

入力文書の印象や感情に基づく楽曲提供の一手法

菅野沙也^{A1} 伊藤貴之^{A2} 高村大也^{A3}

概要：本研究では文書の印象や感情に基づいて楽曲生成を行う。前処理としてコードとリズム進行のデータを作成しておき、さらにユーザーごとの感性データを取得しておく。これは一人ひとりによって異なる音楽的感性を考慮し、ユーザーそれぞれに対応した楽曲を提供するためである。文書を入力するとまず形態素解析と感情極性を用いた文書解析をすることで印象値を取得し、次にこの印象値から場面ごとのコードとリズムの進行を決定する。これらを合成することで生成された楽曲がユーザーごとに提供されることとなる。入力文書の印象や感情を楽曲を用いて表現することでさらなる印象理解を促すことを目的とした。

Music synthesis based on impression and emotion of input documents

Saya Kanno^{A1} Takayuki Itoh^{A2}
Hiroya Takamura^{A3}

Abstract: This paper presents a technique to synthesize the music based on the impression and emotion of the input documents. The technique prepares variety of chord and rhythm progression datasets as MIDI files, and estimates their impressions and emotions as a preprocessing. Given an input document, the technique firstly analyzes the impression and emotion of the document applying a morphological analysis and semantic orientation calculation. Then, the technique specifies the chord and rhythm progressions those impressions and emotions are the closest to the input document. Finally, the technique provides the music by combining the chord and rhythm. We suppose this technique encourages to express impression and emotion of the input documents by music.

1. はじめに

我々は小説などの文書の印象を様々な形で表現する。特にレビューを書くなどの形で別の文章を用いたり、イラスト化するなどの視覚的手法を用いたりすることが多い。しかし、これらの手段には作品のネタバレをさせたり、登場人物の容姿や場所のイメージに関する先入観を与えすぎたりする可能性がある。そこで文書の印象を表現する別の手段として、昔から我々の生活や感情表現に欠かせない要素である「音楽」で文書の印象を表現することを提案する。音楽にはイメージ誘導効果・行動誘導効果・感情誘導効果があり、これは音楽への興味の有無に関わらず、すべての人に効果があると行動心理学の観点[1]からも語られている。さらに、文学と音楽という異なる種類の芸術をリンクさせることで、新たな感動や刺激が得られることも期待できる。このような効果は、オペラやバレエ音楽などで既に実現されている。しかし楽曲の制作には専門的知識や経験が必要である。特に、ある特定のイメージをもとに楽曲を制作するためには高度な知識と経験を要することが多く、誰でも気軽に実現できるとは言いがたいのが現状である。そこで本研究では、文書を入力するとその印象や感情に基づいた楽曲を提供する一手法を提案する。本手法のような

システムにより、文章作品の印象や感情の間接的な伝達が音楽によって促進できると期待できる。

2. 関連研究

文書から楽曲を生成する研究として、物語の論旨や文法の解析結果に基づいて楽曲を生成する手法[2]がある。しかしこの手法は、読者が受けた印象や感情に基づいたものではない。一方で、文章の印象や感情に基づいた楽曲生成手法として、文章の印象値からの機械的な音列生成によって楽曲を自動生成する手法[3]がある。しかしこの手法では、機械的な音列生成であるがゆえに、音楽的表現に限りがあり、また例えば自分の好みの音楽的素材を採用することが難しい。それに対して本研究では、あらかじめ用意したコード進行とリズム素材のマッシュアップによって楽曲を生成する、という点が関連研究と異なる。

Ricardo らの研究[4]では本研究と同様に文書の感情から楽曲を生成するといった試みがされている。しかしこちらの研究ではユーザーごとの感性を楽曲に適応するような処理はされていないため、本研究とは異なる点がある。

3. 提案手法

本手法は大きく分けて3つの処理段階で構成される。具体的には、

- ・事前調査：感性語と音楽的特徴量の選出
- ・学習：コードとリズムの印象の関係性算出
- ・対話処理：ユーザーの文書入力に対する楽曲生成

^{A1} お茶の水女子大学大学院
Ochanomizu University Graduate School
^{A2} お茶の水女子大学
Ochanomizu University
^{A3} 東京工業大学
Tokyo Institute of Technology

