

# 感性による音楽検索のための 音響特徴からの感性ベクトル自動生成

上野 智子 相川 清明

東京工科大学大学院バイオ情報メディア研究科メディアサイエンス専攻  
〒192-0982 八王子市片倉町 1404-1  
aik@media.teu.ac.jp

**あらまし** 感性表現による効果音楽検索システムのデータベースである楽曲ごとの感性ベクトルを、学習用楽曲の音響特徴と感性ベクトルとの重回帰分析に基づき自動生成する方法と、そのための音響特徴の選択方法について述べる。その結果、パワー、ピッチ、スペクトル変化を含む 19 種類の音響特徴の組み合わせにより未学習データに対して良好な感性ベクトルを生成できることが分かった。本報告ではさらに、自動生成した感性ベクトルによる検索性能の評価実験結果について述べる。

## Automatic Generation of Emotional Feature Vectors from Acoustic Features for Emotion-based Music Retrieval

Tomoko Ueno and Kiyooki Aikawa

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

**Abstract** An emotion-based music retrieval system has been developed using an emotional vector database. This report describes a method for automatic generation of these vectors using multiple regression analysis between emotional feature vectors and acoustic features. The best emotional feature vectors were obtained using 19 acoustic features including power, pitch and spectrum change. This report also describes the results of subjective tests on music retrieval.

## 1. はじめに

音楽データの検索では、ある時の気分や状況などに合わせた人間の感性を生かした検索が有効である。感性表現を用いた音楽検索として Sound Advisor システムがある [1]。「楽しさ」「悲しさ」など 8 つの感性項目を 5 段階で入力する事で探している感性に合った曲を提示するシステムである。検索要求とデータベースの楽曲を多次元空間中の感性ベクトルとして表し、最も類似度の高いものを検索結果として選択するベクトル空間法を利用している。

このシステムで使用する感性ベクトルデータベースは、使用曲全てについて 8 つの感性項目をそれぞれ 5 段階で評価してもらうアンケート調査を行って作成されたものである。これらのデータベースに楽曲を追加する際はこの評価アンケートを行う必要がある。この方法では時間や作業効率を考えるととても現実的ではなく、感性ベクトルデータベースを自動的に作成できる必要がある。

そこで、各種の音響特徴と現在ある感性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パラメータへの変換行列を導き、自動的に感性ベクトルを生成する方法を提案した [3]。音響特徴としてピッチ、パワー、スペクトル変化の大きさを表す特徴量である動的尺度 [5] の特徴量を使い、データベース内の全曲において予測式を用いて感性ベクトルテーブルに変換したところ、元の感性ベクトルテーブルと比較的一致するものを自動作成することができた [4]。

本稿では、音響特徴量の数や組み合わせに最適なものを導けば性能が向上すると考え、主成分分析などを用いて音響特徴の選定を行った。さらに、それによって作成したテーブルを組み込んだシステムの検索性能の評価実験を行った。

## 2. Sound Advisor システム

8 つの感性項目「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 40 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。

### 2.1 使用曲

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 効果音楽から 86 曲。それぞれ 10 秒から 1 分程度の曲があるが、本研究では全曲、曲頭から 10 秒をカットして利用した。

### 2.2 感性ベクトルデータベースの作成

全曲に対して、8 つの感性項目を 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、の 5 段階で評価してもらうアンケートを行った。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を使用した。1 曲に対して 40 次元のテーブルである。

表 1 感性ベクトルテーブルの一部

	楽しさ					悲しさ					恐怖					...
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	
曲1	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5	...
曲2	0	0	4	9	1	6	5	3	0	0	12	1	0	1	0	...
曲3	1	6	2	3	2	5	0	4	4	0	10	2	2	0	0	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...

