

# 色彩と音楽のクロス・メディア連結機構による 自動音楽生成システムの実現

春日優希<sup>†</sup> 倉林修一<sup>‡</sup> 清木康<sup>†</sup>

<sup>†</sup> 慶應義塾大学環境情報学部 〒252-0805 神奈川県藤沢市遠藤 5322

<sup>‡</sup> 慶應義塾大学大学院政策・メディア研究科 〒252-0805 神奈川県藤沢市遠藤 5322

E-mail: <sup>†</sup> <sup>‡</sup> { t07184yk, kurabaya, kiyoki }@sfc.keio.ac.jp

**あらまし** 本稿では、色彩と音楽を同一の計量空間上に写像し、色彩の時系列推移に応じて、色彩と音楽間の相関量計量により色彩の印象推移に応じた音楽を自動的に検索、生成するシステムを提案する。本システムの特徴は、音楽理論における音楽構造分析の手法、音楽心理学、および、色彩心理学の研究成果をメディア分析・連結関数群として実装し、それらの関数群を用いて、色彩推移に応じた楽曲データの自動生成を実現する点にある。本システムの利用者は、映像（静止画像、動画像）を入力し、それら映像メディアデータがあらゆる色彩印象に応じた自動作曲結果を得ることが出来る。本稿では、色彩と楽曲を対象とした音楽のクロス・メディア連結機構自動作曲システムの実現方法を示す。

**キーワード** 音楽データベース、印象検索、クロスメディア、マルチメディアデータベース

## A Cross-Media Association Engine for Automatic Music Generation

Yuki KASUGA<sup>†</sup> Shuichi KURABAYASHI<sup>‡</sup> and Yasushi KIYOKI<sup>†</sup>

<sup>†</sup> Faculty of Environment and Information Studies, Keio University 5322 Endo, Fujisawa, Kanagawa, 252-8520 Japan

<sup>‡</sup> Graduate School of Media and Governance, Keio University 5322 Endo, Fujisawa, Kanagawa, 252-8520 Japan

E-mail: <sup>†</sup> <sup>‡</sup> { t07184yk, kurabaya, kiyoki }@sfc.keio.ac.jp

**Abstract** A large number of music resources exist in the world-wide scope. This paper presents a novel music database system equipped with a musical knowledge base that assists database processing automatically by providing formal intelligence about music. The music knowledge base system realizes a simple and intuitive mechanism for general users to retrieve music compositions. One of the most significant advantages of this system is that it supports various combinations of media analysis functions that are implemented in the knowledge base. This system makes it possible to express users' intensions to retrieve music data as a combination of query sequence and its interpretation methods.

**Keyword** music database, impression-based retrieval, cross-media, multimedia database

### 1. はじめに

今日、デジタル・メディア・コンテンツのネットワークを通じた流通の普及に伴い、異種のメディアデータを横断的に利用するクロス・メディア生成環境を構築が活発である。色彩心理学、および、音楽心理学の分野では、色彩と色調感覚には強い関連があることが音刺激によって色覚を伴う共感覚として知られており、色彩と音楽のクロス・メディア連結の実現は、新たなメディア利用の手法として期待されている。色彩と音楽の関連を計量するためには、楽曲中の時間軸に沿って表現されたデータの物理的な特徴量や、感性的な印象、構造的な意味を、利用者が楽曲を検索・鑑賞・理解する際の意図や視点に応じて計量する機構の実現が重要である。そこで、本稿では、色彩情報と音楽を同

一の計量空間上に写像し、色彩の時系列推移に応じて、色彩と音楽フレーズ間の相関量を計量することにより、色彩の印象推移に応じた音楽を自動的に検索、生成するシステムを提案する。本システムの利用者は、映像（静止画像、動画像）を入力し、それら映像メディアデータがあらゆる色彩印象特徴量に応じた自動作曲結果を得ることが出来る。

本システムの特徴は、音楽理論における音楽構造分析の手法や、音楽心理学、および、色彩心理学の研究成果を反映したメディア分析・連結関数群を用いて、色彩に相関の高い楽曲フレーズを選択し、それらを時系列に沿って組み合わせ、楽曲データの自動生成を実現する点にある。本システムの利用者は、映像（静止画像列、動画像）を入力し、それら映像メディアデー

タがあらわす色彩印象特徴量に応じた自動作曲結果を得ることが出来る。このメディア分析関数群を組み合わせる方式の利点は、1) 楽曲調性に応じた特徴量生成、2) 色彩と音楽（調性）との相関量計量、3) 色彩の時間的变化に応じた楽曲の自動生成に関する3機能を実現する点にある。本音楽生成データベースシステムでは、それらの機能を本方式による音楽自動生成の重要な3機能として実現する。

## 2. 色彩と音楽のクロス・メディア連結機構による自動音楽生成システム

本システムは、1) 色彩感性空間、2) 色彩空間への印象語および色彩のマッピング・テーブル、3) 音楽感性空間、4) 時系列色彩推移に対応する音楽生成機構の4サブシステムから構成される。図1に、本方式の概要を示す。

