

画像の特徴を用いたメロディ自動生成システムの提案

若林 努[†]

小林 孝浩[†]

平林 真実[‡]

本研究では、iPhone による画像の特徴を用いたメロディ自動生成システムを開発する。いつでもどこでも思いついた時に操作可能な携帯端末の利点を生かした、エンターテインメント性のある自動作曲システムを目指す。手始めに、システム構築および動作テストの行いやすさを考慮した環境により、自動作曲アルゴリズムの模索を行う。本稿では、主に自動作曲アルゴリズムの模索に関する内容を述べる。なお、色情報と画素の連なり方、今回はこの2つを画像の特徴として用いた。

Proposal of an Automatic Melody Generation System by using Features of an Image

TSUTOMU WAKABAYASHI[†]

TAKAHIRO KOBAYASHI[†]

MASAMI HIRABAYASHI[‡]

In the present study, the automatic melody generation system that uses the features of an image on the iPhone is being developed. This study aims at the automatic composition system with entertainment factors by using a smart phone that can be used at anytime and anywhere. As the first step of the development, the automatic composition algorithm was considered on the test environment. In this paper, the automatic composition algorithm is mainly discussed. The two features such as color and chain condition of pixels in an image are employed.

1. はじめに

1.1 本研究について

音楽を作り出す手法には、人の手による作曲以外にも計算機を用いた自動作曲が存在する。現在、自動作曲の手法は数多く存在しており、マルコフ過程やニューラルネットワークなどを用いた手法がある。

既存の自動作曲システムの中には、画像やテキストファイルなどの情報を音に変換するもの¹⁾や画像の色情報の特徴として利用するもの²⁾なども存在する。このようなシステムは、画像を入力するだけで音楽が得られる簡易性が一つの長所としていえる。また、画像から奏でられる音楽の想像と挑戦が可能であり、このような能動的な楽しみ方ができる点も大きな長所である。

近年、音楽を能動的に楽しむためのツールとして、Apple 社が提供するスマートフォンである iPhone が注目を浴びている。マルチタッチディスプレイなど、iPhone ならではの要素を利用した音楽系アプリケーションが公開されており、作曲や演奏などを手軽に楽しむことが可能である。

本研究では、iPhone による画像の特徴を利用した

メロディ自動生成システムの実現を目指す。システムにより生成された音楽を、入力に用いた画像のイメージに近づけることが主たる目標である。また、ジェスチャーなどの直感的操作により様々なアクションを起こす。人が集う場で即興として楽しむ。融通の利かない空き時間を有意義に埋める。このようなことが可能な携帯端末である iPhone ならではの利点を生かした、場所を選ばないエンターテインメント性のある自動作曲システムを目指す。

本研究で想定しているシステムの利用方法、および、利用シーンを図1に示す。また、システムを利用して音楽を生成するまでの流れを図2に示す。

1.2 関連研究

画像の分析から取得した情報を音に結び付ける研究は、旧来より行われている。音楽の推薦と提示を目的としたものでは、風景画像の色彩情報から判定された環境状況と感性情報との相関量から音楽データを推薦する研究³⁾がある。その他にも、画像の対象物および色分布から得られる印象に合わせた音楽の自動アレンジ手法に関する研究⁴⁾などがある。音楽の推薦や編曲を目的としたこれらの研究に対し、本研究は、画像から抽出した特徴により作曲を行うことが目的である。音楽を自動生成するには、画像の特徴を調性や和音などの音楽を構成する要素に結び付ける。

また、最近では iPhone を利用した能動的音楽シス

[†] 情報科学芸術大学院大学

Institute of Advanced Media Arts and Sciences

[‡] 国際情報科学芸術アカデミー

International Academy of Media Arts and Sciences

テムの研究も行われており、マルチタッチパネルディスプレイの特性を用いて擦弦楽器を疑似的に再現する電子楽器のインタフェースについての研究⁵⁾などがある。音楽の能動的な楽しみ方としては演奏が最も分かりやすい例であるが、本研究では自ら用意したスコア（画像）から生成される音楽を聴く形で能動的な楽しみを体験する。

