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

大山 喜冴 [1)]　　伊藤 貴之 [2)]

1)　お茶の水女子大学大学院人間文化研究科
2)　お茶の水女子大学理学部情報科学科
E-mail: {kisa, itot}@itolab.is.ocha.ac.jp

概要

映画やテレビ CM の製作、およびマルチメディア技術において、画像の印象に合った音楽を用いることは非常に重要である。そこで我々は、ユーザーが任意の音楽と画像を入力した際に、画像の印象に合わせて音楽を自動アレンジする手法の研究を進めている。本報告ではその初期成果として、画像の色分布および対象物からその印象を推測し、その印象に合ったリズムパターンで音楽を自動アレンジする手法を提案する。

本手法では前処理として、ユーザーに多数のサンプル色またはサンプル画像を提示し、これらのサンプルの各々から連想されるリズムパターンを回答させる。それと並行して、画像に写る対象物をキーワードとして提示し、この各々から連想されるリズムパターンを回答させる。この回答結果から本手法では、各々のリズムパターンに対して、どのような画像から高い連想度を得られるか、を推測する算出式を導出する。続いて本処理では、任意の入力画像に対して、各々のリズムパターンの連想度を算出し、最も連想度の高いリズムパターンを用いて自動アレンジした音楽をユーザーに提示する。

本報告では、複数の被験者を対象として、本手法によって得られたアレンジ結果の満足度を検証した。その結果として、サンプル色を提示するよりも、サンプル画像を提示するほうが、入力画像と音楽自動アレンジ結果に被験者間の相関性が高くなり、実験結果として有効である傾向が現れた。また、被験者が持つ画像に対する印象には、色から影響されやすい被験者と、対象物から影響されやすい被験者と、大きく二つに分かれる傾向が現れた。そのため被験者の傾向によって算出式の係数を調整することで、より有効な実験結果が得られた。

1. はじめに

映画やテレビＣＭの制作において、画像と音楽は密接な関係にある。また例えば個人でも、「自分で制作したホームページに印象の合った音楽を載せたい」という感想は、多くの人が一度は持つような感想だろう。しかし音楽に精通していない人は、画像に印象の合う音楽を選べない場合があり、画像と音楽の相関性のない一見不釣合いなホームページをつくる場合が多い。

この問題を解消するために我々は、ユーザーが任意の音楽と画像を入力した際に、画像の印象に合わせて音楽を自動アレンジする手法「DIVA: Digital Image Varies Arrangement」の研究を進めている。DIVA は以下の 2 段階の処理により、任意画像の印象に合わせた音楽自動アレンジ手法を特定するものである。

[前処理] サンプルとなる視覚情報と音楽情報を被験者に提示し、その印象をユーザーに回答させることで、そのユーザーにおける画像と音楽の印象の関連性を数式化する。

[本処理] 任意の画像と音楽を入力したときに、その画像に対してユーザーがもつ印象を計算機が類推し、それに合うように音楽を自動アレンジする。

本手法では、サンプル画像や任意画像に実写画像を仮定し、その色分布から画像の印象を推定する。またあらかじめ、同じメロディに対して数種類のリズムパターンを用いてアレンジされた楽曲を用意する。この前提のもとで本手法の[本処理]では、以下のような手順によって、画像の印象に合わせた音楽アレンジ方法を特定する。

- 画像から色分布を算出し、その色分布から各々のリズムパターンの連想度を算出する。
- 画像に写る対象物(「海」「山」等)をキーワードとして提示し、その対象物から各々のリズムパターンの連想度を算出する。
- 上記 2 種類の連想度の合計値が最も高いリズムパターンを用いて、音楽を自動アレンジする。

このような技術を確立するために本報告では、以下の 2 種類の前処理を実装・実験している。

[前処理 1] 所定のサンプル色（66 色）および、メロディは同じだがリズムパターンが違う音楽を用意する。被験者にサンプル色を提示して、サンプル色か

ら連想されるリズムパターンを選んでもらう。この回答結果から、各々のサンプル色に対するリズムパターンの連想度を算出する。それに加えて、風景画像に写る対象物(「山」「木」「海」等)の単語のみから連想されるリズムパターンを選んでもらい、これを対象物に対するリズムパターンの連想度とする。

