公立はこだて未来大学 2021 年度 システム情報科学実習
グループ報告書

Future University-Hakodate 2021 System Information Science Practice
Group Report

プロジェクト名

脳をつくるプロジェクト

Project Name

Make Brain Project

グループ名

ソマティック・マーカー仮説に基づいた GAN による音楽生成

Group Name

Music Generation with GAN based on the Somatic Marker Hypothesis

プロジェクト番号/Project No.

20-C

プロジェクトリーダ/Project Leader

石川慶孝　Yoshitaka Ishikawa

グループリーダ/Group Leader

高野凌太　Ryota Takano

グループメンバ/Group Member

伊村尚矢　Takaya Imura
釜石健太郎　Kentaro Kamaishi
高野凌太　Ryota Takano
松田祐輔　Yusuke Matsuda

指導教員

香取勇一　佐々木博昭　佐藤直行　ヴラジミール リアボフ

Advisor

Yuichi Katori　Hiroaki Sasaki　Naoyuki Sato　Volodymyr Riabov

提出日

2022 年 1 月 19 日

Date of Submission

January 19, 2022

概要

本プロジェクトは、新しい人工知能を作成することや脳の仕組みの現実問題への応用を目標としている。その中で、C グループでは人工知能に脳の仕組みを取り入れることを目的に活動していた。私たちは、脳の仕組みの中でも脳の感情処理をテーマに選び、感情と意思決定に関する理論である「ソマティック・マーカー仮説」に着目した。ソマティック・マーカー仮説は、簡単に言えば「感情が身体的反応として現れることで、脳がそれを知覚して意思決定を効率化している」という仮説である。「ソマティック・マーカー」は、この身体的反応によって脳に送られる信号である。ソマティック・マーカー仮説と組み合わせる人工知能には、ディープラーニングの手法である GAN（Generative Adversarial Networks）を選んだ。GAN とは生成系のネットワークの一つで、主に画像の生成や変換に用いられる。GAN には、GAN 自身が生成したデータと訓練データを判別する Discriminator という部分がある。私たちはその判別を意思決定と見なし、ソマティック・マーカー仮説に基づいた意思決定の効率化ができると考えた。実装は、既存の音声生成用の GAN である WaveGAN[1] に音楽から受ける感情を分類するブロックを追加し、分類された感情をソマティック・マーカーとして Discriminator に渡すことで行った。分類結果の感情は、一般的な One-Hot ベクトル表現ではなく、Softmax 関数が出力する確率分布として表現され、条件付けに使われるのが特色である。また、評価として、WaveGAN でも同様に音楽生成を行い、学習曲線や評価実験の結果を比較した。すると、学習速度は劣るが、ある程度長く学習を行えば有意な差があるとは言えなくなることが明らかになった。また、生成される楽曲の感情を指定できる点は従来の GAN にはない特徴である。今後は、ベースとなる GAN の種類や学習データ、訓練データの楽曲に対する感情の決め方を再検討し、他の手法を上回るようにモデルを改良することが望まれる。

キーワード　人工知能, 脳の感情処理, ソマティック・マーカー仮説, GAN, 音楽生成

（文責: 釜石健太郎）

Abstract

The goal of this project is to create a new artificial intelligence and to apply the brain mechanism to real world problems. In this context, Group C worked to incorporate brain mechanisms into artificial intelligence. We chose the brain's emotional processing as one of the brain mechanisms, and focused on the Somatic Marker Hypothesis, a theory about emotions and decision making. The Somatic Marker Hypothesis is a hypothesis that the brain perceives emotions as physical reactions, which in turn improves the efficiency of decision making. Somatic markers are the signals sent to the brain by these physical reactions. We chose GAN (Generative Adversarial Network), a deep learning method, as the artificial intelligence to combine with the Somatic Marker Hypothesis. GAN is a generative network, mainly used for image generation and transformation. GAN has a part called Discriminator that discriminates between the data generated by the GAN itself and the training data. We considered the discrimination as decision making, and thought that we could improve the efficiency of decision making based on the Somatic Marker Hypothesis. The implementation was done adding a block that classifies the emotions received from music to the GAN, and passing the classified emotions to the Discriminator as somatic markers. The emotions in the classification results are represented as probability distributions output by the Softmax function, rather than the general One-Hot vector representation, and are used for conditioning. As an evaluation, we also performed music generation on a conventional GAN as well, and compared the learning curves and the results of evaluation experiments. It was found that the learning speed was inferior, but the difference became less significant after a certain length of training. In addition, the ability to specify the emotion of the generated music is a feature not found in conventional GANs. In the future, it is hoped that the type of base GAN, the training data, and the method of determining the emotion for the songs in the training data will be re-examined, and the model will be improved so that it can outperform other methods.

Keyword Artificial Intelligence, Emotional Processing in the Brain, Somatic Marker Hypothesis, GAN, Music Generation

（文責: 釜石健太郎）

目次

第 1 章　はじめに

1.1　背景

本プロジェクトは、新しい人工知能を作成することや脳の仕組みの現実問題への応用を目標として活動している。その中で、本グループは脳の感情処理に焦点を当て、その仕組みを人工知能に活用したいと考えた。そこで、脳の感情処理に関する仮説である「ソマティック・マーカー仮説」に着目した。

ソマティック・マーカー仮説は、Damasio らによって提唱され、Damasio の著書『生存する脳 —— 心と脳と体の神秘』[2] によって知られるようになった。Damasio は、情動とは単なる心の評価的なプロセスではなく、そのプロセスに対する身体的反応を含むものであると考えた。また、感情の本質は、情動や情動の微妙な変化などを引き起こしたメンタルイメージを並置しながら、身体的変化をモニターするプロセスであるとしている。例えば、久しぶりに好きな人に会えて生じて「嬉しい」という状況を考えると、その情動は鼓動が速くなること、顔が赤らむこと、表情が緩むことなどを含み、「好きな人」から生じるメンタルイメージと共にその変化を知覚することが感情にあたるだろう。ソマティック・マーカー仮説とは、この考えに基づいた「意思決定の際に、以前その情報によって経験した身体的反応が呼び起こされ、それを脳が知覚することで決定が効率化される」という仮説である。

より具体的には以下のように説明される [3]。初めに扁桃体と VMPFC(腹内側前頭前皮質) が過去の経験から対象に感情的な意味づけをし、身体的反応を形成する。この身体的反応による信号は「ソマティック・マーカー」と呼ばれる。脳には、体性感覚皮質や島をはじめとする身体状態を監視し続けている部分があり、ソマティック・マーカーが投射されたこれらの部位が、推論を司る前頭領域にさらに投射を行うことによって、意思決定にバイアスを与える。このバイアスが選択肢を減らすことで、意思決定の効率化が起こる。ソマティック・マーカーは選択肢を減らすものであるから、コストや利益の計算を伴う考慮よりも先に、無自覚に生じると仮定する。また、扁桃体や VMPFC が起こるはずの身体的反応の予測をし、身体をバイパスして、体性感覚皮質に起こるはずだった活動パターンを作り出すように命じる「あたかも身体ループ」があることも Damasio は指摘している。この「あたかも身体ループ」は、ある意思決定と罰や報酬と関連した身体状態が成長する中で繰り返し現れることで、身体状態に依存することをやめたことによるものであると Damasio はしている。

人工知能に関しては、「GAN(Generative Adversarial Networks)」がディープラーニングの分野で近年注目を集めている。GAN は生成系のネットワークの一種である。生成物やネットワークの構造によって種類が多くあるが、基本的には学習データに近いものの生成を行う。GAN では、Generator と Discriminator という二つのネットワークからなり、Generator は「偽物」のデータを生成し、Discriminator は与えられたデータが「本物」である訓練データか「偽物」のデータであるかを判別する。学習を進める中で、Generator が Discriminator を欺くのが上手くなっていくことで、訓練データに近い精巧なデータを生成できるようになる。

私たちは、この Discriminator の本物か偽物かという判別を意思決定とみなし、ソマティック・マーカー仮説による意思決定の効率化の仕組みを取り入れることができると考えた。

（文責: 釜石健太郎）

1.2 目的

本グループは、ソマティック・マーカー仮説に基づいた GAN のアーキテクチャを作成し、音楽生成を行わせることを目的としている。既存の音声生成のための GAN として、WaveGAN が知られている。与えた音楽から受ける感情を分類する「感情分類器」を WaveGAN へ組み込み、その感情を Discriminator に与えるようにすることで、ソマティック・マーカー仮説の状況を再現する。後述するように WaveGAN は音声ファイル形式の一種 wav をデータ形式として扱っており、本プロジェクトの音楽生成における音楽は MIDI(Musical Instrument Digital Interface) のような離散的な音階表現ではなく、連続的な波として表現される。音楽を選んだのは、画像や文章よりも感情と結びつけやすいためである。また、楽曲の質を調査するための評価実験を、提案する GAN と WaveGAN のそれぞれにより生成された楽曲を用いて行い、結果を比較をすることで評価を行う。また、Generator と Discriminator の損失をプロットした学習曲線の変化についても比較を行う。