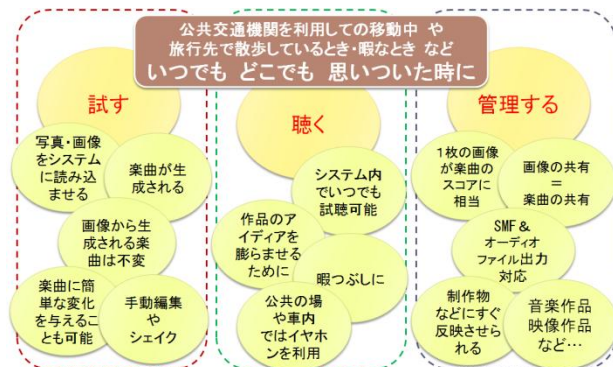


図1 利用方法と利用シーン

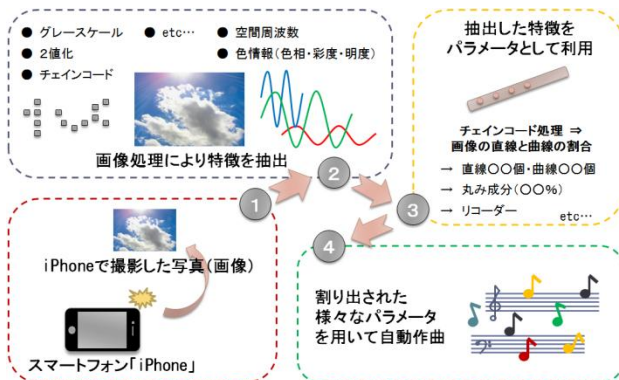


図2 音楽を生成するまでの流れ

2. 画像の対象について

画像には、空が写った写真の様に対象が曖昧であるもの、また、一輪の花が写った写真の様に対象が明瞭であるものなど様々なものがある。本研究で扱う画像の対象については、以下の様なものが候補として挙げられた。

- 風景（写ったものの全て）
- 人の顔
- 自然の繰り返し

風景を対象とする手法では、画像中の色や形などの様々な情報を特徴として抽出する。抽出した特徴は、テンポや調性など音楽を構成する要素のパラメータとして利用する。

人の顔（肌色）を対象とする手法では、画像中に写った顔の数や大きさ、位置を特徴として利用する。このように対象を何かに限定した場合、その数が多けれ

ば賑やか、少なければ静かなメロディ、などというように、画像の特徴と音楽のパラメータとの結び付けが比較的容易である。

自然の繰り返しとは、木目模様などの様な 1/f 揺らぎを持った対象のことを指している。1/f 揺らぎは人間にとって、心地の良いリズムであるとされている。この特性を利用して、画像から繰り返しパターンの特徴を抽出し、癒しの効果を持った音楽の生成を試みる。

いずれは、画像の対象に合わせてユーザが任意でモード切り替えできるシステムにする予定であるが、今回は、風景を対象とした自動作曲システムを構築する。

3. 自動作曲アルゴリズムの模索

iPhone にシステムを実装する前に、システム構築およびテストの行いやすい環境を用いて自動作曲アルゴリズムの模索を行う。自動作曲アルゴリズムの模索には、ビジュアルプログラミング言語である Max を用いる。Max は、拡張モジュールである MSP と Jitter が用意されているため、オーディオ処理や画像・映像処理が容易である。

本章では、この開発環境を利用した音楽生成手法について述べる。

3.1 生成される音楽について

システムにより生成される音楽は、以下の様なルール・要素を持つ。

- 4 小節（音楽の長さ）
- 4/4 拍子（拍子数）
- 40～200bpm（テンポ）
- MIDI 音源（音色）
- 長調と短調（調性）
- メジャースケールとナチュラルマイナースケール（音階）
- メジャーコードとマイナーコード（和音）
- 音価と音高（メロディ）
- カデンツの原理に基づいた和音進行
- 640×480 の画像からなる音楽（画素数）
- 同じ画像からは同じ音楽を生成する

生成される音楽の長さは 4 小節に限定し、和音はセブンスコードやテンションコードを考慮せず 3 和音に限定した。また、生成するメロディは音の強弱を考慮せず音価と音高のみを設定した。

3.2 抽出する特徴

今回、自動作曲を行うために画像から以下の様な特徴を抽出した。

- 色情報（色相・彩度・明度）
- 画素の連なり方（画像内の直線と曲線の割合）

3.2.1 1枚の画像による色情報

色情報は、画像全体もの、もしくは、画像を分割した後にその領域内のものを利用する。

調性（長調と短調）の決定は、分割のされていない1枚の画像を利用する。1枚の画像の全画素から、RGBごとに色の明度の平均値を割り出す。このとき、3色の明度の平均値を足し合わせ、 256×3 段階中 200 以上の場合には調性を長調に決定する。また、200 未満の場合には、調性を短調に決定する。しかし、長調や短調といっても様々な種類があり、長調を例とした場合、ハ長調（C メジャー）やト長調（G メジャー）などがある。その部分については、参考資料⁶⁾⁷⁾にない決定する。白が目立つ画像はハ長調、青が目立つ画像はト長調などとし、この色（調）を判定するためには RGB ごとに幾つかの段階を設けて色の閾値処理を行う。なお、今回は、A~G の長調、Am~Gm の短調、計 14 種類の調を用いる。