### 2.3 システム評価

このシステムを使用し、システムの評価実験を行ったところ、検索結果が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 [2]。

### 3. 感性ベクトル自動生成

Sound Advisor システムの感性ベクトルデータを自動的に作成するために、音響特徴を利用して感性ベクトルに変換する方法を提案した [3]。既にアンケート調査により 86 楽曲の感性ベクトルが得られている。これらの楽曲からいくつかの音響特徴を求め、重回帰分析により変換行列と定数行列を求めれば、未知楽曲についても音響特徴から感性ベクトルを予測することができる。

これらをシステムのデータベースとして組み込めば、未知楽曲であってもアンケート調査を行うことなく楽曲検索を行うことが可能となる。

#### 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8次元のテーブルを使用する。  
(1~5-の5段階評価を-2~2に変更)

表2 感性ベクトルテーブル

	楽しさ	悲しさ	恐怖	落ち着き	怒り	不気味	明るさ	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	-1.00	0.29	-0.63	-0.21	-1.21	-0.50	-0.57	-0.29
...	...	...	...	...	...	...	...	...

### 3.2 感性ベクトル自動生成の式

感性ベクトルを  $E$ 、音響特徴ベクトルを  $A$  とすると、変換行列  $C$  と定数ベクトル  $B$  が求まり、 $E$  を導くことができる。

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

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

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

$M$  は感性ベクトルの次元数、 $N$  は音響特徴ベクトルの次元数である。ここでは、重回帰分析により変換行列  $C$  とバイアスベクトル  $B$  を求める。1つの感性ベクトル  $e_j$  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。

### 3.3 使用する音響特徴

音量やピッチの変化を使用する。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、以下の様な音響特徴を使用した。

- ・ピッチ 平均・分散
- ・パワー(対数パワー) 平均・分散

- ・動的尺度・1秒間のピークの数・平均・分散

動的尺度のケプストラム次数と分析窓幅を変えたものを追加して分析を行った。

- ・ケプストラム次数 1-16次~129-256次
- ・分析窓幅 11フレーム 33フレーム 66フレーム(フレーム周期10ms)

### 3.4 評価方法

提案した変換により適切な感性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。以下の式を用いて元のテーブルと、予測して求めたテーブルの距離を計算する。(テーブル間の距離が0ならば元のベクトルが完全に再現できている)

( $A$ =元の行列、 $B$ =変換して求めた行列、 $N$ =感性項目数、 $I$ =曲数)

$$D = \sqrt{\frac{1}{NI} \left( \sum_{i=1}^I \left( \sum_{n=1}^N (A-B)^2 \right) \right)}$$

### 3.5 感性ベクトルテーブル評価

変換式を用いてデータベースの全曲に対して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結果、0.29であった。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、元の感性ベクトルと変換して求めた感性ベクトルが比較的近いものになっているといえる。この方法で感性ベクトルがある程度導ける事が分かった。

### 4. 曲の追加

分析を行うために、元の Sound Advisor システムで使用していた86曲に245曲を追加した。感性ベクトルデータベースを作成するために、評価アンケートを実施した。

#### 4.1 使用曲

- ・効果音楽CD
  - ・KING RECORD SE SERIES -17
  - ・King Sounds & Music Effect
  - ・KINGRECORD マルチメディア音ネタシリーズ1
  - ・KING RECORD マルチメディア音ネタシリーズ2
  - ・PCを百倍楽しむ効果音
  - ・幽霊の音楽 四谷怪談 コロムビア
- ・インターネットのフリー音源

から245曲を使用した。各楽曲1分~3分程度であるが、曲頭から10秒をカットして使用する。

#### 4.2 データベース作成アンケート

被験者は 15 人、各曲を聴き、8 つの感性項目に対して 5 段階で評価を行った。20 曲を 1 グループとし、そのグループをランダムな順序で聴いて評価してもらった。

