Vol.2017-MUS-115 No.59
2017/6/18

カラー画像内の対象物と背景からの印象語抽出に基づく楽曲の半自動生成

藤井 ほのか[1]　齋藤 康之[1,a)]　嵯峨山 茂樹[2]

概要：本研究では，誰でも，手軽に，画像の印象に合う楽曲を生成できる方法について議論する．画像，音楽は人の情動に大きな影響を与える．画像を見る際にその画像の印象に合った音楽が流れれば，より深く画像の印象を伝えられると考えられる．それに適する楽曲の選定は，時間的・労力的な困難さを伴い，著作権の問題も発生しうる一方，作曲は概してさらに難しい．そこで，画像の印象に合う楽曲を適切な楽曲を手軽に得るべく，我々は画像からの楽曲生成を行ってきた．先行研究では，カラー画像の領域分割方法や，画像の配色と印象語との対応づけに改善の余地があったため，本研究では，画像内の対象物領域と背景領域との分割や，配色に対応する印象語の割り当てをユーザ自身が行うこととした．システムは，各領域の色情報に当てはまる「印象語」を 5 つ抽出し，それらの特性から調性，コード進行，テンポを決定する．ユーザは，イントロ，A メロ，B メロ，サビ，アウトロの 5 つの楽曲パートに印象語を割り当て，また，各パートの伴奏パターンを決定する．最終的に，システムはメロディと伴奏を持つ部分楽曲を連結して 1 つの楽曲を生成する．このような，ユーザ介入型の半自動楽曲生成システムを構築した．主観評価実験結果では，5 段階評価で 3 以上が得られ，概ね画像の印象に合った楽曲を生成できている．

キーワード：カラー画像，半自動楽曲生成，対象物と背景，配色，印象語

1. はじめに

画像や音楽は人の情動に大きな影響を与える．そして，人が絵画や写真などの画像を見る際に，その画像に合った音楽を流すことができれば，画像の印象がより深まると考えられる．既存の楽曲を選定する場合，その楽曲を知る人との共感が得られやすいであろう．しかしながら，画像に合う適切な楽曲の選定作業には時間的・労力的な困難さを伴うばかりか，著作権上の問題が生じる場合もありうる．その一方で，自ら作曲をすれば，著作権に関する問題は回避できるものの，概して楽曲の選定に比べてさらに難しい．

そこで我々は，画像の印象に合う楽曲を適切な楽曲を手軽に得るべく，画像からの楽曲生成を行ってきた．前田 [1] は，カラー画像を領域分割・統合して，最も大きな領域と次に大きな領域を結ぶ「メロディ・ライン」を引き，その線上の輝度変化をメロディの音高の変化に対応づけ，全自動で楽曲を生成する方法について検討した．また，根本 [2] は，画像の色情報を「配色の印象」に変換し，心理的な実験理論を参考に，ユーザが介入して，半自動で楽曲を生成するシステムを構築した．これらの研究に共通する改善点として，画像の内容を考慮せずに画像の領域分割を行っていることが挙げられる．また，根本の研究では，配色と印象語との関係づけは，根本の主観により決定した彩度と明度の範囲の組み合わせを用いていた．

本研究では，ユーザが手動で画像を対象物領域と背景領域に分けるようにし，配色と印象語との関係づけもユーザ自身が行うように改める．ユーザの好みを反映しつつ，誰でも，手軽に，画像の印象に合う楽曲を生成できるシステムの構築を目指す．

2. 配色，楽曲と心理学的知見

画像の配色や楽曲は人の感情に関与していると言われており [3], [4]，心理学的知見により画像と楽曲との対応を見出せると考えられる．

2.1 PCCS

PCCS（Practical Color Coordinate System）は，財団法人 日本色彩研究所が定義した色彩調和を主な目的としたカラーシステムである．明度と彩度を「トーン」という概念でまとめ，「色相」「トーン」の 2 系列で色彩調和の基本形

[1] 木更津工業高等専門学校 情報工学科
NIT, Kisarazu College, Kisarazu, Chiba 292–0041, Japan
[2] 明治大学 総合数理学部 先端メディアサイエンス学科
Meiji University, Nakano-ku, Tokyo 164–8525, Japan
[a)] saito@j.kisarazu.ac.jp

