

メディアコンテンツを対象とした統計的一般化逆作用素構成方式とその楽曲メディアコンテンツ生成への適用

岡田 龍太郎^{1,2,a)} 中西 崇文^{1,b)} 本間 秀典^{2,†1} 北川 高嗣²

受付日 2015年7月30日, 採録日 2016年2月8日

概要: 本稿では, メディアコンテンツを対象とした統計的一般化逆作用素の構成方式を示す. これは, 与えられた印象を表す単語とその重みで表される印象メタデータからメディアコンテンツを生成する機構を実現するものである. 本方式は, 従来, 我々が提案してきた手法であるメディアコンテンツから印象を表す言葉をメタデータとして抽出する, メディアコンテンツを対象としたメタデータ自動抽出方式手法の逆演算として構成される. しかしながら, 本逆演算においては一般的に, 不良設定問題が発生する. この不良設定問題を解決するために, 制約条件としてそのメディアの種類に応じた統計情報やそのメディアを対象とした研究成果・理論を用いる. これにより, メディアコンテンツと言葉の間の相互変換を可能とする. さらに, 本方式を楽曲メディアコンテンツに適用し, 印象語で表現された印象メタデータから楽曲メディアコンテンツを生成するシステムを実装する. これを用いて実験を行い, 提案方式が実現されていることを示す.

キーワード: 統計的一般化逆作用素, メディアコンテンツ生成, 不良設定問題, 楽曲コンテンツ

A Construction Method of Stochastic Generalized Inverse Operator and its Application for Music Data Generation

RYOTARO OKADA^{1,2,a)} TAKAFUMI NAKANISHI^{1,b)} HIDENORI HOMMA^{2,†1} TAKASHI KITAGAWA²

Received: July 30, 2015, Accepted: February 8, 2016

Abstract: In this paper, we present a construction method of Stochastic generalized inverse operator for media contents. This method realizes automatic media contents creation depending on impression words as an inverse operation of our automatic metadata extraction method which we have proposed. This is an inverse operator which extracts words from media contents. However, this creation mechanism contains ill-posed problems. In order to solve them, we introduce stochastic data about the media and its studies. Moreover, we construct a mutual conversion between impression words and media contents. Furthermore, we apply our method to music data creation system. We performed verification experiments, and showed the effectiveness of our method.

Keywords: stochastic generalized inverse operator, media contents creation, ill-posed problem, music data

¹ 国際大学グローバル・コミュニケーション・センター (GLOCOM) Center for Global Communications, International University of Japan, Minato, Tokyo 106-0032, Japan

² 筑波大学大学院システム情報工学研究科 Graduate School of Systems and Information Engineering, University of Tsukuba, Tsukuba, Ibaraki 305-8573, Japan

^{†1} 現在, 東京医科歯科大学 Presently with TOKYO MEDICAL and DENTAL UNIVERSITY

^{a)} rokada@glocom.ac.jp

^{b)} takafumi@glocom.ac.jp

1. はじめに

近年, コンピュータが人間の生活する環境に遍在するようになり, コンピュータとのコミュニケーションにおける人間の感性が着目されつつある. これまでのコンピュータと人間の論理情報の伝達だけでは, その操作による人間の負荷が大きくなっており, 人間の感性や直感に合致した, ユーザへの負担が少ないコミュニケーションメディアの実

現が重要な課題となってきた。コンピュータと人間がより密接に関わるようになっていくにつれて、人間の感性に基づく情報の獲得方式やコンピュータによる感性の表現方式はいっそう重要度を増していくことが予想される。

我々のコミュニケーションにおいて、互いの感情を正確に理解することは重要であり、言葉のほか、顔の表情、ジェスチャーなど非言語行動が様々な情報を相手に効果的に伝えるメディアとして重要な役割を担っている。このような感性的な非言語行動にあたるマルチメディアコンテンツが人間とコンピュータの間におけるコミュニケーションメディアとして導入されれば、人間の感性や直感に合致した、ユーザへの負荷が少ないインタフェースの実現が可能となると考えられる。

我々はこれまでに、様々なメディアを対象として、そのメディアコンテンツが持つ印象や感性的な情報を、言葉のメタデータとして自動抽出する枠組みである“Media-lexicon Transformation Operator” (“ ML ”) [1] の研究を行ってきた。それぞれの ML は、対象となるメディアコンテンツに対応する心理学などの専門分野における研究成果を用いることにより構成されている。 ML によって抽出される、メディアコンテンツの持つ印象や感性的な情報を、以下では“印象メタデータ”と呼ぶ。印象メタデータは、単語とその重みの組によって構成されている。これまでの“ ML ”の実現例としては、楽曲を対象としたもの [3], [4], 画像の色を対象としたもの [5], “音相”と呼ばれる言葉の発音情報を対象としたもの [6] などがある。

本稿では、上記のようなメディアコンテンツから印象メタデータを自動抽出する手法とは逆に、印象メタデータを入力として、メディアコンテンツを自動生成する新たな枠組みである、メディアコンテンツを対象とした統計的一般化逆作用素“Stochastic Generalized Inverse Media-lexicon Transformation Operator” (“ iML ”) を定義する。統計的一般化逆作用素 iML は、 ML の逆演算を一般化したものである。

そもそも、メディアコンテンツから印象メタデータを自動抽出する方式の逆演算によって印象語を入力としたコンテンツ自動生成システムを構成する際には、印象語を表す次元数とメディアコンテンツの特徴を表す次元数は異なることがほとんどであり、その次元数の不一致から、逆作用素を構成した際に、不確定要素が発生する。この不確定要素を、一般化作用素を生成する際に発生する不良設定問題と呼ぶ。そこで、不確定要素となった部分を対象とするメディアに関連する統計情報やそのメディアを対象とした研究成果・理論を導入することにより、対象メディア依存の制約条件を設けることができる。その制約条件によって、不確定要素を最小限に抑えながら、印象を表したメディアコンテンツを生成することが可能となる。メディアコンテンツ生成の過程に現れる不良設定問題を、対象とするメ

ディアに関連する統計情報を導入することによって解決することを本稿では統計的 (“stochastic”) と名付けている。

本稿では、 iML の適用事例として、楽曲メディアコンテンツ生成を対象とする。本方式は、音楽心理学者 Hevner の研究を用いて、楽曲メディアに対する ML と、その逆演算である iML を構成する。Hevner の研究 [7], [8], [9] 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 つをあげており、この 6 つの楽曲構造要素と 8 つの印象語群によって表現される印象との相関関係を調べ、表にまとめられている。Hevner の研究は楽曲の構造要素と、それに対して人間のいづく感情との対応関係を調べたものである。楽曲メディアに関する ML と、その逆演算として構成された iML を組み合わせることで、入力した楽曲メディアコンテンツの持つ印象の抽出、さらに印象を表す印象語からそれに合致した楽曲メディアコンテンツとして構成する双方向のシステムが実現できる。

本稿の構成は以下のとおりである。2章では、関連研究を示す。3章では、提案方式である iML の定義について述べる。4章では、提案方式の適用事例として、楽曲を対象とした iML の構成方式について述べる。5章では、実験により、4章で提案した方式の有効性を確認する。6章で本稿をまとめる。

2. 関連研究

本稿は、メディアコンテンツを対象とした統計的一般化逆作用素“Stochastic Generalized Inverse Media-lexicon Transformation Operator” (“ iML ”) 構成方式を示しているが、本稿の対象メディアの楽曲に焦点を絞ると、楽曲自動作曲システムの 1 つと位置付けることができる。

自動作曲の現状については、文献 [10] が明るい。

自動作曲のはじまりは、1957年に発表された「イリアック組曲」[11] であるといわれている。これは、音高のマルコフ連鎖と乱数生成によって音高列を決定するモジュールと生成結果が妥当であるかを検証するモジュールで構成されるものである。妥当性の判断は音高どうしの不協和回避ルールというルールベース的なアプローチで行われる。

現在、自動作曲システムについては様々なアプローチが提案されている。たとえば、音楽には和声法や対位法などの方法論や理論が存在することから、これらをルールとして計算機上に実装することによって作曲システムを構築する、ルールに基づくアプローチが存在する。たとえば、旋律生成システムの Ebcioğlu の研究 [12] が、それにあたる。

また、遺伝的アルゴリズムを用いたアプローチも提案されている [13], [14]。実際にユーザに聴かせたときの評価を用いることで、自然かつ嗜好に合った旋律を得ていくというものである。これらのアプローチの欠点としては、ユーザの負担が大きいことが問題としてあげられている。

