修士論文

敵対的生成ネットワークにより生成された旋律画像の補正手法

2021 年度

鈴木 大河

明星大学大学院
情報学研究科 情報学専攻
20MJ-005

概要

画像を利用した自動作曲の分野では，音楽の生成に特化した敵対的生成ネットワークを用いた手法が提案されている．この手法により生成された画像は，生成された旋律において前半部と後半部の音階が異なる状態で出力される可能性がある．この問題は，先行研究により GAN は音楽の時間的方向に対する規則を認識できないことが示されている．そこで本研究では，汎用敵対的生成ネットワークを用いて生成された旋律画像に対し，生成された画像の特徴を損なうことなく音階にあった音高に補正する手法を提案する．この手法は，画像が持つ濃度情報と学習データから生成した音高遷移確率の 2 種類の指標を利用し，生成した画像の線形をなぞる旋律を生成する．単純な GAN モデルを利用し生成された画像に対して本手法による補正を行い，生成した旋律 6 本に対し 5 段階の主観評価を行ったところ，最低評価となった旋律が平均 2.50 点だったのに対し，最高評価となった旋律は平均 4.75 点となった．また，生成旋律に対し音階とコード進行の観点から評価を行ったところ，提案手法により生成された旋律の中にはコード進行の観点と楽曲構成の観点から，表現力のある旋律が生成されたことが確認された．

目次

1 序論 1

1.1 背景 1

1.2 目的 2

1.3 本論文の構成 2

1.4 本研究の意義 2

2 音楽理論と和声法 4

2.1 音程 4

2.2 音名 5

2.3 音階 6

2.4 和声法 9

3 関連研究 16

3.1 和声を軸にした旋律の自動生成 16

3.2 旋律の自動生成 16

3.3 旋律概形 19

3.4 コード進行の自動付与 21

4 提案手法 23

4.1 学習対象となる旋律の画像化 24

4.2 旋律画像の生成 25

4.3 生成された画像の補正 26

4.4 コード進行の付与 30

5 プロトタイプシステムの実装 32

5.1 学習楽曲の加工 32

5.2 GAN による画像の生成 34

5.3 生成された画像の補正 35

6 評価実験 39

7 考察 47
7.1 全体の傾向 47
7.2 旋律 1 48
7.3 旋律 2 49
7.4 旋律 3 52
7.5 旋律 4 53
7.6 旋律 5 54
7.7 旋律 6 55

8 結論 57
8.1 結論 57
8.2 今後の課題 57

表目次

1	2 音間の音程名	5
2	音名の種類 (中央ドと 440Hz を含むオクターブ)	6
3	半音操作を含んだオクターブ内 12 種の音名	7
4	短音階	8
5	主なコードネームとその表記	15
6	主な旋律生成手法の比較	24
7	主なコード進行付与手法の比較	31
8	実験被験者詳細	40
9	主観評価結果	43
10	評価対象の旋律に出現した表現力	48

図目次

1 鍵盤楽器における全音と半音の関係 4
2 ハ長調 7
3 ハ短調 9
4 ハ短調 (旋律的短音階) 9
5 ハ短調 (和声的短音階) 9
6 和声法で許容される和音進行 10
7 ドを根音とした長三和音 (C) 12
8 ドを根音とした短三和音 (Cm) 12
9 ドを根音とした属七の和音 (C7) 12
10 ドを根音とした長七の和音 (CM7) 13
11 ハ長調の和音記号 13
12 ハ長調のダイアトニックコード (5 和音である G9 を除く) 14
13 コードネーム「C」と「C/E」の違い 15
14 GAN の構成 17
15 Cheng らが提案したモデル構造 [5] 19
16 MuseGAN のモデル構造 [17] 19
17 MuseGAN の出力 [17] 20
18 MuseGAN の生成過程 [17] 20
19 旋律概形の例 [18] 21
20 遺伝的アルゴリズムによるコード進行付与の個体表現の例 [25] . . . 22
21 形状モデルへの和音の登録例 [25] 22
22 提案手法の概要 24
23 補正イメージ 25
24 画像化で表現可能な旋律特徴の例 26
25 MidiNet の 3 種類のモデルにより生成された旋律 [16] 27
26 同じ生成画像に対し，4 パターンの二値化を行った例 28
27 画像の補正方向 29
28 グレースケールを利用した音高推定値のイメージ 30

29 補正値の算出イメージ 31
30 楽曲加工・旋律画像化の手順 33
31 旋律画像の形式 34
32 生成されたテーブルの一部 37
33 二値化による音高推定値 G_{eval} 38
34 生成された旋律画像の例 39
35 評価対象旋律 1 40
36 評価対象旋律 2 41
37 評価対象旋律 3 41
38 評価対象旋律 4 41
39 評価対象旋律 5 42
40 評価対象旋律 6 42
41 評価対象旋律 1 の主観評価 43
42 評価対象旋律 2 の主観評価 44
43 評価対象旋律 3 の主観評価 44
44 評価対象旋律 4 の主観評価 45
45 評価対象旋律 5 の主観評価 45
46 評価対象旋律 6 の主観評価 46
47 補正前後の旋律画像-旋律 1 49
48 補正前後の旋律画像-旋律 2 49
49 補正前後の旋律画像-旋律 3 49
50 補正前後の旋律画像-旋律 4 49
51 補正前後の旋律画像-旋律 5 50
52 補正前後の旋律画像-旋律 6 50
53 旋律 1 の音階の変化 50
54 変ホ短調への転調部 51
55 変ホ短調への転調部-補正前画像との比較 51
56 旋律 2 のコード進行 52
57 旋律 2 の低評価区間 52
58 旋律 2 の高評価区間 53
59 音高遷移確率による補正が有効的に行われている例 54

60 コード進行の表現力が高いと考えられる区間 55
61 旋律 6 のパターン分類 56
62 パターン A とその再現部であるパターン A’の形状比較 56
63 pix2pix による入力画像 (Input) と対応する生データ (Ground truth) と生成画像 (Output)[4] 60

1 序論

1.1 背景

クラシック音楽に代表される西洋音楽における音楽の要素として，律動（リズム），旋律（メロディー），和声（ハーモニー）の 3 要素が挙げられる．自動作曲は過去にさまざまな手法が検討されてきたが，多くの手法では，和声要素にあたるコード進行やベースラインを生成し，和声構成音を軸に旋律を生成する作曲手法が取られている．自動作曲のアルゴリズムとして音高推移を確率モデルで処理する手法や，時系列処理が可能である再帰型ニューラルネットワーク (Recurrent Neural Network : RNN) や，長・短期記憶を利用した再帰型ニューラルネットワーク (Long Short-Term Memory : LSTM) などが旋律の生成アルゴリズムとして多く使われている．これらの手法では，旋律に使われる音が和声構成音と一部の非和声音に限定されやすい．また，自動作曲の分野では機械的に作曲されたような印象が強い出力が多い．このことから，主観評価で人力作曲手法と肩を並べることのできる決定的なアルゴリズムは現状では存在しない．

自動生成される旋律はあたりはずれが激しく，単純な乱数で音を並べた状態に等しい旋律もあるが，主観評価で高い評価をえる表現力の高い旋律が出力されることもある．近年，画像の生成分野では敵対的生成ネットワーク [1] が提案され，畳み込みニューラルネットワークを構造に取り込むことで，学習画像に近い画像を大量に生成できる手法として注目されている [2]．この手法により特定の画像や動画 [3]，ある関係性を持つ対の画像 [4] など，多くの画像生成が可能になった．音楽は時間と音高の 2 次元展開することで画像化できる．このことから，畳み込みニューラルネットワークを用いた敵対的生成ネットワークにより音楽の生成が可能である．GAN を用いた旋律生成システムを用いることで，ベースラインやコード進行などの特定の入力なくノイズから旋律を生成することが可能となる．

一方で，畳み込みニューラルネットワーク (Convolutional Neural Network : CNN) が一方向に流れる時間構造を認識できない [5] ことから，GAN により生成された旋律は，出力の前後で大きく音階が狂うことが確認されている．この手法を用いた旋律生成を行うには，生成時に音に制限を与える手法や生成後に補正を行うな

ど音階の調整が必要である．

1.2 目的

本研究では，画像ベースの旋律生成手法として，敵対的生成ネットワークにより生成された旋律画像に対し補正を行う手法を提案する．敵対的生成ネットワークによる旋律生成は音階が崩れることがあるため，出力画像に対し画像のグレースケールの値および旋律の音高遷移確率の 2 点に着目し，出力画像の特徴を損なうことなく音階に合った旋律に補正する．また，近年の楽曲の旋律は，音楽理論的に正しくないようなものも含むことがあるため，本研究では音楽理論による許容と禁止を徹底した生成手法は取らない．また，本研究では，表現力として次の点を重視する．

- 旋律内で一定の音階である．またはある地点で転調する．
- 同じパターンが 3 回以上連続しない[*1]．
- 変化に富んだコード進行が付与可能である[*2]．

1.3 本論文の構成

本論文の以降の構成は次の通りである．第 2 章では，本研究における議論を進めるにあたり必要な音楽理論と和声法について述べる．第 3 章では，先行研究と関連研究について，第 4 章では提案手法の詳細，第 5 章では提案手法の具体的な実装方法，第 6 章では提案手法の有効性を検証するための実験内容と実験結果，第 7 章は実験結果に対する考察を述べる．そして第 8 章では，導出される結論と提案手法の課題について述べる．

1.4 本研究の意義

敵対的生成ネットワークによる旋律の生成手法では，音階が崩れる点が課題となっていた．提案手法では，敵対的生成ネットワークにより生成された旋律画像に対し，補正を行うことで音階の崩れを抑制することができる．生成された画像に対する補

[*1] 主題に対する模倣と変形で構成されるポリフォニー音楽への考慮.
[*2] 付与対象コードがダイアトニックコード (第 2.4.5 項) に限定されない状態を指す.

正を行うため，さまざまな形式により生成された旋律画像に対応することが可能である．また，旋律画像以外の入力から旋律の生成も可能であることから，画像と音楽を組み合わせた芸術分野への応用が考えられる．

2　音楽理論と和声法

2.1　音程

音程とは，ある 2 音間のピッチ差を指す．西洋音楽ならびに西洋音楽をベースとする現代の多くの音楽では，周波数比が 1 : 2 である 2 音間 (オクターブ) を 12 等分した 12 平均律を利用する．

音程の単位には度を用いる．2 音間の周波数比が 1 : 1 となる場合に 2 音間の音程は完全 1 度に，1 : 2 となる場合に 2 音間の音程は完全 8 度となる．完全 8 度の音をオクターブと呼ぶ．西洋音楽における音程の最小単位は半音[*3]と呼び，半音の関係にある 2 音間の周波数の比率は 1 : $\sqrt[12]{2}$ となる．また，半音 2 つ分の音程を全音と呼ぶ．図 1 に鍵盤楽器における全音と半音の関係を，表 1 に 2 音間の半音の数で整理した主要な音程名を示す．

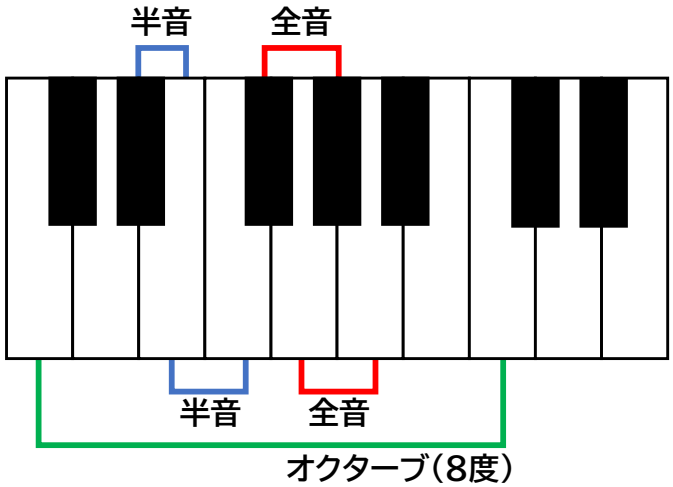


図 1　鍵盤楽器における全音と半音の関係

[*3] ピアノなどの鍵盤楽器では，白鍵と黒鍵を含むすべての隣り合った鍵盤の音程が半音となる．

表 1　2 音間の音程名

半音の数	音程名
0 個	完全 1 度
1 個	増 1 度，短 2 度
2 個	長 2 度
3 個	短 3 度
4 個	長 3 度
5 個	完全 4 度
6 個	増 4 度，減 5 度
7 個	完全 5 度
8 個	短 6 度
9 個	長 6 度
10 個	短 7 度
11 個	長 7 度
12 個	完全 8 度，オクターブ

2.2　音名

各音には，音の高さに応じた名前である音名が付与されている．音名の種類には，日本語式・英語式・イタリア語式などが存在する．440.0Hz をラの音とした場合[*4]の種類別の音名を表 2 に示す[*5]．表中の音は，周波数 261.6Hz のドの音から全音・全音・半音・全音・全音・全音・半音の順で並べたものである．

各音から半音の上下を表現する場合，日本語式では下がる場合に変を，上がる場合には嬰をその音名の前につける．英語式とイタリア語式では，下がる場合は♭を，上がる場合には♯をその音名の後ろにつける．上下に操作された音名を含めたものを表 3 に示す．

[*4] 440.0Hz が国際標準であるが，オーケストラや一部のメーカーなど 442.0Hz で調整する場合もある．

[*5] 表中の音は，鍵盤楽器における白鍵を並べたものである．また，第 2.3 項で後述するハ長調と呼ばれる音階の構成音である．

表 2　音名の種類 (中央ドと 440Hz を含むオクターブ)

周波数 (Hz)	前音との関係	ドからの度数	日本語式	英語式	イタリア語式
261.6	–	1 度	ハ	C	Do(ド)
293.7	全音	2 度	ニ	D	Re(レ)
329.6	全音	3 度	ホ	E	Mi(ミ)
349.2	半音	4 度	ヘ	F	Fa(ファ)
392.0	全音	5 度	ト	G	Sol(ソ)
440.0	全音	6 度	イ	A	La(ラ)
493.9	全音	7 度	ロ	B	Si(シ)
523.3	半音	8 度	ハ	C	Do(ド)

本研究では，音単体を指す場合に固定ド[*7]の「ドレミ」表記を用いる．

2.3　音階

ある規則により並べられた音高の集合を音階と呼び，西洋音楽では長音階 (長調) と短音階 (短調) などが存在する．音階は 1 オクターブを単位とした繰り返しの構造を持つ．音階の基準となる音を主音と呼ぶ．音階名は，主音となる音と音階の種類の組み合わせで構成される．本研究では音階名を日本語式で表現し，主音がドの音である長音階の場合の音階名はドの音を日本語式に直した「ハ」に長音階である事を示す「長調」を加えた「ハ長調」と呼ぶ．本研究では，調性を含め広義の意味での音階としてあつかう．

音階を構成する音の中で，音階の主音となる音をトニック，5 度上にある 5 音目をドミナント，ドミナントの下側に隣り合った 4 音目をサブドミナントと呼び，特別な役割を持つ．

[*7] 本来移動ドの「ドレミ」は音階中の音に対応する音名であり，絶対的な音名ではない (移動ド) が，一般的には絶対的な音名として扱われることが多い．

表 3　半音操作を含んだオクターブ内 12 種の音名

日本語式	英語式	イタリア語式	MIDI ノートナンバー[*6]
ハ	C	ド	0 + 12 * 音の高さ
嬰ハ，変ニ	C ♯，D ♭	ド♯，レ♭	1 + 12 * 音の高さ
ニ	D	レ	2 + 12 * 音の高さ
嬰ニ，変ホ	D ♯，E ♭	レ♯，ミ♭	3 + 12 * 音の高さ
ホ	E	ミ	4 + 12 * 音の高さ
ヘ	F	ファ	5 + 12 * 音の高さ
嬰ヘ，変ト	F ♯，G ♭	ファ♯，ソ♭	6 + 12 * 音の高さ
ト	G	ソ	7 + 12 * 音の高さ
嬰ト，変イ	G ♯，A ♭	ソ♯，ラ♭	8 + 12 * 音の高さ
イ	A	ラ	9 + 12 * 音の高さ
嬰イ，変ロ	A ♯，B ♭	ラ♯，シ♭	10 + 12 * 音の高さ
ロ	B	シ	11 + 12 * 音の高さ
ハ	C	ド	12 + 12 * 音の高さ

図 2　ハ長調

2.3.1　長音階 (メジャースケール)

1 オクターブ内の音を 7 分割し，基準音から全音・全音・半音・全音・全音・全音・半音の順番で並べたものを，長音階と呼ぶ．図 2 にドの音を主音とした長音階であるハ長調の音を示す．

表 4　短音階

	前音との関係		
	自然短音階	和声的短音階	旋律的短音階
1 度	–	–	–
2 度	全音	全音	全音
3 度	半音	半音	半音
4 度	全音	全音	全音
5 度	全音	全音	全音
6 度	全音	全音	**半音**
7 度	全音	**半音**	**全音**
8 度	半音	全音	全音

2.3.2　短音階 (マイナースケール)

1 オクターブ内の音を 7 分割し，基準音から全音・半音・全音・全音・半音・全音・全音の順番で並べたものを，短音階と呼ぶ．図 3 にドの音を主音とした短音階であるハ短調の音を示す．

短音階には，旋律に用いられる通常の短音階である自然短音階と旋律的短音階，和声に用いられる和声的短音階がある．旋律的短音階では 7 度の音と 6 度の音を半音上げ，和声的短音階では 7 度の音を半音上げる．旋律的短音階は基準音から全音・半音・全音・全音・半音・全音・全音の順番で並べ，和声的短音階は基準音から全音・半音・全音・全音・半音・全音の順番で並べる．表 4 に短音階の音程を示す．太字の音は，自然短音階と異なる音を示す．図 3 にドの音を主音としたハ短調の旋律的短音階の音を，図 5 にドの音を主音としたハ短調の和声的短音階の音を示す．