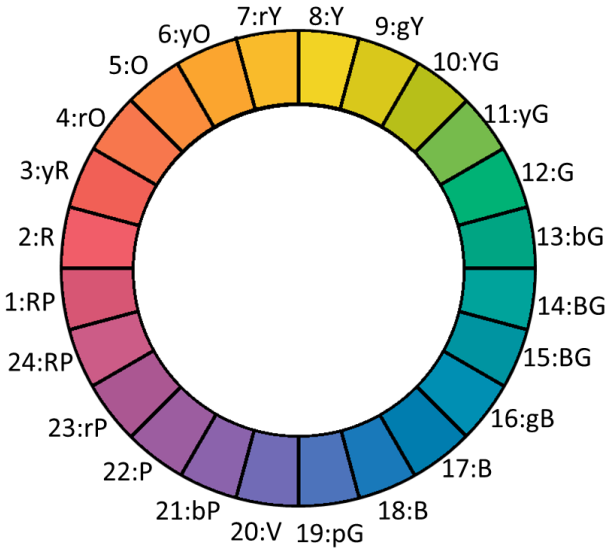

図 1　PCCS 色円環

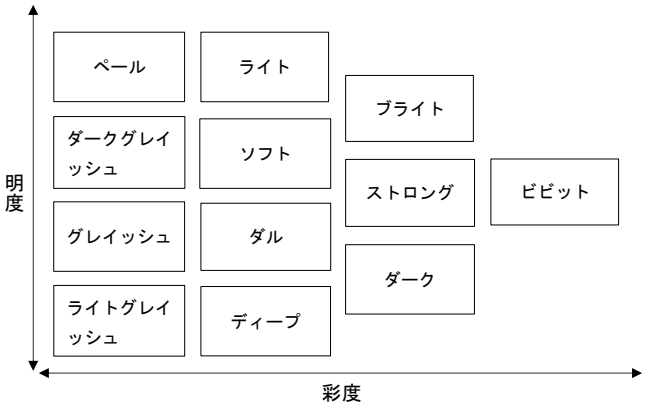

図 2　PCCS トーン

列を表している（図 1，図 2）．

色彩から多くの人が共通して受ける印象として，高明度は「柔らかい色」「膨張色」「軽い色」，低明度は「硬い色」「収縮色」「重い色」とされる．また，高彩度は「派手」，低彩度は「地味」などが一般的に挙げられ，高彩度・暖色系は「興奮色」，低彩度・寒色系は「沈静色」とされる．

2.2　Hevner の研究

音楽心理学者の Hevner の研究 [5] では，楽曲構成要素として調性・テンポ・音高・リズム・和声・旋律の 6 つを挙げている．Hevner は，この 6 つの楽曲構成要素と 8 つの印象語群によって表現される印象との相関係を調べた（表 1）．8 つの印象語群は図 3 のように円形に並べられ，各群の中の形容語は互いに類似性が高く，隣り合う群はやや関連するが類似性はそれほど高くない．そして，円形の反対側に位置する群の形容語は反対の意味を持つように並べられている．

表 1　楽曲構成要素と印象語群の相関係

楽曲構成	印象語群名							
要素名	C1	C2	C3	C4	C5	C6	C7	C8
key	長調	短調	短調	長調	長調	長調	—	—
tempo	遅い	遅い	遅い	遅い	速い	速い	速い	速い
pitch	低い	低い	高い	高い	高い	高い	低い	低い
rhythm	固定	固定	流動	流動	固定	流動	固定	固定
harmony	単純	複雑	単純	単純	単純	単純	複雑	複雑
melody	上昇	—	—	上昇	下降	—	下降	下降

