

任意の言葉の音の印象に合致した楽曲の自動生成方式

岡田龍太郎[†] 芳村 亮^{††} 本間 秀典^{††} 北川 高嗣^{††}

[†] 筑波大学第三学群情報学類

^{††} 筑波大学大学院システム情報工学研究科

あらまし 本稿では、言葉の音声表現により人間が受ける印象を用いて、任意の日本語の印象に合致した楽曲自動生成方式を提案する。我々は、音相理論と呼ばれる理論を用いて、入力された任意の言葉から、その音声表現から受ける印象に合致した印象語をメタデータとして抽出する方式を実現している。また、Hevner の研究を用いて、楽曲の印象を表す 8 つの印象語群と、6 つの楽曲構造要素の相関を求めることにより、楽曲を自動生成する方式を実現している。本稿では、この印象語抽出方式と楽曲自動生成方式を意味の数学モデルを用いてメタレベルで連結することにより、任意の言葉の音の印象に合致した楽曲自動生成方式を実現する。

キーワード 自動作曲, 情報生成, 楽曲メディアデータ, 音相, メタデータの抽出, 感性

An Automatic Music Composition Method Corresponding to the Impression of the Sounds of Arbitrary Words

Ryotaro OKADA[†], Ryo YOSHIMURA^{††}, Hidenori HOMMA^{††}, and Takashi KITAGAWA^{††}

[†] College of Information Sciences, Third Cluster of Colleges, University of Tsukuba

^{††} Graduate School of Systems and Information Engineering, University of Tsukuba

Abstract This paper presents an automatic music composition method corresponding to the impression of arbitrary Japanese words by using the impressions received from the sound expression of the words. Founded on the “onso” theory, we have realized an automatic extraction method which extracts impression words as metadata corresponding to the impression by the sound expression of arbitrary words. And we have also realized a method of an automatic composition method of music, by searching for correlations between 8 groups of words about musical impressions and 6 musical structural elements by using the reserches done by Hevner. In this paper, we realize the method of automatic composition of music corresponding to the impression of the sounds of arbitrary words by connencting the impression words extraction method and the method of automatic composition of music on meta-level by using a mathmatical model of meanings.

Key words automatic music composition, generation of infomation, music media data, “Onso”, extraction of metadata, kansei

1. はじめに

現在、ネットワークの高速化および広域化やコンピュータの廉価化に伴い、多種多様なメディアデータ群がネットワーク上に散在しつつある。このため、人間がコンピュータを介してこれらの情報群と接触する機会は増大する傾向にある。しかしながら、これらを対象としたシステムはメディアごとに独立して存在しており、これらを統一的に扱うことが可能なメディアデータ連携統合方式の実現が重要となっている。

我々はこれまでに、言葉と言葉の関係の計量による検索機構として、意味の数学モデルによる意味的連想検索を提案している。これは、単語群を文脈として解釈する機構により、言葉と

言葉、あるいは言葉とメディアデータ、ドキュメント間の相関を文脈に応じて動的に計算することを可能とするモデルである。

また、これまでに、メディアデータから人間の心理的要素が反映されたメタデータを抽出する方式として、「音相理論 [5]」と呼ばれる理論を用いて、日本語の音の印象に合致したメタデータを抽出する方式 [4] を実現しており、また、印象を表すメタデータからメディアデータを生成する方式として、音楽心理学者 Hevner による研究 [1] ~ [3] を用いて、任意の印象語から楽曲を自動生成する方式 [9] を実現している。

これらの方式を意味の数学モデルのような意味的連想検索を用いてメタレベルで連結することにより、独立に実装されたメディア情報の相互運用性を実現し、既存のデータベース群の利

用価値を高め、新たな価値の創造が可能になると考えられる。

本稿では、音相と楽曲メディアデータに着目し、言葉の音声情報の持つ印象から、音楽を生成する楽曲自動生成方式について述べる。本方式は、ユーザから与えられた任意の日本語から、任意の言葉を対象とした音の印象によるメタデータ自動抽出方式 [4] と任意の印象語を対象とした楽曲自動生成方式 [9] を意味の数学モデル [6], [7] を用いてメタレベルで連結することによって、任意の日本語をコンテキストとして楽曲を自動生成する。