2.3.3　その他の音階

西洋音楽以外の音楽では，7 音で構成されない音階が用いられる場合が多い．代表例として，日本固有の音階であるヨナ抜き音階や，琉球音階などの 5 音音階が挙げられる．西洋音楽以外の民族音楽では，西洋音楽の音階から音が抜かれた音階が数多く使われる．本研究では，すべての音階を西洋音楽における長音階か短音階のど

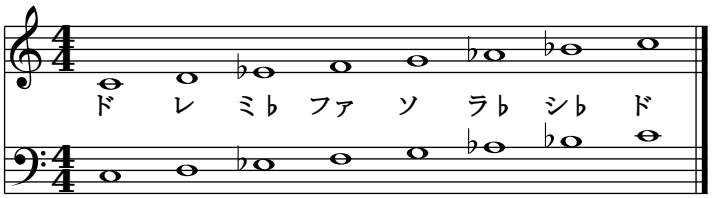


図 3　ハ短調

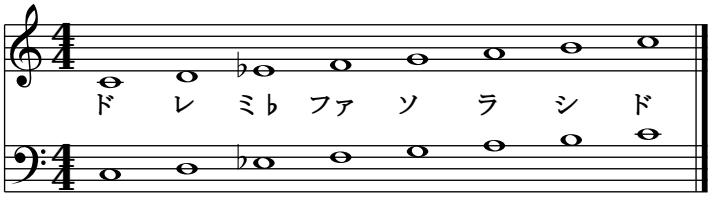


図 4　ハ短調 (旋律的短音階)

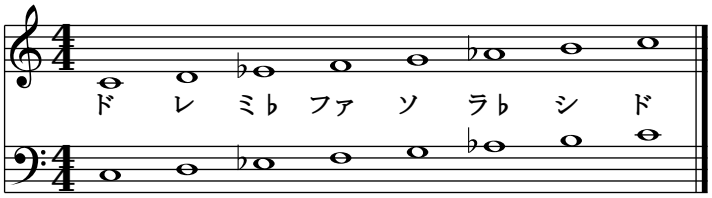


図 5　ハ短調 (和声的短音階)

ちらかに属するとみなし，長音階と短音階以外の音階は考慮をしない．

2.4　和声法

和声法 [6] とは，17 世紀から 19 世紀に発達したクラシック音楽を始めとした西洋音楽の音楽理論で，律動と旋律と並ぶ重要な要素として挙げられる和声の進行と配置を，許容と禁止で制限する理論である．和声法の例として，和声法で許容される和音進行を図 6 に示す[*8]．なお，本研究では生成される旋律に対し付与されたコード進行に対する可否は考慮しない．本研究で考慮する和声法の一部を以下に示す．

[*8] 図中の I から VI の数字はダイアトニックコード (第 2.4.5 項) を指し，T，D，SD は和音の役割 (第 2.4.2 項) を示す．

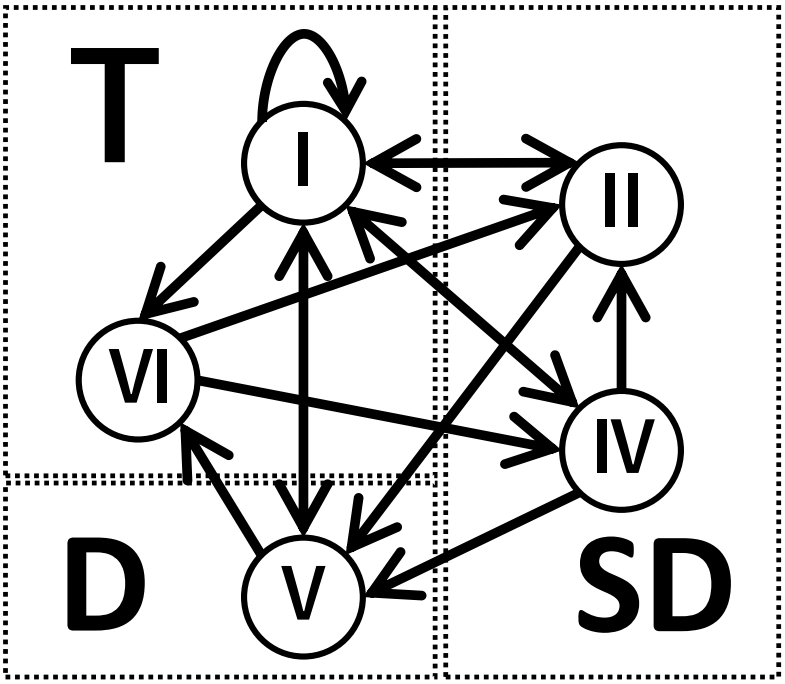


図 6　和声法で許容される和音進行

2.4.1　和声音と非和声音

旋律を構成する音には，和声音と非和声音がある．和声音は旋律に対するその瞬間の和音内に構成される音で，非和声音はそれ以外の音を指す．ある和声に対し非和声音が含まれると不協和音となり，響きの印象が悪くなることがある．旋律でも同様で，非和声音を旋律に構成する場合には一定の規則を取り入れた旋律構成が必要で，これにより非和声音を含んだ旋律から違和感をなくすことが可能となる．規則外の非和声音を多く含んだ旋律は，不協和音となると同時に旋律全体を通して一定の音階として認識することが困難になり，美しい旋律と呼べなくなる．旋律を生成するにあたり考慮すべき主な非和声音の規則は次の 6 種類である．

- 刺繍音
  和声構成音から非和声音を挟み，同じ音に戻る旋律内の非和声音を指す．
- 経過音

和声構成音から非和声音を挟み，同じ方向の異なる和声構成音へ到達するときに経過する旋律内の非和声音を指す．

- 倚音
  直後に隣の和声構成音に移動する旋律内の非和声音を指す．
- 逸音
  直前の和声構成音から非和声音を挟み, 次の和声に含まれる和声構成音に進む旋律内の非和声音を指す．
- 掛留音
  直前の和声構成音が，旋律内で次の和声まで保持された場合の非和声音を指す．
- 先取音
  次の和声に変化する直前に，次の和声から和声構成音が前倒しで旋律に含まれる場合の非和声音を指す．

2.4.2 和音

和音とは，高さの異なる音を同時に響かせた状態を指し，3 音から構成される和音を 3 和音，4 音から構成される和音を 4 和音と呼ぶ．また，旋律中の音の内，和音を構成する音を和声構成音，和音を構成しない音を非和声音と呼ぶ．各和音は役割を持ち，トニック・コード (以下トニック (T) とする)，サブドミナント・コード (以下サブドミナント (SD) とする)，ドミナント・コード (以下ドミナント (D) とする) の 3 種類に分類できる．また，和声法では，和音の遷移を役割により制限する．トニックは音階の主音に関連する和音であるため，安定的な響きを持ち，曲の頭と終端で多くの場合に用いられる．また，すべての和音へ遷移することができる．ドミナントは緊張状態にある和音で，トニックへの解決が必要である．また，すべての和音から遷移することができる．サブドミナントはトニックとドミナントの間の役割を持ち，トニックへ解決すべき場合とドミナントへの遷移が必要となる場合がある．クラシック音楽などの古典的な西洋音楽に適応される和声法では，3 和音を基準とした作曲手法が取られる．一方，ポピュラー音楽では 4 和音を基準とした作曲手法が取られることが多い．

和音の最低音を根音と呼び，和音には，根音と根音からの音程に応じた音の配置か

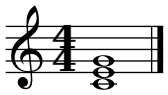


図 7　ドを根音とした長三和音 (C)

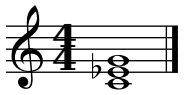


図 8　ドを根音とした短三和音 (Cm)

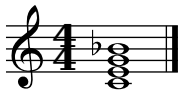


図 9　ドを根音とした属七の和音 (C7)

らの和音名が付けられている．長音階に対応する和音をメジャーコード，短音階に対応する和音をマイナーコードと呼ぶ．以下に主な和音を示す．

- 長三和音
  根音と根音から長 3 度の音と完全 5 度の音からなる和音を長三和音と呼ぶ．長調の場合に基本となるメジャーコードであり，最もシンプルな形をした和音である．ドを根音とした長三和音を五線譜上に置いたものを図 7 に示す．
- 短三和音
  根音と根音から短 3 度の音と完全 5 度の音からなる和音を短三和音と呼ぶ．短調の場合に基本となるマイナーコードであり，最もシンプルな形をした和音である．ドを根音とした短三和音を五線譜上に置いたものを図 8 に示す．
- 属七の和音
  長三和音に短 7 度の音を追加した和音を属七の和音と呼ぶ．ドを根音とした属七の和音を五線譜上に置いたものを図 9 に示す．
- 長七の和音
  長三和音に長 7 度の音を追加した和音を長七の和音と呼ぶ．ドを根音とした長七の和音を五線譜上に置いたものを図 10 に示す．

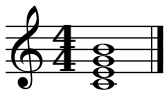


図 10　ドを根音とした長七の和音 (CM7)

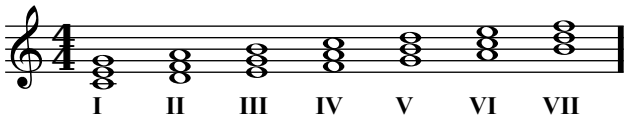


図 11　ハ長調の和音記号

和声法では，旋律を含むソプラノ・アルト・テノール・バスの下 3 声を和音としてるため，本研究では和音の根音をバスとみなし，ベースと表現する．

2.4.3　和音記号

和音を簡単に表現するために，その和音の根音と音階の主音の音程の度数をローマ数字で表した和音記号を利用することが多い．ハ長調における基本的な 3 和音[*9]の和音記号を図 11 に示す．

2.4.4　カデンツ

カデンツとは和声の遷移可否を機能別に整理したもので，役割を自然な流れで遷移させたコード進行であり，すべてのカデンツは和声法で許容される進行に当てはまる．楽曲中のコード進行は，カデンツ任意な順序で連結させることで成り立つ．カデンツは次の 3 種類に分類される．

カデンツ第 1 型　T→D→T 比較的強い進行
カデンツ第 2 型　T→SD→D→T 非常に強く，安定的な進行
カデンツ第 3 型　T→SD→T 比較的弱い進行

[*9] 音階内に存在する音のみで構成された和音で，3 和音を利用してダイアトニックコードを構成する．

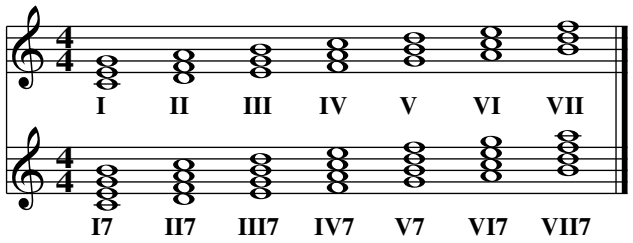


図 12　ハ長調のダイアトニックコード (5 和音である G9 を除く)

2.4.5　ダイアトニックコード

ある音階に存在する音のみを利用して，基本的な形の和音をすべての音に応じて配置したものをダイアトニックコードと呼ぶ．ポピュラー音楽の作曲で広く用いられる．基本的な形は，音階中の全音に対応する 3 和音と 4 和音などが対象となる．ダイアトニックコードから構成されるコード進行をベースに，和音構成音から旋律構成音を選択し非和声音などで装飾する旋律作曲手法が一般的である．この方法で作曲することにより不協和音の発生を大幅に抑えることが可能となる．5 和音となる V9 を除いたハ長調のダイアトニックコードを図 12 に示す．

2.4.6　コードネーム

コードネームとは，ポピュラー音楽やギター演奏などで用いられる和音の形と根音を示す記号である．和音の根音を英語式音名で表記し，その後ろに形に応じて種類を記述する．旋律の和声音は和音の種類により左右されるため，コード進行中のコードネームの種類が多いほど旋律内で使うことのできる和声構成音が増えることから，表現力が高いと言える．表 5 に主なコードネームの表記と別名を示す．根音がドの音である指すフォーのコードネームは，ドを指す「C」にサスフォーを指す「sus4」を加えた「Csus4」となる．また，ある和音に対し最低音を変更し和音を転回させたコードを分数コードと呼び，コードネームにスラッシュ「/」とベース音を表記する．コードネーム「C」の和音に対しベース音をミに変更した場合のコードネームは，ミを指す「E」を加えた「C/E」となる．コードネーム「C」と「C/E」の違いを図 13 に示す．

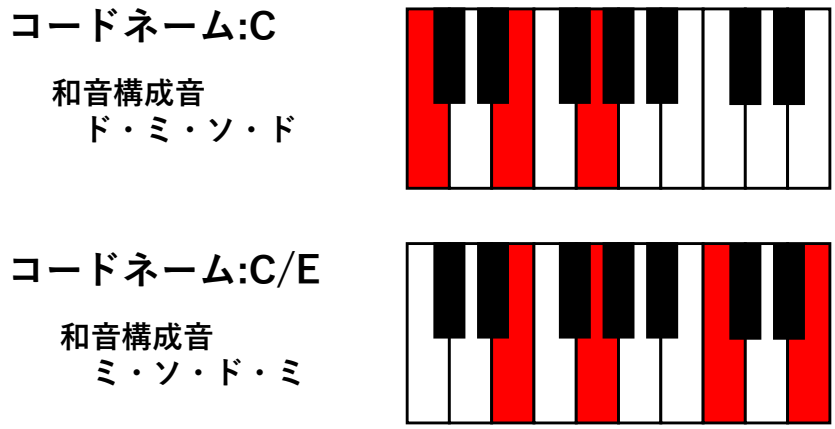


図 13　コードネーム「C」と「C/E」の違い

表 5　主なコードネームとその表記

コードネーム	略式表記　根音がドの場合の表記		別名
メジャー・トライアド	M	C	長三和音
マイナー・トライアド	m	Cm	短三和音
サスフォー (サスコード)	sus4	Csus4	–
ディミニッシュト・トライアド	dim	Cdim	減三和音
オーグメンテッド・トライアド	aug	Caug	増三和音
セブンス	7	C7	属七の和音
メジャー・セブンス	M7	CM7	長七の和音
マイナー・セブンス	m7	Cm7	短七の和音

3　関連研究

旋律の自動生成手法には大きく分けて次の 2 種類が存在する．

- 和声を軸にした旋律の自動生成
- 旋律の自動生成

旋律の自動生成では，多くの場合律動も含めた生成を行うため，律動を軸とした生成手法は一般的ではない．本章では，本研究と関連する旋律の自動生成手法を示し，本研究との類似と相違点を議論する．

3.1　和声を軸にした旋律の自動生成

和声を軸にする生成手法では，旋律の生成に利用する音を絞ることが可能となり，音楽理論に忠実で音階が崩れない旋律を生成することができる．この方針に適した手法として，遺伝的アルゴリズム [7] を用いた手法 [8][9][10] や，RNN を用いた手法 [11] などが挙げられる．一方で，旋律生成に入力としてコード進行が必要な点と，旋律構成音が和声構成音から強く制限される点が欠点として挙げられる．特に，ポピュラー音楽では旋律に音階に含まれない音を使うことも多い．コード進行が限定された状況では，転調や借用和音の位置が固定されることから，表現力のある楽曲にすることが困難である．本研究では，旋律を直接生成する手法を取る．

3.2　旋律の自動生成

旋律の自動生成手法として，歌詞を入力とし言葉の音律から音高の上下を推定する手法 [12] や，遺伝的アルゴリズムを用いた旋律と律動の組み合わせによる作曲手法 [13] などが提案されているが，和声を軸にした手法と比較すると広く提案されていない．深山ら [12] は，日本語の音律を基に旋律の音高を設定する楽曲生成システム Orpheus を提案した．入力された日本語文章のイントネーションから音高の上下を推定し，旋律として生成する．和声付与は一定のパターンが格納されたコード進行データベースから，利用者が任意のコード進行を付与する．本研究とは入力の形式が異なる．田中ら [13] は，既存楽曲のリズムパターンと音高列をデータベース化

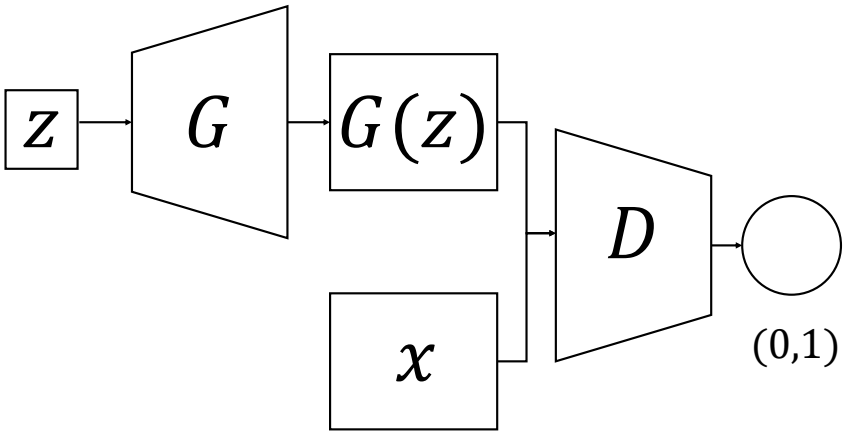


図 14　GAN の構成

し，遺伝的アルゴリズムにより組み合わせることで新しい楽曲を生成する手法を提案した．既存楽曲の音高列を用いることから，遺伝的アルゴリズムによる楽曲生成 [8][9] で課題となる不協和音の発生を抑制することができる．リズムパターンと音高列を組み合わせた際にはみ出した音高は無視する．この手法は既存楽曲の一部のパターンをそのまま流用することから，本研究の目指す方向性とは異なる[*10]．

3.2.1　GAN を利用した手法

何かを生成するネットワークの代表例として，敵対的生成ネットワーク (Generative Adversarial Network:GAN)[1] が挙げられる．GAN は，互いに敵対的に設計されたディスクリミネータとジェネレータの対により構成される生成ネットワークである．GAN の構成を図 14 に示す．ディスクリミネータ D はジェネレータ G の生成と学習データ x を比較し区別することを，ジェネレータ G はディスクリミネータ D で判別が不能な物をノイズ z から生成することを目的とする．ジェネレータがディスクリミネータにおいて学習データと判別がつかない出力が可能となると，学習データに似た出力が可能となる．