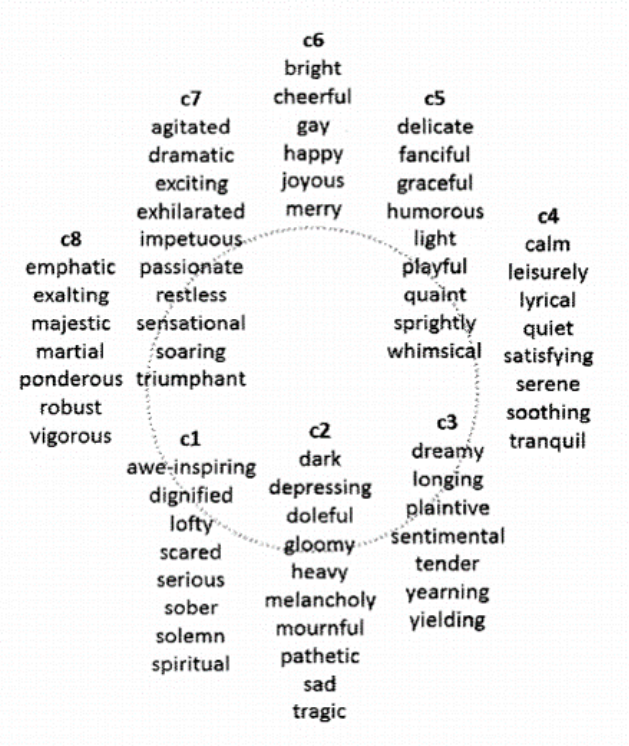

図 3　8 つの印象語群

3.　画像からの半自動楽曲生成方法

本システムは，入力画像から得られる配色の印象を楽曲に変換し，電子音楽ファイル SMF（standard MIDI file）を出力する．システムの流れを図 4 に示す．処理過程は，画像処理部と楽曲生成部の 2 つからなる．

3.1　画像処理部

入力画像の持つ色情報は，RGB 各 8 ビットとする．各画素の RGB 値が，どの色相・トーンに相当するのかを分析し，配色から得られる印象語を抽出する．印象語は Hevner の 8 つの印象語群の中から「awe-inspiring，dark，dreamy，calm，delicate，bright，dramatic，majestic」を引用することとする．

3.1.1　対象物と背景の分離

入力画像を対象物と背景に分離する．入力画像を読み込み，マウス・ペイントを用いて画像内の対象物としたい部

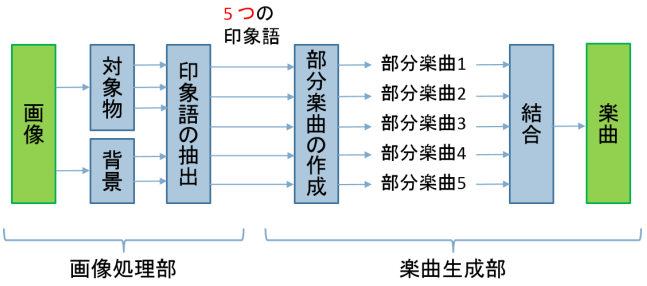

図 4　システムの流れ図

分を白く塗りつぶす．その後，その白く塗りつぶした部分をマスク・パターンとして入力画像に対してマスク処理を行い，対象物領域を抽出する．そして，その領域のみを含む画像を，対象物画像として生成する．

最初に白く塗りつぶしていない領域を背景領域として扱い，背景画像内に含める．

3.1.2　色の分析と減色

配色の組み合わせを分析しやすくするため，減色処理を施す．各画像内の全画素に対し，各画素の色が PCCS 色相環，PCCS トーンの概念のどこに位置するか分析する．

色相は有彩色の場合，24 の色相のどれであるか，トーンは 12 種のどれであるかを調べる．あらかじめ色見本 [6] に従い，PCCS 色相環，トーンの各色において代表色の色相番号をデータベースに登録しておく．

RGB 色空間における色同士の距離を調べることによって注目画素の色がデータベース中のどの色に最も近いかを分析する．その距離 d は，調べたい画素の RGB の値を R_j, G_j, B_j とし，あらかじめ登録されている RGB の値を R_k, G_k, B_k とすると，式 (1) で求まる．

$$d = \sqrt{(R_j - R_k)^2 + (G_j - G_k)^2 + (B_j - B_k)^2} \qquad (1)$$

3.1.3　印象語の抽出

減色した各画像を基に，以下の観点から対象物画像より 3 つ，背景画像より 2 つの合計 5 つの印象語を抽出する．

(1) ユーザが対象物自体に抱く印象
(2) 対象物画像内で最も多く含まれているトーンの配色
(3) 対象物画像内で 2 番目に多く含まれているトーンの配色
(4) 背景画像内で最も多く含まれているトーンの配色
(5) 背景画像内で 2 番目に多く含まれているトーンの配色

(1) はユーザが印象語を直接 割り当てる．(2)〜(5) は各トーンの配色に対する印象を事前にユーザに指定してもらい，その情報と照らし合わせて当てはまる印象語を抽出する．