本グループの取り組みによって、ソマティック・マーカー仮説を GAN に組み込む有用性の判明や新しい音楽生成手法の確立が期待される。

（文責: 釜石健太郎）

第 2 章　序論

2.1　関連研究

2.1.1　GAN

GAN は敵対的生成ネットワークと呼ばれ、生成系ネットワークの一つである。GAN の特徴は従来の角度や色を変えたりするだけでなく、訓練データの特徴をもった新しいデータや情報を生成することである。一般的な GAN の構成は図 2.1 の通りである。GAN は Generater と Discriminator の二つからなっている。ノイズである乱数により Generator から偽物のデータを生成し、Discriminator で訓練データと比較し判断する。これを繰り返すことで、Generator は訓練データに近い精巧な偽物を作るようになり、Discriminator はより判断が上手になっていく。これを繰り返すことで、出力が訓練データに近いものが生成されるようになる。

（文責: 伊村尚矢）

2.1.2　DCGAN

DCGAN(Deep Convolution GAN)[5] とは GAN を改良させたモデルである。現在ある高解像度の生成ネットワークの根底にある GAN である。Generator に畳み込みニューラルネットワーク (CNN) を導入したことにより GAN の学習を安定させている。と Discriminator の両方に Batch Normalization を導入している。Generator では活性化関数に ReLU 関数を使用し、出力層のみ Tanh を使用している。Discriminator では活性化関数に LeakyReLU が全層にわたって使用されている。このようにして学習の安定化と高解像度の生成が可能となったものが DCGAN である。

（文責: 伊村尚矢）

2.1.3　WaveGAN

WaveGAN[1] は音声の生成として使われる GAN の手法の一つである。学習方法は GAN と同じである。訓練データにサンプルごとの時系列データとして wav ファイルを使用する。WaveGAN は DCGAN[2] を踏襲した上で音の特徴に合わせた変更がされている。DCGAN では 5 × 5 の 2 次元の畳み込みのところを、WaveGAN では 25 の 1 次元の畳み込みに変更されている。DCGAN よりもレイヤーを増やすことで、16kHz のサンプリング周波数で 1 秒程度の音を生成できるように構成されている。

（文責: 伊村尚矢）

2.1.4 Conditional GAN

Conditional GAN[6] の構成は図 2.2 の通りである。Conditional GAN は生成するクラスを指定できる GAN であることが特徴になる。DCGAN では、mnist データを学習に用いることで、高解像度な手書き文字の生成に成功している。しかし、DCGAN では使用される用途が限られてしまう。例えば、「5」という手書き文字を作りたいと思っても、DCGAN だと生成する文字を指定することが不可能であり、「5」が偶然生成されるまで生成器を動かし続ける必要がある。ここで、Conditional GAN を導入することで、「5」を指定して生成することが可能になる。GAN ではノイズが 1 つの入力であったが、Conditional GAN ではノイズベクトル：Z と、条件ベクトル：Y の 2 つのノイズを用いて Generator と Discriminator を学習させている。使用される訓練データは予めクラス分けを行い使用する。

（文責: 伊村尚矢）

2.1.5 Hevner による音楽の感情分類

音楽と感情を結び付ける際、感情は様々な形容詞で表現される。それらの形容詞には楽しい、陽気などとニュアンスがとても似ているものや、楽しい、悲しいといったように反対の意味を持つものがある。Hevner はこれら形容詞の相互関係を調べた。Hevner は被験者に音楽を聴いてもらい、その音楽に適していると思うものを 66 個の形容詞の中から選ばせた。その結果から似ている形容詞を 8 グループに分けることで分類を決定した。形容詞を個別に扱うのではなくグループに分けることで、音楽の感じ方のばらつきを少なくしている。グループは以下のように分けられている。

Group1: spiritual, lofty, awe-inspiring, dignified, sacred, solemn, sober, serious
Group2: pathetic, doleful, sad, mournful, tragic, melancholy, frustrated, depressing, gloomy, heavy, dark
Group3: dreamy, yielding, tender, sentimental, longing, yearning, pleading, plaintive
Group4: lyrical, leisurely, satisfying, serene, tranquil, quiet, soothing
Group5: humorous, playful, whimsical, fanciful, quaint, sprightly, delicate, light, graceful
Group6: merry, joyous, gay, happy, cheerful, bright
Group7: exhilarated, soaring, triumphant, dramatic, passionate, sensational, agitated, exciting, impetuous, restless
Group8: vigorous, robust, emphatic, martial, ponderous, majestic, exalting

グループ番号の差が 4 である感情は反対のものになるように分けられている。例えば、Group1 では spiritual, serious などが分類されており、Group5 は humorous, fanciful などが分類されている。

（文責: 松田祐輔）

2.2 課題の設定

プロジェクトの目標達成のために、ソマティック・マーカー仮説をはじめとする脳科学を学ぶ必要がある。そのために論文や参考書を読んで学習を行う。脳科学の知識を得るために「メカ屋のための脳科学入門: 脳をリバースエンジニアリングする」を参考にした。 また、既存のディープラーニングの手法を学ぶ必要があるため、「ゼロから作る Deep Learning」を輪読することによって、その仕組みを知り、自分たちの手で実装できるようになることを目指す。既存の GAN モデルを作るために、夏休みに「生成 Deep Learning 絵を描き、物語や音楽を作り、ゲームをプレイする」を輪読することを行う。

GAN の学習に使われる訓練データとして wav 形式のデータを採用した。学習データを選ぶ注意点として、ノイズが少なく聞き取りやすい音であるか、感情の分類に偏りはないか、などが挙げられる。モデル実装の準備として、学習に用いる音楽データに感情のラベルを手動で振らなければならない。そこでは客観性のある感情の分類手法を用いる必要がある。実装段階 では、初めに感情分類のモジュールを持たない既存の GAN モデルによる音楽生成を実現しなけれ ばならない。並行して感情分類のネットワークを別に作成することが必要である。この二つを実装した後に、これらを組み合わせることで提案するモデルの実装を行うことができる。

実装が終了したら 二つのモデルの学習効率や生成される音楽の比較・評価を行わなければならない。生成された音楽を被験者に聞かせ、楽曲の質についての設問に答えてもらう評価実験を行い、その結果に対して統計的手法を用いることで新しいモデルの評価を行う。さらに、音楽データの質ではなく、学習モデルの学習効率を測るために学習曲線から考察する。

ディープラーニングがどのような仕組みなのか理解するためには数学を理解している必要がある。 また、実装するにはプログラミングの技術が必要不可欠である。よって、関連する本学の講義として「解析学 I, II」、「線形代数学 I, II」、「情報処理演習 I」が挙げられる。

（文責: 高野凌太）

2.3 訓練データ

CNN や GAN の学習を進めるにあたって大量のデータが必要となる。そのデータ数は 2、3 万にもなる。データ数が少ないと過学習になり未知なデータに対して正確な予測が困難になる。そこで、大量のデータを確保することができるデータセットを用意することは重要である。また、データを十分に用意できなくてもデータを加工することによって数を増やす手法がある。

データを入手するのに必要なコストとして有償か無償で提供されているか気にする必要がある。予算に余裕があれば有償のデータを採用してもよい。しかし、クラッシク曲は無償で手に入れて活用できることが多かったので私たちは無償で手に入れることのできるデータセットのみを扱った。Google の機械学習のプロジェクトである Magenda が maestro というデータセットを提供している。maestro には 200 時間分のピアノ演奏した楽曲が保存されている。ファイル形式は wav、csv、MIDI であり、そのまま訓練データとして扱いやすい。これほどの大量のデータがあれば学習に困ることはない。

maestro はコンテスト発表の時に録音されたものでありノイズが混じっており、私たちの作成するモデルに適さなかった。しかし、先行研究で maestro が使われていることもあるので大量のデー

タを手に入れる手段として候補に挙げるべきである。機械学習の技術として augmentation というものがある。これもデータを水増しする手法であるがこれはデータを加工する必要がある。例えば、画像を反転させたり一部分を拡大、縮小したりして別の画像データとして扱うことができる。扱うデータが音であっても音をメルスペクトログラムと呼ばれる波形に変換して画像として扱うことができる。よって、augumentaion を音データを扱う上で活用することもできるが採用には至らなかった。そこで採用したデータの水増し方法は曲の編集である。曲を 1 秒ごとに区切り、重複しないように組み合わせることで一つの曲からたくさんのデータ数を生み出すことができた。機械学習のモデルの学習をする際に、どのデータを扱うか、どのようにしてデータを集めることができるか慎重になることが良いモデルを作る上で重要となる。

（文責: 高野凌太）

第 3 章　方法

3.1　提案手法

3.1.1　構成

今回使用するモデルの構成を図 3.1 に示す。また使用するモデルの名称をソマティック・マーカーから頭文字をとり、SM-GAN(Somatic Marker GAN) とする。