[*10] 本研究は音高列の生成を目的としているため，音高列のデータベースの幅を広げることが可能となる．

ジェネレータとディスクリミネータにさまざまなモデルを適用することで，組み込んだモデルに応じた生成ネットワークとしてあつかうことができる．GAN のバリエーションとして，CNN を組み込み画像生成を可能にした DCGAN(Deep Convolutional GAN)[2]，RNN を組み込み時系列データの生成を可能にした C-RNN-GAN[14] などが挙げられる．本研究では，画像ベースの旋律生成を行うため，DCGAN の派生である WGAN(Wasserstein GAN)[15] を利用する．

Yang ら [16] は GAN を利用した音楽生成モデルである MidiNet を提案した．複数種類のモデルを含み，8 小節の旋律か旋律と和音の対を生成することができる．このモデルは，MIDI ファイルを小節単位に分割しているため，細かい音価を表現した生成が可能となる．専門家への主観評価で，RNN を利用した Google Magenta[11] の MelodyRNN と比較して同様の旋律の繰り返しが少ない出力がされていることから，この手法による出力がより興味深いと示されている．一方でこのモデルは前後の小節に依存した生成を行う影響から，8 小節を通して評価した場合に音楽的な正しさがなく，予期せぬ音が発生するため人工的に生成された音楽である印象が強いことが示された．

Cheng ら [5] は，音楽の自動生成を行う Temporal GAN(TGAN) モデルを提案した．このモデルは，旋律と対応する和音のデータセットから，短い数小節単位の旋律と和音の対を生成するビートモデル G_{pre} と，ビートモデルによって生成された旋律と和音の対を時間軸方向へ順序付けする時間モデル G_t の 2 種類により構成される．提案されたモデル構造を図 15 に示す．ビートモデルが出力された画像を分析した結果，和声と旋律が音楽的に正しい出力が見られたことから，音楽理論をある程度認識することができたと考えられた．一方で，時間モデルは，時間モデルのディスクリミネータに対し，ビートモデルにより生成された楽曲と人間により作成された音楽的に正しい例と悪い例を入力した場合に，log 損失が大差がないことから時間軸方向の音楽的な正しさをディスクリミネータが理解したとは言えないことを示した．

GAN による統合的な作曲手法として，Dong ら [17] は複数の楽器に対応した譜面出力を同時に行うことができる MuseGAN を提案した．MuseGAN には，入力に対応した 2 種類[*11]のモデルと，楽器種類に対応した 3 種類[*12]のモデルを合わせた合計

[*11] ノイズから完全新規で生成するモデルと入力として旋律の一部などを利用するモデルの 2 種類.

[*12] パート別にジェネレータとディスクリミネータを使用するモデル，すべて同じジェネレータとディスクリミネータを使用するモデル，組み合わせたモデルの 3 種類.

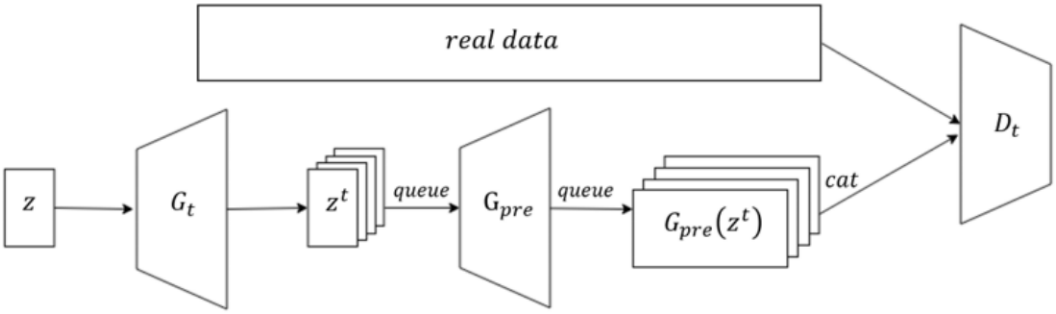


図 15　Cheng らが提案したモデル構造 [5]

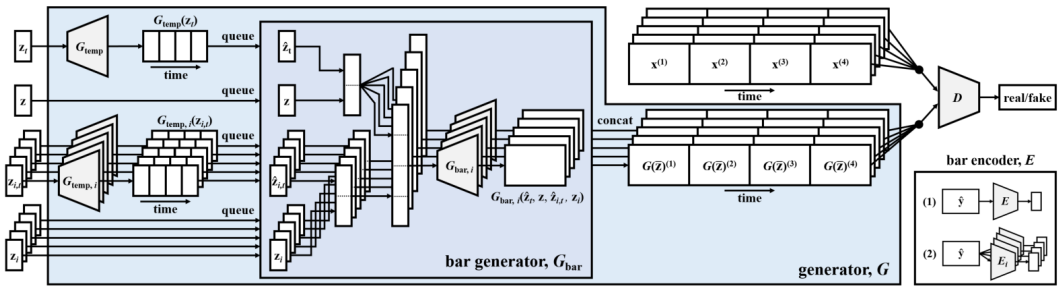


図 16　MuseGAN のモデル構造 [17]

6 種類のモデルを含む．MuseGAN のモデル構造を図 16 に，出力を図 17 に，生成過程を図 18 に示す．

これらの研究は，小節や楽曲の構造などの細かい単位で生成した旋律画像を並べ替えることで旋律を生成する．本研究では，GAN による出力を慣性系とみなさず，旋律概形を模した画像を GAN により生成した後に補正を行う．最終的に補正した旋律画像を変換することで，音楽的に正しい旋律の生成を試みる．

3.3　旋律概形

土屋ら [18] は，音楽の非専門家を対象とし旋律概形を利用した作曲手法を提案した．旋律概形は，各音の音高と音価が陽に表現されず，旋律を 1 本の曲線で示した

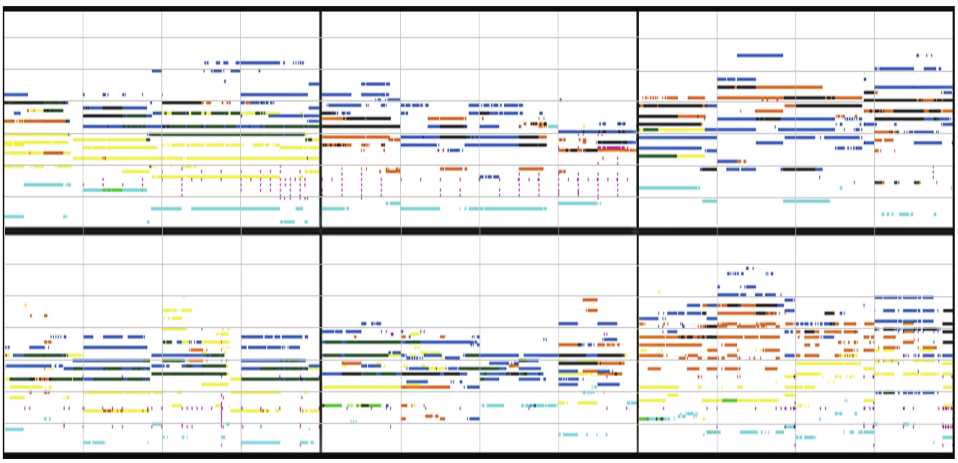


図 17　MuseGAN の出力 [17]

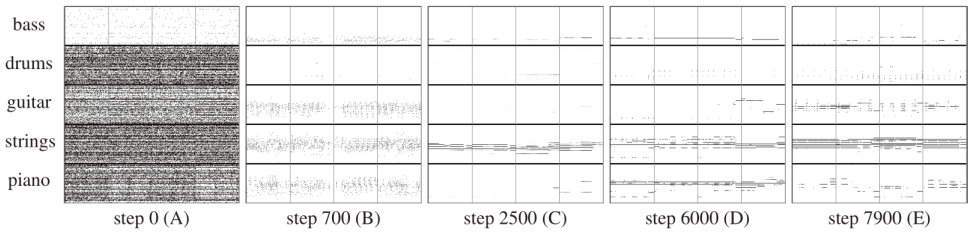


図 18　MuseGAN の生成過程 [17]

ものである．旋律概形は旋律の音高の時系列に対しフーリエ変換を行うことで生成できる．また，生成された旋律概形を音名に単純変換すると連続した音となり音階が崩れてしまうが，この過程に隠れマルコフモデルを用いることで，各音階における正しい音名を推定し旋律に変換する．主観評価から，旋律概形には旋律の大まかな特徴を反映することが可能であると示された．旋律概形の曲線は画像で表現することが可能である．本研究では旋律を画像で表現し，旋律の大まかな特徴を捉えた生成方法として旋律概形に類似したもの[*13]を GAN により生成する．

[*13] 本研究では，土屋らの研究で定義される旋律概形と違い，音高と音価を陽に表現した一般的に「ピアノロール」と呼ばれる形式の画像を生成する．

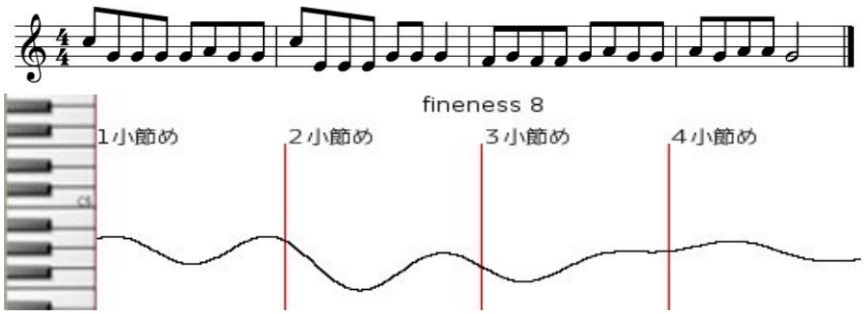


図 19　旋律概形の例 [18]

3.4　コード進行の自動付与

コード進行の自動生成は自動作曲の一分野として研究されているが，多くの場合コード進行から生成する．本研究では，事前に生成された旋律に対してコード進行を付与するため，生成手順が異なる．旋律に対するコード進行の自動付与では，ポピュラー音楽で広く使われる定型コード進行を集めたデータベースからコード進行を充てる手法 [12][19] や，事前に学習された確率遷移モデルから，入力された旋律に対し自動で適合するコード進行を付与する手法 [20][21][22][23] や，ルールベースで付与する手法 [24] などが提案されている．しかし，これらの手法では，付与されるコード進行が事前に用意したパターンに限定されてしまう問題がある．

筆者の先行研究として，遺伝的アルゴリズムによる旋律へのコード進行付与手法を提案した [25]．MIDI ノートナンバーで構成される 4 音を遺伝子として持ち, コードネームとして表現される和声以上の表現を可能とした. 個体表現の例を図 20 に示す．評価関数として，入力旋律の構成音と探索中の和声構成音との一致度, 和声の構成形状 (図 21), 和声の進行の 3 要素の評価を総合した適応度を利用する. この手法では 4 和音の探索を行うため，既存のポピュラー音楽で多く使われるコードに対応する．また，コードネームを利用せずコードの形状を音の組み合わせで評価するため，任意の形状を登録することでより複雑な和音の付与が可能となり，より多くのコードネームを付与対象とすると同時に転回形の表現も可能となることから表現力で優

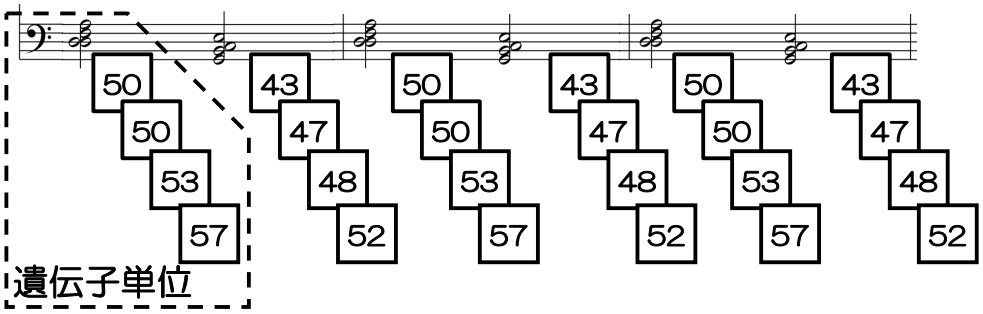


図 20　遺伝的アルゴリズムによるコード進行付与の個体表現の例 [25]

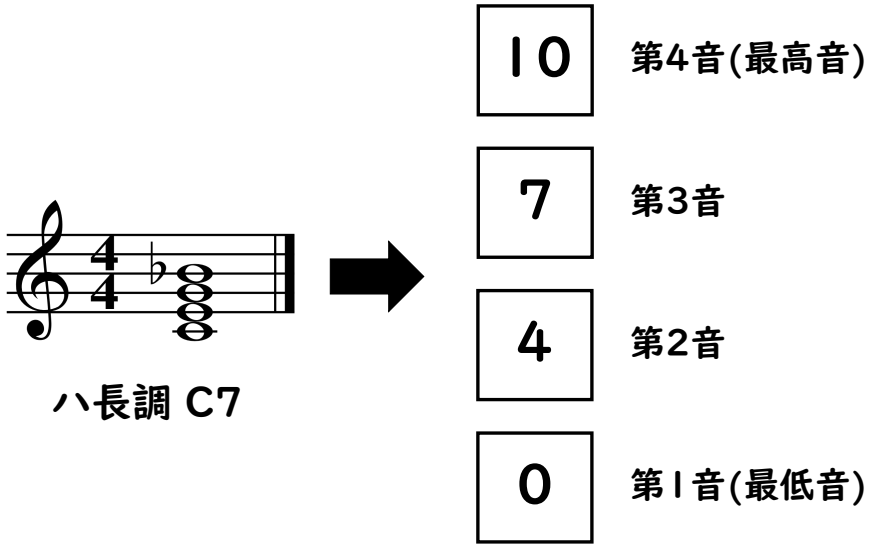


図 21　形状モデルへの和音の登録例 [25]

位性がある．本研究では，遺伝的アルゴリズムを用いたコード進行付与手法を採用する．

4　提案手法

自動作曲の分野では，主観評価で人間によって作曲された楽曲と肩を並べる楽曲を生成する決定的なアルゴリズムは，現状存在しない．主な旋律生成手法の比較を表 6 に示す．自動生成される旋律はあたりはずれが激しく，単純な乱数で音を並べた状態に等しい旋律もある．一方で，主観評価で高い評価をえる旋律が出力されることもある．HMM や RNN を用いた手法で旋律を生成するには，生成のベースとなる短区間の旋律かコード進行の一部を入力する必要がある．ノイズから旋律を生成することができる GAN では，計算機の性能を生かし大量の旋律を生成し，その中から最も良い旋律を選択する手法を取ることが可能である．また，RNN で生成された旋律は，同じ旋律構成の繰り返し[*14]になることが多く，GAN による生成手法はこの点の改善が確認されている [16]．さらに，旋律を画像化することで旋律の特徴を捉えた概形として表現することが可能になる．上記から本研究では，旋律概形に重きを置いた作曲手法として，GAN による画像ベースの旋律生成手法を提案する．

一方で，GAN による出力旋律は音階が崩れる可能性がある [5] ことから，出力画像に対し音階に合わせた補正を行うことで，この欠点を補う．補正には生成された旋律画像内の濃度情報と，学習データから生成した音高遷移確率を利用する．

提案手法の概要を図 22 に，補正イメージを図 23 に示す．黒と灰色の箱が音を示し，曲線が想定される旋律概形のイメージを示す．左側が補正前，右側が補正後を示す．補正対象の画像を生成するモデルは音高と時間方向への展開を行い，白色で表現した旋律部分と黒色の背景に二値化した画像を学習および生成する．画像は左上を原点とし，下方向に MIDI ノートナンバーによる音高情報を，横方向に時間情報を格納する．

本手法では以下の手順で楽曲生成を行う．本章では，各手順について説明する．

1. 旋律の画像化
2. 旋律画像の生成

[*14] 主題に対する模倣と変形で構成されるポリフォニー音楽では問題にはならないが，近年のポピュラー音楽において支配的な作曲手法とは言えない．また，繰り返しあらわれる旋律が学習データに大きく影響される可能性がある．

表 6　主な旋律生成手法の比較

	ルールベース	GA	RNN	GAN
0 からの生成が可能か	no	yes	no	yes
音階外の音を許容できるか	no	yes	yes	yes
表現力があるか	no	yes	no	yes

3. 生成された画像の補正
4. 音価の付与と小節の決定
5. コード進行の付与

4.1　学習対象となる旋律の画像化

学習対象とする旋律画像の作成のために，学習対象の楽曲 MIDI ファイルから MIDI ファイル内の旋律を抽出し画像化する．画像ベースで旋律を表現することにより，最高音・音高の推移などの旋律の特徴があらわれることから，その特徴を持った画像を生成することを期待する．画像で表現可能な旋律特徴の例を図 24 に示す．

