

地図標高データを用いたメロディ生成の試み

梅村 祥之^{1,a)} 伊達 彩斗¹

概要：国土地理院の地図情報から山の形状を表す標高データのカーブを得て、メロディラインを生成するシステムを開発した。そのようにして大量に作成したメロディの中に良いフレーズが何箇所か存在する。それらを人が選定して、つなぎ合わせることで曲にまとめた。このようにして作成された曲2曲と、プロの作曲家によって作曲された既存曲3曲の計5曲について、26名の評価者に曲の良さを5段階評価してもらった。その結果、本法による生成曲の良さは既存曲と同程度であることが確認された。

キーワード：自動作曲，作曲支援，旋律，主観評価

A trial to generate melodies using altitude data of mountains

YOSHIYUKI UMEMURA^{1,a)} AYATO DATE¹

Abstract: We have developed a system that generates melody lines using curves of altitude data representing the shape of mountains provided by Geographic Information System of the Geographical Survey Institute. There are some good phrases in many melodies created by the system. An operator chose them and combined them together so that the tune becomes a good music. Twenty-six participants evaluated the preferences of the songs by five categories. The songs include two songs created in this way and three songs composed by professional composers. As a result, it was confirmed that the preferences of the generated songs by our method is about the same as the existing songs.

Keywords: automatic composition, assistance of composition, melody, subjective evaluation

1. はじめに

これまで、自動作曲技術の研究開発の中で、自然界に存在する信号や、自然界の形状を模倣した生成モデルによる信号を基にして曲を生成する方法や声のような人の発する情報を基に曲を生成する方法などが研究開発されている。自然界の形状を模倣した生成モデルによる生成法として、フラクタル図形によって生成された樹木の形状や雪の結晶の形状を基に曲を生成する様々な方法が研究されている [1]。我々も、人と人の音声対話における音声の韻律情報からメロディラインを生成する研究を行った [5]。対話音声素材とした理由は、文献 [2] の1頁記載の「音楽は、コ

ミュニケーションの基本的チャネルの一つである。それは人々が共有することの出来る感情、意図、意味の手段を与える。」との説に基いて、音声対話データの中の韻律情報に、感情を伝達する成分が含まれると考えたからである。

次に、先行研究の問題点と、それに対する本研究の狙いを述べる。自然界ないし人間の発する信号を基に楽曲生成を行う場合、楽曲を大量に生成させると、その中に、良い箇所が少量含まれるであろう。そのため、大量に生成させる必要がある。したがって、基にするデータを大量に手間を掛けずに入手できることが必要である。本研究において用いる地形データは、国土地理院から大量に提供されておりこの点で他の方法に比べて優れている。

元のデータを手間なく大量に入手できても、メロディラインの素材となる時系列に変換する際の手間がかかったり、誤認識などの誤りが生じると問題である。先の対話音声の韻律情報に基づく曲生成においては、音声波形データ

¹ 広島工業大学
Hiroshima Institute of Technology, Hiroshima 731-5193, Japan

^{a)} y.umemura.im@it-hiroshima.ac.jp

から音声のピッチの時系列を得る際に、背景ノイズなどによって誤検出する問題があった。カメラで得た画像から形状データを得て音楽に変換する場合にも誤認識の問題がある。それに対して、本研究では、国土地理院から提供される地形の高度データをそのまま用いるため、変換する手間と誤認識が発生する問題がない。

先行研究において、生成曲の評価を行う場合に、生成曲の良さをプロの作曲家による曲と比較して、どの程度のレベルであるのかを示した研究が少ないように思われる。本研究では、本手法による生成曲と作曲家による既存曲の良さに関する主観評価実験を行い、統計的検定と推定を行って両者の違いの程度を示す。

