

任意の言葉の印象と音楽心理学に基づく楽曲自動生成方式

芳村 亮[†] 中西 崇文^{††} 北川 高嗣[†]

[†] 筑波大学大学院システム情報工学研究科

^{††} 独立行政法人情報通信研究機構

E-mail: [†]yoshimura@mma.cs.tsukuba.ac.jp, ^{††}takafumi@nict.go.jp, ^{†††}takashi@cs.tsukuba.ac.jp

あらまし 本稿では、任意の言葉を入力することにより、その印象を音楽心理学に基づいて表現する楽曲を自動生成するための方式について述べる。先行研究において、音楽心理学者 Hevner の研究成果を用い、与えられた言葉と音楽に関する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、与えられた任意の言葉からの楽曲自動生成を実現するための手法について示している。本方式では、これに音楽心理学分野における個別の楽曲構造要素に関する研究成果を反映し、与えられた任意の言葉の印象により合致した楽曲自動生成を可能としている。本方式の実現により、利用者の意図する感情に合致した楽曲メディアデータの自動生成が可能となる。こ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、本方式の基本機能の実装について述べ、実験によってその実現性および有効性について示す。
キーワード 自動作曲, 情報生成, 楽曲メディアデータ, 感性

Automatic Music Composition Method Based on Impressions of Arbitrary Words and Music Psychology

Ryo YOSHIMURA[†], Takafumi NAKANISHI^{††}, and Takashi KITAGAWA[†]

[†] Graduate School of Systems and Information Engineering, University of Tsukuba

^{††} National Institute of Information and Communications Technology

E-mail: [†]yoshimura@mma.cs.tsukuba.ac.jp, ^{††}takafumi@nict.go.jp, ^{†††}takashi@cs.tsukuba.ac.jp

Abstract This paper presents an automatic composition method of the music expressing the impression of given arbitrary words based on music psychology. The automatic composition method we proposed in previous work, using researches done by Hevner who was a music psychologist, measuring relation between impression of words given by the user and 8 groups of words about musical impression, and searching for correlation of 6 musical structural elements, can compose the music dynamically corresponding to the impression of the arbitrary words. This method enable to compose the music more corresponding than the previous method, by using the researches about respective musical structures. When this method is realized, musical media data conform to emotion intended by the user is composed automatically, computers will be able to share sensitivities with humans. In this paper, we implement basic functions of this method and clarify the possibility of this method by showing experimental results.

Key words automatic music composition, generation of information, musical media data, kansei

1. はじめに

近年、ネットワーク技術の発達や計算機資源の増大により、コンピュータネットワーク環境には様々なメディアデータが分散・蓄積されている。しかし、これらのメディアデータを獲得する機会の増加に伴い、膨大な情報を検索するための負担も大きくなっている。このような状況である一方で、将来、人間とコンピュータがその場に応じた感性を共有し、増幅しあうシス

テムの実現が重要になると考えられる。

これまで、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 [1]~[3] 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。

また, [6], [7] では, 楽曲メタデータを自動抽出する手法について述べられている。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である。なお, これらを用いて, 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 [8] を実現している。

さらに, 感性の共有・増幅という点では検索以上に感性情報からのメディアデータの生成および提示が有効であると考えられることから, [9] では, 入力された任意の言葉の印象に合致した楽曲の自動生成方式が提案されている。

本稿では, 任意の言葉の印象と音楽心理学に基づいた楽曲自動生成の実現方式について述べる。本方式は, [9] で述べられている 6 つの楽曲構造要素に加え, 新たな楽曲構造要素を加味することで, 与えられた任意の言葉の印象により合致した楽曲の自動生成を可能とする。これは, [6]~[8] 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, 楽曲の構造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を生成できる点, さらに, 言葉同士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一般的な語を意味的に計量でき, ほぼすべての言葉からそ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できる点がある。本方式は任意の一般的な単語を入力するという点で, 生成する曲のジャンルを指定する「自動作曲システム」[10] などのような既存の自動作曲手法とは異なる。ただし既存の手法と組み合わせることで, より自然な楽曲生成が可能になると考えられる。

