

任意の言葉を対象とした楽曲自動生成方式

芳村 亮[†] 中西 崇文^{††} 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、任意の印象を表す言葉により音楽を生成する楽曲自動生成の実現方式について述べる。本方式は、音楽心理学者 Hevner による研究を用い、与えられた任意の言葉を対象として、音楽に関する 8 つの印象語群との関係を計量し、さらにそれと 6 つの楽曲構造要素との相関を求めることにより、与えられた言葉の印象に合致した楽曲を自動生成することを可能とする。本方式は、言葉と言葉、あるいは言葉とメディアデータ間の意味的な関係を与えられた文脈や状況に応じて動的に計算することが可能である意味の数学モデルを用いることにより、任意の言葉からの楽曲自動生成を実現している。本方式が実現されることで、利用者の意図とする感情に合致した楽曲メディアデータが自動生成され、それによりコンピュータ環境の中で人間の感性を増幅し、コンピュータが人間と密着した感性を共有することが可能になると考えられる。本稿では、本方式の基本的な機能の実装を行い、実験によって本方式の有効性を示す。

キーワード 自動作曲, 情報生成, 楽曲メディアデータ, 感性

An Implementation Method of Automatic Composition of Music by Arbitrary Impression Words

Ryo YOSHIMURA[†], Takafumi NAKANISHI^{††}, 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 a composition method of the music by arbitrary impression words. This method can compose the music dynamically corresponding to the impression by using for researches done by Hevner, who was a psychologist of music, measuring relation between impression of given words and 8 groups of words about musical impression, and searching for correlation of 6 musical structural elements. By using a mathematical model of meaning which can calculate semantic relations between languages or a language and media data dynamically according to the context and the situation, this method can compose automatically from arbitrary words. When this method realizes, musical media data which conform to emotion intended by the user is composited automatically, computers will be able to share sensitivities with humans. In this paper, we implement basic functions of this method, and clarify the effectiveness of this method by showing experimental results.

Key words automatic composition of music, generation of information, music media data, sensitivity

1. まえがき

現在、ネットワークの高速化および広域化、計算機資源の高速化および廉価化により、コンピュータネットワーク環境には多数のサイトに様々な形でメディアデータが分散、蓄積されている。メディアデータ群を対象とした情報獲得の機会の可能性が増大する一方、膨大な情報の検索の負担が大きくなってきている。その環境下において、利用者の感性的な問い合わせに合致した、メディアデータをリアルタイムに自動生成する方式の実現が重要となってきた。

これまで、文献 [1]~[3] で、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、多変量解析による空間生成を用いた検索方式である LSI(Latent Semantic Indexing) [4], [5] とは異なり、言葉と言葉、あるいは、言葉と検索対象のメディアデータ、ドキュメント間の関係を検索者が検索語として与えた単語の集合(以下、コンテキスト)に応じて動的に計量することを可能とする。意味の数学モデルと LSI の違いについて、詳細は文献 [6] で報告されている。

また、これまで、文献 [7], [8] で、楽曲メディアデータのメ

タデータを自動抽出するための実現方式について述べている。これは、メディアデータが人に与える印象を心理学の研究成果に基づいて、言葉によって表現されるメタデータとして自動抽出するための実現方式について述べている。さらに、これらを用いた、楽曲メディアデータを対象とした人間の感性に合致した意味的連想検索 [9] を実現している。このように人間の心理的要素が反映されたメタデータを対象としたメディアデータ検索機能の実現により、広域ネットワーク上に散在する膨大なメディアデータ群の中から適切なメディアデータを獲得する機会が増大すると考えられる。

本稿では、任意の印象を表す言葉による楽曲自動生成の実現方式について述べる。本方式は、与えられた任意の言葉を対象として、その言葉の印象に合致した楽曲を自動生成することを可能とする。つまり、本方式は文献 [7]~[9] の方式の逆演算を行うことにより、ユーザが発行した問い合わせに合致した楽曲メディアデータを生成するものである。

本方式の特徴として、楽曲の構成要素とその印象の関係を示した専門家の研究を用いることにより、その印象に合致した楽曲の構成要素を構成し、楽曲が生成できる点、さらに、言葉と言葉の意味的な関係を計量可能な意味の数学モデルを用いることにより、専門家が提示した印象を表す語と名詞、動詞などを含む一般的な語を意味的に計量することが可能となり、ほぼ全ての言葉を対象として、その言葉の印象に合致した楽曲を生成できる点、任意の言葉を用いることで、利用者の意図する複雑な感情に対応することができる点にある。本方式は、任意の一般的な単語を入力するという点で、DTM マガジンの連載「自動作曲研究所」などで紹介されている、生成する曲のジャンルを指定する「自動作曲システム」[10] などのような従来手法とは異なる。

