

音楽と印象に関する一分析

井上 湧哉^{1,a)} 植村 あい子^{1,b)} 北原 鉄朗^{1,c)}

概要: 本稿では、「陽気な」「力強い」などの楽曲を聴いたときの印象を印象語で表したものと その楽曲の MIDI データから抽出した特徴量の関係を調査する。これまで様々な自動作曲システムが提案・開発されてきたが、生成される楽曲の特徴に関する要件をユーザがシステムに伝えるのは簡単ではなかった。その問題を解決する方法として、上で述べた印象語を作曲のパラメータを増やすことが考えられるが、それを実現するには、これらの印象語が表す印象を楽曲 (MIDI データ) のどんな特徴が作り出しているのかを明らかにする必要がある。我々は、河村らが策定した「厳格な」「憂鬱な」「憧れる」「穏やかな」「風変わりな」「陽気な」「興奮させる」「力強い」の 8 つの印象語に対して、各々が表す印象の高低を MIDI データから自動識別する決定木を構築する。J-POP 楽曲 86 曲分の MIDI データの各々に、上述の印象語に対する印象の正解ラベルを聴取実験をもとに付与し、様々な特徴量セットを用いて J4.8 で識別を行った。その結果、各印象語が表す印象と MIDI データから抽出した特徴量の結果が一部明らかになった。

1. はじめに

近年、コンピュータの発達によって、コンピュータ上での創作活動が盛んになっている。創作活動の例として、作曲と編曲 (以下、作編曲) が取り上げられ、ボーカロイドや DTM などコンピュータを利用した作編曲が数多く存在する。また、音楽ゲームを作っている会社が一般人に向けて楽曲を募集^{*1}したり、ゲーム会社に楽曲を提供する活動を行なっているアーティストも存在しており、楽曲の創作活動がより一層盛んになってきている。

作編曲は、未経験者にとっては難しいことだと予想される。このような初心者に対して、歌詞の韻律に基づく制約条件下で歌唱曲を自動生成 [1] したり、遺伝的アルゴリズムを用いてコード進行から自動作曲 [2] をするなど、作編曲を自動的に行なうシステムを提案している研究は多く存在する。研究以外にも既存の製品で楽曲の自動生成を行なうものも存在する。例えば、「Band in a Box」はコード進行を入力し、生成する伴奏スタイルを選ぶだけでそのスタイルに合わせた楽曲が生成される。

しかしながら、これらを用いてユーザが所望する楽曲を生成するのは容易ではない。その理由の一つに、音楽の専門家ではないユーザが、自分の欲しい楽曲の特徴をシステムに伝えることが難しい点があげられる。たとえば、Orpheus[1]

は歌詞の他にリズムやコード進行を選択することができるが、音楽知識のないユーザがこれらを適切に選択するのは容易ではない。Band in a Box は 6000 種類以上あるなかから楽曲のスタイルを選択することができるが、スタイルはジャンルを元にした名称が付与されているだけであり、ジャンルの知識がなければ結局 1 つ 1 つ試す他ない。

本研究では、「陽気な」「力強い」などの印象語に着目し、これらを自動作曲時のパラメータに追加することで、この問題の解決を目指す。たとえば、陽気な楽曲が欲しければ、「陽気な」にチェックを入れたり、スライダなどで「陽気な」に高い値を与えると、その指示に沿った旋律などが生成されるシステムを構築する。これを実現するには、各印象語の高低 (当てはまる度合い) と楽曲特徴量の関係を明らかにする必要がある。たとえば、「陽気な」が高い楽曲は長調であることが容易に想像できるが、必ずしも聴取実験によって裏付けられているわけではない。本稿では、上述の目的に向け、楽曲に対する印象語を策定し、その印象語が表す印象と楽曲 (MIDI データ) から抽出可能な特徴との関係を分析する。

2. 実験方法

2.1 実験の基本的な方針

実験は、パターン認識的アプローチによって行う。MIDI データベースの各楽曲に対して、特定の印象語 (たとえば「陽気な」) が表す印象の高低をラベル付けする。このラベル付けは、後述の聴取実験に基づいて行う。この印象ラベル付き MIDI データベースに対して、ごく少数の特徴量の

¹ 日本大学文理学部

a) inoue@kthrlab.jp

b) uemura@chs.nihon-u.ac.jp

c) kitahara@chs.nihon-u.ac.jp

*1 <https://chunithm-special.sega.jp/1st/>

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

図 1 Hevner の印象語群 [5]

みを用いて印象ラベルの分類を行う。もしも高精度で分類ができれば、そのときに用いた特徴量はその印象と密接な関係にあることになる。様々な特徴量に対してこれを繰り返すことで、各印象語に関連する特徴量を探索する。