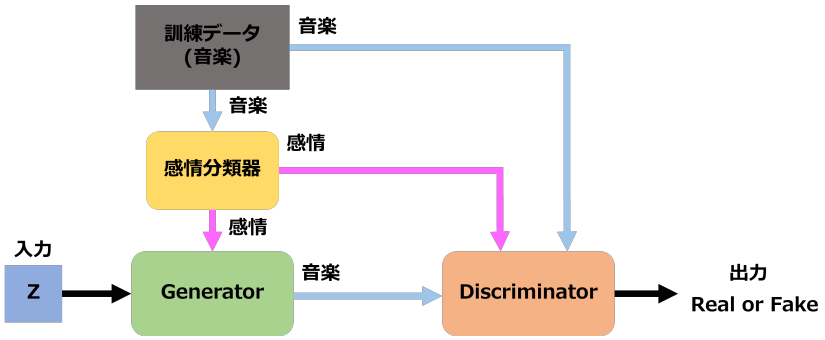

図 3.1　使用する GAN(SM-GAN) のモデル図

従来の GAN と異なるのは、感情分類モジュールが追加されている点である。感情分類器は、あらかじめ感情を判断できるように訓練データを用いて学習させた。訓練データを感情分類器に入力することで、各感情の確率分布のベクトルが得られる。そして Generator には分類されたラベルと乱数を入力し、Discriminator には Generator で生成された音楽データと分類されたラベルを入力する。

Generator と Discriminator に訓練データに付随しているラベルを入力するという方法も考えられたが、確率としてラベルを渡すことで訓練データにはどのような特徴を用いてラベルをつけられたのかを学習することができる。そのため、感情分類器を用いることとした。また、初期のモデル図では乱数からなるノイズを Generator と感情分類器に入力され、生成された音楽データに感情分類器が分類したラベルを付けて Discriminator に入力する。しかし、この構成では Discriminator に入力される Generator からの入力されるラベルと訓練データのラベルが一致せず、損失が大きくなってしまう。よって初期のモデルでは学習を進めることが困難であった。そこでモデルの改善案として今回の SM-GAN を提案した。

3.1.2　感情分類器

感情分類器の実装には、畳み込みニューラルネットワーク（以下 CNN）を用いた。今回使用した CNN の構成を図 3.2 に示す。

畳み込みを行う際、フィルタの大きさを変更して畳み込みを行う際、フィルタの大きさを変更して畳み込みを行う。これは音楽の時系列の関係性を考慮するためである。フィルタの大きさは (1, 4) と (4, 1)、(1, 8) と (8, 1) の 2 種類である。フィルタを Conv2D 層は異なるフィルタを持つ畳み込み層に活性化関数として ReLU 関数を用いて BatchNormalization を適用した。

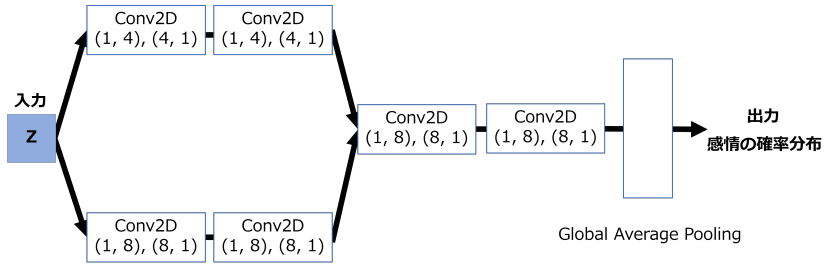

図 3.2　感情分類器に用いた CNN

BatchNormalization とはバッチ分のデータの同じチャンネルを平均 0、分散 1 として正規化する処理である。この処理によって学習を安定させ、学習速度を高めることができる。フィルタが異なる 2 つの層それぞれに訓練データが入力され、出力を結合しさらに畳み込みを行う。最後に Global Average Pooling 層に入力される。Gloval Average Pooling 層は各チャンネルの画素平均を求め、一次元ベクトルにする層である。この処理によって、最後の全結合層でのパラメータ数を減らすことができる。感情分類器の損失関数には交差エントロピー損失を使用した。また、最適化関数には Adam を使用した。最適化関数を決める際、Adam 以外にも SGD や最急降下法、Radam を使用し学習を試した。最も良い結果となったのは Radam であったが、python のバージョンを最新にする必要があった。バージョンを最新にした場合、GAN でバージョン違いによるエラーが発生したため、次に良い結果となった Adam を使用した。また、学習終了時の重みデータを pth ファイルで保存した。SM-GAN の学習時には、この pth ファイルで重みを固定した感情分類器を用いた。

（文責: 松田祐輔）

3.1.3　GAN

Generator は乱数で構成された一次元ベクトルを全結合層に出力する。その後、1 次元の転置畳み込み層に入力する。転置畳み込み層は 6 層あり、活性化関数に ReLU 関数を使用した。また最後の層には Tanh 関数を使用した。Discriminator は Generator から出力された音楽データを畳み込み層に入力する。活性化関数には LeakyReLU 関数を使用し、a 値を 0.2 とした。LeakyReLU 関数は入力された値が 0 未満である場合、入力に a 値を掛けた値を出力とする。また次の層に入力する前に Phase Shuffle を行う。Phase Shuffle は入力された特徴量に対し、ランダムに要素をずらす処理である。要素をずらして空いた部分には Reflection Padding を適用して埋める処理を行う。Generator が生成した音楽データには特定の位相で checkerboard artifacts という市松模様のようなパターンが出現してしまう。checkerboard artifacts は訓練データには普通現れないため、Discriminator が偽物だと判断してしまう。そのため Phase Shuffle を行うことで多様な位相で checkerboard artifacts を発生させ、Discriminator の学習を進めるようにした。Generator と Discriminator の損失関数には WGAN-GP を使用した。また最適化手法には Adam を使用した。

（文責: 松田祐輔）

3.1.4 訓練データ

訓練データは先行研究で使用されていた音楽データを用意した。用意したデータの中から感情を読み取れる部分をトリミングし、ラベルを付けた。ラベル付けはグループメンバー 4 人で行った。ラベル付けには、Hevner による音楽から受ける感情の分類方法を使用した。ここで使用した音楽データで該当する感情が極端に少ないグループは使用しなかった。学習を行う際、トリミングされたデータからランダムに 2 秒間でさらにトリミングを行い、元のラベルを付けた。これを設定したデータ数の分だけ行い訓練データとした。CNN で用いた訓練データの数は 2000 とし検証用データは 400 とした。また、SM-GAN の訓練データには、同様の方法で逐一バッチサイズ分のデータを切り出して作ったバッチが使用された。

（文責: 松田祐輔）

3.1.5 ソマティック・マーカー仮説との関連

ノイズからなる乱数より感情分類が行われ、生成された感情と乱数をもとに Generator で音楽を生成する。感情分類器は SM-GAN を学習する前に学習をしておくため、過去の経験と捉えることができる。ラベルとして付けられた感情は生成された音楽とともに Discriminator に入力されるため、過去の経験から感情を決定し、意思決定を支えている様子を再現していると考えることができる。

（文責: 松田祐輔）

3.2 評価手法

3.2.1 評価実験 1

評価実験 1 では、更新回数を 50000 回にして学習した WaveGAN と SM-GAN のそれぞれで生成した音楽について、「聞きやすさ」「ピアノ曲らしさ」についてアンケート形式で回答してもらい、比較することを目的とした。

被験者は、グループ C のメンバー以外の大学生 22 名、教員 2 名の計 24 名であった。実験は、2021 年 12 月 1 日から 12 月 4 日の被験者に都合の良い時間に行われた。装置として、インターネットが使える端末、イヤフォン、アンケートサービス「Google フォーム」を用いた。音声は、動画共有サイト「YouTube」上に動画としてアップロードすることで共有した。

刺激として、学習済みの WaveGAN と SM-GAN で生成された 2 秒のピアノ楽曲をそれぞれ 10 曲ずつ用いた。WaveGAN については、曲を 100 曲生成し、グループ C のメンバー 4 人で多数決を行うことで優れた 10 曲を選んだ。SM-GAN については、5 つある各感情のグループについて 20 曲ずつ生成し、WaveGAN と同様の方法でそれぞれから 2 曲ずつ選んだ。各楽曲は「Youtube」にアップロードするために、音声ファイルから動画に変換した。動画は、黒い背景に白文字で「実験音声」と書かれた画面が表示され続けるものであった。3.3 に Google フォームのスクリーンショットを示す。

図 3.3　アンケート画面

実験は以下の手順で行われた。被験者は初めに「得られたデータは授業のため以外で使わないこと」「2 秒のピアノ曲を動画で聴くこと」「計 20 試行からなり、約 20 分ほどかかること」を教示された。次のページにて、後述する質問項目となるべく一人で実験を行ってほしいことを教示された後、実験を開始した。初めの 10 試行では SM-GAN の楽曲、残りの 10 試行では WaveGAN の楽曲が使われた。また、提示される曲の順序は全被験者で共通であった。なお、被験者には、聞いている楽曲が WaveGAN、SM-GAN のいずれであるかを教示しなかった。各試行は、動画内の楽曲について、「聞きやすさ」「ピアノ曲らしさ」のそれぞれを 1 から 5 までの 5 段階評価で回答するものであっていた。なお、1 が「聞きづらい」「ピアノらしくない」、5 が「聞きやすい」「ピアノらしい」に対応していた。各試行の最後には、ピアノ曲を聞いて感じたことを任意の自由記述で回答する欄が設けられていた。最後の試行が終わると、任意で感想を自由記述する欄が設けられたページへ遷移し、被験者が送信ボタンを押すことで実験は終了した。