本方式の実現により, 利用者の意図する感情に合致した楽曲メディアデータの自動生成ができ, 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でき, 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, 任意の言葉の印象と音楽心理学に基づいた楽曲自動生成方式について述べ, さらに実験によりその実現性・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 [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。

(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。

3. 提案方式

本節では, 入力された任意の言葉からその印象を表すメタデータを抽出し, その印象に合致した楽曲を音楽心理学に基づいて自動生成する方式について述べる。

3.1 本方式の概要

本方式の全体図を図 1 に示す。この方式は, 以下に示す手順により実現される。

Step1: 楽曲印象語群ベクトルの抽出 意味の数学モデルにより, 入力された言葉 (コンテキスト) の Hevner の 8 つの楽曲印象語群それぞれに対する相関を計量し, これを楽曲印象語群ベクトルとする。詳細は 3.4.2 で述べる。

Step2: 楽曲構造要素ベクトルの抽出 楽曲印象語群ベクトルに, 印象語群と楽曲構造要素との相関を示す変換行列を作用させることにより, 各楽曲構造要素への重みを表す楽曲構造要素ベクトルを求める。詳細は 3.4.3 で述べる。

Step3: 楽曲の特徴値の抽出 Step2 で抽出した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求める。詳細は 3.4.4 で述べる。

Step4: 楽曲の生成・出力 Step3 で抽出した特徴値をもとに, それに沿った楽曲を生成する。詳細は 3.4.5 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)・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 2) 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, さらに印象語群間で類似性があるものを隣接するよう, 円形に配置し

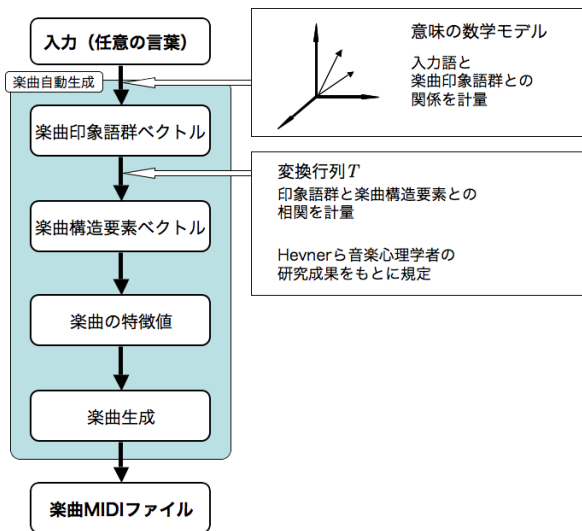


図1 楽曲自動生成方式の全体図

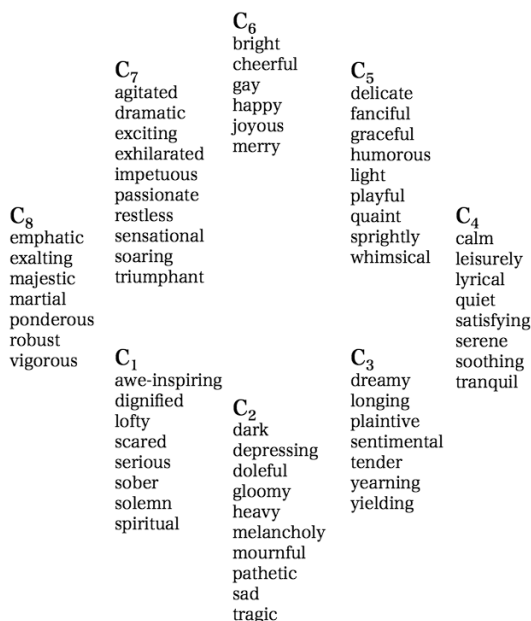


図2 Hevner による 8つの印象語群

表1 各印象語群に対する楽曲構造要素の相対重要性の表

	楽曲印象語群							
	C ₁	C ₂	C ₃	C ₄	C ₅	C ₆	C ₇	C ₈
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
harmony	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

