

## 音声入力を用いた感性特徴による 図柄パターン検索

鈴木教子<sup>†</sup> 相川清明<sup>†</sup>

本報告では、音声によるクエリ入力と感性特徴ベクトルを用いて図柄パターンを検索する方法を提案する。検索クエリと図柄パターンの持つ特徴を感性特徴ベクトルで表し、それらを照合して検索を行う。感性特徴ベクトルの比較法としては、4種類の類似性尺度算出法を提案する。本方法の特徴は音声ドキュメント処理のように一旦テキストにするのではなく、データの特徴のレベルで検索を行うことである。動作確認のために小規模な検索実験を行った。

## Voice-Input Symbolic Pattern Retrieval using Emotional Feature Vectors

YUKIKO SUZUKI<sup>†</sup> KIYOAKI AIKAWA<sup>†</sup>

This report proposes a symbolic pattern retrieval method using emotional feature vectors. Spoken queries and symbolic patterns are represented by emotional vectors. Four similarity measures were compared on a set of collected test queries. The proposed method is characterized by the search using feature vectors close to raw data. This point is different from conventional spoken document search using transcribed text.

### 1. はじめに

最近、音声ドキュメント検索が2つの観点から注目されている[1-2]。1つはコンピュータ内やネットワーク上の音声ドキュメントの爆発的な拡大である。もう1つは、声認識の新たな応用の可能性を秘めていることである。

検索対象が音声の場合には音声認識を用いてテキスト化し、テキストに対して検索を行うことができる。しかし、検索対象の中には図案のようにその物理的特徴と言語的表現が確立されていない場合があり、音声ドキュメント処理の方法が直ちに適用できない。

また、音声ドキュメント処理は図1(a)のように一旦テキストに変換を行うので、全体に雑音を重ねているデータなどでは正しいテキストに変換されにくい。さらに、テキストに変換されなければ検索することができない。一方、テキストになりにくい楽曲の検索などで、感性特徴を用いた検索が研究されてきた[3-6]。

そこで、本報告では、クエリに基づき特徴パラメータレベルで検索を行う方式を提案する。特徴パラメータレベルで検索を行う方式では、雑音に埋もれた特徴をPMC[7]のように特徴パラメータを合成して検索することが可能となることが考えられる。また、似たような特徴をもつ別の候補の追加検索や、候補のもつ特徴の度合いを指定した検索も可能になる。

本方法の特徴は認識によりテキストにしてから検索するのではなく、図1(b)のように特徴で検索するので、特徴を合成して検索する、検索閾値を下げてあいまいな特徴を再検索する、などが可能となる。

本報告では従来の音声ドキュメント処理のように認識によってテキストに変換できない図柄パターンを検索対象とし、感性特徴を特徴パラメータとする検索アルゴリズムを提案し、図案検索タスクによる動作例を紹介する。

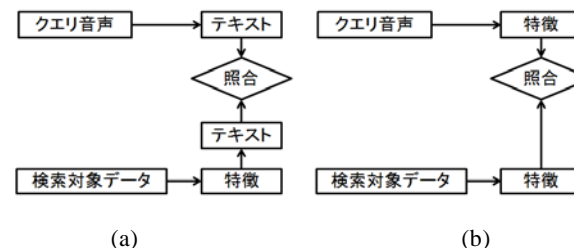


図1 音声ドキュメント処理と特徴パラメータによる検索のフローの比較

<sup>†</sup> 東京工科大学 メディア学部

Tokyo University of Technology, School of Media Science

## 2. 感性特徴

今回特徴パラメータとして用いた感性表現は、「複雑さ」「明るさ」「柔らかさ」「温かさ」「軽さ」「曖昧さ」「迫力」「勢い」の8項目である。まずは100種類の図柄パターン(図1参照)を用意して、これを被験者に見てもらい、検索クエリを集めた。

集まった58種類のクエリについて対し、上記の8種類の感性項目がどの程度当てはまるかを、表1に示す5段階で答えていただいた。その結果からそれぞれのクエリについて感性ベクトルを求めた。

また図柄パターンについてもクエリ表現と同様に、図1に示すリストを見ながら、それぞれの図柄パターンについて8種類の感性項目の当てはまる程度を5段階で答えていただいた。