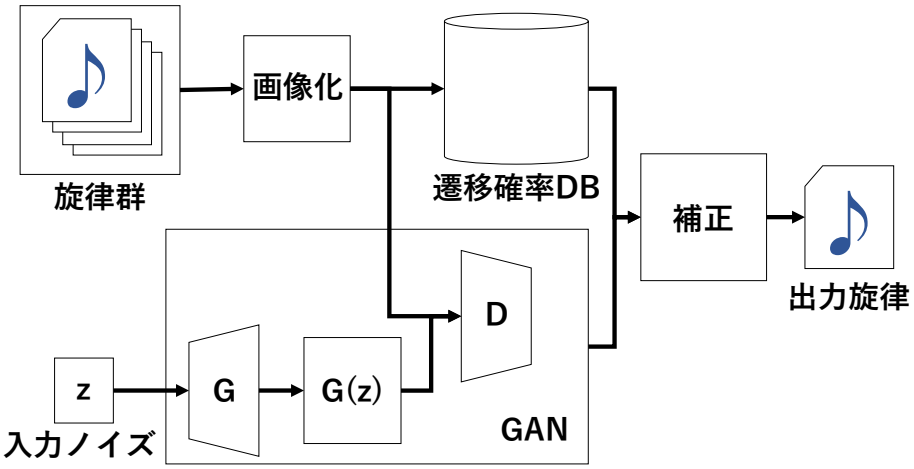


図 22　提案手法の概要

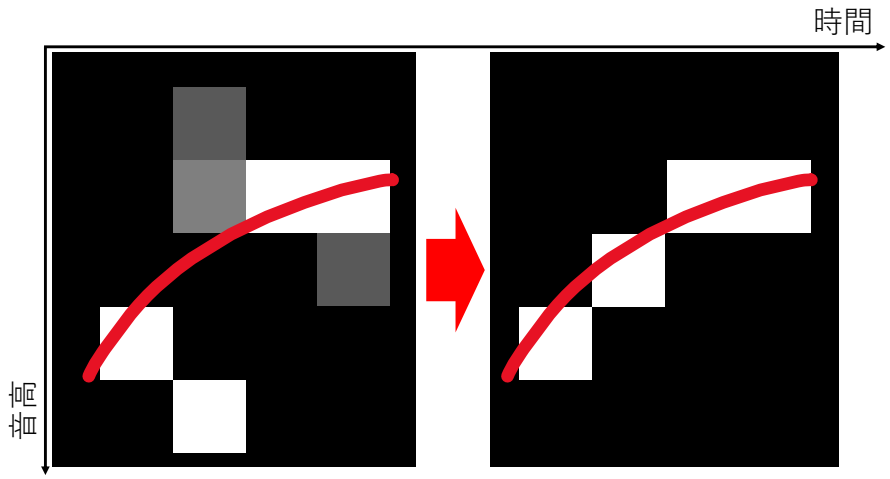


図 23　補正イメージ

本手法の旋律画像は，横軸に時間軸，縦軸に音高を取る．また，より長い区間の旋律の音高情報を含んだ旋律画像とするため，元の MIDI ファイル内の旋律が持つ音価は無視し圧縮する．圧縮された音価は第 4.3.3 項で復元する．画像化された旋律画像は，元データとして GAN の学習とディスクリミネータの評価に利用する．旋律を画像化する際に，2 音間の遷移確率をデータベース化し音高遷移確率データベースを生成する．このデータベースは旋律画像の補正で用いる．

4.2　旋律画像の生成

本フェーズでは，第 4.1 節で加工した旋律画像を利用し，敵対的生成ネットワークを利用した旋律画像の生成を行う．画像の生成を目的とすることから，ネットワークとして WGAN-gp を利用する．敵対的生成ネットワークでは，乱数によるノイズから画像を生成することが可能で，RNN と比べて表現力がある [16]．また，GAN により生成された旋律画像は，概ね旋律の範囲内の出力が行われることが筆者の先行研究 [26] により示されている．一方で，CNN をディスクリミネータとジェネレータに用いた GAN を利用する課題として，GAN の生成器や判別機が音楽の基本的構造を理解しているとは言えない [5] 点と，生成された旋律の前後で音階が崩れてし

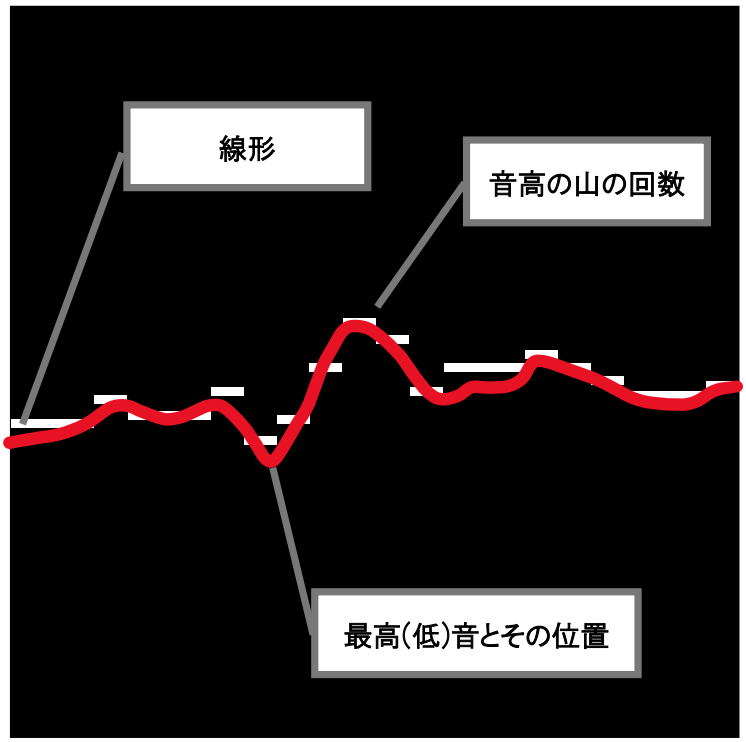
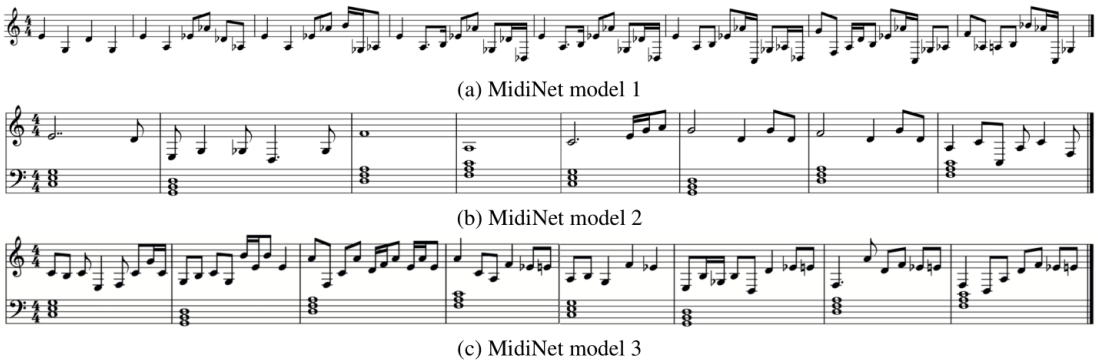


図 24　画像化で表現可能な旋律特徴の例

まうケースが多くみられる点が挙げられる．MidiNet[16] により生成された旋律 (図 25(a)) には，後半にかけてフラットとナチュラルの臨時記号が数多く出現しており，音階が崩れていることが示されている．

4.3　生成された画像の補正

GAN により生成された旋律画像には音階の崩れが発生する可能性があり，汎用的な画像生成モデルを利用したことから鮮明な画像にならず滲みが発生するため，GAN により生成された画像に対し画像の特徴を生かしつつ音階を崩さない補正を行う．候補となる音高に対して，グレースケール濃度と前後の遷移確率から算出する値が最も大きい音高を出力する．GAN による出力でグレースケールを許容すること



(a) MidiNet model 1

(b) MidiNet model 2

(c) MidiNet model 3

図 25　MidiNet の 3 種類のモデルにより生成された旋律 [16]

により，二値化出力では表現できな音階の幅を出力することができるため，特徴を大きく損なうことなく音階的に正しい音高を推定することができる．グレースケールを許容したことにより，生成画像にはノイズが混ざる事や滲む部分があるため，音源化するにあたり二値化する必要がある．単純な二値化の方法として，単純な閾値を利用した二値化の手法と濃度と音高に対し加重平均を取る手法など，比較的単純な画像加工の手法も考えられるが，本研究ではグレースケールの値と学習旋律全体の 2 音間における音高遷移確率の二点に着目した補正を行う．同じ生成画像に対し，OpenCV[27] による単純な閾値を利用した二値化[*15]，濃度と音高に対し加重平均を取った二値化 [26]，単純最大濃度を取る二値化，提案手法による二値化をそれぞれ行った画像を図 26 に示す．

比較的単純な画像加工の手法を取ることで，GAN が局所的な前後の関係性から生成時に反映された音階関係が失われることを避けることができる．音高推定が同方向に長く続く事で音階が狂う可能性があることから，中央から左に音高推定を行い，左半分の音高が確定次第右半分の音高推定を行う．画像の補正にあたり，最初に推定する時刻 t_0 では，グレースケールによる音高推定値のみを利用する．音高推定の方向を図 27 に示す．

[*15] 閾値は自動決定.

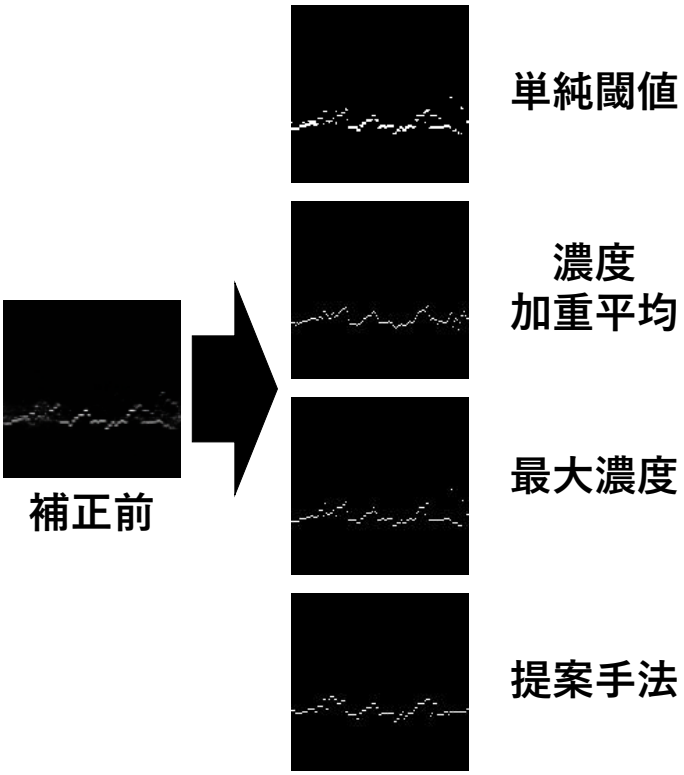


図 26　同じ生成画像に対し，4 パターンの二値化を行った例

4.3.1　グレースケールによる音高推定値

補正手順として，始めに GAN により生成された画像の色濃度を利用した音高推定値を算出する．この推定値は GAN により出力された画像の濃度から算出される推定値である．音高推定値の算出イメージを図 28 に示す．濃度は青枠内の 9 ピクセルに対するグレースケールの値を指す．この推定値は，色濃度が高いと上昇し低いと減少する．また，ある程度の濃度以上になると上昇割合を抑え気味にし，濃度が低い場合には急激に上昇するように補正を行う．この操作は，生成画像で明確に濃度が高くなる区間が限られていることから，ある程度の濃度がある場合に算出されやすくするために行う．

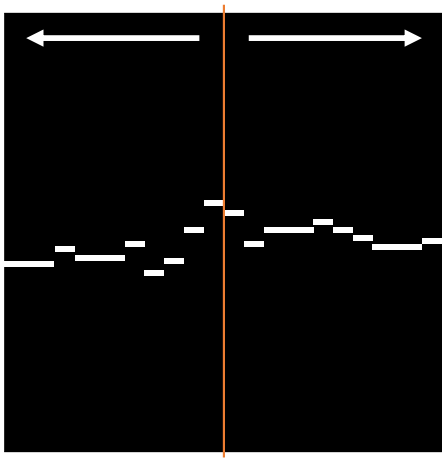


図 27　画像の補正方向

4.3.2　音高遷移確率を考慮した音高推定値

補正手順の第二段階として，グレースケールによる音高推定値と音高遷移確率から最終的な音高推定値を算出する．画像内にある時間方向で直前の決定された音高から，現在の音高への遷移確率とグレースケールによる音高推定値を乗算したものを最終的な音高推定値とする．なお，この遷移確率を考慮した音高推定値は曲の構造に対する補正を行う目的ではなく，グレースケールによる音高補正に対し音階の観点から補助を行う目的で導入したもので，旋律生成手法としての隠れマルコフモデルに類似する扱いは行わない．遷移確率は学習画像生成時に同時に生成した音高遷移確率データベースから取得する．最終的に，音高推定値が最大となる音高を時間単位の音高として決定する．音高推定値の算出イメージを図 29 に示す．prev time は直前の画像の濃度情報を，this time は現在の画像の濃度情報を，G_{eval} はグレースケールによる音高推定値[*16]を，P_{eval} は音高遷移確率を，$Eval$ は最終的な音高推定値を示す．

[*16] この値は濃度に補正をかけた値である．

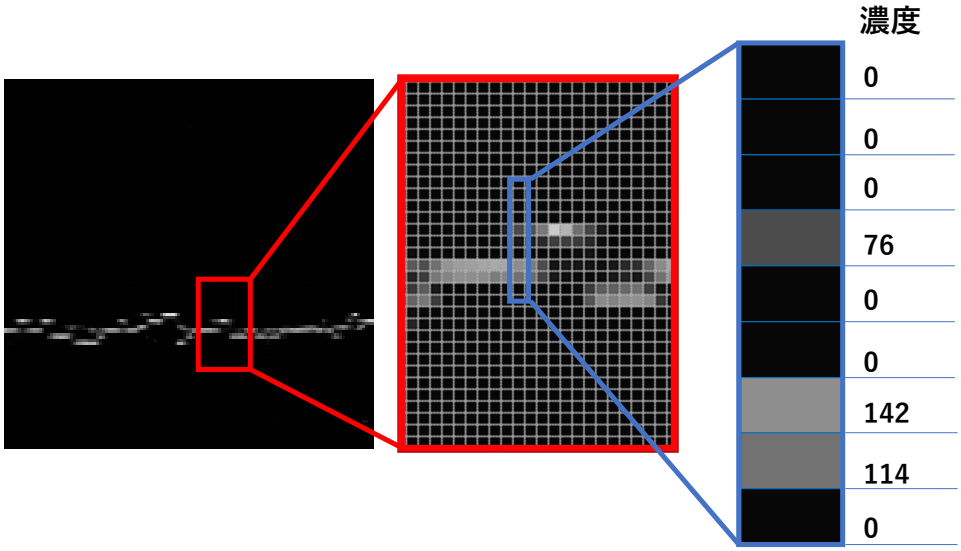


図 28　グレースケールを利用した音高推定値のイメージ

4.3.3　音価の付与と小節の決定

補正手順の最終フェーズとして，生成された音高列に対し音価の付与とそれに伴う小節の決定を行う．事前に生成されたリズム列に対し，生成された音高列を埋め込む手法 [13] も存在するが，リズム列の内容により音高列を削る必要があるため，本研究では生成された画像内にある連続した音高が続く区間を長音とみなす．

4.4　コード進行の付与

本研究では旋律をベースとした楽曲生成を行うため，生成された旋律に対しコード進行を付与する必要がある．

コード進行の付与手法には遺伝的アルゴリズムによる手法と GAN による手法の二種類を使い，最終的な適合度の高い方を出力とする．適合度はコード進行の形状と旋律との一致度により算出される．ルールベース・HMM・RNN による手法に対し，これらの手法はコードネームに限定されることなく探索できることから，より多くの可能性を含んだ出力が可能となる．主なコード進行付与手法との比較を表 7 に

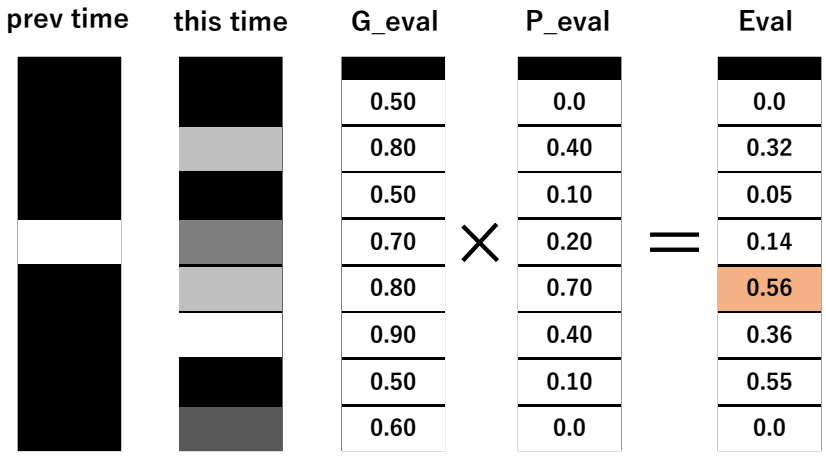


図 29　補正値の算出イメージ

表 7　主なコード進行付与手法の比較

	ルールベース	GA	RNN	GAN
旋律と同時に生成することが可能か	no	no	no	yes
コードネームに限定されない探索が可能か	no	yes	no	yes
借用和音を許容できるか	no	yes	yes	yes
表現力があるか	no	yes	no	yes

示す.

