

# メディアトランスフォーメーションと音楽の感性&理論知識ベースによるメロディー作曲方式

宮本 健夫<sup>†</sup> 清木 康<sup>‡</sup>

<sup>†</sup> 慶應義塾大学総合政策学部 〒252-0882 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 〒252-0882 神奈川県藤沢市遠藤 5322

E-mail: <sup>†</sup> takeoff\_point@ksj.biglobe.ne.jp, <sup>‡</sup> kiyoki@sfc.keio.ac.jp

**あらまし** 人間の持つ感性や感情を芸術分野へと変換することは、人々の内的衝動を平和的な形で昇華する行為であり、健全な自己実現の形である。本論文で提案する作曲システムは、形容詞を介して画像から音楽へとメディア間の情報変換をすることで、全ての人へ作曲手段を提供するものだ。本システムは、まず画像ピクセルの色情報とカラーイメージスケールの距離計算を行い、画像を形容詞の数値に変換したメタデータを生成し、それから音楽のリズム、テンポ、音色、音の動き方などを定義していき、さらに音楽理論に基づき音の動きを自然に選択するための音楽理論データベースを利用し、旋律を作る。このシステムの新規性は、感性情報処理と音楽家の持つ高度な専門知識を組み合わせる点にある。それによって得られる結果は、人々の心理状態や印象を西洋音楽の理論に基づいた音楽性を持った作品として表現することが実現されたものである。

**キーワード** multi-database, image processing, music composition, Mathematical Meaning for Multimedia Information Systems, music theory, impression, semantic analysis, media-transformation, knowledge-base, melody

## 1. はじめに

抽象的な表現手法として心象を発する文化を広く根付かせることは、人々を救う。自己の内的衝動を発散するとともに、主張や不遇な状況を、不快さを抱かせない形で他者に示すことが実現されるからだ。

これは芸術家達が日常的に行なっていることである。その中でも音楽は最も抽象度の高い表現方法である。旧ソ連下で自由が極めて制限された日々の中でもショスタコーヴィチがソ連政府に対する皮肉や日々の葛藤を音楽で描き続けていたことは、内的心象の発散として作曲行為が役立つという意味においても注目に値する [1]。

このような行為が誰にでもできるように支援する機能があれば、多くの人々がもっとうまく生きていく方法を発見することができるようになる。

そこで本論文では、感性や感情を旋律として表現するシステムを提案し、内容の解説と、実行結果を示す。本システムは、人間の内面を表現するにふさわしい感性語、感情語からなる形容詞群を介して音楽に変換することを根幹部分とする。また、本論文ではインプットとして画像のピクセルデータを用いた。

## 2. 基本方式

本システムは、感性データ抽出、感情データの抽出、音楽基本データの定義、メロディの制作に分けられる。感性語はカラーイメージスケールによる客観的な形容詞群から 16 語[2]を用いるのに対して、感情語は正負の 10 の軸[3]である。これらを組み合わせて計量を行うことで音楽の持つ表現に迫ることを実現した。

また、メロディ制作はさらに細分化される。息継ぎの設定、リズムの決定、音の選択のプロセスである。

### 2.1. 感性データの抽出

本システムは画像ピクセルの色情報を利用し、カラーイメージスケールで定義された各形容詞の 3 つの色の組み合わせと HSV 方式による距離計算を行うことで、各形容詞の度合いを数値化することで画像の印象を抽出する。本システムは音楽要素に変換するためのデータ定義を実現可能にするため、利用する感性語をカラーイメージスケール内の 16 の主要クラスタ[2]の中で中央域に位置する以下の 16 語に絞った。

- ・プリティな ・ロマンチックな ・カジュアルな
- ・ナチュラルな ・クリアな ・エレガントな
- ・シックな ・スポーティな ・ダイナミックな
- ・ゴージャスな ・ワイルドな ・クラシックな
- ・重厚な ・ダンディな ・フォーマルな ・モダンな

