

音楽認知モデルによる感性情報抽出

武藤 誠, 半田 伊吹, 坂井 修一, 田中 英彦

東京大学大学院工学系研究科
〒113-8656 東京都文京区本郷7-3-1

{muto, handa, sakai, tanaka}@mtl.t.u-tokyo.ac.jp

あらまし:

楽曲の感性情報抽出の従来研究では主に相関分析の手法が用いられてきたが、有効な特微量の設定が困難であることなどの問題から十分な抽出精度が得られていない。本研究では人間の音楽認知と情緒反応をモデル化し、そのモデルに適切な計算手法と考えられるニューラルネットワークモデルを用いて感性の計算モデルを得た。その際、音楽の段階的構成構造の特性や人間の知覚特性を考慮してニューロン間の結合に適当な制約を設け、また中間層に適切な競合系を設けた。この計算モデルが通常のニューラルネットワークモデルに比べ、より正確で効率的な学習能力を有することと、感性情報処理一般に有効であることを示す。

キーワード: 感性情報処理, 音楽認知モデル, ニューラルネットワーク, 近傍法

Sentimental-information detection by music cognition model

MUTO Makoto, HANNA Ibuki, SAKAI Shuichi and TANAKA Hidehiko

The university of Tokyo
7-3-1 Bunkyo-ku, Tokyo, 113-8656
{muto, handa, sakai, tanaka}@mtl.t.u-tokyo.ac.jp

ABSTRACT:

To detect sentimental-information of musical songs, multivariant analysis or neural network model have been used for analyzing the correlation between musical parameters and music listener's subjective report. But these methods are improper for Sentimental-information detection. So we use a music cognition model for sentimental information detection. The computational model is 3 layer neural network model - input layer represents perception of musical components, output layer represents the sentimental reaction evoked by the musical songs. Network structure is properly organized as to learn more precisely and effectively. In this paper, we discuss the modeling of music cognition and computational model of sentimental reactions.

KEY WORDS:

Sentimental-information processing, music cognition model, Neural Networks, Nearest Neighbor

1はじめに

今日、人間にとて感性が必要不可欠であることが次第に明らかになり、感性に関する研究の重要性が増している。音楽を対象とする研究では、楽曲の特徴とその曲を聴いた被験者による感性的特徴の評価との間の相関関係を調べる研究が行われている。これらの研究では、曲の特徴を特微量で表し、被験者による曲の感性的特徴の数値評価との間の相関関係を重回帰分析、またはニューラルネットワークによって分析している。このような手法は対象の間に比較的明確な対応関係がある分析対象に対して是有効であるが、感性と楽曲のように対応関係が不明確なものに対して用いるには適さないことが一般的に知られている。また、感性は主観的、状況依存的であるなど、複雑で定式化の困難な性質を有しているので、従来のような統計的な相関分析の手法では限界があると考える。

そこで本研究では、認知科学や音楽心理学などの知見に基づいて音楽聴取時における感性のモデルを設定し、これに基づいた計算モデルによって、楽曲が持つ感性情報¹を抽出することを目的とする。このように人間の感性を計算機上でモデル化することによって、従来手法では扱うことが困難であった感性を、計算機によって解析することが可能になると見える。本研究で扱う感性は、人間が音楽音響信号を聴き、その曲の印象を感性語²に対する主観的強度を尺度評価するまでを指す。

感性のモデルには実際の人間の認知機構に基づいて、音楽の構成要素の並列的知覚、知覚された構成要素と類似した記憶中の構成要素の想起、音楽の構成要素の間の相互作用、感性情報による情緒反応をそれぞれ構成要素とした。

感性モデルの計算モデルには3層の階層型ニューラルネットワークを用いた。入力層は音楽の各構成要素の主観的類似度マップの配列とし、出力層は5種類の形容詞に対する感性情報の強度とした。中間層は、構成要素の記憶の想起を行う部分と音楽の構成要素間の相互作用を行う部分とからなり、人間の知覚特性を考慮して、入力層、出力層との結合に適当な制約を設けた。また、中間層に競合系を設けることにより、近傍法による推論を可能にした。これにより、通常の階層型ニューラルネットワークと比べて、より正確で効率的な学習を行うことができ、また感性情報の特性に応じた情報抽出が可能となる。