3.2　楽曲生成部

画像処理部で得られた 5 つの印象語を基に楽曲生成を行う．楽曲の構成は一般的な楽曲に用いられるイントロ，A メロ，B メロ，サビ，アウトロで構成し，得られた印象語をユーザがそれらに対応づける．その後，対応に従って楽曲を作成し，最後に楽曲を繋げ 1 つの楽曲として出力する．

表 2　調性の決定

調性	印象語
ハ長調	delinate, bright, dramatic, majestic
イ短調	awe-inspiring, dark, dreamy, calm

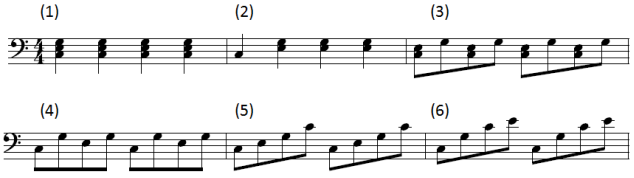

図 5　伴奏パターン

表 3　テンポ決定で用いる重み

抽象的なテンポ	印象語群	印象語	重み
遅い	c1	awe-inspiring	-14
	c2	dark	-12
	c3	dreamy	-16
	c4	calm	-20
速い	c5	delicate	6
	c6	bright	15
	c7	dramatic	20
	c8	majectic	10

3.2.1　調性，コード進行，伴奏パターン，テンポの決定

Hevner は印象語を長調と短調に分類した．本研究では，ハ長調とイ短調を扱うこととし，得られた印象語が属するグループの多い方の調性を楽曲の調性とする．表 2 に調性と印象語の対応を示す．たとえば，「bright, calm, delicate, calm, calm」が得られた場合，ハ長調とする．

ハ長調ならば C，F，G，イ短調ならば Am，Dm，E としてコードを決定する．この際，コード進行を HMM（隠れマルコフモデル）でモデル化する．確率的に状態遷移することで，印象の異なるコード進行になると考えられる．

伴奏は図 5 に示した 6 つのパターンの中からユーザが指定し，コードに合わせて決定する．ある伴奏パターンを複数のパートで重複して用いてよい．

楽曲全体のテンポは Hevner の研究より印象語により決定する．まず，得られた印象語が属するグループの多さにより，楽曲全体のテンポを「速い」「遅い」というような抽象的なテンポとして決定する．次に，表 3 に示した Hevner の研究により得られた重みを用いて，式 (2) により具体的なテンポ t の決定を行う．

$$t = 88 + w_{sum} \quad (44 \leq t \leq 184) \qquad (2)$$

ここで，「遅い」テンポの上限の 60 と，「速い」テンポの下限の 130 の中間となる 88 を基準のテンポとしている [7]．また，w_{sum} は重みの合計を示すが，決定された抽象的な

表 4　印象語によるメロディの決定

メロディのとり方	印象語
固定	awe-inspiring, dramatic, dark, delicate, majestic
流動	bright, dreamy, calm

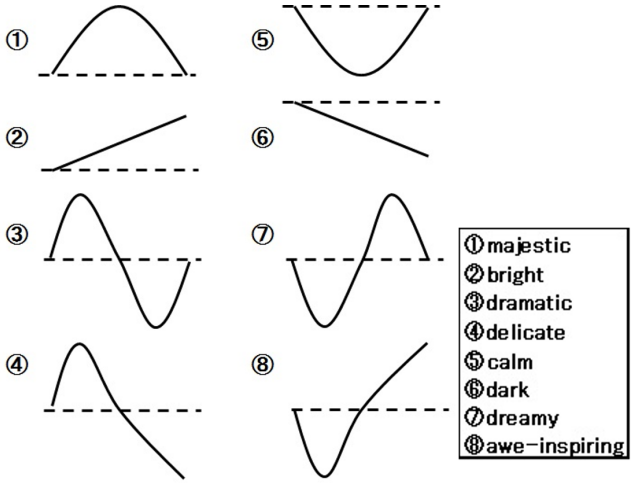

図 6　印象語と音高パターンの対応

テンポのグループに属する印象語の重みだけを用いる．そして，通常の楽曲で用いられるテンポの範囲に収まるように，t の値域には下限と上限を設け，各々 44, 184 とする．