2. 楽曲生成

2.1 標高データ取得

国土地理院の Web での基盤地図情報ダウンロードサービス [8] により標高データを取得する。このデータは、日本の地形に関して、5m 刻みの格子点毎に標高をデータ化したものである。東西 11.3km × 南北 7.5km の区画毎にダウンロードする形式となっています。1 区画のデータの中に、東西方向 2,250 サンプルからなる標高系列が南北に 1,500 本含まれる。ここで、区画という用語は、同サイトで使われる用語であり、以下、この用語を使用する。1 例として南アルプス市と伊那市にまたがる南アルプス山脈の北岳の標高データを示す。図 1 に、その標高データに基いて 3 次元グラフィックス表示した画像を示す。山の形状は、フラクタル図形が持つ自己相似性とランダム性を有する。これをメロディライン生成の原データとして用いる。標高系列を抽出するにあたり、東西、南北等の直線で走査することもできるし、山の稜線を追跡する方法等も考えられる。ここでは、単純な方法として最南端から東西に走査し、5m 北上して東西に走査してゆく方法を採用する。ただし、隣接走査線間の距離が 5m ではカーブ形状の違いが乏しいため、今回、走査線 8 本 40m 刻みに飛び越し走査し、1 区画あたり 188 本の標高系列を得る。

2.2 高域強調

山の形状は通常、なだらかなため、標高データをそのまま使って音符系列を作成すると同音進行がいくつも続く緩慢な系列になってしまう。それを調整するために高域強調フィルタで低周波成分を除去し高周波成分を強調する。先の標高データの最南端の東西線に対して、強調具合を 3 種類変化させてフィルタを適用した結果を図 2 に示す。このパラメータを変えることで生成する曲調を調整できる。以下、「高域強調小」のパラメータで高域強調したデータを用いて処理を継続する。

2.3 音階処理及び標高から音高への変換

まず、メロディ生成において使用する音階について述べる。音階は、国別、地域別の曲の雰囲気や醸し出す重要な要素である。しかし、今回、基礎検討の段階であることから、最も代表的なダイアトニックスケールに固定する。そして、ソフトウェア内部では、データ処理を簡略化するため、ダイアトニックスケールを構成する音高に 1 ずつ増加する整数値を割り当てて表現する。処理の最終段で、半音刻みで 1 ずつ増加する整数値に変換して、一般的な演奏ソフトとのデータ受け渡しをしやすくする。

さて、標高の数値から音高に変換する方法は、単純に、切片と傾きの 2 つのパラメータを持つ 1 次変換にて変換する方法である。切片と傾きは音高が演奏に適した音域に入るように調整する。例えば、E4 を中心とした上下 2 オクターブ程度に収まるように設定する。

2.4 音長系列の生成

我々が先の行った音声のピッチの時系列からメロディラインを生成する研究の中で開発した音長系列の生成方法と同じ方法を用いる [9]。基本的な考え方として、勾配の小さいカーブに対して長い音長の音符を割り当て、勾配の大きいカーブに対して短い音長の音符を割り当てる。ここで、音符の音長として、8 分音符、4 分音符、付点 4 分音符、2 分音符、付点 2 分音符、全音符の 6 種類を使用する。割当にあたって、音符の継続時間中に小節線が来ないようにすることと、できるだけ長い音符のオンセットが小節線位置に来るようにする。これによって、ダウンビートのリズム感を持たせるようにする。今回 4/4 拍子としている。図 1 に先の系列に対して本処理を行った結果の系列を示す。同音進行の場合に、音符の区切りが見えないため、同図の最後に示したピアノロール表示を参照されたい。

2.5 コード進行に合わせた音高の調整

多くの自動作曲手法において、コード進行に基づいて、コード構成音が出現しやすくするように音高を変更する処理が行われている。本法でも、そのような処理を含める。今回、コード進行として J-POP で頻繁に使用される「王道進行」ないし「小悪魔コード進行」と呼ばれる $F \rightarrow G \rightarrow E_m \rightarrow A_m$ のコード進行を用い、1 小節毎に 1 コードを配置する。これまでに生成した音符の音高をコード進行に沿って変更する。小節の第 1 音符の場合、最寄りの内音に変更する。最寄りの内音が複数あれば、直前の音符との音程が狭い方の音符を優先する。小節の第 2 音符以降の場合、奇数番目の音符を内音にする。最寄りの内音が複数あれば、直前の音符との音程が狭い方の音符を優先する。偶数番目の音符は変更しない。その理由は、奇数番目の音符を内音としているので、偶数番目の音符は、内音と内音をつなぐ経過音の役割を持つ音符となることを期待した。