3.2.2　評価実験 2

評価実験 2 では、更新回数を 200000 回に変更して学習した WaveGAN と SM-GAN のそれぞれで生成した音楽について、「聞きやすさ」「ピアノ曲らしさ」「楽曲から受ける感情」についてアンケート形式で回答してもらい、比較することを目的とした。

被験者は、グループ C のメンバー以外の大学生 17 名であった。実験は、2021 年 12 月 24 日から 2022 年 1 月 7 日の被験者に都合の良い時間に行われた。装置、刺激については評価実験 1 と同様であった。実験の手順もおおむね同様であったが、以下の点が異なっていた。新たにサンプル音

声が聞けるページを試行の前に設け、曲の雰囲気の確認や音量調整をできるようにした。また、選ばれた 20 曲は、SM-GAN か WaveGAN かで分けず、ランダムな順で用いた。なお、曲の順序自体は全被験者で共通であった。さらに、各試行のページに楽曲から受ける感情を 5 つの選択肢から選ぶ項目を追加した。3.1.4 で述べた分類の各グループについて、含まれる感情の中でも代表的と思われる 3 つを以下のようにそれぞれ残して提示した。( ) 内は元々のグループ番号を示している。

- 哀れな、悲しい、暗い (Group2)
- 夢のような、優しい、感傷的な (Group3)
- 満足な、穏やかな、静かな (Group4)
- ユーモアのある、軽快な、優美な (Group5)
- 爽快な、情熱的な、興奮した (Group7)

（文責: 釜石健太郎）

3.3 開発プロセス

3.3.1 開発ツール

開発を進めるにあたって情報の共有方法や進行方法を効果的にすることは重要である。私たちが主に使用したコミュニケーションツールは LINE、Slack である。そして、開発方法としてアジャイル開発を採用した。LINE は日常的に使用しているコミュニケーションツールであり情報がすぐに伝わるという利点があった。しかし、重要なファイルを管理したり、異なる議題を同時に進行することは難しかった。その点は Slack を活用することで解消された。どちらにも優れた点があり同時に運用していくと効果的なコミュニケーションが可能となる。特にソースコードの共有方法として GitHub を活用した。自分のソースコードを管理しやすくなるだけでなく、他人のソースコードを参考にすることもできた。本グループは人工知能のモデルに csv ファイルや大量の wav データが必要になったのでそれらの管理にも貢献をもたらした。

3.3.2 進行方法

本グループはリーダーを決めたが、全員が同じ発言力を持ち責任を負うことを話し合いで決めた。そのおかげで他人任せをするのではなく全員が意見することが求められる。スケジュールを細かく設定していたのでリーダーがいなくても進行に困ることはなかった。アジャイル開発とは、計画、実装、テストまでを短期間で行い開発を進めていくことである。何度もテストすることをできたので多くの失敗作ができたとともに、締め切りまでに完成したモデルを作ることができた。また、この作業の過程を全員で共有することにより開発者が気付かなかったフィードバックが与えられた。また、他人の考え方から自分の開発に新たなアイデアが生まれたりした。アジャイル開発を通じて全員が批判や失敗を多く経験しながらより良いアイデアを生み出すことができた。

（文責: 高野凌太）

第 4 章　結果

4.1　学習曲線

評価実験 1 で用いた更新回数が共に 50000 回のときの WaveGAN と SM-GAN における損失の変化を表す学習曲線をそれぞれ図 4.1、図 4.2 に示す。点は更新回数 100 回ごとにプロットされている。Discriminator については、両グラフともに始めの数点を除き、損失がほぼ 0 の状態が続いていた。また Generator については、WaveGAN はおおよそ 0 から 10 の範囲を、SM-GAN はおおよそ −10 から 10 の範囲を振動していた。

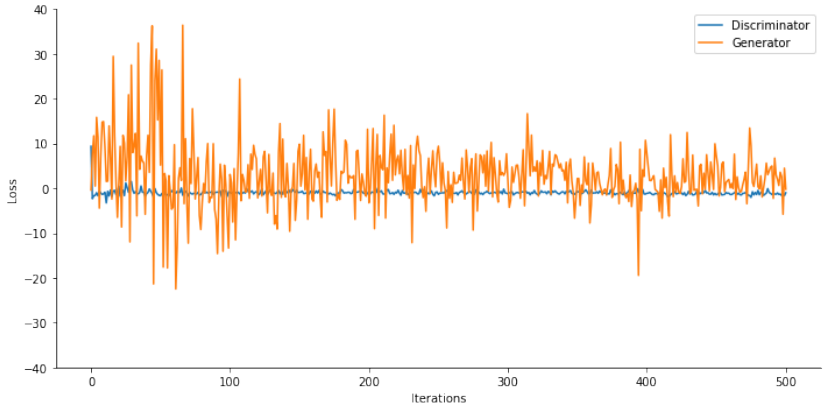

図 4.1　WaveGAN の学習曲線 (更新回数=50000)

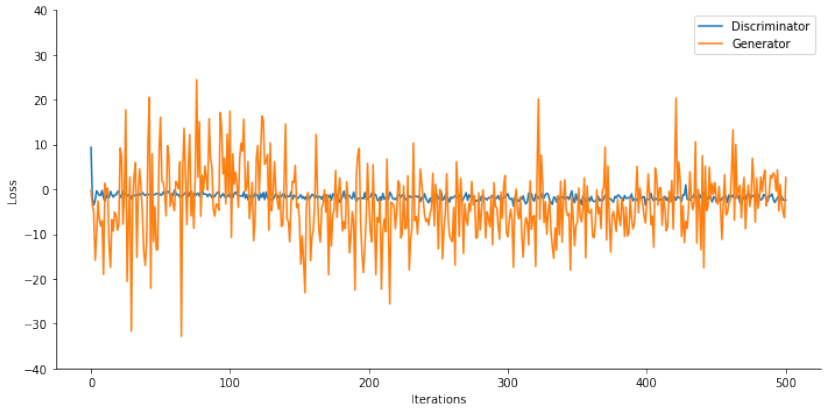

図 4.2　SM-GAN の学習曲線 (更新回数=50000)

次に、評価実験 2 で用いた更新回数が共に 200000 回のときの WaveGAN と SM-GAN における損失の変化を表す学習曲線をそれぞれ図 4.3、図 4.4 に示す。こちらも点は更新回数 100 回ごとにプロットされている。Discriminator については、両グラフともに 50000 回以降損失が徐々に下がり、最終的には −2 のあたりを小さい幅で振動する状態が続いていた。また Generator については、両グラフとも 50000 回以降は徐々に振動の幅が小さくなり、最終的には 0 から 5 のあたりを振動していた。

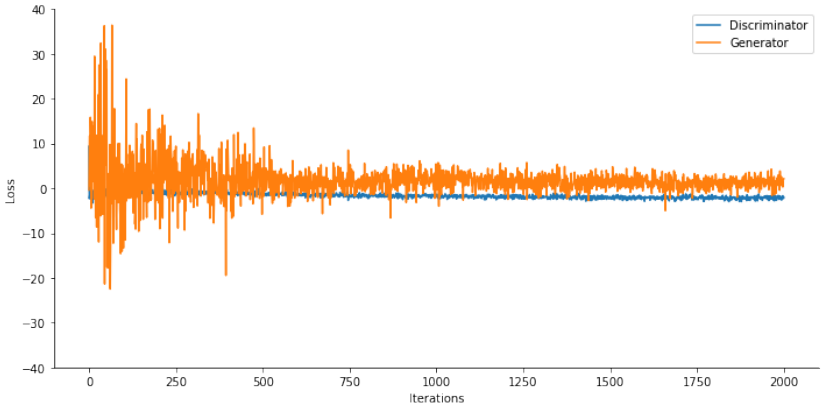

図 4.3　WaveGAN の学習曲線 (更新回数=200000)

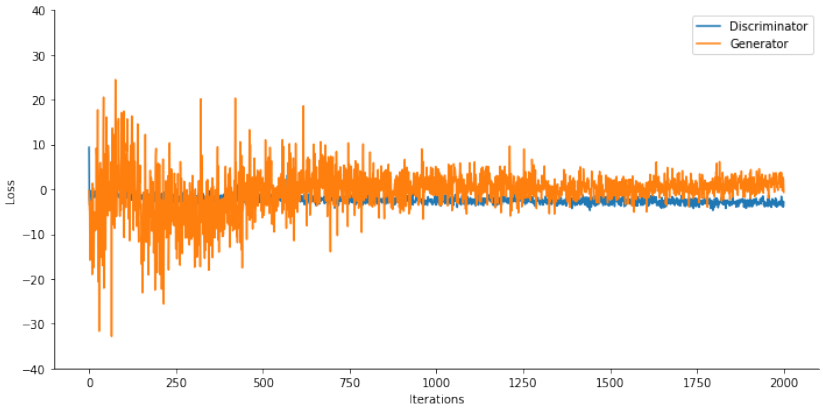

図 4.4　SM-GAN の学習曲線 (更新回数=200000)

