

音楽検索のための感性表現ベクトルと音響特徴量の関係の分析

上野 智子、相川 清明

東京工科大学メディア学部
〒192-0982 八王子市片倉町 1404-1
aik@media.teu.ac.jp

あらまし 楽しい、悲しい、落ち着いたなどの感性表現で効果音楽を検索する Sound Advisor システムについてすでに報告している。本研究ではこれらデータベースの曲ごとの感性表現ベクトルを楽曲から自動生成するために音響特徴量と感性表現ベクトルとの関係の分析をおこなった。効果音楽のパワー、ピッチなどの音響特徴と現在ある感性表現ベクトルとの重回帰分析を行うことで、効果音楽の音響パラメータから感性表現パラメータへの変換行列を導く方法を提案する。

The Relation between Emotional Representation Vectors and Acoustic Features for Sound Retrieval

Tomoko Ueno and Kiyoaki Aikawa

School of Media Science, Tokyo University of Technology
1404-1 Katakuracho, Hachioji, Tokyo 192-0982, Japan

Abstract Emotional representations are more effective than conventional keywords such as genres and artist names in retrieving music. A vector-based Sound Advisor system has been reported for retrieving background music in emotional phrases such as “happy”, “sad”, “calm” or “angry”. This report analyzed the relation between emotional representation vectors and acoustic features for automatically generating emotional representation vector table of the sound retrieval system. This report proposes a method for deriving a transformation matrix from acoustic feature parameters to emotional parameters based on regression analysis.

1. はじめに

近年、デジタル技術の進歩に伴い音声、画像、動画、音楽、テキスト文書など、大量かつ多様な情報コンテンツを扱う機会が増え、大量のデータベースの中からユーザが所望のコンテンツを素早く、そして正確に取り出せる検索技術の開発が望まれている。音楽データの検索では、アーティストやタイトルといったキーワードの文字情報による検索が主流である。しかし音楽を検索する際、ある時の気分や状況に合わせた音楽の内容で検索できる必要性が高いと考えられる。音楽は聴いた時に何かしら印象を持つ。こういった人間の感性を生かした検索ができる必要性がある。

感性語を用いた検索として様々なアプローチで研究が行われている。旋律の音高推移に着目し音高推理特徴量と感情価との関係を分析した方法では音高推移特徴量と高揚・抑鬱の感情価との間に相関関係があると予測できるという結果が得られた[1]。リズムパターンに着目し、人間が感じる感情反応との分析を行った研究では、結果研究曲調変化の感知者数と主旋律のリズムパターンのマッチングに強い相関があった。ここでは楽曲から喚起される感情反応には 50～100%の共通性があると述べられている [2]。さらに、イメージによる検索システムにおいて楽譜情報と感情価との関係を求め、新たな作品の高揚、抑鬱、強さ、軽さなどの感情価を自動生成する方式の提案[3]や、音楽作品の感情的側面を種類と量を表す感情価に着目し感情価尺度の概念を用いた音楽検索の研究[4]などがある。

本研究では楽しい、悲しいなどの感性表現を用いた音楽データの検索を対象とする。ベクトル空間法を使った感性表現での音楽検索として Sound Advisor システムについて既に述べている [5]。楽しい、悲しい、落ち着いたなどの感性表現によってデータベース内の効果音楽を検索するシステムである。評価実験の結果、このシステムの検索結果の満足度は 74%であった。

この感性表現ベクトルのテーブルは、検索対象の曲すべてについてアンケート調査を行って求めている。しかし、アンケート調査は時間のかかる作業であり、自動的にこの感性表現ベクトルテーブルが得られることが望ましい。

各種の音響特徴と現在ある感性表現ベクトルとの回帰分析を行うことで効果音楽の音響パラメータから感情パラメータへの変換行列を導く方法を提案する。今回は音響特徴として暫定的にピッチ、パワー、動的尺度の特徴量を使い分