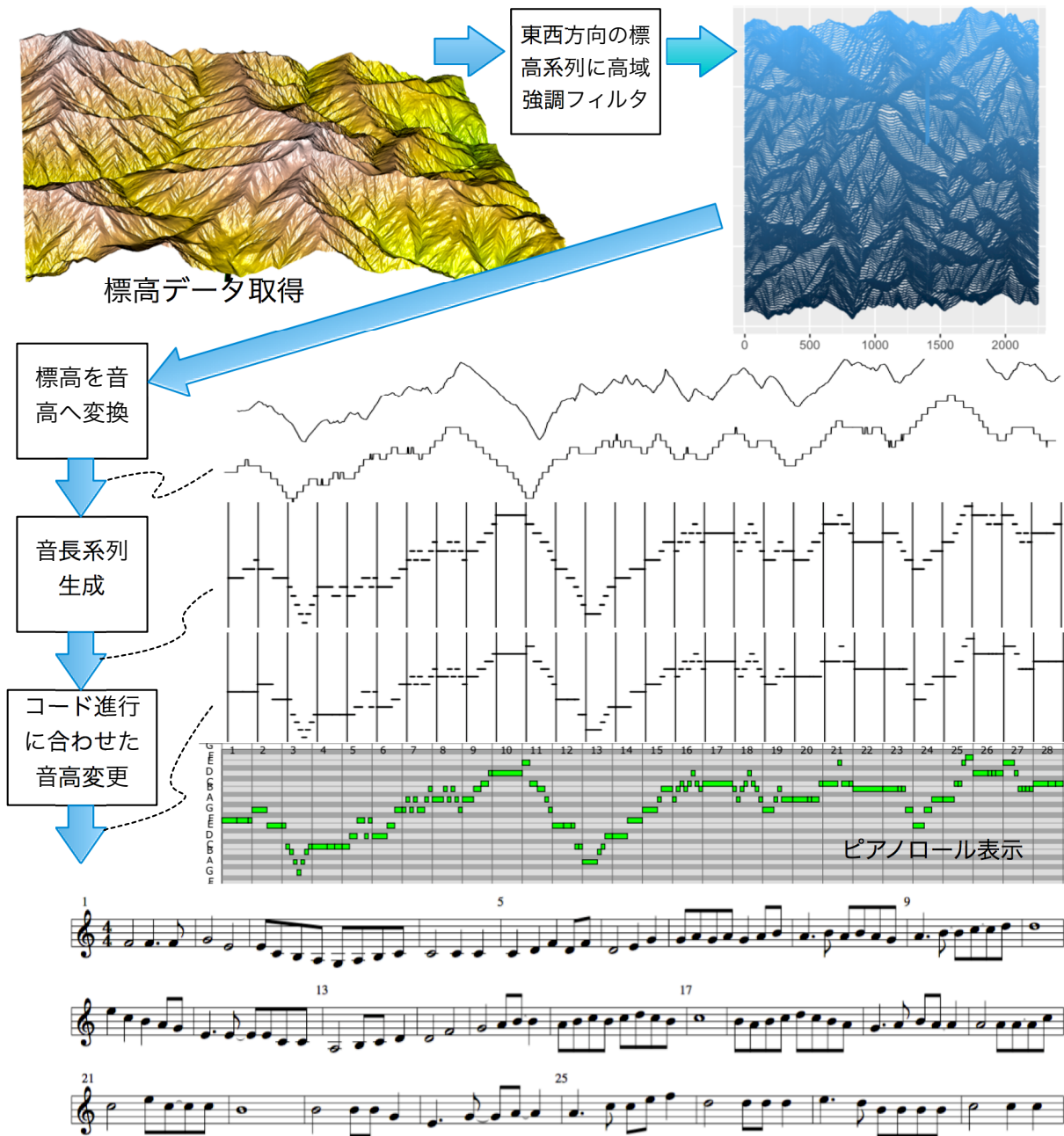


図 1 楽曲生成のブロック図

Fig. 1 The block diagram of our music generation system.