5　プロトタイプシステムの実装

提案手法の基本的な性能を評価するために，簡易的な単純 WGAN-GP[28] を利用したプロトタイプシステムを実装した．システムは Python3.6.9 を用いて実装した．

プロトタイプシステムは次の要素により構成される．

- 学習画像の加工
- 旋律画像の生成
- 生成された画像の補正

旋律の評価を目的としたプロトタイプシステムのため，コード進行の自動付与は行わない．

5.1　学習楽曲の加工

学習対象とする旋律画像の作成のために，学習対象の MIDI ファイルから旋律画像の生成を行う．本フェーズの入力は，学習対象となる楽曲群の MIDI ファイルであり，出力は旋律画像群と音高遷移確率データベースとなる．MIDI ファイルの解析には Python のライブラリである pretty_midi[29] を利用した．複数のパートを含む MIDI ファイルの場合は，旋律に相当する部分を抽出し利用する．楽曲加工と旋律の画像化の手順を図 30 に示す．

本システムの学習画像は，Nottingham Database[30] から旋律を抽出し，MIDI 形式に変換したもの[*17]を利用した．また，抽出した旋律はハ長調に正規化する．

5.1.1　旋律画像の生成

本フェーズの入力は単旋律の MIDI ファイルであり，出力は旋律画像となる．旋律画像は MIDI ファイルを音高と時間の 2 方向に展開し画像化したものを利用する．音高情報を二次元展開し，縦軸は音高である MIDI ノートナンバーを 0 から 127 まで格納，横軸は時間軸とする．画像背景は RGB 値で (0,0,0) である黒を利用し，

[*17] このデータベースは ABC 記譜法により表現されているため，MIDI 形式に変換することなく画像化することが可能であるが，システムの汎用性を確保する観点から MIDI 形式への変換を行う．

ノートが存在する区間を RGB 値 (255,255,255) で指定する．MIDI ファイルを画像にするにあたり，音価の情報は無視しすべて同一とする．縦方向は MIDI ノートナンバーに合わせ 120 ピクセル，横方向は音の数のピクセル数の画像を生成する．128 音で構成されない旋律の場合，画像が長方形になる場合があるため，ImageMagick[31] の Convert コマンドを利用して正規化を行う．最終的に 128*128 の画像を学習画像として利用する．また，画像化する際に音高の推移回数と MIDI ノートナンバー単位での 2 音間の推移回数を記録する．この情報は，音高遷移確率データベースの作成に利用する．旋律画像の形式を図 31 に示す．

5.1.2　音高遷移確率データベースの生成

本フェーズの入力は第 5.1.1 項で旋律を画像化する際に取得した音高の遷移回数であり，出力は音高遷移確率データベースとなる．画像化した旋律に対し，2 音間の遷

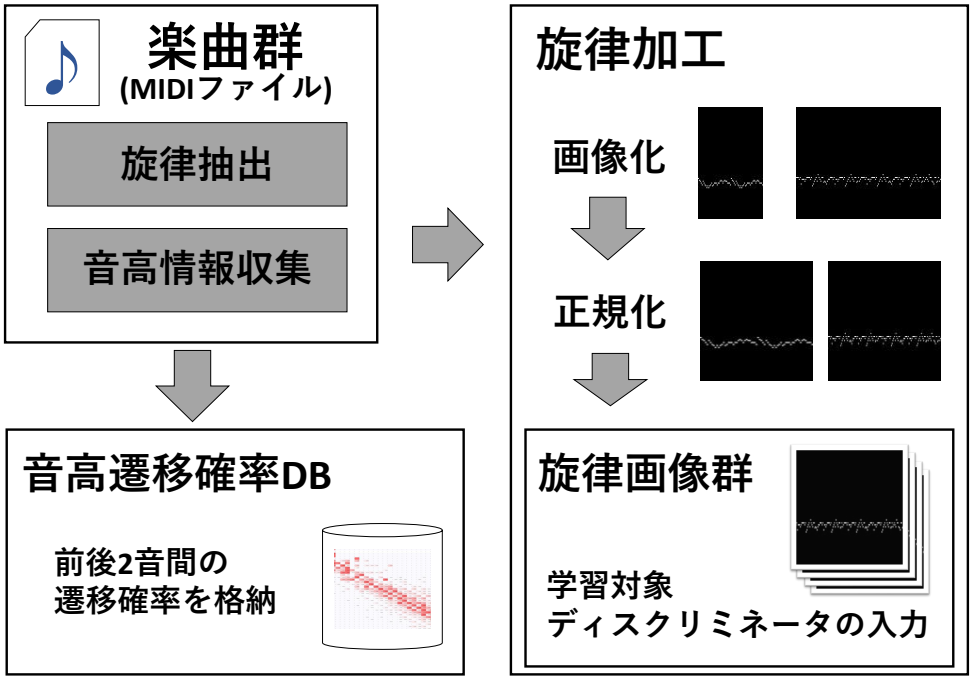


図 30　楽曲加工・旋律画像化の手順

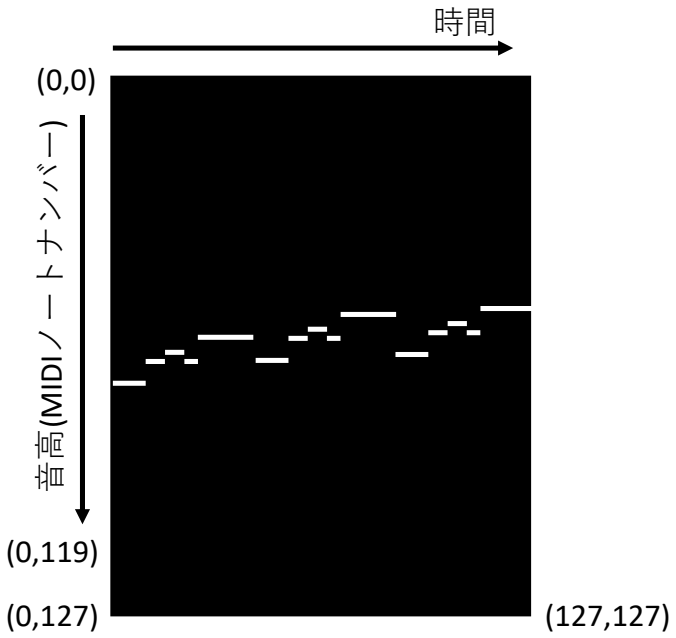


図 31　旋律画像の形式

移回数を二次元テーブルに格納する．二次元テーブルへの格納方法を式 (1) に示す．格納した遷移回数を用いて総遷移回数を用いた確率を算出し，音高遷移確率データベースを生成する．生成されたテーブルの一部を図 32[*18]に示す．

$$bigram[prev_{note}][current_{note}] \tag{1}$$

5.2　GAN による画像の生成

生成した学習対象となる旋律画像を用いて，GAN による学習と旋律画像の生成を行う．本フェーズの入力は第 5.1.1 項で生成した旋律画像群とノイズ[*19]で，出力は未補正の旋律画像となる．本システムでは単純 WGAN[28] を利用した簡易的な画像生成モデル[*20]を利用する．実装には TensorFlow[32]1.14.0 を用いた．ジェネレー

[*18] 本来のテーブルには確率である数字のみが格納されており，本図は可視化目的ですべてのセルに対し確率が 1 に近い場合に赤色になるように着色した．

[*19] ノイズの生成には乱数を利用する．

[*20] この生成モデルは旋律画像以外の画像生成にも対応する汎用的なモデルである．

タとディスクリミネータの畳み込み層は 4 層で構成される．学習画像は 1037 枚である．

5.3 生成された画像の補正

GAN により生成された画像に対し，滲みの除去と音階に合わせた音高推定を行う．本フェーズの入力は第 5.2 節で生成した未補正の旋律画像で，出力は補正後の旋律画像と旋律の MIDI ファイルである．補正後の画像は学習対象となる旋律画像の形式に従う．生成した画像の横軸を時間軸と扱い，1 ピクセル単位で縦方向の音高の濃度情報と音高遷移確率を利用して時間単位の音高を決定する．

時刻 t における音高候補 $note_t$ は，画像のグレースケールの値から算出される G_{eval} と直前の時刻 $t-1$ で確定した音高空の遷移確率である P_{eval} を用いて推定される．

画像濃度による音高推定値 G_{eval}

画像濃度からの補正値 G_{eval} は，式 (2) により算出される．$greyscale$ は時刻 t における音高の濃度を示す．$grayscale$ は音高軸である縦軸 1 ピクセル目から 120 ピクセル目まで取得する[*21]．また，濃度に対しシグモイド関数を適応することで，画像濃度が一定の値以上に推定値の伸びを小さくし，濃度が一定未満の場合の推定値の伸びを大きくする．中濃度から高濃度の場合の音高推定値の差を少なくすることで，より多くの音高候補を残し，遷移確率の影響を高める事ができる．また，画像濃度が 0 の場合にある程度可能性を残すことで，音階に合った音高を推定されやすくする．G_{eval} のグラフを図 33 に示す．

$$G_{eval} = \frac{1}{1 + exp(-(greyscale/100))} \tag{2}$$

遷移確率を考慮した音高推定値 P_{eval}

音高遷移確率からの補正値 P_{eval} は，決定済みの時刻 $t-1$ の音高 $note_{t-1}$ から，時刻 t の音高候補 $note_t$ への遷移確率を用いる．また，t_0 となる 64 ピク

[*21] 121 ピクセル目以降は MIDI ノートナンバーに対応しないため省く．

セル目は $P_{eval} = 1$ とし，G_{eval} による推定を行う．遷移確率は，第 5.1.2 項で生成した音高遷移確率データベースを利用し取得する．

P_{eval} と P_{eval} を組み合わせた音高推定

各時刻 t における補正された音の候補 $NoteEval_t$ は式 (3) を用いて導出される．P_{eval} と P_{eval} に対し重みである G_{weight} と P_{weight} を設定することができる．$NoteEval_t$ が最大となる音高が時間単位での補正された音 $note_t$ として算出される．

$$NoteEval_t = G_{eval} \cdot G_{weight} * P_{eval} \cdot P_{weight} \tag{3}$$

音高推定は，画像の中央である 64 ピクセル目から左方向に行い，0 ピクセル目に達した段階で 65 ピクセル目から右方向へ推定を行う．方向が変わる 65 ピクセル目は 64 ピクセル目を時刻 $t-1$ と扱い，64 ピクセル目から 65 ピクセル目への音高遷移で P_{eval} を算出する．

音価の付与

本システムでは，連続した音高が続いた場合には長音とあつかう．生成画像の時間軸が 128 ピクセルであることから，八分音符を基本音価とし，標準状態で生成旋律は 16 小節で構成される[*22]．基本音価 NV_d に同じ音価が続いた回数 C_n を掛け合わせたものを最終的な音価 $NoteValue$ とする．音価の付与方法を式 (4) に示す．

$$NoteValue = NV_d \cdot C_n \tag{4}$$

[*22] 評価試験では旋律の評価のみを行うため，プロトタイプシステムでは小節の設定は行わない．

図 32　生成されたテーブルの一部

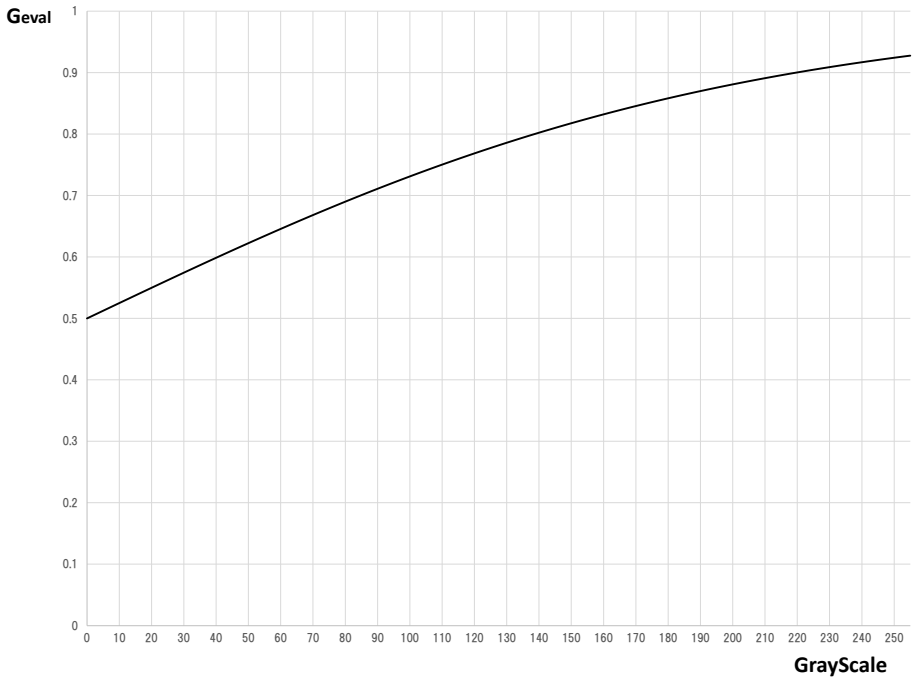


図 33　二値化による音高推定値 G_{eval}

6　評価実験

プロトタイプシステム[*23]を用いて旋律を生成し，その中からランダムに選択した旋律に対して主観評価実験を行った．生成した旋律画像の例を図 34 に示す．

実験被験者は 20 代の男性 8 名で，そのうち楽器経験者が 3 名である．また，8 名は作曲・編曲の経験がなく，簡易的なコード進行の付与もできない．実験被験者の詳細を表 8[*24]に示す．評価対象旋律は，出力旋律からランダムで選択した図 35 から図 40 の 6 種類である．この旋律は，小節の設定を行っておらず，音部記号の指定も行っていない[*25]．また，右側の補正画像は上下が反転した状態[*26]で表示する．

実験は評価対象の譜面と音源を提示し，旋律全体に対する 5 段階評価と任意の区間に対し良/悪を選択する区間評価の 2 点による評価を行った．5 段階評価は「とて

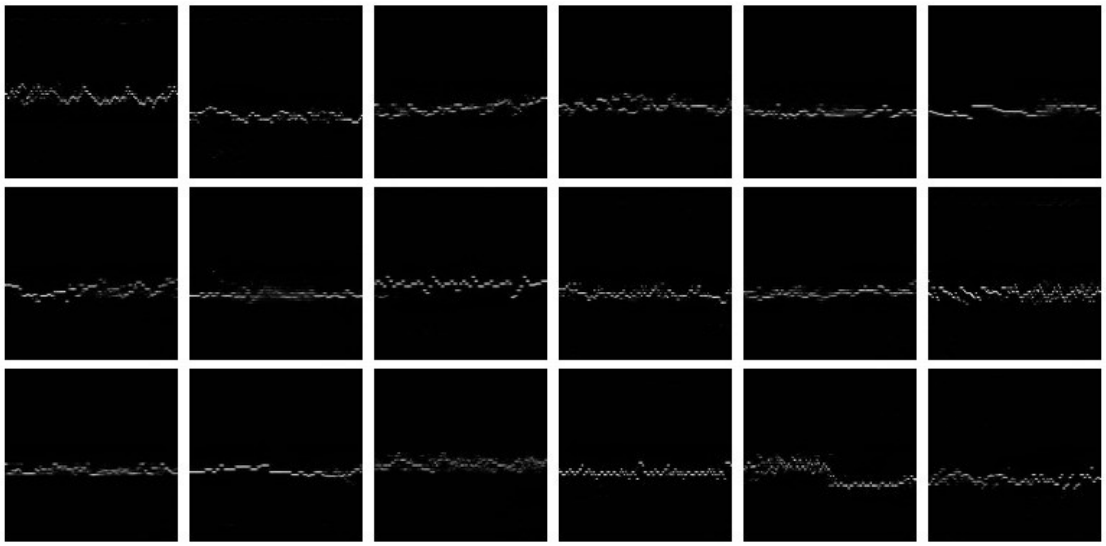


図 34　生成された旋律画像の例

[*23] 重み G_{weight}，P_{weight} はそれぞれ 1 を設定した．

[*24] 被験者 F は楽器の演奏経験に「ピアノ (片手)0.5 年」と記述したが，演奏経験が音感に与える影響を考慮するための確認であることから，本人の確認のもと演奏経験がないとみなした．

[*25] ハ長調の譜面として表示する．

[*26] 音高が高い場合が上側になる状態．

表 8　実験被験者詳細

	楽器の演奏経験	調性	不協和音	作曲経験
被験者 A	トランペット 4 年	–	感じる	無
被験者 B	ピアノ 5 年	感じる	–	無
被験者 C	箏 3 年	感じる	感じる	無
被験者 D	無	–	感じる	無
被験者 E	無	–	感じる	無
被験者 F	無	–	感じる	無
被験者 G	無	–	–	無
被験者 H	無	–	–	無

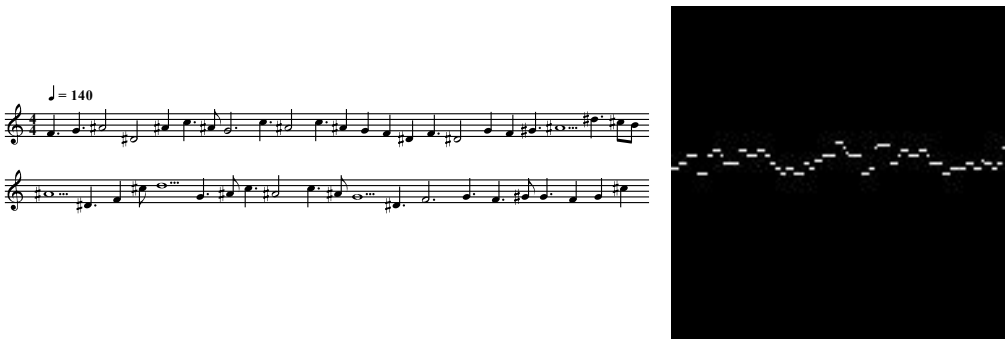


図 35　評価対象旋律 1

も悪い」を 1 とし，「とても良い」を 5 とする．5 段階評価の結果を表 9 に，区間評価の結果を旋律 1 から順に図 41 から図 46 に示す．区間評価の結果の図中左側の数字は 5 段階評価を示し，青箱部が良い，赤箱部が悪いを示す．