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。

(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、長調楽曲は短調へ、短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。

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

3.3 楽曲の音量・強弱に関する研究

上記の Hevner の研究では、8つの印象語群との関係が述べられているのは6つの楽曲構造要素のみで、楽曲の音量 (loudness) や強弱の幅 (loudness variation) といった要素には触れられていない。

[14] では、楽曲の音量や強弱の幅とその印象の関係について調べた音楽心理学分野の研究が Hevner の印象語群に対応づけて示されている。

3.4 楽曲自動生成の実現方式

ここでは、提案方式の各機能の詳細について示す。

3.4.1 楽曲の仕様

本方式において生成する楽曲の仕様は以下の通りである。

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_m⁽⁻⁵⁾、短調ではIm, II_m⁽⁻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、旋律の長さは四分音符と八分音符のみとする。

3.4.2 楽曲印象語群ベクトルの出力 (Step1の詳細)

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。この各相関値を

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、8次元のベクトル c' を生成する。

$$c' = (v_{c_1}, v_{c_2}, \dots, v_{c_8}) \quad (2)$$

ただし v_{c_i} ($i = 1, 2, \dots, 8$) は図2の語群 C_i への重みを表す。さらに、この c' を2つ結合させたものを、16次元の楽曲印象語群ベクトルとする。

$$c = (c', c')^T \quad (3)$$

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
vol	音量の基本値
velvar	強弱の割合

なお、相関値の正規化の方式は、い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、ここでの定式化においては上記に限定しない。

3.4.3 楽曲構造要素ベクトルの出力 (Step2 の詳細)

まず、表1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、各行を1ノルムで正規化したものを行列 A とする。また、[14]に示されている音量や強弱の幅に関する研究成果をもとに定めた、楽曲印象語群と楽曲の音量・強弱の幅との相関を示す行列（各行について1ノルムで正規化）を B とする。このふたつの行列 A と B より、変換行列 T を図3のように定める。

さらに、楽曲印象語群ベクトルに対して変換行列 T を作用させることにより8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 を生成する。

$$f = (v_{f_1}, v_{f_2}, \dots, v_{f_8})^T = Tc \quad (4)$$

ただし、 v_{f_i} ($i = 1, 2, \dots, 8$) は、それぞれ楽曲構造要素 key, tempo, pitch, rhythm, harmony, melody, loudness, loudness variation への重みを表す。なお、本方式では、変換行列 T の各行を1ノルムで正規化しているが、この規定方式は予備実験によるものであり、ここでの定式化においては上記に限定しない。

3.4.4 楽曲の特徴値の出力 (Step3 の詳細)

出力された楽曲構造要素ベクトルの値をもとに、生成する楽曲の特徴値 $kn, tem, mnap, unac, ac, tc, oc, um, dm, lm, vol, velvar$ 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[15]に、それ以外の各特徴値のうち、 $tem, mnap, unac, ac, tc, oc, um, dm, lm$ の決定方法は[6]中の式に基づいて定めている。また、 vol および $velvar$ の決定方法については、予備実験をもとにして定めている。

表2に楽曲の各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[15]による表3から楽曲構造要素ベクトルにおける調性の値 v_{f_1} 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 [6]による以下の式から、生成する楽曲の

表3 各調の調性度を示す表

長調の調性度									
C	G	D	A	E	B	F#	F	B \flat	E \flat
0.9375	1.0000	0.7500	0.7500	0.4375	0.6250	0.5000	0.6875	0.1250	0.4375
A \flat	D \flat								
0.3125	0.0000								
短調の調性度									
a	e	b	f#	c#	d	g	c	f	b \flat
0.0000	-0.6250	-0.6250	-0.6250	-0.8750	-0.7500	-0.5000	-1.0000	-0.7500	-0.8750