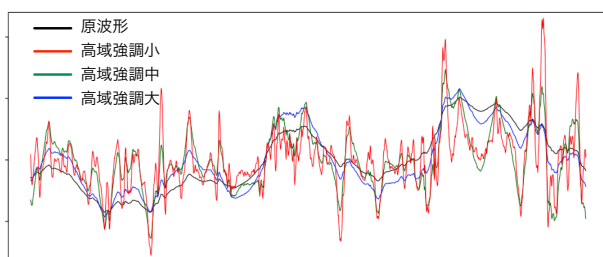


図 2 高域強調の違いによる標高カーブの変化

Fig. 2 Attitude curves emphasized by high-frequency emphasis filters.

今回採用した方式の他に、メロディラインの音高を更に内音に近づけることも可能であるし、逆に、内音に変更する場合の条件を緩和して、コードと協和しない音符を増やすことも可能である。前者の場合、全体になめらかに進行する音楽となり、後者の場合、ところどころの音符で、悪く言えば、違和感を持つ音、良く言えば新鮮で意外性を持つ音が出現する。今回採用する方式は、両者の中間的な設定である。

2.6 生成結果

図 1 に先の系列に対して本処理を行った結果の系列を示

す。曲の演奏に際して、テンポを設定する必要がある。数通り聴き比べて150に設定した。この曲を演奏して試聴したところ、違和感のないメロディであるという印象を得た。標高データ1区画から188曲が生成され、全体で約1時間30分の演奏時間となる。各々の曲の中に、ところどころ良いフレーズが含まれているようであった。そこで、今回、1名の作業者によって、全188曲を試聴し、良いフレーズと感じられる部分がどの曲のどこに出現するかを大まかに調べた。場所の特定は、小節番号単位で行った。その結果、全体の約3%分が良いフレーズとして抽出された。なお、具体的なフレーズの例は、次章でそのフレーズを使った作曲例として図3に示される。同図における記号Pプラス数字が付記された小節範囲が、本法によって生成されたフレーズ素材である。

3. 生成フレーズによる作曲支援

3.1 企画

前章の方法で作曲すると、部分的に良いフレーズが含まれているものの、1曲全体が良い曲となる可能性は低い。

この技術をベースに、もし、曲全体を全自動で作曲する技術を開発しようとすれば、まず、良いフレーズを機械選定する技術が必要となる。我々の研究室で、楽曲のフレーズをコンピュータで判定する技術の開発を行っているものの、完成には至っていない[4]。今回行ったように、1時間30分の生成曲から人手で良いフレーズを選定するのは、2時間程度の作業で完了する。また、選定作業は、作曲経験のない初心者でも可能である。

そこで、良いフレーズを選定を人手で行い、作曲のためのフレーズ素材を用意しておく。これを用いて、曲全体を作曲しようとする場合、機械処理で行うことを考えると、フレーズの繰り返しや、イントロ、サビ、エンディングといった曲構成に関する知識処理および、イントロ向き、サビ向き、エンディング向きのフレーズの判定処理が必要である。これらは、必ずしも高度な技術を要しないと推測される。