一方で、既存の楽曲の断片を事例として部分的に用いて、組み合わせて用いる手法 [15] も提案されている。

さらに、近年では確率モデルを用いたアプローチも提案されている。たとえば、白井らの研究 [16] では、旋律の音の遷移の偏りを階層 Pitman-Yor 言語モデルを用いて学習し、そのモデルから旋律を生成している。また、深山らの研究 [17] では、日本語歌詞から、音楽要素の組合せによる制約条件の下で、動的計画法の音高系列の探索問題として作曲を定式化することで歌詞の韻律に合った旋律を生成することを実現している。

近年、ユーザ視点に立ち、ユーザとのインタラクションを考え、作曲の創作支援を考える実用的な研究も増えてきている。たとえば、Tsuchiya らの研究 [18] では、旋律の大まかな音の流れを曲線で入力させることにより、旋律を作曲させる手法がある。この研究の場合、ユーザ意図をどのように伝えるかというユーザインタフェースを研究の主眼としている。また、北原ら [19] は、生成結果をユーザがどう修正するかということに着目し、システムが生成した楽曲をシステムの支援を受けながら編集できる human-in-the-loop 型の自動作曲というものを提案している。

本稿で提案する我々の手法は、上記の手法とは異なり、そのメディアコンテンツが持つ印象や感性的な情報を、言葉のメタデータとして自動抽出する枠組みである ML の逆作用素 iML を構成することによって自動作曲を行うシステムである。本方式では、音楽心理学者 Hevner の研究 [7], [8], [9] を用いる。Hevner の研究は、楽曲構造要素というメディアコンテンツから抽出しうる特徴量と人間のいさだ感の関係性を心理学的に調査した研究である。また、Hevner は、クラシックのピアノ曲に対して調査を行っているため、本システムの適用範囲はクラシックのピアノ曲に限定される。

本稿では、音楽心理学者 Hevner の研究をもととして、統計的一般化逆作用素 iML を構成する。 iML を用いることで、印象メタデータを入力とし、その印象に対応する楽曲構造要素を持つ楽曲を生成することで、入力した印象を表現するクラシックピアノ曲を生成する自動作曲システムとなる。

3. メディアコンテンツを対象とした統計的一般化逆作用素 (iML)

本章では、我々がこれまで提案してきたメディアコンテンツから人間が受ける印象を言葉として自動的に抽出する方式である Media-lexicon Transformation Operator (ML) [1] を示したうえで、本稿の提案方式である、印象を表現した言葉からメディアコンテンツを生成する統計的一般化逆作用素 Stochastic Generalized Inversed Media-lexicon Transformation Operator (iML) の性質と構成方

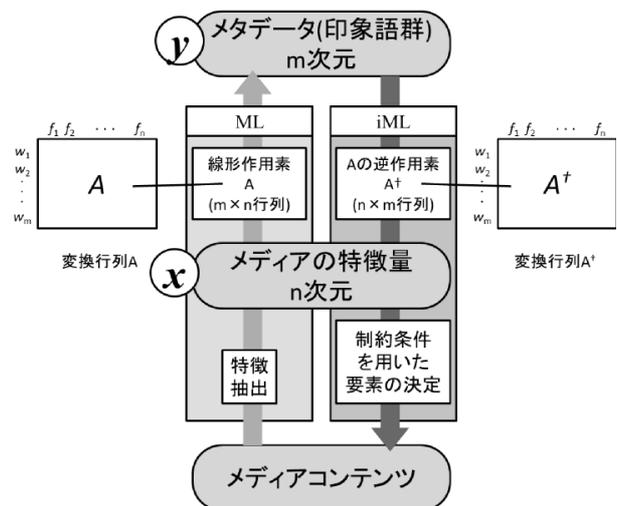


図 1 ML と iML
Fig. 1 ML and iML .

式を述べる。 iML は ML の逆演算である。 ML と iML の概略図を図 1 に示す。

3.1 節では、Media-lexicon Transformation Operator (ML) の概要を示す。3.2 節では、 ML の逆作用素であり、かつ、本稿の提案方式である、Stochastic Generalized Inversed Media-lexicon Transformation Operator (iML) について示す。

3.1 Media-lexicon Transformation Operator (ML)

Media-lexicon Transformation Operator (ML) は、対象とするメディアコンテンツに関する分野の専門家による研究や評論、統計などを用いることにより、人間がそのメディアコンテンツから受ける印象を表す言葉（以下“印象語”と呼ぶ）の抽出を実現する機構である。

ML は言葉どうしの相関を計量する機構とセットで考案されたものである。様々な種類のメディアコンテンツを言葉という統一的なメタデータで表現し、さらにそれらのメタデータどうしの関係を計量する機構と組み合わせ、メディアコンテンツの分野をまたいだ統一的な操作を行うことを目指している。

ML は一般に次のように表される。

$$ML(Md) : Md \mapsto Ws. \tag{1}$$

(Md : メディアコンテンツ, Ws : (重み付き) 印象語群)

図 1 に示すとおり、 ML によるメタデータ抽出は、メディアコンテンツから特徴量への変換と、特徴量から印象語への変換という 2 段階で構成される。

特徴量から印象語への変換には、その分野の専門家による研究や評論、統計などによる人間がそのメディアコンテンツから受ける印象を表す印象語の関係を示す研究を用いて変換行列を構成する。これらの専門家による研究はアン

ケート調査によるものが多いが、今日的にはインターネット上の情報を収集することによる手法も考えられる。変換行列は、メディアコンテンツから抽出可能な特徴量（画像であれば色彩情報など、楽曲であればテンポ・音程など）と印象語の関係を表す研究（カラーイメージスケール [5], Hevner の研究 [7], [8], [9] など）を用いて、変換行列を作成する。

変換行列は、特徴量と印象語の関係を表現したものとなり、特徴を n 個、印象語を m 個とするとき、 $m \times n$ 行列となる。ここで、 $w_1 \sim w_m$ は印象語、 $f_1 \sim f_n$ は特徴として、変換行列 A を図 1 中に示すように作成する。

変換行列 A によって、 n 次元の特徴量ベクトル x を、 m 次元の印象語ベクトル y に変換することができる。

$$y = Ax \tag{2}$$

ML により、メディアコンテンツから想起されうる印象を表すメタデータを付与することが可能となり、印象からメディアコンテンツを検索するような感性検索 [4], [5], [6] を実現することができる。

3.2 Stochastic Generalized Inversed Media-lexicon Transformation Operator (iML)

本節では提案方式である、統計的一般化逆作用素、Stochastic Generalized Inversed Media-lexicon Transformation Operator (iML) の構成方式について述べる。 iML は、印象メタデータからその印象を表現するメディアコンテンツを生成する機構である。 iML は、 ML の一般化逆作用素であり、これらは互いに対をなす存在である。 iML は一般に次のように表される。

$$iML(Ws) : Ws \mapsto Md. \tag{3}$$

(Md : メディアコンテンツ, Ws : (重み付き) 印象語群)

この逆作用素は、3.1 節に示した ML の逆演算により構成される。図 1 に示すとおり、 iML によるメディアコンテンツ生成は、印象語から特徴量への変換と、特徴量からメディアコンテンツへ変換という 2 段階で構成される。

印象語から特徴量への変換は、 ML における変換行列 A の一般化逆行列を構成することで実現する。3.1 節で示した変換行列 A から、ムーア・ペンローズ一般化逆行列を計算し、逆変換行列 A^\dagger を導出する。 A^\dagger とは、以下の性質を満たす行列のことである。

$$AA^\dagger A = A. \tag{4}$$

$$A^\dagger AA^\dagger = A^\dagger. \tag{5}$$

また、 AA^\dagger , $A^\dagger A$ はエルミート行列となり、以下の性質を持つ。なお、行列 A の随伴行列を A^* と表す。

$$(AA^\dagger)^* = AA^\dagger. \tag{6}$$

$$(A^\dagger A)^* = A^\dagger A. \tag{7}$$

逆変換行列 A^\dagger を図 1 中に示す。逆変換行列 A^\dagger によって、 m 次元の印象語ベクトル y を、 n 次元の特徴量ベクトル x に変換することができる。

$$x = A^\dagger y \tag{8}$$

ML の逆演算である iML を構成することにより、印象メタデータからメディアコンテンツの生成を実現する。

特徴量からメディアコンテンツへの変換は、 ML における特徴抽出の逆演算に対応するが、その演算は一般に不良設定問題となる。一般化逆行列のような一意な逆演算のみでメディアコンテンツを生成できればよいのだが、後半の特徴量からメディアコンテンツへの変換については自由度が大きく、一意なコンテンツを作成することができない。そこで、対象とするメディアにおける統計情報やそのメディアを対象とする研究成果・理論を新たに制約条件として導入することで、自由度を抑え、メディアコンテンツの生成を可能とする。