¹人間の感性に働きかける情報。本研究では特に人間に情緒反応を引き起させる情報を扱う。

²形容詞など感性を表現するのに用いられる言葉

第2章では、感性情報抽出に関する関連研究と、従来手法の問題点について述べる。第3章では、本研究で用いる音楽認知モデルと、音楽の段階的構成構造について説明する。第4章では、感性の計算モデルについて述べる。第5章はまとめである。

2 楽曲の感性情報抽出

2.1 本研究の目的

本研究では、楽曲の譜面情報から、その楽曲が持つ「明るさ」などの感性情報を計算機で抽出することを目的とする。ここで、本研究で想定する楽曲は図1のような旋律パート、伴奏パート、打楽器パートの3パート構成のものとし、コード進行は無く、2小節(約4秒)のものとする。

また、本研究で扱う感性情報は、音楽の感情価に関する研究[谷口]に基づき、「明るい」、「穏やかな」、「強い」、「軽い」、「崇高な」のそれぞれの強度を図1のように0から1までで表したものとする。ここで、これらの感性情報は、曲全体が持つものとし、曲の進行に伴う時間変化は考慮しない。

また、感性情報抽出の応用分野としては感性データベース³や楽曲の特徴視覚化などがある。

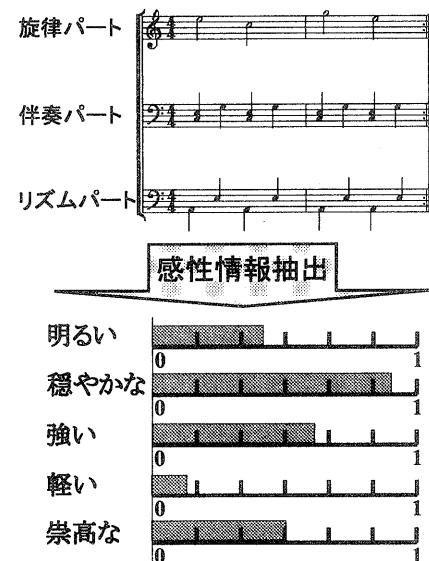


図1: 楽曲の感性情報抽出

³感性語を用いて検索、加工、変形などを行なうことのできるデータベース

2.2 従来手法とその問題点

楽曲の感性情報に関する関連研究には大きく分けて、被験者による情緒反応の報告によるものと、被験者の生理学的な測定⁴によるものがある。このうち、後者の生理学的測定による研究は多く行われているが、楽曲と生理学的反応との間の関係は未だ明らかになっていない [Radocy]。

一方、被験者による情緒反応の報告による研究も多く行われている。そのような研究では、楽曲の譜面情報・音響情報の各種特徴量と、アンケート被験者による感性的評価値に基づいて設定された感性空間との間の相関関係を、多変量解析で分析 [池添], [武藤]、またはニューラルネットワークによる機械学習によって解析 [坂本], [日間] という手法が主に用いられている。しかしながら、これらの手法には以下のような問題点があるので、その抽出精度には限界があると考える。

- (問題点 1) 感性情報抽出に有効な特徴量を設定するのが難しい。

上のような相関分析の手法においては特徴量の設定が重要であるが、従来手法で用いられている特徴量と感性情報との対応関係は定かではない。

- (問題点 2) 音楽の段階的構成構造を考慮していない。

音楽は音色、旋律、リズムなど複数の構成要素から成り立っており、それらの間の相互作用が感性情報に影響を与えることが確かめられている [梅本]。しかしながら、従来手法では音楽の階層構造を考慮していない [坂本], [池添], [武藤]。また一方で、階層構造を持たない単旋律のみを扱った研究 [辻] も行われているが、このような音刺激はもはや楽曲とはいえないでの、音楽の研究素材として不適切である。

- (問題点 3) 人間の持つ知識、記憶を考慮していない。

音楽の認知には、音楽に関する知識や過去に聴いた曲の記憶が影響することが確かめられている [Radocy]。音楽に関する知識を楽曲の感性情報抽出に用いた研究 [片寄] では、特定の音楽の構成要素から特定の情緒反応を活性化させるルール群を分析対象の楽曲に適用して、感性情報を抽出することが行われているが、ルールの設定が経験的であることや、感性情報の定量的な側面を扱えないなどの問題点がある。

