

# 音声入力による図柄パターン検索における 未知語処理

鈴木教子<sup>†</sup> 相川清明<sup>†</sup>

音声入力による図柄パターン検索では、擬態語を含む未知語がクエリとして入力されることが多い。しかし、これらの未知語は誤認識されたとしても類似した音素を含んでいる。そこで、誤認識単語から既知語を推定し、それを用いて図柄パターン検索を行う方法を提案する。動作確認のために小規模な検索実験を行い、未知語と既知語の場合で検索精度の差はあまりみられないことがわかった。

## OOV Processing for Voice-Input Symbolic Pattern Retrieval

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

This paper proposes an OOV processing method for voice-input symbolic pattern retrieval. The queries for pattern retrieval include various mimetic words which are not included in the ASR vocabulary. However, misrecognized words include fragmental phoneme sequences which reflect acoustic features in the query words. This paper describes a method for estimating similar vocabulary words close to the OOV word using DP matching of phoneme sequences. Experimental results indicated there is no noticeable difference between the retrieval accuracy of known words and that of OOV words.

## 1. はじめに

前回の報告では、音声入力によるクエリに基づき特徴パラメータレベルで図柄パターンの検索を行う方法を提案した[1]。前報告でのアンケートで得られたクエリ表現において、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ものの数は全 150 語中 72 語、48%であった。半数以上のクエリは、通常の音声認識では正しく認識されず、これらの未知語を処理しなければ、図柄パターンの効果的な検索はできない。

これまでの音声ドキュメント等における未知語処理の手法としては、サブワード単位での認識や形態素解析を用いた認識、文脈に基づく認識などがあるが[2-8]、図柄パターンや効果音等は擬音語や擬態語で表現されることが多い。さらに、その表現に用いる擬音語・擬態語は人によってまちまちであり、また独特なものである。それらの語を全て音声認識単語辞書に掲載することは非常に困難である。

しかし似た音響特徴をもつ擬音語・擬態語は、それらの表すイメージも似ていることが多い。そのため、入力されたクエリと類似性の高い音響特徴をもつクエリで図柄パターンの検索を行えば、検出される図柄パターンも類似性が高く、求めるものが検出できると考えられる。そこで、本報告ではクエリの音素列情報で登録単語を推定することによる未知語処理の手法を提案する。

## 2. 前回方法

### 2.1 感性ベクトル

前回検索を行う上での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 項目であった。各図柄パターンと、アンケート調査で得たクエリについてアンケートを行い、それぞれ表 1 に示す 5 段階評価で感性ベクトルを求めた。得られた感性ベクトルを、確率分布間距離を用いて比較し、検索結果を得た。

<sup>†</sup> 東京工科大学 メディア学部

Tokyo University of Technology, School of Media Science



図1 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

## 2.2 クエリと図柄パターンの感性特徴ベクトル比較方法

今回図柄パターンとクエリの感性ベクトルを比較するのに用いた類似度は、以前の実験における正解パターンの検出で、最もF値が高かった確率分布間距離である。今回は各クエリにおいて図柄パターンとの距離の最も大きいものの値を1とし、上位0.45に属するパターンを検索結果として抽出した。

表2に示す例のような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。

表2 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

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。

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  $p_{nk}(i)$ 、 $p_{nk}(i)$  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を  $q_k(i)$ 、 $q_k(i)$  における平均値を  $\nu_k$ 、分散を  $\rho_k$  とすると、計算式は

$$\mu_{nk} = \sum_{i=1}^L i p_{nk}(i)$$

$$\nu_k = L \sum_{i=1}^L i q_k(i)$$

$$\sigma_{nk}^2 = \sum_{i=1}^L i^2 p_{nk}(i) - \mu_{nk}^2$$

$$\rho_k^2 = \sum_{i=1}^L i^2 q_k(i) - \nu_k^2$$

(1)

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

(2)

となる。全体としての平均距離は

$$d_n = \sum_{k=1}^N d_{nk}$$

(3)

により求められる。なお、ここで、 $N$  は感性要素数で 8、 $L$  はレベル数で 5 である。

### 3. クエリ認識方法

この方法では検索語そのものを用いて図柄パターンを検索を行うのではなく、検索語を感性ベクトルに変換し、それを用いて検索を行う。従って、クエリが未知語であっても、類似した感性ベクトルが生成できれば、目的の図柄パターンを検索できると考えられる。

前回のアンケート調査で得られた 150 語のクエリを登録クエリとする。音声入力したクエリに対して、まず Julius を用いて音声認識を行い、認識結果内の phseq 文を用いて音素列に変換した。そして入力されたクエリと登録クエリの音素列を照合し、最も音素列の距離が近いものを認識結果とした。登録語を導いた後に感性ベクトルを得て、それを用いて同じ感性ベクトル表現をされた図柄パターンを検索する。その際、第 1 章で述べた類似を用いるというところまで一連の検索アルゴリズムとして記述する。