テンポを決定する。

$$tem = \begin{cases} tstand + (tmax - tstand)v_{f_2}, & v_{f_2} \geq 0 \\ tstand + (tstand - tmin)v_{f_2}, & v_{f_2} < 0 \end{cases} \quad (5)$$

ただし、 v_{f_2} は楽曲構造要素ベクトル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）、 $tstand$ はテンポの標準値（ここでは88）である。また、 $tem, tmax, tmin, tstand$ 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。

$$mnap = \begin{cases} pstand + (pmax - pstand)v_{f_3}, & v_{f_3} \geq 0 \\ pstand + (pstand - pmin)v_{f_3}, & v_{f_3} < 0 \end{cases} \quad (6)$$

ただし、 v_{f_3} は楽曲構造要素ベクトルにおける音高の値、 $pmax, pmin, pstand$ はそれぞれ音高の最大値、最小値、標準値（88 鍵ピアノの最高音、最低音、中央A音にあたるMIDIノート番号）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。

$$v_{f_4} = \frac{unac - ac}{unac + ac} \quad (7)$$

ただし、 v_{f_4} は楽曲構造要素ベクトル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。

$$v_{f_5} = \frac{tc - oc}{tc + oc + nc} \quad (8)$$

ただし、 v_{f_5} は楽曲構造要素ベクトルにおける和声の値、 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。

$$v_{f_6} = \frac{um - dm}{um + dm + lm} \quad (9)$$

ただし、 v_{f_6} は楽曲構造要素ベクトルにおける旋律の値である。[15]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。[15]の内容は次節の中で示す。

音量 vol の決定 以下の式より、生成する楽曲の音量の基本値 vol を決定する。

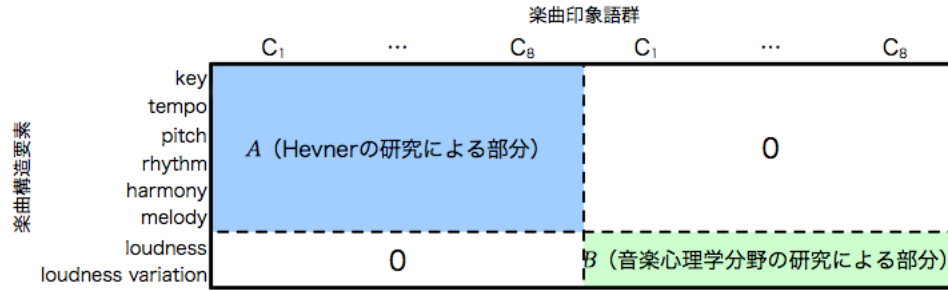


図3 変換行列 T

$$vol = \begin{cases} vstand + (vmax - vstand)v_{f_i}, & v_{f_i} \geq 0 \\ vstand + (vstand - vmin)v_{f_i}, & v_{f_i} < 0 \end{cases} \quad (10)$$

ただし、 v_{f_i} は楽曲構造要素ベクトルにおける音量の値、 $vmax$ は音量の最大値、 $vmin$ は最小値、 $vstand$ は標準値である。なお、本方式では $vmax = 110$ 、 $vmin = 65$ 、 $vstand = 85$ と規定している。 vol は楽曲全体の音量を定めるもので、MIDI ではコントロールナンバー7の volume に相当する。

強弱 $velvar$ の決定 以下の式より、生成する楽曲の強弱の幅 $velvar$ を決定する。

$$velvar = \begin{cases} vvstand + (vvmax - vvstand)v_{f_8}, & v_{f_8} \geq 0 \\ vvstand + (vvstand - vvmin)v_{f_8}, & v_{f_8} < 0 \end{cases} \quad (11)$$