我々はこの分析を自動作曲に適用することを想定しているが、多くの自動作曲システムでは、主旋律の生成と伴奏部の生成は別の処理になっている。そこで、主旋律のみの MIDI データと伴奏部のみの MIDI データを用いることで、各印象語が表す印象が主旋律の特徴によるものなのか伴奏の特徴によるもののかも調査する。

2.2 印象語の選定

河村ら [3] が用いた 8 つの印象語 (表 1) を扱う。これは、芳村らの研究 [4] を参考に Hevner [5] の印象語群 (図 1) を用いている。この印象語群を引用した日本語の先行研究 [6], [7], [8] では、8 つのグループそれぞれに対応する 8 つの日本語の印象語を用いているが、1 グループで複数の印象語が定義されていたり、各研究で日本語訳が異なったりするため、アルクが提供する英辞郎 on the WEB [9] によってすべて翻訳し、グループごとに最も出現している日本語訳を選定した。

2.3 楽曲データ

河村ら [3] が用いた J-POP 楽曲 86 曲である。ヤマハミュージックデータショップから購入した MIDI ファイルに対して、次の 3 種の MIDI ファイルを生成して用いた。

- (1) 主旋律 (トラック 1) 以外を消去したもの
- (2) 主旋律を消去したもの

表 1 印象語

| | 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 |

表 2 印象語 2 値化のしきい値

| 印象語 | しきい値 | 個数比 |
|-------|------|-------|
| 厳格な | 1.8 | 44:42 |
| 憂鬱な | 2.0 | 45:41 |
| 憧れる | 2.8 | 41:45 |
| 穏やかな | 3.5 | 42:44 |
| 風変わりな | 2.5 | 42:44 |
| 陽気な | 3.0 | 44:42 |
| 興奮させる | 2.3 | 44:42 |
| 力強い | 3.0 | 42:44 |

(3) (1) に対してテンポを指定する命令を除去したもの
 各々を「主旋律」「伴奏」「主旋律 (テンポなし)」と呼ぶ。

2.4 印象ラベルの付与

河村ら [3] が行った主観評価に基づいて印象ラベルを付与する。河村らは、大学生 19 名を対象に、上述の 86 曲からランダムに選ばれた 20 曲に対して、表 1 の各印象語が表す印象の強さを 1~5 の 5 段階で答えてもらった。各楽曲に対して 1 曲当たり 3~6 人程度の回答があるため、その平均値をその楽曲の各印象語に対する印象値とする。その後、表 2 をしきい値として 2 値化を行った。しきい値は、High に属する楽曲数と Low に属する楽曲数ができるだけ同じになるように設定した。

2.5 楽曲特徴量

MIDI データからの楽曲特徴量の抽出には jSymbolic2 [10] を使用した。これは、MIDI データから様々な特徴量を抽出できるソフトウェアである。抽出可能な特徴量は、

- P:** Overall pitch statistics
- M:** Melodic intervals
- C:** Chords and vertical intervals
- R:** Rhythm
- RT:** Rhythm and Tempo
- I:** Instrumentation
- T:** Musical texture
- D:** Dynamics

の 8 つのカテゴリーに分かれており、特徴量の種類は 200 種類以上、特徴量の総次元数は 1494 である。特徴量の具体的な内容は、次の URL に書かれている。

http://jmir.sourceforge.net/manuals/jSymbolic_manual/featureexplanations_files/featureexplanations.html

2.6 分類器

分類器には、データマイニングツール「weka」に実装されている決定木 J4.8 を用いる。また、分類実験には 10-fold cross validation を用いる。

(a) 主旋律
Combined_Strength_of_Two_Strongest_Rhythmic_Pulses <= 0.3437:
| LOW (30.0/6.0)
Combined_Strength_of_Two_Strongest_Rhythmic_Pulses > 0.3437
| Note_Density <= 2.111: HIGH (49.0/11.0)
| Note_Density > 2.111: LOW (7.0/1.0)
(Correct : 0.7558)

(b) 伴奏
Variability_of_Number_of_Simultaneous_Pitch_Classes <= 1.089
| Beat_Histogram_81 <= 0.001849: HIGH (58.0/16.0)
| Beat_Histogram_81 > 0.001849: LOW (12.0/2.0)
Variability_of_Number_of_Simultaneous_Pitch_Classes > 1.089:
LOW (16.0/1.0)
(Correct : 0.7558)