[前処理 2] 数枚のサンプル画像および、メロディは同じだがリズムパターンが違う音楽を用意する。被験者にサンプル画像を見てもらい、そのサンプル画像から連想されるリズムパターンを選んでもらう。さらに、サンプル画像における各サンプル色の重要度を算出する。この算出結果と回答結果から、各々のサンプル色に対するリズムパターンの連想度を算出する。それに加えて[前処理 1]と同様に、風景画像に写る対象物(「山」「木」「海」等)の単語のみから連想されるリズムパターンを選んでもらい、これを対象物に対するリズムパターンの連想度とする。

本報告では、以上の処理によって実現された自動アレンジに対して、被験者の満足度を調査することで、本研究の有効性を検証している。

2. 関連研究

本報告の提案手法に類似した研究として、画像に合った音楽を検索できるシステムの研究[1]がある。また、音楽と画像の持つ印象をマッチングにする手法として、通常の検索に加え感性検索も可能とするマルチメディア感性データベース管理システム[2]や、言語情報と画像情報に関わる自然言語を検索キーとしたシステムの研究[3]などがある。しかし、入力画像の印象に合う楽曲が、必ずしもシステムに登録されているとは限らない。また、このシステムにおいて選曲された楽曲が、ユーザーの好きな作曲者・演奏家による楽曲である保証はない。これとは別に、画像に合った音楽の自動作曲システムの研究[4]が報告されている。

しかし、ユーザーの好みのメロディをあらかじめ指定し、これを画像の印象に基づいて自動アレンジする研究は、我々が調べた限り見つかっていない。

3. 提案手法の概要

3.1 基本構想

本報告で提案する「画像の印象に合わせて音楽を自動アレンジする手法」は、1 章で論じた[前処理][本処理]の 2 段階処理によって、音楽を自動アレンジする手法であると定義づけられる。

一方、音楽や画像には以下のような多種の構成要素がある。音楽や画像の印象は、これらの構成要素と大きな関連性がある。

[音楽の構成要素] 「調(長調/単調)」「テンポ」「旋律(上昇/下降)」「音高(高/低)」「和声(単純/複雑)」「リズム(固定/流動)」「歌詞の内容」など。

[画像の構成要素]「色彩」「構図」「対象物」「場面設定」など。

音楽と画像の印象を計算機が類推するためには、これらの構成要素すべてを計算機が分析することが望ましいが、現実的にはそれは難しい。そこで現時点での実装では、上記の多種の構成要素の中から、人間がもつ印象を特に大きく左右する構成要素だけを用いている。

3.2 現時点の実装の概要

現時点での実装では、前節で紹介した音楽の構成要素のうち、「リズム」を用いている。また前節で紹介した画像の構成要素のうち、「色彩」と「画像に描かれている対象物」を用いている。以下に、なぜこれらの構成要素が最も重要であると判断したか、について論じる。

論文[3]では、音楽の印象と色彩の印象に相関性があることを実証する実験結果が示されている。例えば色相と音楽の印象の相関性には、「赤には迫力のある音楽が似合う」「緑には明るい音楽が似合う」という回答が多いとの結果が出ている。また、明度や彩度と音楽の印象の相関性には、「明度が高くなるに連れて音楽を明るく感じる傾向にある」「彩度の高低と音楽の力強さの度合いが対応する傾向にある」という回答が多いとの結果が出ている。このことから画像の持つ「色」が音楽に与える影響は大きいと考えられる。このことから現時点での実装では、画像の色彩から主に、その印象を類推することにした。また文献[5]においても、画像の持つ印象は主に配色が重要である、という実験結果が報告されている。このことから本研究では、画像の特徴のうち主に配色に着目した。

しかし、それでも色単体のみから画像の印象を特定するのは不確実であると考えられる。例えば、赤い花と赤い夕日では、同じ色分布を有する画像であっても印象が大きく変わることが多い。そこで本研究では、画像に描かれている対象物(「海」「山」等)を、キーワードとしてあらかじめ提示する手法を併用した。

