った。しかし、動画サイトやSNSでは中学生や高校生、30歳以上の人々、女性の方も動画

を投稿している。そのため、性別や年齢の異なる人々を対象者として評価実験を行うことで、より一般的な評価結果を得られると考える。加えて、今回は対象者の結果全体を平均して評価を行ったが、対象者の個々の結果に注目して評価することも必要だと考えている。

7. 結言

本研究では、Transformerモデルを用いて4種類の感情語(楽しい、穏やか、悲しい、恐ろしい)からBGMを生成し、生成された楽曲に対して評価実験を行うことで動画に適したBGMの生成ができていないか判定した。今回は動画に含まれる感情を基に生成を行うため、データセットを感情ごとの楽曲に分けて4種類作成した。楽曲の学習、生成にはMusic Transformerモデルを使用し、4種類のデータセットをそれぞれ学習させることで4種類のモデル(FUNモデル、RELAXモデル、SADモデル、FEARモデル)を作成した。各モデルに感情語を入力し、既存楽曲の冒頭数秒に続くメロディーを生成したものがオリジナル楽曲として出力された。楽曲は各モデルで20曲(合計80曲)生成を行った。

生成された楽曲に対しては、楽曲単体の評価実験と動画BGMとしての評価実験を行った。楽曲単体の評価実験では、対象者4名に生成した80曲の中から選定した16曲を聴いてもらい、評価してもらった。評価指標には曲に含まれる各感情に1~4の4段階の値を設定した。値が4に近いほど感情を含んでおり、1に近いほど感情を含んでいないものとし、対象者には各感情の値を決定してもらうことで、楽曲にどの感情が多く含まれているかを評価した。評価結果より、作成した4種類のモデルのうち、3種類(RELAXモデル、SADモデル、FEARモデル)は各感情を持つ楽曲が生成された。しかし、FUNモデルからは期待した感情と異なる感情を多く含んだ楽曲が生成された。

動画BGMとしての評価実験では、対象者4名に音の無い動画4本を視聴してもらい、動画の持つ感情を決定した。その後、評価してもらった動画4本と各感情を含んだ楽曲4曲を組み合わせてBGM付き動画を16本作製した。そして、作製した16本の動画を対象者に視聴してもらい、動画に対してどのBGMが最も適しているかを評価してもらった。評価結果より、恐ろしいBGMは暗い感情を持った動画に対して最も高い適合率が得られた。しかし、悲しいBGMは暗い感情を持った動画に対して適合率が低く、反対に明るい感情を持った動画に対して適合率が高い結果が得られた。また、楽しいBGM、穏やかなBGMは多くの動画に対して適合率が低い結果となった。

本研究では、感情を基にして動画に適した BGM を自動生成するシステムの開発を研究目的とした。評価実験の結果から、作成した一部のモデルからは期待した感情を含み動画に適した楽曲が生成されたが、全てのモデルからは動画に適した楽曲が生成できなかったため、学習楽曲の再評価や楽曲数の増加など、学習データセットの改良が必要であると考えた。また、評価実験の対象者の年齢層の拡大や対象者個々の結果に対して評価を行うことで、より一般的な評価を得ることが必要だと考える。

また、今回は感情ごとにデータセットを作成し、学習させた 4 種類のモデルから楽曲を生成した。しかし、学習させていたデータセットを一つにまとめてモデルに学習させ、生成した楽曲の評価を行うことで、どちらの形態が研究目的である動画に適した BGM を自動生成するシステムに適切か比較する必要がある。

参考文献

- [1] "YouTube をめぐる 16 の統計データ", <https://www.infocubic.co.jp/blog/archives/15518/> (2022/4/11)
- [2] Oxford Economics, "YouTube Impact Report : 2021 年 日本における YouTube の 経 済 的 ・ 社 会 的 ・ 文 化 的 影 響", <https://www.oxfordeconomics.com/resource/a-platform-for-japanese-opportunity-assessing-the-economic-societal-and-cultural-impact-of-youtube-in-japan-in-2021-jp/>(2022/4/11)
- [3] Business of Apps, "TikTok Revenue and Usage Statistics (2022)", <https://www.businessofapps.com/data/tik-tok-statistics/>(2022/4/11)
- [4] Curtis Hawthorne, Andriy Stasyuk, Adam Roberts, Ian Simon, Cheng-Zhi Anna Huang, Sander Dieleman, ... Douglas Eck, "Enabling Factorized Piano Music Modeling and Generation with the MAESTRO Dataset", In ICLR(2019)
- [5] Cheng-Zhi Anna Huang, Ashish Vaswani, Jakob Uszkoreit, Noam Shazeer, Ian Simon, Curtis Hawthorne, ... Douglas Eck, "Music Transformer: Generating Music with Long-Term Structure", In ICLR(2019)
- [6] "無料のオンラインファイルコンバータ-AnyConv", <https://anyconv.com/ja/>(2022/6/18)
- [7] "高品質なフリー動画素材-Pixabay", <https://pixabay.com/ja/>(2022/6/20)