(c) 主旋律 (テンポなし)
なし

図 2 印象語「憂鬱な」に対する決定木生成結果

2.7 実験の手順

「主旋律」「伴奏」「主旋律 (テンポなし)」の 3 種の MIDI データそれぞれに対して、次の手順で実験を行う。

- (1) すべての特徴量から 1 つのみ選んで分類を行い、識別率が 65% 以上のもののみを残す。
- (2) (1) で残った特徴量を対象に、特徴量を 2 つにして同様の分類を行う。
- (3) (2) で最も識別率が高かった特徴量の組み合わせを調べ、考察の対象とする。最も高い識別率が 65% 未満の印象語は、考察対象から除外する。

3. 実験結果・考察

生成された決定木およびその考察を以下に記す。なお、「厳格な」は 3 種の MIDI データのいずれに対しても識別率が 65% 以上の決定木が得られなかったため、省略する。

3.1 「憂鬱な」

「憂鬱な」に対する決定木を図 2 に示す。主旋律に対しては、Combined Strength of Two Strongest Rhythmic Pulses (RT-26), Note Density (RT-5) という 2 つの特徴量を用いた時が最も識別率が高かった。RT-26 はビートヒストグラムの上位 2 つの強度の和である。ビートヒストグラムは、隣り合うオンセット間の時間に基づいて計算され、各ビンの強度は総和が 1 になるように正規化されている。RT-26 の値が高いことは、ビートヒストグラムの特定のビンの値が特に高いことを表し、同じ IOI (inter-onset interval) が何度も出てくる、すなわちリズムが比較的単純であることを示す。つまり、リズムの複雑さが主旋律の憂鬱さに関連するといえる。RT-5 は、1 秒あたりの音符数である。この値が 2.111 以下のときに「憂鬱な」が高いという結果が得られた。1 秒あたりの音符数が多いと憂鬱さを感じられないのは、一般的な考え方に合致すると言える。

(a) 主旋律
なし

(b) 伴奏
Pitch_Class_Histogram_9 <= 0.09524: LOW (35.0/7.0)
Pitch_Class_Histogram_9 > 0.09524
| Pitch_Class_Kurtosis <= 4.868: HIGH (41.0/9.0)
| Pitch_Class_Kurtosis > 4.868
| | Pitch_Class_Histogram_9 <= 0.1157
| | | Pitch_Class_Kurtosis <= 4.912: LOW (2.0)
| | | Pitch_Class_Kurtosis > 4.912: HIGH (2.0)
| | Pitch_Class_Histogram_9 > 0.1157: LOW (6.0)
(Correct : 0.7442)

Pitch_Class_Skewness <= 0.1432
| Pitch_Class_Kurtosis <= 4.847: HIGH (43.0/10.0)
| Pitch_Class_Kurtosis > 4.847: LOW (11.0/2.0)
Pitch_Class_Skewness > 0.1432: LOW (32.0/6.0)
(Correct : 0.7442)

(c) 主旋律 (テンポなし)
なし

図 3 印象語「憧れる」に対する決定木生成結果

伴奏に関しては、Variability of Number of Simultaneous Pitch Classes (C-5) と Beat Histogram 81 (RT-16) を用いたときの識別率が一番高かった。C-5 は、同時発音ピッチクラス数が楽曲中で変化するとき高い値を取る。つまり、メリハリのある編曲かどうかを表す 1 つの指標になっていると考えられる。編曲にメリハリがあれば憂鬱さが低くなるのは、一般的な考え方に合致する結果と言える。

3.2 「憧れる」

「憧れる」に対する決定木を図 3 に示す。主旋律は、識別率が 65% 以上になる決定木は得られなかった。伴奏に対しては 2 つの決定木が得られ、Pitch Class Histogram 9 (P-2), Pitch Class Skewness (P-28) がそれぞれ選ばれた。他、Pitch Class Kurtosis (P-31) が両方に共通して選ばれた。これらの特徴量はいずれもピッチクラスのヒストグラムから計算されるが、このヒストグラムは最も出現頻度の高い音名が 0 番目に来るように移動されている。伴奏で最も出現頻度の高い音名は、その楽曲の調の主音であると推測できるので、Pitch Class Histogram 9 は、主音から長 6 度上の音名の出現頻度の割合ということになる。この特徴量が低いとき、その曲は短調である可能性が高い。つまり、楽曲が短調であることが、「憧れる」を低くする傾向があるといえる。P-31 は、ピッチクラスヒストグラムの尖度を表す。最頻の音名は主音である可能性が高いため、P-31 が高いとき、主音の出現割合が高いことを意味すると推測される。つまり、コード進行が比較的シンプルであることが予想され、そのことが「憧れる」を低くした可能性がある。