次に、楽曲の構成要素について論じる。文献[4]では、この構成要素の中でも「リズム」→「旋律」→「和声」→「音高」の順で印象に残りやすく、楽曲を特徴付ける大きな要素となっていると述べている。このことから現時点での実装では、音楽の構成要素

のうちリズムパターンを差し替えることで自動アレンジを行う。他の構成要素は、現時点ではアレンジには用いていない。

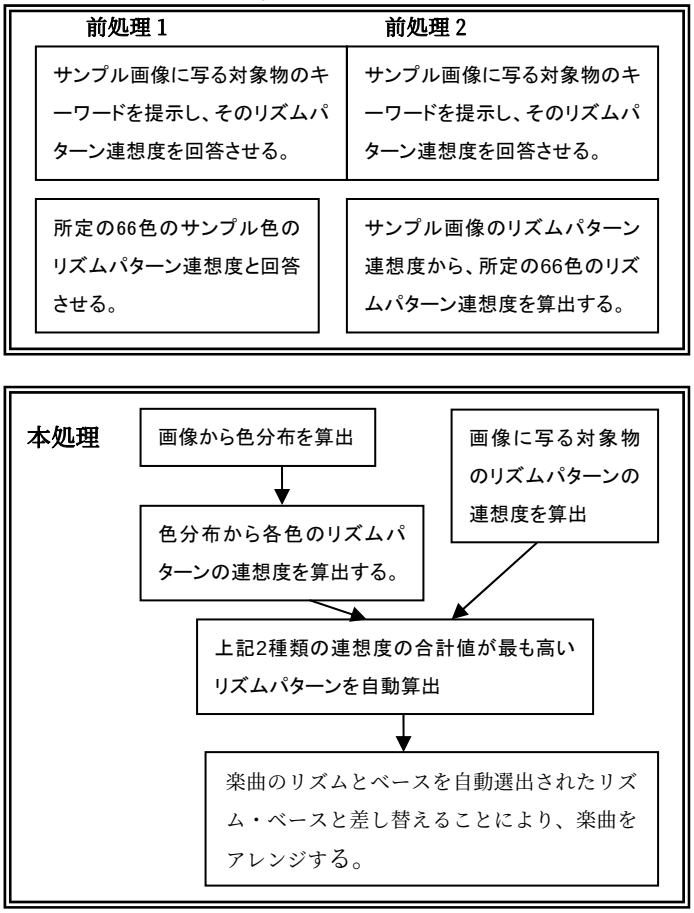

図 1：本手法の概要

以上の前提によって構築された、現時点での実装の概要を、図 1 に示す。現時点での実装では、二種類の前処理のいずれかを用いて、画像を構成する色と対象物からのリズムパターン連想度を算出する。続いて本書では、入力された任意の画像の色分布を算出し、次に画像に描かれている対象物(「海」「山」等)から連想されるリズムパターンを算出する。そして、色分布から連想されたリズムパターンと、画像に描かれている対象物から連想されるリズムパターンの二つを用いて、この画像からの連想度が高いと思われるリズムパターン自動選択し、このリズムパターンを用いて音楽を自動アレンジする。

4. 提案手法の実装

4.1 サンプル色の選定

筆者らの現時点での実装では、66 色のサンプル色を規定し、この 66 色の各々に対するリズムパターンの連想度を算出する。筆者らの実装における 66 色の内訳は以下の通りである。まず有彩色を HSV 色空間に配置し[6]、その色相を 7 段階、彩度および明度を 3 段階に分割する。以上の処理により、HSV 空間を 66 個の部分空間に分割し、各々の部分空間内部に 1 個ずつサンプル色を選ぶ。以上によって 66 色のサンプル色を選定する。なお色相の分割数を決定する際には、太陽光が分光して見られる虹の色彩を 7 色で表現する文化[7]がある、という事実を参考にした。

