

## 統計的手法に基づいた画像・音声情報からの概念獲得

長島 徹† 荒井 秀一†

† 武蔵工業大学

〒158-8557 東京都世田谷区玉堤 1-28-1

E-mail: †tohru@la.cs.musashi-tech.ac.jp

あらまし これまでのメディア認識・理解の研究の多くは、シンボルを用いてより複雑なシンボルの学習、シンボル間の関係学習などを行っていた。しかし、実環境において計算機が人間の提示する情報を理解するためには、言葉の意味そのものを前もって獲得している必要がある。そこで、我々はシンボリック処理を用いずに画像・音声刺激という実メディアから、言葉の意味そのものである言語シンボル概念を獲得する枠組を提案してきた。しかし、画像・音声といった異なったメディア情報を扱うため、それぞれの認識結果の尺度が異なり、直接的な比較を行うためには、何らかの変換が必要となる。そこで、本研究では画像・音声メディアの認識結果を統計的な手法を用いて、保持している知識と入力情報の適合確率として表現する手法を提案し、適合確率を用いたモデル化を行う。

キーワード 概念獲得, 言語概念, 学習, 人工知能

## The Concept of Linguistic Symbol acquisition from Image and Speech based on statistical technique

Tohru NAGASHIMA† and Shuichi ARAI†

† Musashi Institute of Technology,

Tamazutsumi 1-28-1, Setagaya-ku, Tokyo, 158-8557 Japan

E-mail: †tohru@la.cs.musashi-tech.ac.jp

**Abstract** Former research of media recognition and an understanding learned the complicated symbol and The relation between symbols using symbol. However, the computer needs to acquire sense of language for understands information which man presents in the real world. So, we proposed a framework that acquires the concept of linguistic symbol from image and voice media not using symbolic processing. It is necessary to make the measure of the result of recognition of some media the same for comparison. In this paper, we propose a method to translate the probability of conformity of knowledge and input based on statistical technique from the image and voice media and a framework using it.

**Key words** Concept acquisition, Concept of Linguistic Symbol, Study, Artificial Intelligence

### 1. ま え が き

計算機に画像・音声などのメディアを認識・理解させる研究が数多く行われている。計算機に認識・理解を行わせる際には、必ず認識対象に関する知識が必要となるが、どのような知識を用意すべきかはそのタスク毎に異なる。一般にこれらの知識は、様々な目的に対応できる普遍的なものが望まれるがそういった知識の構築は非常に困難であり、実際のパターン理解に関する研究では、対象に依存した知識を用いて成功している場合が多い。しかし、専門家が対象・目的に応じた知識を用意しなければならず、認識対象の追加やタスクの変更を行うには、その都度、

知識の追加や変更が必要であった。

このような問題を解決するには、計算機自身が事例より自動的に学習し、知識を生成・変更できる事が望まれる。さらに、柔軟な認識・理解を実現するためには、対象依存の知識を排除し、学習により対象に関する知識獲得ができることが必要である。またタスクに応じて、獲得した知識の変更・生成を行って知識の再利用が行えるような、抽象度の高い知識を獲得することも必要である。

そこで、これまでに画像・音声の実メディアを用いて言語シンボル概念を獲得するための枠組を提案してきた [1] [2] [3]。概念とは言語シンボルが内包する“シンボルの意味”そのもので

ある。明示的な知識を与えずに、計算機自身が与えられたメディア情報から自身で学習し、概念の獲得を行うことでシンボルを用いない認識が可能である。さらに画像・音声のメディア情報を抽象化して保持することで、再利用性のある知識を構築している。本枠組では、同時に入力された画像・音声メディア間に何らかの関係があるとし、メディア情報のみでなく、その関係を含めて概念を定義している。つまり概念の認識において、画像・音声メディア毎の入力と知識の比較結果を直接的に比較する必要がある。しかしながら、それぞれのメディアにおける処理は独立に行われるため比較結果の尺度が異なり、直接比較のためには尺度を統一化させる必要がある。