(a) 主旋律

```
Chord_Duration <= 0.961
| Polyrythms_-_Tempo_Standardized <= 0.25
| | Polyrythms_-_Tempo_Standardized <= 0.1538: HIGH (2.0)
| | Polyrythms_-_Tempo_Standardized > 0.1538: LOW (16.0/4.0)
| Polyrythms_-_Tempo_Standardized > 0.25: HIGH (36.0/4.0)
Chord_Duration > 0.961: LOW (32.0/4.0)
(Correct : 0.8023)
```

```
Beat_Histogram_Tempo_Standardized_80 <= 0.2062
| Polyrythms_-_Tempo_Standardized <= 0.25
| | Polyrythms_-_Tempo_Standardized <= 0.1538: HIGH (2.0)
| | Polyrythms_-_Tempo_Standardized > 0.1538: LOW (20.0/4.0)
| Polyrythms_-_Tempo_Standardized > 0.25: HIGH (39.0/6.0)
Beat_Histogram_Tempo_Standardized_80 > 0.2062: LOW (25.0/3.0)
(Correct : 0.8023)
```

```
Mean_Tempo <= 119.9
| Polyrythms_-_Tempo_Standardized <= 0.25
| | Polyrythms_-_Tempo_Standardized <= 0.1538: HIGH (2.0)
| | Polyrythms_-_Tempo_Standardized > 0.1538
| | | Mean_Tempo <= 79.16: HIGH (4.0/1.0)
| | | Mean_Tempo > 79.16: LOW (10.0/1.0)
| Polyrythms_-_Tempo_Standardized > 0.25: HIGH (39.0/6.0)
Mean_Tempo > 119.9: LOW (31.0/3.0)
(Correct : 0.8023)
```

(b) 伴奏

```
Mean_Tempo <= 119.3
| Melodic_Interval_Histogram_20 <= 0: LOW (7.0/1.0)
| Melodic_Interval_Histogram_20 > 0: HIGH (48.0/10.0)
Mean_Tempo > 119.3: LOW (31.0/3.0)
(Correct : 0.8023)
```

```
Mean_Tempo <= 119.3
| Seventh_Chords <= 0.1395
| | Mean_Tempo <= 74.97: HIGH (2.0)
| | Mean_Tempo > 74.97: LOW (11.0/2.0)
| Seventh_Chords > 0.1395: HIGH (42.0/7.0)
Mean_Tempo > 119.3: LOW (31.0/3.0)
(Correct : 0.8023)
```

(c) 主旋律 (テンポなし)

```
Beat_Histogram_80 <= 0.2062
| Polyrythms_-_Tempo_Standardized <= 0.25
| | Polyrythms_-_Tempo_Standardized <= 0.1538: HIGH (2.0)
| | Polyrythms_-_Tempo_Standardized > 0.1538: LOW (20.0/4.0)
| Polyrythms_-_Tempo_Standardized > 0.25: HIGH (39.0/6.0)
Beat_Histogram_80 > 0.2062: LOW (25.0/3.0)
(Correct : 0.8023)
```

図 4 印象語「穏やかな」に対する決定木生成結果

3.3 「穏やかな」

「穏やかな」に対する決定木を図 4 に示す。主旋律に対しては、3 つの決定木が得られ、Chord Duration (C-27), Beat Histogram Tempo Standardized 80 (R-53), Mean Tempo (RT-2) がそれぞれ選ばれた他、Polyrythms-Tempo Standardized (R-66) が共通して選ばれた。C-27 はコードチェンジ間の時間長の平均であるが、ここでは主旋律のみの入力のため、実際には主旋律のピッチクラスが変化する時間

長、つまり、同じピッチクラスの音符の連続を 1 つの長い音符とみなしたときの主旋律の音長の平均である (四分音符を 1.0 とする)。この値が高いときに「穏やかな」が低いのは、テンポに理由があるからと考えられる。テンポが速い曲はその分 1 音 1 音の音価が長めになる傾向がある。テンポが速い曲は穏やかな印象を与えにくいと考えられるため、主旋律の音長と穏やかさに負の相関が見られたと考えられる。テンポとの関連は、RT-2 が選ばれたことから分かる。

穏やかさは 3 連符や付点音符など複雑なリズムの有無とも関連があると言える。R-53 は、楽曲のテンポを 120 にしたときにテンポ 80 つまり 3/4 秒が周期の音符 (付点 8 分音符に相当) がどの程度含まれているかを表す。R-66 は、基本的な音価 (4 分音符) の整数倍や整数分の 1 倍ではない音がどれだけ含まれているかを表す。3 連符や 5 連符などが多いときにこの値が高くなる。このような複雑なリズムを含む主旋律は、穏やかな印象を与えにくいと予想される。