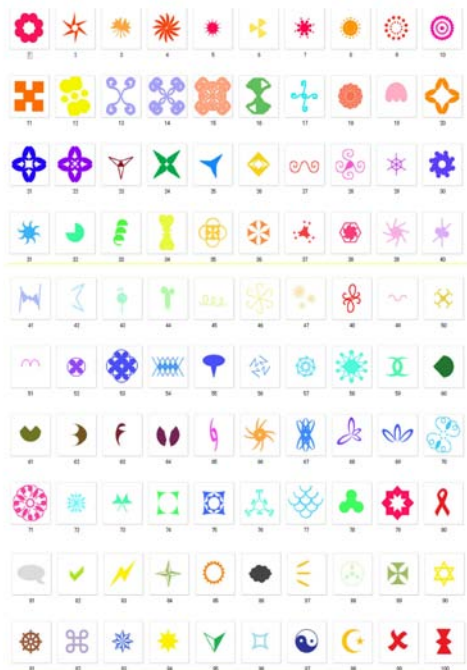


図2 図柄パターン

表1 5段階評価の内容(複雑さの例)

度合	内容
1	全く複雑ではない
2	あまり複雑ではない
3	ある程度複雑
4	かなり複雑
5	非常に複雑

## 3. 感性ベクトルの類似度

今回用いた感性ベクトルの比較の手法について、以下に簡略化した例を挙げる。

今、クエリ表現の感性ベクトルが表2、図柄パターンの感性ベクトルが表3のように求まっているとする。両者は、クエリ表現の感性ベクトルをそれぞれ5段階評価で求めたものを最大値5で割って算出した。

表2 クエリ表現とその感性ベクトルの例

クエリ中の検索語	感性表現		
	丸っこい	角ばった	複雑
丸っこい	1	0	0
角ばった	0	1	0
複雑	0	0	1
スムーズ	1	0	0
ざらざら	0	0.5	0.5
単純	0.5	0.5	0

表3 図柄パターンとその感性ベクトルの例

図柄パターン	感性表現		
	丸っこい	角ばった	複雑
□	0	1	0
⊗	0.5	0	0.5
☺	0.9	0	0.1

それぞれの感性ベクトルが求まったら、クエリを音声認識する。例えば、クエリが「スムーズで単純」である場合、「スムーズ」の感性ベクトル[1 0 0]と、「単純」の感性ベクトル[0.5 0.5 0]の和をとって

$$[1 + 0.5 \quad 0 + 0.5 \quad 0 + 0] / 2$$

というベクトルを合成して各パターンとの類似度を計算する。

もし、アンケート調査の結果例の統計分布をそのまま保存するのであれば、表4のようになる。

表4 アンケート調査結果例の頻度分布

感性表現 評価値 図柄パターン	丸っこい					角ばった					複雑				
	1	2	3	4	5	1	2	3	4	5	1	2	3	4	5
□	20	0	0	0	0	0	0	0	2	18	19	1	0	0	0
⊗	0	0	5	10	5	10	5	3	2	0	0	0	3	6	11
☺	0	0	1	3	16	19	1	0	0	0	17	3	0	0	0

それぞれを確率密度関数として扱うには回答数20で割る必要がある。

クエリが「丸くて複雑」であれば、検索ベクトルは

$$00001 \quad 10000 \quad 00001$$

この時単に15次元のベクトルのcos距離を求めると、部分空間の00001と00010、00100は比較的類似しているにも関わらず、すべて距離が0になってしまう。

そこで、以下の4種類の方法で検索ベクトルの類似度を算出する。

### 3.1 方法1

一つ目は、アンケート調査結果を標準偏差1の混合ガウス分布で表現し、距離ではなく確率を用いて類似性を評価する方法である。

n番目の図案のk番目の感性特徴の確率密度関数を

$$p_{nk}(i) \quad i=1,5$$

クエリのk番目の感性特徴の確率密度関数を

$$q_k(j) \quad j=1,5$$

とすると、n番目の図案とクエリの類似性は以下の式で表される。

各感性特徴の評価値の確率分布を混合ガウス分布で表わすときには、 $i=1-5, j=1-5$ の座標軸の各点で分散1の正規分布があるとすると、

$$d(p_n, q) = \sum_{k=1}^N \sum_{i=1}^L \sum_{j=1}^L p_{nk}(i) q_k(j) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(i-j)^2}{2}\right) \quad (1)$$