音階（スケール）は調性に合わせて決定する。したがって、調性が長調の場合はメジャースケール、短調の場合はナチュラルマイナースケールとなる。

和音進行のパターンについては、長調と短調合わせて数十種類を用いる。決定方法は、まず、3 色の明度の平均値を足し合わせ、その値を今回用いる和音進行のパターン数で除算する。そして、その計算によって算出された剰余の値を用いることで、和音進行のパターンを決定する。

1 小節中の和音（コード）は、和音進行のパターンに応じて自動的に設定される。例えば、調がハ長調（C メジャー）、和音進行のパターンが $I \rightarrow IV \rightarrow V \rightarrow I$ である場合、I には C、IV には F、V には G が設定される。

テンポについては、まず、3 色の明度の平均値を足し合わせ、その値を 3 で除算する。算出された整数値を利用し、計算により発生した剰余は無視する。なお、テンポは下限を 40、上限を 200 と決めているため、算出された整数値が 40 未満の場合は 40、200 を超えた場合は 200 に設定される。

3.2.2 分割した画像による色情報

次に、画像を分割した特徴抽出法について述べる。まず、分割した領域には時間軸を設け、小節数として利用する。今回生成する音楽のルールとして、小節数は 4 に限定しているため、図 3 に示す通り、画像も等間隔で横に 4 分割する。

1 小節中のメロディの音価を決定する場合、領域内の全画素から、RGB ごとに色の明度の平均値を割り出す。その後、それぞれの平均値の中で最も値の大き

い色を選出し、R（赤）ならば 8 分音符、G（緑）ならば 4 分音符、B（青）ならば 2 分音符とした。例えば、G（緑）と判定された場合、その小節中は 4 つの単音が奏でられる。ただし、選出された色以外でも平均値の大きさ次第では音価に反映される。例えば、R（赤）と G（緑）、平均値の大きさが 15 以内の差で G（緑）と判定された場合、最初の 2 拍は 4 分音符、後の 2 拍は 8 分音符となり、その小節中は 6 つの単音が奏でられる。今回は、赤は活発なイメージ、緑は平凡なイメージ、青は静寂な（音が少ない）イメージであるなど、主観に基づいて設定した。

音高の決定のため、すでに分割された領域を 1 小節中で奏でられる音の数分だけ更に縦に分割する。1 小節中 4 つの音が奏でられるとすれば、縦方向に等間隔で 4 分割することになり、一番上の領域がその小節中の 1 つ目の音、その下の領域は 2 つ目の音という形で割り当てていく。その後、分割された領域の RGB ごとに色の明度の平均値を割り出す。B（青）が最も高い場合はルート音（コードが C ならば C）を割り当て、G（緑）の場合は第 3 音（コードが C ならば E）、R（赤）の場合は第 5 音（コードが C ならば G）を割り当てる。

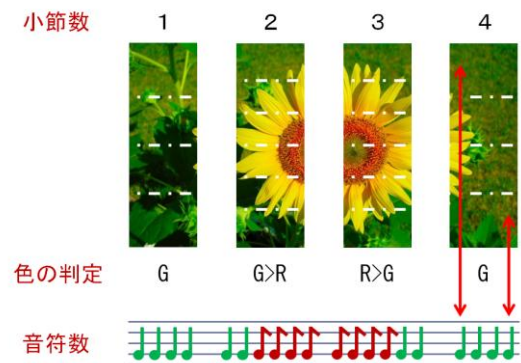


図3 分割した画像の特徴抽出

3.2.3 画像内の直線と曲線の割合

この処理には、1 枚の画像を用いる。前処理として、1 枚の画像を一定の閾値で 2 値化処理し、その画像に対しエッジ検出の処理をかける。すると、画像中の対象のエッジが検出され、画素の連なりが確認できる。

音色を決定するために直線と曲線を割り出す処理を行う。まず、画像の左上から右下へ向かって、エッジとして検出された画素が見つかるまで検索を実行する。検出後、画素をカウントしていき、画素数が 40 に達した時点で、その画素の連なりを 1 本の線 (ll) とみなす。ただし、画素数が 40 に満たない画素の連なりはゴミとみなされ、直線や曲線に含まれない。次に、