の3段階である。詳細について以下に論述する。

3.1 事前調査

まず、文書解析の基準となる感性語対と、コードやリズムの選出基準となる音楽的特徴量を選出する。図1に列挙するように感性語対と音楽的特徴量を候補にあげた。音楽的特徴量については、長谷川ら[5]が列挙した特徴量を参考にして選出し、さらに我々の主観から印象や感情の表現に向いていると思われる特徴量として、コードの種類、和音の不協和音度、ドラムの音色の割合を加えた。感性語については、池添ら[6]が列挙した感性語を参考にしながら我々の主観を交えて候補を列挙した。

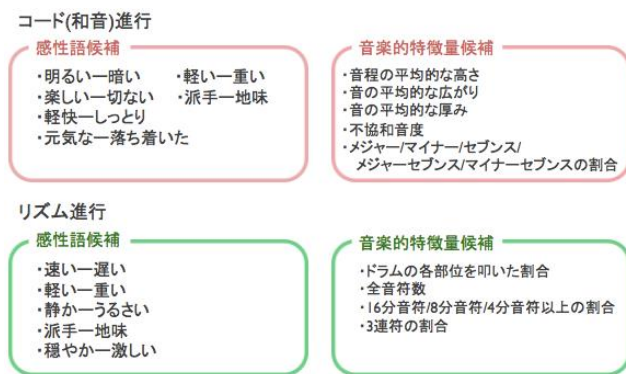


図1：感性語対と音楽的特徴量の候補

続いて、感性語対と音楽的特徴量の選出のためにユーザーアンケートを実施した。被験者にはコードとリズムの音源を聴いてもらい、それぞれ候補となる感性語対の度合いを5段階評価で回答してもらった。この回答をもとに、感性語対に関する音源Nの適合度と音源Nについての音楽的特徴量の相関性を相関係数により算出し、どの感性語対とも相関性が小さい音楽的特徴量、またはどの音楽的特徴量とも相関性が小さい感性語対を対象から除外した。以上の結果として、コード進行およびリズム進行の感性語対と音楽的特徴量を図2のように選出した。

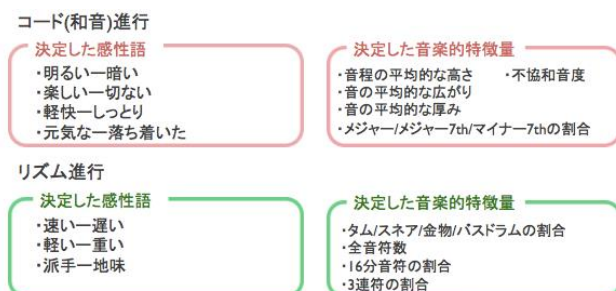


図2：決定した感性語対と音楽的特徴量

続いて、日本語の各内容語(名詞、動詞、形容詞、副詞)が、選出した各感性語対のいずれの感性に近いかを数値化する。具体的には感性語対の一方を+1に、もう一方を-1に対応させ、-1から+1の実数値を各内容語に対し算出する。この算出には感情極性抽出手法[7]を適用し、各感性語対を種単語として用いた。

3.2 学習

続いて、コードおよびリズムの音楽的特徴量とそれに対する各ユーザーの印象の関係を学習する。現時点での我々の実装では、各ユーザーにサンプルコードおよびサンプルリズムを聴いてもらい、3.1節で選出した感性語への適合度を回答してもらい、以後、この適合度を印象値と称する。また、印象値を推測するための式を音楽的特徴量の線形結合(図3参照)と仮定し、重回帰分析を用いてこの式の各係数を算出する。この式を用いてコード進行とリズム進行の印象値を算出し、あらかじめ格納する。以上の学習により、ユーザーごとの印象の違いを考慮した楽曲生成が可能になる。

