

## 音楽に合わせた画像生成のための基礎検討

兵頭 博成 齋藤 隆文<sup>†</sup>

東京農工大学大学院工学研究科電子情報工学専攻

<sup>†</sup> 東京農工大学工学部情報コミュニケーション工学科

〒184-8588 東京都小金井市中町 2-24-16

E-mail: hyodo@vc.cs.tuat.ac.jp, <sup>†</sup> txsaito@cc.tuat.ac.jp

本研究では、音楽に含まれる情報をもとに、それにあった画像の生成を試みた。音楽データとして SMF (Standard MIDI File) を用い、画像生成はその中に含まれる楽音のパラメータと二次元画像オブジェクトのパラメータをマッピングすることで行なう。いくつかのマッピング方法について試行したが、パラメータ同士の単純なマッピングでは音楽・画像の感性レベルでの調和は難しい。このため、より高次の情報として、和声構造を把握するためのマッピング方法についても検討した。完全五度の上下進行を環状に並べること(五度圏)により、音楽の和声・調性といった情報を視覚的にとらえることができる。

## Basic Examination for Image Generation Matched with Music

Hironari HYODO Takafumi SAITO<sup>†</sup>

Div. of Electronic and Information Engineering, Graduate School of Technology,  
Tokyo University of Agriculture and Technology

<sup>†</sup> Dept. of Computer, Information and Communication Sciences,  
Tokyo University of Agriculture and Technology

2-24-16 Nakamachi, Koganei-shi, Tokyo, 184-8588 Japan

This paper presents an trial to generate images matched and synchronized with music. We use Standard MIDI Files (SMF) as input musical data, and generate images by mapping musical sound parameters to 2D object parameters. We examined several mapping combinations, but it is difficult to obtain sensuous matching by using these simple mapping methods. Therefore, we examined the mapping methods to recognize harmonic information. By using "circle of fifth", harmonic information and key in music can be visualized.

## 1. はじめに

映画・TV 番組などの映像作品では、必ずといってもよいほど BGM が使用されている。ドラマなどの悲しいシーンで奏でられる悲しげな曲に思わず涙したり、あるいはホラー映画を観ているとき、BGM によって恐怖感が増したりということは多くの人を経験していることだろう。このことは映像における BGM の重要性を示唆しているといえる。なお、以上の例では映像に音楽を付加するという映像主導の場合を示しているが、逆に音楽に映像を付加するという音楽主導の場合も考えられる。現在盛んに行われている音楽作品に対するビデオ・クリップの制作などがそれにあたりといえる。いずれにしても、こうした映像作品の制作者は、意図的にこれら映像と音楽を組み合わせ、お互いがうまく作用して視聴者に対し効果的に情報伝達がなされるよう工夫している。これは作品を制作する際の大きな手間となる。

そこで、映像からその雰囲気に合った音楽、あるいは音楽からその雰囲気に合った映像を自動的に生成することはできないか、という要求が生じる。本研究では、このうち音楽に合わせた映像の生成を目指す。

しかし、音楽と映像の間で、雰囲気の対応が何をもってなされるのかが明確でない。そこで本研究では、音楽と画像の基本的な要素のマッピングによりそれらの雰囲気をマッチさせることはできないか、という検討を行なう。

画像の生成は、まず図 1 の方法で行うこととする。すなわち、与えられた音楽の中から、これを構成するパラメータ(音高、音量、音色、テンポ、リズムなど)を抽出し、これを何らかの形で二次元オブジェクトの要素(色、形状、大きさ、位置など)に対応付け、リアルタイムで画像を生成するというものである。