ただし、 v_{f_8} は楽曲構造要素ベクトルにおける強弱の値、 $vvmax$ は強弱の幅の最大値、 $vvmin$ は最小値、 $vvstand$ は標準値である。本方式では $vvmax = 0.65$ 、 $vvmin = 0$ 、 $vvstand = 0.225$ と規定している。 $velvar$ は vol とは違い、ひとつひとつの音符ごとの強弱について特徴づけるもので、MIDI では各音の velocity に対応する。

3.4.5 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」のものには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、(3)において「三和音」に固定した、最後の2拍または2.5拍は必ずI度の和音とする。

隣接する旋律音間の音程に関する研究 [15]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4)。

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、水平音の割合を決定する。
- (3) 上昇音、下降音、水平音それぞれに属する四分音符・八分音符の割合をランダムに決定する。
- (4) 上で定めた上昇音、下降音、水平音の順序を決定する。
- (5) 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。上昇音と下降音ではその音程の変化を表4に示した音程の出現頻度に基づいた確率でランダムに決定する。

音量・強弱の幅の決定 特徴値 vol および $velvar$ から、各和音・旋律音の音量・強弱の幅を決定する。曲全体を通した旋律音の音量 (MIDI の volume) を vol 、和音の音量を $0.75vol$ とする。また、各音の強弱 (MIDI の velocity) については、通常の強さを $velo$ (本方式では80に規定) としたとき、最も強い音を $(1 + velvar)velo$ 、最も弱い音を $(1 - velvar)velo$ とする。また、各和音・旋律音の強弱は楽典における強拍・弱拍の基本概念をもとに決定する。

4. 実 験

4.1 実験目的

本方式に基づく実験システムを構築し、実験を行った。

実験1では、実際に楽曲自動生成を行い、コンテキストの持つ印象と生成された楽曲の構造について考察することにより、本方式の実現性・有効性を示す。

また実験2では、 $loudness$ や $loudness variation$ が考慮されていない、[9]の方式 (以下、従来方式) で生成される楽曲との比較を行う。

なお、本方式で用いている、Hevner らによる音楽心理学の研究は、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 らの研究成果の検証に過ぎない) ため、ここではそのような検証は行わないものとする。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。

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

表4 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

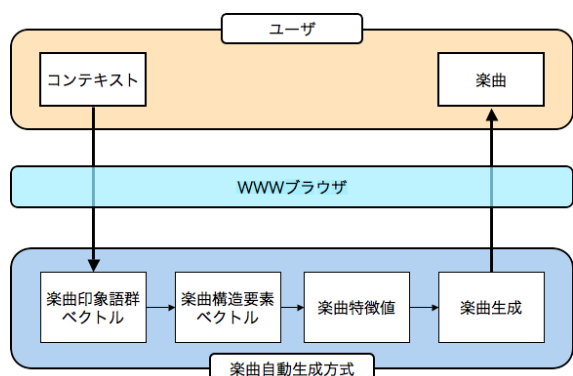


図4 実験システム

いては、“Longman Dictionary of Contemporary English” [16] 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的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”として、2.節(1)のデータ行列 M を作成した。これにより、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.節および上記より“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 [17] 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

本方式に基づく実験システムを、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生成・提示する。実験システムの全体図を図4に示す。

4.4 実験 1

4.4.1 実験方法

誰にとっても明らかな印象を持つと思われる語をコンテキストとして楽曲の生成を行い、生成された楽曲の構造について調べる。

ここでは1つ目のコンテキストを「fun」、2つ目のコンテキ



図5 コンテキスト「fun」から生成された楽曲



図6 コンテキスト「terrible」から生成された楽曲

ストを「terrible」とした。

4.4.2 実験結果

コンテキスト「fun」から得られた楽曲の譜面を図5に、「terrible」から得られた楽曲の譜面を図6に示す。また、コンテキスト「fun」および「terrible」から抽出された楽曲印象語群ベクトル、楽曲構造要素ベクトル、特徴値はそれぞれ表5、表6のとおりである。