析を行い、「楽しい」「悲しい」の項目については、これらの音響特徴である程度説明できる事が分かった。

2. Sound Advisor システム

感性表現の 8 つの項目「楽しさ、悲しさ、落ち着き、怒り、恐怖、不気味、明るさ、爽やかさ」をそれぞれ 5 段階で入力することで探している感性にあった効果音楽を検索できるシステムである。外部から取り込んだ情報と、内部で保持される情報とを共に感性表現ベクトルとして表現し、二つの類似度を計算し最も類似度の高いものを選択するベクトル空間法を利用した。音の特徴は N 次元空間中のベクトルとして表現され、検索要求に対し、最も類似度が高いテンプレートベクトルに対応する音が検索結果として示される。

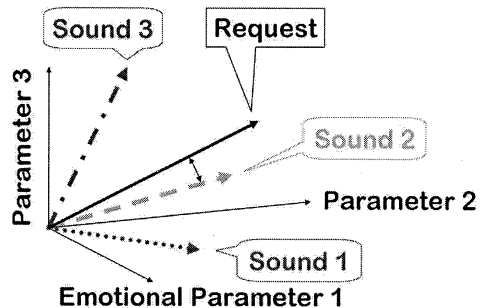


図1ベクトル空間法による音検索の模式図

2.1 使用曲

楽しい曲や悲しい曲など様々な場面で用いられる曲があり、歌謡曲など幅広い範囲の音楽より感情からの検索をする上で効率が良いため、データベースには効果音楽を使用した。

「ビクター効果音ライブラリー10 効果音楽」
ビクターエンターテイメント

テーマ音楽を集めた CD であり、アニメーション、ドラマ、ニュース、明るさ、怒り、落ち着きなどの 18 種類の感情がテーマとなっておりあらかじめ雰囲気分けられているものである。各曲は 10 秒～60 秒程度である。この CD から 88 曲を使用する。参考に以下のようなタイトルの効果音楽である。

- ・おどけた
- ・お笑い
- ・サスペンス
- ・悲しみ
- ・ゆったりした
- ・ホラー
- ・にぎやかな
- ・ホームドラマ

2.2 データベースの作成

様々な種類の音源を聞いてもらう簡単なアンケートを行い、聴いた音を言葉で表現してもらった。その結果出た言葉と今回使用した効果音CDテーマ音楽で用いられている18種類の言葉の中から「楽しい」「悲しい」「怖い」「落ち着く」「怒り」「不気味」「明るい」「爽やか」8つの項目を選出した。この8つの項目を使い使用曲88曲について5段階評価(1.非常に遠い 2.やや遠い 3.どちらともいえない 4.やや近い 5.非常に近い)を行った。被験者数は14名である。ベクトルはアンケートを取った人数分の合計数値を使用する。

1曲に対して

感性項目(各5段階評価)×8項目=40次元

40次元×88曲 のテーブルを使用する。

表1 感性表現ベクトルテーブル

	楽しい					悲しい					怖い					...
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	
曲1	4	3	4	2	0	6	2	3	2	0	3	2	3	3	1	...
曲2	3	3	5	2	0	2	1	4	5	1	8	3	2	1	0	...
曲3	11	2	0	0	0	0	0	0	3	10	2	5	5	2	1	...
...

2.3 結果

全体で見たとき、自分のイメージした曲に近いと答えた割合は74%であった。

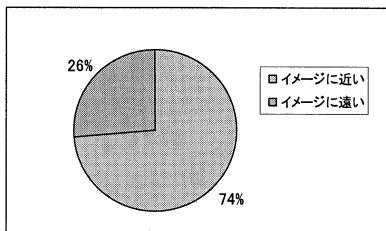


図2 Sound Advisor システム満足度

3. 提案方法

Sound Advisor システムの音楽データベースと対応した感性表現ベクトルのテーブルについては(2.3項)で挙げた通りである。この感性ベクトルのテーブルは検索対象の曲すべてについてアンケート調査を行って求めていた。しかし、ア

ンケート調査は時間のかかる作業であり、楽曲から自動的にこの感性ベクトルテーブルが得られることが望ましい。本報告では、各曲ごとの感性表現ベクトルと、音響特徴との関係付けを行うことで感性表現ベクトルのテーブルを自動的に算出する方法を提案する。