特徴量からメディアコンテンツを生成する段階において、ルールに基づくアプローチである、たとえば旋律生成システムの Ebcioğlu の研究 [12] などに応用することは可能である。

つまり、 iML による印象メタデータから特徴量への変換によって、印象に合致した特徴量を抽出することができ、その特徴量から、そのメディア依存のルールなどによって、メディアコンテンツの要素を確定し、メディアコンテンツを構成することができる。本方式によって、印象メタデータからその印象に合致するようなメディアコンテンツを構成可能となる。

4. 楽曲コンテンツを対象とした統計的一般化逆作用素 (iML) の構成

本章では、3 章で示した iML の具体的事例として、印象メタデータから楽曲メディアコンテンツを自動生成する方式を示す。

4.1 節では、対象とする楽曲コンテンツに関する分野の専門家による研究として用いる、Hevner の研究について述べる。4.2 節では、4.1 節で述べた Hevner の研究を用いて楽曲からメタデータを自動抽出する方式、つまり、楽曲コンテンツを対象とした ML について述べる。さらに、4.3 節で、印象メタデータから楽曲メディアコンテンツを生成する方式、つまり、本稿の提案方式である、楽曲メディアコンテンツ自動生成方式 (iML) について示す。

4.1 Hevner の研究

Hevner の研究 [7], [8], [9] では、楽曲メディアにおける特徴量である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

表 1 Hevner による 8 つの印象語群

Table 1 Hevner's 8 categories of impression words.

印象語群	含まれる印象語
C1	Dignified, Serious, Awe など
C2	Sad, Dark, Heavy など
C3	Dreamy, Tender, Sentimental など
C4	Serene, Calm, Lyrical など
C5	Graceful, Humorous, Light など
C6	Happy, Bright, Joyous など
C7	Exciting, Agitated, Dramatic など
C8	Vigorous, Emphatic, Majestic など

	key	tempo	pitch	rhythm	harmony	melody
e1	0.070	-0.244	-0.174	0.314	0.052	0.070
e2	-0.349	-0.209	-0.331	0.052	-0.122	0.000
e3	-0.209	-0.279	0.105	-0.157	0.070	0.000
e4	0.052	-0.349	0.139	-0.035	0.174	0.052
e5	0.366	0.105	0.279	-0.139	0.209	-0.052
e6	0.418	0.349	0.105	-0.174	0.279	0.000
e7	0.000	0.366	-0.137	0.035	-0.244	-0.122
e8	0.000	0.105	-0.227	0.174	-0.139	-0.139

図 2 変換行列 T

Fig. 2 Transformation Matrix T .

(melody) の 6 つをあげている。Hevner は、この 6 つの楽曲構造要素と 8 つの印象語群 (表 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。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 1 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に、円形に配置している。

4.2 楽曲メディアコンテンツを対象としたメタデータ自動抽出方式 (ML)

前節で述べた Hevner の研究を用いて楽曲から印象メタデータを自動抽出する方式の概要について述べる。詳細は文献 [3], [4] に述べられている。この方式では、Hevner の研究を用いて、楽曲と印象語の関係性を自動で計量できるように、Standard MIDI ファイルから印象メタデータを計量するまでの手順を具体的に定めている。

表 2 に、楽曲より直接抽出可能な各楽曲特徴値およびその意味を示す。さらに、表 3 に楽曲特徴値と楽曲メディアにおける Hevner の研究で定められた特徴量である楽曲構造要素の対応関係を示す。これらの関係から、楽曲を特徴値へ、特徴値を楽曲構造要素へと変換する。この変換方法の詳細については、文献 [3] に示されている。また、変換行列 T を図 2 に示す。変換行列 T は、Hevner が示した、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、さらに行列全体を 2 ノルム正規化をすることにより構成される。

4.3 楽曲メディアコンテンツ自動構成方式 (iML)

ここでは、印象メタデータから楽曲メディアコンテンツを構成する方式の各機能の詳細について示す。その流れを

表 2 楽曲の特徴値

Table 2 Features of music data.

<i>kn</i>	楽曲の調性
<i>mt</i>	拍子
<i>tem</i>	テンポ
<i>mnap</i>	旋律の音高の平均
<i>unac</i>	四分音符以上の長さの和音の総演奏時間
<i>ac</i>	四分音符未満の長さの和音の総演奏時間
<i>tc</i>	三和音の総演奏時間
<i>oc</i>	三和音以外の和音の総演奏時間
<i>nc</i>	三和音未満の和音の総演奏時間
<i>um</i>	上昇する旋律音の総演奏時間
<i>dm</i>	下降する旋律音の総演奏時間
<i>lm</i>	水平な旋律音の総演奏時間

表 3 楽曲の特徴値と楽曲構造要素の対応関係

Table 3 Relationships among features and structural elements of music data.

楽曲特徴値	楽曲構造要素
<i>kn</i>	<i>key</i>
<i>tem</i>	<i>tempo</i>
<i>mnap</i>	<i>pitch</i>
<i>unac, ac</i>	<i>rhythm</i>
<i>tc, oc, nc</i>	<i>harmony</i>
<i>um, dm, lm</i>	<i>melody</i>

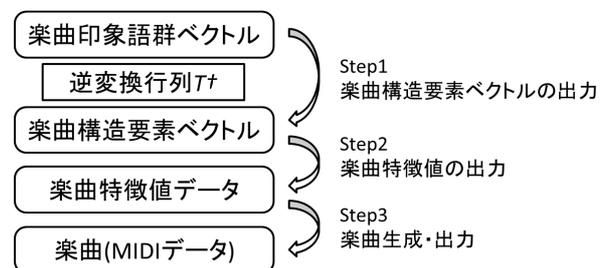


図 3 楽曲メディアコンテンツ自動構成方式の流れ

Fig. 3 Overview of the media contents generation method for music.

図 3 に示す。図 3 のとおり、「楽曲構造要素ベクトル出力」、「楽曲特徴値データ出力」、「楽曲生成・出力」の大きく 3 つのステップからなる。

4.3.1 項では、Step1 として、楽曲構造要素ベクトル出力について示す。4.3.2 項では、Step2 として、楽曲特徴値出力について示す。さらに、4.3.3 項では、Step3 として、楽曲生成・出力について示す。

4.3.1 楽曲構造要素ベクトル出力 (図 3 : Step1)

逆変換行列 T^\dagger の生成

3.2 節で示した方式を用いることによって、図 2 に示した変換行列 T から、ムーア・ペンローズ一般化逆行列を計算し、逆変換行列 T^\dagger を導出する。具体的には図 4 のようになる。これが、iML のコアとなるべき、印象メタデータから楽曲構造要素に変換する行列

	c1	c2	c3	c4	c5	c6	c7	c8
key	-0.048	3.466	-4.947	6.726	3.141	-0.775	5.398	-3.403
tempo	0.063	-2.550	1.882	-4.760	-2.146	1.279	-2.820	1.519
pitch	-0.155	-4.226	2.028	-2.954	-0.631	-1.560	-2.465	1.209
rhythm	2.096	-4.914	3.016	-5.756	-2.188	-0.303	-4.799	3.909
harmony	1.295	-3.138	5.905	-7.057	-2.625	3.301	-7.006	5.116
melody	0.376	1.602	-5.154	2.235	-1.780	-0.027	1.956	-7.191

図 4 変換行列 T^\dagger

Fig. 4 Transformation matrix T^\dagger .