本方式が実現されることにより、利用者の意図とする感情に合致した楽曲メディアデータの自動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となる。また、本方式によりコンピュータ環境の中で人間の感性を増幅することができるシステムとなり、コンピュータが人間の生活を豊かにし、人間と密着した感性を共有するパートナーとして新しい関係を築く第一歩となると考えられる。

本稿では、任意の言葉による楽曲自動生成方式について示し、さらに実験により本方式の有効性を示す。

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 [1]~[3] 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。

具体的には以下の通りメタデータ空間 MDS を設定する。

初めに、 m 個の基本データについて各々 n 個の特徴 (f_1, f_2, \dots, f_n) を列挙した特徴付ベクトル $\mathbf{d}_i (i = 1, \dots, m)$ が与えられているものとし、そのベクトルを並べて構成する $m \times n$

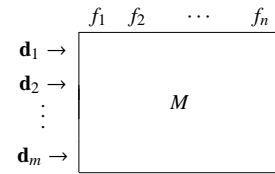


図1 データ行列 M によるメタデータの表現。

行列を M とおく (図1)。このとき、 M は、列ごとに2ノルムで正規化されている。

- (a) データ行列 M の相関行列 $M^T M$ を計算する。
- (b) $M^T M$ を固有値分解する。

$$M^T M = Q \begin{pmatrix} \lambda_1 & & & \\ & \ddots & & \\ & & \lambda_\nu & \\ & & & 0_{\cdot 0} \end{pmatrix} Q^T, \quad (1)$$

$$0 \leq \nu \leq n.$$

ここで行列 Q は、

$$Q = (\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_n) \quad (2)$$

である。この $\mathbf{q}_i (i = 1, \dots, n)$ は、相関行列の正規化された固有ベクトルである。相関行列の対称性から、この固有値は全て実数であり、その固有ベクトルは互いに直交している。

- (c) メタデータ空間 MDS を以下で定義する。

非ゼロ固有値に対応する固有ベクトルによって形成される正規直交空間をメタデータ空間 MDS と定義する。この空間の次元 ν は、データ行列 M のランクに一致する。この空間は、 ν 次元ユークリッド空間となる。

$$MDS := span(\mathbf{q}_1, \mathbf{q}_2, \dots, \mathbf{q}_\nu). \quad (3)$$

$\{\mathbf{q}_1, \dots, \mathbf{q}_\nu\}$ は MDS の正規直交基底である。

(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。

具体的には、メディアデータのメタデータを以下のようにベクトル化する。

- (a) **メディアデータの特徴づけ**

メディアデータ P を t 個の印象語 (あるいは、 t 個のオブジェクト) $\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_t$ を用いて、次のように特徴づける。

$$P = \{\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_t\}. \quad (4)$$

ここで、各印象語 \mathbf{o}_i は、データ行列の特徴と同一の特徴を用いて表現される特徴付ベクトルである。

$$\mathbf{o}_i = (o_{i1}, o_{i2}, \dots, o_{in}) \quad (5)$$

- (b) **メディアデータ P のベクトル表現**

メディアデータ P を構成する t 個の印象語 $\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_t$ が、それぞれ n 次元のベクトルで定義されている。印象語 $\mathbf{o}_1, \mathbf{o}_2, \dots, \mathbf{o}_t$ は、合成することで n 次元ベクトル表現され、メディアデータベクトル \mathbf{p} を形成する、さらにこのメディアデータベクトル \mathbf{p} をメタデータ空間 MDS に写像する。これにより、同じ空間上に言葉とメディアデータが配置されることになり、言葉とメディアデータの関係性を空間上の距離として動的に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 の部分空間 (意味空間) 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。

3. 任意の印象語を対象とした楽曲自動生成方式

本章では、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。

3.1 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

本方式の全体図を図 2 に、本方式での処理の流れを図 3 に示す。この方式は、次のような手順により実現される。

- Step1: 楽曲印象語群ベクトルの出力

意味の数学モデルにより、入力された言葉 (コンテキスト) と Hevner の 8 つの印象語群それぞれとの相関を示す楽曲構造要素ベクトルが出力される。詳細は 3.3.2 で述べる。

- Step2: 楽曲構造要素ベクトルの出力

楽曲印象語群ベクトルと、8 つの印象語群と 6 つの楽曲構造要素との相関を示す変換行列により、6 つの楽曲構造要素によって特徴付けされた楽曲構造要素ベクトルが出力される。詳細は 3.3.3 で述べる。