（文責: 釜石健太郎）

4.2　評価実験 1

「ピアノ曲の聞きやすさ」と「ピアノ曲らしさ」の 2 項目について、WaveGAN と SM-GAN の 5 段階評価の平均値と標準偏差をそれぞれ表 4.1、表 4.2 に示す。両項目について、SM-GAN よりも WaveGAN の方が平均値が高く、ばらつきも少なかった。さらに両モデルについて、有意水準を 5% とした 2 つの母平均の差の両側 t 検定を行った。すると、「ピアノ曲の聞きやすさ」「ピアノ曲らしさ」について、有意確率は 0.005%、4.9% となり、WaveGAN の方が SM-GAN よりも平均値が有意に高かったことが示された。

表 4.1　WaveGAN の平均値と標準偏差 (評価実験 1)

	ピアノ曲の聞きやすさ	ピアノ曲らしさ
平均値	3.19	3.29
標準偏差	1.14	1.13

各楽曲への自由記述では、前半の SM-GAN の 10 曲には「子供が適当に遊んでいるみたい。」「音割れが大きい。」「汚い」といった否定的なものが多い中、「いい感じに聞こえる！」といった肯定的なものもいくつかはあった。後半の WaveGAN の 10 曲には「比較的聞ける」「屋外カフェに流

表 4.2　SM-GAN の平均値と標準偏差 (評価実験 1)

	ピアノ曲の聞きやすさ	ピアノ曲らしさ
平均値	2.83	3.07
標準偏差	1.17	1.27

れてそう」「Nothing special」といったやや肯定的、または中立的な物が多かった。

実験の感想には主に以下の 3 種類のものがあった。初めに、「少し音割れしているように感じた。でもメロディーがピアノぽいところもあって面白かった。」「音質悪くないか？反響してる気がする。」という音の悪さについての言及が 2 件あった。次に、「短くて判断するのが難しかった」「Looks great to me. perhaps somewhat longer pieces might be better」という曲の短さについての指摘が 2 件あった。最後に、「似た音楽が続くので、後半になるほど何を記述すべきかが曖昧になってくる。」「だんだん音に慣れてきて基準が下がっていった気がした。」という慣れによる評価の変化についての言及が 2 件あった。

（文責: 釜石健太郎）

4.3　評価実験 2

評価実験 2 では SM-GAN を改良後どれほどピアノ音の質を向上させることができたか、意図した感情をどれほど再現できているかアンケートを実施して調べるものであった。ピアノ曲の質を調べるのは評価実験 1 と同様に平均値の差の検定を行って調べる。感情を付与した音の再現性は χ^2 検定を行うことで明らかにする。SM-GAN と WaveGAN の生成する「ピアノ曲の聞きやすさ」と「ピアノ曲らしさ」の分析結果は下の図のようになった。

表 4.3　WaveGAN の平均値と標準偏差 (評価実験 2)

	ピアノ曲の聞きやすさ	ピアノ曲らしさ
平均値	3.44	3.52
標準偏差	0.95	0.98

表 4.4　SM-GAN の平均値と標準偏差 (評価実験 2)

	ピアノ曲の聞きやすさ	ピアノ曲らしさ
平均値	3.36	3.47
標準偏差	1.04	1.03

また、SM-GAN と WaveGAN のピアノ曲の聞きやすさとピアノ曲らしさの平均値の差の検定を行い、それぞれの有意確率は 45%、63% となった。どちらも有意確率が大きく生成するピアノ曲の質に違いが見られないことが分かる。評価実験 1 では WaveGAN よりも SM-GAN のほうが生成するピアノ曲の質が低いという結果であった。よって、SM-GAN を改善したことにより生成するピアノ曲の質が向上したと言える。次に、χ^2 検定を行うためのアンケート結果を分析すると以下の図になった。

有意確率は 0.0013 となり、「SM-GAN が意図して感情を付与させたピアノ曲」と「被験者がピ

表 4.5　被験者が意図した感情をもつピアノ曲をどのように感じたかアンケートの結果

	哀れな (Intended)	夢のような (Intended)	満足な (Intended)	ユーモアのある (Intended)	爽快な (Intended)
哀れな (Perceived)	29%	9%	6%	12%	6%
夢のような (Perceived)	38%	24%	18%	24%	15%
満足な (Perceived)	23%	12%	15%	18%	6%
ユーモアのある (Perceived)	9%	29%	35%	32%	44%
爽快な (Perceived)	0%	26%	26%	12%	29%

アノ曲の感情の感じ方」について独立性がないことが認められる。よって、SM-GAN が特定の感情を持つピアノ曲を選んで生成できていて、被験者はそれを特定の感情だと言い当てることができることが分かる。しかし、表 4.5 から対角成分を見ると必ずしも割合が最も高いというわけではなく、意図した感情を被験者が正確に聞き分けできているわけではない。感情の分類の境界が曖昧であるために被験者が感情を正確に判断することが難しかったと予想される。例えば、爽快な感情を持つピアノ曲を意図して生成して、被験者はユーモアのある、爽快な、のいずれかに 73% が答えている。

各楽曲の自由記述について、意図して生成した哀れな感情を持つピアノ曲について「哀愁漂う感じだった」、「暗い感じがした」という回答が寄せられた。また、意図して生成したユーモアのある感情を持つピアノ曲について「軽快さが印象的だった」、「明るい感じがした」という回答が寄せられた。このことから被験者は感情の違いがはっきりしている楽曲に関しては正確に分類できていることが分かる。また、実験への感想について「前回よりもピアノらしい曲が多かった」という回答があり、SM-GAN の性能の向上がコメントからも判断できる。

（文責: 高野凌太）

第 5 章　考察

5.1　ネットワーク

SM-GAN は WaveGAN よりラベルがついているため特徴量が多くなっていた。学習は同じ回数を繰り返したため、ラベルの特徴量分だけ SM-GAN の学習が遅かったと考えらえる。そのため WaveGAN の方が生成された音楽の質が良かったと考えられる。よってラベルの特徴量を考慮しネットワークの構成を変更する必要があると考えられる。生成された音楽はどちらも楽器の特徴は現れていた。しかし、元の音楽のようにメロディアスなものではなかった。これはネットワークで音楽データの横の関係を学習できるように改善する必要があると考えられる。ネットワークの構造として訓練データの音楽データが入力されるのは感情分類器と Discriminator である。Discriminator に入力されるラベルは感情分類器から出力されたラベルであるため、感情分類器が誤って分類したラベルを Discriminator に入力すると本来のラベルがついていない訓練データが Discriminator で扱われてしまう。頻繁にこの現象が起きるわけではないため、ある程度の音楽を生成することができている。しかし、学習が上手くいかない原因の一つだと考えられる。

SM-GAN は訓練データに付随しているラベルを利用し学習を効率化するネットワークである。そこで SM-GAN の性能を向上させる手法としてラベルごとに Discriminator を用意し、それぞれを学習させる方法が考えられる。Discriminator をラベル分作成し、感情分類器で分類された感情をもとに Discriminator を選択する。選択された Discriminator で本物か偽物化を判断させ学習を行う。個別の Discriminator を用意し学習をすることで、それぞれが一つのラベルに特化した判断をすることができるようになる。よって、それぞれのラベルに特化した学習を進めることができ、指定した感情の音楽の質が向上すると考えられる。

（文責: 松田祐輔）

5.2　訓練データ

訓練データとラベルとの必然性がない、1 つのラベルに対しての音楽データが多様であることが SM-GAN の学習が上手くいかない原因であると考えられる。例えば動物を訓練データにした場合は、犬の画像に「犬」というようなラベルが付けられ、訓練データとして扱われる。このとき、画像に映っているのが犬であることは、誰が見ても変わらない事実である。しかし、音楽データに感情のラベルを付ける場合、必然性のある決め方はできず、どうしても主観が介入する。また、音楽データは異なるメロディやリズムでも「楽しい」や「悲しい」というようにラベルが付けられる。そのため「楽しい」とラベル付けされた音楽の中でも、アップテンポな音楽もあれば、ゆっくりでワルツのような音楽もあり、音楽的な構造が共通していなかった。これらが原因で学習が上手くいかなかったとも考えられる。

訓練データにはピアノの演奏をマイクで録音したデータセットである maestro を使用する予定であった。しかし、そのデータセットでは WaveGAN と SM-GAN のどちらも学習が上手くいかなかった。このデータセットはピアノの反響が大きく、音が重なっていることが多かった。反響した音はサンプリングした次の状態で処理されるため、本来譜面上にはない音がサンプリングされ

てしまっていたと考えれる。よって反響した音も訓練データとして扱われ、学習に悪影響を及ぼしたと考えられる。このことから音楽データを訓練データとする場合は、リバーブやディレイなどのエフェクトがない音楽を用意する必要がある。今回の音楽データは wav ファイルを使用したが、midi ファイルとして扱ったり、反響や遅延をなくした状態で wav ファイルを作成し音楽データとするべきだと考えられる。

（文責: 松田祐輔）

5.3 評価実験

評価実験 2 での SM-GAN の結果について、評価実験 1 よりも「ピアノ曲の聞きやすさ」「ピアノ曲らしさ」ともに平均値が向上し、統計的にも WaveGAN と比べた母平均の有意な差が見られなくなった。これには以下の 2 つの理由が考えられる。

一つは、Generator の損失がより収束に近づいたことである。4.1 で述べたように、SM-GAN の Generator の損失は、評価実験 1 では WaveGAN よりも振動の幅が大きかったが、評価実験 2 ではほぼ同じ幅になっていた。更新回数が増え、学習状況が追いついたことで、WaveGAN と同程度の性能が得られたと考えられる。