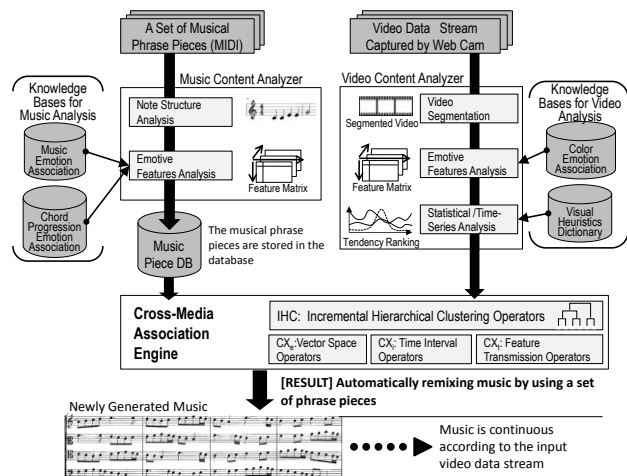


図1 システム概要図

### 2.1. 空間生成

#### 2.1.1. 色彩特徴の設定による色彩感性空間の構築

本システムでは、動画データから静止画像群として分析する。動画から静止画像へのフレーム分析は、コーデックを用いて1秒1フレームを抽出する。静止画像から色彩印象抽出のルールは、カラーイメージスケールをもとに、マンセル表色系の代表130色を用いた配色セット（カラースキーマ）を作成した。カラーイメージスケールによると、マンセル代表130色の1色ずつが、複数の様々な印象を有している。そのうちの1つの印象に着目し、その印象を有するマンセル色を、ひとつのカラースキーマとして作成する。それにより、計183セット作成した。そのため、静止画像からの色彩情報抽出は、マンセル表色系を参考し、有彩色120色（図3）、無彩色10色の計130色のみを用い、上述ステップでは静止画像をRGBモデルからHSVモ

デルに変換している。なお、本システムでは、静止画像中の色彩を、上述のとおり色彩の印象（語）に変換しているが、演算や検索の段階では、その印象を直接的に利用している機能はない。ストーリーの変化をたどる要素として、特に制作者により意図的に彩色されたアニメーションなどにおいて、動画データ中の色彩変化を計ることは、ストーリーの変化としてたどるに非常に有意な情報と考えるが、一方で、色彩の持つ印象が、静止画像、および、動画データのコンテキストと必ずしも合致するわけではないからである。



図2 配色と色彩印象の関係の定義

#### 2.1.2. 音楽の色彩空間へのマッピング

楽曲のMIDIデータから調性を自動的に判定するために、調性自動判定アルゴリズムであるKrumhansl-Schmucklerアルゴリズム[4][5]を用いる。4節で述べる実装システムでは、Krumhansl-Schmucklerアルゴリズムを実装している。調の基本は音階であるため、Krumhansl-Schmucklerアルゴリズムは、音高(ピッチ)情報から調を推測する。具体的には、長調・短調における各調性のプロファイルを統計により算出し、楽曲中の音高のヒストグラムから、最も長い時間鳴っている音高を調性とする。ただし、音高名には、C,D,E,F,G...cdefg という順番で、次のような重みがつけられている。なお、この重みは、文献[5]のサンプルプログラムより引用した。

- MAJOR\_PROFILE = {6.35, 2.23, 3.48, 2.33, 4.38, 4.09, 2.52, 5.19, 2.39, 3.66, 2.29, 2.88}
- MINOR\_PROFILES = {\$6.33, 2.68, 3.52, 5.38, 2.60, 3.53, 2.54, 4.75, 3.98, 2.69, 3.34, 3.17}

#### B. Bishop

c	g	d	a	e	b	f#	c#	ab	eb	bb	f
C	G	D	A	E	B	F#	C#	Ab	Eb	Bb	F

#### A.N. Skryabin

c	g	d	a	e	b	f#	c#	ab	eb	bb	f
C	G	D	A	E	B	F#	C#	Ab	Eb	Bb	F

図3 調性と色彩の関係を定義した調性と色彩のマッピング・スケール

本方式は、楽曲における時系列的・構造的な印象の変

化を、色彩の変化（グラデーション）として表現するために、調性-色彩メタデータを導入する。調性-色彩メタデータは、調性から生じる印象の時系列的な変化を連続的な色彩の変化に変換したものである。楽曲の調性情報から調性-色彩メタデータを生成するため、調性と色彩の関係を定義した調性と色彩のマッピング・スケール（図 3）を用いる。心理学の分野では、聴覚と色覚に密接な関係が存在することが、共感覚として知られており、本方式におけるカラーマッピング・スケールは、共感覚を有する人物による定義を利用したものである。具体的には、作曲家である A.N. Skryabin によるマッピング・スケール[2][3]や、B.Bishop によるマッピングスケール[3]を実装している。