ひるがえって、素材となるフレーズを使って、曲構成を人手で行うことを考えると、必ずしも作曲能力を必要とせず、初心者でも可能で、かつ、作業自体が楽しい作業である。

以上から、フレーズ選定に手間がかかるという問題はあまるものの、人手で選定したフレーズ素材を使って、初心者が作曲できる作曲支援技術を開発するのが良いと考える。

3.2 作曲方法

前章で生成し、良いフレーズとして選定したフレーズを基に、作曲経験のない作業者1名が、1分強の曲を2曲作曲する。作曲に際して、以下の条件を課す。以下の条件であれば、作曲のスキルのない初心者であっても作業ができ

ることを保証するものである。生成システムで生成されたメロディに含まれる良いフレーズをつなぎ合わせるだけで曲を作る。つなぎ合わせる部品を、以下、「断片」と称する。断片として、音符単位といった短い単位を許せば、自由度が増大して初心者向きにならないため、断片は生成システムで生成したメロディ内の小節単位の部分とする。通常は、フレーズが断片になるものの、フレーズの一部を断片に使う場合もある。以下に、条件の詳細を記載する。

- 生成システムで生成されたメロディから小節単位で切り出された断片のみを用い、それらをつなぎ合わせて曲を作る
- 小節単位の切り出しに際し、フレーズ境界と小節境界が一致しないなどの理由で、小節間の接続に違和感のある場合がある。その場合、小節の最後の音符と同じ音高の全音符を1小節追加しても良いとする
- 断片全体をオクターブ単位で上下させても良い
- 同じ断片を複数回使用しても良い

以上の基準であれば、初心者でも可能であると考え、2曲作曲する時間はおよそ6時間であった。作曲後、約2時間、作曲経験のあるアドバイザー1名（音楽教育を受けておらず、趣味として作曲を行っている人物）にアドバイスを貰い、曲構成の見直しを行った。見直しにあたって、音符の加工に関して、初心者でも作業可能のように設定した先の基準に従った。

4. 主観評価実験

4.1 素材の選定、加工方法、演奏方法

前章の方法で作曲した曲を以下、生成曲と称する。生成曲2曲の良さに関し、36名の評価者に評価してもらう。比較のために、プロの作曲家による既存曲3曲を含める。既存曲の選定は、ジャンルをBGM曲とし、TVアニメの挿入曲から2曲、ポピュラーソングのピアノ曲から1曲を選定する。これら3曲のプロフィールを表1に示す。アニメ挿入歌の「緑色の記憶」はゆったりとした曲で、高い評価点とはならないであろうと予想した曲であり、他の2曲「始まりの音色」と「Joy」は良い曲として評価されるであろう曲を選定している。

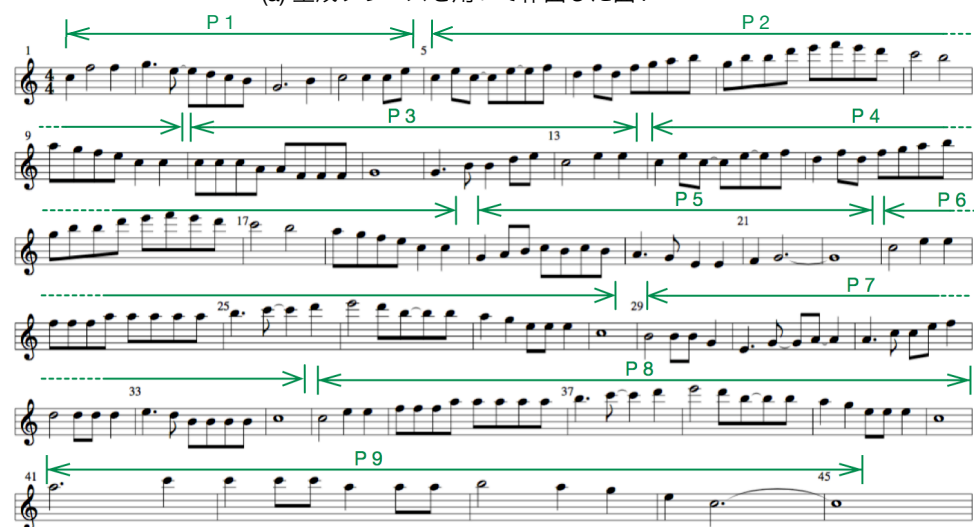
曲をそのまま使用すると時間が長すぎるため、作業者が適切だと思われる部分を切り出すことによって、1分強の曲にして、5曲それぞれの演奏時間が同じくらいになるようにする。抜き出した箇所について、表1に記載する。