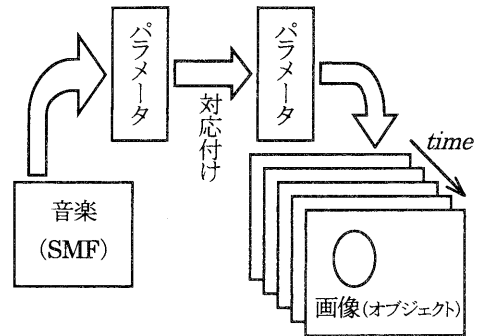


図 1. 画像生成の方法

## 2. 楽曲中の情報とその抽出

### 2.1. 音楽を構成する要素

音楽の世界では、「リズム」「メロディ」「ハーモニー」の三つが音楽の主要な要素であるとされている。これに対し、音楽の基本単位である「音」は、「音高」「音量」「音色」の三つの要素を持つ。本研究では、これら「音の三要素」に着目し、音楽の基本となるパラメータとして扱う。

### 2.2. 入力データ

入力データには SMF (Standard MIDI File) を用いる。SMF は、MIDI 機器による演奏情報を異なるシーケンサ同士でやり取りするために定められたファイル・フォーマットである。MIDI では機器の操作(演奏)を「鍵盤を弾く」「音色を変える」といった情報を持つ「MIDI メッセージ」によって行う。SMF には、こうした MIDI メッセージとそのタイミングの情報(イベント)が記録されている。つまり、楽譜のように整理・抽象化されている情報とは異なり、演奏上のニュアンスなども含めた具体的な演奏の動作そのものが記録されている。

### 2.3. SMF からのパラメータの抽出

SMF から各要素を抽出する方法を以下に示す。

#### (1) 音高

SMF での一音の情報は、「ノート・オン」と「ノー

ト・オフ」という二つのイベントで構成される。「ノート・オン」で指定された音高の音が鳴り始め、「ノート・オフ」で鳴り止む。音高は、ノート番号(中央 C 音を 60 として半音ごとに数字が割り当てられている)という形でこれらのイベントに含まれる。

### (2) 音量

音量は、「ノート・オン」に含まれる「ベロシティ」の値と、チャンネルごとに設定される「チャンネル・ボリューム」(各チャンネル同士の音量バランスをとるために用いられる)と「エクスプレッション」(クレッシェンド、ディミヌエンドなどの強弱表現を表すために用いられる)で決定される。

### (3) 音色

音色は、「プログラム・チェンジ」イベントによって MIDI 機器に内蔵されている音色番号で選択される。ここでは GM (General MIDI: MIDI 音源の標準化規格)を想定し、音色番号から対応する音色を決める。

## 3. 画像生成の方法

### 3.1. 画像を形成する要素

画像の生成は、基本的に音楽の要素を画像の要素にマッピングさせることで行う。画像は二次元オブジェクトの組み合わせと考え、楽曲中の一音を一つのオブジェクトに対応付ける。画像オブジェクトの要素は、次のようなものを考える。

- 位置
- 色
- 大きさ
- 形状

### 3.2. 要素同士のマッピング

音楽にマッチした画像を生成するには、適当なマッピング方法を考える必要がある。適当と考えられるマッピング方法を以下に示す。

#### (1) 音高 → 位置

音高は音の「高い」「低い」を示すパラメータであるため、位置情報との結びつきが強いといえる。事実、楽譜での記述でも音高は上下方向の位置

により表現されている。また、ピアノをはじめとする鍵盤楽器は左に低音、右に高音を配置しているため、左右方向の位置に高音をマッピングすることも考えられる。また、12 半音を環状に並べ、同じ音名の音を同じ位置に表示することは、和声的な意味をもつ可能性がある。

#### (2) 音高 → 色

音はその周波数の増加に応じて高くなる。これに対し、色は周波数に応じて赤から紫へと色相を変化させる。これを利用し、低音を赤、音高が高くなるにつれ色相を変化させ高音を紫にマッピングさせるという方法も考えられる。

また、12 半音を色相環に対応させるマッピングも考えられる。同じ音名の音に同じ色を割り当てることは、生成画像に調性を反映させる可能性があり、何らかの意味を持つといえる。