66 色にポスタライズした画像の例を、図 2 に示す。著者らは、この 66 色によるポスタライズ結果と、さらに少ない色数でポスタライズした結果を主観比較した結果、66 色では画像の印象をほぼ損なわないが、さらに減色すると画像の印象を損ないやすくなる、と判断した。以上の判断結果により著者らは、ポスタライズの色数を 66 色と決定した。

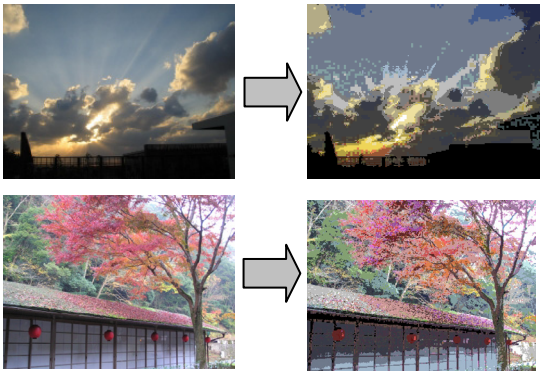

図 2：66 色の妥当性の検証例

本研究では 66 色のサンプル色それぞれに対して、リズムパターンの連想度を算出する。以下、その理由を論じる。文献[8]では、色相が与える印象の平均的傾向が記されている。しかし一方で文献[9]では、色相が同じ色であっても、明度・彩度が異なると、心理的にそれぞれ違う影響も与えることも報告している。このことから本研究では、色相・彩度･明度の異なる 66 色のサンプル色それぞれに対して、リズムパターンの連想度を算出する必要がある、と考えた。

4.2 色彩対比を利用した色重要度の算出

本手法では、画像の印象を類推するために「色重要度」という概念を導入している。色重要度とは、その色が画像の印象を支配する度合いを推定する数値である。本手法では色重要度を、

- 画像中に占める面積
- 隣り合う色との差分
- 色自体の持つ印象

により算出する。

現時点の実装では、論文[9]に紹介される色彩対比のうち、以下の二項目を実装することで、「隣り合う色との差分」「色自体の持つ印象」を加味して色重要度を算出する。

[明度対比] 暗い色に囲まれた色の方が明るく見える現象。この現象を色重要度に反映するために、現時点での実装では、面積が小さいにも関わらず、周囲と比較して極端に明度が高い領域において、その色彩の重要度をあげる。本手法では、この条件に該当するとき b>1,それ以外のとき b=1 となる変数 b を導入する。

[進出・後退] 暖色系が進出し手前に見え、寒色系の色が後退して見える現象。この現象を色重要度に反映するために、現時点での実装では、暖色は寒色に比べ進出して見えるため重要度を上げている。本手法では、この条件に該当するとき p>1,それ以外のとき p=1 となる変数 p を導入する。

以上の二項目を用いて本手法では、66 色に減色された画像に対して、i 番目のサンプル色の色重要度 C_i を、以下の式で算出する。

$$C_i = \sum_{i=1}^{N} b_i p_i f_i \quad (1)$$

ここで N は画像の総画素数、b_i および p_i は i 番目の画素における b および p の値、f_i は i 番目の画素の画素値が i 番目のサンプル色であれば f_i =1、さもなければ f_i =0 となる二値変数である。

4.3 画像に描かれている対象物の印象

画像全体から色分布をあらわす特徴量のみを抽出しただけでは、画像の雰囲気を特長付けるのは不十分であると考える。そこで本研究では、画像に表現されている「物体」に関する特徴量も算出する。以下、その理由を論じる。

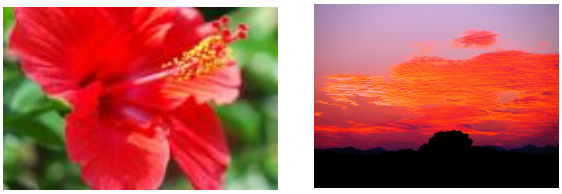

図 3：赤い画像の比較

赤い対象物を含む 2 種類の画像を、図 3 に示す。この図を見てもわかるように、赤い花の画像と、赤い夕焼けの空の画像では、同じ赤でも被験者が受ける印象は大きく変わる可能性がある。そこには「花」または「夕焼け」という、個々の対象物が持つ印象が大きく関係していると考えられる。このことから本研究では、これらの印象の違いを反映するには、画像に写る対象物（例えば「海」「山」等）から連想されるリズムパターンを算出する必要がある、と考えた。