- Step3: 楽曲の特徴値の出力

楽曲構造要素ベクトルの値をもとにして、生成する楽曲の特徴値を決定する。詳細は 3.3.4 で述べる。

- Step4: 楽曲の生成・出力

Step3 で決定した特徴値をもとに楽曲を生成する。詳細は 3.3.5 で述べる。

3.2 Hevner の研究

Hevner の研究 [11]~[13] では、楽曲構造要素として調性 (key)・

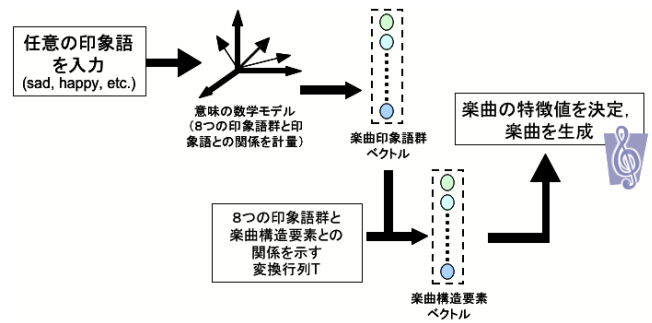


図 2 楽曲自動生成方式の全体図

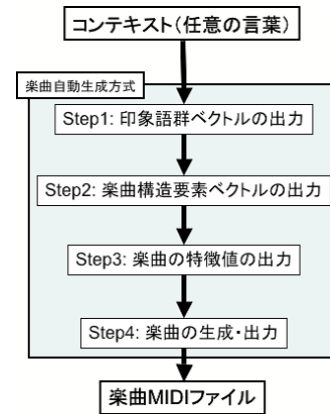


図 3 楽曲自動生成方式の処理の流れ

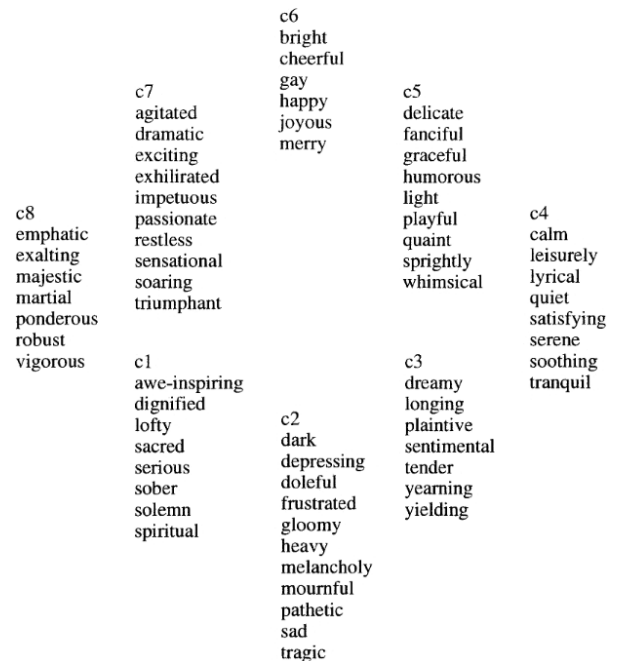


図 4 Hevner による 8 つの印象語群

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 4) によって表現される印象との相関関係を調べた。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、さらに印象語間で類似性があるものを隣接するよう、円形に配置している。

Hevner は、各印象語群によって表現される印象と楽曲構造要素との相関関係を調べるにあたり、次のような実験を行っている。

印象語群名	c1	c2	c3	c4	c5	c6	c7	c8
構造要素名								
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5 各印象語群に対する楽曲構造要素の相対重要性の表

(1) 任意の楽曲を任意の数, 被験者に聴かせ, 各楽曲の印象を表す印象語を円形の印象語群リストから選ばせる.

(2) (1) で用いた楽曲について, 他の楽曲構造要素はできる限りそのままに保ち, ひとつの楽曲構造要素のみを変化させたものを被験者に聴かせ, 1. と同様にその印象を表す印象語を選ばせる. 楽曲構造要素の変化は, 各楽曲構造要素において対になる変化をさせる. 例えば, 調性においては, 長調の楽曲は短調へ, 単調の楽曲は長調へ変化させる.

(3) (1), (2) で選ばれた印象語を各印象語群ごとにまとめて集計した結果を $vote_{kij}$ (k : 印象語群番号, i : 楽曲構造要素番号, j : 変化番号; $k = 1, \dots, 8, i = 1, \dots, 6, j = 1, 2$) とし, 各印象語群ごとの総標本数を $spec_{kij}$ とする.