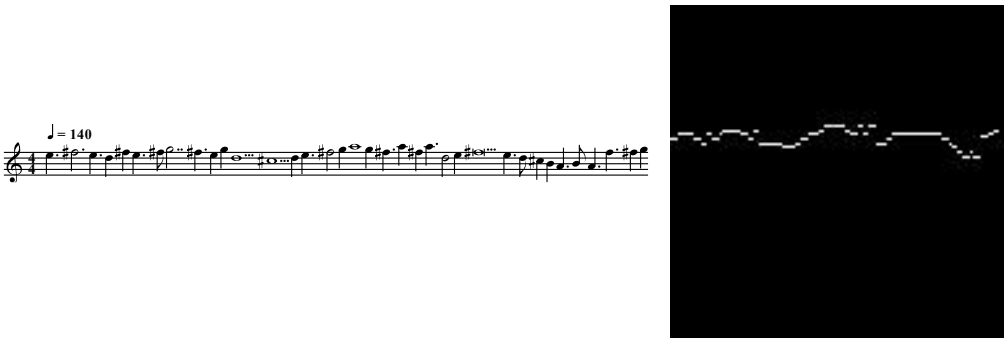


図 36　評価対象旋律 2

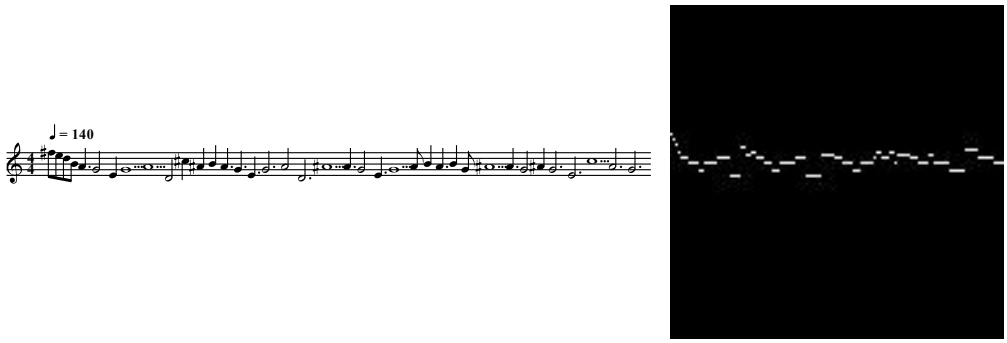


図 37　評価対象旋律 3

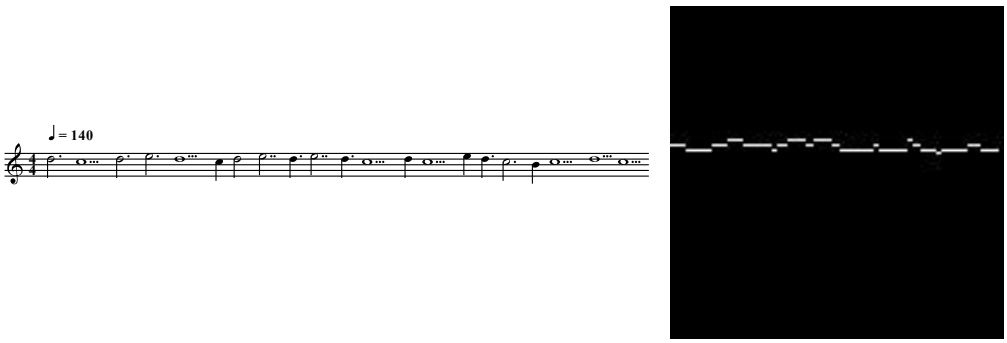


図 38　評価対象旋律 4

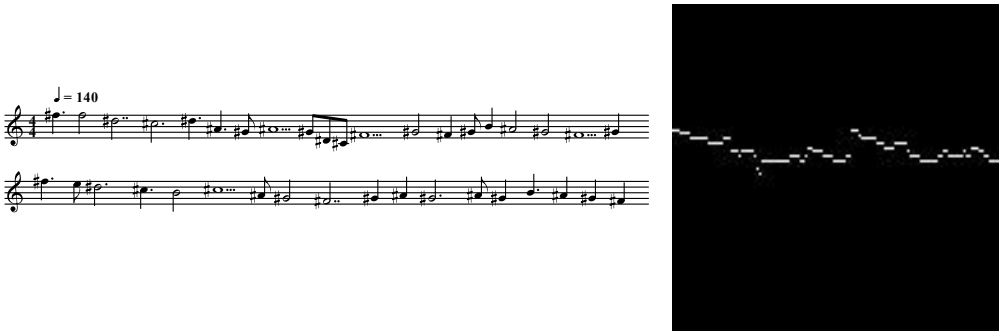


図 39　評価対象旋律 5

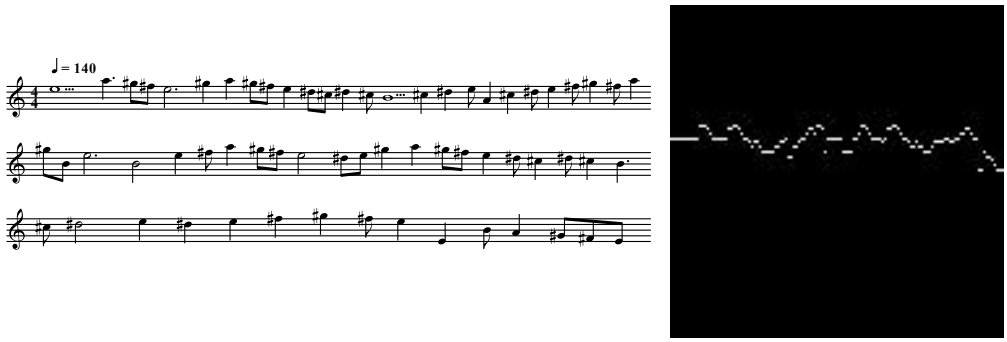


図 40　評価対象旋律 6

表 9　主観評価結果

	旋律 1	旋律 2	旋律 3	旋律 4	旋律 5	旋律 6
被験者 A	3	3	2	3	3	4
被験者 B	4	3	1	3	4	5
被験者 C	4	3	2	1	2	5
被験者 D	3	2	3	2	3	4
被験者 E	3	2	4	3	4	5
被験者 F	4	3	2	1	2	5
被験者 G	4	3	5	3	5	5
被験者 H	3	2	3	4	4	5
平均点	3.50	2.63	2.75	2.50	3.38	4.75

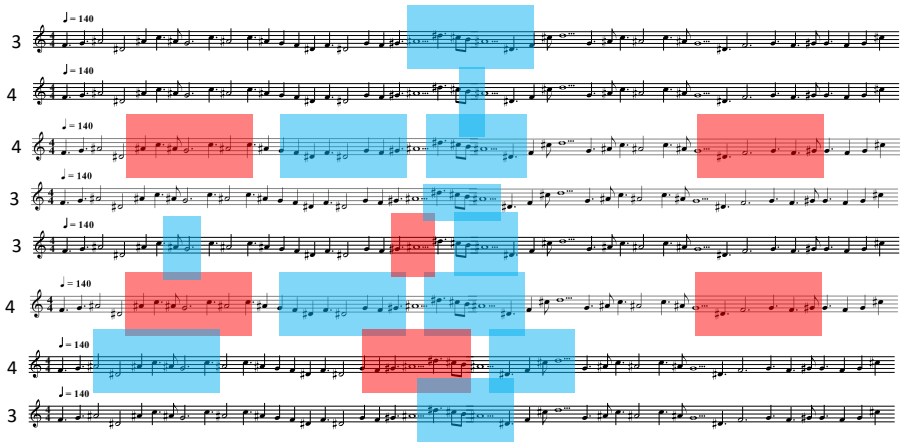


図 41　評価対象旋律 1 の主観評価

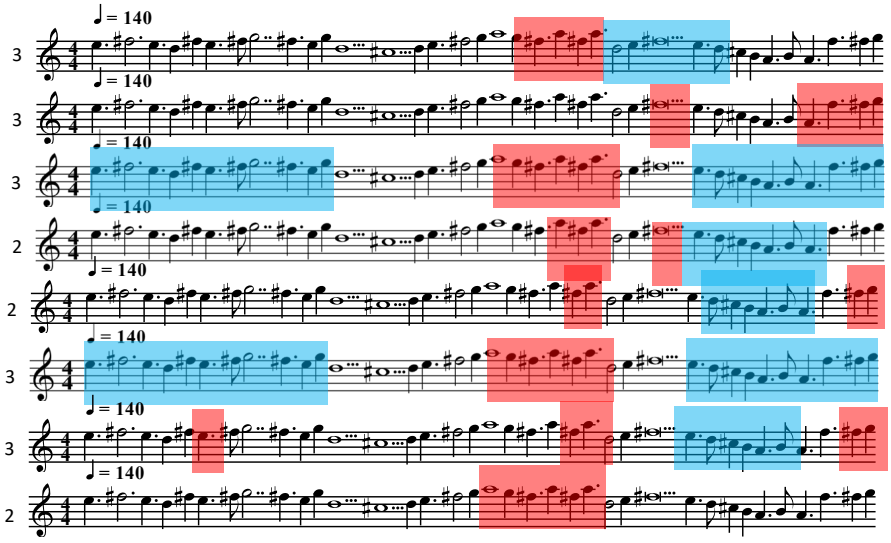


図 42　評価対象旋律 2 の主観評価

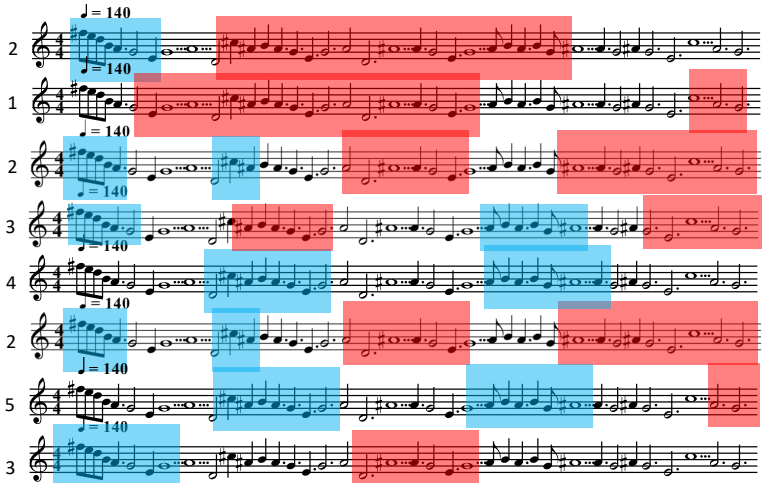


図 43　評価対象旋律 3 の主観評価

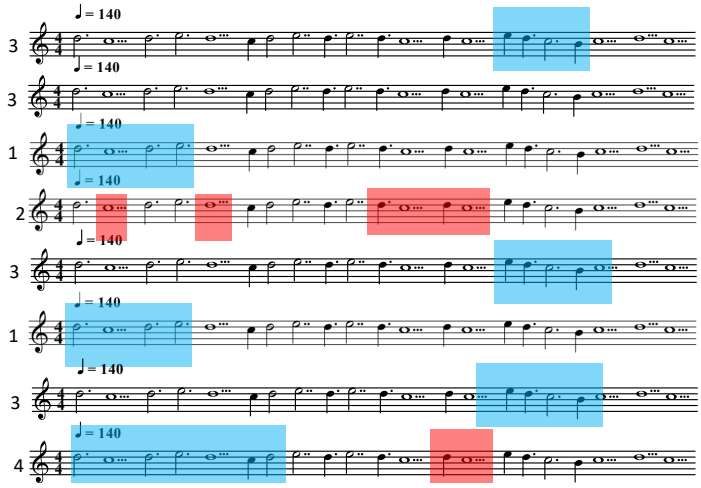


図 44　評価対象旋律 4 の主観評価

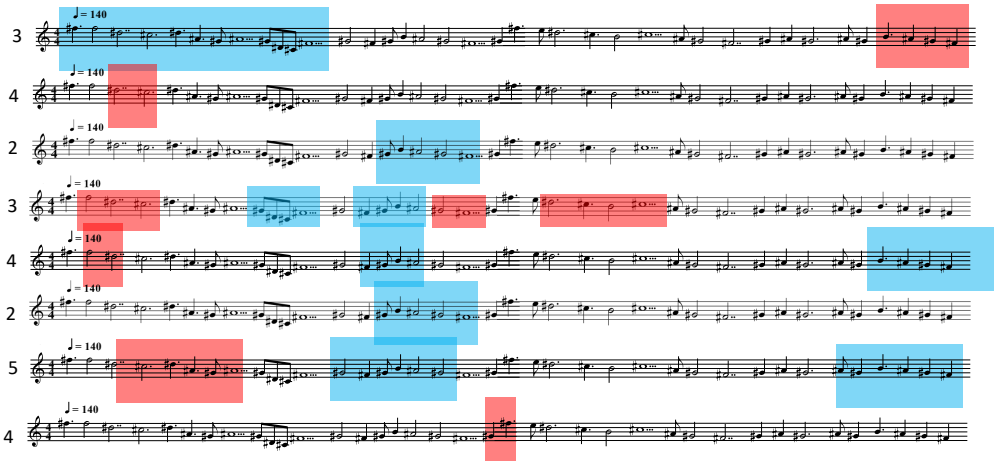


図 45　評価対象旋律 5 の主観評価

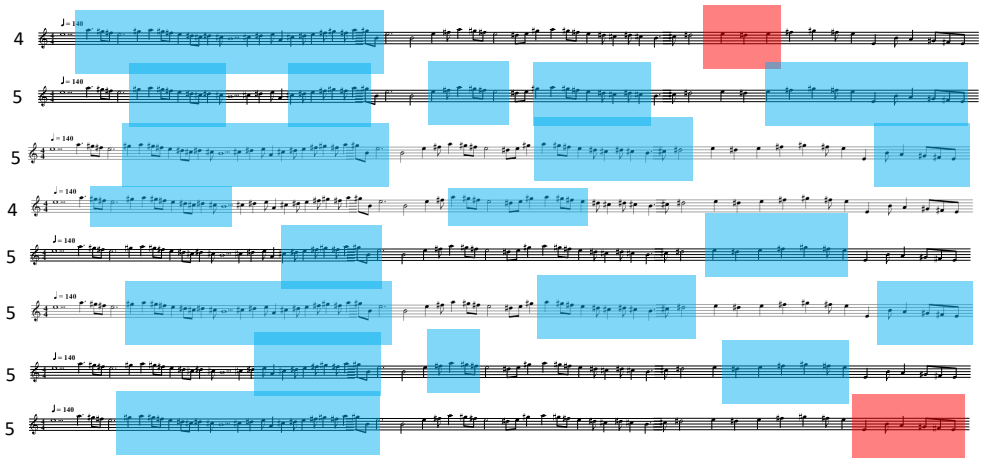


図 46　評価対象旋律 6 の主観評価

7　考察

本章では，生成された旋律に対する主観評価と区間評価から，生成された旋律の表現力に関する考察を行う．調性の推定とコード進行に関する考察は，筆者が対象となる区間に対して手動でコード進行の付与を行った[*27]．第 7.1 節では評価対象旋律全体に関する傾向を述べ，第 7.2 以降では各旋律個別の考察を述べる．

7.1　全体の傾向

この節では，各旋律に対し本研究で重視する表現力 (第 1.2 節) との比較を述べる．重視する表現力との比較を以下に示す．

旋律内で一定の音階である．またはある地点で転調する

旋律 3 を除く旋律が音階の推定が可能であり，旋律 1 では明確な音階の変化が見られた．このことから，旋律 3 を除きこの表現力を持つ旋律が生成されたと考えられる．

同じパターンが 3 回以上連続しない

項目 2 に対し，旋律 4 を除く旋律が複数のパターンを持つ旋律となった．旋律 6 は，同様のパターンの繰り返し構造が見られるが，繰り返しが 1 回であることと，主題に対する再現部とも呼べる構造を持つことから，極めて表現力が高いと考えられる．旋律 4 は旋律構成音が 4 音に限定され，音階の動きも一定のパターンに限定されたためこの項目に対する表現力を持たない．

変化に富んだコード進行が付与可能である

旋律 4 は旋律構成音が 4 音に限定されたことから，付与対象となるコード進行のバリエーションが少ないことから，この項目に対する表現力を持たないと考えられる．

評価対象の旋律に出現した本研究で重視する表現力と，特筆すべき和声法の項目を表 10 に示す．特筆すべき和声法の項目は，「カデンツ」は和音進行にカデンツが

[*27] 機械的な小節の区切りによるコード進行の付与を避け，旋律の塊に合わせたコード進行を付与するため，手動で行った．

表 10　評価対象の旋律に出現した表現力

	旋律 1	旋律 2	旋律 3	旋律 4	旋律 5	旋律 6
項目 1	o	o	−	o	o	o
項目 2	o	o	o	−	o	o
項目 3	o	o	o	−	o	o
カデンツ	−	o	o	−	o	o
終止形	−	−	o	−	o	o
楽曲構造	−	−	−	−	−	o

存在する場合，「終止形」は旋律末尾の和音進行が終止形で表現されている場合，「楽曲構造」は旋律が優れた楽曲構造を持つ場合を示す．これらの詳細は，各旋律に対応する第 7.2 節以降で述べる．

7.2　旋律 1

旋律 1 は変ホ長調の旋律であり，途中から変ホ短調に転調し，最終的に変ニ長調になる旋律であり，音階の変化に富んだ出力となっている．補正前後の旋律画像を図 47 に示す．旋律画像の着色部が補正後の旋律である．

音階が変ホ長調から変ホ短調へ変化する前は，レの音は通常であるのに対し，音階が変化した直後にはレに♭[*28]が付く変化がみられる．変ホ長調のレには♭がつかないため，明確に音階が変化したと考えられる．旋律 1 の音階の変化を図 53 に示す．旋律 1 のこの特徴は補正前の画像から存在し，補正は比較的画像濃度による音高推定値 G_{eval} が強く評価されたものと考えられる．補正前の画像から転調部を拡大したものを図 55 に示す．この転調部は，区間評価で図 54 に示されるように高評価を示す青箱が被験者 B を除いて設置された．このことからも，転調部の表現力が高いと言える．

