

任意の言葉の印象に合った楽曲生成についての検討

稲垣 有紗† 酒向 慎司† 北村 正†

名古屋工業大学大学院工学研究科†

1 はじめに

ユーザのイメージに合った楽曲を生成する場合、イメージを表現する言葉を介することで、任意の言葉と音楽を対応づける方法が考えられる。しかし、言葉は莫大な数が存在し、またそれらの意味も多様であるため、音楽と対応させることは困難である。そこで、言葉の表す意味をできる限り損なわない、より小さな空間として印象を設定することを考える。イメージを表現するための空間を介させることで、任意の語彙と印象の対応と、印象と音楽の対応を分離して考えることができる。

先行研究では、英語の印象に合った自動楽曲生成方式が提案され、その有効性が示されている [1]。この方式は、意味の数学モデル [2] と呼ばれる単語の関係を計測するモデルと、音楽心理学に基づく研究を用いることにより実現されている。ただし、言語が限定されているため、その他の言語を使用したい場合は語彙を英語に置き換える必要がある。他言語への拡張を試みる場合、英単語と対象とする言語の単語は必ずしも一対一で対応するわけではなく、単語間の翻訳だけではうまく意味を表現することができない。その問題に対しては、文脈を考慮することで解決できると考えた。そこで、本研究では日本語の印象に合った楽曲生成方式を提案し、聴取実験によりその有効性を検討する。

2 言葉と音楽の印象表現

言葉の印象に合った楽曲を生成するために、印象情報を付加した楽曲構成要素を算出する。その要素から生成される楽曲は、それぞれの印象に合ったものになると考えられる。本研究では、図 1 のように印象語群ベクトル MCV を介して楽曲構成要素 MFV を任意の言葉により生成する。

2.1 日本語による語彙の意味抽出

莫大な数存在する言葉を全て同じ条件で扱うため、本研究では辞書を用いる。辞書には各単語の意味が表記されている。その文章を用いることで、その単語と関連のある単語を抽出することができると考えられる。

まず、全ての見出し語が決められた語彙のみで説明されているという特徴をもつ英英辞書『Longman Dictionary of Contemporary English』を用いて、英語による単語の意味を調べる。そして、それらを日本語へ対応させていくことを考える。一般的に、異なる言語間で単語の意味が一対一に対応しないことがある。例えば、英語と日本語の場合「right」や「space」などが挙げられ、使われ方により複数の意味を持つ。これらの単語を正確に日本語へ変換するためには、文章全体から判断する必要がある。

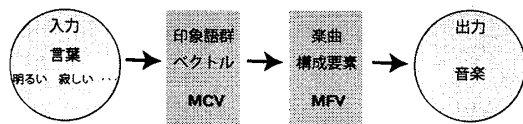


図 1: 全体の流れ

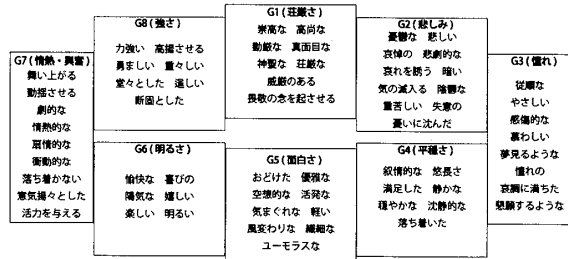


図 2: 印象語群

そこで、一対一で変換可能である単語についてはそのまま対応する日本語に変換し、一対一で変換不可能な単語については、前後の文脈に合わせて人が意味を検証し、適切な日本語に変換する。以上より、日本語による言葉の意味を取り出す。

2.2 意味空間の設定

語彙の関係を計測し、言葉の意味を扱うために意味空間 S [2] を設定する。この空間に各単語を射影することにより、単語同士の相関をみることができると考えられる。

まず、m 個の単語をそれぞれ n 個の特徴で表現した m × n 行列 M について相関行列を求め、固有値分解をする。

$$M^T M = Q \begin{pmatrix} \lambda_1 & & & \\ & \lambda_2 & & \\ & & \ddots & \\ & & & \lambda_n \end{pmatrix} Q^T \quad (1)$$

ただし、Q = (q1, q2, ..., qn) であり、相関行列の固有ベクトルを表している。この固有ベクトルは、相関行列の対称性により互いに直交している。次に、固有ベクトルにより張られる空間を次式で定義する。

$$S = \text{span}(q_1, q_2, \dots, q_n) \quad (2)$$

2.3 印象語群

本研究で扱う印象には、心理学者 Hevner により定義された印象語群 (図 2) を用いる [3]。これは、感情を表現するとされる 66 語の形容詞を、意味が類似する語が同じグループに所属されるように分類している。合計 8 個のグループで構成されており、隣接するグループは離れたグループより似た意味を持ち、向かい合うグループはほぼ反対の意味を持つように配置されている。

入力語 w に対し、空間 S における各印象語群との距離を相関の強さとして 8 次元の印象語群ベクトル MCV を求める。以上により、言葉の印象を抽出する。

2.4 印象に合った楽曲構成要素

人が音楽を聴いたときに抱く印象は明るい、寂しいなど音楽により異なる。音楽は、音高やテンポなど複数の音響的特徴により構成されており、その違いが音楽の印象に大きく影響を与えていると考えられる。