3.2.2　メロディ・トラックの作成

あらかじめ 1 小節間のリズムパターンを数十通り用意し，平均音長が長い順にソートしてデータベースに登録しておく．そこに，事前に印象語と対応づけておいた音高パターンを重ね合わせて音高を持ったリズムパターンを生成し，これをメロディ・トラックとする．音高の変化パターンの単位としては，多くのフレーズの変化単位である 2 小節とする（図 6）．

また，Hevner の研究から，メロディを「固定」または「流動」に決定する．固定は並び替えられた 3 パターンのリズムを 8 小節間繰り返し，流動は 1 小節ごとにリズムの変更を行う．表 4 にメロディのとり方と印象語との対応を示す．

3.3　システムの実装

OpenCV [8] の HighGUI モジュールを用いてシステムを作成した．トラックバーによる選択と，マウス・イベントによるマウス・ペイントを実装した．説明文は画像として作成し，`imshow()` で表示した．

まず，ユーザは 12 種のトーンに対し抱いた印象に合う印象語をあてはめていく（図 7）．この情報は保存しておいて，次回のシステム使用時には自動的に呼び出される．

続いて，入力画像をキャンバスとしたマウス・ペイント画面が表示され，ユーザは対象物としたい部分を白く塗りつぶす．手動による対象物の抽出例を図 8 に示す．ペイント

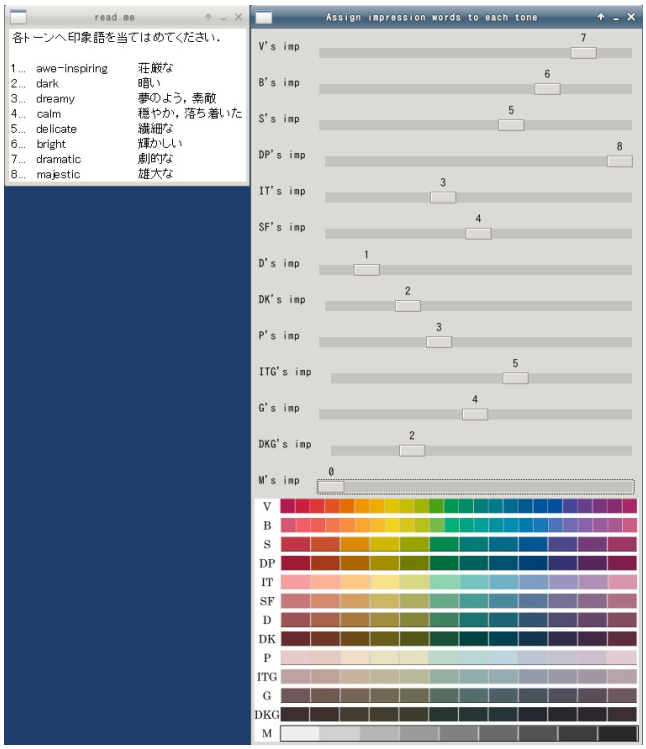

図 7　トーンの印象の選択．右のウィンドウの下段に示されている各トーンの印象が，左のウィンドウに示されている印象語のどれに該当するかを選択する．図中，一番下の「M」は初期値の 0 となっており，これは未選択状態であることを示す．

トが完了すると，生成された対象物画像が表示され，そこで対象物に抱いた印象に合った印象語を当てはめる（図 9）．

そして，抽出された 5 つの印象語が表示され，ユーザは各パートに印象語を 1 つずつあてはめる（図 10）．最後に，各パートに伴奏パターンを当てはめ（図 11），ユーザからの入力が全て完了すると楽曲が生成される．

4.　実験結果

実際にユーザにシステムを利用してもらい，主観評価実験を行った．木更津高専の吹奏楽部員 4 名に「1 回目と 2 回目とで異なる対象物を選ぶ」ことを条件とし，すなわち，同一の画像に対し，(a) 画像内のある領域 A を対象物とした場合と，(b) 画像内の B ($A \neq B$) を対象物とした場合の各々で「楽曲 (a)」と「楽曲 (b)」を生成した．評価項目は，以下の 5 つである．

(1)　画像に合っているか
(2)　メロディが自然か
(3)　伴奏が自然か
(4)　ユーザの好みが反映されているか
(5)　対象物の違いが楽曲に表れているか

実験の流れとしては，楽曲 (a) を生成して評価項目 (1)〜(4) を答えてもらい，その後，楽曲 (b) を生成して評価項目 (1)〜(4) について答えてもらった．最後に，楽曲 (a) →