もう一つは、楽曲を提示する順序を変えたことである。両モデルが生成する音楽は「2 秒のピアノ曲」という聞き馴染みのない形式であり、音質もノイズが混じっていたり、音が割れたりしていたため、普段から親しまれている音楽より劣って聞こえやすかったと考えられる。実際 4.2 で述べたように、実験の感想にはそういった内容のものがあった。評価実験 1 では、初めに SM-GAN による 10 曲を、その後に WaveGAN による残りの 10 曲を聞かせた。そのため、初めに連続して現れ、イメージしていた音楽とギャップのある SM-GAN の方を低く評価したと推察される。または、慣れにより WaveGAN への評価が優しくなってしまったとも考えられる。4.2 で挙げた感想でも慣れによる基準の低下を述べたものがあった。一方、評価実験 2 では完全にランダムな順で提示した。さらに、実験の前にサンプルの音声を聞くことが可能であった。そのため、評価実験 1 のように順序による不当な低評価がされにくくなったのだと考えられる。

（文責: 釜石健太郎）

第 6 章　外部評価

6.1　中間発表

前期の成果を発表する場として中間発表会が行われた。新型コロナウイルスの影響により前年度と同じく Zoom を用いたオンラインでの開催となった。中間発表会では、1 時間の各プロジェクトの動画や Web サイトを見て内容を把握する時間が設けられた。その後、質疑応答の時間は、前半 15 分を 3 回、後半 15 分を 3 回で行われた。発表についての評価アンケートを実施し、集計された結果を用いて今後の課題として議論を行った。

（文責: 伊村尚矢）

6.1.1　発表準備

ポスターには必要最低限の情報をのせ、他は動画で内容を補っていた。ポスターと動画だけを見てすべて理解することは難しく、質疑応答がメインであると考え、事前に来る質問を予測し、リストアップした。ポスターのデザインや配色は単調ではあるが、説明内容としては端的に述べられており、理解のしやすい内容にした。動画では関連研究と提案手法の説明をメインとし、例えを入れることで噛み砕いた内容を説明した。

（文責: 伊村尚矢）

6.1.2　発表評価

中間発表では、Google フォームを用いたアンケートを実施した。アンケート内容は、発表技術と発表内容についてである。それぞれ 10 段階での評価と、なぜそのように評価したのかという理由や感想を自由記述してもらった。最終的に 36 人の方々の回答を得ることができた。まず、発表技術についての評価を示す（6.1）。36 人の評価平均値は 7.4 であった。評価理由の記述をいかに一部抜粋する。

- c グループの色が青色に対し、タイトルも青色なので、c グループの活動がメインのプロジェクトのように感じた。
- 内容理解に必要な知識の解説が丁寧で、プロジェクトの理解がしやすかった。ただ、c 班「ソマティック・マーカー仮説に基づいた GAN による音楽生成」では図解が多く、実際の作業プロセスが見えにくかったので、どのような作業・開発・分析を行っているのかのスクリーンショットでもなんでもいいのではしかった。

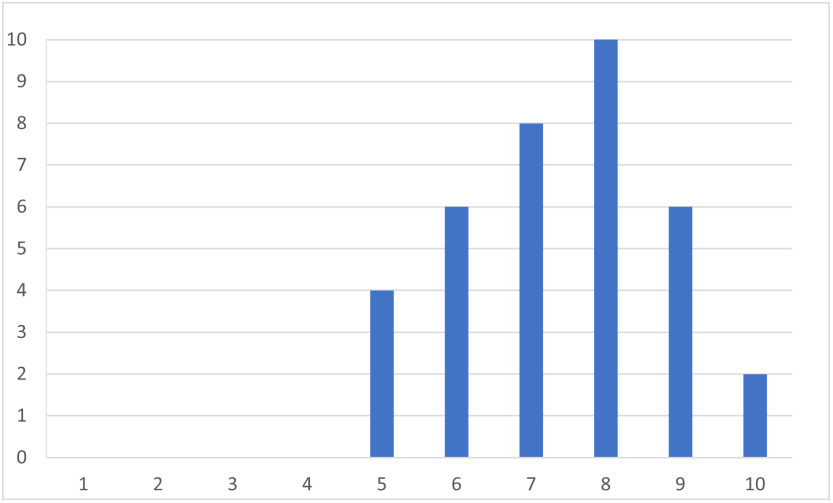

図 6.1　中間発表における発表技術の評価

総評として、見にくさとわかりにくさの 2 点が改善点として指摘された。ポスターや動画で用いたスライドの色には同系色が多く、見にくいものとなってしまった。そして内容の理解には動画の尺では足りず、成果発表会では Web に掲載したほうがいいのではないかという話がグループの中で出た。他にも、「一人の質疑応答の時間が長い」「沈黙が発生して質問しにくい空気ができている」といった点が指摘された。これらの意見、反省をいかして成果発表会への準備をしていくことが課題であると考えられる。次に発表内容についての評価を示す（6.2)。36 人の評価平均値は 7.4 という結果であった。評価理由の記述を以下に一部抜粋する。

- 感情と人工知能を融合し、感情を音にするという考え方はあまり浸透していないので、良いアプローチであると言えます。
- GAN のモデル構築の構想などが良かった。
- 目標や計画がはっきりとしていてよかった。

などの意見があったが一方で、

- 嬉しいや悲しいなどの感情をどのように認識させ、制御するのかがわからないです。
- Hevner model は音楽と情動では良く参照されるモデルのようですね.Hodgkin-Huxley model のようにその分野で確立されているモデルに依拠するのであれば、出典を示して発表した方がしっかり関連研究調査をしていることを示す点でも良いと思います。
- 計画について、いつまでに何をする、のような具体的計画案があると良いと思った。

などの意見もあった。

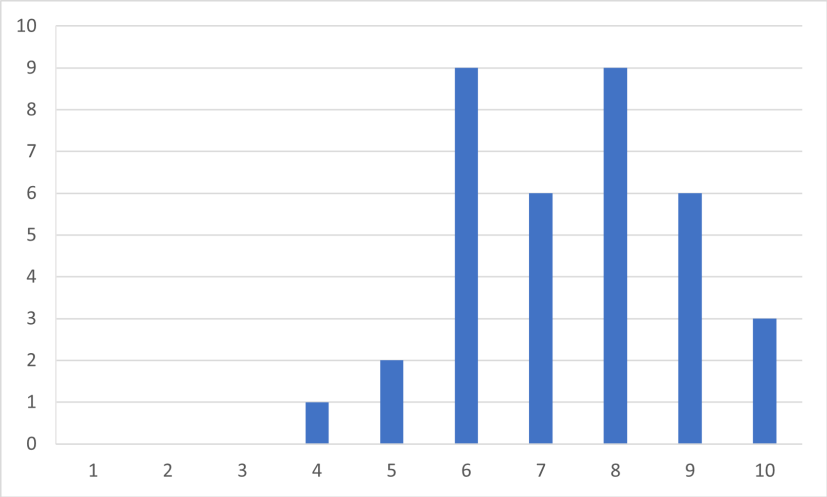

図 6.2　中間発表における発表内容の評価

総評としては、より具体的な内容を伝えることが課題としてあげられた。それぞれの過程や根拠の説明が足りておらず、悩ませてしまうことが多かった点が今後の課題となる。しかし我々のグループが伝えたかったことはポスター、動画、質疑応答でしっかり理解して貰えたという点は、今後の成果発表会にいかしていきたい。

（文責: 伊村尚矢）

6.2　最終発表

今年度の一年間の活動を発表するため、2021 年 12 月 10 日に Zoom にて最終成果発表会が行われた。最終成果発表会では 1 時間の各プロジェクトの動画や Web サイトを見る時間が設けられた。その後、質疑応答の時間が前半 15 分を 3 回、後半 15 分を 3 回で行われた。途中で Zoom 会場が落ちてしまったため、前半 3 回目はとても短い時間で行われた。発表についての評価アンケートを実施し、集計された結果を用いて、今後の課題として議論を行った。

（文責: 伊村尚矢）

6.2.1　発表準備

発表準備としては、ポスター、動画の作成と質問内容を事前に予測しリストアップすることを行った。中間発表では、内容が専門的すぎて理解することが難しいという評価も多かった。そのため、最終成果発表では説明のメインである動画を決められた時間内で噛み砕いた内容を伝えることに焦点を当てた。ポスター内に QR コードを用いることで、見てほしい資料、聞いてほしい音源にすぐに飛ぶことができるようになった。専門的な内容を噛み砕いて説明するために、紹介用 Web サイトを作成するという案があった。しかし他のグループと共通して行わなければいけないため困難と考え廃案となった。

（文責: 伊村尚矢）

6.2.2　発表評価

最終成果発表では、Google フォームを用いたアンケートが実施された。アンケート内容は、発表技術と発表内容についてである。それぞれ 10 段階での評価と、なぜそのように評価したのかという理由や感想を自由記述した。最終的に 37 人の方々の回答を得ることができた。発表技術についての評価を示す（**??**）。37 人の評価平均値は 7.8 であった。中間発表会では評価平均値が 7.4 であり、0.4 ポイント増加したことが分かる。評価理由の記述を以下に一部抜粋する。