3.1 データベース変換

各感性語あたり5次元のベクトルになっていたものを分布関数とみなして平均値をもとめ、40次元のベクトルを8次元に変換した。この変換により、5次元中のベクトルの方向で表していた情報を1次元のベクトルの長さで表現することになる。

1曲に対して

感性項目 × 8項目 = 40次元

8次元 × 88曲 のテーブルを使用する。

表2 感性表現ベクトルテーブル(提案方法)

	楽しい	悲しい	怖い	落ち着	怒り	不気味	明るい	爽やか
曲1	0.79	-1.21	-1.71	0.07	-1.79	-1.57	1.07	1.57
曲2	-0.07	-0.46	-1.57	0.64	-1.93	-1.79	0.31	0.07
曲3	-1	0.29	-0.63	-0.21	-1.21	-0.5	-0.57	-0.29
...

3.2 変換アルゴリズム

88曲すべてについて感性項目と関連のありそうな音響特徴を抽出する。

さらにそれらの特徴量と感性項目との回帰分析を行い、特徴量から感性項目を求める。

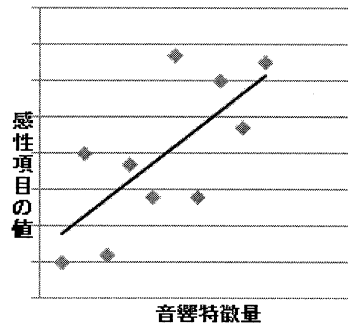


図3 提案方法

感性表現ベクトルを 音響特徴ベクトルを

$$E = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_M \end{pmatrix} \quad A = \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{pmatrix}$$

とする。音響特徴ベクトルから感性表現ベクトルを求める変換行列 C と定数ベクトル B を求めることができれば、式(1)のように、楽曲から音響特徴ベクトルを求め、さらに感性表現ベクトルを求めることができる。

$$E = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_M \end{pmatrix} = CA + B \quad (1)$$

$$= \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1N} \\ c_{21} & c_{22} & \dots & c_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ c_{M1} & \dots & \dots & c_{MN} \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_N \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_M \end{pmatrix}$$

本研究では重回帰分析により変換行列 C を求める。1つの感性表現ベクトル e_j の要素について音響特徴に関する重回帰分析を行えば、式(2)のように音響特徴の一次式を求めることができる。

$$e_j = \sum_{i=1}^N c_{ji} a_i + b_j \quad (2)$$

既にアンケート調査により 88 楽曲の感性表現ベクトルが得られている。これらの楽曲の音響特徴を求めれば、重回帰分析により変換行列 C と定数行列 B が求まる。これらが求まれば、未知楽曲についても音響特徴から感性表現ベクトルを求めることができる。これらを Sound Advisor のデータベースとして組み込めば、アンケート調査を行うことなく未知楽曲を含めて感性表現により楽曲検索を行うことが可能となる。

変換の流れとしては図4のようになり、未知曲を追加する際に音響特徴から自動的に感性表現ベクトルに変換し、曲に対応した感性表現ベクトルテーブルを生成していくという流れになる。

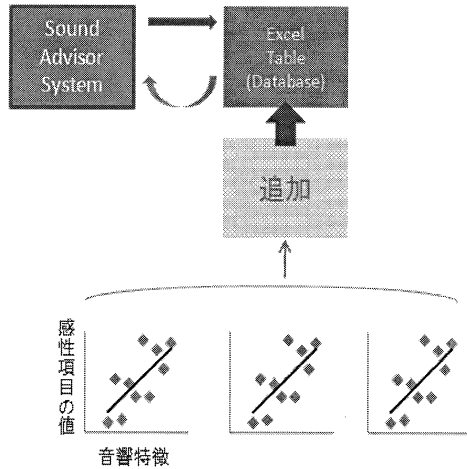


図4 感性ベクトル自動生成方法のイメージ

4. 分析

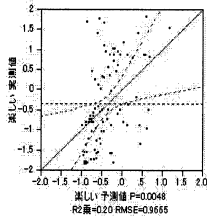
分析には MATLAB を使用する。

パワーの分散が強弱の激しさ、パワーの平均が曲全体を通しての音量、ピッチの分散が音高変化の大きさ、ピッチ平均が曲全体を通しての音高、低次ケプストラムによる動的尺度のピーク数が音色の変遷などを反映するのではないかと考えた。本報告ではこれらの値を用いて、各値が曲ごとに違いを表わせる特徴となるかどうかを調べた。