#### 4.3 アンケート結果

##### 4.3.1 曲の分布

曲の感性の内訳は以下の通りであった。各感性 0 以上 2 以下を表す曲は平均して全体の 35% だったが、怒りに関しては 1 以上は 0% で、0-1 に関しても 9% しかなかった。「怒り」を感じる曲が少ない事が分かった。

表 3 データベース楽曲の感性内訳

	楽しさ	悲しさ	恐怖	落ち着き	怒り	不気味	明るさ	爽やか
2 ~ 1	8.4(%)	14	8.2	11	0.4	12	17	6.9
1 ~ 0	20	20	13	27	9	22	18	33
0 ~ -1	24.9	29	26	42	26	23	32	33
-1 ~ -2	36.7	36	53	20	65	42	34	28

##### 4.3.2 回答の分布

回答の分散については以下の事が分かった。「楽しさ」の分散が小さい事から、「楽しさ」の感覚については皆が共有できるが「落ち着き」の分散が大きい事から「落ち着き」については人によって感じ方がまちまちである。

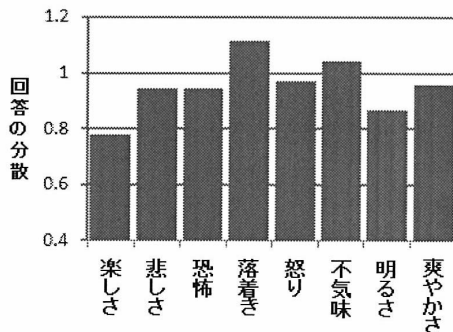


図 1 感性項目ごとの回答の分布

##### 4.3.3 感性項目について

8 つの感性項目に対して主成分分析を行ったところ、「恐怖/不気味」「楽しさ/明るさ」の回答は類似していることが分かった。

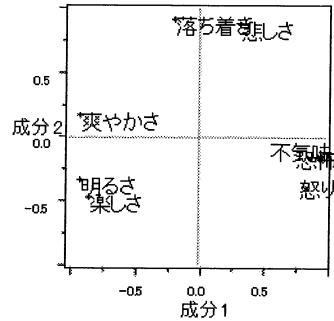


図 2 感性項目主成分分析の負荷量

相関分析を行ったところ高い相関がみられ、それぞれ「楽しさ/明るさ」「恐怖/不気味」に集約できる。

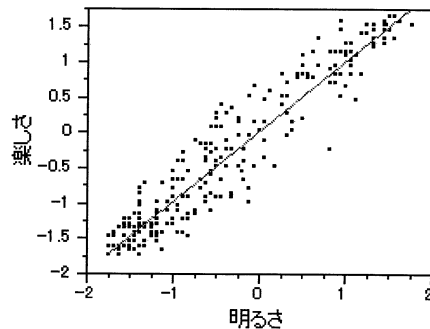


図 3 楽しさと明るさの関係  
R2 乗=0.91

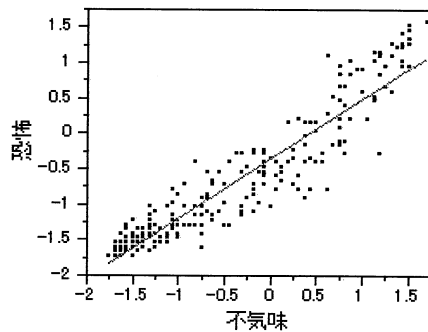


図 4 「恐怖」と「不気味」の関係  
R2 乗=0.88

## 5. 特徴量の選択

### 5.1 分析方法

追加した 245 曲を前半と後半にわけ、それぞれ学習サンプル、テストサンプルとした。学習サンプルで音響特徴から感性ベクトルテーブルへの変換行列を学習し、テストサンプルで感性ベクトルテーブルを生成してアンケート調査結果との比較を行うことにより評価する。学習サンプルとテストサンプルを入れ替えた実験も行い、平均を結果とした。