4.4 前処理：サンプル色・対象物とリズムパターンの連想度の算出式の導出

本手法では前処理として、視覚情報を被験者に提示し、その視覚情報から連想されるリズムパターンを回答させる。この回答結果をもとに本手法では、サンプル色とリズムパターンの連想度の算出式をたてる。以下、本研究で実装している 2 種類の前処理について論じる。

4.3.1 前処理 1

前処理 1 では、画像に写る対象物をキーワードとして提示し、これに対するリズムパターンの連想度を、被験者に回答させる。ここで本報告では、k 番目のキーワードに対する j 番目のリズムパターンの連想度を、Q_{kj} と記述する。

それと同時に前処理 1 では、所定の 66 色のサンプル色を提示し、同時に複数のリズムパターンを被験者に提示する。続いて、各サンプル色に対するリズムパターンの連想度を、被験者に回答させる。ここで本報告では、i 番目のサンプル色に対する j 番目のリズムパターンの連想度（回答結果から得られる値）を、R_{ij} と記述する。

4.3.2 前処理 2

前処理 2 では、前処理 1 と同様に、画像に写る対象物のキーワードに対するリズムパターンの連想度 Q_{kj} を用いる。

それと同時に前処理 2 では、十分な枚数のサンプル画像を提示し、同時に複数のリズムパターンを被験者に提示する。続いて、各サンプル画像に対するリズムパターンの連想度を回答させる。続いて、この回答結果から、66 色の各サンプル色に対する各リズムパターンの連想度 R_{ij} を算出する。

ここで、1 枚の画像に対する各リズムパターンの連想度の算出式を考える。ここで、1 枚の画像に対する j 番目のリズムパターンの連想度（回答結果から得られる値）を、P_j とする。また本手法では、画像を 66 色に減色したときの、画像中の i 番目のサンプル色の重要度を、式(1)から算出した結果として C_i と記述する。このとき P_j と R_{ij} には、以下の式が成立する。

$$P_j = \sum_{i=1}^{66} C_i R_{ij} \quad (2)$$

ここで R_{ij} を求めるために、式(2)を以下のように変形する。合計 m 枚のサンプル画像のうち、k 番目の画像の i 番目のサンプル色の重要度を C_{ki} とする。ま

た、k 番目の画像に対する j 番目のリズムパターンの連想度（回答結果から得られる値）を P_{kj} とする。このとき

$$\mathbf{P}=\begin{bmatrix} P_{1j} \\ \vdots \\ P_{mj} \end{bmatrix}, \mathbf{C}=\begin{bmatrix} C_{11} & \cdots & C_{1i} \\ \vdots & & \vdots \\ C_{m1} & \cdots & C_{mi} \end{bmatrix}, \mathbf{R}=\begin{bmatrix} R_{1j} \\ \vdots \\ R_{ij} \end{bmatrix}$$

とすると、

$$\mathbf{P}=\mathbf{CR} \quad (3)$$

が成立する。このとき 66 枚のサンプル画像に対する回答結果があれば、連立方程式を解くことで R_{ij} の各々の値を得ることができる。しかし 66 枚のサンプル画像に対する回答が必要だとすると、これは被験者に対する負担が大きい。そこで本手法では、式(3)を以下のように変形する。

$$\mathbf{R}=\mathbf{C}^{-1}\mathbf{P} \quad (4)$$

これに m 枚の画像における P および C の値を代入することで、サンプル画像が 66 枚未満である場合にも、j 番目のリズムパターンに対する R_{ij} 値の近似値を得ることができる。このような近似解算出手法は、本手法の目的以外にも実績のある解法である。例えば 3 次元 CG の代表的なレンダリング手法であるラジオシティ法においても、漸近的かつ高速な照度算出のために、これと等価な解法が用いられている。