調性の五度圏（Circle of Fifth：図 4）は、全 24 種類の調性間の近さや類似性といった関係を定義したモデルである[6]。本システムでは、この五度圏を調性の変化によって引き起こされる、印象の変化を計量するメトリクスとして利用する。五度圏は環状に配置された調性の移動により、時計回りに調性が移動した場合エネルギーが上昇し、反時計回りはエネルギーが減少する。また、短調(minor)への移動は暗い印象となり、長調(Major)への移動は明るい印象となる。

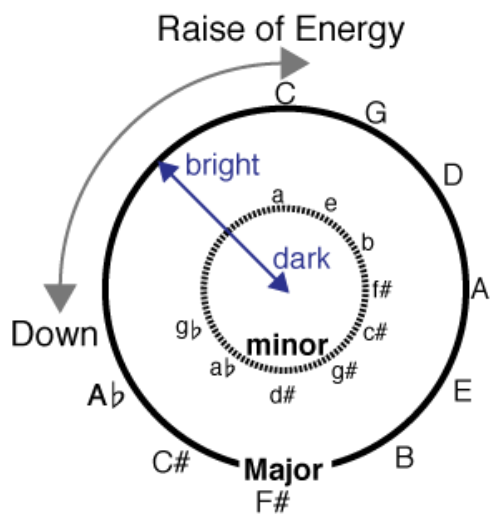


図 4 調性の五度圏

## 2.2. 音楽シーケンスの自動生成

PG Music 社の Band-in-a-Box を用いて、音楽フレーズを自動生成した。Band-in-a-Box は、調性列を入力として、その調性列を満たす楽曲を自動生成する機能を有する。本システムでは、自動生成機能を用いて獲得した音楽フレーズ軍を対象として、図 5、図 6 に示す色彩・音楽調性変換マトリクスを適用し、利用者から入力された静止画像列と、生成した各音楽フレーズとの相関量を計量し、最も相関の強いフレーズ群を合成し、新たな楽曲を生成する。

## 2.3. プロトタイプシステム

本節では、プロトタイプシステムの実装について述べる。実装には Java 言語を用いた。本節では、既に実装の完了している、音楽データを対象とした調整分析と、その可視化システムを示す。可視化動作中の本システムのスクリーンショットを図 5 に示す。

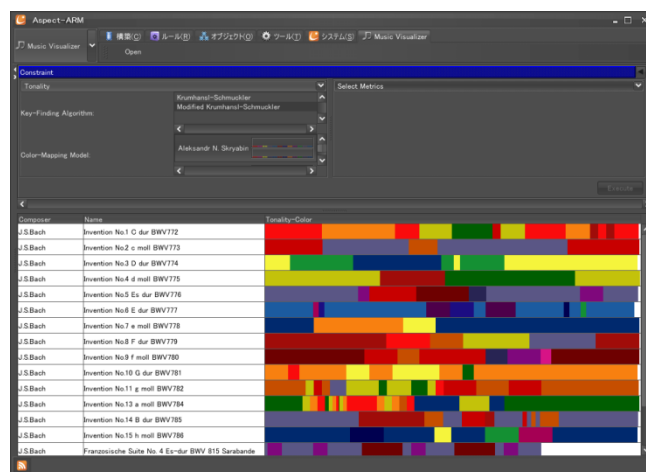


図 5 可視化動作中のプロトタイプシステム

## 2.4. 時系列色彩推移に対応する音楽生成プロセス

- (1) 時系列変化を伴う映像（静止画像列，動画像）を受け取る。
- (2) 映像における時系列色彩推移における各色彩情報を色彩感性空間にマッピングし、それぞれに相関の高い色彩を抽出する。
- (3) 抽出された時系列色彩列（色彩推移）を音楽感性空間へマッピングする。
- (4) 各色彩と楽曲フレーズ間において相関量計量を行い、各色彩に相関の高い楽曲フレーズを選択する。
- (5) 推移に沿って、それらの楽曲フレーズを時系列に沿って組み合わせ、楽曲データの自動生成を実現する。形成された楽曲フレーズ列を出力（演奏）する。

## 3. 結論

本稿では、色彩と音楽を同一の計量空間上に写像し、色彩の時系列推移に応じて、色彩と音楽間の相関量計量により色彩の印象推移に応じた音楽を自動的に検索、生成するシステムを示した。本システムは、映像を表す色彩の推移を物理量とし表現し、その色彩によって受ける感性的な印象を分析し、映像における印象推移に応じて、印象推移を反映した楽曲を自動的に生成する機能を有する音楽データベースシステムの実現方法を示した。本システムの特徴は、音楽理論における音楽構造分析の手法や、音楽心理学、および、色彩心理学の研究成果をメディア分析関数群として実装し、それら分析関数を用いて楽曲データの内容変化をデータ

