歌詞と音楽が与える印象の分析 *

☆河村 翔太　植村 あい子　北原 鉄朗 (日大)

1　はじめに

近年，コンピュータの発達により，創作活動がコンピュータ上で行われる機会が多くなってきている．音楽における作曲活動においても，同様のことが言える．DAW やボーカロイドの出現などにより，より一層コンピュータに頼った作曲活動が盛んになっている．それに伴って，コンピュータ自身が作曲活動を行う研究 [1,3,9] についても盛んにおこなわれている．

また，歌唱曲における作曲活動において，歌詞の内容・意味というのは重要である．とりわけ日本人の趣味嗜好において，歌詞の意味を重視した音楽づくりは重要視されている [2]．しかしながら現段階における自動作曲システムにおいては韻律やモーラ数から音楽を創り上げる [1] といったものや，リズムや小節数を限定することによって音楽を創り上げる [3] 手法などがあげられるが，これらは聴取者が歌詞の内容・意味から受ける印象と，音楽から受ける印象との差異について取り扱っていない．

本研究では，各種の特徴ベクトルの比較によって分析する．歌詞の印象と音楽の印象の特徴ベクトルを生成し，印象について印象語を用いたアンケートにより楽曲の歌詞と音楽それぞれに対して正解ラベルを付与し，印象語ベクトルを生成する．それらの特徴をニューラルネットワークや重回帰分析によって比較し，関係を分析する．

2　分析手法

本研究では，n 次元の印象語ベクトルを考え，歌詞・楽曲の双方から抽出した特徴ベクトルをこの印象語ベクトルに変換することで，歌詞・楽曲の特徴と印象の関係を分析する．つまり，歌詞の特徴ベクトルを $\boldsymbol{x}$，楽曲の特徴ベクトルを $\boldsymbol{y}$，印象語ベクトルを $\boldsymbol{z}$ とし，2 つの写像 $f : \boldsymbol{x} \to \boldsymbol{z}, g : \boldsymbol{y} \to \boldsymbol{z'}$ を考える．$\boldsymbol{x}$ と $\boldsymbol{y}$ を同じ楽曲から抽出した場合，$\boldsymbol{z}$ と $\boldsymbol{z'}$ が近いほど，歌詞と楽曲との間で印象の違いが小さいといえる．

2.1　特徴ベクトル・印象語ベクトルの設計

2.1.1　歌詞からの特徴抽出

歌詞の特徴抽出には Bag-of-words，TF-IDF，word2vec を用いた．どれも日本語を形態素解析したうえで，特徴をベクトル化するものである．TF-IDF，Bag-of-Words は歌詞本文のみでベクトルを出力することができる．一方で，word2vec は単語の意味を指定された辞書の語彙全体から算出する手法を取っているため，辞書が必要である．歌詞に特化した辞書を生成するのは困難であったため，今回は最もメジャーな wikipedia から生成した辞書を用いることにした．

word2vec は単語ごとに特徴抽出するものである．本研究では具体的な意味を含まない助詞・助動詞を避け，名詞・動詞・形容詞を使用した．また，そのうえで全単語の特徴の平均ベクトルを歌詞の特徴ベクトルとした．

本研究の場合，Bag-of-words の次元数は 3690，TF-IDF の次元数は 3734，word2vec の次元数は 200 であった．

2.1.2　楽曲からの特徴抽出

音楽の特徴抽出には jSymbolic[10] を使用した．これは MIDI ファイルの特徴ベクトルを出力してくれるシステムで，その特徴の種類は，Basic Pitch Histogram，Rhythmic Variability などがあげられるが，200 以上に及ぶので割愛する．

本研究の場合，jSymbolic の次元数は 1023 であった．

2.1.3　印象語の選定

印象語については，Hevner[5] の感情環 Fig. 1 を用いた．ジャンルごとの 8 つそれぞれに対応する日本語を，翻訳により最も意味が重複し，かつ Hevner の研究 [5] を引用している日本語の先行研究 [6-8] で最も出現している単語を選定基準として，Table 1 に示す 8 つの印象語を選定した．