#### (3) 音高 → 大きさ

音は高いほど波長が短く、低いほど波長が長くなる。このことから、低い音を大きなオブジェクトに、高い音を小さなオブジェクトにマッピングすることも考えられる。

#### (4) 音色 → 位置

オーケストラなどでは楽器によって配置位置がある程度決まっている。これを利用し、音色によってオブジェクトの表示位置を変えることも一つの方法として考えられる。

#### (5) 音色 → 色

音色はそれぞれが異なる雰囲気を持っている。音色を近い雰囲気を持つ色にマッピングさせることも考えられる。しかし、どの音色がどの色と雰囲気がマッチするか、何らかの方法で検討しなければならない。

#### (6) 音色 → 形状

音色の持つ印象は、たとえば「やわらかい音」「鋭い音」といった表現も可能である。こうした印象を形状へのマッピングに利用することも考えられる。音色の違いは、音波の倍音成分の違いが大きく関わっているため、音色の倍音成分の違いと形状の空間周波数の違いは対応する可能性がある。しかし、どの音色とどの形状を対応させるかについて

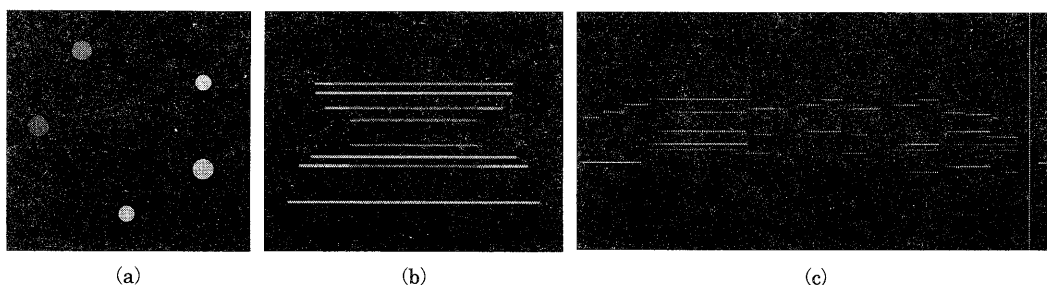


図 2. 検証したマッピングの一部

(a)は「音高(12音階)→位置(環状)」「音高(12音階)→色(色相環)」「音量→大きさ」「形状は円形に固定」、(b)は「音高→位置(上下方向)」「音量→色(輝度)」「音量→大きさ(長さ)」「形状は矩形に固定」、(c)は「音高→位置(上下方向)」「音高→色(色相)」「音量→色(輝度)」「形状は矩形に固定」「履歴を表示」といったマッピングで画像生成を行なった。

ては、音色と形状についての印象の違いを検討する必要がある。

#### (7) 音量 → 色

音量は他の要素と異なり唯一物理的エネルギーを持っている。よって音量を同じくエネルギーを持つ光の強さ、すなわち輝度や明度に対応付ける方法が考えられる。

#### (8) 音量 → 大きさ

音楽における音量の大きさは、ときには圧倒感にもつながる。音量をオブジェクトの大きさにマッピングすることは、相互に印象の強さを高める可能性がある。

### 3.3. 時間の表現

基本的に、画像生成は音楽と同期してリアルタイムで行なう。しかし、音楽は時間軸上での芸術表現であり、時間の概念を画像に加えることは大きな意味があるといえる。人間の感覚では、未来の刺激を予測することはできないが、過去の刺激に関しては「余韻」として残ることも十分ありうる。

そこで、上記のマッピング方法に加えてオブジェクトの履歴を残す方法についても検証した。

### 3.4. マッピングの妥当性

以上に示したマッピング方法から、いくつかの組み合わせで検証を行なった。また、組み合わせの一部を図 2 に示す。

#### • 音量 → 大きさ

このマッピングは、感覚的に合っているといえる。特にエクスペッションを多用したダイナミックレンジの広い曲では、大きな効果を得ている。ただ、音色ごとのエンベロープ(音色ごとにみられる発音中の音量の変化)を考慮する必要性にも気付かされる。ここでは、減衰音であるピアノもノート・オフまで音が持続するバイオリンなども同様に(持続音として)扱われているためである。