となる。

楽曲印象語群ベクトルの構成

印象メタデータは、8つの楽曲印象語群に対応する要素からなるベクトルとして表現される。楽曲印象語群ベクトル c を構成する。 c は次のように定義される。 C_1, \dots, C_8 はそれぞれ、Hevner が定めた楽曲印象語群に対応する。

$$c = (C_1, C_2, \dots, C_8)^T. \tag{9}$$

ここで、要素である C_1, \dots, C_8 をそれぞれ $-1 \sim 1$ として設定し、 c を構成する。これが印象メタデータに相当する。

楽曲構造要素ベクトルの構成

式(10)のように、楽曲印象語群ベクトルに対して変換行列 T^\dagger 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 を構成する。

$$f = (\text{key}, \text{tempo}, \text{pitch}, \text{rhythm}, \text{harmony}, \text{melody})^T = T^\dagger c. \tag{10}$$

楽曲構造要素ベクトルの正規化

楽曲構造要素ベクトルの各要素は、楽曲特徴値との関係上、各要素に対応する性質が最大である状態を1、最小である状態を-1とし、-1~1の値を持つように定義されている。しかし入力する楽曲印象語群ベクトルの値によってはその範囲に収まらなくなるため、正規化を行う。正規化後の楽曲構造要素ベクトルを f' とし、以下の式で表す。

$$f' = \begin{cases} f & \|f\|_\infty < 1, \\ \frac{f}{\|f\|_\infty} & \text{otherwise.} \end{cases} \tag{11}$$

4.3.2 楽曲特徴値出力 (図3: Step2)

出力された楽曲構造要素ベクトルの値をもとに、表2に示した、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。これらを、文献[3]に示されている抽出操作の逆手順になるように定義していく。この定義は、心理学の研究成果や楽典を参考に構成している。これによって、自然な楽曲メディアコンテンツ生成を可能とする。

調性 kn の決定

調性度 key の値から、生成する楽曲の調 kn を定める。 key は、楽曲データの調名データ kn および、調性度の

表 4 各調の調性度 (key)

Table 4 Tonality of each key.

調	調性度 (key)	key の範囲	
G	1	0.96875	~ 1.00000
C	0.9375	0.89583	~ 0.96875
A	0.85416667	0.81250	~ 0.89583
D	0.77083333	0.72917	~ 0.81250
F	0.6875	0.65625	~ 0.72917
B	0.625	0.56250	~ 0.65625
F#	0.5	0.46875	~ 0.56250
E	0.4375	0.40625	~ 0.46875
Eb	0.375	0.34375	~ 0.40625
Ab	0.3125	0.21875	~ 0.34375
Bb	0.125	0.06255	~ 0.21875
Db	0.0001	0.00000	~ 0.06255
Am	-0.0001	-0.10422	~ 0.00000
Gm	-0.2083333	-0.31250	~ -0.10422
G#m	-0.4166667	-0.48958	~ -0.31250
Bm	-0.5625	-0.59375	~ -0.48958
Em	-0.625	-0.65625	~ -0.59375
F#m	-0.6875	-0.69492	~ -0.65625
Dm	-0.7023333	-0.74100	~ -0.69492
Fm	-0.7796667	-0.80650	~ -0.74100
C#m	-0.8333333	-0.87500	~ -0.80650
Bbm	-0.9166667	-0.95833	~ -0.87500
Cm	-1	-1.00000	~ -0.95833

表4を利用して決定する。この表は、梅本の研究[20]による、各調の長調らしさ、短調らしさを表す値を-1~1に正規化して作成された、文献[3]に示されている表を拡張したものである。元々の表では2つ以上の調に同一の調整度が割り当てられている場合があり、そのままでは逆変換の際に調が一意に復元できなくなってしまうため、すべての調が異なる調性度を持つように修正を加えた。具体的には、2つ以上の調に同一の調性度が割り当てられている場合には、前後の調の持つ調性度の間隔を同一の調性度に割り当てられている調の数で等分割し、分割したそれぞれの点に1つの調をランダムに選んで対応させた。

テンポ tem の決定

曲のテンポの楽曲群における相対的な速さに関する楽曲構造要素 $tempo$ と、生成される楽曲のテンポの値 tem の関係式は文献[3]中で次のように定義される。

$$tempo = \begin{cases} \frac{tem - tstand}{tmax - tstand} & tem - tstand \geq 0, \\ \frac{tem - tstand}{tstand - tmin} & tem - tstand < 0. \end{cases} \tag{12}$$

ここで、 $tmax, tmin, tstand$ は、それぞれ M.M (メルツェル・メトロノーム) の最大値 (=184)、最小値 (=40)、基準値 (=88) とした。M.M は一般的なメトロノームにおけるテンポを示している。上式を変形し、 $tempo$ から tem を求められる形にする。変形した式を以下に示す。

$$tem = \begin{cases} tstand + (tmax - tstand)tempo & (tempo \geq 0), \\ tstand + (tstand - tmin)tempo & (tempo < 0). \end{cases} \quad (13)$$

音高 $mnap$ の決定

楽曲の楽音の高低の楽曲群における相対的な高さに関する楽曲構造要素 $pitch$ と、生成される楽曲の楽音の高さの平均 $mnap$ の関係式は文献 [3] 中で次のように定義される。

$$pitch = \begin{cases} \frac{mnap - pstand}{pmax - pstand} & mnap - pstand \geq 0, \\ \frac{mnap - pstand}{pstand - pmin} & mnap - pstand < 0. \end{cases} \quad (14)$$

ここで、 $pmax$, $pmin$, $pstand$ は、それぞれ $mnap$ が最大のとき、最小のとき、標準のときの MIDI ファイルの音高番号であるが、これは楽器に依存する値である。今回はピアノ曲を対象としているため、ピアノ曲においてとりうる値の最大最小を考慮して決定する。ピアノの最高音にあたる MIDI ファイルの音高番号は 108、最低音は 21 であるが、本稿では予備実験により範囲を狭めて、 $pmax$ (= 84) (最高音から 2 オクターブ下のド)、 $pmin$ (= 45) (最低音から 2 オクターブ上のラ)、 $pstand$ (= 64) (中央ミ) とした。上式を変形し、 $pitch$ から $mnap$ を求められる形にする。変形した式を以下に示す。

$$mnap = \begin{cases} pstand + (pmax - pstand)pitch & (pitch \geq 0), \\ pstand + (pstand - pmin)pitch & (pitch < 0). \end{cases} \quad (15)$$

また、文献 [3] 中では $mnap$ は全楽音データの長さの平均と定義されていたが、予備実験において、伴奏の和声が複雑な場合に音が多くなり、平均が下に引っ張られるため、音高値 $mnap$ が同じ曲でも主旋律の高さが高くなってしまふことから、 $mnap$ を、旋律音の音高平均 $mnap_{melody}$ と伴奏音の音高平均 $mnap_{chrord}$ の平均へと定義を変更した。

$$mnap = \frac{mnap_{chrord} + mnap_{melody}}{2}. \quad (16)$$

リズムの特徴 $unac$, ac の決定

和音の音価 (音の長さ) ごとの総演奏時間である固定、流動リズム値 $unac$, ac を定める。 $unac$ は楽曲における四分音符以上の音価を持つ楽音の総演奏時間、 ac は四分音符未満の音価を持つ楽音の総演奏時間と定義する。 $rhythm$ は $unac$ と ac から定義される、楽曲群における相対的なリズムの固定性に関する楽曲構造要素である。

$$rhythm = \frac{unac - ac}{unac + ac}. \quad (17)$$

式 (17) を変形し、 $unac$ と ac を求める。ここで、 $rhythm$ は $unac$ と ac の比率さえ保持していれば同一の値になるため、 $unac$ と ac の関係を比率ととらえることとし、仮に $unac + ac = 1$ とすると $unac$, ac を一意に定めることができる。

$$\begin{aligned} unac &= \frac{1 + rhythm}{2}, \\ ac &= \frac{1 - rhythm}{2}. \end{aligned} \quad (18)$$

和声の特徴 tc , oc , nc の決定

和音の種類ごとの演奏時間である単純、複雑、ノーハーモニー値 tc , oc , nc を定める。 tc はダイアトニックコードに含まれる 3 和音の総演奏時間、 oc は tc 以外の 3 和音以上の音数を持つ和音の総演奏時間、 nc は 3 和音未満の音数の和音の総演奏時間と定義する。 $harmony$ は tc , oc , nc から定義される、楽曲群における相対的な和声の単純さに関する楽曲構造要素である。 $harmony$ と tc , oc , nc の関係式は、文献 [3] 中で次のように定義される。

$$harmony = \frac{tc - oc}{tc + oc + nc}. \quad (19)$$

式 (19) を変形し、 tc , oc , nc を求める。ここで、 $harmony$ は tc , oc , nc の 3 つの値の比率さえ保持していれば同一の値になるため、 tc , oc , nc の関係を比率ととらえることとし、仮に $tc + oc + nc = 1$ とすると、 tc , oc は以下のように表現できる。

$$\begin{aligned} tc &= \frac{1 - (nc - harmony)}{2}, \\ oc &= \frac{1 - (nc + harmony)}{2}. \end{aligned} \quad (20)$$