本方式は印象語の抽出に言葉の音声表現を用いる方式であるので、利用者の意図する印象を持つ語の印象に合致した楽曲の生成が実現されることで、音楽を通じた情報の新しい価値を生み出すことが可能になる。さらに、未知語や擬音語など、普段我々がその言葉が持つ印象を意識しない言葉でさえ、その印象を表した楽曲を生成することができるので、人間の感性に新たな発見を見出す手助けになると考えられる。

2. 意味の数学モデルの概要

本節では、言葉と言葉の関係の計量を実現する意味の数学モデルの概要を示す。詳細は、文献 [6], [7] に述べられている。

(1) メタデータ空間 MDS の設定

検索対象となるメディアデータをベクトルで表現したデータにマッピングするための正規直交空間（以下、メタデータ空間 MDS ）を設定する。

(2) メディアデータのメタデータをメタデータ空間 MDS へ写像

設定されたメタデータ空間 MDS へ、メディアデータのメタデータをベクトル化し写像する。これにより、検索対象データのメタデータが同じメタデータ空間上に配置されることになり、検索対象データ間の意味的な関係を空間上での距離として計算することが可能となる。

(3) メタデータ空間 MDS の部分空間（意味空間）の選択
検索者は与える文脈を複数の単語を用いて表現する。検索者が与える単語の集合をコンテキストと呼ぶ。このコンテキストを用いてメタデータ空間 MDS に各コンテキストに対応するベクトルを写像する。これらのベクトルは、メタデータ空間 MDS において合成され、意味重心を表すベクトルが生成される。意味重心から各軸への射影値を相関とし、閾値を超えた相関値（以下、重み）を持つ軸からなる部分空間（以下、意味空間）が選択される。

(4) メタデータ空間 MDS の部分空間（意味空間）における相関の定量化

選択されたメタデータ空間 MDS において、メディアデータベクトルのノルムを検索語列との相関として計量する。これにより、与えられたコンテキストと各メディアデータとの相関の強さを定量化している。この意味空間における検索結果は、各メディアデータを相関の強さについてソートしたリストとして与えられる。

2.1 音相理論

木通はそれぞれの言葉が持つ音の構造の違いによって生まれる表情を音相と呼んだ。音相理論では、音相を捉えるための単

位として、それぞれの音の最小単位である音素と、それらの音素を発音するために使われる調音器官を表す調音点、および発音の方法にあたる調音法が挙げられている。さらに、各音相基がどのような印象を表すときに使われているかを調査することにより、その音素の動性（強さ）と輝性（明るさ）を明らかにした。音相理論において、音相の単位として挙げられたこれらの要素は音相基と呼ばれており、40 の音相基が定義されている。そして単独の音相基、または 2 つの音相基の組み合わせと、音相理論では“表情”と呼ばれる、音相基から生じる印象の相関関係、および各表情とそれらに合致する印象語群についてまとめられている [5]。

2.2 任意の言葉の音の印象に合致したメタデータ自動抽出方式の概要

本節では、音相理論に基づいて、言葉の音を持つ印象を表す言葉をメタデータとして自動抽出するための方式を示す。

(1) 音相基の抽出

この方式では、初期パラメータとして、入力語から 40 種の音相基を表現するのに必要十分な 29 の要素からなる初期ベクトル v_i を抽出し、音相基と表情の関係を用いて 40 種の音相基を表す 40 次元のベクトル v_b へと変換する。

(2) v_b の拡張

それぞれの表情が有効であるかを判断する尺度として各音相基の標準使用率が示されているので、これを用いて v_b を補正する。さらに、2 つの音相基の組み合わせによりさらに生じる 38 種の表情の判定を行い、 v_b を 78 次元のベクトル v_o に拡張する。

(3) 印象語群とその重みの出力

それぞれの表情とそれらから抽出される印象語群の関係から変換行列 T を作成し、以下により 40 語の印象語を表す 40 次元のベクトル w_o に変換する。

$$w_o = T v_o.$$

ここでの 40 語の印象語は、2 語を 1 組として 1 種の表現属性を表しており、20 種の表現属性を表している。表情語と表現属性表を表 1 に示す。この w_o が音相の印象に合致したメタデータ抽出方式により抽出される、言葉の音の印象を表す重み付き印象語群によるメタデータである。