(a)

(b)

(c)

図 8 手動による対象物の抽出例．(a) 入力画像，(b) 対象物領域を白色で塗りつぶした例，(c) 対象物領域．なお，(b) にて白色で塗りつぶしていない領域が背景領域として扱われる．

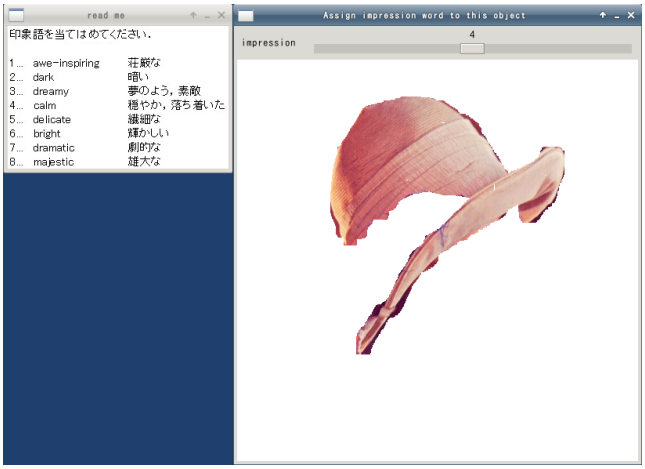

図 9 対象物に対する印象語の割り当て

楽曲 (b) の順で楽曲を流し，評価項目 (5) を 1 回 答えてもらった．実験に利用した画像を図 12 に，被験者 4 名をそれぞれ S_1，S_2，S_3，S_4 とした評価結果を表 5 に示す．なお，表中の (a) と (b) は，被験者が作成した楽曲 (a)，楽曲 (b) を指す．概ね画像に合った楽曲を生成することができたといえる．また，伴奏は安定して自然であり，対象物の違いが楽曲にもある程度は表れている．

5. 考察

画像から受ける印象に合う楽曲の生成が概ねできているといえる．しかしながら，メロディの自然さやユーザの好みの反映については改善の余地がある．今後，リズムや音

図 10 印象語のパートへの割り当て．同一の印象語が抽出されることがあるが，それらの区別はない．この例では calm が 2 つ現れており，各々 A メロとサビに割り当てているが，逆に割り当てても同じ結果となる．

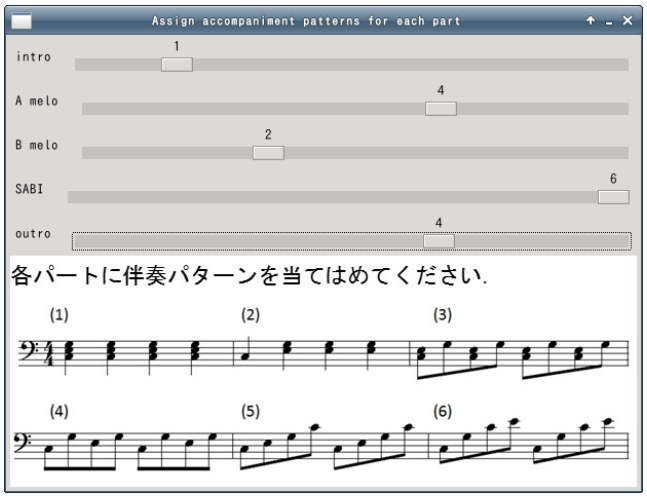

図 11 伴奏パターンの割り当て．あるパターンを複数のパートで重複して用いてもよい．

高変化のパターンをさらに充実させていくことが必要であり，また，それらをユーザが追加できる仕組みも整えたい．それと並行して，楽曲が単調にならないような工夫も施す必要があるだろう．今回は，楽曲の調性にハ長調とイ短調のみを用い，それを楽曲を通じて固定して適用していたが，他の調性を含めることのほか，転調できるようにするとともに，その遷移が自然となるような音高変化となるように改善したい．

また，テンポの変化がほとんどなかったという意見が多

図 12　主観評価実験で用いた画像 [9]