ここでNは感性特徴数、Lは感性特徴の段階数である。

本研究では

$$N=8, L=5$$

とした。

表5 アンケート結果

$p_k(i) = p(n, k, i)$	丸っこい k=1	角ばった k=2	複雑 k=3
図案 n	i=1,2,3,4,5	i=1,2,3,4,5	i=1,2,3,4,5

表6 クエリ

$q_k(i) = q(k, j)$	丸っこい k=1	角ばった k=2	複雑 k=3
	j=1,2,3,4,5	j=1,2,3,4,5	j=1,2,3,4,5

### 3.2 方法2

こちらは混合ガウス分布ではなくアンケート調査の結果の頻度分布を確率分布とし、クエリも複数の検索語から導かれる感性特徴の確率分布が得られているとし、分布間の平均2乗距離の期待値を求める方法である。

非常に単純な例としては、アンケート調査結果がある感性表現で00001、クエリの感性表現が00100であれば、1の位置が2離れているので、距離は2の2乗。これに5段階のレベルそれぞれに確率が付いているから、期待値はすべてのレベルの組み合わせの2乗距離のアンケート調査結果とクエリの各感性表現レベルの確率が乗算されたものとなる。

$p_n(x)$   $p_n(i)$  n番目の図案のアンケート回答の確率分布 (評価値iの)

$q(x)$   $q(j)$  クエリの特徴分布

pの座標 (レベルまたは評価値) を x、 qの座標 (レベルまたは評価値) を y、

n番目の図案のk番目の感性要素の確率分布  $p_{nk}(x)$  における平均値を  $\mu_{nk}$ 、分散を  $\sigma_{nk}$ 、クエリのk番目の感性要素の確率分布  $q_k(x)$  における平均値を  $\nu_k$ 、分散を  $\rho_k$  とすると、

$$\begin{aligned}
 d_{nk}(y) &= \int (x-y)^2 p_{nk}(x) dx \\
 &= \int ((x-\mu_{nk})-(y-\mu_{nk}))^2 p_{nk}(x) dx \\
 &= \int \left\{ (x-\mu_{nk})^2 - 2(x-\mu_{nk})(y-\mu_{nk}) + (y-\mu_{nk})^2 \right\} p_{nk}(x) dx \\
 &= \int (x-\mu_{nk})^2 p_{nk}(x) dx - 2(y-\mu_{nk}) \int (x-\mu_{nk}) p_{nk}(x) dx + (y-\mu_{nk})^2 \int p_{nk}(x) dx \\
 &= \sigma_{nk}^2 - 2(y-\mu_{nk}) \left\{ \int x p_{nk}(x) dx - \mu_{nk} \int p_{nk}(x) dx \right\} + (y-\mu_{nk})^2 \\
 &= \sigma_{nk}^2 - 2(y-\mu_{nk}) \{ \mu - \mu_{nk} \} + (y-\mu_{nk})^2 \\
 &= \sigma_{nk}^2 + (y-\mu_{nk})^2
 \end{aligned} \tag{2}$$

$$\begin{aligned}
 d_{nk}(y) &= \sigma_{nk}^2 + (y-\mu_{nk})^2 \\
 d(p_{nk}, q_k) &= \int q_k(y) d_{nk}(y) dy \\
 &= \int q_k(y) \{ \sigma_{nk}^2 + (y-\mu_{nk})^2 \} dy \\
 &= \sigma_{nk}^2 \int q_k(y) dy + \int q_k(y) (y-\mu_{nk})^2 dy \\
 &= \sigma_{nk}^2 + \int q_k(y) \{ (y-\nu_k) - (\mu_{nk} - \nu_k) \}^2 dy \\
 &= \sigma_{nk}^2 + \int q_k(y) (y-\nu_k)^2 dy - 2(\mu_{nk} - \nu_k) \int q_k(y) (y-\nu_k) dy + (\mu_{nk} - \nu_k)^2 \int q_k(y) dy \\
 &= \sigma_{nk}^2 + \rho_k^2 + (\mu_{nk} - \nu_k)^2
 \end{aligned} \tag{3}$$