### 2.1.1. 画像の代表色を抽出

まず、与えられた画像の各ピクセルの色情報を色相 & トーンシステム(図 1)による 130 色[2]へ近似する。そこから頻度が上位の色を 12 色ピックアップし、与えられた画像の代表色とする。

R/V	YR/V	Y/V	GY/V	G/V	BG/V	B/V	PB/V	P/V	RP/V	N/10
R/S	YR/S	Y/S	GY/S	G/S	BG/S	B/S	PB/S	P/S	RP/S	N/9
R/B	YR/B	Y/B	GY/B	G/B	BG/B	B/B	PB/B	P/B	RP/B	N/8
R/P	YR/P	Y/P	GY/P	G/P	BG/P	B/P	PB/P	P/P	RP/P	N/7
R/Vp	YR/Vp	Y/Vp	GY/Vp	G/Vp	BG/Vp	B/Vp	PB/Vp	P/Vp	RP/Vp	N/6
R/Lgr	YR/Lgr	Y/Lgr	GY/Lgr	G/Lgr	BG/Lgr	B/Lgr	PB/Lgr	P/Lgr	RP/Lgr	N/5
R/L	YR/L	Y/L	GY/L	G/L	BG/L	B/L	PB/L	P/L	RP/L	N/4
R/Gr	YR/Gr	Y/Gr	GY/Gr	G/Gr	BG/Gr	B/Gr	PB/Gr	P/Gr	RP/Gr	N/3
R/Di	YR/Di	Y/Di	GY/Di	G/Di	BG/Di	B/Di	PB/Di	P/Di	RP/Di	N/2
R/Dp	YR/Dp	Y/Dp	GY/Dp	G/Dp	BG/Dp	B/Dp	PB/Dp	P/Dp	RP/Dp	N/1
R/Dk	YR/Dk	Y/Dk	GY/Dk	G/Dk	BG/Dk	B/Dk	PB/Dk	P/Dk	RP/Dk	
R/Dgr	YR/Dgr	Y/Dgr	GY/Dgr	G/Dgr	BG/Dgr	B/Dgr	PB/Dgr	P/Dgr	RP/Dgr	

図 1: 有彩色 120 色および無彩色 10 色からなる全 130 色のマンセル基本色彩セット[2]

### 2.1.2. 感性語との距離を抽出

カラーイメージスケールで定義されている感性語 A の i 色目と、得られた代表色から 3 つを選んだ順列 P を全て生成し、その i 番目と HVS 要素による距離計算を行い、結果の和が最短だったものをその形容詞との距離と定義する。

なお、同一の形容詞でも複数のカラーパターンが定義されているため、本システムでは全パターンに対して試行した最短距離の結果を各形容詞との距離とした。

また、HVS 空間における距離計算には、*Godlove* 色差式を用いた[4]。各感性語との距離計算式と、*Godlove* 色差式  $\Delta_{godlove}$  を以下に示す。

$$D_{AP} = \sum_{i=1}^3 \{ \Delta_{godlove}(A_{iH}, A_{iV}, A_{iS}, P_{iH}, P_{iV}, P_{iS}) \}$$

$$\Delta_{godlove}(H_1, V_1, S_1, H_2, V_2, S_2) = \frac{2S_1S_2 \left( 1 - \cos \left( 2\pi \frac{|H_1 - H_2|}{100} \right) \right) + (|S_1 - S_2|)^2 + (4|V_1 - V_2|)^2}{2}$$