生成曲は単旋律であり、それと合わせるため、既存曲も主旋律のみを単旋律で演奏する。演奏にあたって、MIDIファイルをDTMソフトであるLogic Pro Xで演奏させる。その際、楽器をSteinway Grand Pianoに統一し、ベロシティを100に統一する。

既存曲を知っている評価者もいるため、聴いたことがある曲か申告させ、聴いたことがある場合には、その評価者



(a) 生成フレーズを用いて作曲した曲1



(b) 生成フレーズを用いて作曲した曲2

図 3 生成したフレーズから作曲した曲の楽譜. 記号 P で示す小節範囲は、生成曲からの抽出単位を示す.

Fig. 3 Scores composed using generated phrases.

を不適確者として集計から除外する.

4.2 評価者

工学系大学の1年から4年までの一般の学生34名である. 18才から24才まで含まれ, 平均が20.1才であった. 男女の構成は, 女性3名男性31名であった.

4.3 曲の提示方法と評価方法

提示方法は, 5曲の提示順をランダムとし, スピーカーで1曲演奏し, 1曲毎に約10秒間のポーズをとり, その間に評価結果を用紙に記入してもらう. 全5曲提示後, 同じ提示順で, 2回めの提示を行い, 評価を見直してもらう. 評価実験全体が約20分で終了する.

評価方法は, 5段階評価とし, 「悪い」, 「やや悪い」, 「ど

ちらとも言えない」, 「やや良い」, 「良い」の5つの評価語から1つを選択する.

4.4 評価結果

既存曲3曲のうち, 1曲でも聴いたことがあると申告した評価者は8名であった. この8名を除き, 有効データが26名分となった.

「悪い」から「良い」までの5段階評価の評価語に対して, 1から5までの1刻みの評点の数値を割り当てる.

使用した5曲に対して, 生成曲2曲をまとめて扱い, 「生成曲」と称し, 既存曲のうちの「始まりの音色」と「Joy」をまとめて扱い, 「既存良」と称し, 既存曲うちの「緑色の記憶」を単独で扱い, 「既存否」と称する.

3水準(曲のグループが3種類)に対する評点の平均値

表 1 比較対象とする既存曲 3 曲のプロフィール
Table 1 Profiles of 3 reference existing tunes.

曲名	Joy
グループ名	既存良
作曲家	George Winston
収録アルバム売上	米国で 300 万枚
提示曲の小節番号	1-12, 23-41
曲名	始まりの音色
グループ名	既存良
作曲家	浜口史郎
テレビ番組挿入歌	日本テレビ系列 アニメ「花咲くいろは」 2011 年に 6 ヶ月放送
提示曲の小節番号	1-18, 21-37, 39-54
曲名	緑色の記憶
グループ名	既存否
作曲家	吉森信
テレビ番組挿入歌	TBS 系列アニメ「デュララ」, 2011 年, 2015 年, 2016 年に計 1 年 3 ヶ月放送
提示曲の小節番号	1-25

の差の検定を分散分析によって行う。各評価者が全曲を評価するため、対応のあるデータに関する分散分析を行う。使用する分散分析のソフトはデータ解析用言語 R の基本パッケージ stats で提供される関数 TukeyHSD である。

曲の良さ位に有意差があるか否かを調べる他に、有意差がない場合、曲の良さが同等であるかどうか調べる。同等性の分析にあたっては、文献 [6] の方法に従い、平均値の差の 95%信頼区間の下限が評価語の隣接カテゴリ間の距離にあたる -1 を下回ることなく、かつ、95%信頼区間の上限が +1 を上回ることがなければ同等性を有すると判断する。