### 5.2 音響特徴の個数

現在使用している音響特徴は 31 種類である。何種類の特徴を使用するのが一番良いのかを調べた。

音響特徴が 19 種類するとき、テストサンプルでのベクトル間の距離が一番小さいことが分かった。この時ベクトル間の距離は約 0.7 である。

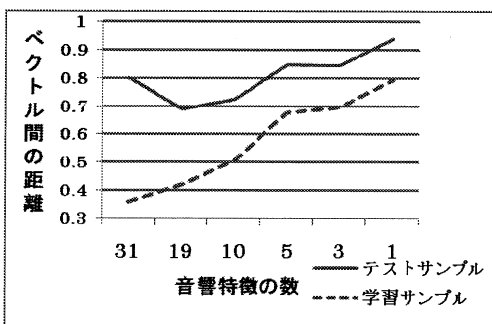


図5 音響特徴の個数とベクトル間の距離

### 5.3 音響特徴の選択

音響特徴 31 種類全てを使用した時と、音響特徴を選択したときの感性ベクトル間の距離を調べた。音響特徴の選択は、主成分分析、ステップワイズ、p 値が小さい順に 19 種類の特徴を使用、類似性の低い特徴を 19 種類使用してベクトル間の距離を比較した。(主成分分析では第 9 主成分までで寄与率が 80%を超えるため、第 9 主成分まで使用)

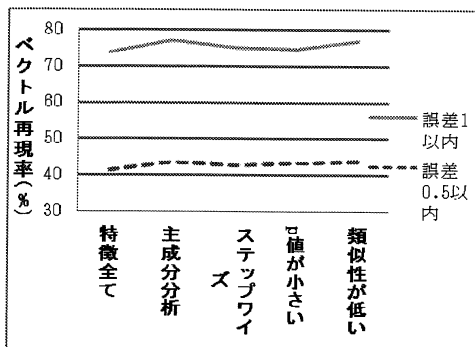


図6 音響特徴の選択方法と感性ベクトルの再現率

主成分分析を用いたものと類似性の低い特徴を用いた時、誤差が 1 以内の曲は約 77%、誤差が 0.5 以内の曲は 45% になった。以上から、誤差 1 以内を含めれば約 80% が予測できることが分かった。

### 5.4 感性項目ごとの距離

「恐怖」についてはアンケート調査結果が 2 に近いほど(「恐怖」を強く感じるほど)誤差が大きく、-2 に近いほどよく予測できている。アンケート結果が 2~1 までの誤差の平均は 1.68、-2~-1 では 0.41 であった。「恐怖」を感じる曲が全体的に少なかったためだと考えられる。

「落ち着き」と「不気味」についてはあまり良い結果が得られなかった。アンケート回答の分散が大きく、両者とも感覚が千差万別である。両者のイメージとして、テンポが遅い、又はテンポがなく(変化がない)、さらに和音の響きが暗いか明るいかで「落ち着き」又は「不気味」どちらかに感じると考えられる。「落ち着き」と「不気味」は感性が裏腹にあり、今回用いた特徴では再現しづらいといえる。

## 6. システム評価実験

以上の結果から、追加した 245 曲を学習サンプルとし、元の 86 曲について主成分分析を用いた予測と、類似性の低い特徴を使用し感性ベクトルの予測を行った。この時、ベクトル間の距離はそれぞれ 0.84、0.77 であった。

予測したベクトルをシステムに組み込み、評価実験を行った。元の感性ベクトル、主成分分析により予測したベクトル、類似性の低い特徴を使って予測したベクトルの 3 つの Sound Advisor システムを実際に使ってもらい検索結果の満足度を記入してもらった実験を行った。

使用曲ははじめのシステムで使用していた86曲である。感性項目は「楽しさ/明るさ」「悲しさ」「恐怖/不気味」「落ち着き」「怒り」「爽やかさ」の6項目である。検索アルゴリズムとして、類似度を用いた方法とユークリッド距離を用いた方法で比較を行った際、満足度は変わらないという結果が得られているため[2]、はユークリッド距離により検索を行った。