伴奏に対しては、2 つの決定木が得られ、Melodic Interval Histogram 20 (M-1), Seventh Chords (C-32) がそれぞれ選ばれた他、RT-2 が両方に共通して選ばれた。RT-2 は主旋律とほぼ同じ閾値になっている。Melodic Interval Histogram は、MIDI チャンネル (10 以外) ごとに発音時刻が隣り合う音符の音高差を求めて作ったヒストグラムを全チャンネル分合算したものである。M-1 の 20 は、半音 20 個分の音高の遷移が出現する割合を表している。この特徴量が 0 よりも高いときに「穏やかな」が高いと判断されたのは、ピアノ伴奏の曲においてオブリガートを多用する編曲になっていたからと考えられる。ピアノが主たる伴奏を奏でるバラード調の楽曲では、低めの音域で伴奏をしつつ高めの音域でオブリガートを入れることが頻繁にあり、そういった場合に 1 オクターブ半を超える音高の遷移が起こりうる。そのため、この特徴量が 0 より高いときにバラード調、つまり穏やかな曲であると判断されたと考えられる。これに該当する楽曲としては、『I LOVE YOU』『366 日』などがあげられる。

C-32 は、同時に鳴った音が長 7 度もしくは短 7 度である頻度を表している。平均テンポが 74.95 以上 119.3 以下であり、この特徴量が 0.1395 以下のとき「穏やかな」は低いと判定された。この条件を満たす楽曲の多くは、『北の旅人』や『天城越え』など短調の演歌曲であった。これらの楽曲は暗い雰囲気強調したものが多いため、「穏やか」の対義語である) 激しさはないものの、穏やかな楽曲とも受け取られなかったと思われる。演歌ではセブンスコードはあまり使われないため、この特徴量が選ばれたと考えられる。

主旋律 (テンポなし) では、テンポありの主旋律の 2 つ目の決定木と同じものが得られた。R-53 と RT-16 はテンポを 120 に正規化しているかそうでないかの違いである。MIDI データではテンポ情報がない状態で特徴抽出してい

```
(a) 主旋律
Major_or_Minor <= 0
| Range <= 15: HIGH (13.0/2.0)
| Range > 15: LOW (57.0/15.0)
Major_or_Minor > 0: HIGH (16.0)
(Correct : 0.7674)

(b) 伴奏
Major_or_Minor <= 0
| Melodic_Interval_Histogram_18 <= 0.000273: HIGH (19.0/5.0)
| Melodic_Interval_Histogram_18 > 0.000273: LOW (51.0/12.0)
Major_or_Minor > 0: HIGH (16.0)
(Correct : 0.7907)

(c) 主旋律 (テンポなし)
Major_or_Minor <= 0
| Range <= 15: HIGH (13.0/2.0)
| Range > 15: LOW (57.0/15.0)
Major_or_Minor > 0: HIGH (16.0)
(Correct : 0.7674)
```

図 5 印象語「風変わりな」に対する決定木生成結果

るので、BPM がデフォルトで 120 に設定されるため、R-53 と RT-16 は同じ意味を持つ。

3.4 「風変わりな」

「風変わりな」に対する決定木を図 5 に示す。主旋律に対しては、テンポありとテンポなしの両方において、Major or Minor (P-33) と Range (P-8) という 2 つの特徴を用いたときの識別率が最も高かった。P-33 は 0 のときに長調、1 のときに短調を表す。一般的に短調な曲は暗いイメージがあるとされており、その暗さが風変わりという印象に影響を及ぼした可能性がある。P-8 は出現する最高音と最低音のノートナンバーの差を表しており、この値が 12 以下のときに「風変わりな」は高くなった。通常の J-POP では、A メロから B メロ、サビと進むにつれ主旋律の音域が高くなるが多いため、主旋律全体の音域は 18 程度になるのが普通である。P-8 が 12 以下というのは、サビになってもあまり音域が高くないことを意味し、その点が典型的な J-POP と異なる印象を与えた可能性がある。

伴奏では、主旋律と同様に P-33 が選ばれた一方、P-8 の代わりに Melodic Interval Histogram 18 (M-1) が選ばれた。旋律で音高が半音 18 個分遷移することはまずないが、同一チャンネル内で和音を奏でている場合はありうる。D2-A2-D3-F#3 の直後に C2-G2-C3-E3 が来る場合である (『さくらんぼ』の 36~37 小節目)。このコード進行やヴォイシングは珍しくないため、これが一度も出現しないほうが風変わりな印象を与えやすかったと考えられる。