(4) $prob_{kij} = \frac{vote_{kij}}{spec_{kij}}$ としたとき, 各印象語群と各楽曲構造要素との相関度 cor_{ki} を次の式で求める.

$$cor_{ki} = \frac{|prob_{ki1} - prob_{ki2}|}{\sqrt{\frac{prob_{ki1}(1-prob_{ki1})}{spec_{ki1}} + \frac{prob_{ki2}(1-prob_{ki2})}{spec_{ki2}}}} \quad (6)$$

Hevner は, 以上の実験から得られた cor_{ki} をまとめ, 印象語群によって表現される 8 つの印象に対する各楽曲構造要素の相対重要性の表 (図 5) [12] を作成している.

3.3 楽曲自動生成の実現方式

本節では, 提案方式の各機能の詳細について示す.

3.3.1 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.

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, 旋律音の長さは四分音符のみとする.

3.3.2 楽曲印象語群ベクトルの出力 (Step1 の詳細)

2. 章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のを 1 に写像, 2) 相関値のうち最小のものを -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 $\mathbf{c} = (v_{c1}, v_{c2}, \dots, v_{c8})$, (v_{ci} は語群 c_i への重み) を生成する. なお, この正規化の方式は, い

楽曲構造要素	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6 変換行列 T

表1 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

長調の調性度									
C	G	D	A	E	B	F#	F	B _b	E _b
0.9375	1.0000	0.7500	0.7500	0.4375	0.6250	0.5000	0.6875	0.1250	0.4375
A _b	D _b								
0.3125	0.0000								

短調の調性度									
a	e	b	f#	c#	d	g	c	f	b _b
0.0000	-0.6250	-0.6250	-0.6250	-0.8750	-0.7500	-0.5000	-1.0000	-0.7500	-0.8750

図7 各調の調性度を示す表

くつかの予備実験の結果をもとに定めたものである. また, ここの定式化においては上記に限定しない.

3.3.3 楽曲構造要素ベクトルの出力 (Step2 の詳細)

図 5 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 6). 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 $\mathbf{f} = (\text{key}, \text{tempo}, \text{pitch}, \text{rhythm}, \text{harmony}, \text{melody})^T$ を生成する.

$$\mathbf{f} = T\mathbf{c} \quad (7)$$

なお, 変換行列 T の規定方式は予備実験によるもので, ここの定式化においては上記に限定しない.

3.3.4 楽曲の特徴値の出力 (Step3 の詳細)

出力された MFV 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. 調性とテンポはここで定める値を楽曲に使用し, 音高・リズム・和声・旋律については, ここで定める特徴値をもとに次の行程でより具体的に決定する. kn の決定方法については文献 [14] に, それ以外の各特徴値の決定方法は文献 [7] 中の式にもとづいて定めている. 表 1 に楽曲の特徴値の意味を示す.

調性 kn の決定

調性の値から, 生成する楽曲の調 kn を定める. 梅本の研究 [14] による表 (図 7) から特徴値に近いものを探し, それを生成する楽曲の調性とする. 候補となる調性が複数ある場合は, その中からひとつをランダムに選択する.

テンポ tem の決定

以下の式により、生成する楽曲のテンポ tem を定める。

$$tem = \begin{cases} tstan + (tmax - tstan)tempo, & tempo \geq 0 \\ tstan + (tstan - tmin)tempo, & tempo < 0 \end{cases} \quad (8)$$

ただし、 $tempo$ は MFV におけるテンポの値、 $tmax$ はテンポの最大値（ここでは 184）、 $tmin$ はテンポの最小値（ここでは 44）、 $tstan$ はテンポの標準値（ここでは 88）である。また、 tem , $tmax$, $tmin$, $tstan$ の単位は拍/分である。

音高 $mnap$ の決定

以下の式により、生成する楽曲の旋律音の高さの平均 $mnap$ を定める。

$$mnap = \begin{cases} pstand + (pmax - pstand)pitch, & pitch \geq 0 \\ pstand + (pstand - pmin)pitch, & pitch < 0 \end{cases} \quad (9)$$

ただし、 $pitch$ は MFV における音高の値、 $pmax$ は音高の最大値（ここでは 88 鍵ピアノの最高音にあたる MIDI ノートナンバー） $pmin$ は音高の最小値（88 鍵ピアノの最低音にあたる MIDI ノートナンバー）、 $pstand$ は音高の標準値（88 鍵ピアノの中央 A 音にあたる MIDI ノートナンバー）である。

リズムの特徴 $unac, ac$ の決定

以下の式により、生成する楽曲のリズムの特徴 $unac, ac$ を定める。

$$rhythm = \frac{unac - ac}{unac + ac} \quad (10)$$

ただし、 $rhythm$ は MFV におけるリズムの値である。 $unac + ac$ は曲全体の演奏時間に相当するため、これを 1 とすると $unac, ac$ を一意に定めることができる。