以上の問題に対処するため、本研究では音楽の認知モデルを用いた感性情報抽出を行う。

⁴ 血液の循環、心拍数、呼吸などの測定

3 音楽認知モデルを用いた感性情報抽出

3.1 音楽認知と情緒反応のモデル

人間が音楽音響信号を聴き、その曲の印象を感性語で表現するまでの感性の過程として、以下のものが寄与していると考え、図 2 のような音楽認知モデルを設定した。

音楽の構成要素の並列的知覚

我々の脳の情報処理能力には限界があることから、脳では分散・並列的な処理が行われていると考えられている [安西]。また、人間は無意識に知覚されたものに一貫した意味を与え、不安定な知覚を避けていると考えられている [安西]。特に音楽は複数の旋律、音色、歌詞などの構成要素から成り立っており、それぞれ独立に一貫した意味を持つ。人間はそれぞれの構成要素の選択的知覚を並列的に行っていると考える。

音楽に関する記憶・知識

音楽に関する記憶は曲の認知の際の連想などと関係し、曲の認知のしかたに影響を与えることが実験で確かめられている [Radocy]。例えば、未知曲と既知曲とでは認知のしかたが異なることが知られている。また、音楽の知識が曲の認知のしかたに影響を与えることが確かめられている [Aiello]。

音楽の構成要素間の相互作用

音楽の構成要素としては旋律、音色などが考えられるが、それらはそれぞれ完全に独立に認知されるわけではなく、お互いに相互作用を及ぼすことが知られている [Radocy], [梅本]。例えば、主旋律と伴奏の関係は旋律の解釈などに影響を及ぼすと考えられている。

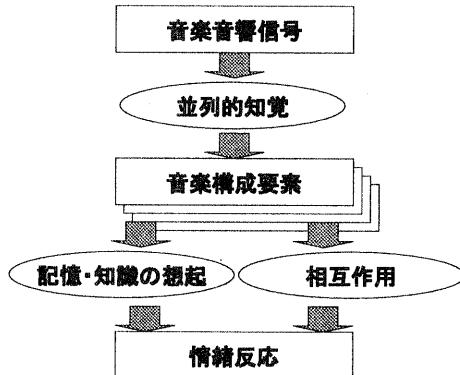


図 2: 音楽認知と情緒反応の心理的モデル

3.2 音楽の構成要素

本研究で想定する2小節の曲には表1のような階層構造がある[梅本]。ここで、A「音響」の属性の音高、長さ、大きさは曲の形で知覚される場合は、B「知覚対象」の属性の旋律、リズムとしてまとめて知覚されるので、旋律、リズムとしてまとめて扱うことができる。そこで、本研究では音楽の持つ属性として、音色、旋律、リズム、和声を扱う。

表1: 音楽の階層構造

	次元	属性
A	音響信号	音色、ピッチ、長さ、大きさ
B	知覚対象	旋律、リズム、和声

次に、それぞれの属性について本研究で想定する3パート構成の曲に対し、各パート(旋律パート、伴奏パート、リズムパート)毎に構成要素を設定する。

音色属性

各パートの楽器音を構成要素とする。

旋律属性

旋律パートにおいて旋律属性として重要なものに、輪郭⁵と音程の幅⁶がある[Aiello]。この輪郭と音程の幅を構成要素とする。

伴奏パートにおける旋律属性は、伴奏パターンとする⁷。リズムパートの旋律属性は無しとする。

リズム属性

各パート毎に音符の時系列パターンを、構成要素とする。

和声属性

伴奏パートにおける和声属性として、旋律の音名の頻度⁸ [Aiello]を設定する。伴奏パートにおける和声属性として、コード名を設定する⁹。リズムパートの和声属性は無しとする。