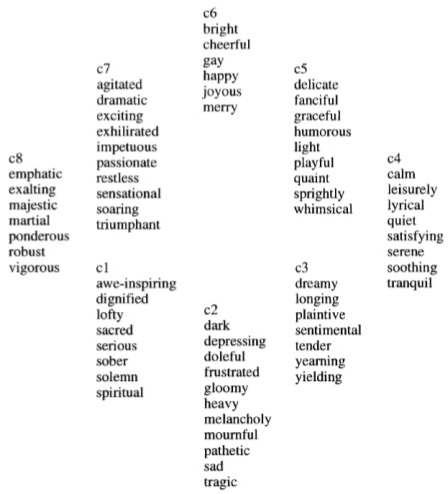

Fig. 1　Hevner の感情環

2.2　歌詞・楽曲データの収集

対象とする楽曲は，ヤマハミュージックデータショップより購入した MIDI ファイル 86 曲分の J-POP で

---

* “An analysis of the impression given by lyrics and music” by KAWAMURA, Shôta, Uemura, Aiko, KITAHARA, Tetsurô (Nihon University)

Table 1　選定した印象語

感情環	印象語
c1	厳格な
c2	憂鬱な
c3	憧れる
c4	穏やかな
c5	風変わりな
c6	陽気な
c7	興奮させる
c8	力強い

ある．第 1 節でも述べたが，本研究では日本人の趣味嗜好にあった歌唱曲を，歌詞と音楽の二側面から捉え関連付けを行うという点で，対象とする歌唱曲は邦楽のみとした．

これらには歌詞データが付随していなかったため，Web ブラウザ上で楽曲名から検索し，歌詞タイムや Uta-Net，J-Lyric.net などのサイトから収集した．

2.3　印象データの収集

印象語ベクトル $\boldsymbol{z}$，$\boldsymbol{z'}$ の正解データを得るため，大学生 35 人（重複あり）を対象に，歌詞を読んだときおよび音楽を聴いたときの印象を答えてもらう実験を行った．

2.3.1　歌詞からの印象の収集

各被験者には，86 曲分からランダムに選択された 20 曲分の歌詞（A4 用紙に印刷されたもの）を読み，楽曲ごとにその歌詞全体が Table 1 の 8 つの印象語それぞれにどの程度当てはまるかを 5 段階で答えてもらった．予備実験により，1 曲 1 分 30 秒で読み終えることができ，回答は 30 秒程度で可能であることが判明していたので，目安としてその旨を伝えていたが具体的な制限時間は設けなかった．

これにより取得した印象の各曲の平均を印象語ベクトル $\boldsymbol{z}$ とした．

2.3.2　音楽からの印象の収集

各被験者には，86 曲分からランダムに選択された 20 曲分の音楽を聴いてもらい，楽曲ごとに音楽全体が Table 1 の 8 つの印象語それぞれにどの程度当てはまるかを 5 段階で答えてもらった．歌詞の意味に印象が左右されないよう，ボーカルパートを楽器音に置き換えた MIDI データを，YAMAHA-MU500 により再生した．スピーカーには Bose SoundLink Color Bluetooth speaker を用いた．また，聴取場所の暗騒音は 20〜40dB 程度であった．

多くの歌唱曲は 1 番と 2 番に分かれているが，これらは繰り返すようなメロディであることを考慮し，聴かせるのは 1 番のみとした．予備実験により，1 曲 1 分〜2 分程度で聴き終え，その後 30 秒で回答できるというのが判明していたので，目安としてその旨を伝えていた．回答時間は明確に 30 秒という制限時間を設けた．

これにより取得した印象の各曲の平均を印象語ベクトル $\boldsymbol{z'}$ とした．

2.4　PCA による次元圧縮

Bag-of-words や TF-IDF はその性質上要素数が多いかつゼロが多く出現するため，PCA による主成分分析（次元圧縮）を行う．PCA による特徴損失というデメリットを累積寄与率によって確認しながらパラメータを調整するが，これら 2 つのシステムについては，データの最適化によるメリットの方が強く顕れると考えられる．

また，本研究において PCA を利用した場合の累積寄与率は，Bag-of-words，TF-IDF，word2vec，jSymbolic どれにおいても 6 割〜9 割を示した．