3. 任意の印象語を対象とした楽曲自動生成方式

本節では、入力された言葉から印象のメタデータを抽出し、それに合致した楽曲を自動的に生成する、楽曲自動生成方式について述べる。詳細は文献 [9] に述べられている。

3.1 Hevner の研究

Hevner の研究 [1] ~ [3] では、楽曲構造要素として調性 (key)・テンポ (tempo)・音高 (pitch)・リズム (rhythm)・和声 (harmony)・旋律 (melody) の 6 つを挙げている。Hevner は、この 6 つの楽曲構造要素と 8 つの印象語群 (図 1) によって表現される印象との相関関係を調べ、表にまとめた (図 2)。8 つの印象語群は、印象語間で類似性があるものをまとめて一つの印象語群を作り、さらに印象語群間で類似性があるものを隣接する

表 1 音相を表す表情語と表情属性表 .

表情語群	表情属性	訳語
A	シンプルな, 明白さ	plain, obvious
B	躍動感, 進歩的	vibrant, advance
C	新鮮さ, 新奇さ	fresh, unprecedented
D	動的, 活性的	dynamic, active
E	派手さ, 賑やかさ	florid, bustle
F	軽やかさ, 軽快感	light, trippingly
G	若さ, 澆刺さ	young, effervescent
H	現代的, 都会的	modern, urban
I	明るさ, 開放的	bright, open-minded
J	合理的, 現実的	reasonable, real
K	個性的, 特殊的	individual, special
L	強さ, 鋭さ	powerful, sharp
M	適応性, 庶民的	adaptable, popular
N	清らかさ, 爽やかさ	pure, brisk
O	健康的, 清潔感	healthy, clean
P	暖かさ, 安らぎ	warm, comfortable
Q	安定感, 信頼感	stable, confidence
R	高級感, 充実感	expensive, fulfil
S	高尚な, 優雅さ	profound, elegant
T	静的, 非活性的	static, inactive

	印象語群							
	c1	c2	c3	c4	c5	c6	c7	c8
key	0.0476	-0.1429	-0.2381	0.0357	0.2500	0.2857	0.0000	0.0000
tempo	-0.1217	-0.1043	-0.1391	-0.1739	0.0522	0.1739	0.1826	0.0522
pitch	-0.1149	-0.2184	0.0690	0.0920	0.1839	0.0690	-0.1034	-0.1494
rhythm	0.2903	0.0484	-0.1452	-0.0323	0.1290	-0.1613	0.0323	0.1613
hamony	0.0405	-0.0946	0.0541	0.1351	0.1622	0.2162	-0.1892	-0.1081
melody	0.1600	0.0000	0.0000	0.1200	-0.1200	0.0000	-0.2800	-0.3200

図 3 変換行列 T

表 2 楽曲の特徴値

kn	楽曲の調性
tem	テンポ
$mnap$	旋律の音高の平均
$unac$	四分音符の和音の総演奏時間
ac	八分音符の和音の総演奏時間
tc	三和音の総演奏時間
oc	三和音以外の和音の総演奏時間
um	上昇する旋律音の総演奏時間
dm	下降する旋律音の総演奏時間
lm	水平な旋律音の総演奏時間

- Step1: 楽曲印象語群ベクトルの出力

2. 節に示した意味の数学モデルを用いて, 入力に与えられた任意の言葉と, Hevner による 8 つの各印象語群との相関を求める. さらに, その各相関値を, 1) 相関値のうち最大のもを 1 に写像, 2) 相関値のうち最小のもを -1 に写像 することにより正規化し, 楽曲印象語群ベクトル $c = (v_{c1}, v_{c2}, \dots, v_{c8})^T$, (v_{ci} は語群 c_i への重み) を生成する.