## 6.1 実験方法

被験者は15人、システムを提示する順序はランダム化した。  
 (指定した項目による検索を3回+自由な項目による検索を3回)×3システム=合計18回の検索を行ってもらった。  
 項目の指定に関しては曲の印象と数値を指定した。ネガティブイメージ(「怖い」など)とポジティブイメージ(「明るい」など)の両方が入るように設定した。  
 指定した項目は以下のものである。

表4 評価実験の指定項目

検索する曲の印象	明るさ	楽しさ	悲しさ	不気味	恐怖	落ち着き	怒り	爽やかさ
1.非常に楽しい曲	2	-2	-2	-2	-2	-2	0	
2.なんとなく不気味で怒りに満ちている曲	-2	-2	1	-2	2	-2		
3.非常に悲しく落ち着いている曲	-2	2	-2	2	-2	-2		

検索項目4から6については曲の印象、感性項目の値も自由に決定してもらった。

1回の検索で意図していた効果音が提示されなかった場合は相対検索を用いて[1]、2回までさらに検索を行ってもらった。

## 6.2 実験結果

「楽しさ」「悲しさ」の検索については予測したベクトルでの平均評価値が5段階評価でそれぞれ4.4、4.27であり、3システムとも4を超えた。また、「怖い」の検索、反する感性による検索(「楽しくて怖い」など)は予測したシステムの平均評価値が3.94、3.93、3.88、3.6であり、4に達しなかった。

満足度が5段階中4か5だったのは元の感性ベクトルでの検索が83%、予測したシステムが全検索の73%であった。  
 以上の結果により、元のベクトルが比較的再現できており、検索システムでの検索結果においても比較的満足できるといえる。

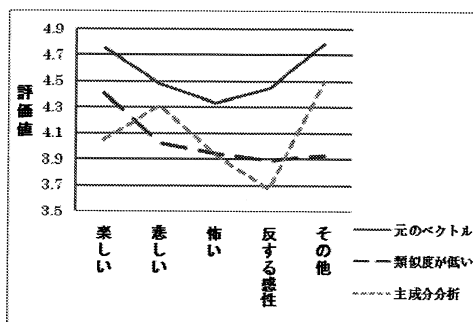


図7 3システムの満足度比較

## 7. むすび

各種の音響特徴と感性ベクトルとの重回帰分析を行うことで感性ベクトルを自動的に生成する方法と、その改良について述べた。音響特徴を19種類に選定した際に、アンケート調査と最も類似した感性ベクトルが得られることがわかった。

19種類の音響特徴から予測した感性ベクトルをシステムに組み込み、検索性能の評価実験を行った結果、予測したテーブルを使用したシステムの満足度が5段階評価で4.06であり、被験者に行ってもらった全ての検索のうち、約73%の検索について提示された曲の満足度が4以上であった。  
 本方法により、未学習の音楽データに対しても検索用の感性ベクトルをある程度自動生成できることがわかった。

## 参考文献

- [1] 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)
- [2] 上野智子, 相川清明, “感性空間の距離に基づく相対的音楽検索” 日本音響学会講演論文集, 3-5-2, pp. 925-926, (2008-03).
- [3] 上野 智子相川 清明, “音楽検索のための感性ベクトルと音響特徴量の関係の分析”, 情報処理学会研究報告, 2008-MUS-74, 2008-SLP-70, Vol. 2008, No. 12, pp. 211-216, (2008-02).
- [4] 上野 智子, 相川 清明, “感性表現による音検索のための音響特徴量の分析”, 第2回音声ドキュメント処理ワークショップ講演論文集, p. 91-96, (2008-03).
- [5] 嵯峨山 茂樹, 板倉 文忠, “音声の動的尺度に含まれる個性情報”, 日本音響学会講演論文集, pp. 589-590, 1979-06.