この場合、 nc を決めない限り tc , oc を一意に決定することはできない。また、 $harmony$ の大きさによって nc のとりうる値が変化する。ここで、 $|harmony| \leq 1$ であるため、この条件と式 (20) より以下の関係が導かれる。

$$0 \leq nc \leq |harmony|. \quad (21)$$

nc の値がこの条件を満たさない場合、 tc , oc , nc から $harmony$ を導出する際に値を保持することができなくなる。よって nc の値は、この範囲における任意の値をとることができる。本稿では、生成される楽曲にバリエーションを持たせる目的から、とりうる値の範囲の中心を中心とした正規分布から確率的に決定することとした。 nc の値を、とりうる値の範囲の中心を中心とした正規分布から確率的に決定する。 nc を決定することができれば、式 (20) 中の tc , oc は一意に決定することができる。

旋律の特徴 um , dm , lm の決定

旋律の音程の種類 (次の音への変化による分類) ごと

の総演奏時間を表す上昇, 下降, 水平旋律値 um , dm , lm を定める. um , dm , lm はそれぞれ, 旋律を構成している楽音を, 注目している楽音の次の楽音が, 注目している楽音に対して高い音であるか, 低い音であるか, 同じ高さであるかに分類した際の, それぞれの楽音の総演奏時間と定義する. $melody$ は um , dm , lm から定義される, 楽曲群における相対的な旋律の上昇性に関する楽曲構造要素である. $melody$ と楽曲データの um , dm , lm の関係式は, 文献 [3] 中で次のように定義される.

$$melody = \frac{um - dm}{um + dm + lm}. \quad (22)$$

ここで, $melody$ は式 (22) より, $-1 \sim 1$ の値をとるが, $melody$ がその上限値または下限値であるということはきわめて特殊な状態である. たとえば $melody$ が 1 であるということは旋律音がすべて上昇列で構成されているということになるが, それではすぐ使用可能な音域の上限に達してしまう.

ここで仮にメロディー生成の制約として, “長音階の構成音のみ用いる”, “音程は最大 8 度 (1 オクターブまで)” の 2 つを与えたとすると, $melody$ が最大になるのは, 上昇はすべて 2 度音程で, 下降はすべて 8 度音程になるときである. このとき, 8 度音程は 2 度音程 7 つ分であるので, $um : dm = 7 : 1$ の状態, つまり, $um = 0.875$, $dm = 0.125$, $lm = 0$ のとき $melody$ は最大値をとり, 式 (22) により $melody = 0.75$ となる. um と dm の関係が逆になる場合には $melody$ の符号は反転し絶対値は同じとなるため, 上記の制約における $melody$ のとりうる値の範囲は, $-0.75 \sim 0.75$ となる. 以上をふまえ, 制約によって決まる $melody$ の上限値と下限値のうち上限値を $limit$ として定義する. $limit$ の値は, 統計的な研究などから音程の出現確率を与えた際の, 上昇音と下降音の生起確率の比 ($um : dm$) が最大, かつ, 音程の移動量の期待値が 0 となる場合の, $melody$ の値とする. $limit$ を用いて $melody$ を以下の式に従って正規化する. 正規化後の値を $melody'$ とする.

$$melody' = limit \cdot melody. \quad (23)$$

実際に用いる $limit$ の値については, 旋律生成の制約条件に依存するため, 楽曲の生成・出力 (4.3.3 項) 時に述べる.

式 (22) を変形し, um , dm , lm を定める. ここで, $melody$ は um , dm , lm の 3 つの値の比率さえ保持していれば同一の値になるため, um , dm , lm の関係を比率ととらえることとし, 仮に $um + dm + lm = 1$ とすることで, um , dm , lm を一意に定めることができる.

$$\begin{aligned} um &= \frac{1 - (lm - melody')}{2}, \\ dm &= \frac{1 - (lm + melody')}{2}. \end{aligned} \quad (24)$$

和声の場合と同様に, 式 (24) 中の水平旋律音 lm を決定する. ここで, $|melody'| \leq 1$ であるため, この条件と式 (24) より以下の関係が導かれる.

$$0 \leq lm \leq |melody'|. \quad (25)$$

よって lm の値は, この範囲における任意の値をとることができる. 予備実験により, 本稿では lm の値を, とりうる値の範囲の中心とした正規分布から確率的に決定する. それによって um , dm を一意に決定することができる.

4.3.3 楽曲生成・出力 (図 3 : Step3)

楽曲の特徴値に沿って楽曲を生成する. 以下に楽曲データの生成手順を示す.

生成する楽曲の概要

生成する楽曲は有限の小節数を持ち, 調性を持ち, 旋律と伴奏を持つとする. 途中で調やテンポ, 拍子の変更は含まれない. また, 音色にあたる楽器はピアノを想定する. 旋律および伴奏に用いる楽音の音価は四分音符と八分音符に限定する. ただし, 最後の 1 小節における旋律の楽音の音価は 1 小節分とする.

テンポ

tem の値をそのまま楽曲の M.M (メルツェル・メトロノーム) におけるテンポの値とする. M.M は一般的なメトロノームにおけるテンポを示している.

小節数

楽曲の小節数を $size$ とする. 本稿における実験システムでは $size$ は 8 とした.

拍子

楽曲の拍子を mt を分子分母に分け n/m とする. 本稿における実験システムでは, $4/4$ とした.

和音による伴奏

決定した特徴値のうち, kn , $unac$, ac , tc , oc , nc 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) 曲中で使用する和音を kn によって決定する. 伴奏に用いる和音はその調のダイアトニックコードとする. なお, 短調の曲に関しては和声的短音階に準じる.
- (2) $unac$ は四分音符以上の音価を持つ楽音の総演奏時間を表し, ac は四分音符未満の音価を持つ楽音の総演奏時間を表すが, ここでは, $unac$ を四分音符の演奏時間割合, ac を八分音符の演奏時間割合と限定して割り当てる. $size$ と $unac$ 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. $unac$ と ac の値を, それぞれ四分音符と八分音符の和音の演奏時間の割合に対応させ

る。四分音符と八分音符の割合が決まったら、音価の配列として表現し、配列をランダムにシャッフルする。

- (3) 作曲に用いる和音は、和声的短音階における 7th を含むダイアトニックコードの I~VII の和音を用いる。なお、ここで短調の I の和音のみ、和声的短音階におけるダイアトニックコードの 7th が長七度であるところを短七度としている。和声法を参考にし、コード進行はトニック・サブドミナント・ドミナントの関係 considering して決定する。1 小節に入る和音の種類としては、I~VII のどれか 1 種類だけとし、最後の 1 小節は必ず I 度の和音とする。
- (4) *tc*, *oc*, *nc* の値をもとに和音を割り当てる。それぞれ、*nc* には「2 和音 (root と 3rd)」, *tc* には「3 和音 (2 和音 + 5th)」, 「4 和音 (3 和音 + 7th)」を割り当てる。

旋律の生成

特徴値のうち *kn*, *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。旋律に使用する音はその調の音階に沿ったものとする。
- (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) *size* と *um*, *dm*, *lm* から、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し、音価の配列を作る。
- (4) 配列をシャッフルし、上で定めた上昇音、下降音、水平音の順序を決定する。
- (5) 曲の最後の音が調の基音となるようにし、旋律の上下幅はそこから上下 1 オクターブの、最大 2 オクターブとする。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、旋律の長さは原則として四分音符と八分音符のみとするが、曲が終わった感じを出すため、最後の 1 小節は主音とし、1 小節分の長さとする。
- (6) 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。上昇音と下降音における音程の変化量は、統計的な情報を用いて決定する。統計情報として文献 [20] における、隣接する 2 つの旋律音の音程とその出現する割合についての研究を用いる。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 5)。この統計情報を制約条件として旋律を確率的に決定していく。この際、4.3.2 項の式 (23) における *limit* は、0.5538

表 5 歌曲の旋律音における各音程の出現頻度

Table 5 Occurrence rate of each pitch of the melody.