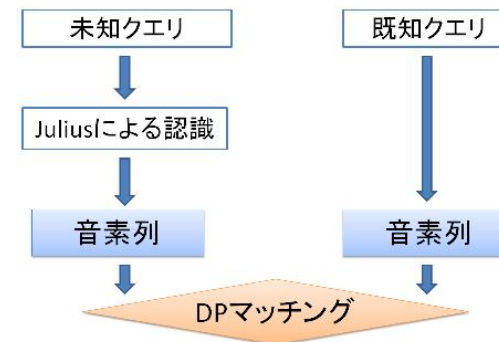


図 2 クエリ認識方法のアルゴリズム

音素列の比較は、DP マッチングによって行う。同じ母音同士、同じ子音同士である場合は距離 0、異なる分類の子音同士、または母音と子音、異なる母音同士の場合は距離 1 とした。また、似た音響特徴をもつもの同士でグループを作成し、同じグループに属するもの同士の場合は距離 0.5 とした。その分類は表 3 の通りである。DP パスは図 3 に示す最も基本的なものである。

表 3 音素分類表

1	a
2	i,y
3	u,w
4	e
5	o
6	p, t, k, q
7	s, h
8	m, n
9	b, d, g, z, r



図 3 DP パス

## 4. 実験

### 4.1 実験方法

#### 4.1.1 クエリの未知語処理結果の評価

前回の実験で得られた 150 語のクエリを音声入力して未知語処理を行い、入力されたクエリと未知語処理結果がどの程度一致するのかを調べた。各クエリについて 5 回ずつ音声入力を行い、一致する割合を調べた。以下、入力されたクエリと未知語処理結果が一致するクエリを「一致クエリ」とする。

また一致しなかった場合、その未知語処理結果が入力したクエリと類似した音響特徴をもち、その語のもつイメージも似ているものについては、これを「類似クエリ」としてその数を調べた。1 度でも一致クエリまたは類似クエリとなったものは、そのクエリを検索において有効であるものとし、その有効クエリの数と割合を調べた。

#### 4.1.2 検出されたパターンに含まれる最適パターン数・適切パターン数の分析

前回の実験で得られた 92 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 段階で被験者に評価してもらうことにより、感性ベクトル値を得た。残りの 58 語の検索クエリ表現と 100 種類の図柄パターンについての感性ベクトル値は、前回の実験で得られたものを使用した。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、類似度の上位 45% 以内に入るものを検索結果とした。

また、各クエリ表現に対し最もふさわしい図柄パターン「最適パターン」と、それらの次にふさわしい図柄パターン「適切パターン」を設定した。最適パターンは各クエリにつきパターンずつ設定し、適切パターンについてはパターン数を制限せずに設定した。

予めクエリを読み上げた音声を録音しておき、その音声ファイルに対し音声認識・未知語処理を行うことで検索クエリを得た。そしてその検索クエリから得られた検索結果と、設定された「最適パターン」「適切パターン」を比較し、検索結果内に含まれる「最適パターン」と「適切パターン」の数を調べた。用いたクエリと図柄パターンは前回のアンケート時に得られたものであり、被験者数は前回実験時と同様に 20 名である。

### 4.2 結果分析

#### 4.2.1 クエリの未知語処理結果

音声入力したクエリの数と、1~5 回目の実験において入力したクエリと未知語処理結果が一致したクエリの数、処理結果が類似するクエリとなった数、またそれぞれの割合は、既知語、未知語、全体のそれぞれにおいて表 4 のようになった。

表4 クエリ未知語処理結果

		1回目	2回目	3回目	4回目	5回目	計
既知語	総数	72	72	72	72	72	360
	一致クエリ数	66	62	62	65	61	316
	一致クエリ率	0.92	0.86	0.86	0.90	0.85	0.88
	類似クエリ数	0	0	0	0	0	0
	類似クエリ率	0	0	0	0	0	0
未知語	総数	78	78	78	78	78	390
	一致クエリ数	56	51	62	44	40	253
	一致クエリ率	0.72	0.65	0.79	0.56	0.51	0.65
	類似クエリ数	12	13	11	11	12	59
	類似クエリ率	0.15	0.17	0.14	0.14	0.15	0.15
全体	総数	150	150	150	150	150	750
	一致クエリ数	122	113	124	109	101	569
	一致クエリ率	0.81	0.75	0.83	0.73	0.67	0.76
	類似クエリ数	12	13	11	11	12	59
	類似クエリ率	0.08	0.09	0.07	0.07	0.08	0.08