本研究では、音楽の印象に関係があると考えられる構造要素として文献 [3][4] で挙げられている調性、テンポ、音高、

リズム、和声、旋律という 6 種類の要素に着目する。その構成要素と印象語群との相関を表す行列 T を用いることで、次式より印象を考慮した楽曲構成要素 MFV が得られる。

$$MFV = T \cdot MCV \quad (3)$$

$$= (key, tempo, pitch, rhythm, harmony, melody)^t$$

3 楽曲の生成方式

式 (3) より得られる構成要素を基に、楽曲の生成に必要な音高の平均や和音の構成比など、より詳細な特徴量を文献 [1] を参考に算出する (表 1)。そして、得られる特徴量に従って、楽曲を生成する。

伴奏の生成は、特徴量の内 kn , $unac$, ac , tc , oc を用いる。まず調性 kn により使用する I 度から VII 度の和音を決定し、 $unac$ と ac の値に従った個数の四分音符と八分音符を配置する。和音の並びは一般的な和声進行 [5] と、 tc と oc の値による主要三和音 (I, IV, V) とそれ以外の和音の割合から決定する。

また、旋律の生成は kn , $mnap$, um , dm , lm を用いる。使用する音は調性 kn の音階のものとし、四分音符と八分音符を配置する。次に um , dm , lm の比率に従って上昇音、下降音、水平音の個数を算出し、各音符に割り当てる。ここで、旋律の最後は調の主音となるため、各音を最後から順に決定していく。ただし、先に生成された伴奏と調和させるため、旋律の各音は同じタイミングで演奏される和音に含まれる音階を使用している。

4 実験

本手法の有効性について検討する。まず、生成される楽曲の特徴量と単語の印象について検討する。そして、聴取実験により主観評価を行う。

4.1 生成される楽曲について

本手法を用いて楽曲を生成した。例として、「喜び」と「悲しい」から得られた楽曲を図 3 に示す。印象語群との相関を表す印象語群ベクトル (表 2) から、「喜び」については G6 (明るさ) と正の相関が強く、G2 (悲しみ) と負の相関が強いことがわかる。一方、「悲しい」については G2 (悲しみ) と正の相関が強く、G7 (興奮) と負の相関が強いことから、どちらも単語の印象を表現できていると考えられる。

4.2 主観評価実験

生成される楽曲について、実際に印象が表現されているかをみるため、聴取実験を行った。まず入力した単語と得られた楽曲を提示し、それらの印象がどの程度一致しているかを 5 段階で評価するという方法で主観評価実験を実施した。ただし、対象とする楽曲は 50 曲とし、被験者は 8 名である。

実験の結果、最も評価が高かった楽曲は「暗い」で、続いて「陽気な」「楽しい」「孤独な」となった。一方、評価が低かった楽曲は「危険な」「忌まわしい」「不安定な」などであった。図 4 から、評価の高い楽曲については印象を与え易いテンポや音高に大きく違いが表れているのに対し、評価の低い楽曲についてはそれらが平均的な値を示していることがわかる。

5 むすび

本研究では、日本語を対象とした印象に合致する楽曲生成を行った。聴取実験の結果、テンポや音高で印象が表現される楽曲については高い評価が得られ、手法の有効性が示された。しかし、印象をうまく表現できない語彙も存在する。それらに対応するため、特徴量から楽曲を生成する手法の改善が課題として挙げられる。

その他の課題として、意味空間の設定に必要な行列 M の作成に使用する辞書の検討が挙げられる。また、印象に合わせて音に強弱を付けたり、使用する楽器を多様化したりというように、より自然に印象を表現した楽曲の生成が考えられる。



(a) 「喜び」



(b) 「悲しい」

図 3: 生成された楽曲

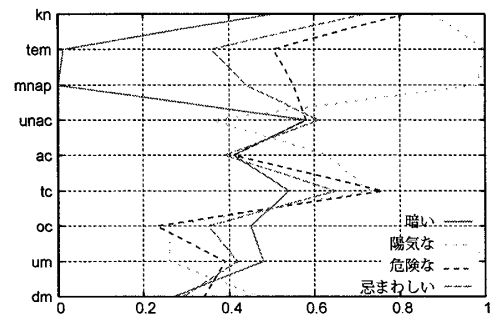


図 4: 特徴量の算出結果

表 1: 楽曲の特徴量

kn	調性
tem	テンポ
$mnap$	旋律の音高平均 (ノートナンバー)
$unac$	四分音符以上の長さである和音の総演奏時間 (ms)
ac	四分音符未満の長さである和音の総演奏時間
tc	主要三和音の総演奏時間
oc	主要三和音以外の総演奏時間
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

表 2: 印象語群ベクトル

グループ	入力語	喜び	悲しい
G1 (荘厳さ)		-0.96	0.65
G2 (悲しみ)		-1.00	1.00
G3 (憧れ)		-0.67	0.45
G4 (平穏さ)		-0.01	0.44
G5 (面白さ)		0.27	0.29
G6 (明るさ)		1.00	-0.10
G7 (興奮)		-0.81	-1.00
G8 (強さ)		-0.35	0.65

参考文献

- [1] 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論文集, 2006.
- [2] Y.kiyoki, T.kitagawa, H.Takanari: “A Metadatabase System for Semantic Image Search by a Mathematical Model of Meaning”, ACM SIGMOD Record, vol.23, pp.34-41, dec. 1994.
- [3] Hevner, K.: “Experimental studies of the elements of expression in music”, *American Journal of psychology*, Vol.48, pp.246-268, 1936.
- [4] Henkin, R. I.: “A Factorial Study of the Components of Music”, *Journal of Psychology: Interdisciplinary and Applied*, Vol.39, pp.161-181, 1955.
- [5] 石折 真礼生, 末吉 保雄, 丸田 昭三, 飯田 隆, 金光 威和雄, 飯沼 信義: “新装版 楽典 理論と実習”, 音楽之友社, 2008.