4.4 本処理：任意の入力画像に対する音楽の自動アレンジ

4.4.1 任意の入力画像に印象の合うリズムパターンの自動選定

本処理では任意の画像を入力すると、まずその画像を 66 色に減色し、各サンプル色の色重要度 C_i を算出する。また R_{ij} および Q_{kj} は前処理で算出された値をそのまま用いる。画像に写る対象物のキーワードを n 個であるとすると、本処理では以下の式

$$P_j = a\sum_{i=1}^{66} C_i R_{ij} + b\sum_{k=1}^{n} Q_{kj} \quad (5)$$

により、入力画像に対する j 番目のリズムパターンの連想度 P_j を導出する（ただし a,b は任意の正実数）。本処理では、この値が最大であるリズムパターンを選ぶことで、入力画像に印象が合うと思われるリズムパターンを特定する。

4.4.2 音楽の自動アレンジ

我々の現時点での実装では、音楽情報に GM (General MIDI) を想定し、メロディ、和音、ベース、リズムの 4 パートが所定のチャンネルに記録された SMF (Standard MIDI File) 形式の楽曲ファイルを用いる。画像の印象に合うリズムパターンを特定すると、我々の実装は、それにしたがってリズムとベースを差し替えることにより、楽曲をアレンジする。

なお現時点では単純のために、アレンジ時にメロディ・和音のパートを一切変更しない。これらをリズムパターンに応じて操作する実装ができれば、より効果的な音楽アレンジが実現できると考えられる。

5. 実行結果

5.1 実験方法

本研究ではまず準備段階として、同じメロディ A に対して異なる 7 種類のリズムパターンを適用した 7 曲を、被験者に鑑賞させた。続いて以下の 3 種類の実験を用い、式(5)によりリズムパターンの連想度 P_j を求めた。実験 1～3 で使用された楽曲は一貫して、著者の一人である伊藤によって作曲された物である。

[実験 1] 所定の 66 色を被験者に提示し、各々の色からどのリズムパターンを連想したかを被験者に回答させた。この回答結果からリズム連想度 R_{ij} を得た。

[実験 2] サンプル画像 25 枚を被験者に提示し、各々の画像からどのリズムパターンを連想したかを被験者に選択させた。この回答結果から式(4)を用い、リズム連想度 R_{ij} を得た。

[実験 3] [実験 2]に加えて、画像に写る対象物のキーワードを被験者に提示し、各々のキーワードからどのリズムパターンを連想したかを回答させた。この回答結果からリズム連想度 Q_{kj} を得た。

[実験 1]～[実験 3]のいずれにおいても、本処理では任意の画像を入力した際に、式(5)を用いてその画像から各リズムパターンへの連想度 P_j を算出し、この値が最大であるリズムパターンを用いて楽曲をアレンジした。ただし[実験 1][実験 2]では、式(5)において b=0 とした。筆者らの実験では本処理において、様々な風景が写った、サンプル画像とは異なる 25 枚の画像を用いた。

以上の実験結果の一部は、以下の URL に掲載されている。http://ito.is.ocha.ac.jp/kisa/diva1/

5.2 実験結果の分析

本研究では、提案手法の本処理によるリズムパターン選択結果の妥当性を検証するために、本処理で用いた画像を被験者にも提示し、被験者が自分で選択したリズムパターンと比較し、両者の一致率を算出した。この結果を以下に記述する。

前処理に[実験 1]を採用した際には、本処理の結果と被験者の回答との一致率は 28%であった。前処理に[実験 2]を採用した際には、本処理の結果と被験者の回答との一致率は 48%であった。このことから、

筆者らの実験では、リズムパターンの自動選出に画像の色分布のみを参照するのでは不十分であり、後述するように画像中に写る対象物の種類も参照するほうが望ましいことがわかった。