既知語と未知語において、1度でも一致クエリとなったクエリの種類数と、1度でも類似クエリとなったクエリの種類数は表5の通りであった。またそれらのクエリを検索において有効であるクエリとする場合、その有効クエリの数とその割合は同表のようになった。ただし未知語の有効クエリ数については、一致クエリとなることができたものと類似クエリとなることができたものが一部重なっているため、一致クエリ種類数と類似クエリ種類数の合計と、有効クエリ数は一致していない。

表5 クエリ未知語処理結果2

	総数	一致クエリ種類数	類似クエリ種類数	有効クエリ数	有効クエリ率
既知語	72	67	0	67	0.93
未知語	78	59	20	71	0.91
合計	150	126	20	138	0.92

入力したクエリと未知語処理結果クエリが類似した例としては表のようなものがあった。

表6 類似クエリ例

入力クエリ	未知語処理結果クエリ
terereN	tere:N
kiraraN	kira:N
kiraN	kirakira
chiku	chikuchiku
pika	pikapika
gunya	gunyagunya
bishi	pishi
hyoro	hyorohyoro

大多数のクエリが有効クエリとなっており、この未知語処理方法は図柄パターンの検索において、非常に有効であるといえる。

#### 4.2.2 検出性能の評価

入力されたクエリに対して設定された「最適パターン」「適切パターン」を検出することのできたクエリ数とその割合は、既知語の場合と未知語の場合でそれぞれ表7の通りであった。

表7 予め設定されたパターンを検出することのできたクエリ数

	総クエリ数	最適パターン 検出クエリ数	最適パターン 検出クエリ 割合	適切パターン 検出クエリ数	適切パターン 検出クエリ 割合
全体	150	30	0.20	77	0.51
未知語	78	13	0.17	37	0.47
既知語	72	17	0.24	40	0.56

最適パターンを検出することのできたクエリの割合、適切パターンを検出することのできたクエリの割合は、どちらも未知語と既知語の間で大きな差は見られなかった。また適切パターン数に最適パターン数を加えたものを「正解数」とし、事前に設定した正解数、検索で得られた回答数、検出された正解項目数、再現率、適合率、F値を求めたところ、表8の通りとなった。

表8 予め設定されたパターンと検出されたパターンの比較

	総クエリ数	予め設定 した正解数	回答数	検出された 正解項目数	再現率	適合率	F値
全体	150	2339	1447	318	0.15	0.24	0.18
未知語	78	1143	679	128	0.12	0.21	0.15
既知語	72	1196	768	190	0.17	0.27	0.21

F値を比較しても既知語と未知語の間にはそれほど大きな値の差はみられなかった。よって表7、表8より本未知語処理法が効果を有するといえる。

## 5. おわりに

本報告では、音声入力による図柄パターン検索において、未知語によるクエリに対し認識結果の音素列と既知語の音素列からDPマッチングを行い、最も類似したクエリに変換して未知語処理を行う手法を提案した。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行い、その結果について提示した。既知語と未知語の間には大きな検索精度の差がみられず、本手法は音声入力による図柄パターン検索における未知語処理に効果を有するといえる。

## 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]鈴木 教子、相川 清明，“音声入力を用いた感性特徴による図柄パターン検索”，2010-SLP-85, No.4, pp. 1-6, (2011-02)
- [2]小野寺 悠二、伊藤慶明、小嶋 和徳、石亀 昌明、田中 和世、李時 旭，“複数のサブワード・言語モデルを用いた音声中の検索語検出の高精度化”，SDPWS2010-14, pp. 1-7, (2010-02)
- [3]高橋将史、藤井 康寿、山本 一公、中川 聖一，“音声ドキュメントに対する頑健な検索方法の検討”，日本音響学会講演論文集, 3-Q-35, pp.285-288, (2009-03)
- [4]嶋田 和孝、遠藤 勉，“対話型ロボットのための複数の音声認識器を利用した発話理解”，第60回 人工知能学会 言語・音声理解と対話処理研究会, SIG-SLUD-B002-06, pp.27-30, (2010-10).
- [5]笹野 遼平、黒橋 禎夫，“形態素解析における連濁および反復形オノマトペの自動認識”，言語処理学会 第13回年次大会, pp.819-822 (2007-03)
- [6]鍛冶 伸裕、喜連川 優，“文脈にもとづく未知語獲得における識別モデルの適用”，言語処理学会 第15回全国大会, (2009-03)
- [7]福岡 知隆、税田 竜一、久保村 千明、服部 峻、亀田 弘之，“文の類似性を用いた未知語処理手法の提案とそれに基づく円滑な対話応答システムの作成”，情報処理学会創立50周年記念(第72回)全国大会 6X-2, pp.619-620, (2010-03)
- [8]吉野 幸一郎、森 信介、河原 達也，“述語項の類似度に基づく情報推薦を行う音声対話システム”，2011-SLP-87, no.11, pp.1-6, (2011-07)