- Step2: 楽曲構造要素ベクトルの出力

図 2 で示した, Hevner による各印象語群に対する楽曲構造要素の相対重要性の表に正負の符号をつけ, 行ごとに正規化したものを変換行列 T とする (図 3). 楽曲印象語群ベクトルを入力として, 変換行列 T を作用させることにより, 6 つの楽曲構造要素への重みを表す楽曲構造要素ベクトル $f = (key, tempo, pitch, rhythm, hamony, melody)^T$ を生成する.

$$f = Tc.$$

- Step3: 楽曲の特徴値の出力

出力された f の値をもとに, 生成する楽曲の特徴値 $kn, tem, mnap, unac, ac, tc, oc, um, dm, lm$ を決定する. 表 2 に楽曲の特徴値の意味を示す.

- Step4: 楽曲の生成・出力

Step3 で決定した特徴値をもとに, 3.3 節に示す楽曲の仕様にしたがって楽曲を生成する. 特徴値のうち, 和音による伴奏の生成には $kn, mnap, unac, ac, tc, oc$ を使用し, 旋律の生成には $kn, mnap, um, dm, lm$ を使用する.

3.3 楽曲の仕様

ここでは, 本方式において生成する楽曲の仕様について示す. 楽曲の拍子・長さ 本方式では本格的な曲を生成する必要はなく, 入力された言葉の持つ意味や感情が短いフレーズに反映されればよいので, 生成する楽曲は 4/4 拍子, 長さは 4 小節とする.

	c6 bright cheerful		c5 delicate fanciful graceful humorous light	
	c7 agitated dramatic exciting exhilarated impetuous passionate restless sensational soaring triumphant		c4 calm leisurely lyrical quiet satisfying serene soothing tranquil	
c8 emphatic exalting majestic martial ponderous robust vigorous				
	c1 awe-inspiring dignified lofty sacred serious sober solemn spiritual	c2 dark depressing doleful frustrated gloomy heavy melancholy mournful pathetic sad tragic	c3 dreamy longing plaintive sentimental tender yearning yielding	

図 1 Hevner による 8 つの印象語群

印象語群名	c1	c2	c3	c4	c5	c6	c7	c8
key	長調 4	短調 12	短調 20	長調 3	長調 21	長調 24	---	---
tempo	遅い 14	遅い 12	遅い 16	遅い 20	速い 6	速い 20	速い 21	速い 6
pitch	低い 10	低い 19	高い 6	高い 8	高い 16	高い 6	低い 9	低い 13
rhythm	固定 18	固定 3	流動 9	流動 2	固定 8	流動 10	固定 2	固定 10
hamony	単純 3	複雑 7	単純 4	単純 10	単純 12	単純 16	複雑 14	複雑 8
melody	上昇 4	---	---	上昇 3	下降 3	---	下降 7	下降 8

図 2 各印象語群に対する楽曲構造要素の相対重要性の表

よう, 円形に配置している.

3.2 任意の印象語を対象とした楽曲自動生成方式全体の概要

本方式は次のような手順により実現される.

和音 一つの楽曲に使用しうる和音は非常に多いので制限を与える．使用する和音は I 度から VII 度までの 7 つで，和音の構成音として使用する音はその調の音階に沿ったものとする．長調では I, II_m, III, IV, V7, VI_m, VII_m7⁽⁻⁵⁾，短調では I_m, II_m7⁽⁻⁵⁾, III, IV_m, V7, VI, VII である．また，和音の長さは四分音符と八分音符の 2 種類のみとする．

旋律 旋律に使用する音はその調の音階に沿ったものとする．また，Hevner の研究では旋律音の長さやリズムには言及していないので，ここでは旋律音の長さは四分音符と八分音符の 2 種類のみとする．

4. 任意の言葉の音の印象に合致した楽曲の自動生成方式

本節では，任意の言葉の音声情報からその印象に合致した楽曲を自動生成する方式について示す．本方式の概要は以下の通りである．全体の概要図を図 4 に示す．

(1) 言葉の印象の抽出

3. 節に示した，対象となる語の音相に合致した印象語の抽出方式を用いて，ユーザから入力された任意の言葉の音相から，印象語ベクトル w_o を抽出する．

(2) 重みの小さい印象語の削除 (提案手法 1)

(1) により得られた印象語ベクトル w_o は，そのままでは多数の印象語が選択されているので，Hevner の指定する 8 つの各楽曲印象語群との相関を計量した際に，特徴を打ち消しあい，入力語ごとの差を生じにくくさせる恐れがある．そこで，重みの大きい印象語だけを言葉の音の印象として選択し，重みの小さい印象語を削除する．