のように求まる。

離散的な 5 段階の評価レベルを用いた計算では、 $n$  番目の図案の  $k$  番目の感性要素の離散分布を  $p_{nk}(i)$ 、クエリの  $k$  番目の感性要素の確率分布を  $q_k(i)$  とすると、計算式は

$$\begin{aligned}
 \mu_{nk} &= \sum_{i=1}^L i p_{nk}(i) \\
 \nu_k &= L \sum_{i=1}^5 i q(i) \\
 \sigma_{nk}^2 &= \sum_{i=1}^L i^2 p_{nk}(i) - \mu_{nk}^2 \\
 \rho^2 &= \sum_{i=1}^L i^2 q(i) - \nu_k^2 \\
 d(p_{nk}, q_k) &= \sigma_{nk}^2 + \rho_k^2 + (\mu_{nk} - \nu_k)^2
 \end{aligned} \tag{4}$$

$$\begin{aligned}
 d_{nk} &= \sum_{i=1}^L \sum_{j=1}^L p_{nk}(i) q_k(j) (i-j)^2 \\
 &= \sum_{i=1}^L p_{nk}(i) \sum_{j=1}^L q_k(j) (i^2 - 2ij + j^2) \\
 &= \sum_{i=1}^L p_{nk}(i) \left\{ i^2 \sum_{j=1}^L q_k(j) - 2i \sum_{j=1}^L q_k(j) j + \sum_{j=1}^L q_k(j) j^2 \right\} \\
 &= \sum_{i=1}^L p_{nk}(i) \{ i^2 - 2i\nu_k + \rho_k^2 + \nu_k^2 \} \\
 &= \sum_{i=1}^L i^2 p_{nk}(i) - 2\nu_k \sum_{i=1}^L i p_{nk}(i) + (\rho^2 + \nu^2) \sum_{i=1}^L p_{nk}(i) \\
 &= \sigma_{nk}^2 + \mu_{nk}^2 - 2\nu_k \mu_{nk} + (\rho_k^2 + \nu_k^2) \\
 &= \sigma_{nk}^2 + \rho_k^2 + (\mu_{nk} - \nu_k)^2
 \end{aligned} \tag{5}$$

となり同じ式となる。

全体としての平均距離は

$$d_n = \sum_{k=1}^N d_{nk} \quad (6)$$

により求められる。

### 3.3 方法 3, 4

これらは確率分布情報を使わず平均ベクトルとの類似度や距離を用いる方法である。

$r_{nk}$  は  $n$  番目の図案のアンケート結果の  $k$  番目の感性要素の全調査者の 5 段階評価の平均値、 $s_k$  はクエリの  $k$  番目の感性要素の複数検索語の 5 段階評価の平均値である。

#### 3.3.1 方法 3

この方法では、 $\cos$  距離  $c_n$  を用いて類似性を評価する。ベクトル  $a$  と  $b$  の内積を  $\text{dot}(a,b)$  で記述すると

$$\begin{aligned} d(a,b) &= \frac{\text{dot}(a,b)}{\sqrt{\text{dot}(a,a)\text{dot}(b,b)}} \\ &= \frac{\sum_{k=1}^N a_k b_k}{\sqrt{\sum_{k=1}^N a_k^2 \sum_{k=1}^N b_k^2}} \end{aligned} \quad (7)$$

となるから、実際の演算は

$$\begin{aligned} c_n &= \frac{r_n \cdot s}{\sqrt{r_n \cdot r_n \sqrt{s \cdot s}}} \\ &= \frac{\sum_{k=1}^N r_{nk} s_k}{\sqrt{\sum_{k=1}^N r_{nk} r_{nk} \sum_{k=1}^N s_k s_k}} \end{aligned} \quad (8)$$

#### 3.3.2 方法 4

この方法ではユークリッド距離を用いる。この場合、

$$d(a,b) = \sum_{k=1}^N (a_k - b_k)^2 \quad (9)$$

となるから、実際の演算は

$$v_n = \sum_{k=1}^N (r_{nk} - s_k)^2 \quad (10)$$

となる。

## 4. 実験

### 4.1 実験方法

#### 4.1.1

図 2 と回答用紙を被験者に配布し、100 種類の図柄パターンと 58 語の検索クエリ表現について、それぞれに対して感じた感性を 8 感性項目 5 段階で評価してもらうことにより、感性ベクトル値を得た。クエリ表現と図柄パターンのそれぞれに対して得られた感性ベクトル値を 3. の計算方法で比較し、類似度が最も高いものを検索結果とした。

検索によって各クエリ表現に対しどの図柄パターンが検出されるかを予め 3 パターンずつ予測しておき、その予測と検出結果を比較した。また、検出された図柄パターンとクエリ表現を照合し、適切な検索が行われた項目数を調べた。