なお前述の実験結果からは、前処理において被験者に、色を提示するよりも画像を提示するほうが、良好な結果が現れている。ここで、同一の色や画像に対する被験者間の回答の相関性について観察した。すると、[実験 1]の回答は、被験者ごとにあまり似通っていなかったのに対して、[実験 2]の回答は被験者ごとに似ている傾向があった。この事実からも、[実験 1]より[実験 2]のほうが安定した結果が得られそうであることがわかった。

続いて[実験 3]を行い、その過程において被験者ごとに回答を観察した。その結果、

[傾向 1] 色とリズムパターン連想度の相関性が比較的大きい被験者

[傾向 2] 対象物とリズムパターン連想度の相関性が比較的大きい被験者

の 2 グループに分かれる傾向が顕著に現れた。そこで[実験 3]では、式(5)の係数 a,b について、[傾向 1]の被験者に対しては a>b, [傾向 2]の被験者に対しては a<b となるように、a,b の値を使い分けた。その結果、一致率は 71%に上昇した。以上の結果により、リズムパターンの自動選出には、画像の色分布だけでなく、画像に写る対象物の種類も参照したほうが望ましい、ということがわかった。

続いて、本実験で用いたリズムパターンの特徴と、実験結果の関係について論じる。リズムパターンに対する被験者間の回答の相関性について観察したところ、各リズムパターンの選択される回数に、被験者間の偏りの大きいものと小さいものが見られた。選ばれる回数の少ないリズムパターンは、もともと入力されたメロディに対してアレンジしにくいリズムパターンである、ということも想定される。

続いて、本実験で用いた画像の特徴と、実験結果の関係について論じる。 [実験 1][実験 2]において、被験者の回答との一致率の高い画像は、彩度や明度の低い色の重要度が高い画像であった。反対に明度・彩度が高い色の重要度が高い画像では、一致率が極端に低かった。逆に[実験 3]において b 値を大きくした場合には、「海」「月」など、色との相関性が高いキーワードをもつ画像の一致率が高かった。反対に「花」などのように、色との相関性の低いキーワードをもつ画像の一致率は低かった。

6. まとめ

本論文では、画像の色分布と、画像に写る対象物の種類から連想されるリズムパターンを用いることで、与えられた画像に印象が合うように音楽の自動アレンジ手法を特定する手法を提案した。また筆者らの実験においては、本手法の前処理では被験者に色を提示するより実写画像を提示するほうが良好な結果が得られること、また画像の色分布だけでなく画像に写る対象物の種類も考慮したほうが良好な結果が得られること、がわかった。

今後の課題として

- さらに多くの被験者を対象とした実験
- 楽曲の持つ雰囲気も考慮し、より多彩な楽曲にアレンジができるようなリズムパターンの選定
- リズムパターン以外の構成要素（メロディ、和音、音色）を変化させる音楽の自動アレンジ
- 学習アルゴリズムなどの適用により、よりユーザーの好みを反映するシステムの研究開発

などに着手していきたいと考えている。

謝辞

本研究の被験者の方々に感謝の意を表します。

参考文献

[1] 古賀, 下塩, 画像に合った音楽の選定技術, ヒューマンコミュニケーション基礎研究会技報, 平 11-9, 1999.

[2] 坂井, 大塚, 宮崎, マルチメディア感性データベース YAMAKAN, 第 13 回データ工学ワークショップ (DEWS2002), 2002.

[3] 安達, 岩宮, 色彩と音楽が互いに及ぼす影響--ショパンのエチュードを手がかりに, 第5回学生のための研究発表会講演論文集(日本音響学会九州支部), pp.13-16, 2003.

[4] 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
http://www.forest.impress.co.jp/article/2003/04/21/okiniiri.html

[5] 感性(印象語)語による検索
http://www.slis.keio.ac.jp/~ueda/semi/99onsei.html

[6] 原田, 感性語句を用いた自然言語文による画像データベースの対話的検索, 静岡大学博士論文, 工博甲第 175 号, 1995.

[7] 大林, 銀河の道 虹の架け橋, 小学館, 1999.

[8]色について http://park11.wakwak.com/~d-o-b/color/

[9]色の見え方
http://www.colordream.net/taihi.htm