4.4.3 考察

表5での楽曲印象語群ベクトルの値を見ると、語群 C_6 への重み、つまり“happy”, “merry”などを表す印象が強くなっている。また、特徴値では、調 kn が長調の B、テンポ tem が 139 など、多くが明るい印象を与えるものとなっている。図5に示した楽譜でも、そのような構造の楽曲であることが確認できる。なお、各音の強弱について、最も強く演奏された音の velocity は 107、最も弱い音は 52 であった。これらの結果より、入力したコンテキスト「fun」の印象に合致した楽曲が生成できていると言える。

表5 コンテキスト「fun」の各値

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
C ₁ : -1.000000	key: 0.615808	kn: B
C ₂ : -0.229047	tempo: 0.367089	tem: 139
C ₃ : -0.585757	pitch: 0.459551	mnap: 86.922486
C ₄ : -0.435320	rhythm: -0.347019	unac: 0.326491
C ₅ : 0.884293	harmony: 0.436543	ac: 0.673509
C ₆ : 1.000000	melody: 0.030979	tc: 0.718272
C ₇ : -0.721753	loudness: -0.055010	oc: 0.281728
C ₈ : -0.460130	loudness variation: 0.272150	um: 0.404903
		dm: 0.373924
		lm: 0.221174
		vol: 83
		velvar: 0.340664



図7 従来方式でコンテキスト「fun」から生成された楽曲

表6 コンテキスト「terrible」の各値

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
C ₁ : -0.660379	key: -0.586986	kn: F#m
C ₂ : 1.000000	tempo: -0.118718	tem: 82
C ₃ : 0.084317	pitch: -0.309524	mnap: 54.142862
C ₄ : -0.965606	rhythm: -0.280123	unac: 0.359938
C ₅ : -1.000000	harmony: -0.330687	ac: 0.640062
C ₆ : -0.378423	melody: 0.203701	tc: 0.334657
C ₇ : -0.607578	loudness: -0.058942	oc: 0.665343
C ₈ : -0.422226	loudness variation: -0.011584	um: 0.447845
		dm: 0.244145
		lm: 0.308010
		vol: 83
		velvar: 0.222394



図8 従来方式でコンテキスト「terrible」から生成された楽曲

一方、表6の楽曲印象語群ベクトルでは、語群C₂への重み、つまり“dark”, “sad”などを表す印象が強い。特徴値を見ると、調knは短調のF#m、音高mnapは約54など、暗く重々しい印象を与えるものとなっている。図6に示した楽譜でもそのような構造上の特徴が確認できるので、コンテキスト「terrible」に対してもその印象に合致した楽曲が生成できていると言える。なお、各音の強弱について、最も強い音のvelocityは97、最も弱い音は62であった。

4.5 実験2

4.5.1 実験方法

提案方式と[9]の方式(従来方式)とで同一のコンテキストに対して生成されたふたつの楽曲について、その構造の比較を行う。

従来方式では、音量(loudness)と強弱の幅(loudness variation)というふたつの楽曲構造要素については考慮されていない。

ここでは、従来方式に実験1で用いたコンテキスト「fun」および「terrible」を与え、実験1の結果と比較する。

4.5.2 実験結果

従来方式においてコンテキスト「fun」および「terrible」から生成された楽曲をそれぞれ図7と図8に示す。また、各楽曲の楽曲印象語群ベクトル、楽曲構造要素ベクトル、特徴値をそれぞれ表7と表8に示す。なお、従来方式で生成された楽曲の音量および強弱は、また生成されたすべての楽曲において同一の値であり、各楽曲中でも変化はなく一定値となっている。

表7 従来方式でのコンテキスト「fun」の各値

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
C ₁ : -1.000000	key: 0.615808	kn: B
C ₂ : -0.229047	tempo: 0.367089	tem: 139
C ₃ : -0.585757	pitch: 0.459551	mnap: 86.922486
C ₄ : -0.435320	rhythm: -0.347019	unac: 0.326491
C ₅ : 0.884293	harmony: 0.436543	ac: 0.673509
C ₆ : 1.000000	melody: 0.030979	tc: 0.718272
C ₇ : -0.721753		oc: 0.281728
C ₈ : -0.460130		um: 0.377179
		dm: 0.346201
		lm: 0.276620