分散分析の結果、2 グループの既存曲間「既存良」と「既存否」に有意差があった ($p = 0.017 < 0.05$)。生成曲と「既存良」、生成曲と「既存否」の間には有意差が見られず、同等性が成立していた。平均値の差の信頼区間の計算出力を表 2 に示す。曲のグループ毎の評点の平均値に標準偏差をエラーバーとして付加したグラフを図 4 に示す。以上の結果、生成曲の良さはプロの作曲家による曲と同等のレベルであることが示された。

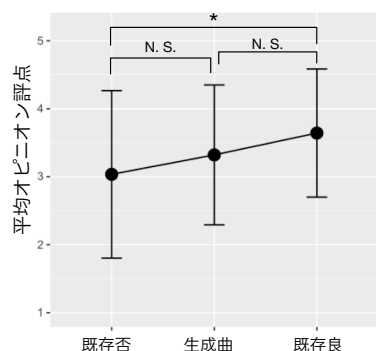


図 4 主観評価結果に対する平均オピニオン評点

Fig. 4 Mean opinion scores for the subjective evaluations.

表 2 分散分析結果, lwr は 95%信頼区間の下限値, upr は上限値を示す。

Table 2 The result of ANOVA.

	diff	lwr	upr	p adj	同等性
既存良-既存否	0.6071	0.0906	1.1236	0.0168	
生成-既存否	0.2857	-0.2308	0.8022	0.3903	あり
既存良-生成	0.3214	-0.1003	0.7432	0.1708	あり

5. まとめ

地形の標高データを基に自動作曲する技術を開発した。特徴としては、元となるデータを大量に容易に入手できること、処理のステップの中で、例えば機械認識等の過程がないため、全自動でメロディラインの生成ができることである。生成した 1 時間 30 分の曲を試聴したところ、良いフレーズが 3%程度含まれていた。それらの良いフレーズを単純に組み合わせるのみという方法で作曲未経験者が 1 分強の曲を 2 曲作曲した。プロの作曲家による既存曲 3 曲と本技術による生成曲 2 曲の曲の良さを実験参加者に 5 段階で主観評価してもらい、評価結果を分散分析した結果、生成曲は、既存曲と同等であることが確認できた。

本技術を発展させれば、作曲の初心者に対して、容易に作曲できるフレーズ素材を提供でき、有用な作曲支援技術に発展すると期待している。

謝辞 本研究に先立ち、予備的な検討を広島工業大学梅村研究室福本亮太君が 2016 年度卒業研究として行った。作曲経験のある当研究室卒業生松崎裕佑君に生成フレーズからの作曲に関して貴重なアドバイスを頂いた。研究の進め方に関し名古屋工業大学白松俊准教授より貴重なご助言を頂いた。関係各位に深謝します。

参考文献

- [1] Gerhard Nierhaus: *Chaos and Self-Similarity. Algorithmic Composition*, Springer(2009).
- [2] D. Miell(ed), R. Macdonald(ed) and D. J. Hargreaves(ed): *Musical Communication*, Oxford University Press(2005), 星野悦子 (監訳):音楽的コミュニケーション, 誠信書房 (2012).
- [3] 国土地理院:基盤地図情報ダウンロードサービス (online), 入手先 <<https://fgd.gsi.go.jp/download/menu.php>>(2017).
- [4] 梅村祥之:楽曲コーパス中の 8 音符からなるフレーズに対する心地よさの客観評価法の提案, 第 105 回情報処理学会音楽情報科学研究会 (2014).
- [5] 大野拳汰, 山本樹, 佐々並賢志, 梅村祥之: 対話音声の韻律情報を用いたメロディ生成の試み, 第 110 回情報処理学会音楽情報科学研究会 (2016).
- [6] 新谷歩: 今日から使える統計学講座: 歯学統計チェックリスト, 日本補綴歯科学会誌, Vol.6, No.1, pp.1-11(2014).