2.5　特徴ベクトルから印象語ベクトルへの変換

特徴ベクトル $\boldsymbol{x}$，$\boldsymbol{y}$ から印象語ベクトル $\boldsymbol{z}$，$\boldsymbol{z'}$ を得る手法として，ニューラルネットワークおよび重回帰分析を用いる．次の 2 つの手法を用いる．

手法 1：$\boldsymbol{x}$ または $\boldsymbol{y}$ を入力ノードとし，$\boldsymbol{z}$ または $\boldsymbol{z'}$ を出力ノードとする多層パーセプトロン（ハイパーパラメータを Table 2 のように設定する）

手法 2：$\boldsymbol{z}$ または $\boldsymbol{z'}$ の各要素を目的変数とし，$\boldsymbol{x}$ または $\boldsymbol{y}$ を説明変数とした重回帰分析（印象語ベクトルの各要素は独立に扱われる）

Table 2　Table 3 で設定した値

パラメータ名	値
入力層のデータサイズ	各種ベクトル依存
隠れ層のノード数	20
訓練データ始点	0
訓練データ終点	50
テストデータ始点	50
テストデータ終点	86
学習回数	4000
バッチサイズ	8
学習率	0.02
PCA 使用時の次元数	20

3　分析結果

第 2 節で示した歌詞や音楽に関する特徴ベクトルを事前に準備したうえで，歌詞と音楽の印象に関する分析を行った．本研究では歌詞特徴ベクトル $\boldsymbol{x}$ と歌詞印象語ベクトル $\boldsymbol{z}$，音楽特徴ベクトル $\boldsymbol{y}$ と音楽印象語ベクトル $\boldsymbol{z'}$ をニューラルネットワークや重回帰分析で関連付ける．

そして，ニューラルネットワークや重回帰分析における推論によって出力された値と各印象語ベクトル $\boldsymbol{z}$ 及び $\boldsymbol{z'}$ の平均二乗誤差平方根や平均相関係数により計算する，また，印象語どうしの $\boldsymbol{z}$ と $\boldsymbol{z'}$ の平均二乗

誤差平方根や平均相関係数を計算することで，音楽と歌詞の印象の相違についても考察する．

3.1 手法 1：ニューラルネットワーク

関連付けの手法としてニューラルネットワークを用いた．印象語ごとの平均二乗誤差平方根を求めた結果として Table 3 が，平均相関係数を求めた結果として Table 4 が得られた．Table 3 について，歌詞の 3 種を比べたときに，最も誤差が小さいのは Bag-of-words であった．また，最も誤差が大きい印象語には 4 種でばらつきがあったが，最も誤差が小さいものは “厳格な” で一致していた．Table 4 について，負の相関はないものの，非常に低い結果となった．特に jSymbolic における “憂鬱な” は 0.045 とほぼゼロである．

3.2 手法 2：重回帰分析

もう一つの関連付けの手法として重回帰分析を用いた．印象語ごとの平均二乗誤差平方根を求めた結果として Table 3 が，平均相関係数を求めた結果として Table 4 が得られた．Table 3 について，歌詞の 3 種どれを見ても突出して誤差が小さかったり，大きかったりするものはなかった．各印象語間の誤差において、大小関係の傾向に共通性を見ることはできたが，誤差が最大の印象語，最小の印象語が一致するのは TF-IDF と jSymbolic における最大の誤差が “憂鬱な” であるという 1 つのみであった．また，Table 4 について，すべての値で比較的高い 0.6〜0.9 の値を示している．

3.3 印象語ベクトル

正解データ間である歌詞印象語ベクトル $\boldsymbol{z}$ と音楽印象語ベクトル $\boldsymbol{z'}$，および推論データ間である重回帰分析の word2vec で推論した印象語ベクトル，重回帰分析の jSymbolic で推論した印象語ベクトルについて比較を行った．比較には平均二乗誤差平方根を用いた．これにより Table 5 が得られた．正解データ間に比べ，推論データ間の誤差はすべての印象語でほぼ半分程度になった．