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

となる。

音高の修正

音の高さの平均値が *mnap* に等しくなるように伴奏と旋律の高さを調節する。伴奏や旋律をシフトまたは変形して、音高平均値が *mnap* が等しくなるよう近づけていく。まず、旋律は伴奏より上にくるように 1 オクターブずつ平行移動する。次に、伴奏の和音 (楽音のセット) を 1 つずつ 1 オクターブ分シフトして、平均値を *mnap* に近づけていく。その処理では発散してしまう場合は、さらに和音の最低音を固定したまま、その他の音を 1 つずつ 1 オクターブ分シフトし、音高を微調節する。

5. 実験

本稿で示した方式のシステムを構築し、検証実験を行う。

実験 1 では、*iML* を実現したシステムによって実際に楽曲が構成されることを実証する。楽曲印象語群ベクトルを入力として、楽曲を構成する例を示す。

実験 2 では、*ML* と *iML* の逆作用素としての関連性の検証を行う。既存の楽曲から *ML* を用いて印象を抽出し、その印象を *iML* の入力として楽曲を生成することで、逆演算において保持される情報と保持されない情報について比較検討する。具体的には、抽出時と生成時に現れる楽曲構造要素ベクトル、楽曲特徴値、楽曲をそれぞれ比較する。

実験 3 では、実験システムによって構成された楽曲が、指定した印象を判別可能であるほどの表現力を有しているかを検証する。方法としては、2 つの異なる楽曲をもとに *ML* を用いて楽曲印象語群ベクトルを抽出し、片方の楽曲印象語群ベクトルから *iML* を用いて楽曲を生成し、元になった楽曲をあてる被験者実験を行う。

5.1 実験システム

実験システムを、Python 言語および Perl 言語を用いて実装した。本システムは、WWW ブラウザをユーザインタフェースとし、任意の楽曲印象語群ベクトルを入力として、楽曲を生成し、提示する。楽曲印象語群ベクトルの入力には、テキストファイルで事前に用意して与えることと、各語群 $C1, \dots, C8$ のそれぞれについて、ブラウザ上から -1 から 1 の値を選んで入力とすることができる。楽曲は Standard MIDI File の形式で出力される。

5.2 実験 1 (楽曲生成の実証)

5.2.1 実験方法

楽曲印象語群ベクトルを指定し、実験システムを用いて楽曲を 2 曲生成する。今回、印象メタデータとして入力とした楽曲印象語群ベクトルは、明るい曲とその反対の暗い曲を示すことを想定したものである。明るい曲を表現する楽曲印象語群ベクトルの要素は、 $C6$ (“Happy, Bright” など) と $C5$ (“Graceful, humorous” など) の値を “1” とし、 $C2$ (“Sad, Dark” など) を “ -1 ”， $C3$ (“Dreamy, Tender” など) を “ -0.5 ” とし、残りは “0” とした。反対の暗い曲については、明るい曲の楽曲印象語群ベクトルの要素の符号をそれぞれ反転し、 $C6$ および $C5$ を “ -1 ”， $C2$ を “1”， $C3$ を “ 0.5 ” とした。

5.2.2 実験結果

楽曲生成の中間データである楽曲構造要素ベクトルと楽曲特徴値の各値を表 6 に示す。生成された楽曲の楽譜を図 5、図 6 に示す。

図 5 および図 6 から、実際に楽曲が生成されていることが確認できる。また、楽曲の仕様として指定した制約を逸脱しておらず、想定どおりの楽曲生成が行われていることが確認できた。表 6 の楽曲構造要素ベクトルを比較すると、各構造要素についての絶対値が等しく、符号は逆になっていることが分かる。

明るい曲を作ることを想定した曲（以下、明るい曲）では、音高 $mnap$ が大きく、暗い曲を作ることを想定した曲（以下、暗い曲）では $mnap$ は小さい。楽曲においてもそれが反映されており、明るい曲の楽音は全体的に高い音で構成されており、反対に暗い曲では低い音で構成されている。2 つの曲の楽曲特徴値を比べると、調性 kn は明るい曲では長調のト長調になっており、暗い曲では短調のハ短調になっている。テンポ tem は、標準である 88 から、明るい曲では 140 と速く、暗い曲では 62 と遅くなっている。これらの楽曲特徴量は、入力の印象メタデータに対して想定したとおりとなっている。

「明るい」「暗い」という印象とは直接関係しないと思われる、その他の特徴について述べる。 $rhythm$ に対応する ac と $unac$ は、この 2 曲ではちょうど逆になっている。それを反映して、明るい曲では伴奏に四分音符が多く、暗い

表 6 実験 1：作曲した二曲の各値

Table 6 Each value of the Experiment 1.

楽曲構造要素ベクトル	楽曲特徴値	
	明るい曲	暗い曲
key	1.0000	-1.0000
tempo	0.5398	-0.5398
pitch	0.7432	-0.7432
rhythm	0.6669	-0.6669
harmony	0.6271	-0.6271
melody	-0.6056	0.6056

	楽曲特徴値	
	明るい曲	暗い曲
kn	G	Cm
mt	4/4	4/4
tem	140	62
mnap	79	50
unac	0.8335	0.1665
ac	0.1665	0.8335
tc	0.7127	0.0354
oc	0.0856	0.6625
nc	0.2017	0.3021
um	0.3008	0.6273
dm	0.5848	0.3433
lm	0.1143	0.0295



図 5 実験 1：明るい印象メタデータから生成した楽曲

Fig. 5 Experiment 1: Generated music data from bright impression metadeta.

曲では八分音符が多い。和声 $harmony$ に関する特徴値 tc , oc , nc について見ていくと、絶対値は異なるものの、3 和音にあたる tc と 4 和音にあたる oc の値の大小関係が逆転している。楽曲でもそれが反映され、明るい曲は 3 和音が多く、暗い曲では 4 和音が多い。同様に、旋律 $melody$ に関する特徴値 um , dm , lm についても、旋律の上昇に相当する um と下降に相当する dm の割合が逆転している。

以上の結果から、実験によって、入力した印象メタデータを反映した楽曲が生成できているといえる。しかしながら、本方式においては、 ML と iML が正しく逆の演算になっていることが、印象メタデータと楽曲の印象が正しく対応しているということによって示されることが重要である。この点について、実験 2、実験 3 でより深く検討していくこととする。

5.3 実験 2 (保持される情報の検証)

5.3.1 実験方法

既存の楽曲から ML を用いて印象を抽出し、その印象を iML の入力として楽曲を生成することで、逆演算において保持される情報と保持されない情報について比較検討する。具体的には、抽出時と生成時に現れる楽曲構造要素ベクトル、楽曲特徴値、楽曲をそれぞれ比較する。実験に



図 6 実験 1：暗い印象メタデータから生成した楽曲

Fig. 6 Experiment 1: Generated music data from dark impression metadata.



図 7 「ハンガリー舞曲 No.5」の冒頭 8 小節

Fig. 7 Opening 8 bars of Hungarian Dance No.5.



図 8 「ハンガリー舞曲 No.5」を入力として生成された楽曲

Fig. 8 Generated music data in the case of Opening 8 bars of Hungarian Dance No.5.

用いる既存の楽曲として、「ハンガリー舞曲 No.5」と「家路」の 2 曲について、印象抽出と楽曲生成を行った。

5.3.2 実験結果

印象抽出の元になった楽曲と、抽出された印象メタデータから生成された楽曲の例を示す。「ハンガリー舞曲 No.5」の冒頭 8 小節の楽譜を図 7 に、生成された楽曲を図 8 に示す。「家路」の冒頭 8 小節の楽譜を図 9 に、生成された楽曲を図 10 に示す。印象抽出時と楽曲生成時の楽曲特徴値、楽曲構造要素ベクトルをそれぞれ表 7、表 8 に示す。

抽出と生成の変換を経る中で、楽曲の情報として保持されているものと、保持されていないものに分けて述べる。2 つの楽曲について、楽曲構造要素ベクトルはすべて保持されている。わずかな差は丸め誤差によるものと思われる。楽曲特徴値の中で情報が保持されているのは、調 (kn)、テンポ (tem)、音高 ($mnap$)、伴奏の四分音符の演奏時間割合 ($unac$)、伴奏の八分音符の演奏時間割合 (ac) となる。

保持されていないのは、拍子、和声に関する特徴値 (tc , oc , nc)、旋律に関する特徴値 (um , dm , lm) となる。この中で、生成のたびに値が変化するものに“*”を付けて示



図 9 「家路」の冒頭 8 小節

Fig. 9 Opening 8 bars of From the New World.