和声の特徴 tc, oc の決定

和声の値から、生成する楽曲に使用する和声の配分を定める。

$$harmony = \frac{tc - oc}{tc + oc + nc} \quad (11)$$

ただし、 $harmony$ は MFV における和声の値、 nc は和声のない時間の合計（今回は 0 に規定）である。このとき、リズムの場合と同様に tc, oc は一意に定まる。

旋律の特徴 um, dm, lm の決定

以下の式により、生成する楽曲の旋律の特徴を定める。

$$melody = \frac{um - dm}{um + dm + lm} \quad (12)$$

ただし、 $melody$ は MFV における旋律の値、である。文献[14]より、次の旋律音が同じ音となる割合は 26%なので、 lm が 0.26 に近い値となるように乱数を用いて定める。ただし、 $lm > 1 - |melody|$ となったときは、 $lm \leq |melody|$ となるように乱数を用いて定める。 lm が定まると、リズムや和声の場合と同様に um, dm は一意に定まる。文献[14]の内容は次節の中で示す。

3.3.5 楽曲の生成・出力 (Step4 の詳細)

楽曲の特徴値に沿って、新たな楽曲を生成する。以下に楽曲データの生成手順を示す。

和音による伴奏の生成

決定した特徴値のうち、 $kn, mnap, unac, ac, tc, oc$ を用いて、

和音を構成する音を以下の手順により生成する。

(1) 使用する和音を kn によって決定する。

(2) $unac$ と ac の値から四分音符の和音と八分音符の和音の個数・配列を決定する。 $unac$ の値を四分音符の和音の個数、 ac を八分音符の個数に対応させる。

(3) tc, oc の値をもとに各和音を「三和音」と「それ以外の和音」に振り分ける。ただし、最後の 2 拍または 2.5 拍は必ず「三和音」とする。

(4) 各和音の具体的な和音名を決定する。「三和音」のものは I 度、IV 度、V 度から、「それ以外の和音」は II 度、III 度、VI 度、VII 度からランダムに選択する。ただし、最後の 2 拍または 2.5 拍 (3. において「三和音」に固定した部分) は必ず I 度の和音とする。

旋律の生成

特徴値のうち $kn, mnap, um, dm, lm$ を用いて旋律音を以下の手順により生成する。

(1) kn より使用する音階を決定する。長調ならば長音階、短調ならば短音階（和声的短音階）とする。

(2) um, dm, lm の割合に応じて上昇音、下降音、水平音の個数を決定する。

(3) 旋律の大まかな流れを決定するため、各旋律音を上昇音、下降音、水平音のいずれかに振り分ける。ただし、最後の音については、後ろの音が存在しないので水平音と定める。

(4) 旋律を具体的に決定する。最後の音をその調の主音に固定して後ろから順に決定していく。ここで、文献[14]には、隣接する二つの旋律音の音程とその出現する割合についての研究が示されている。この研究では、歌曲の旋律について、1 度から 8 度までの音程の出現頻度を調べている (表 2)。上昇音と下降音ではその音程の変化を表 2 に基づいた確率でランダムに決定する。

(5) 生成された旋律が $mnap \pm 1$ オクターブ以内に収まらない場合は、(3) からやり直す。

4. 実験

4.1 実験目的

本方式の有効性を検証するため、本方式に基づく実験システムを構築し、検証実験を行った。

実験 1 では、楽曲自動生成を行い、生成された楽曲の構造要素について考察することにより、本方式の有効性の検証を行う。

また、本方式は、楽曲の特徴値や旋律音、和音などにおいてランダムに決定する要素を含むため、同一のコンテキストに対して生成される楽曲にばらつきが生じる可能性がある。そこで、実験 2 では、一つのコンテキストから生成された複数の楽曲について、それらの印象を考察し、コンテキストの印象に合致した楽曲が生成されることを示す。

なお、Hevner の研究は一般の被験者に対する聴取実験の結果に基づいており、本方式の検証として聴取実験を行うことは意味を持たない (Hevner の研究成果の検証にすぎない) ため、ここではそのような検証方法は用いないこととした。また、本方式に関する適切な評価方法の確立は今後の課題である。

表2 歌曲の旋律音における各音程の出現頻度

音程	1度	2度		3度		4度	
		短2度	長2度	短3度	長3度	完全4度	増4度
出現頻度 (%)	26	17.5	24	10	6	9	0.5
音程	5度		6度		7度		8度
	減5度	完全5度	短6度	長6度	短7度	長7度	
出現頻度 (%)	0.5	3	1	1	0.5	0	1