## 2.2. 感情データの抽出

本システムでは、ネガティブ-ポジティブ感情を表す語として、以下の 10 語を採用している。

- ・ 冷たい-熱い ・ 暗い-明るい ・ 重い-軽い
- ・ 醜い-美しい ・ 悲しい-嬉しい ・ 遠い-近い

- ・ 弱い-強い ・ 静かな-騒がしい ・ 悪い-良い
- ・ 男らしい-女らしい

これらを、稲浪らの実験[3]により定義されている数値を用いて計量を行う。内容を図 2 に示す。

表 2 平均値

	白	赤	黄	緑	青	紫	桃色	茶	橙	黄緑	灰色	黒
冷-熱	2.27	4.84	3.74	2.91	1.94	2.45	3.60	2.85	4.20	3.00	1.66	1.81
暗-明	4.31	4.49	4.46	3.07	2.97	2.27	4.43	2.08	4.48	4.17	1.60	1.17
重-軽	4.20	2.66	4.17	2.71	2.73	2.02	4.36	2.05	3.79	4.09	1.90	1.18
醜-美	4.46	3.69	3.69	3.67	3.85	2.99	3.93	2.49	3.77	3.79	2.38	2.34
悲-嬉	3.20	3.75	4.06	3.15	2.42	2.35	4.25	2.64	4.08	3.69	1.77	1.78
遠-近	2.72	3.58	3.46	2.80	2.35	2.59	3.40	2.87	3.67	3.28	2.29	2.38
弱-強	2.39	4.58	3.09	3.34	3.15	3.60	2.39	3.33	3.59	2.73	3.45	4.08
静-騒	1.56	4.35	3.65	2.20	1.89	3.16	3.48	2.74	3.93	2.59	2.02	2.10
悪-良	4.39	3.10	3.31	3.50	3.41	2.63	3.53	2.87	3.60	3.49	2.41	2.08
男-女	3.59	3.22	3.20	2.77	2.30	3.04	4.60	2.64	3.44	3.00	2.65	2.20

第 1 段階を 1 点、第 2 段階を 2 点、第 3 段階を 3 点、第 4 段階を 4 点、第 5 段階を 5 点として、それぞれの色の平均値を計算した。

図 2: 感情語と色の関係[3]

感情語メタデータを抽出するにあたっては、まず各ピクセルの色情報を *Godlove* 色差による距離計算を用いて原色 12 色に近似する。

画像 I における各ピクセル j の色  $C_{ij}$  に対応する感情語 a の値  $C_{ik}^n$  を使い、全ての感情語 a の値  $W^n$  を算出する。データ構造と、 $W^n$  を求める計算式を以下に示す。

$$\begin{aligned} & \begin{matrix} \text{感情語} \end{matrix} \begin{matrix} \begin{matrix} a_1 \text{ (冷-熱)} \\ \vdots \\ a_{10} \text{ (男-女)} \end{matrix} \end{matrix} \begin{matrix} \begin{matrix} k_1 \text{ (白)} & k_{12} \text{ (黒)} \\ \left[ \begin{matrix} C_{i1}^1 & \cdots & C_{i12}^1 \\ \vdots & \ddots & \vdots \\ C_{i1}^{10} & \cdots & C_{i12}^{10} \end{matrix} \right] \end{matrix} \end{matrix} \\ & \begin{matrix} <\text{冷-熱}> W_1 = \frac{\sum_{j=1}^N (C_{i1}^1, C_{i2}^1, \dots, C_{in}^1)}{N} \\ & \begin{matrix} <\text{暗-明}> W_2 = \frac{\sum_{j=1}^N (C_{i1}^2, C_{i2}^2, \dots, C_{in}^2)}{N} \\ & \vdots \\ & \vdots \\ & <\text{女-男}> W_{10} = \frac{\sum_{j=1}^N (C_{i1}^{10}, C_{i2}^{10}, \dots, C_{in}^{10})}{N} \end{matrix} \end{aligned}$$

## 2.3. 音楽基本設定の定義

以上より得られた各形容詞の値より、各音楽要素に変換を行う。その中でも基本的な音楽の設定として、BPM(Beats per Minute)、音階の使用する割合やリズムの選び方、メロディの動き方を定義する。

これらはいずれも各音楽要素と関連する形容詞を複数選択し、各語と選択する音楽要素によって定義されたデータベースを構築し、入力値として与えられた各形容詞の数値と距離計算を行うことで、最も印象に近い選択を行なっていくものである。