以上をまとめ、本研究では音楽の構成要素として、図3の10種類のものを用いる。

3.2.1 構成要素間の相互作用

音楽の構成要素には様々な相互作用¹⁰があることが知られている。例えば、旋律輪郭は知覚と記憶において、調性とリズムパターンの双方と相互作用し

⁵隣接する音符の上昇、下降の時系列のパターン。上昇を「+」、下降を「-」で表す。

⁶隣接する音符の音高の幅

⁷伴奏はある一定のパターンが繰り返されることが多く、そのため伴奏パターンは旋律の属性を持つ。

⁸各音名の旋律中の出現頻度

⁹伴奏パートはコードに基づいて演奏され、和声属性が強い

¹⁰複数の構成要素が同時に知覚されることで、曲全体の感性情報に変化が生じること

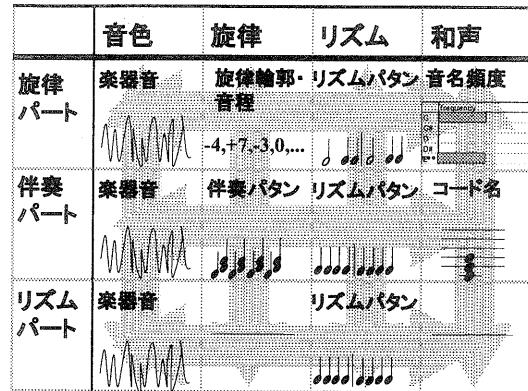


図3: 音楽構成要素のその相互作用

ている。また、旋律と音色との間に相互作用があることが示されている[梅本]。

このような研究から、音楽の構成要素の間には相互作用があることがわかる。そこで、本研究では図3の矢印のように、各属性についてパート間の相互作用(図3の縦の矢印)と、各パートについて各属性間の相互作用(図3の横の矢印)を扱う。

3.2.2 各構成要素の主観的類似度の計算

本研究では楽曲の感性情報の抽出を行うが、その際、システムに音楽の構成要素に関する記憶と知識を学習によって予め与える。しかし、感性情報抽出対象の曲(未知曲)の構成要素がシステムの記憶・知識の中に存在するとは限らない。このような場合、記憶・知識を用いて未知曲の感性情報を推測する必要がある。その際、図4のような各音楽構成要素の類似度関係を用いて、未知曲の感性情報を推測する。このような類似度マップの計算方法は様々提案されている[栗田],[鈴木]。本研究ではこれらの研究を参考にして、それぞれの音楽構成要素について類似度マップを作成し、感性情報抽出の計算モデルにおいて用いている。

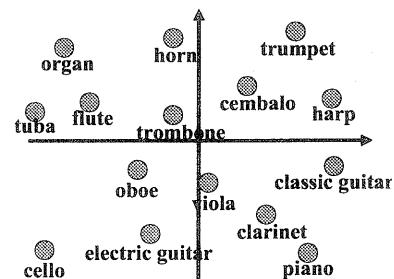


図4: 音色の主観的類似度マップ

4 感性の計算モデル

本章では前章で得られた、音楽認知と情緒反応の心理学的なモデルに計算手法を割り当て、感性の計算モデルを構成する。

4.1 計算モデルの全体構成

本研究で用いる計算モデルは、図 5 のような 3 層からなるニューラルネットワークモデルである。

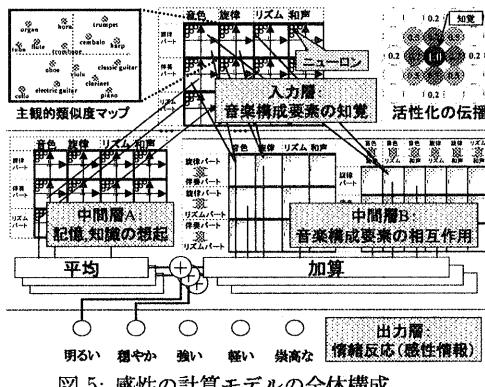


図 5: 感性の計算モデルの全体構成

入力層: 音楽構成要素の知覚

入力層は、図 5 のように各構成要素に対応して分割されていて、各音楽構成要素の類似関係を表す類似度マップ上にニューロンが並べられている。各構成要素が知覚されると対応するニューロンが発火し、活性度が 1 になる。また、図 5 右上のようにニューロンの活性化は周辺のニューロンに伝播する。

中間層: 音楽構成要素の記憶想起と相互作用

中間層は構成要素の記憶想起を行う部分（中間層 A）と構成要素間相互作用を行う部分（中間層 B）とから成る。

中間層 A のニューロンの配列は入力層と同じである（図 5）。中間層 B は、音楽の各属性間相互作用を行うものと、各属性に対するパート間の相互作用を行うものがある。