上記のことから，本研究の表現力として定義された「旋律内で一定の音階である．またはある地点で転調する．」において，旋律 1 は表現力のある旋律と言える．曲全

[*28] 楽譜はハ長調で表記されており臨時記号をすべて♯で表現しているため，レ♭は譜面中のド♯に相当する

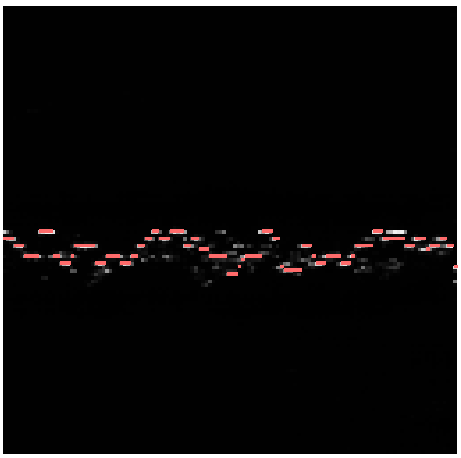


図 47　補正前後の旋律画像-旋律 1

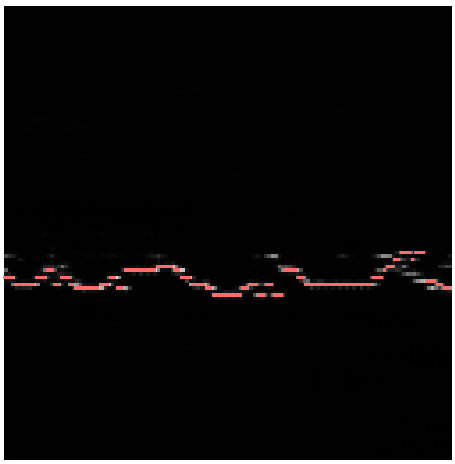


図 48　補正前後の旋律画像-旋律 2

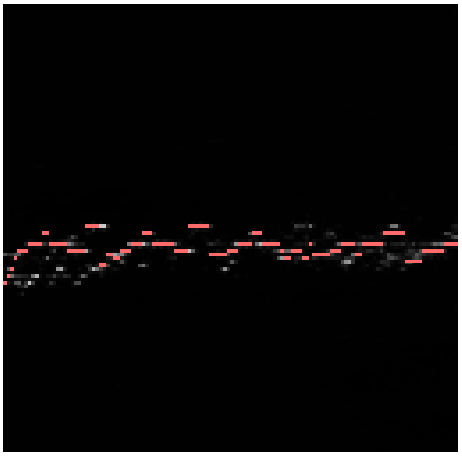


図 49　補正前後の旋律画像-旋律 3

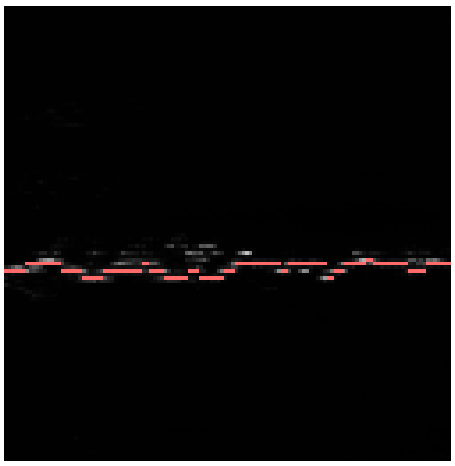


図 50　補正前後の旋律画像-旋律 4

体を通して明確な転調がみられ，音階の推定が困難である区間が殆どないことからか，5 段階評価では平均 3.50 点となった．

7.3　旋律 2

旋律 2 はニ長調の旋律である．補正前後の旋律画像を図 48 に示す．旋律前半はトニックである D からサブドミナントの G，ドミナントの G，トニックの D とカデン

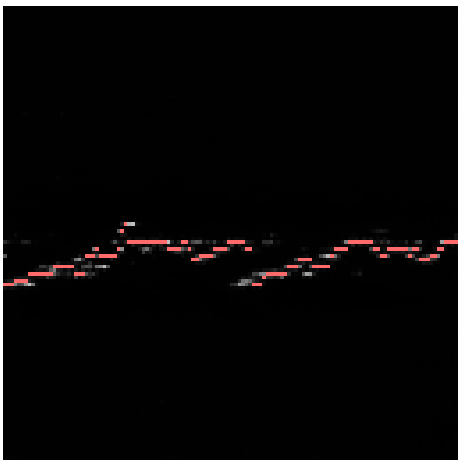


図 51　補正前後の旋律画像-旋律 5

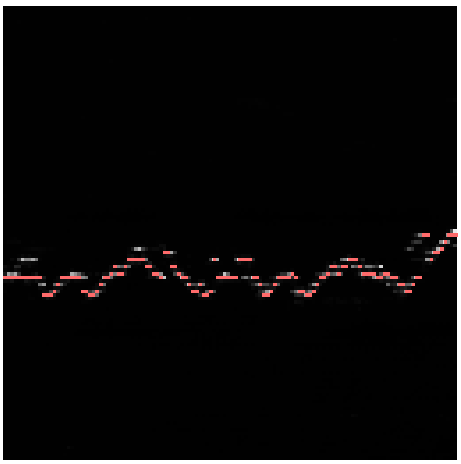


図 52　補正前後の旋律画像-旋律 6

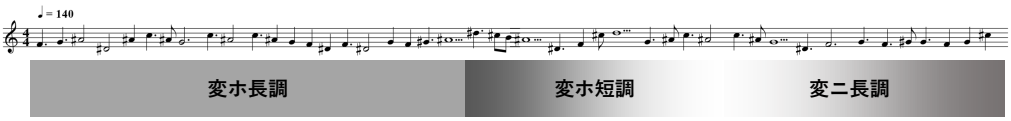


図 53　旋律 1 の音階の変化

ツ第二型に当てはまり，非常に強く安定的な進行と言える．一方で，普遍的なコード進行で全体的に単調であることから，主観評価の 5 段階評価では平均で 2.63 と低い値になったと考えられる．区間評価で被験者 B 以外が低評価を付けた区間 (図 57) は，音階内で旋律構成音が構成されているものの，変化が乏しく同じパターンの連続になったことから低評価になったと考えられる．

5 名の被験者が高評価を付けた区間 (図 58) は，トニックである D からドミナントの A，サブドミナントである G への流れとなっている．トニックからドミナントへ繋がる区間の旋律はすべての非和声音が経過音で構成されており，自然な流れとなっている．一方で，旋律の終端である区間がサブドミナントになっており，3 名の被験者が低評価となった．プロトタイプシステムでは終端に関する補正を行っていなかったため，サブドミナントの音が唐突に発生し終了する事から全体の評価が下がったと考えられる．ポピュラー音楽ではサブドミナントで終了する楽曲も存在す

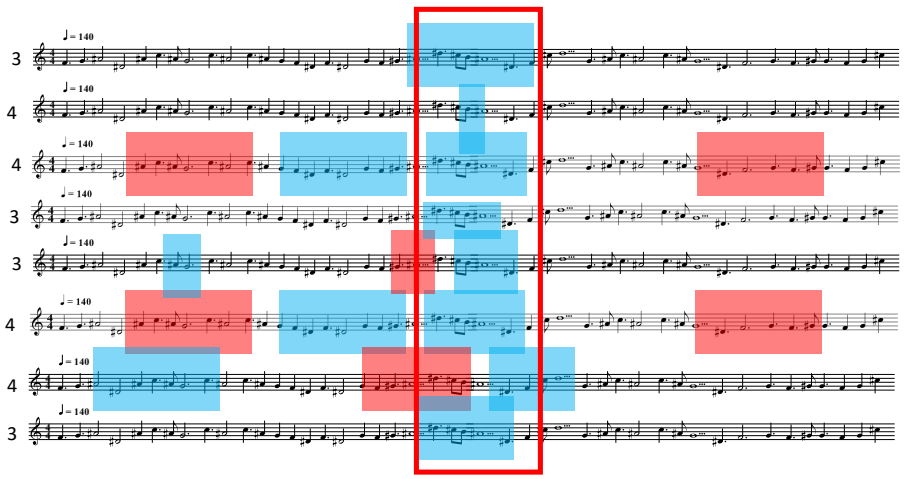


図 54　変ホ短調への転調部

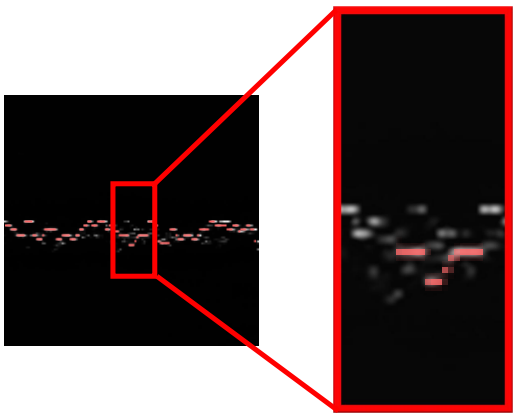


図 55　変ホ短調への転調部-補正前画像との比較

ることから，最終音をロングノーツにするなどの対策をすることでこの違和感を和らげることが可能である．

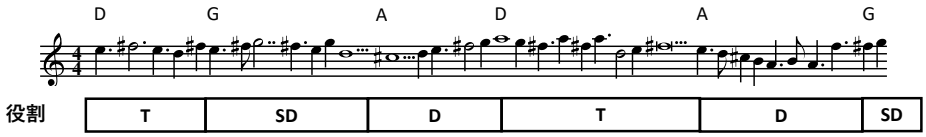


図 56　旋律 2 のコード進行

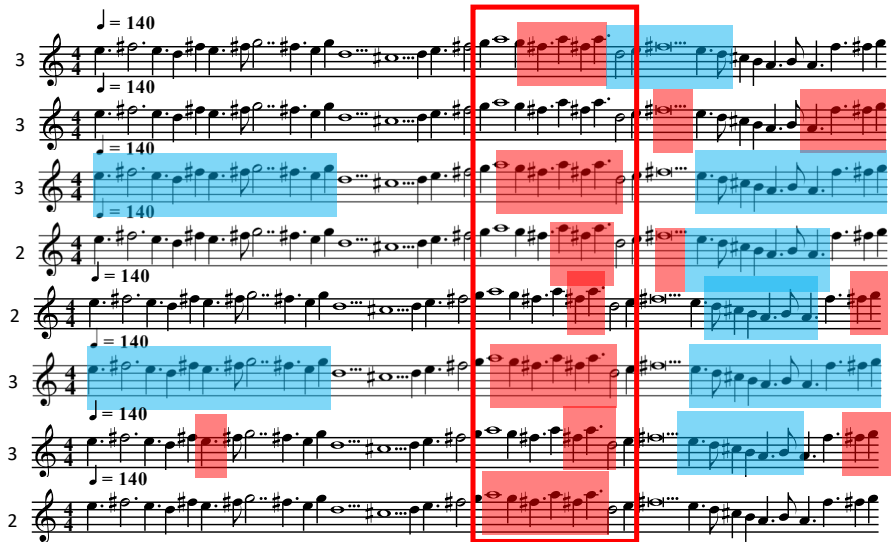


図 57　旋律 2 の低評価区間

7.4　旋律 3

旋律 3 は音階の特定が困難である旋律である*29．補正前後の旋律画像を図 49 に示す．音階が定まらないことからか，5 段階評価では 2.75 と低い評価となっている．旋律冒頭の区間が，コードネーム D に対し旋律構成音中の非和声音がすべて経過音で構成されており自然な流れであることから，区間評価にて 5 名の被験者が高評価としている．冒頭以外の区間は音階が定まらない事から，楽器演奏経験者である被験者 A から C は残りの区間全体を低評価としており，5 段階評価も 1 から 2 と低い．一方で，被験者 E と G はこの区間を高評価とし，5 段階評価が 4 から 5 と高い．

*29 D の和音から始まり比較的ホ短調かト短調に寄った旋律である．

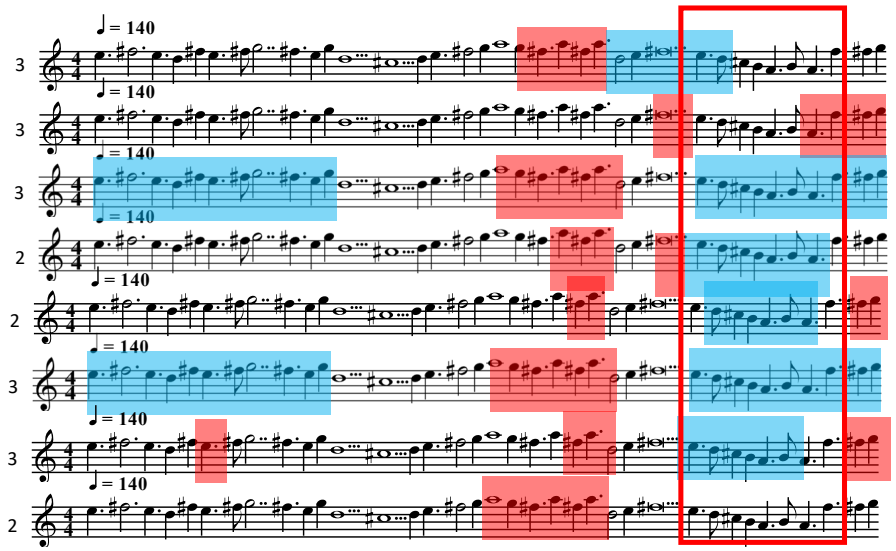


図 58　旋律 2 の高評価区間

この区間の和音は Edim7 であると推定され，ポピュラー音楽では比較的出現する和音である．このことから，好みが分かれた結果だと考えられる．

7.5　旋律 4

旋律 4 はハ長調の旋律である．補正前後の旋律画像を図 50 に示す．補正前の生成画像を含め旋律全体を通して音高の変化に乏しく，旋律は 4 種類の音高のみで構成される．被験者 D は区間評価において，旋律中最も長い音に対し悪いの評価を付けており，5 段階評価では平均 2.50 と 6 本の旋律の中で最低の評価となった．

一方で，この旋律は音階の狂いが発生することなく，すべての区間で音階内の音高が選択されている．図 59 で拡大した区間において，比較的濃度が薄い部分が補正結果として出現している．横軸 52 ピクセル前後では最大濃度 111 の音高が選択されず，赤丸付近の濃度 12 の音高が選択された．このことから，音高遷移確率による補正が有効的に行われた事例と言えるが，補正前の画像の旋律概形を無視した補正となっており，音高遷移確率推定値の重みである P_{weight} の調性が必要である．ただし，旋律 4 は補正前の旋律画像から音高の変化に乏しいことから，生成画像の影響が強く補正の影響は少ないと考えられる．

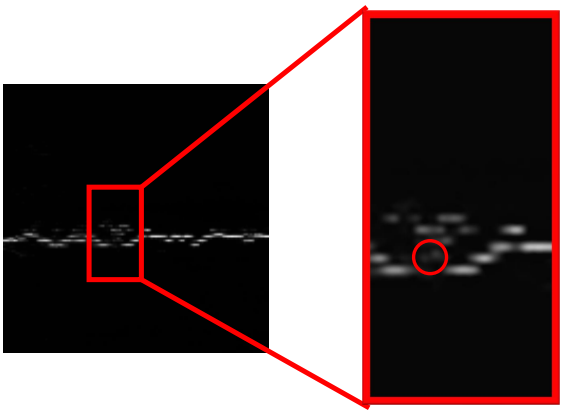


図 59　音高遷移確率による補正が有効的に行われている例

7.6　旋律 5

旋律 5 は変ト長調の旋律である．補正前後の旋律画像を図 51 に示す．区間評価にて被験者 5 名が高評価とした区間 (図 60) は，和声構成が G ♭ sus4/D ♭から G ♭/D ♭に繋がり，その後サブドミナントである B/E ♭に繋がるため，表現力が高いと考えられる．旋律冒頭の区間は音階構成音から構成されているが，E ♭ m のコードに類似した旋律構成で，変ト短調の旋律の開始和音としては一般的でないことから 4 名の被験者が低評価にしたと考えられる．被験者 G は総合評価で 5 を付けたものの，この区間に低評価を付けている．一方で，メジャーではないものの変ト短調における E ♭ m はトニックに当たるため，考えられない構成ではない．区間評価でもトランペット演奏歴のある被験者 A はこの区間を旋律中唯一の高評価としている．一般的に変ト長調の旋律では，I の和音である G ♭を頭のコードに持ってくることが多いことから，この事例は「変化に富んだコード進行が付与可能である.」という点において表現力があると言える．

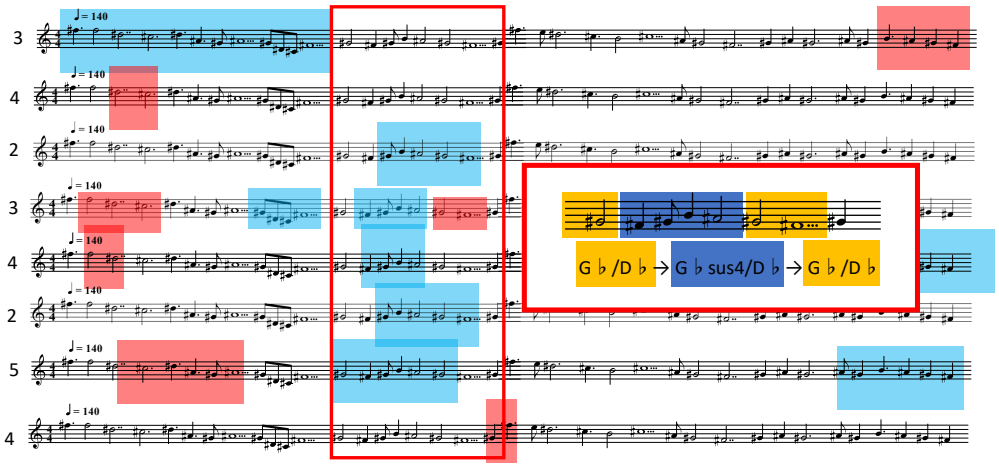


図 60　コード進行の表現力が高いと考えられる区間

7.7　旋律 6

旋律 6 はホ長調の旋律である．補正前後の旋律画像を図 52 に示す．この旋律には主題と呼べるパターンが存在し，「主題部 → 再現部 → 終端部」に近い形式となっている．主題部にはパターン A とパターン B が存在し，パターン A は同様の形状の旋律で構成され，パターン B は音高が上昇する形状の旋律で構成される．パターン A の形状比較を図 62 に示す．また，2 回目の出現である再現部ではそれぞれのパターンにアレンジを加えたような旋律となっている．さらに，終端部がオクターブを下げた音高により出現しており，非常に表現力のある旋律と言える．旋律 6 に対しパターン分類をした状態の譜面を図 61 に示す．これは，音価の圧縮により長い区間の旋律を学習させた効果であると考えられる．