#### • 音量 → 色(輝度)

大きさに対応させた場合と同様に感覚にマッチしたマッピングであるといえる。特に履歴を残す方法とともにこのマッピングを使用した場合、音量の変化がグラデーションにより表示されるため、視覚的に興味をひくものとなっている。

#### • 音高 → 位置(上下)

音高を上下方向に配置する手法は、直感的であるといえる。音の高さを「高い」「低い」と表現する人間の感覚は、空間的な「高い」「低い」と密着しているものと考えられる。

#### • 音高 → 色(色相)

このマッピングは、例えば「音高→位置(上下)」のマッピング方法と組み合わせると、色相の変化による美しいグラデーションを呈するが、それが感覚的であるかどうかについては疑問を残した。

#### • 音高(12音階) → 位置(環状)

このマッピングでは、12の半音が順に環状に配置されているため、直感的ではないが和音を構成音のパターンとしてとらえることができるとい

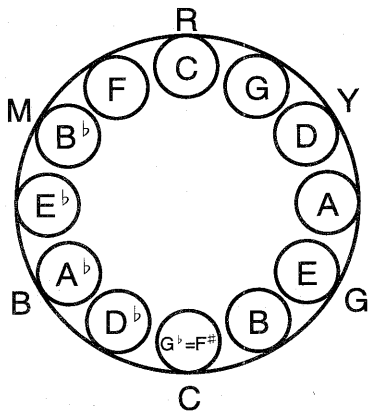


図 3. 五度圏と色の対応(RGB, CYM で表示)

る。

• 音高(12 音階) → 色(色相)

各音名が一つの色に対応するため、色の組み合わせの変化により和音の変化を見ることが出来る。上記の「音高(12 音階)→位置(環状)」のマッピングと組み合わせると、その効果がより明確になる。

以上、考えられる候補のいくつかについてその妥当性を検証した。

しかし、これらのようにパラメータを加工せず使用了場合、音楽の感性に合った画像は表現しにくい。音楽と画像との感性レベルでのマッチングを期待するならば、こうしたパラメータからより高次の情報(和声、調性、拍など)を取得し、画像に反映させることが求められるといえる。

## 4. 和声のマッピング

### 4.1. 五度圏の利用

和声を考える上で、音楽理論では「五度圏」がよく用いられる。五度圏は、ある音から完全五度の上下進行を環状に並べた図である。例えば、C から上方に五度ずつ進行させると C-G-D-A-E-B-F#となり、下方に進行させると C-F-B<sup>b</sup>-E<sup>b</sup>-A<sup>b</sup>-D<sup>b</sup>-G<sup>b</sup>となる。F#と G<sup>b</sup>は異名同音(記譜上では異なるが鍵盤上では同じ音)となるため、オクター

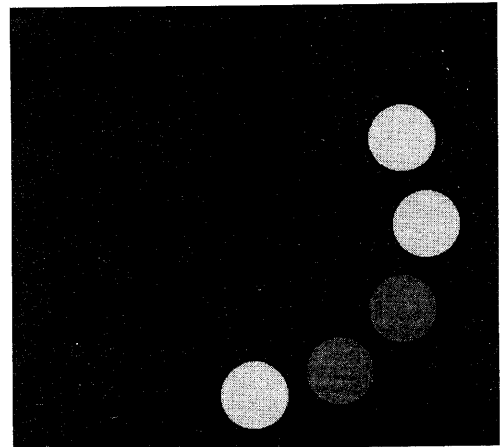


図 4. 表示例(D, A, F#が発音中)