## 2.4. メロディの制作

メロディ制作は、息継ぎ情報の決定、リズムの決定、音の選択の 3 プロセスに細分化され、音楽基本設定を用いて各プロセスは進行していく。

### 3. 実現方式

本システムは図 5 に示すように、画像処理による形容詞抽出機構とメロディ生成機構の 2 パートに音楽理論データベースが介入する形によって構成される。

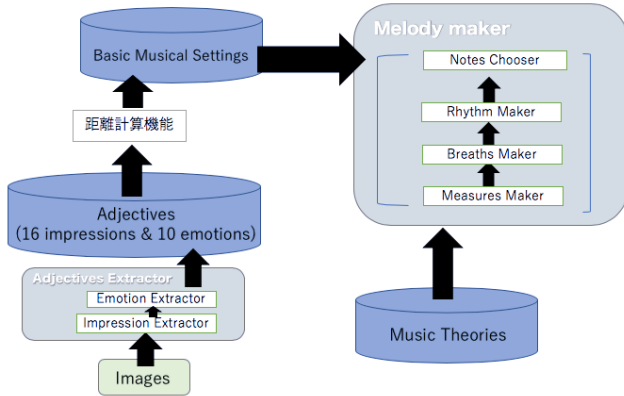


図 5: システム構成図

#### 3.1 Basic Musical Settings(基本音楽設定)

Adjectives Extractor によって得られた形容詞群を元に音楽で使える情報として取り出すために、楽曲作りの中で発生する選択肢と形容詞との関係性を Basic Musical Settings としてデータ定義する。それらには、Adjectives and BPMs, Length of Inhale, Length of Exhale, Adjectives and Booth Making, Emotions and Scale などがある。形容詞群のデータに対して MMM の方式[8]を適用することによって、これらの値を定義していく。

以下の図 6 に、本システム内にデータベースとして導入してある基本音楽情報の全体像と Melody Maker のどの部分で利用されるかを示す。

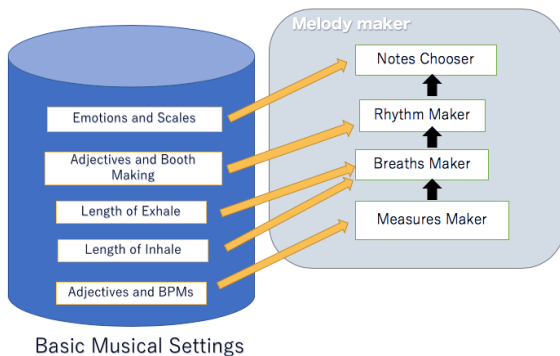


図 6: 基本音楽設定(Basic Musical Settings)の詳細と Melody Maker 内の各プロセスとの関係

#### 3.2 Notes Chooser(音選択器)

発音部分に対して、楽曲基本情報で指定された音階の割合に従い音階を割り振る。さらに同様に指定されている楽器の音域内[6]に候補を絞る。候補に残った選択可能音からランダムで選ぶ。

この音選択を 1 曲の最後まで行う。結果を図 15~16 に示すルールによってスコア化する行程を個体数分行

い、図 16 の設定による遺伝的アルゴリズムによってスコアが最大の個体を完成品として選択する。

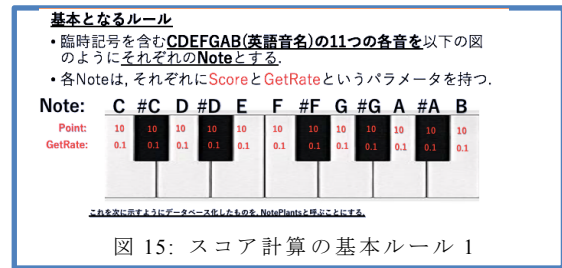


図 16: スコア計算の基本ルール 2

また、旋律が音楽的に不自然な進行をするのを避けるために以下のルールを設ける。

- Music Theories によって定義された禁則[5] [7]を犯した場合、スコアを激減させる
- 同じ音が連続で選ばれた場合、他の音の GetRate が通常より多く減少する
- 全ての音数に対して順次進行が Emotion and Smooth Rate によって得られた割合から離れていた場合、スコアを減少させる