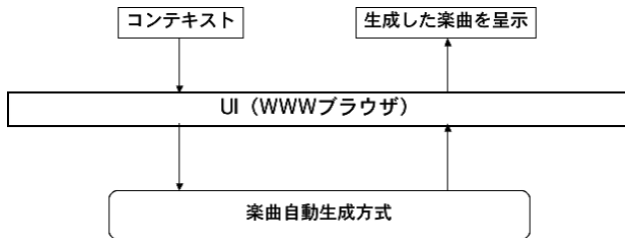


図8 実験システム図

4.2 実験環境

意味の数学モデルにおける、メタデータ空間 MDS 作成については、“Longman Dictionary of Contemporary English” という英英辞書を使用した。同辞書は、約 2,000 の基本語だけを用いて約 56,000 の見出し語を説明している。ここで基本語を特徴と見なし、各見出し語を説明する基本語が肯定の意味に用いられていた場合 “1”，否定の場合 “-1”，使用されていない場合 “0”，見出し語自身が基本語である場合その基本語の要素を “1” とし、2. 章 (1) のデータ行列 M を作成した。これより、約 2000 次元の正規直交空間であるメタデータ空間 MDS を生成した。約 2000 次元のメタデータ空間 MDS では、約 2^{2000} 通りの意味の様相が表現可能である。

問い合わせとして用いるコンテキストを構成する印象語は、2. 章および上記より “Longman Dictionary of Contemporary English” の約 2,000 語からなる基本語で特徴づけ可能な語はすべて利用可能にすることができる。実装では、“Longman Dictionary of Contemporary English” の見出し語、約 56,000 語をコンテキストとして利用可能である。つまりわれわれが普段用いる単語をほぼ網羅している。

生成した楽曲データは、いったんフリーソフト MF2T/T2MF [15] で規定された形式のテキストファイルを出力し、それを T2MF によりスタンダード MIDI ファイルに変換する。

4.3 実験システム

この実験システムを、C 言語および Perl を用いて実装した。このシステムでは、WWW ブラウザをユーザインタフェースとし、問い合わせたコンテキストの印象に合致した楽曲を提示する。実験システムの全体図を図 8 に示す。

4.4 実験 1 (コンテキストの印象と生成した楽曲の印象との比較に関する実験)

4.4.1 実験方法

誰にとっても明らかな印象を持つ言葉をコンテキストとして与えて楽曲を生成し、楽曲印象語群への相関値のうち最大のものと最小のものに対して強い影響を持つ楽曲構造要素について調べる。

ここでは、図 4 の Hevner による印象語群に含まれない任意

の語からでも楽曲を生成可能であることを示すため、コンテキストを「enjoy」、「cry」、「silent」の 3 種類とした。コンテキストを「enjoy」としたときに生成された楽曲を図 9 に、「cry」としたときの楽曲を図 10 に、「silent」としたときの楽曲を図 11 に示す。また、3.3.2 節の過程で生成される、それぞれのコンテキストに対する楽曲印象語群ベクトルを表 3, 4, 5 に示す。

4.4.2 実験結果

コンテキスト「enjoy」を与えたときの楽曲印象語群ベクトルは、 $c_6 = \{bright, happy, \dots\}$ が最大、 $c_2 = \{dark, sad, \dots\}$ が最小となっている。また、「cry」の場合については $c_2 = \{dark, sad, \dots\}$ が最大、 $c_1 = \{sacred, serious, \dots\}$ が最小である。「silent」の場合については、 $c_4 = \{calm, quiete, \dots\}$ が最大、 $c_7 = \{dramatic, exciting, \dots\}$ が最小である。

図 5 および図 6 より、印象語群 c_1 に関しては、楽曲構造要素 $tempo, pitch$ と強い負の相関が、 $rhythm$ と強い正の相関がある。 c_2 に関しては、 $key, tempo, pitch$ と強い負の相関がある。 c_4 に関しては、 $tempo$ に強い負の相関、 $pitch, harmony$ に強い正の相関がある。 c_6 は、 $key, tempo, harmony$ と強い正の相関がある。 c_7 は、 $tempo$ に強い正の相関、 $pitch, harmony$ には強い負の相関がある。このことより、コンテキスト「happy」の楽曲では $key, tempo, pitch, harmony$ に関する特徴値 $kn, tem, mnap, tc, oc$ を、コンテキスト「cry」の楽曲では $key, tempo, pitch, rhythm$ に関する特徴値 $kn, tem, mnap, unac, ac$ を、コンテキスト「silent」の楽曲では $tempo, pitch, harmony$ に関する特徴値 $tem, mnap, tc, oc$ を調べた。

「enjoy」の楽曲における調性 kn は A で、曲のテンポ tem は 127、旋律音の音高の平均 $mnap$ は約 86、 tc は八分音符 23 個分、 oc は八分音符 9 個分であった。