出力層: 情緒反応

出力層は、図 5 のように、「明るい」、「穏やかな」、「強い」、「軽い」、「崇高な」の各情緒反応の強度を表す 5 つのニューロンから成る。それぞれのニューロンの活性度がそれぞれ対応する感性情報を表す。

4.2 結合の制約と競合系の設定

記憶の検索と情緒反応（入力層 - 中間層 A - 出力層）

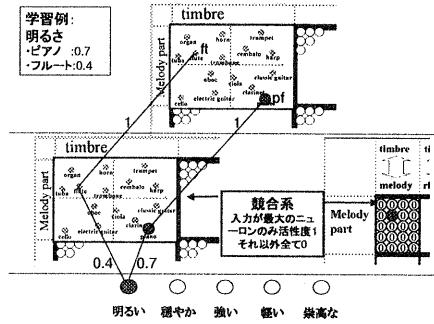


図 6: 感性の計算モデルの例

入力層と中間層 A との間では、記憶している音楽構成要素¹¹に対応するニューロン同士が図 6 のように 1 対 1 関係で結合されていて、それ以外の結合はない。結合度は全て 1 で固定である。中間層 A と出力層との結合は、記憶のある構成要素に対応するニューロンにのみあり、結合度は対応する構成要素の感性情報の強度を表す。

また、中間層の各ニューロン群（図 5 で黒の四角で囲まれたニューロン群）は競合系にあり、入力が最大のニューロンだけが活性度が 1 となり、それ以外のニューロンの活性度は 0 となる（図 6）。

この競合系によって、知覚された構成要素と最も類似した記憶中の構成要素に対応するニューロンが選択的に活性化される。これによって、知覚された構成要素と最も類似した記憶中の構成要素の検索が実現される。

また、中間層の競合系により、本研究で想定する 3 パート構成の楽曲が知覚されると中間層 A では必ず 10 個¹² のニューロンが活性化する。本研究では、相互作用の無い構成要素から成る楽曲の感性情報は、それぞれの構成要素の感性情報の平均であると仮定する。そのため、中間層 A からの出力を平均して、相互作用を行う中間層 B の出力を加算し（次項参照）、出力層のニューロンの活性度とする。

音楽構成要素の相互作用（入力層 - 中間層 B - 出力層）

入力層と中間層 B の間の結合は、相互作用の対象となる 2 つの構成要素に対応したニューロン群からそれぞれ 1 つずつ結合する形に限られる。結合度は全て 1 で固定である。

中間層の競合系の設定は、中間層 A と同様である。

また、音楽構成要素の相互作用は感性情報の強度の変化の度合とする。そのため、中間層 A の出力に対して、中間層 B の出力を加算して出力層のニューロンの活性度とする。

¹¹学習データに含まれている音楽構成要素

¹²音楽構成要素の数が 10 個のため。

4.3 感性の計算モデルの学習

学習用曲 100 曲程度と、それを図 7 のように構成要素毎に順次分解した曲¹³を全て用いて被験者にアンケートを行ない、学習データを得る。その学習データを用いて、各層間に結合を設定する。その際、結合度はの Perceptron や Hebb の学習則¹⁴を用いて設定する。

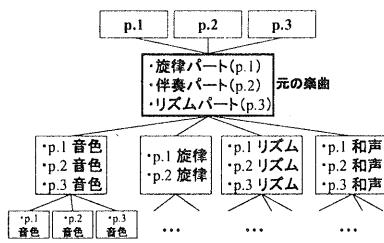


図 7: アンケートに用いる曲刺激

4.4 感性情報処理における本手法の有効性

ニューロン間の結合の制約による学習の効率化

一般に、感性情報処理ではアンケート結果を学習データとして用いる。そのため、学習データの数は比較的少なく、また主観的評価によっているために誤差が大きい。そのため、学習は効率的に正確に行なう必要がある。

通常の階層型ネットワークでは、隣接する層の間のニューロンはお互いに総当たりに結合されているが、本研究で用いるニューラルネットワークでは結合が意味のあるものに制約されている。そのため、本手法によって効率的で正確な学習を行なうことができる。このことから、本手法は感性情報処理に有効であるといえる。