- ポスターが少し見にくいと感じた。英文のフォントを少し小さくするなどの対処をした方がいいと思う。生成された音声が少し聞き取りにくかった。音量を上げる・生成音の長さを伸ばすなどの対策が必要かと思った。
- 内容をじっくり早くするためにポスターの存在は意外に重要です。プレゼンはどのグループも分かりやすく、内容を把握することができたのですが、ポスターの配色が対色で配置され、文字扱いが悪く、とても読み難く感じました。「レイアウト 4 原則」という極めてスタンダードなレイアウトの方法が一般化していますので、この法則を援用されることをオススメします。せっかくの良い内容が聴講者にスムーズに伝わらないというのは残念でなりません。
- ポスターには内容が簡潔にまとめられており、詳しい部分は QR コードを使い、別部分で補うという形は良かった。
- C グループに関しては感情処理を取り入れた音楽生成をしていたが、課題として既存の音楽に劣ってしまうという点を改善することができれば実践的な成果物になると思いました。
- 生成された音声が少し聞き取りにくかった。音量を上げる・生成音の長さを伸ばすなどの対策が必要かと思った。

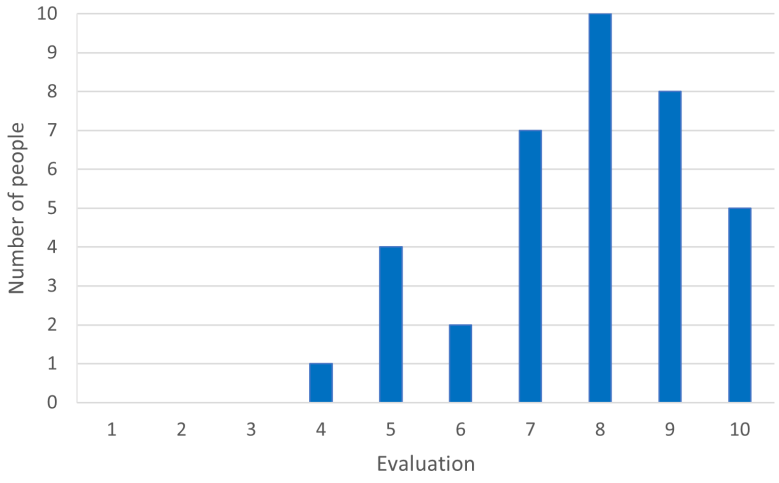

図 6.3　最終発表における発表技術の評価

評価点数は高かったが、評価理由を見ると「聞きにくい」「わかりにくい」という意見もあり、人によって評価は様々であったといえる。ポスターと C グループの生成音が問題視されていた。ポスターは配色による見づらさ、わかりにくさがコメントにあげられ、改善の必要があると考える。しかし QR コードを使い、ポスターには書ききれない内容を補える点は評価されたと考える。中間

発表ではポスターへのコメントが無かったことから改善がなされなかった。これが今回の評価を招いたと考えられるため、プロジェクトメンバー内での評価を細かくするべきである。生成音に関しては、現状の 2 秒を超える生成は難しかったため、このままでは今後聞き手が評価しづらい状況が続くと予想される。

次に発表内容についての評価を示す（6.4）。36 人の評価平均値は 8.2 という結果であった。中間発表会では評価平均点が 7.4 であり、0.8 ポイント増加したことが分かる。評価理由の記述を噛み以下に一部抜粋する。

- 目的で「理解を深める」としていたグループは、最終的に理解を深めることができたのか、なにを理解で、それをどうするのかなどを伝えてほしいです。
- C：感情という不安定で扱いにくいものの生成が出来ているのはすごいと思う。
- 有識者が多くない分野だと思うのでもう少し砕いた説明があると嬉しいです。
- 筋道を立てて成果物に使用する技術選定をしているというのが伝わってきて良かったです。
- 専門用語が少し多いように感じたので詳しい説明が少しほしいと感じました。発表内容は各グループとても面白いものでデータや映像として成果物ができていたので良かったと思います。
- 全体的に研究要素が強いテーマだったので、従来の研究成果に対する新規性を期待してしまいますね。
- C グループのみ背景に課題が無く、どうしてそのテーマを選んだのか不明瞭でした。成果物に関しては目的に対してある程度の解答を出せた良いものだと思います。

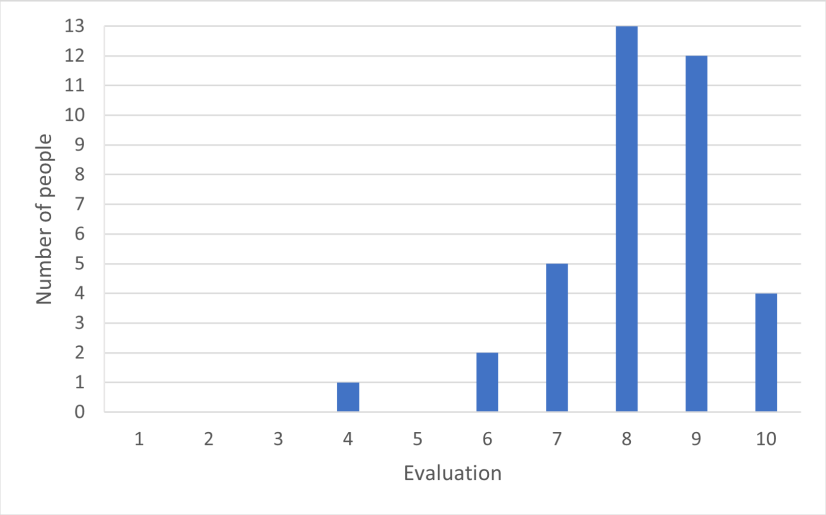

図 6.4　最終発表における発表技術の評価

総評としては、内容の興味深さが主に評価された。短い時間の中で、噛み砕いた内容を伝えるということが成功したため、内容を面白く感じた方が多くいたのだと考える。専門知識の多いグループではあるが、初めて聞く人たちは何を知りたいのか、有識者の人たちは何を聞いてくるのかを事前に予測し、リストアップするという準備がとても役立った。感情という難しい分野に対しての成果を評価してもらった点は、成果発表会での大きな成果であると考える。

（文責: 伊村尚矢）

6.3 グループ内での評価

最終成果発表会や今までの活動の結果を踏まえて、プロジェクトグループ内で自分たちのグループ活動について評価を行った。「目的」「現状の把握」「今後の計画と具体性」「表現力」「協調性」の5つを評価基準とし、それぞれ5段階で評価を行った。結果は以下の通りである。

目的：4

- 脳と人工知能を組み合わせることができた。しかし脳の要素が少なく、ただの GAN との区別が少し弱い。
- 感情を音楽に反映させたいという目標を達成することができた。

現状の把握：4

- 対面時には会話を多くし、それ以外は Slack や LINE を使って現状把握をしていた。しかし一部取り組みの共有ができていないことがあった。
- 自分たちが何をしてきたか、今後何をするかを、定期的に話し合うことで、お互いのスケジュールを把握することができた。
- GitHub を活用することで、オンラインでの開発でも問題なく行うことができた。

今度の計画と具体性：4

- 感情の扱い方と検証用のモデルを作るなどモデルの改善。
- 正確に判断できる定量的評価を選択し、実装する。

表現力：3

- 図や表を使いながら端的にまとめることができた。しかしポスターのデザイン力をすぐに習得するのが困難で、単調な配色・配置となってしまった。
- 専門性が高く、短い時間で内容を伝えるのが難しく、動画だけでは伝えきれない情報を Web サイトを作って、補うべきであった。

協調性：4

- 意思疎通はかなりできていた。しかし役割分担が微妙な時があった。
- 誰かが困ったときは、どこに困っているのかを聞き全員で話し合って解決するときがあった。

（文責: 高野凌太）

第 7 章　まとめ

7.1　成果

本プロジェクトは既存の人工知能に脳の仕組みを取り入れるという目的であった。そこで脳の感情処理に着目し、ソマティックマーカー仮説を取り込むことができれば従来よりも効率の良い情報処理ができることを期待した。そして、感情を指定して曲を生成するような新しい音楽生成手法の確立を目指した。