詳細は 4.1 節に示す．

(3) 印象語ベクトルと各楽曲印象語群との相関の計量

2. 節に示した意味の数学モデルを用いて，(2) で得られた印象語ベクトル w_o と Hevner の指定する 8 つの各楽曲印象語群との相関を計量する．

(4) 音相を表す印象語の特性に対する相関値の補正 (提案手法 2)

音相を表す印象語として w_o に選ばれた印象語は，Hevner の指定する 8 つの楽曲印象語群との相関を計量した際に，特定の印象ばかりが強調されやすい語が選ばれている可能性があり，出力される楽曲において入力語毎の印象の違いを反映できていない恐れがある．そこで w_o に選ばれた印象語と各楽曲印象語群との相関を計量した際に強調される割合に応じて，相関値の重みを補正する．

詳細は 4.2 節に示す．

(5) 楽曲の生成

(4) で得られた相関値から楽曲印象語群ベクトルを構成し，3. 節に示した任意の印象語を対象とした楽曲自動生成方式を用いて楽曲を生成する．

4.1 重みの小さい印象語の削除 (提案手法 1)

入力語の音相から得られた印象語ベクトル w_o は，そのままでは多数の印象語が選択されているので，Hevner の指定する 8 つの楽曲印象語群における相関を計量した際に，特徴を打ち

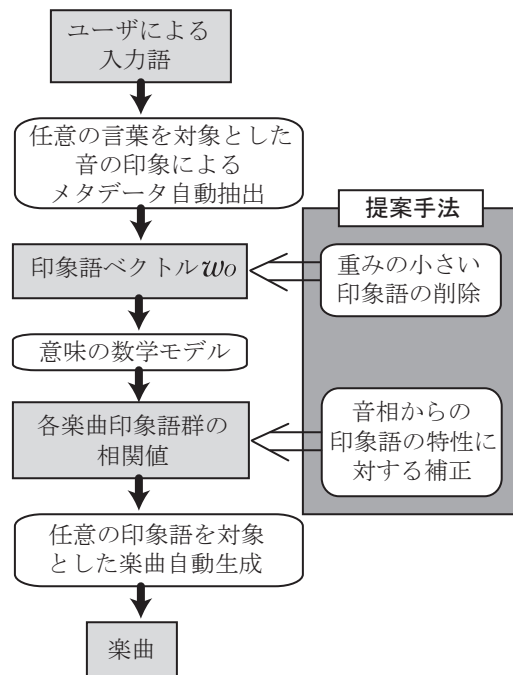


図 4 システム全体の概要図

消しあい，入力語ごとの差を生じにくくさせる恐れがある．そこで，重みの大きい印象語だけを言葉の音の印象として選択し，重みの小さい印象語を削除する．

入力語の音相から抽出された 40 次元の印象語ベクトル w_o の各印象語の重みを $w_{oi} (i = 1, 2, \dots, 40)$ とするとき，各要素の重みを最大の重みで割り，閾値 ε 未満の印象語の重みを 0 とする．実行後の w_o の要素 w'_{oi} は以下の式で表される．

$$w'_{oi} = \begin{cases} 0 & \text{if } \frac{w_{oi}}{\max(w_o)} < \varepsilon \\ \frac{w_{oi}}{\max(w_o)} & \text{if } \frac{w_{oi}}{\max(w_o)} \geq \varepsilon \end{cases},$$

$$0 \leq \varepsilon \leq 1.$$

(ただし， $\max(w_o)$ は w_o の要素のうち最大のもの)

4.2 音相を表す印象語の特性に対する相関値の補正 (提案手法 2)

音相を表す印象語として w_o に選ばれた印象語は，Hevner の指定する 8 つの楽曲印象語群との相関を計量した際に，特定の印象ばかりが強調されやすい語が選ばれている可能性があり，出力される楽曲において入力語毎の印象の違いを反映できていない恐れがある．そこで w_o に選ばれた印象語と各楽曲印象語群との相関を計量した際に強調される割合に応じて，相関値の重みを補正する．