表8 従来方式でのコンテキスト「terrible」の各値

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
C ₁ : -0.660379	key: -0.586986	kn: F#m
C ₂ : 1.000000	tempo: -0.118718	tem: 82
C ₃ : 0.084317	pitch: -0.309524	mnap: 54.142862
C ₄ : -0.965606	rhythm: -0.280123	unac: 0.359938
C ₅ : -1.000000	harmony: -0.330687	ac: 0.640062
C ₆ : -0.378423	melody: 0.203701	tc: 0.334657
C ₇ : -0.607578		oc: 0.665343
C ₈ : -0.422226		um: 0.467866
		dm: 0.264166
		lm: 0.267968

4.5.3 考察

まず、従来方式でのコンテキスト「fun」に対する表7の各値と、提案方式で生成された「fun」に対する楽曲の各値(表5)

とを比較すると, [9] で考慮されていない loudness と loudness variation および *vol* と *velvar*, そしてランダムな要素を含む *um*, *dm*, *lm* を除いたすべての値が一致している. コンテキスト「terrible」に関しても, 表 6 および表 8 より同様である. 生成された楽曲についても, 従来方式で考慮されていない部分とランダムな要素を除けば, 本方式による楽曲・従来方式による楽曲のいずれも同様の特徴となっている.

本方式での変換行列 T は, 図 3 での行列 A の部分に相当する従来方式での変換行列に図 3 の行列 B の部分を付加したものであるが, この実験結果より, 本方式での変換行列 T の定義方法が従来方式で定めた演算に影響を及ぼさないことが確認できる.

また, 生成された各楽曲に対する主観的な印象では, 従来方式では音量は一定で各音の強弱はなく機械的であるが, 提案方式では抑揚があり, よりコンテキストの印象に合致していると感じる.

4.6 実験全体の考察

実験 1 では, 本方式によって楽曲生成を行い, 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, 従来方式での楽曲生成との比較を行い, 従来方式で定めた演算に影響がないこと, 本方式では従来方式に比べコンテキストの印象により合致した楽曲を生成できることを確認した.

以上の実験結果により, 本方式の実現性・有効性を示すことができたと言える.

なお, 本実験で示した 4 つの楽曲の MIDI ファイルは, <http://www.mma.cs.tsukuba.ac.jp/~yoshimura/dews07.html> にて聴取可能である.

5. おわりに

本稿では, 任意の言葉の印象と音楽心理学に基づく楽曲自動生成方式について述べた. また本方式に関する実験を行い, 利用者の意図する感情に合致した楽曲メディアデータの自動生成が可能であることを示した.

本方式の実現により, 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することの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.

今後の課題としては, 本方式に関する適切な評価方法の確立, 他の作曲方式との連携によるより自然な楽曲生成機構の実現, 画像など他メディアを対象とした連携・統合方式の実現などが挙げられる.

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. Takanari, "A Metadatabase System for Semantic Image Search by a Mathematical Model of Meaning", Multi-

media 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] T. Kitagawa, Y. Kiyoki, "Fundamental framework for media data retrieval system using media lexco transformation operator", Information Modeling and Knowledge Bases, IOS Press, 2000.
 [7] 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.
 [8] 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.
 [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 (DEWS2006) 論文集, 電子情報通信学会, 2006.
 [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] P. N. Juslin, J. A. Sloboda, "Music and Emotion," Oxford University Press, 2001.
 [15] 梅本亮夫 (編), "音楽心理学," 誠信書房, 1966.
 [16] Longman Dictionary of Contemporary English, Longman, 1987.
 [17] P. van Oostrum, "MF2T/T2MF," <http://rd.vector.co.jp/soft/dos/art/se021866.html>