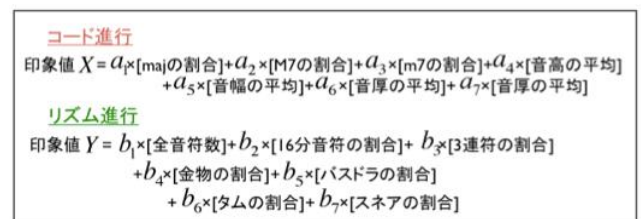


図3：コード進行およびリズム進行の印象推測のための式

3.3 対話処理

最後に、実際にユーザーが文書を入力すると楽曲を提供する対話処理について論じる。対話処理はさらに文書解析、コードとリズムの選出、楽曲生成の3段階に分けられる。

3.3.1 文書解析

まずユーザーが入力した文書に関して文書解析を適用する。現時点の我々の実装では、文書中でコード・リズムを切り替える点をユーザー自身によってタグ等で指定したテキストファイルを前提とし、これを入力することで場面ごとに文書の印象値を算出する。我々の実装では、オープンソース形態素解析エンジン MeCab[8]を用いて入力文書を形態素解析し、そのうち名詞、動詞、形容詞、副詞のみを抽出して、感性極性辞書を用いて数値化する。これの分散をとることによって、各感性語対に対する印象値を算出する。

3.3.2 コードとリズムの選出

次に楽曲の素材となるコードとリズムを選出する。ここでは3.3.1項で算出した印象値を多次元ベクトルとして扱い、これと3.2節で算出したコードおよびリズムの印象値を比較して、ユークリッド空間上で最も距離の近いコードおよびリズムを、入力文書の印象に沿った楽曲の素材とする。現時点での実験環境では、コードについては「明るい⇔暗い」「楽しい⇔切ない」「軽快⇔しっとり」「元気な⇔落ち着いた」の4次元を採用している。リズムについては「速い⇔遅い」「軽い⇔重い」「派手⇔地味」の3次元を採用している。

次章にて紹介する実験では、コード進行のMIDIファイル

ルにてすべての和音を2分音符に限定し、さらにプログラムナンバーによって指定される音色をピアノのみとした。また、コード進行・リズム進行ともに MIDI ファイルの長さを8小節とし、テンポをBPM120と設定した。

3.3.3 楽曲生成

選出したコードとリズムを合成することで楽曲を生成する。我々の実装ではコードとリズムの MIDI ファイルはそれぞれ単一のトラックで構成されることを前提とし、この各々のトラックを新しく生成する MIDI ファイルの1番目、2番目のトラックにコピーする。現時点での我々の実装ではメロディーを扱っていないが、今後はメロディーの選出とコピーを実装したい。

4. 実行結果

本手法によってユーザーAとユーザーBの感性をもとに童話「シンデレラ」を楽曲化した結果を図4に示す。ユーザーAとユーザーBはそれぞれ同じコード・リズム素材の評価をすることで感性データを取得した。また入力文書も全く同じものを用いた。現段階では各場面ではコード・リズム素材が1つずつ選ばれる。

	ユーザーA	ユーザーB
場面1	chord10.mid, rhythm2.mid	chord8.mid, rhythm12.mid
場面2	chord13.mid, rhythm2.mid	chord5.mid, rhythm15.mid
場面3	chord14.mid, rhythm2.mid	chord5.mid, rhythm12.mid
場面4	chord13.mid, rhythm2.mid	chord5.mid, rhythm12.mid

図4：童話「シンデレラ」を楽曲化した結果

ユーザーAとユーザーBでは異なる楽曲素材が選ばれており、学習段階の影響によりユーザーごとの印象の違いを考慮した楽曲が生成されていると考えられる。しかし両ユーザーともに異なる場面でありながらも全く同じ楽曲素材が選ばれていることがあり、文書の細かな印象や感情の違いを表現できるような改良が必要である。