そこで本研究では、メディア毎の入力情報と知識との比較結果の尺度を統計的な手法に基づいて統一化すると共に、その尺度を用いた言語シンボルの概念獲得のモデル化を行う。

## 2. 言語シンボル概念

本研究において獲得している言語シンボル概念について、図1の概略図を基に説明する。

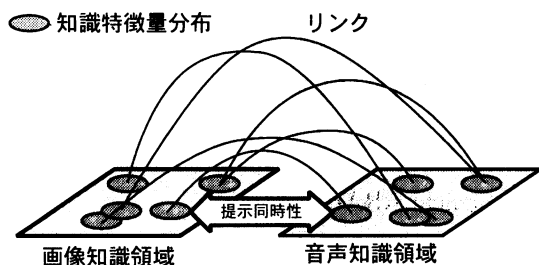


図1 提示同時性による言語シンボル概念の概略図

Fig. 1 Outline figure of the concept of linguistic symbol by time synchronization

一般に概念とは、“経験される多くの事物に共通の内容を取り出し(抽象)、個々の事物のみに属する偶然的な性質を捨てる(捨象)ことによって形成されるもの”である[5]。言語シンボルは、抽象化した事物の概念に単にラベルとして振っているにすぎない。例えば、英語の言語シンボル“Dog”と日本語の言語シンボル“犬”は動物の犬を指しているとする。この場合、“動物の犬”という事物そのものが言語シンボル概念であると言える。つまり言語シンボル概念を獲得することができれば、“動物の犬”という事物のものを理解できることになるため、言語シンボルに依存すること無く、“動物の犬”は認識が可能なのである。

このような言語シンボル概念を獲得するために、本枠組ではまず、外部から入力された画像・音声情報を各々抽象化し、特徴量の分布で表現する。

複数のメディアを用いている理由としては、単一のメディアからでは言語シンボル概念の特定ができないためである。一般的に一つのメディアには複数の概念が内包されているため、複数のメディアを用いてその関係に概念を関連付けることで表現したい概念を特定する。

そこで本研究では、同時に与えられた刺激には何らかの関係

があるという提示同時性を利用することにより、図1のように画像知識と音声知識との間にリンクを張り、そのリンクも含めた画像・音声情報をまとめて言語シンボル概念と定義する。

複数のメディア情報を用いることで、言語シンボル概念を特定している事を既に述べたが、そもそも実メディア情報には、言語シンボルの意味を持たない。例えば、リンゴの画像をメディア情報として提示したとする。それは、色の“赤”という概念なのか、それとも形状の“丸い”という概念なのかは判断できない。それは、教示を行う上でそういった概念を意図して伝達することができるにすぎず、メディア情報はあくまで概念を伝達するための手段であり、それそのもので言語シンボル概念を表現することはできないのである。

そこで、画像メディアだけでなく音声メディアを用いて、双方のメディアに関連を持たせることで初めてシンボライズが可能となる。つまり、画像・音声のメディア情報とその関係を示すリンクを、提示したメディアで表現可能な言語シンボル概念が複数あった場合、その中から概念の特定ができるだけでなく、メディア情報を単独に扱ったのでは、言語シンボル概念の表現そのものが不可能なのである。

画像・音声の入力情報を抽象化した画像・音声情報表現についてそれぞれ説明する。

### 2.1 画像情報の抽象化表現

本研究では、線画像を画像メディアの入力刺激とした。人間は何らかの物体を目にしたとき、大雑把な輪郭を認識し、明確に認識できない場合には徐々に細部の詳細な部分情報を用いた認識へと移るといったトップダウン的な処理を行っている。これに基づき、本研究でも画像情報を図2に示す階層的な表現を用いて記述した。

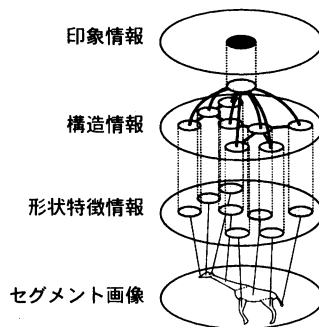


図2 画像情報の階層構造

Fig. 2 Layered structure of information of image

- **印象情報** 物体の輪郭から、輪郭線の変化を HMM を作成し、物体全体の概形を表現する。
- **構造情報** 物体を部分領域に分割し、部分領域間関係によって構造を表現する。部分領域は領域の分割線を教示することによって学習し、複数の分割線を教示することによって物体の詳細な知識の記述を可能としている。
- **形状特徴情報** 部分領域毎に形状的な特徴量(面積・骨

格長・幅・周囲長・最外殻距離)を用いて物体の詳細な形状を表現する。

- **セグメント画像** 入力画像をセグメント化する事によって作成し、上記の各情報の生成に用いる。

### 2.2 音声情報の抽象化表現

本研究では単語発話音声を音声メディアの入力刺激として、その抽象的な表現形態の知識をHMMによって表現する。具体的には、音声刺激から特徴量(LPCケプストラム16次元、パワー、ピッチ周波数、有声無声音の度合)を算出し、HMMを作成することで概念を構築する。