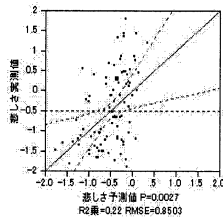
パワー(対数パワー(db))
分析窓幅: 46.4ms
ケプストラム次数 1-16 次

統計分析ツール JMP により各感性項目ごとに重回帰分析を行った結果を示す。

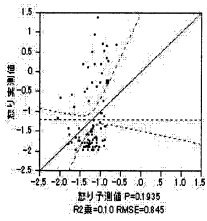
(a) 楽しい



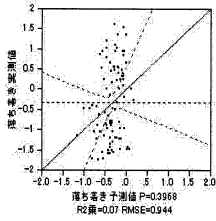
(b) 悲しい



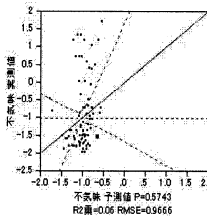
(c) 怒り



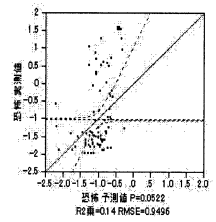
(d) 落ち着き



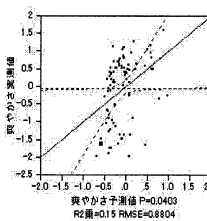
(e) 不気味



(f) 恐怖



(g) 爽やか



(h) 明るい

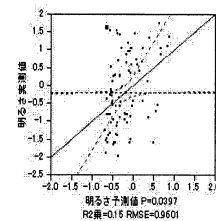


図5 モデル全体による予測値と実測値

4.2 分析結果

以上の結果から、「楽しい」「悲しい」については全体での破線が横軸を回帰直線（実線）と95%信頼曲線（破線）が平均値を通る水平方向の直線（破線）と交差しているの、これらの音響特徴である程度説明できる事が分かった。

5. 変換行列の分析

以上の結果から「楽しい」「悲しい」の項目について予測を行った。

線形重回帰モデルは以下の様になる。

$$y_i = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i \quad (3)$$

$(i = 1, 2, \dots, n)$

「楽しい」についての予測式は以下のようになる。

表3 楽しいについての予測係数

項	推定値	標準誤差	t値	p値 (Prob> t)
切片	-1.7618	1.0127	-1.74	0.0857
動的尺度平均	0.1146	0.0327	3.51	0.0007
動的尺度個数	0.0061	0.0063	0.96	0.3375
パワー平均	-0.0123	0.0211	-0.58	0.5603
ピッチ平均	0.0105	0.0223	0.47	0.6403
パワー分散	0.0008	0.0007	1.07	0.2868
ピッチ分散	-0.0007	0.0007	-0.94	0.3504

$$\begin{aligned} \text{楽しい} = & (-1.7618) + 0.1146 \times \text{動的尺度平均値} \\ & + 0.0061 \times \text{動的尺度個数} + (-0.0123) \times \text{パワー} \\ & \text{平均} + 0.0105 \times \text{ピッチ平均} + 0.0008 \times \text{パワ} \\ & \text{一分散} + (-0.0007) \times \text{ピッチ分散} \end{aligned} \quad (4)$$

「悲しい」についての予測式は以下のようになる。

表4 悲しいについての予測係数

項	推定値	標準誤差	t値	p値 (Prob> t)
切片	1.3367	0.8809	1.5	0.1374
動的尺度平均	-0.08	0.0267	-2.78	0.0067
動的尺度個数	-0.0042	0.0055	-0.76	0.451
パワー平均	-0.0164	0.0186	-0.89	0.378
ピッチ平均	-0.0478	0.0196	-2.43	0.0173
パワー分散	-0.0009	0.0006	-1.33	0.1858
ピッチ分散	0.0005	0.0006	0.81	0.4205

$$\begin{aligned} \text{悲しい} = & 1.3367 + (-0.08) \times \text{動的尺度平均} + \\ & (-0.0042) \times \text{動的尺度個数} + (-0.0164) \times \text{パワ} \\ & \text{ー平均} + (-0.0478) \times \text{ピッチ平均} + (-0.0009) \times \\ & \text{パワ} \\ & \text{ー分散} + 0.0005 \times \text{ピッチ分散} \end{aligned} \quad (5)$$