「cry」の楽曲については、調性 kn が Dm、テンポ tem は 76、旋律音の音高の平均 $mnap$ は約 61、 $unac$ が八分音符 10 個分、 ac は八分音符 21 個分であった。

また、「silent」の楽曲については、テンポ tem が 73、音高の平均 $mnap$ は約 80、 tc が八分音符 18 個分、 oc は八分音符 14 個分であった。

4.4.3 考察

コンテキストが「enjoy」の場合、図 5 および表 3 から、生成する曲には調性 kn が長調、テンポ tem 、音高 $mnap$ が標準値より大きな値で、三和音の演奏時間 tc がそれ以外の和音 oc より長いことが望まれる。この実験においてコンテキスト「happy」より生成された楽曲は、 kn が A (イ長調)、 tem が 127 (標準値は 88)、 $mnap$ が約 86 (標準値は 69)、 tc が八分音符 23 個分、 oc が八分音符 9 個分であり、要件を満たしている。

表3 楽曲印象語群ベクトル

(コンテキスト = enjoy)

c1	0.196688
c2	0.156225
c3	0.162491
c4	0.193579
c5	0.201239
c6	0.211153
c7	0.179321
c8	0.188390

表4 楽曲印象語群ベクトル

(コンテキスト = cry)

c1	0.124194
c2	0.249028
c3	0.223308
c4	0.149383
c5	0.131908
c6	0.135087
c7	0.156837
c8	0.166790

表5 楽曲印象語群ベクトル (コンテキスト = silent)

c1	0.078322
c2	0.138262
c3	0.107311
c4	0.276828
c5	0.115048
c6	0.092444
c7	0.061845
c8	0.088478

enjoy_s.mid Conv

♩ = 127

図9 実験結果：コンテキスト = enjoy

cry_s.mid Conv

♩ = 76

図10 実験結果：コンテキスト = cry

「enjoy」とは逆に、コンテキストが「cry」の場合は、*kn* は短調で、*tem* と *mnap* が標準より小さい値で、*ac* が *unac* より長いことが望まれる。ここでコンテキスト「cry」より生成された

silent_s.mid Conv

♩ = 73

図11 実験結果：コンテキスト = silent

楽曲は、*kn* は Dm (二短調)、*tem* が 76、*mnap* が約 61、*unac* が八分音符 10 個分、*ac* が 22 個分で、この楽曲についても要件を満たしている。

また、「silent」の場合、*tem* は標準よりも小さい値、*mnap* は標準より大きい値で、*tc* が *oc* より長いことが望まれる。生成された楽曲を見ると、*tem* が 73、*mnap* が約 80、*tc* が八分音符 18 個分、*oc* が八分音符 14 個分であり、要件を満たしている。

以上の結果より、入力したコンテキストの印象に合致した楽曲を生成できること、図4の楽曲印象語群に含まれない任意のコンテキストから楽曲を生成できることが確認できた。

4.5 実験2 (同一コンテキストから生成した楽曲の印象に関する実験)

4.5.1 実験方法

コンテキストを「enjoy」として生成された3つの楽曲について、それぞれの楽曲構造要素を調べ、3曲すべてがコンテキスト「enjoy」の印象に合致した楽曲となっているかを検証する。

実験1で用いたコンテキスト「enjoy」の楽曲を enjoy-1 とし、あらたに生成した2曲を enjoy-2 および enjoy-3 とする。enjoy-2 と enjoy-3 の楽譜をそれぞれ図12と図13に示す。楽曲印象語群ベクトルについては、いずれの曲も表3と同一である。

4.5.2 実験結果

実験1のコンテキスト「enjoy」による楽曲と同様、生成された楽曲の *kn*, *tem*, *mnap*, *tc*, *oc* について調べた。

enjoy-1の楽曲構造要素については実験1で述べたとおりである。

enjoy-2は、*kn* が A、*tem* が 127、*mnap* が約 87、*tc* が八分音符 23 個分、*oc* は八分音符 9 個分であった。

enjoy-3は、*kn* が A、*tem* が 127、*mnap* は約 81、*tc* が八分音符 23 個分、*oc* 八分音符 9 個分であった。

4.5.3 考察

生成した3曲の構造要素のうち、*tem*, *tc*, *oc* はすべて同じ値であった。*tem* については楽曲構造要素ベクトルから一意に定まるため、また *tc* と *oc* については、実際に使用する和音はランダムに決定するものの、和音の割合であるこれらの値は楽曲構造要素から一意に定まるため、このような結果になったと思われる。また *kn* は enjoy-1 と enjoy2 では A、enjoy-3 では D (二

enjoy_2_s.mid Conv

♩ = 127

図 12 楽曲 enjoy-2

enjoy_3_s.mid Conv

♩ = 127

図 13 楽曲 enjoy-3

長調)となっている。基本的に kn は一意に決まるが、ここでは図 7 における A と D の調性度が等しいためにいずれかがランダムに選択され、このような結果になった。 $mnap$ の値はそれぞれの曲によって異なるが、いずれも標準値より大きな値となっている。