4 考察

Table 3 の平均二乗誤差平方根において，ニューラルネットワークと重回帰分析を比較してみると，ニューラルネットワーク全体の平均が 0.848 に対して，重回帰分析の全体の平均が 0.491 と，重回帰分析のほうが誤差が小さくなっている．同様に Table 4 の相関係数についても，ニューラルネットワークでは 0.045 とほぼゼロの値も出ているのに対し，重回帰分析では最大で 0.902 と極めて高い相関係数を示すこともあるほど，全体の相関係数が高かった．これは学習データの量が関係していると考えられる．本研究で使用した楽曲は 86 曲であり，それを訓練データ 50 個とテストデータ 36 個に分けたので，学習に用いるデータ量が不足していたと考えられる．重回帰分析とニューラルネットワークの違いは，線形か非線形解析であるかの違いともいえる．ニューラルネットワークは学習データの量が多ければ多いほどより高い精度で判別することができるが，一方で学習データの量が少ない場合は精度が低くなったり，過学習が起きてしまったりする．それに引き換え重回帰分析は学習データの量の多寡の影響は比較的少ないと言える．つまりは，学習データの量が少なかったためにニューラルネットワークではうまく分類できなかったと考えられる．よって，今回のような学習データの量が極めて少ない場合は Table 3 のような重回帰分析に軍配が上がる結果になるのが必然であると考えられる．

また，Table 5 において，実験により収集した歌詞の印象語ベクトル $\boldsymbol{z}$ と音楽の印象語ベクトル $\boldsymbol{z'}$ の正解データ間にはどれも平均二乗誤差平方根において 1.000 前後の誤差があった．つまり，歌唱曲における歌詞と音楽における印象の差異は，皆無ではなく，ある程度の差異があることが正解である可能性も考えられる．それを踏まえて正解データ間と推論データ間を比較してみると，推論データ間のほうが誤差が小さくなっていることがわかる．これは学習過程において，歌詞と音楽とが保有しているべき差異が推論では損失してしまっていることを表していると考えられる．歌詞と音楽の分析を行っていくうえで，必ずしも歌詞と音楽の印象とで差異がなくなることが正解ではないとするならば，Table 5 のように正解データ間と推論データ間を比較することこそが，より歌唱曲らしい楽曲を導くのに重要な指標となってくることが考えられる．

5 おわりに

本研究では，歌唱曲の歌詞と音楽があたえる印象について分析した．印象語は Hevner の感情環 [5] から選定し，印象の正解ラベル $\boldsymbol{z}$，$\boldsymbol{z'}$ はその印象語を用いた聴取実験によって付与した．歌詞の特徴ベクトル $\boldsymbol{x}$ は Bag-of-words，TF-IDF，word2vec によって抽出し，音楽の特徴ベクトル $\boldsymbol{y}$ は jSymbolic によって抽出した．

歌詞と音楽との間に印象の差は少なからず存在することが確認できた．これにより，歌詞と音楽との印象で差異がなくなることが，必ずしも楽曲としての正解ではないということが示された．

現段階の手法では重回帰分析において，相関係数が高く誤差が少ない有用な結果が得られた．しかしながらニューラルネットワークではその精度は低く，今後の課題としてあげられる．分析をより深く正確なものにするためにも，他の文章特徴ベクトルを取得できる SCDV や Doc2vec などの検討や，学習データの量の増加，ニューラルネットワークの中間層（隠れ層）を増やすなど，精度を改善する手法の検討が必要である．また，特徴ベクトル間を関連付けるほかの手法として，印象語の強弱を離散化してクラスと考え，パターン認識を行う手法があげられる．

これらの手法も検討し，既存の手法と組み合わせて

Table 3　推論と z，z' の平均二乗誤差平方根

	ニューラルネットワーク				重回帰分析（PCA 有）			
	歌詞			音楽	歌詞			音楽
印象語	BoW[1]	TF-IDF	word2vec	jSymbolic	BoW[1]	TF-IDF	word2vec	jSymbolic
厳格な	0.580	0.639	0.720	0.708	0.448	0.343	0.403	0.352
憂鬱な	0.815	0.886	1.011	0.791	0.647	0.659	0.582	0.694
憧れる	0.817	0.863	1.076	0.743	0.288	0.469	0.414	0.523
穏やかな	0.786	0.826	1.044	0.765	0.419	0.551	0.499	0.532
風変わりな	0.805	0.829	1.032	0.791	0.637	0.549	0.649	0.633
陽気な	0.824	0.837	1.052	0.804	0.382	0.442	0.398	0.531
興奮させる	0.803	0.820	1.066	0.854	0.387	0.393	0.356	0.403
力強い	0.801	0.812	1.067	0.859	0.681	0.541	0.607	0.348