同パターンでは同じコード進行が成り立ち，終端部では「ドミナント → トニック」の終止形が成立していることから，コード進行に一貫性がある．旋律の上下動が整っていることから，ベースラインが連想しやすくさまざまなコード進行の付与も可能であると考えられる．これらの事象から，この事例は「変化に富んだコード進行が付与可能である.」という点において非常に表現力があると言える．5 段階評価では実験対象となる旋律中最高である平均 4.75 となっており，人間の主観評価からも

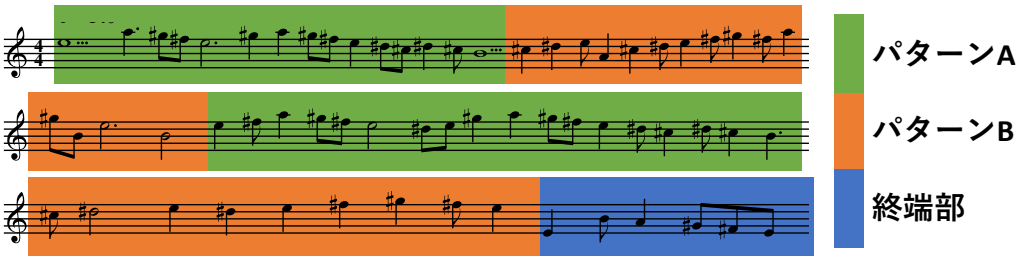


図 61　旋律 6 のパターン分類

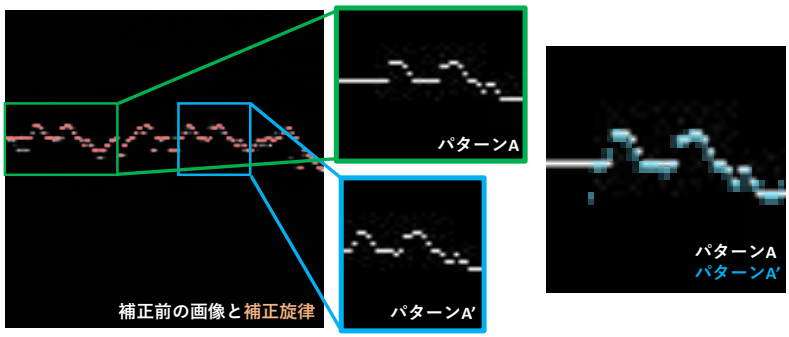


図 62　パターン A とその再現部であるパターン A’の形状比較

非常に表現力のある旋律と言える.

8　結論

8.1　結論

本研究では，敵対的生成ネットワークにより生成された旋律画像に対し，画像濃度と音高遷移確率の観点から補正を行う手法の提案を行った．

本手法を実装したプロトタイプシステムにより行った生成旋律の 5 段階の主観評価実験では，6 曲中 3 曲の旋律が評価の平均点が，評価尺度の中央値である 3 を上回る評価となり，内 1 曲は平均で 4.75 点と非常に良い旋律と評価された．この旋律は整った楽曲構造を持ち表現力があると考えられた．一方で，3 曲は 8 名の平均点が中央値である 3 点を下回り，低評価となった．音階が定まらない旋律や，音高の種類が少ない旋律がこれに相当する．音高の種類が少ない旋律は，音高遷移確率による推定値の影響が強くみられ，補正前の旋律画像にある旋律概形を無視した補正を行っていた．提案手法による旋律画像の補正は，形状と音高の面から生成画像の特徴を反映しており，表現力の高い旋律は生成画像の段階からその要素が読み取れるのに対し，生成画像の段階で音高の動きに乏しい旋律では主観評価で低い評価となり，表現力が乏しい結果となった．

生成旋律に対し行った区間評価実験では，高評価が集中した区間にはコード進行のバリエーションが豊富であることから表現力の高い区間が見られた．生成旋律に対し付与することが可能であるコード進行は，カデンツや終止形の構造を持ち，和声法に比較的合った旋律であることが確認された．一方で，対応する和音が dim7 コードと推定される区間の旋律は，和声のない旋律単体で見ると不協和音にも聞こえることから，低評価となった事例も見られた．また，サブドミナントで終了する旋律も確認され，変化に富んだ和声が付与できることが確認できた．

8.2　今後の課題

8.2.1　手法の妥当性の検証

本研究の提案手法は，「音価を無視した旋律の画像化」，「GAN による旋律画像の生成」，「濃度情報に着目した旋律画像の補正」，「2 音間の音高遷移確率に着目した旋律画像の補正」の 4 要素で構成される．この要素と実装の妥当性について，検証を

行う必要がある．本研究では，表現力に着目した提案手法の主観評価を行い，旋律 6 が特徴的な楽曲構造を持つことから，音価を無視したことでより長い区間の旋律中の音高情報から反映されたものであると分析した．また，「GAN による旋律画像の生成」と「濃度情報に着目した旋律画像の補正」の妥当性は，筆者の先行研究 [26] にて検証を行った．

今後，「音価を無視した旋律の画像化」と「学習対象旋律内の音高遷移確率に着目した旋律画像の補正」の 2 要素について，さらなる検証を行う必要がある．「音価を無視した旋律の画像化」の検証には，本研究のプロトタイプシステムで学習対象とした旋律群と，その旋律群を小節単位に加工した旋律群の 2 グループによる比較評価を行うことで妥当性を示すことができると考えられる．同様に，「2 音間の音高遷移確率に着目した旋律画像の補正」の検証には，提案手法の確率データベースを 3 音間に拡張した音高遷移確率データベースに変更した補正と，確率モデルの一種である隠れマルコフモデルによる旋律生成手法との比較評価を行う．

8.2.2　他手法との比較

本研究では，主観評価として提案手法の基本的な性能を調査した．今後，提案手法による出力旋律と RNN による手法や人間による手法による出力旋律の間で，表現力に着目した比較評価実験を行う必要がある．音楽生成の分野では，それぞれの研究で独自の尺度で手法の性能の評価を行うことが多いが，比較評価実験では本研究で重要視した表現力基準の他に，生成した旋律を定量的に評価することのできる基準が必要である．表現力を定量的に評価するための指標を検討し，実装後に比較評価を行う．

8.2.3　音高遷移確率データベースの問題

本研究で作成した音高遷移確率データベースは，ハ長調における音高の遷移をベースとしたものとなっている．一方で，出力旋律はハ長調に限定されず，さまざまな音階の出力になっている．これは，t_0 における濃度が最も高い音高を基準とした補正になっていることが原因と考えられる．ハ長調に正規化された旋律の音高遷移は，転調や借用和音などの要因で音階外の音である黒鍵の間でも行われるため，音高確率データベースには記録されるものの，母数が大幅に少ないことから正確なデータと言えない．補正がある程度進んだ段階で音階推定を行い，音高遷移確率データベー

スをその音階に合わせたものに変化させる事でこの問題が解消されると考えられる．音高遷移確率データベースは MIDI ノートナンバーを基準とした形式のため，音階にあわせて MIDI ノートナンバーをずらすことで対応することができる．

8.2.4　色情報の活用

本研究では，画像を生成する汎用的な GAN を利用した旋律画像生成を行った．汎用的な GAN であるため，グレースケール以外の色を利用する事が可能である．色に対応させた GAN 関係の研究として，Isola ら [4] が提案した pix2pix が挙げられる．pix2pix は，色に応じて対応する要素を認識し，入力画像内の色情報から要素に応じた画像の生成を行うネットワークである．pix2pix の入力と対応する生データ，生成画像を図 63 に示す．本研究ではコード進行を色で表現する事や，転調部の背景に色を付けた状態で学習をするなどが応用法として考えられる．転調部の背景に色を付け学習させる事で，ハ長調に正規化した状態でも転調した区間を区別することができるため，評価実験の対象旋律 1 のような旋律の生成がより発生しやすくなると考えられる．

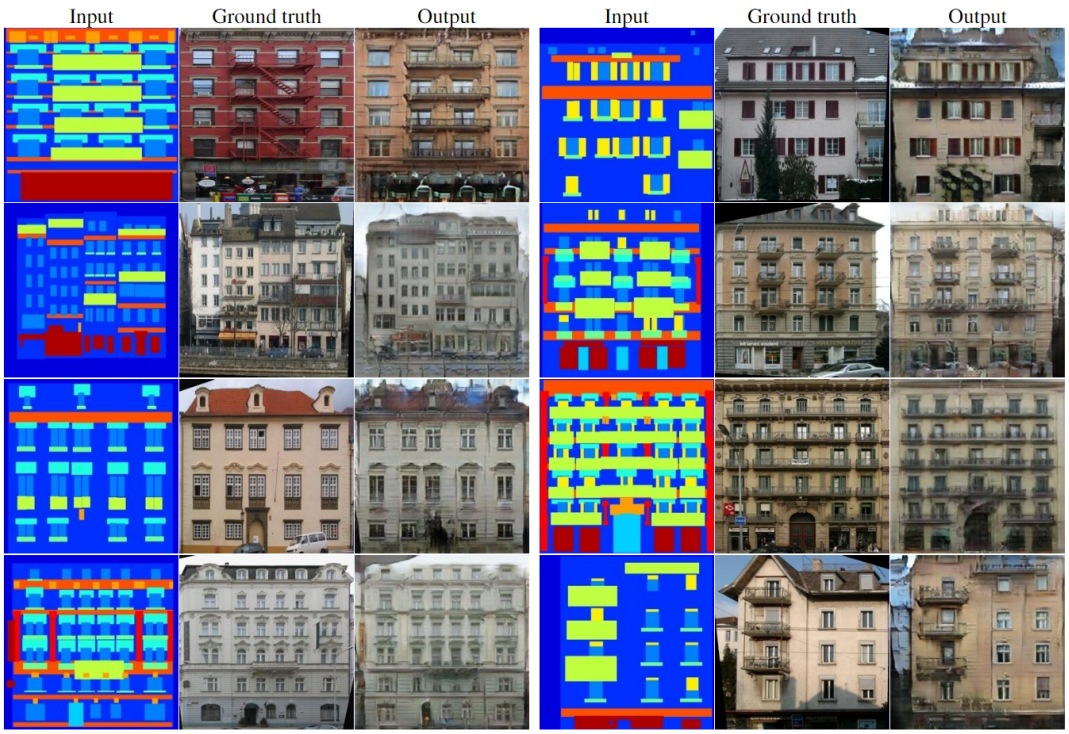


図 63　pix2pix による入力画像 (Input) と対応する生データ (Ground truth) と生成画像 (Output)[4]

参考文献

[1] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio: Generative Adversarial Nets, Advances in Neural Information Processing Systems, pp.2672-2680 , 2014.

[2] Alec Radford, Luke Metz, Soumith Chintala: Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, 4th International Conference on Learning Representations, 2016.

[3] Aayush Bansal, Shugao Ma, Deva Ramanan, Yaser Sheikh: Recycle-GAN: Unsupervised Video Retargeting, European conference on Computer Vision, pp.119–135 , 2018.

[4] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, Alexei A. Efros: Image-To-Image Translation With Conditional Adversarial Networks, IEEE Conference on Computer Vision and Pattern Recognition, pp.1125–1134 , 2017.

[5] Ping-Sung Cheng, Chieh-Ying Lai, Chun-Chieh Chang, Shu-Fen Chiou, Yu-Chieh Yang: A Variant Model of TGAN for Music Generation, 2020 Asia Service Sciences and Software Engineering Conference, pp.40–45 , 2020.

[6] 池内友次郎，島岡譲ほか: 和声 理論と実習 I，音楽之友社， 1964.

[7] 北野宏明: 遺伝的アルゴリズム，産業図書， 1993.

[8] 今井繁，長尾智晴: 遺伝的アルゴリズムを用いた自動作曲，電子情報通信学会技術研究報告， Vol. 98， No. 58，pp. 59–66， 1998.

[9] 山田拓史，椎塚久雄: 遺伝的アルゴリズムを用いた自動作曲について，情報処理学会研究報告音楽情報科学（MUS)， Vol. 1998， No. 96， pp. 7–14，1998.

[10] 羽鳥喜紀，涌井広大，長名優子: 遺伝的アルゴリズムを用いたコード進行を考慮した自動作曲，第 78 回全国大会講演論文集， Vol. 2016， No. 1， pp. 451–452， 2016.

[11] Google AI: Magenta ，https://magenta.tensorflow.org/ (最終アクセ

ス:2022.01.10)

[12] 深山覚，中妻啓，米林裕一郎，酒向慎司，西本卓也，小野順貴，嵯峨山 茂樹: Orpheus：歌詞の韻律に基づいた自動作曲システム，情報処理学会研究報告音楽情報科学（MUS)， Vol．30， No．78， pp．179–184， 2008.

[13] 田中健， 外山史，東海林健二: 遺伝的アルゴリズムを用いたメロディー進行とリズムの組合わせによる自動作曲，情報処理学会研究報告音楽情報科学（MUS)，Vol41，pp.43–48，2001.

[14] Olof Mogren: C-RNN-GAN: A continuous recurrent neural network with adversarial training，Constructive Machine Learning Workshop (CML) at NIPS 2016，2016.

[15] Martin Arjovsky ， Soumith Chintala ，Léon Bottou: Wasserstein Generative Adversarial Networks，34th International Conference on Machine Learning，Vol70，pp.214–223，2017.

[16] Li-Chia Yang，Szu-Yu Chou，Yi-Hsuan Yang: MidiNet: A Convolutional Generative Adversarial Network for Symbolic-domain Music Generation, 18th International Society for Music Information Retrieval Conference, pp.324–331 ，2017.

[17] Dong Hao-Wen ， Hsiao Wen-Yi ， Yang Li-Chia ， Yang Yi-Hsuan: MuseGAN: Multi-Track Sequential Generative Adversarial Networks for Symbolic Music Generation and Accompaniment，Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence ，No.5，pp.34–41，2018.

[18] 土屋裕一, 北原鉄朗: 音符を単位としない旋律編集のための旋律概形抽出手法, 情報処理学会論文誌, Vol4, No54, pp.1302–1307 ，2013.

[19] 東山恵祐: 作曲支援システムにおけるコード進行及びキーの決定方法，情報処理学会研究報告音楽情報科学（MUS)， Vol．2014， No．52， pp．1–6，2014.

[20] Ian Simon，Dan Morris and Sumit Basu: MySong: Automatic Accompaniment Generation for Vocal Melodies，Proceedings of the SIGCHI Conference on Human Factors in Computing systems，CHI ’ 08，New York，

NY，USA，ACM， pp．725–734， 2008.
[21] 森篤史，新井イスマイル: 2 階マルコフ過程を用いた HMM によるコード付与手法の提案，情報処理学会研究報告音楽情報科学（MUS)， Vol．2017， No．18， pp．1–9， 2017.
[22] 川上隆，中井満，下平博，嵯峨山茂樹: 隠れマルコフモデルを用いた旋律への自動和声付け，情報処理学会研究報告音楽情報科学（MUS)， Vol．19， No．34 ， pp．59–66， 2000.
[23] 西田正洋，菊地進一，中西正和: リカレントニューラルネットワークを用いたコード進行の自動生成，情報処理学会研究報告音楽情報科学（MUS)， Vol．2000， No．76， pp．67–72， 2000.
[24] 三浦雅展，青山容子，谷口光，青井昭博，尾花充，柳田益造: ポップス系の旋律に対する和声付与システム: AMOR，情報処理学会論文誌， Vol．46， No．5， pp．1176–1187， 2005.
[25] 鈴木大河，丸山 一貴: 遺伝的アルゴリズムによる旋律への和声付与, 情報処理学会 インタラクション 2020，pp.523–527 ，2020.
[26] 鈴木大河，横山真男: 汎用敵対的生成ネットワークにより生成された旋律画像の補正，情報処理学会研究報告音楽情報科学（MUS)，Vol.2021-MUS-131，no.43，pp1–4，2021.
[27] OpenCV team: Home - OpenCV，https://opencv.org/ (最終アクセス:2022.01.10)
[28] Ishaan Gulrajani, Faruk Ahmed, Martin Arjovsky, Vincent Dumoulin, Aaron C. Courville: Improved Training of Wasserstein GANs, 31st International Conference on Neural Information Processing Systems, pp.5769-5779 ，2017.
[29] Colin Raffel，Daniel P.W.Ellis: Intuitive Analysis, Creation and Manipulation of MIDI Data with pretty_midi．15th International Conference on Music Information Retrieval Late Breaking and Demo Papers ，2014.
[30] The ABC Music Project: The Nottingham Music Database，http://abc.sourceforge.net/NMD/ (最終アクセス:2022.01.06).
[31] ImageMagick Studio LLC: ImageMagick Convert, Edit, or Compose Digital Images，https://imagemagick.org/index.php (最終アクセス:2022.01.09)

[32] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek G. Murray, Benoit Steiner, Paul Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, Xiaoqiang Zheng: TensorFlow: A System for Large-Scale Machine Learning, 12th USENIX Symposium on Operating Systems Design and Implementation , pp. 265–283, 2016.

謝辞

明星大学横山真男教授には，本研究に関して多くのアイディアと音楽理論に関する助言とご指導をしていただきました．ここに感謝の意を示します．明星大学丸山一貴准教授，植木一也准教授には，本研究に関して多くのアイディアとご助言をいただきました．ここに感謝の意を示します．また，学内中間発表会で有意義な議論をさせていただいた明星大学長慎也教授，和田康孝准教授にも感謝の意を示します．最後に，本研究において積極的な議論とアイディアを頂いた明星大学 2020 年度ならびに 2021 年度横山研究室，丸山研究室の皆様にも感謝いたします．