#### 4.1.2

2. で述べた事前アンケートと同様に、被験者に図 1 を見てもらい検索クエリを新たに集めた。そして、2. と 5.1.2. で抽出されたクエリ表現の中で、音声認識によって正しく認識される語数を調査した。

その手法としては、2. で得られた 58 語と本アンケートで得られた 92 語、計 150 語の検索クエリ中、何語が Julius の辞書内に掲載されているかを調べた。

## 4.2 結果分析

### 4.2.1 クエリ表現と検出された図柄パターンの照合

各クエリで検出される図柄パターンの事前予測と実際の検出結果を照合したところ、一致した項目数と一致した図柄パターンの種類数は、表8の『事前予測との比較』の欄に示す結果となった。また、事前予測と検出結果が一致しなかった項目において、検出された図柄パターンと検索クエリを照合し、適切な検索が行われたと考えられる項目数と、そこに含まれる図柄パターンの種類数については、同表の『適切と思われる検出結果』の欄に示す通りであった。

上記の一致項目数と適切項目数を合計し、方法ごとの総項目数 58 で割ることで、適合率を算出した。また、各方法で事前予測と一致した図柄パターンと適切に検出された図柄パターンの総種類数を計算し、それを図柄パターンの総数 100 で割ることで、再現率を算出した。そして求めた適合率と再現率から、F 値を算出した。

表 8 検索結果

		方法 1	方法 2	方法 3	方法 4
事前予測との比較	一致項目数	3	8	13	11
	一致種類数	3	7	13	10
適切と思われる検出結果	適切項目数	16	33	31	33
	適切種類数	6	17	23	23
合計	正解項目数	19	41	44	44
	正解種類数	7	21	28	27
	適合率	0.33	0.71	0.76	0.76
	再現率	0.07	0.21	0.28	0.27
	F 値	0.12	0.32	0.41	0.40

以上の結果より、感性ベクトル比較法は、方法3、方法4、方法2、方法1の順に検出精度が高いことがわかった。また、今回の実験では検索要求に偏りが出てしまった結果、再現率が低くなってしまったと考えられる。

### 4.2.2 正しく認識されるクエリ表現数

Julius の認識単語辞書の中に含まれており、そのままの表記で音声認識される検索クエリの数は、表9に示す通りであった。

表 9 辞書に掲載されていた検索クエリ語数

	2.で検出されたクエリ	5.1.2で検出されたクエリ	合計
辞書掲載語数	36	36	72
辞書掲載率	0.62	0.39	0.48

検索クエリの中には独特な表現が目立ち、音声認識することが難しい単語も多かった。

## 5. おわりに

本報告では、音声認識によるクエリ入力を用いた、図柄パターンの感性特徴ベクトルによる検索方法を述べた。感性特徴ベクトルの比較法として、確率、距離、cos 距離、ユークリッド距離を用いた4種類の類似性尺度算出法を提案した。それらの動作確認のために小規模な検索タスクで実験をし、その結果について提示した。今回の実験でもっとも精度の高い類似性尺度算出法は、cos 距離を用いたものであった。

## 6. 謝辞

本研究には音声認識に Julius4.1.3 を使わせていただいた。ここに深く感謝する。

## 参考文献

- [1] 秋葉 友良、西崎 博光、相川 清明、河原 達也、松井 知子、伊藤 慶明、胡 新輝、中川 聖一、南条 浩輝、山下 洋一、“音声検索語検出と音声ドキュメント検索の評価枠組みの設計”，Vol-2010-SLP-84, No.18, pp. 1-5, (2010-12)
- [2] 秋葉 友良，“音声ドキュメント検索の現状と課題”，Vol.2010-SLP-82, No.10, pp. 1-8, (2010-07)
- [3] 相川 清明、矢島 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”，2007-SLP-65, pp.5-10, (2007-02)
- [4] “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”
- [5] 上野 智子、相川 清明，“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”，2008-SLP-79, pp.211-216, (2008-02)
- [6] 佐藤 真梨、相川 清明，“受信感性情報量の計測とその解釈条件による違い”，2009-SLP-75, pp.1-6, (2009-02)
- [7] Gales, M.J.F., Young, S.J., "Robust continuous speech recognition using parallel model combination", IEEE Transactions, Speech and Audio Processing, Vol.4. No.5, 352-359, 1996-09