5. 評価実験と考察

本手法について、評価実験を行った。手順について以下に示す。

1. 学習段階として各被験者にコード・リズム (以下サンプル楽曲とする)を10曲ずつ聴いてもらい、感性語対の度合いを6段階評価してもらう。以上によってユーザーごとの感性データをとる。
2. 対話処理によって出力した楽曲と元の文書の印象が一致するかどうか、各被験者に4段階評価してもらう。手順1で聞かせるサンプル楽曲はすべてのユーザーで同じものを使用した。また、元の文書と場面の分け方は出力された楽曲をユーザーが視聴してから提示することとした。入力文書には童話「シンデレラ」を用い、場面分けは以下の通りである：

1. シンデレラが継母や姉たちにいじめられる

2. 舞踏会へ行かずに留守番をしていると魔法使いが表れて魔法をかけてくれる
3. 舞踏会に参加して王子様とダンスする
4. 王子様が探しに来てプロポーズされる

上記の実験の回答者は計17人で、すべて20代の女性である。各シーンの長さは各 MIDI ファイルの長さである16秒にすべて統一されているため、本手法の対話処理によって生成されてユーザーに提示された楽曲はすべて64秒である。コード・リズムの素材はそれぞれ25個ずつ用意した。また、各場面の単語数はそれぞれ278, 208, 256, 203であった。

場面1から場面4に対する楽曲生成結果を被験者に評価してもらった結果を、それぞれ図5～8に示す。すべての場面において過半数以上の回答者が「場面と合っている」と回答しており、現時点でおおむね文書の印象を表現出来る楽曲が生成されたと考えられる結果となった。

さらに以下、回答者から得られたコメントから本手法の有効性と課題について議論する。

場面に合っていないと評価した回答者には、場面ごとの変化が少なかったとコメントする人が多く、逆に、場面に合っていると評価した回答者には場面ごとに楽曲の雰囲気はかなり変わっていたとコメントする人が多かった。このことから、場面ごとの楽曲の変化や楽曲の流れが印象表現に大きく影響していると考えられる。そこで今後の課題として、個々の場面の印象値を算出するだけでなく、場面と場面の対比も考慮した印象値算出手法を構築して検証を進めたい。

また、学習段階について、「評価の基準がないため難しい」「サンプル楽曲を聴いて行くうちに基準が変わってしまう」といったコメントが、ほぼすべての回答者から得られた。このことから、学習段階にはサンプル楽曲を評価するための基準が必要であると強く考えられる。しかし特定の基準を提示しても感性によるずれが生じてしまう恐れがあるため、2つのサンプル楽曲を聴き比べて評価するようなトーナメント式での評価が有効ではないかと検討している。