[1] BoW は「Bag of Words」の略称である

Table 4　推論と z，z' の平均相関係数

	ニューラルネットワーク				重回帰分析（PCA 有）			
	歌詞			音楽	歌詞			音楽
印象語	BoW[1]	TF-IDF	word2vec	jSymbolic	BoW[1]	TF-IDF	word2vec	jSymbolic
厳格な	0.319	0.234	0.205	0.124	0.726	0.850	0.786	0.841
憂鬱な	0.329	0.291	0.224	0.045	0.708	0.695	0.772	0.653
憧れる	0.242	0.146	0.194	0.437	0.903	0.714	0.786	0.625
穏やかな	0.251	0.168	0.218	0.533	0.819	0.656	0.730	0.685
風変わりな	0.270	0.255	0.237	0.498	0.685	0.778	0.670	0.689
陽気な	0.286	0.240	0.229	0.443	0.868	0.819	0.856	0.724
興奮させる	0.262	0.207	0.252	0.408	0.795	0.787	0.830	0.775
力強い	0.261	0.222	0.231	0.397	0.532	0.740	0.655	0.902

[1] BoW は「Bag of Words」の略称である

Table 5　z，z' の平均二乗誤差平方根

印象語	正解データ間	推論データ間
厳格な	0.763	0.417
憂鬱な	1.076	0.573
憧れる	0.933	0.404
穏やかな	1.115	0.509
風変わりな	0.955	0.544
陽気な	1.013	0.464
興奮させる	0.912	0.453
力強い	0.900	0.647

精度を上昇させていきながら，歌詞と音楽とでどれくらい印象の差異があることが望ましいのか調べることも考慮すべきである．また，音楽のどのような特徴が歌詞の特徴とかかわりを持っているのか jSymbolic の特徴を変更することにより明らかにしていけば，より深く歌詞と音楽の印象について分析することができると考えられる．

謝辞　本研究は，JSPS 科研費 16K16180，16H01744，16KT0136，17H00749 から支援を受けた．

また，事前調査に協力してくれた 33 名の方々に感謝の意を示す．

参考文献

[1] 深山 覚他，“Orpheus : 歌詞の韻律に基づいた自動作曲システム”，情報処理学会研究報告音楽情報科学（MUS），Vol.2008(78(2008-MUS-076)，pp.179-184，2008.

[2] 佐藤 生実，“踊り場における「恥じらい」のコミュニケーション”，コミュニケーション科学，Vol.27pp.51-72，2007.

[3] 芳村亮他，“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”，DEWS2007 論文集，電子情報通信学会，2007.

[4] Tomas Mikolov *et al.*, “Efficient estimation of word representations in vector space”, International Conference on Learning Representations, 2013.

[5] Hevner, K. “Experimental studies of the elements of expression in music”, American Journal of Psychology, Vol.48, pp.246-268, 1936.

[6] 大串健吾，“音楽と感情”，イオメカニズム学会誌，Vol.30(1)，pp.3-7，2006.

[7] 安田晶子他，“音楽聴取による感動の心理学的研究”，日心第７０回大会，2006.

[8] 浅野雅子他，“音楽心理学の動向について：音楽近く，音楽と感情，音楽療法を中心に”，芸術工学研究，Vol.12．pp.83-95，2010.

[9] 菅野沙也他，“入力文書の印象と感情に基づく楽曲提供の一手法”，情報処理学会研究報告音楽情報科学（MUS），Vol.2014-MUS-105(4)，pp.1-6，2014.

[10] `http://jmir.sourceforge.net/manuals/jSymbolic_manual/home.html`