3.5 「陽気な」

「陽気な」に対する決定木を図 6 に示す。主旋律に対しては、2 つの決定木が得られ、Average Time Between Attacks (RT-7)、Average Time Between Attacks for Each Voice

```
(a) 主旋律
Most_Common_Rhythmic_Value <= 0.25: LOW (28.0/6.0)
Most_Common_Rhythmic_Value > 0.25
| Average_Time_Between_Attacks <= 0.5544: HIGH (41.0/6.0)
| Average_Time_Between_Attacks > 0.5544: LOW (17.0/3.0)
(Correct : 0.8140)

Most_Common_Rhythmic_Value <= 0.25: LOW (28.0/6.0)
Most_Common_Rhythmic_Value > 0.25
| Average_Time_Between_Attacks_for_Each_Voice <= 0.5544:
| HIGH (41.0/6.0)
| Average_Time_Between_Attacks_for_Each_Voice > 0.5544:
| LOW (17.0/3.0)
(Correct : 0.8140)

(b) 伴奏
Note_Density <= 28.45: LOW (30.0/3.0)
Note_Density > 28.45
| Duration_in_Seconds <= 293.5: HIGH (44.0/5.0)
| Duration_in_Seconds > 293.5: LOW (12.0/2.0)
(Correct : 0.8488)

(c) 主旋律 (テンポなし)
Prevalence_of_Most_Common_Melodic_Interval <= 0.3866
| Rhythmic_Value_Histogram_4 <= 0.2052: LOW (45.0/10.0)
| Rhythmic_Value_Histogram_4 > 0.2052: HIGH (8.0/1.0)
Prevalence_of_Most_Common_Melodic_Interval > 0.3866
| Prevalence_of_Most_Common_Melodic_Interval <= 0.4774:
| HIGH (29.0/3.0)
| Prevalence_of_Most_Common_Melodic_Interval > 0.4774:
| LOW (4.0/1.0)
(Correct : 0.7558)

Melodic_Embellishments <= 0.01504: HIGH (50.0/14.0)
Melodic_Embellishments > 0.01504
| Rhythmic_Value_Histogram_4 <= 0.2007: LOW (32.0/4.0)
| Rhythmic_Value_Histogram_4 > 0.2007: HIGH (4.0)
(Correct : 0.7558)
```

図 6 印象語「陽気な」に対する決定木生成結果

(RT-8) がそれぞれ選ばれた他、Most Common Rhythmic Value (R-26) が共通して選ばれた。RT-7 と RT-8 は MIDI チャンネル別に計算するかどうかの違いなので、1 チャンネルしかない主旋律の場合等価である。この特徴量は、隣り合う音符の発音時刻の差 (秒単位) を示しており、この値が高いほど主旋律に短い音符がたくさんあることを示す。Most Common Rhythmic Value (R-26) は、最も多く使われている音価を表す。この値が 0.25 であれば、16 分音符が一番多く使われていることを示す。RT-7/RT-8 が高くて R-26 が低いほど「陽気な」が高くなるのは矛盾するように見えるが、これはテンポに依存する特徴量だからである。テンポが速いと 16 分音符を歌唱・演奏するのが困難になるので、相対的に 16 分音符よりも 8 分音符が増え、R-26 は高い値になる。一方、RT-7/RT-8 は秒単位での発音時刻の差であり、テンポが速くなればその分値は小さくなる。実際、平均テンポである RT-2 が 89 以上のとき 44 曲中 37 曲は「陽

(a) 主旋律
Note_Density <= 1.616: LOW (31.0/4.0)
Note_Density > 1.616
| Most_Common_Rhythmic_Value <= 0.25: LOW (15.0/3.0)
| Most_Common_Rhythmic_Value > 0.25: HIGH (40.0/3.0)
(Correct : 0.8837)

(b) 伴奏
Mean_Tempo <= 111.6
| Basic_Pitch_Histogram_85 <= 0.005344: LOW (45.0/7.0)
| Basic_Pitch_Histogram_85 > 0.005344: HIGH (9.0/2.0)
Mean_Tempo > 111.6: HIGH (32.0/2.0)
(Correct : 0.8140)

(c) 主旋律 (テンポなし)
Most_Common_Rhythmic_Value <= 0.25: LOW (28.0/5.0)
Most_Common_Rhythmic_Value > 0.25: HIGH (58.0/19.0)
(Correct : 0.7209)

図 7 印象語「興奮させる」に対する決定木生成結果

気な」は高い曲である。また、RT-2 と R-26 は相関係数が 0.5427 であることから、テンポが速い曲には陽気な曲が多く、R-26 が高くなるとの結果が得られた。