また、無数のパターンから優良なメロディを探索するために遺伝的アルゴリズムを採用した。遺伝的アルゴリズムの流れを以下に示す。

- 次世代に移るにあたり、前世代のスコアが上位 1 つの個体を継承する
- 全世代にルーレット関数を適用し 2 つの親を決定し、それらを次世代へ継承する
- どちらかの親の突然変異を適応した個体を指定した数生成し続ける
- 突然変異の際は、Note Chooser の機能を使い、選択可能範囲から音を選択していく
- 指定した数の個体数に達したらそこまでをひとつの世代とする

### 4. システムの実行

#### 4.1. 実行結果

入力した画像に対して、得られた形容詞情報、次いで 4 小節+アウフタクト 1 拍を指定し、GA において指定する個体数を 10、世代数を 1000 とし生成されたメロディを数パターン以下に示す。

## 画像



参照(<https://www.photo-ac.com>)

### ・結果

・画像 B より得られた各形容詞の相関

プリティな	98
ロマンチックな	99
カジュアルな	94
ナチュラルな	95
クリアな	100
エレガントな	90
シックな	80
スポーティな	88
ダイナミックな	19
ゴージャスな	79
ワイルドな	58
クラシックな	49
重厚な	0
ダンディな	25
フォーマルな	50
モダンな	58
冷たい-熱い	23
暗い-明るい	50
重い-軽い	9
醜い-美しい	10
悲しい-嬉しい	28
遠い-近い	-25
弱い-強い	-39
静かな-騒がしい	6
悪い-良い	-19
男らしい-女らしい	25

・画像 B より生成された曲 1



・画像 B より生成された曲 2



・画像 B より生成された曲 3



## 5. 結論

本論文では、新たな自動作曲システムの実現可能性について検証した。本システムでは、画像から得られたそれぞれの形容詞群データから、違和感のない形容詞データの数値を抽出することに成功し、それぞれの画像ごとに特徴的なメロディを生成し分けられることを示した。

また、遺伝的アルゴリズム内での適応度計算の手法を音楽理論にかなうように音楽家の専門知識と経験に基づいてスコア計算の規則の設定することで、遺伝的アルゴリズムの持つ欠点(ランダム性による音楽性の欠如傾向)をかなり補い、音楽性の高い結果が得られていることが観察できた。

同様に、コード(和音)という概念を一切利用しないで音楽生成を試みたとしても、結果的にコード感を有し、聴き心地よく、編曲の発展性を十分に持った、結果として有効性のあるメロディを生み出せることを示した。

## 6. 今後の展望

今後は、SNS などのテキストデータ、心理テストの結果、脳波や体温などの身体的情報などをセンシングして形容詞群に落とし込むシステムが、本システムと統合されていくことが望まれる。

また、本システム内の音楽基本情報は、本質的には個人によって変化するものだから、性格診断や利用者に関連する情報のセンシングをし、これらを活用してデータが作成されると良い。

## 参考文献

- [1] ソロモン・ヴォルコフ編, 水野忠夫訳: "ショスタコーヴィチの証言" pp.160-258 19801030
- [2] 小林重順: "カラーイメージスケール改訂版講談社". 2001
- [3] 稲浪正充, 栗山智子, 安部美恵子: "色彩と感情について(3)" In 島根大学教育学部紀要(人文・社会科学) 第 28 巻 pp.35-50 December, 1994
- [4] Godlove. I. H. "Improved Color-Difference Formula, with Applications to the Perceptibility and Acceptability of Fadings". J. Opt. Soc. Am., 41, 11, pp.760-770, 1951
- [5] 石桁真礼生 他: "楽典 理論と実習" 第 5 刷 pp.93-133 20020531
- [6] 日本作曲家協会著"編曲の本" pp.188-276
- [7] ノエル・ギャロン マルセル・ビッチュ、矢代秋雄訳: "対位法" pp.16-17
- [8] Y. Kiyoki, T. Kitagawa, and T. Hayama, "A metadata system for semantic image search by a mathematical model of meaning," *ACM SIGMOD Rec.*, 1994