場面1の評価

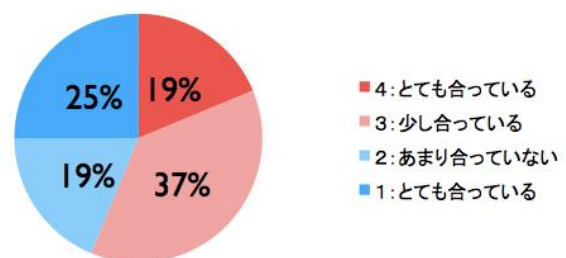


図5：場面1の評価

場面2の評価

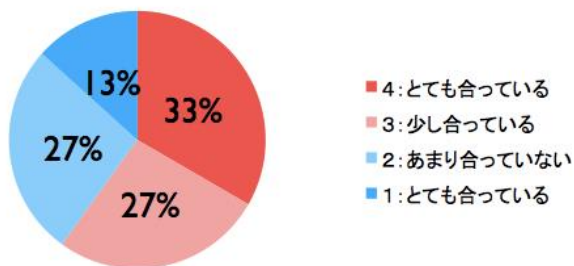


図6: 場面2の評価

場面3の評価

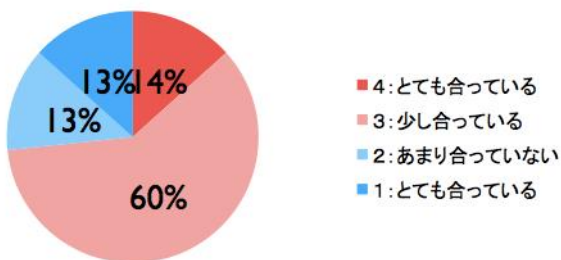


図7: 場面3の評価

場面4の評価

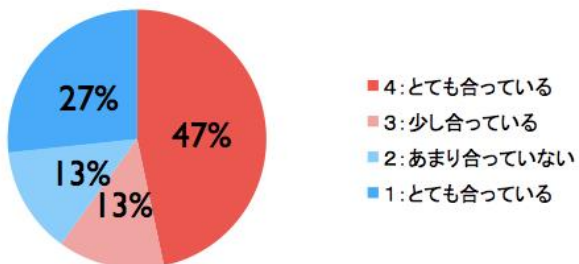


図8: 場面4の評価

6. まとめと今後の課題

本報告では、文書の印象値と音楽的特徴量の比較により、入力文書の印象や感情に基づいた楽曲を提供する事で文書の印象表現を促す手法を提案した。これに伴い、相関性の算出から感性語対と音楽的特徴量の関連について独自に調査を行った。また、ユーザーの感性に対応して一人一人異なる楽曲が生成されるようにした。

今後の課題は以下の通りである。

- ・ 感性語と音楽的特徴量の対応付けを見直す。
- ・ 場面の前後の関係性を考慮して素材（コード進行やリズム進行）を選択する。
- ・ 文書中の場面の区切りを自動認識する手法を実装する。
- ・ 文書の展開と楽曲の展開に関連を持たせる。

- ・ 楽曲にメロディーを付与する。

音楽的特徴量については、現時点での我々の実験に用いたコード進行やリズム進行の MIDI ファイルにおいて一定値とした要素（例えば和音のベロシティ）を排除して候補を列挙した経緯がある。そのため、現時点で採用している音楽的特徴量には印象や感情の表現において偏りがあると考えられ、見直しの必要があると考えられる。また5章での考察でも挙げたように、場面と楽曲素材の個々の印象の一致だけではなくストーリーの流れを意識した素材選択を可能にする必要もある。

謝辞

本研究を進めるにあたり、評価実験にご協力頂いた皆様に、謹んで感謝の意を表する。

参考文献

- 1) 齊藤寛, 心を動かす音の心理学, ヤマハミュージックメディア, 2011.
- 2) 遠藤順, 北館拓真, 小方孝, 物語生成システムにおける音楽の生成・表現機構, 日本認知科学会第29回大会, pp. 3-29, 2012.
- 3) 北原和也, 渡邊英徳, 安藤大地, ウェブアクティビティを反映させた楽曲の自動生成手法, 先端芸術音楽創作学会会報, Vol. 1, No. 1, pp. 8-11, 2012.
- 4) Ricardo Cruz, Antonio Brisson, Ana Paiva, and Eduardo Lopes, "I-Sounds - Emotion-Based Music Generation for Virtual Environments", ACHI 2007 Springer, LNCS 4738, pp. 7696770, 2007.
- 5) 長谷川隆, 西本卓也, 小野順貴, 嵯峨山茂樹, 音楽知識に基づく音高・音長の組合せ特徴量を用いた MIDI データからの作曲家判別. 情報処理学会研究報告.[音楽情報科学], 2009.13: pp. 47-52, 2009
- 6) 池添剛; 梶川嘉延; 野村康雄. 音楽感性空間を用いた感性語による音楽データベース検索システム. 情報処理学会論文誌, 2001, 42.12: 3201-3212.
- 7) 高村大也, 乾孝司, 奥村学, スピンモデルによる単語の感情極性抽出, 情報処理学会論文誌, Vol. 47, No. 2, pp. 627-637, 2006.
- 8) 京都大学情報学研究科 日本電信電話株式会社コミュニケーション科学基礎研究所 共同研究ユニットプロジェクト, MeCab, <http://mecab.googlecode.com/svn/trunk/mecab/doc/index.html>