ソマティックマーカー仮説を取り込んだ新しい人工知能を SM-GAN と呼び、既存の技術である WaveGAN と比較することによって有用性を明らかにしようとした。効率の良い情報処理ができているかそれぞれの学習曲線を比較した。学習曲線から考察すると、学習するまでに SM-GAN のほうが WaveGAN よりも時間がかかることから効率の良い情報処理ができていなかった。WaveGAN と SM-GAN の生成するピアノ曲について「ピアノ曲らしさ」、「ピアノ曲の聞きやすさ」を 5 段階評価で答えてもらうアンケートを行った。いずれも平均値の差に有意さがあるとは言えなかった。SM-GAN が感情をもつピアノ曲を生成できているかカイ二乗検定を行い、有意に感情を持つピアノ曲を生成できていることが明らかとなった。また、学習曲線が SM-GAN と WaveGAN では最終的に振動の幅がほぼ同じになった。よって、SM-GAN は感情処理を取り入れながら WaveGAN と同等のピアノ曲を生成することができると言える。　最終発表会では Google フォームでのアンケートで得られた評価平均値は 0.78 と高めであった。専門家からのコメントで「従来の機械学習のモデルに自分たちの新たな理論を組み込めていてプロジェクトらしい」と好意的なコメントを頂いた。私たちの目的が達成され成果として十分に現れ発表することができたと言える。　学習を効率化させ、少ない時間でモデルの学習をするという点でソマティックマーカー仮説を取り入れることはできなかったが、WaveGAN にはない機能を SM-GAN で実現することに成功した。SM-GAN は感情をラベルとして学習したことにより、感情を指定してピアノ曲を生成することができた。ラベルを指定して生成を行う Conditional GAN があったが同じ学習データを用いても実装ができなかったため、感情を指定してピアノ曲を生成できるという点で新規性が見られる結果となった。この結果から感情が情報処理を効率化させたという推測ができてソマティックマーカー仮説が成り立っていると言える。したがって、ソマティックマーカー仮説を部分的に取り入れて、新たな音楽生成手法の確立をすることができた。

（文責: 高野凌太）

7.2 今後の課題

7.2.1 提案手法について

今回の SM-GAN は WaveGAN をベースとして実装した。しかし生成された音楽がメロディアスではない、音楽の生成可能な時間が短いなどという理由からネットワークを変更する必要がある。SM-GAN は既存の音楽生成ネットワークに感情分類を行う CNN を組み込んだものであるため、ベースとなる音楽生成ネットワークは WaveGAN である必要なない。先行研究では WaveGAN のほかにも SpecGAN[番号] や mp3net[番号] など様々な音楽生成を目的とする GAN が存在するため、これらのネットワークを用いて SM-GAN を実装するべきだと考える。また訓練データに使用する音楽データを wav ファイルではなく midi ファイルとして扱うネットワークも考えられる。midi ファイルは録音された音楽データとは違い外部からのノイズがなく、ピアノロールのように表示した midi ファイルや譜面に起こした画像などを扱うことができる。そのため、今回の SM-GAN とは違う音楽生成をすることができるため、違う結果を見込める。音楽データに感情ラベルを付ける際 Hevner の感情モデルを使用したが、音楽の捉え方は人それぞれなため、完全に客観的な指標とは言えなかった。よって音楽データに感情ラベルを付けるための完全な客観的な指標を作るなど、ラベルに対して音楽データが一様になるような工夫を考えるべきである。

Discriminator で扱われる訓練データが感情分類器の誤った分類により、本来存在しない訓練データを扱う可能性があると考えられた。この対策として、元の訓練データに不随しているラベルと感情分類器から出力されたラベルが一致しない場合の処理が必要となる。ラベルが一致しなければ一致しない訓練データでは学習しないなどといった処理が考えられる。また誤って分類されることが頻繁に起こってしまう場合、学習が遅くなってしまう。よって感情分類器の精度向上、または Discriminator への入力する仕組みを再考するべきだと考えられる。

今回の目的とは離れてしまうが、SM-GAN の性能を調べるため音楽データに付けるラベルの種類を変更した学習も試すべきだと考える。例えば、「3 拍子」や「ハ長調」など音楽の主観性を排した情報をラベルとして扱うことが挙げられる。これにより、現状の SM-GAN のラベリングが評価や性能を悪化させているかを調べることができる。

SM-GAN の性能を向上させる手法として複数の Discriminator を用意し、それぞれを学習させる方法を提案した。しかしこの手法は複数の Discriminator とそれぞれの重みを記録するためのメモリ領域が必要となる。また、学習データをさらに大きくする、ネットワークをさらに深くする際に、プロジェクトで使用したマシンよりも性能の高いマシンを使用する必要がある。ディープラーニングには GPU の tensor コアを使用して訓練データの畳み込みを行う。よって、tensor コアが多い GPU やメモリを増加したマシンを使って学習する必要があると考えられる。

（文責: 松田祐輔）

7.2.2　ソマティック・マーカー仮説について

現状のモデルは、ソマティック・マーカー仮説を構成する要素を十分に再現しきれていないという問題がある。第一に、ソマティック・マーカー仮説は身体的反応の役割を重視しており、現状のモデルでは身体や身体的反応に相当する部分がない。次に、扁桃体や VMPFC、体性感覚皮質といった関連する脳の部位に相当する部分がない。「意思決定にバイアスを与える感情の種類」という結果の部分のみを再現し、ソマティック・マーカーの生成という本仮説の最大の特徴ともいえる過程を再現できていない。さらに、感情を分類している点も仮説の考え方とは合っていないと推察する。文献 [2] での Damasio の説明では、ソマティック・マーカーは選択肢に対する予測がポジティブな結果をもたらすか、ネガティブな結果をもたらすかをマークするものだとされている。予測と並置されるマーカーがポジティブであれば動因、ネガティブであれば警報となるという。この考えに基づけば、現在採用している Hevner により分けられた感情のグループも身体的反応から与えるソマティック・マーカーのポジティブさ・ネガティブさで分けるべきではないだろうか。さらに述べると、グルーピングをやめ、楽曲に対し、ポジティブ・ネガティブに相当する選好のような値を計算する機構を開発し、Discriminator の損失の計算式に組み込む方が仮説に忠実であると考えられる。

これらの問題が残ってしまったのは、ソマティック・マーカー仮説や脳の感情処理についての学習、文献調査が足りなかったためであると考えられる。今後は、仮説や脳への理解を深め、タイトルと合うようにモデルの修正を行うのが望ましいと考える。

（文責: 釜石健太郎）

7.2.3　評価手法について

評価手法については、以下の三点を改善する必要があると考える。

一つ目は、評価実験 1 における曲の順序である。考察で述べた通り、モデルごとに曲を連続させたことが SM-GAN の結果を悪化させた、もしくは WaveGAN の結果を向上させた可能性を否めない。そのため、曲をランダムにする変更を加え、評価実験 1 と同様の実験をもう一度行うことが望ましいと考える。

二つ目は、評価実験で使用する曲の選び方である。私たち実験者が主観に基づいて良い楽曲を選び出しているのは、多数決を取っているとはいえ、実験結果の信頼性を落としてしまう。そのため、完全に無作為に標本を選び出したり、この後に述べる Inception Score のような定量的な評価手法によって高く評価された楽曲を用いるのが望ましいと考える。さらに、全被験者が同じ曲順で同じ曲を聞くのではなく、被験者によって選ばれる曲も曲順も異なるようにすることで、信頼性の向上が期待できると推察する。

三つ目は、定性的な評価に頼っている点である。SM-GAN を定量的に評価する方法として、WaveGAN の元論文でも使われた Inception Score の採用が考えられる。Inception Score は学習済みの分類器である Inception classifier を用いて計算される、GAN によって生成された画像の多様さや区別のしやすさを評価するための指標である。私たちのモデルは画像ではなく音楽を生成するが、音楽をメルスペクトログラムに変換することで、Inception Score の算出が可能になると考えられる。また、Inception Score を改良した Frechet Inception Distance 等の使用も検討される。

（文責: 釜石健太郎）

7.2.4　発表について

本プロジェクトは必要な事前知識が多い。発表を見る人たちは必ずしも事前知識を持っておらず、また事前知識との関連なども理解しなければ何を行っているのかを理解するのは困難であった。そこでポスターや発表動画とは別に、ウェブサイトや書類などで必要な事前知識や関連研究など詳細な資料を用意し配布するべきであると考える。またより理解が深まるような工夫や発表の仕方を変更する必要がある。案として、動画での発表ではなくウェブサイトを作成する、発表を行う際のリハーサルを行うなどが挙げられる。

ポスターが見づらいなどの意見もあったため、発表内容が伝わりやすくなるようにポスターのデザインを考えるべきである。そこでポスターのレイアウトの原則やテンプレートなどを調べ、学ぶ必要がある。また、発表に使用するモデル図なども詳細にかつ見やすいものを作成する必要がある。

発表中に、質問がなく沈黙の時間が続くことがあったため、プロジェクトでの成果物を動画で流したり、プログラムの詳細などを説明するなどして、沈黙の時間を減らすべきであると考える。また質疑応答中で一つ一つが長くなることが多かったため、発表で具体的に内容を伝えることが必要である。これにより、質問にて説明する内容が少なくなるため多くの質問に対応できる。

（文責: 松田祐輔）

参考文献

[1] C. Donahue, J. McAuley, M. Puckette. Adversarial Audio Synthesis. ICLR, 2019.

[2] Damasio, A. R. *Descartes' error:Emotion, reason, and the human brain.* New York:Putnam. 田中三彦 (訳), 2000. アントニオ・ダマシオ 『生存する脳 —— 心と脳と身体の神秘』 講談社.

[3] 大平 英樹. 感情的意思決定を支える脳と身体の機能的関連, 心理学評論, 57(1), 2014.

[4] Hevner.K. Experimental Studies of the Elements of Expression in Music. American Jornal of Psychology, 48, 246-268, 1936.

[5] A. Radford, L. Metz, S. Chintala. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. ICLR, 2016.

[6] M. Mirza and S. Osindero. Conditional Generative Adversarial Nets. arXiv:1411.1784, 2014.