4.2.1 補正值 r_i の導出

音相を表す印象語は，2 語を 1 組として 1 種の表現属性を表しており，20 種の表現属性を 40 語の印象語で表している．ここでは，1 種の表現属性を表す 2 語のみを印象語ベクトル w_o とし，意味の数学モデルで Hevner の指定する各楽曲印象語群との相関を計量する．その際，2 語の重みはそれぞれ 1 とする．20 の表現属性を $o_j (j = 1, 2, \dots, 20)$ と表すとき， w_{o_j} と楽曲印象語群 $c_i (i = 1, 2, \dots, 8)$ との相関値を $v_{c_i o_j}$ と表す．

音相を表す印象語の特性に対する補正值を $r_i (i = 1, 2, \dots, 8)$ としたとき， r_i の導出式を以下に示す．

表 3 補正值 r_i の各値

	補正值
r_1	1.014846497
r_2	0.994237836
r_3	1.023523453
r_4	0.928139303
r_5	0.983858843
r_6	1.025028797
r_7	1.065676845
r_8	0.976531032

$$r_i = \frac{\sum_{k=1}^8 \sum_{l=1}^{20} v_{c_k o_l}}{8 \sum_{l=1}^{20} v_{c_i o_l}}$$

r_i は、音相を表す印象語と Hevner の指定する 8 つの楽曲印象語群との意味的な相関を計量した際の、全ての印象語群の相関値の平均から見た、各印象語群の強調されやすさの割合の逆数である。 r_i の値は入力に関わらず一定であり、それぞれが楽曲印象語群と対応している。 r_i の各値を表 3 に示す。

4.2.2 各楽曲印象語群の相関値の補正

実行時に入力された任意の言葉から得られた各楽曲印象語群の相関値を $cor_{c_i} (i = 1, 2, \dots, 8)$ とする。相関値と r_i との各要素の積をとり、補正を行う。補正後の各相関値を $cor'_{c_i} (i = 1, 2, \dots, 8)$ とするとき、式を以下に示す。

$$cor'_{c_i} = cor_{c_i} r_i.$$

5. 実験例

本節は、本方式の実現性を検証するため、本方式に基づく実験システムを構築した。

5.1 実験環境

本実験では、意味の数学モデルを用いて検索システムの構築を行った。意味の数学モデルの基本構成は 2. 節に示す通りである。メタデータ空間 MDS の設定については、“Longman Dictionary of Contemporary English” [8] という英英辞典を利用した。同辞書は約 2,000 語の基本語を用いて約 56,000 語の見出し語を説明している。この基本語を特徴とみなし、見出し語の説明で肯定的に使われている語を“1”、否定的に使われている語を“-1”、説明に使われていない語を“0”として 2. 節 (1) におけるデータ行列 M を作成した。これにより、約 2,000 次元の正規直交行列であるメタデータ空間 MDS が生成された。 MDS では、約 2^{2000} 通りの意味の様相を表現可能である。

またここでは、重みの小さい印象語の削除 (4.1 節) において、 $\varepsilon = 0.75$ とした。この値は恣意的に決められたもので、この値に限定されるものではない。

5.2 出力例

入力として、文献 [5] において“美しさ、爽やかさ、明白さ、単純さ”に関する語として分類されている“キラキラ”という言葉と、“否定、抗争”に関する語として用いられている“ゲンナリ”という言葉を用いて、楽曲の生成を行った。

まず、重みの小さい印象語の削除 (4.1 節) と音相を表す印象



図 5 “キラキラ”補正前の楽譜



図 6 “ゲンナリ”補正前の楽譜

語の特性に対する相関値の補正 (4.2 節) をどちらも行わないで楽曲を生成した。“キラキラ”を入力とした場合の、印象語ベクトル、楽曲印象語群ベクトル、楽曲構造要素ベクトル、楽曲特徴値の各値の結果を表 4 に、“ゲンナリ”を入力した場合の結果を表 5 に示す。また、実際に出力された楽曲の楽譜を、“キラキラ”を図 5、“ゲンナリ”を図 6 に示す。