## 3. 認識と学習

認識は、図3のように段階的に、メディア毎の入力と知識の比較、さらにその比較結果を用いた総合的な認識を行う。総合的な言語シンボルの認識では、メディア毎の認識結果を統合し、言語シンボル概念そのものについての認識を行う。メディア毎の認識において未知であると判断された入力、新たな知識として学習する。また、新規の知識学習だけでなく、入力がある知識に認識された場合には、その知識に強化学習を行う。

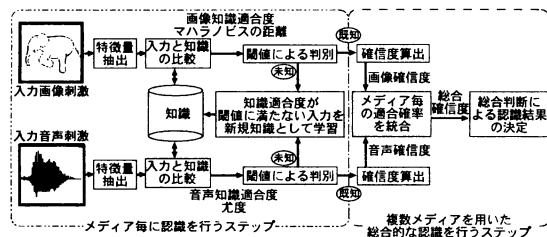


図3 認識・学習の流れ

Fig. 3 The flow of recognition and study

## 4. 認識尺度統一化手法の提案

メディア毎の認識の際に得られる知識と入力の類似度を“知識適合度”と呼ぶことにする。画像・音声の各情報表現が異なるように、各メディア毎の認識もまた独立に行われる。よってメディア毎に得られる知識適合度の尺度は全く異なったものとなる。本研究では、言語シンボル概念を複数のメディアとその関係を示すリンクによって表現していることを既に述べた。各メディアの認識の後に言語シンボル概念の認識が行われるが、各メディア毎に得られた知識適合度をリンク情報に基づいて総合的に判断を行う。メディア毎の知識適合度を直接比較することから、尺度が異なるとは、比較を行うことができない。そこで、知識適合度を統一の尺度にして扱う必要がある。知識適合度は、知識と入力情報との類似度であるので、明確に類似性を判断できる尺度が好ましい。例えば、類似性を示す尺度として距離があるが、知識と入力とが完全に一致する場合は、距離0として明確に示すことが可能である。しかしながら類似していない場合は、値域が無限大に大きくなり評価を行うことが困難である。そこで定量的に類似性を示すことのできる尺度として確率を用いることにした。最も類似している場合が1となり、最

も類似していない場合が0となることから値域も明確である。本節では、各メディアの知識適合度確率化手法と、それに基づくモデル化について解説する。

### 4.1 画像知識適合度の確率化

画像知識適合度のマハラノビスの汎距離  $D$  は、入力を  $f$ 、知識の特徴量分布の重心を  $\mu$ 、分散共分散行列を  $\Sigma$  としたとき式(1)で定式化される。

$$D = \sqrt{(f - \mu)^T \Sigma^{-1} (f - \mu)} \quad (1)$$

またマハラノビスの平方距離  $D^2$  は、 $\chi^2$  分布に従うことが知られている[4]。  $\chi^2$  分布の自由度は特徴量分布の次元数である。画像情報は5次元の形状特徴量であるので自由度5の  $\chi^2$  分布に従う。そこで、画像知識適合度の確率として  $\chi^2$  分布の上側確率を用いることとする。

### 4.2 音声知識適合度の確率化

音声知識適合度の尤度は、音声知識HMMと入力音声をViterbiアルゴリズムを用いて比較して得られる。比較の際に、HMMの遷移確率と出力確率を累積した値であるので、HMMと入力の適合確率である。しかしながら出力確率は、一般的に確率密度で代用される。本研究でも同様の代用を行っているため、知識適合度の確率化にそのまま尤度を採用することはできない。そこで、状態遷移の各状態に特徴量分布を保持していることから画像知識適合度と同様にマハラノビス距離を導出し確率化を行う。音声情報は16次元の特徴量であることから、自由度16の  $\chi^2$  分布に従い、  $\chi^2$  分布の上側確率を求めることが可能である。