伴奏に対しては、RT-5 と Duration in Seconds (RT-4) を使ったときの識別率が一番高かった。RT-4 は楽曲の長さ(秒単位)を表す。RT-4 が 293 以下のときに「陽気な」が高いとの結果であったが、5 分以内の楽曲はテンポが速いものに多いと予想される。また、RT-5 の値が高いほど 1 秒間で鳴る音符数は多い。つまり、伴奏における音の厚さを表しており、この値が高いほど「陽気な」が高いというのは、一般的な価値観に適合した結果といえる。

テンポなしの主旋律からは 2 つの決定木が得られ、Prevalence of Most Common Melodic Interval (M-6)、Embellishments (M-21) がそれぞれ選ばれた他、Rhythmic Value Histogram 4 (R-13) が共通して選ばれた。M-6 は、主旋律の音高遷移のヒストグラムにおける出現割合の最大値を表す。この値が 0.3866 より高いということは、主旋律における音高遷移の 38% 以上がある音程(多くの場合長 2 度)であることを意味する。陽気な曲は比較的単純な旋律が多いと予想され、この値が高い値になると推測される。これに当てはまる曲として『SUM』がある(M-6 は 0.4698)。M-21 は、前後にその音符の 3 倍以上の音価が存在する音符の割合である。比較的短い音符と長い音符の組み合わせで旋律ができているとき、この特徴量は高い値を取る。つまり、旋律のリズムの複雑さを表すと解釈でき、この値が低いとき、つまり比較的単純な旋律ほど「陽気な」が高くなるのは一定の合理性はあると思われる。R-13 は、4 分音符および 4 分休符の出現頻度を表しており、これも旋律のリズムの複雑さを表すと解釈できる。

3.6 「興奮させる」

「興奮させる」に対する決定木を図 7 に示す。主旋律に

(a) 主旋律
Mean_Tempo <= 122.4: LOW (58.0/18.0)
Mean_Tempo > 122.4: HIGH (28.0/4.0)
(Correct : 0.7209)

(b) 伴奏
Average_Rest_Fraction_Across_Voices <= 0.6516
| Major_or_Minor <= 0: LOW (47.0/9.0)
| Major_or_Minor > 0: HIGH (8.0/1.0)
Average_Rest_Fraction_Across_Voices > 0.6516: HIGH (31.0/5.0)
(Correct : 0.8140)

Average_Rest_Fraction_Across_Voices <= 0.6516
| Unpitched_Instruments_Present_17 <= 0: LOW (39.0/5.0)
| Unpitched_Instruments_Present_17 > 0: HIGH (16.0/5.0)
Average_Rest_Fraction_Across_Voices > 0.6516: HIGH (31.0/5.0)
(Correct : 0.8140)

(c) 主旋律 (テンポなし)
Most_Common_Rhythmic_Value <= 0.25: LOW (28.0/5.0)
Most_Common_Rhythmic Value > 0.25: HIGH (58.0/19.0)
(Correct : 0.7209)

図 8 印象語「力強い」に対する決定木生成結果

対して、RT-5 と R-26 の 2 つの特徴量を使ったときの識別率が一番高かった。RT-5 は一秒あたりの音符数、R-26 は最頻の音価を表す。RT-5 が高く R-26 も高いときに「興奮させる」が高くなるのは、3.5 節で述べた「陽気な」と同様にテンポの速い曲である可能性が高い。

伴奏は、RT-2 と Basic Pitch Histogram 85 (P-1) のときの識別率が最も高かった。P-1 の 85 は、ノートナンバー 85 (C \sharp) の出現頻度の割合を表す。嬰ハ短調、変口短調、ホ長調など音階にド \sharp が含む調であるときに「興奮させる」が高くなる傾向がある。実際に『if...』では嬰ハ短調になっており、P-1 の値が 0.07353 と楽曲の中で一番高い値を示している。逆に『粉雪』では P-1 の値が 0 を示しており、ホ短調であった。

テンポのない主旋律では、RT-26 を使ったときの識別率が一番高く、RT-2 や RT-5 は選ばれなかった。RT-2 は平均テンポであり、テンポを除去したデータを使ったため、選ばれないのは当然である。RT-5 は 1 秒あたりの音符数である。この特徴量はテンポによって異なるため値が変化する。テンポ情報のないデータではテンポ 120 を仮定して計算するため、本来の特徴を表現できなかったと考えられる。一方、RT-26 は最頻の音価であるが、3.5 節で述べたように、テンポが遅ければ 16 分音符が増え、テンポが速ければ 16 分音符が減って 8 分音符が増える傾向があり、テンポに依存する特徴を間接的に表しているといえる。そのため、テンポのない主旋律ではこの特徴量が選ばれたと考えられる。