次に、重みの小さい印象語の削除と、音相を表す印象語の特性に対する相関値の補正を行った場合の結果について、“キラキラ”を入力とした場合の、印象語ベクトル、楽曲印象語群ベクトル、楽曲構造要素ベクトル、楽曲特徴値の各値の結果を表 6 に、“ゲンナリ”を入力した場合の結果を表 7 に示す。また、実際に出力された楽曲の楽譜を、“キラキラ”を図 7、“ゲンナリ”を図 8 に示す。

5.3 考察

重みの小さい印象語の削除 (4.1 節) と音相を表す印象語の特性に対する相関値の補正 (4.2 節) をどちらも行わない場合では、楽曲印象語群ベクトルにおいては、“キラキラ”(表 4) においては印象語群 c_4 、すなわち、“calm”、“quiet”などを表す印象が高くなっている。これは語の印象とあまり一致していない。“ゲンナリ”(表 5) においても印象語群 c_4 が高くなっている、これは語の印象とほとんど一致していない。楽曲特徴値においても、“キラキラ”と“ゲンナリ”の双方の調が“Am”であり、テンポも音の高さの平均も非常に似通っている。これは言葉の印象を正確に楽曲に反映できていないと考えられる。

表 4 “キラキラ” 補正前の各値

印象語ベクトル		楽曲印象語群ベクトル		楽曲構造要素ベクトル		楽曲特徴値	
young	9.000000	C1	-0.537486	Key	-0.050635	kn	Am
effervescent	9.000000	C2	-0.241331	Tempo	-0.250188	tem	76
open-minded	8.200000	C3	-0.236541	Pitch	0.204456	mnap	76.973794
bright	8.200000	C4	1.000000	Rhythm	0.113639	unac	0.556819
dynamic	8.200000	C5	0.536656	Harmony	0.114471	ac	0.443181
active	8.200000	C6	-1.000000	Melody	-0.039723	tc	0.557236
plain	8.000000	C7	-0.404560			oc	0.442764
obvious	8.000000	C8	0.383134			um	0.327219
unprecedented	7.000000					dm	0.366942
fresh	7.000000					lm	0.305839
trippingly	6.000000						
light	6.000000						
:							
:							

表 5 “ゲンナリ” 補正前の各値

印象語ベクトル		楽曲印象語群ベクトル		楽曲構造要素ベクトル		楽曲特徴値	
elegant	7.200000	C1	-0.012603	Key	-0.033621	kn	Am
profound	7.200000	C2	-0.685180	Tempo	-0.304845	tem	74
confidence	7.030596	C3	-0.342484	Pitch	0.201665	mnap	76.864922
stable	7.030596	C4	1.000000	Rhythm	0.177761	unac	0.588880
plain	6.333333	C5	0.150210	Harmony	0.145953	ac	0.411120
obvious	6.333333	C6	-1.000000	Melody	0.181317	tc	0.572976
trippingly	6.000000	C7	-0.527704			oc	0.427024
light	6.000000	C8	0.207493			um	0.482286
healthy	5.333333					dm	0.300969
clean	5.333333					lm	0.216745
warm	5.223188						
comfortable	5.223188						
:							
:							

表 6 “キラキラ” 補正後の各値

印象語ベクトル		楽曲印象語群ベクトル		楽曲構造要素ベクトル		楽曲特徴値	
young	1.000000	C1	-1.000000	Key	0.308639	kn	Ab
effervescent	1.000000	C2	-0.089891	Tempo	0.274271	tem	126
bright	0.911111	C3	-0.194059	Pitch	0.262709	mnap	79.245669
dynamic	0.911111	C4	0.529447	Rhythm	-0.121569	unac	0.439216
active	0.911111	C5	1.000000	Harmony	0.000228	ac	0.560784
open-minded	0.911111	C6	0.099062	Melody	-0.467897	tc	0.500114
plain	0.888889	C7	0.714345			oc	0.499886
obvious	0.888889	C8	0.160669			um	0.149798
unprecedented	0.777778					dm	0.617695
fresh	0.777778					lm	0.232508

重みの小さい印象語の削除と音相を表す印象語の特性に対する相関値の補正をどちらも行った場合には、“キラキラ”(表 6)においては、印象語群 c5, すなわち、“delicate”, “light”などを表す印象が高くなっている。これは語の印象に近づいたと考えられる。“ゲンナリ”(表 7)においては、印象語群 c1, すなわち、“serious”, “dignified”などの印象が高くなっている。これは語の印象に近づいたと考えられる。楽曲特徴値においても、“キラキラ”の調は“Ab”というメジャーキーであり、テンポも126と速く、旋律音の平均が高い楽曲が生成されている。また、