その線の両端を結び、線分 (l2) を作成する。その線分の長さ (画素数) が 22 以下であった場合、画素の連なりからできた 1 本の線 (l1) を曲線と判定する。それ以外は直線と判定する。線の判定の例を図 4 に示す。なお、この例には拡大した画像の一部を利用している。その後、画像中に含まれる直線と曲線の割合を計算し、それをもとに音色を決定する。

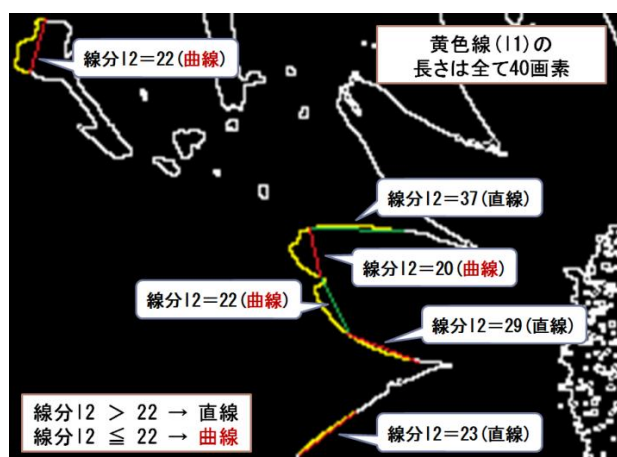


図4 直線と曲線の判定例

3.2.4 画像の特徴から生成された音楽

画像から分析した特徴により生成された音楽を図 5 に示す。

3色の平均値の総和 217 $\rightarrow$ R: 96 + G: 116 + B: 5

調の判定 長調(木長調)

音階の判定 メジャースケール



テンポの判定 $217 / 3 = 72 (\dots 1) \rightarrow 72 \text{ bpm}$

小節数

R→第5音
G→第3音
B→ルート音

色の判定

Key E (I→IV→V→I)

音価と
音高



図5 画像の特徴から生成された音楽

4. おわりに

本稿では、iPhone への実装を最終目的とした、画像の特徴を用いた自動作曲システムを提案した。また、PC 上で Max を利用した自動作曲アルゴリズムの模索についての内容を述べた。引き続き自動作曲アルゴリズムの模索を続け、画像処理による特徴抽出法および音楽を構成する要素と抽出した特徴との結び付けを模索していく。また、西洋音楽理論にとらわれない音楽生成アルゴリズムについても検討する。

音楽を構成する要素として今回は、調性と和音進行、和音とメロディ、そして、音階と音色、テンポを用いた。また、今回メロディにはコード・トーン (和声音) のみを用いた。しかし、これでは音楽を構成する要素数としては不十分であるため、音楽の表現を広げるためにも要素数を増やす必要がある。したがって、今後は、経過音や刺繍音などのノンコード・トーン (非和声音)、また、その他の要素 (音の強弱など) を考慮していく。

これらの実験を進めつつ、iPhone への実装を行い、場所を選ばないエンターテインメント性のある自動作曲システムを実現する。

謝辞 本研究を進めるにあたり、相談に乗っていた IAMAS の教員や学生の皆様に感謝する。

参 考 文 献

- 1) 佐藤武英: 画像から音楽を自動演奏「ピクチャーメロディー」v1.2, (2003).
<http://www.forest.impress.co.jp/article/2003/04/21/okiniiri.html>
- 2) 小島健治: 画像の色情報を音楽へ変換する「RGB_MusicLab」v34, (2010).
http://www.forest.impress.co.jp/docs/review/20100421_362686.html
- 3) 桐本篤, 佐々木史織, 清木康: 風景画像を用いた環境状況コンテキスト対応型音楽推薦システムの実現, 電子情報通信学会第 19 回データ工学ワークショップ(DEWS2008).
- 4) 大山喜冴, 伊藤貴之: DIVA: 画像の印象に合わせた音楽自動アレンジの一手法の提案, 芸術科学会論文誌, Vol. 6, No. 3, pp. 126-135, (2007).
- 5) 安藤大地, 馬場哲晃: マルチタッチディスプレイを利用した電子擦弦楽器とエフェクトコントローラ, 情報処理学会インタラクシオン (2010).
- 6) 最相葉月: 絶対音感, 新潮社, pp. 141-148, (2006).
- 7) 絶対音感持論, (1997).
<http://nucl.phys.s.u-tokyo.ac.jp/torii/music/onkan1.html#colormy>