図 10 「家路」を入力として生成された楽曲

Fig. 10 Generated music data in the case of Opening 8 bars of From the New World.

表 7 「ハンガリー舞曲 No.5」における抽出時と生成時の各値

Table 7 Each value in the case of Hungarian Dance No.5.

	楽曲構造要素ベクトル		楽曲特徴値		
	原曲	生成時	原曲	生成時	
key	-0.6875	-0.6875	kn	F#m	F#m
tempo	0.4583	0.4583	mt	2/4	4/4
pitch	-0.3684	-0.3684	tem	132	132
rhythm	-0.8654	-0.8655	mnap	57	57
harmony	0.1217	0.1217	unac	0.0673	0.0673
melody	-0.0870	-0.0870	ac	0.9327	0.9327
			tc	0.1217	*0.5186
			oc	0.0000	*0.3970
			nc	0.8783	*0.0844
			um	0.4463	*0.2951
			dm	0.4463	*0.3359
			lm	0.0667	*0.3690

した。

楽曲特徴値に着目すると、 $harmony$ に関する特徴値である tc , oc , nc と、旋律に関する特徴値である um , dm , lm は、原曲と生成時で大きく違っている。このようになる理由としては、1 つの楽曲構造要素が 3 つの楽曲特徴値によって決まる関係にあるため、逆演算の際の値の決定に自由度が含まれてしまうからである。この自由度による特徴値の違いがあっても、楽曲の印象が保持されているかについての被験者調査を実験 3 で示す。

5.4 実験 3 (作曲システムの表現力の検証)

5.4.1 実験方法

システムによって構成された楽曲が、指定した印象を表現できているのかを被験者実験によって検証する。2 つの異なる楽曲から ML を用いて楽曲印象語群ベクトルを抽

表 9 実験 3：実験に用いた楽曲

Table 9 Experiment 3: Tunes used in experiments.

ID	曲名 (作者)	調の長短	小節数	楽曲印象語群ベクトルの最大～第三要素
1	ワルツ第 7 番 (ショパン)	短調	32	C7, C2, C8
2	カノン (パッヘルベル)	長調	10	C1, C5, C8
3	ラ・カンパネラ (リスト)	短調	21	C3, C6, C4
4	前奏曲 Op.28-7 (ショパン)	長調	16	C1, C5, C4
5	ウィリアム・テル序曲 (ロッシーニ)	長調	48	C6, C7, C5
6	新世界より「家路」(ドヴォルザーク)	長調	15	C4, C5, C1
7	埴生の宿「イングランド民謡」	長調	16	C6, C5, C4
8	ソナタ「ワルトシュタイン」第一楽章 (ベートーベン)	長調	13	C6, C7, C8
9	ソナタ「月光」第一楽章 (ベートーベン)	短調	16	C2, C1, C8
10	飛翔 (シューマン)	短調	16	C7, C2, C3
11	ハンガリー舞曲 5 番 (ブラームス)	短調	34	C2, C7, C3
12	前奏曲 Op.28-4 (ショパン)	短調	26	C3, C2, C4

表 8 「家路」における抽出時と生成時の各値

Table 8 Each value in the case of the New World.

	楽曲構造要素ベクトル		楽曲特徴値		
	原曲	生成時	原曲	生成時	
key	0.9375	0.9375	kn	C	C
tempo	-0.8750	-0.8750	mt	4/4	4/4
pitch	0.3000	0.3000	tem	46	46
rhythm	0.2942	0.2941	mnap	70	70
harmony	-0.1955	-0.1955	unac	0.6471	0.6471
melody	-0.0018	-0.0018	ac	0.3529	0.3529
			tc	0.3152	*0.1522
			oc	0.5107	*0.3478
			nc	0.1741	*0.5000
			um	0.4463	*0.2496
			dm	0.4472	*0.2504
lm	0.1065	*0.5000			

出し、片方の楽曲印象語群ベクトルを *iML* の入力として楽曲を生成し、元になった楽曲をあてる被験者実験を行う。楽曲の印象を判断する際には、楽曲が長調であるか短調であるかが最も大きな影響を及ぼすと考えられるため、その1点のみの評価とならないように、長調の曲どうしまたは短調の曲どうしの比較においても楽曲の印象が判断できるかを検証する。印象を抽出する既存の楽曲を12曲用意し、「長調-短調」の組を2問、「長調-長調」の組を2問、「短調-短調」の組を2問、計6問の実験とする。実験に用いる、印象抽出の元となる楽曲を表9に示す。また、これらの楽曲が長調であるか短調であるかを表9に示す。またこれらの楽曲は、曲の冒頭から、転調やテンポの大幅な変化などの大きな曲調の変化を含まない範囲を著者が任意の小節数分抜粋したものである。抜粋した小節数を表9に示している。また、表9には、各楽曲から抽出した印象メタデータである楽曲印象語群ベクトルの要素のうち、最大のものからと3番目に大きいものまでを示している。なお今

回のケースでは、これら3番目までの楽曲印象語群ベクトルの要素の値はすべて正となっている。

被験者は、印象の抽出元にあたる楽曲を2曲聴き、続けて、自動生成によって得られた楽曲のうち1曲のみを聴き、初めに聴いた2曲のうちのどちらにより似ているかを答える。選んだ曲が生成元であれば正解とする。被験者は10人とした。また、生成された楽曲の音楽的・芸術的評価についても、自由記述欄を設けて、被験者に記入してもらった。

5.4.2 実験結果 3

実験結果を表10に示す。最も高い正解率は100%、最も低い正解率は70%、平均正解率は90%と、全体的に高い正解率となった。実験結果から、作曲システムが、楽曲印象語群として表された楽曲の印象を、楽曲として表現する能力を持っているといえる。長調どうし、短調どうしの楽曲においても高い正解率となっており、楽曲の印象を楽曲の調の長短以外の要素によっても表現できているといえる。

楽曲印象語群ベクトルに着目すると、ID7とID8の実験セットにおいては、どちらも最大要素がC6となっており、印象メタデータが比較的類似しているケースといえるが、これも正解率は100%となっている。ここから、楽曲印象語群ベクトルの各要素は、特に大きい値を持つ要素だけではなく、各要素が複合的に影響して楽曲の印象を決めているといえる。

ID3「ラ・カンパネラ」とID4「前奏曲 Op.28-7」の実験セットが比較的低い正解率となった原因としては、「ラ・カンパネラ」の抜粋した部分のメロディーのほとんどが16分音符で構成されているにもかかわらず、現状の楽曲生成システムでは4分音符と8分音符のみを用いて楽曲を生成しているため、16分音符で構成されるメロディーを再現していないことなどが考えられる。これは作曲システムの制約を変更することで改善が可能であると考えられる。

表 10 実験 3 の結果

Table 10 Result of Experiment 3.

実験セット	曲の長短	正答率
1 と 2	長-短	80%
3 と 4		70%
5 と 6	長-長	90%
7 と 8		100%
9 と 10	短-短	100%
11 と 12		100%
		平均正答率
		90%

被験者による楽曲の音楽的・芸術的評価としては、1.「メロディーがつかみにくい。和声が悪ければメロディーがより際立つのではないか」、2.「曲調が淡々としている」、3.「楽曲のフレーズが途中から始まっているように感じた」、といった意見があがった。一方 4.「響きがおかしいとは感じなかった」という意見もあった。1.「メロディーがつかみにくい。和声が悪ければメロディーがより際立つのではないか」については、アヴェイラブル・ノートスケールを考慮した制約を加えることで改善が可能であると考えられる。また、本システムにおいては音量が考慮されていないため、メロディー部分の音量を大きくする制約を加えることや、音符の小節内での位置における音量を考慮した制約を加えることによっても改善が可能であると考えられる。2.「曲調が淡々としている」については、そもそも本システムは単一の印象を表す楽曲を生成することを目的としているため一概に悪い結果とはいえないが、本システムを拡張し起伏のある楽曲を構成する方法としては、複数の印象を入力として与え、それらの印象の遷移を連続的に表現するような楽曲を生成するシステムを構築することが考えられる。3.「楽曲のフレーズが途中から始まっているように感じた」については、曲の末尾と同様に曲の冒頭に制約を加えることで改善が可能であると考えられる。これらは今後の課題とする。