値についての考察

「楽しい」では動的尺度平均、動的尺度個数が正であるが、「悲しい」では負である。ピッチ平均、分散、パワー平均、分散に関しては差が見られなかった。

5.2 生成したテーブルの評価

提案した変換により適切なベクトルテーブルが生成できるかどうかをベクトルテーブル間の距離を用いて評価した。8種の感性項目に対して変換式を用いて88種の音の感性ベクトルを生成し、もとの感性ベクトルとの一致度を分析した。

(A=元の行列、B=変換して求めた行列、
n=感性項目数、m=曲数)

$$D = \sqrt{\frac{1}{nm} \left(\sum_{m=1}^m \left(\sum_{n=1}^n (A-B)^2 \right) \right)} \quad (6)$$

の式を用いて行列間の違いを計算した。

D=0.2943

アンケート調査に基づく感性ベクトルの値の範囲は-2~+2のため、元の行列と変換して求めた行列が比較的近いものになっているといえる。

6. むすび

感性項目を用いて効果音楽を検索できる Sound Advisor システムに音楽データベースを追加していく際、自動的にデータベースを作成できる必要がある。音響特徴から感性表現ベクトルに変換する方法として、各種の音響特徴と現在ある感性表現ベクトルとの重回帰分析を行うことで効果音楽の音響パラメータから感性表現パラメータへの変換行列を導く方法を提案した。今回は音響特徴としてピッチ、パワー、動的尺度の特徴量を使い、分析を行った。結果「楽しい」「悲しい」については、これらの音響特徴である程度説明できる事が分かった。予測式を用いて感性表現ベクトルテーブルを生成し、元の感性表現ベクトルとの一致度を分析したところ、比較的近いものになっているといえる結果が得られた。適切な音響特徴の探索が今後の課題である。

参考文献

- [1] 堀野義博, 佐藤聡, 黒木進, 北上始, “旋律の音高変化に基づく音楽特徴量についての検討”, 音楽情報科, 40-5, (2001-05-23), Vol.2001, No.45(20010523) pp. 29-34, 2001-MUS-40-5
- [2] 川野達誠, 亀田昌志, 宮原誠, “楽曲により喚起される感情反応と Rhythm に基づいた曲調変化との関係”, 情報処理学会 音楽情報科学, 42-5, (2001-10-26) Vol.2001, No.103(20011026) pp. 27-34, 2001-MUS-42-5
- [3] 佐藤聡, 菊地幸平, 北上始 “音楽データを対象としたイメージ検索のための感情価の自動生成”, 情報処理学会, データベースシステム 118-8 情報学基礎 54-8, (1999-05-17), Vol.99, No.39(19990517) pp. 57-64, 99-DBS-118-8
- [4] 佐藤聡, 小川潤, 堀野義博, 北上始 “感情に基づく音楽作品検索システムの実現に向けての検討”, 情報処理学会 音楽情報科学, 39-8, (2001-02-22). Vol.2001, No.16(20010222) pp. 51-56 2001-MUS-39-8
- [5] 相川 清明, 谷島加奈子, “ベクトル空間法を用いた相対的感性表現による音検索”, 音声言語情報処理, Vol.2007, No.11(20070209) pp. 5-10 2007-HI-122-(2)
- [6] 福村純也, 川越恭二, “特徴空間とメロディ空間を用いた楽曲の類似検索方法”, データベースシステム研究会, 129-3 (2003-1-23) Vol.2003, No.5(20030123) pp. 17-24 002-DBS-129-3
- [7] 辻康博, 星守, 大森匡, “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・音声 SP96-124(1997-3) Vol.96, No.56 5(19970306) pp. 17-24
- [8] 平江遼, 西隆司, “感性に基づくクラシック音楽の分類”, 音響学会講演論文集(2007-9)