中間層の競合系による近傍法の実現

本手法では中間層に競合系を設定することによって、近傍法の推論形式を実現した。

認知科学によると、人間は知覚対象物を自分が持つ知識、記憶で説明出来るように変形して知覚する傾向がある [安西]。すなわち、知覚されたものに最も類似した記憶中の構成要素が実際には知覚される傾向がある。このように、人間の知覚や感性には近傍法と同じ特性がある。よって、本手法は感性情報処理に適しているといえる。

¹³ 例えれば、図 7 の左下の「p.1 音色」は、旋律パートの楽器音だけを単独で鳴らした音を意味する。

¹⁴ ニューラルネットワークの学習則としてよく用いられる。

5 まとめ

楽曲の感性情報抽出の従来手法では、(1) 相関分析に有効な楽曲の特徴量の設定が困難である、(2) 音楽の段階的構成構造を考慮していない、(3) 音楽聴取者の持つ音楽に関する記憶・知識を考慮していない、などの問題点があった。これらの問題に対応するために本研究では人間の音楽認知と情緒反応をモデル化し、そのモデルに適切な計算手法と考えられるニューラルネットワークモデルを用いて感性の計算モデルを得た。その際、音楽の段階的構成構造の特性や人間の知覚特性を考慮してニューロン間の結合に制約を設け、また中間層に適切な競合系を設けた。この計算モデルが通常のニューラルネットワークモデルに比べ、より正確で効率的な学習能力を有することと、感性情報処理一般に有効であることを示した。

参考文献

- [安西] 安西祐一郎 他:「岩波講座 認知科学 9 注意と意識」, 岩波書店, 1994.
- [池添] 池添剛, 梶川嘉延, 野村康雄:「形容詞対を用いた音楽データベース検索システムに関する研究 (II)」, 情報処理学会第 59 回全国大会講演論文集 (2), pp.21-22, 1999.
- [梅本] 梅本聰夫 編:「音楽心理学の研究」, ナカニシヤ出版, 1996.
- [岡田] 岡田頼宏, 阿部純一:「音楽知覚におけるモジュール性」, 音楽知覚認知研究, Vol.5, No.1, pp.22-43, 1999.
- [片寄] 片寄晴弘, 今井正和, 井口征士:「音楽における感性情報抽出の試み」, 人工知能学会誌, Vol.3, No.6, pp.66-72, 1988.
- [栗田] 栗田多喜夫, 下垣弘行, 加藤俊一:「主観的類似度に適応した画像検索」, 情報処理学会論文誌, Vol.31, No.2, pp.227-237, 1990.
- [坂本] 坂本崇, 梶川嘉延, 野村康雄:「音楽感性空間における非線形判別分析を用いた曲印象別グループの分類」, 情報処理学会論文誌, Vol.40, No.4, pp.1901-1909, 1999.
- [鈴木] 鈴木健嗣, 橋本周司:「ニューラルネットワークを用いた感性情報の数量化」, 電子情報通信学会論文誌, Vol.J82-D-II, No.4, pp.677-684, 1999.
- [辻] 辻康博, 星守, 大森 匠:「曲の局所パターン特徴量を用いた類似曲検索・感性語による検索」, 電子情報通信学会研究技術報告, SP96-124, pp.17-24, 1997.
- [谷口] 谷口高士:「音楽作品の感情価値測定尺度の作成および多面的感情状態尺度との関連の検討」, 心理学研究, 65, 463-470, 1995.
- [日間] 日間賀充寿, 大西昇, 杉江昇:「情動に相関のある楽曲中のパラメータについて」, 情報処理学会研究報告 94-MUS-103, pp.7-12, 1994.
- [武藤] 武藤誠, 木下智義, 半田伊吹, 坂井修一, 田中英彦:「音楽音響信号からの楽曲の感性的特徴の抽出」, 情報処理学会第 59 回全国大会講演論文集 (2), pp.11-12, 1999.
- [Aiello] Aiello, R.: Musical Perceptions, Oxford Univ. Press, 1994.
- [Radocy] Radocy, R. E.: Psychological Foundations of Musical Behavior, Charles C Thomas Publisher, 1979.