3.7 「力強い」

「力強い」に対する決定木を図 8 に示す。主旋律では、RT-2 を用いたときの識別率が最も高かった。つまり、テン

ポが 122.4 より速いときに「力強い」が高くなるとの結果が得られた。

伴奏では、2つの決定木が最も高い識別率を示しており、P-33 と Unpitched Instruments Present 17 (I-2) がそれぞれ選ばれた他、共通して Average Rest Fraction Across Voices (R-43) が選ばれた。I-22 の 17 は、チャイニーズシンバルが鳴っている (1) かそうでないか (0) を表す。チャイニーズシンバルは、楽曲の最後のサビ部分や アクセントを入れたところで使われたりすることが多い。チャイニーズシンバルを鳴らすと迫力が出ると思われるので、この迫力が「力強い」に影響したと考えられる。R-43 は、チャンネルごとの音符のない時間長の平均である。この値が大きいほど、休符が多くなり、静かな楽曲になると予想される。

テンポなしの主旋律では、R-26 を使ったときの識別率が最も高かった。3.5 節でも述べたように、テンポが速いほど 16 分音符などの短い音符を歌唱・演奏しにくくなるため、R-26 (最頻の音価) は値が高くなる傾向がある。実際、RT-2 と R-26 の相関係数は 0.5427 であった。そのため、RT-2 (平均テンポ) の代わりにこの特徴量が選ばれたと考えられる。

4. おわりに

本研究では、楽曲を聴いたときの印象を印象語で表したものと、その楽曲から抽出した特徴量との関係を調査した。印象語として「厳格な」「憂鬱な」「憧れる」「穏やかな」「風変わりな」「陽気な」「興奮させる」「力強い」を採用し、河村ら [3] のデータを用いて MIDI 特徴量から各印象語の印象の高低を識別する決定木を作成したところ、次に挙げる事柄が明らかになった。

- 「穏やかな」「興奮させる」「力強い」では平均テンポ (Mean Tempo) が重要である。
- 最頻の音価 (Most Common Rhythmic Value) は平均テンポと相関があり、「陽気な」「興奮させる」「力強い」において重要である。
- ビートヒストグラムや音価のヒストグラムから抽出した特徴量にはリズムの複雑さを数値化したものと解釈できるものがあり、「憂鬱な」「穏やかな」「陽気な」と関連がある。

印象語を用いた自動作曲を実現するには、これらの結果を自動作曲エンジンに組み込む方法を検討する必要がある。たとえば、遺伝的アルゴリズムを用いた自動作曲を想定するのであれば、上で述べた決定木の内容を適合度関数に含める方法が考えられる。しかし、対象とする自動作曲エンジンが 1 小節ごとに旋律などを生成するタイプのものであれば、上の決定木も 1 小節ずつの旋律データに対するものでなければならない。今後は、必要に応じて決定木を再生成した上で、決定木を自動作曲エンジンに導入する方法の検討を進める。

謝辞 本研究は、JSPS 科研費 16H01744, 17H00749,

19K12288 から支援を受けた。

参考文献

- [1] 深山 覚, 中妻 啓, 米林 裕一郎, 酒向 慎司, 西本 卓也, 小野 順貴, 嵯峨山 茂樹: 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. 76, pp. 179–184 (2008).
- [2] 山田 拓志, 椎塚 久雄: 遺伝的アルゴリズムを用いた自動作曲について, 情報処理学会研究報告音楽情報科学 (MUS), Vol. 27, pp. 7–14 (1998).
- [3] 河村 翔太, 植村 あい子, 北原 鉄朗: 歌詞と音楽が与える印象の分析, 日本音響学会 2019 年春季研究発表会 講演論文集 (2019).
- [4] 芳村 亮, 中西 崇文, 北川 高嗣: 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 (2007).
- [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, *バイオメカニズム学会誌*, Vol. 30, No. 1, pp. 3–7 (2006).
- [7] 浅野 雅子, 古根川 円, 中島 祥好, 蓮尾 絵美: 音楽心理学の動向について: 音楽知覚, 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).
- [8] 安田 晶子: 音楽聴取による感動の心理学的研究: 身体反応の主観的計測に基づいて, *認知心理学研究*, Vol. 6, No. 1, pp. 11–19 (2008).
- [9] 英辞郎 on the WEB (online) 入手先 (<https://eow.alc.co.jp/>) (2018.11.05).
- [10] McKay C., Cumming J. E. and Fujinaga I.: jSymbolic 2.2: Extracting features from symbolic music for use in musicological and MIR research, *Proc. International Society for Music Information Retrieval Conference*, pp. 348–353 (2018).