表 5　主観評価実験結果（5 段階評価）

評価項目	S_1		S_2		S_3		S_4	
	(a)	(b)	(a)	(b)	(a)	(b)	(a)	(b)
(1)	4	3	4	5	5	5	3	4
(2)	3	3	5	3	4	3	4	4
(3)	4	5	5	4	5	4	4	4
(4)	4	4	4	3	3	4	4	5
(5)	5		4		5		3	

かった．その原因としては，基準テンポの 88 [bpm] というのは，下限の 44 と上限の 184 に対してやや「遅い」側に偏っており，「遅い」グループに属する印象語が多く抽出されるとすぐに下限を下回ってしまう（その場合は，下限の 44 [bpm] にクリッピングされる）ことが挙げられる．そこで今後は，基準テンポを見直すことや，表 3 の値に対して 1 未満の乱数を乗ずることなどの修正が必要であると考えられる．

その他，今回はピアノ楽曲の生成を念頭においたが，多くの楽器パートを用いてより豊かな響きとなる楽曲の生成を行いたい．楽器（音色）の選択やドラムスの有無などをユーザが指定できるようにすれば，柔軟性が高まり，ユーザの好みを反映できるだろう．

ユーザ・インタフェースの面では，現在は対象物領域の抽出におけるやり直しの機能が未実装であるので，早急に対応したい．また，注目する対象物領域を白色で塗りつぶすという方法そのものが分かりにくい．たとえば，対象領域を囲む輪郭線を修正しつつ，背景領域の彩度や明度を低下させ，どこが対象領域として扱われているのかを視覚的に鮮明化することなどが考えられる．その作業の支援としては，たとえば，文献 [10] で提案されているように，ユーザが大まかに輪郭を与えた後に自動的に詳細な輪郭線を求める方法などが考えられる．

6. まとめ

本研究では，誰でも，手軽に，画像の印象に合う楽曲を生成できる方法について議論した．ユーザが画像を対象物領域と背景領域に手動で分け，各領域の配色を「印象語」に変換し，楽曲を半自動的に生成するシステムを構築した．主観評価実験を実施したところ，概ね画像の印象に合う楽曲を生成できたといえる結果が得られた．

今後は，5 章で述べたような改善を行うとともに，より多くの被験者に対して主観評価実験を行い，システムの有効性を検証する予定である．また，文献 [11] のように，動画像に適用できるように拡張したい．

謝辞　本研究の一部は，日本学術振興会の科学研究費補助金 16K00501, 17H00749 による．

参考文献

[1] 前田 和博, 齋藤 康之, “カラー画像からのパラメトリック楽曲生成”, 映像情報メディア学会 メディア工学研究会, vol.ME-2010-68, pp.77–80, Feb. 2010.
[2] 根本 彩恵, 齋藤 康之, “カラー画像の印象にマッチした楽曲の半自動生成に関する研究”, 情報処理学会 音楽情報科学研究会, vol.2016-MUS-111, no.28, pp.20–27, May 2016.
[3] 社団法人 日本流行色協会, “色のイメージ辞典”, 同朋舎出版, 1991.
[4] 山崎 晃男, “音楽と感情についての心理学的研究”, 大阪樟蔭女子大学 人間科学研究紀要, 8, pp.221–232, 2009.
[5] Hevner, K., “Expression in music: A discussion of experimental studies and theories”, Psychological Review, vol.42, pp.186–204, 1935.
[6] “WSJ - Good! よいホームページを創ろう講座 5.3”, 入手先 ⟨http://www.wsj21.net/ghp/ghp0c_03.html⟩
[7] “BPM についての簡単な説明”, 入手先 ⟨http://www14.plala.or.jp/nekokirin/02aboutbpm/01aboutbpm.html⟩
[8] “OpenCV”, 入手先 ⟨http://opencv.jp/⟩
[9] “臥竜公園”, 入手先 ⟨http://toriton.blog2.fc2.com/blog-entry-2980.html⟩
[10] 井上 誠喜, “画像合成のための対象物抽出法”, 信学論 D–II, vol.J74-D-II, no.10, pp.1411–1418, Oct. 1991.
[11] 清水 柚里奈, 菅野 沙也, 伊藤 貴之, 嵯峨山 茂樹, 高塚 正浩, “動画特徴量からの印象推定に基づく動画 BGM の自動生成”, 情報処理学会 第 78 回全国大会講演論文集, vol.2016, no.1, 2Q-01, pp.2-447–2-448, Mar. 2016.