ブを構成する 12 半音すべてを循環する図となる(図 2)。

この五度圏を利用し、楽曲中の和声構造をとらえることを考える。

### 4.2. 五度圏を用いた 12 音階の可視化

楽曲の演奏とともに発音中の音を五度圏の配置で表示した。ここでは、それぞれの音を色相環にマッピングしており、例えばCは赤、Eは緑、A<sup>b</sup>は青で表示される(図 3・図 4)。

楽曲中で和音が鳴ると、その構成音によって図 5 に示すようなパターンが表示される。同じパターンであれば、根音が変わってもパターンの位置が変化するだけでパターンそのものは変化しない。このことは、あらかじめ考えられるコードのパターンを登録しておけば、パターンマッチングによりその時点でのコード・ネームを取得することができる可能性を示している。しかし、和音の構成音が同時に鳴っていないなければならないという制約があり、アルペジオや構成音が省略されている場合などは、このままでは正しい和音を得ることができない。

### 4.3. 五度圏-色相環のカラーマッピング

前述の五度圏-色相環の色マッピングを利用して、ピアノ・ロール表示した楽曲データの一音一

音を着色した(ピアノ・ロールとは、縦軸を音高、横軸を時間とし、楽曲中の音を表示する方法である)。図6はその出力結果の一例である。

楽曲中に転調がある場合、音階を構成する音が変わるため、転調以前と転調以後の部分で色味が変化する。このため、直感的に転調の位置を把握することができる。

## 5. まとめ

SMF から音楽情報を抽出し、その演奏に合わせた画像の生成を試みた。画像の生成は、音楽を構成する音の音高、音量、音色といったパラメータを、二次元画像オブジェクトの位置、形状、大きさ、色といったパラメータにマッピングさせることで行なった。その過程で、パラメータのマッピング方法をいくつかピックアップし、おのおの方法について検証を行なった。これらの方法で、音楽に合った画像は生成できたが、感性にマッチした画像は、これらのパラメータの単純なマッピングでは難しいといえる。このため、五度圏の理論を用いて和声をとらえることを試みた。

音楽に合った画像の生成という目的から考えると、本研究はまだ初期段階に過ぎない。今後は、音楽の雰囲気にもマッチした画像を生成することを目的として、新たなパラメータの抽出や、よりよいマッピング方法を考えていきたい。

## 参考文献

- [1] 大蔵康義：音と音楽の基礎知識，国書刊行会，1999.
- [2] 長島洋一，橋本周司，平賀譲，平田圭二 編：コンピュータと音楽の世界—基礎からフロントエンドまで—，共立出版，1999.

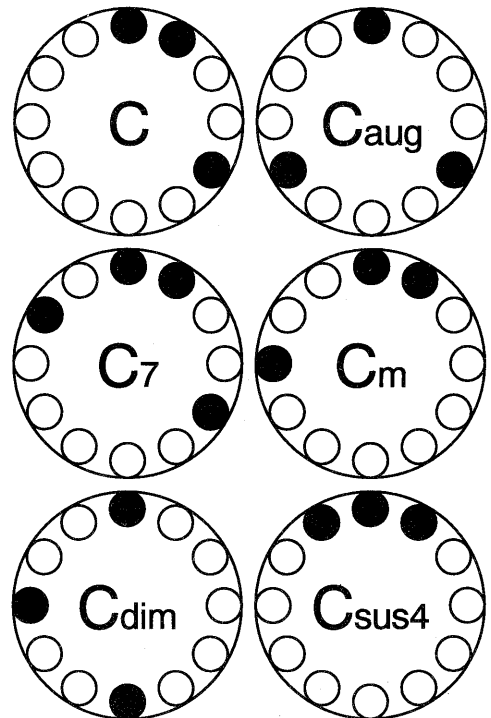


図5. コード・パターンとその可視化例



図6. 五度圏-色相環のカラーマッピングを用いた楽曲のピアノ・ロール表示