5.5 考察

実験 1 により、実験システムが提案方式を想定どおり実装できており、メディアコンテンツ生成において生じる不良設定問題を解決して楽曲が生成できることが確認できた。実験 2 により、 ML による印象抽出と iML による楽曲生成の過程において、保持される情報と保持されない情報があることが確認できた。実験 3 により、変換の過程で保持されない情報があるにもかかわらず、生成した楽曲が印象を判別するのに十分な情報を保持していることが確認できた。以上により、楽曲を対象とした iML が実現されていること、およびその有効性を示した。

今回の被験者実験では高い正解率を得ることができたが、被験者実験の結果は、個々人の感性や音楽的知識の有

無によるばらつきほかに、実験に用いる抽出元の楽曲のタイプに依存すると考えられる。今回実験に用いた楽曲は、Hevner の研究において対象とされていた、クラシックのピアノ曲に限定している。それ以外の形式の楽曲においては、それぞれ新たな ML の構成および、メディアコンテンツ生成における統計情報の選択が必要となると考えられる。ジャンル別、時代別、作曲家別などの ML および iML を構成していくことは今後の課題である。

本稿では、構成された iML が全体として有効性を持つことを示したが、 iML に含まれるいくつかの不良設定問題のうち、今回被験者実験を行って検証したのは、楽曲特徴値から楽曲コンテンツへ変換する際の不良設定問題である。その他の変換において発生する不良設定問題とその解決法が、全体にどのような影響を及ぼしているかについて検証していくことは今後の課題とする。

6. おわりに

本稿では、メディアコンテンツから印象語からなるメタデータを自動抽出する ML の逆演算である、印象メタデータを入力として、メディアコンテンツを自動生成する新たな枠組みである、メディアコンテンツを対象とした統計的一般化逆作用素 “Stochastic Generalized Inverse Media-lexicon Transformation Operator” (“ iML ”) を示した。また、 iML の具体的な実現例として、楽曲メディアコンテンツを対象として、印象メタデータから楽曲メディアコンテンツを生成する方式について示した。また、上記の自動生成方式を実装し、実験を行い、本方式の有効性を示した。

楽曲を対象とした統計的一般化逆作用素 iML の構成によって、印象を表す言葉からの楽曲メディアコンテンツの生成が可能となった。今後、他のメディアを対象とした iML を構築することによって、楽曲だけでなく、画像、動画などの任意の種類のメディアコンテンツを対象としたメディアコンテンツ自動生成方式が構成できると考えられる。

本方式を応用することで、たとえば Web 上の楽曲コンテンツおよび SNS などのデータから、楽曲の特徴と印象の相関が求められるのであれば、楽曲構造要素と印象との関係を表す “ ML ” を Web 上のデータから作り出せる可能性がある。そうなれば、本稿で提案する統計的一般化逆作用素 iML の構成を行うことにより、データドリブンな計算機による自動作曲方式の実現の可能性がある。

今後の課題として、Web 上のデータを利用した ML 、 iML の構成方式、個人化方式、および、ジャンル別、時代別、作曲家別などに対応した、様々な音楽メディアに対する新たな ML および iML の構成方式があげられる。

参考文献

- [1] Kiyoki, Y., Kitagawa, T. and Hayama, T.: A meta-database system for semantic image search by a mathematical model of meaning, *ACM Sigmod Record*, Vol.23, No.4, pp.34-41 (1994).
- [2] Kitagawa, T. and Kiyoki, Y.: The mathematical model of meaning and its application to multidatabase systems, *Proc. 3rd IEEE International Workshop on Research Issues on Data Engineering: Interoperability in Multidatabase Systems*, pp.130-135 (1993).
- [3] 吉野太智, 高木秀幸, 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, Vol.1998-DBS-116, No.2, pp.109-116 (1998).
- [4] 北川高嗣, 中西崇文, 清木 康: 楽曲メディアデータを対象としたメタデータ自動抽出方式の実現とその意味的楽曲検索への適用, 電子情報通信学会論文誌 D, Vol.85, No.6, pp.512-526 (2002).
- [5] 北川高嗣, 中西崇文, 清木 康: 静止画像メディアデータを対象としたメタデータ自動抽出方式の実現とその意味的画像検索への適用, 情報処理学会論文誌, データベース, Vol.43, No.12, pp.38-51 (2002).
- [6] 本間秀典, 中西崇文, 北川高嗣: 任意の言葉を対象とした音韻印象変換作用素の構成とその感性検索への適用, 情報処理学会論文誌, Vol.51, No.5, pp.1294-1309 (2010).
- [7] Hevner, K.: Expression in music: A discussion of experimental studies and theories, *Psychological Review*, Vol.42, pp.186-204 (1935).
- [8] Hevner, K.: Experimental studies of the elements of expression in music, *American J. Psychology*, Vol.48, pp.246-268 (1936).
- [9] Hevner, K.: The affective value of pitch and tempo in music, *American J. Psychology*, Vol.49, pp.621-630 (1937).
- [10] 北原鉄朗, 深山 覚: 自動作曲・自動編曲の現状と課題, 電子情報通信学会学会誌, Vol.98, No.6, pp.475-479 (2015).
- [11] Hiller, L. and Isaacson, L.: Musical composition with a high-speed digital computer, *Journal of Audio Engineering Society*, Vol.6, No.3, pp.154-160 (1958).
- [12] Ebiccioglu, K.: Computer counterpoint, *Proc. ICMC*, pp.534-543 (1980).
- [13] Biles, J.A.: Genjam: A genetic algorithm for generating jazz solos, *Proc. ICMC*, pp.131-137 (1994).
- [14] Ando, D., Dahlstedt, P., Nordaxhl, M.G. and Iba, H.: Computer aided composition by means of interactive GP, *Proc. ICMC*, pp.254-257 (2006)
- [15] Stephen, S. and Levitt, D.: *Machine Models of Music*, MIT Press (1992).
- [16] 白井 亨, 谷口忠大: 階層 pitman-yor 言語モデルを用いたメロディー生成手法の提案, 知能と情報, Vol.25, No.6, pp.901-913 (2011).
- [17] 深山 覚, 中妻 啓, 酒向慎司, 西本卓也, 小野順貴, 嵯峨山茂樹: 音楽要素の分解再構成に基づく日本語歌詞からの旋律自動作曲, 情報処理学会論文誌, Vol.54, No.5, pp.1709-1720 (2013).
- [18] Tsuchiya, Y. and Kitahara, T.: Melodic outline extraction method for non-note-level melody editing, *Proc. Sound and Music Computing*, pp.762-767 (2013).
- [19] 北原鉄朗, 深山 覚, 片寄晴弘, 嵯峨山茂樹: Orpheus-BB: Human-in-the-loop 型の自動作曲システム, インタラクション, pp.57-64 (2011).
- [20] 梅本克夫: 音楽心理学, 誠信書房 (1966).



岡田 龍太郎 (正会員)

2007年筑波大学第三学群情報学類卒業。2009年筑波大学大学院システム情報工学研究科博士前期課程修了。2014年筑波大学大学院システム情報工学研究科博士後期課単位取得退学。現在、筑波大学大学院システム情報工学研究科研究生として在籍中。マルチメディア情報システムの研究に興味を持つ。



中西 崇文 (正会員)

国際大学グローバルコミュニケーションセンター (GLOCOM) 准教授・主任研究員。2006年筑波大学大学院システム情報工学研究科修了。博士(工学)。情報通信研究機構 (NICT) 研究員を経て現在に至る。マルチメディアシステム, データ分析工学, ビッグデータ分析, 統合データベースに関する研究に従事。電子情報通信学会, 日本知財学会, 日本データベース学会各会員。



本間 秀典 (正会員)

2009年筑波大学大学院システム情報工学研究科修了。博士(工学)。現在、東京医科歯科大学特任助教。システムズバイオロジーによる生物学的データ解析に関する研究に従事。異種メディア間検索に関する研究に興味を持つ。電子情報通信学会会員。



北川 高嗣

筑波大学大学院情報工学域教授。1978年名古屋大学工学部卒業。1983年名古屋大学大学院工学研究科博士課程修了。スタンフォード大学計算機科学科客員研究員, 愛媛大学理学部数学科講師, 筑波大学電子・情報工学系助教授を経て, 現在に至る。数値解析, 逆問題, マルチメディア情報システムの研究に従事。日本応用数理学会会員。