“ゲンナリ”の調は“Am”というマイナーキーであり、旋律音の平均も低い楽曲が生成されている。

これらにより、任意の言葉の音の印象に合致した楽曲自動生成方式の実現可能性が示されていることが確認できた。

6. おわりに

本稿では、音相と楽曲メディアデータに着目し、ユーザから与えられた任意の日本語の音声情報の持つ印象に合致した楽曲を自動生成する方式を実現した。

表 7 “ゲンナリ” 補正後の各値

印象語ベクトル		楽曲印象語群ベクトル		楽曲構造要素ベクトル		楽曲特徴値	
elegant	1.000000	C1	1.000000	Key	-0.146430	kn	Am
profound	1.000000	C2	-0.894169	Tempo	-0.141244	tem	81
confidence	0.976472	C3	-0.048231	Pitch	0.047264	mnap	70.843303
stable	0.976472	C4	-0.677318	Rhythm	0.306280	unac	0.653140
plain	0.879630	C5	-0.093467	Harmony	-0.165718	ac	0.346860
obvious	0.879630	C6	-1.000000	Melody	0.350937	tc	0.417141
light	0.833333	C7	-0.116608			oc	0.582859
trippingly	0.833333	C8	-0.713591			um	0.523506
						dm	0.172569
						lm	0.303925



図 7 “キラキラ” 補正後の楽譜



図 8 “ゲンナリ” 補正後の楽譜

本方式により、音相と楽曲メディアデータという独立に実装されたメディア情報の相互運用性を実現し、既存のデータベース群の利用価値を高め、新たな価値の創造が可能になる。また、普段我々がその言葉が持つ印象を意識しない言葉の印象を表した楽曲を生成することは、人間の感性に新たな発見を見出す手助けになると考えられる。

また、本方式を利用して、商品名のイメージに合った CM の楽曲を自動生成するシステムや、詩のイメージに合った楽曲を自動生成するシステムなど、人間の感性を増幅するシステムの開発が可能になると考えられる。

今後の課題としては、本方式の定量的な評価方式、および他メディアを対象としたメディアデータ連携統合方式の実現が挙げられる。

文 献

- [1] Hevner, K., “Expression in music: A discussion of experimental studies and theories”, *Psychological Review*, vol.42, pp.186-204, 1935.
- [2] Hevner, K., “Experimental studies of the elements of expression in music”, *American J. Psychology*, vol.48, pp.246-268, 1936.
- [3] Hevner, K., “The affective value of pitch and tempo in music”, *American J. Psychology*, vol.49, pp.621-630, 1937.
- [4] 本間 秀典, 中西 崇文, 北川 高嗣: “任意の言葉を対象とした音の印象によるメタデータ自動抽出方式”, 電子情報通信学会技術研究報告, Vol.106, No.149, pp.7-12, 電子情報通信学会, 2006.
- [5] 木通隆行: “日本語の音相 — ことばのイメージを捉える技術, 表現する技術 —”, 小学館スクウェア, (2004).
- [6] Kitagawa, T. and Kiyoki, Y.: “The mathematical model of meaning and its application to multidatabase systems”, *Proceedings of 3rd IEEE International Workshop on Research Issues on Data Engineering: Interoperability in Multidatabase Systems*, pp. 130-135(1993).
- [7] Kiyoki, Y., Kitagawa, T. and Hayama, T.: “A Metadatabase System for Semantic Image Search by a Mathematical Model of Meaning”, *Multimedia Data Management – using metadata to integrate and apply digital media –*, McGrawHill, A. Sheth and W. Klas(editors), Chapter 7 (1998).
- [8] *Longman Dictionary of Contemporary English*, Longman (1987).
- [9] 芳村 亮, 中西 崇文, 北川 高嗣: “任意の言葉を対象とした楽曲自動生成方式”, 第 17 回データ工学ワークショップ論文集, 電子情報通信学会, 2006.