したがって、enjoy-1, enjoy-2, enjoy-3 の 3 曲はいずれも入力されたコンテキスト「enjoy」の印象に合致していると言える。

4.6 実験全体の考察

実験 1 では、入力されたコンテキストの印象に合致した楽曲を生成できることを確認した。

実験 2 では、ランダムに決定する要素を含んでいても、一つのコンテキストから同様の印象を持つ楽曲を生成できることを示した。

以上の実験結果は、本方式の有効性を示すものである。

5. あとがき

本稿では、与えられた任意の言葉を対象として、その言葉の印象に合致した楽曲を生成する楽曲自動生成の実現方式について述べた。本方式により、利用者が発行する任意の言葉による問い合わせを対象として、その言葉の印象に合致した楽曲メタデータの生成が可能となった。

また、本方式を実現する実験システムを構築し、実験を行い、

本方式の有効性を示した。

本方式により、利用者の意図とする感情に合致した楽曲メタデータの自動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となる。また、本方式によりコンピュータ環境の中で人間の感性を増幅することできるシステムとなり、コンピュータが人間の生活を豊かにし、人間と密着した感性を共有するパートナーとして新しい関係を築く第一歩となると考えられる。

今後の課題は、画像など他メディアを入力として、その印象と合致した楽曲を生成するような異種メディア間の連携・統合方式の実現、本方式における学習方式、個人差の計量方式の実現、および、本方式の定量的な評価方式の確立が挙げられる。

文 献

- [1] T. Kitagawa, Y. Kiyoki, "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, April 1993.
- [2] Y. Kiyoki, T. Kitagawa, H. Takamari, "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, Amit Sheth and Wolfgang Klas(editors), Chapter 7, 1998.
- [3] 清木康, 金子昌史, 北川高嗣, "意味の数学モデルによる画像データベース探索方式とその学習機構," 電子情報通信学会論文誌, D-II, Vol.J79-D-II, No.4, pp.509-519, 1996.
- [4] M. W. Berry, S. T. Dumains, G. W. O'Brien, "Using linear algebra for intelligent information retrieval," SIAM Review Vol. 37, No.4, pp.573-595, 1995.
- [5] S. Deerwester, S. T. Dumais, G. W. Furnas, T. K. Landauer, R. Harshman: "Indexing by Latent Semantic Analysis," Journal of the American Society for Information Science, Vol. 41, No. 6, pp.391-407, 1990.
- [6] 伊東拓, 中西崇文, 北川高嗣, 清木康, "潜在的意味抽出方式と意味の数学モデルによる意味的連想検索方式の比較," 第 13 回データ工学ワークショップ (DEWS2002) 論文集, 電子情報通信学会, 2002.
- [7] T. Kitagawa, Y. Kiyoki, "Fundamental framework for media data retrieval system using media lexco transformation operator," Information Modeling and Knowledge Bases, IOS Press, 2000.
- [8] 吉野太智, 高木秀行, 清木康, 北川高嗣, "楽曲データを対象としたメタデータの自動生成とその意味的連想検索への適用," 情報処理学会研究報告, 1998-DBS-116(2), pp.109-116, 1998.
- [9] T. Kitagawa, T. Nakanishi, Y. Kiyoki, "An Implementation Method of Automatic Metadata Extraction Method for Music Data and Its Application to Semantic Associative Search," Systems and Computers in Japan, Vol.35, No.6, pp59-78, 2004. Translated from Denshi Joho Tsushin Gakkai Ronbunshi(in Japanese), Vol.J85-D-1, No.6, pp.512-526, 2002.
- [10] (有) ミューテック, "自動作曲システム," <http://hp.vector.co.jp/authors/VA014815/music/autocomp.html>
- [11] K. Hevner, "Expression in music: A discussion of experimental studies and theories," Psychological Review, vol.42, pp.186-204, 1935.
- [12] K. Hevner, "Experimental studies of the elements of expression in music," American J.Psychology, vol.48, pp.246-268, 1936.
- [13] K. Hevner, "The affective value of pitch and tempo in music," American J. Psychology, vol.49, pp.621-630, 1937.
- [14] 梅本堯夫 (編), "音楽心理学," 誠信書房, 1966.
- [15] P. V. Oostrum, "MF2T/T2MF," <http://rd.vector.co.jp/soft/dos/art/se021866.html>