### 4.3 知識適合確率を用いたモデル化

画像・音声の知識適合度より算出された確率値を統計論の分野で用いられる仮説検定を用いて、意味を定義した。仮説を“入力に知識に含まれない”と定義し、分布の上側確率で仮説を棄却する。ここで、上側確率を入力の知識適合確率と定義する。入力と知識との比較において、画像は部位毎に知識適合確率を算出するため、画像全体としてそれらをまとめた確率値を求める必要がある。また音声は、状態遷移の各状態毎に知識適合確率が算出されるため、音声全体としての確率値を求める必要がある。そこで、部分毎の知識適合確率の統合法として相乗平均を用いた。部分  $i$  の知識適合確率を  $p(i)$ 、部分の数を  $N$ 、全体の知識適合確率を  $P$  としたとき式(2)で定義する。

$$P = \left( \prod_{i=0}^N p(i) \right)^{\frac{1}{N}} \quad (2)$$

相乗平均をとることで、各部分の知識適合確率が平均化される。しかしながら、人間は部分的に知識との類似性が低い部分を無視したり、知識との類似性が高い部分を注視したりといった事物の特徴を考慮したふるまいが想定される。しかし今回の手法においては、認識パラメタを確率値に統一した初期のモデル化を目的としているため、そのようなふるまいは考慮しない。

また認識の結果、知識適合確率が高い値を示す知識が複数あった場合には、一概に一つの知識に認識することは誤認識の危険性がある。そこで、入力と全ての知識の知識適合確率を用いて事後確率を求める事により、知識全体を考慮した適合度の“確

信度”を定義する。知識の全体を  $x$ , 入力と知識  $k$  の知識適合確率を  $P(k)$ , 確信度の算出候補として採用されたカテゴリの総数を  $m$  としたとき, 知識群  $x$  に含まれる知識  $k$  と入力との確信度  $B(k|x)$  を式 (3) で定式化した。

$$B(k|x) = \frac{P(k)}{\sum_{j=1}^m P(j)} \quad (3)$$

メディア毎に算出された確信度を用いて言語シンボル概念の認識を行う際の総合的な認識尺度として“総合確信度”を定義する。総合確信度は、メディア毎の確信度の相乗平均とした。画像確信度を  $B^I$ , 音声確信度を  $B^S$  としたとき, 総合確信度  $B^T$  を式 (4) で定式化した。

$$B^T = \sqrt{B^I B^S} \quad (4)$$

最終的な言語シンボル概念の認識では, 最も高い総合確信度を示した画像・音声メディアの組合せが持つ概念を認識結果とする事とした。

## 5. 実験および考察

### 5.1 知識適合確率に関する実験と考察

知識適合確率が適切に導出可能であることを確認するための実験を行った。学習を何度か行った知識に対して

- 入力 1 知識に学習済みの入力
- 入力 2 知識と類似した未学習の入力
- 入力 3 知識と異種の入力

の 3 種類の入力を行い, 画像・音声共に知識適合確率  $P$  (入力 1),  $P$  (入力 2),  $P$  (入力 3) を検証する。結果としては,  $P$  (入力 1) >  $P$  (入力 2) >  $P$  (入力 3) の大小関係となることが予測できる。結果を表 1 に示す。予想のように  $P$  (入力 1) に比べ,  $P$  (

表 1 知識適合確率の実験結果

Table 1 The experimental result of the probability of conformity of knowledge and input

入力番号	画像知識適合確率 (%)	音声知識適合確率 (%)
入力 1	81.358	75.272
入力 2	32.234	44.069
入力 3	4.134	0.0

入力 3) が低くなっている。これは, (入力 3) が知識とは特徴的に似ていない入力であることを示しており, 妥当な結果といえる。(入力 2) は知識と類似した入力ではあるが, (入力 1) の方が知識により類似していることが  $P$  (入力 2) と  $P$  (入力 1) からわかる。知識との特徴的な差を知識適合確率に反映できているといえる。

### 5.2 認識・学習の評価実験

本実験では, 提案手法に基づき, 認識のパラメタに知識適合確率を用いたモデルについて認識および学習が機能するかを確認し, 提案手法の有効性を示す。本枠組において“認識”と“学習”は密接に関係のある処理である。なぜなら, ある入力が与

えられたとき, ある知識に認識されれば学習がなされ, 認識が行われなければ学習はなされない。また学習がなされなければ, 知識状態が変化しないので, 一定の入力を繰り返す限り認識結果も変わらない。つまり互いに影響を持ちあう関係であるので, 併せて評価を行うこととした。

#### 5.2.1 認識評価実験

認識は, 学習による知識の変化によって徐々に変化していく。学習状態が初期段階の場合, 画像知識は状態が不安定であるが, 学習を繰り返すことによって判断が明確になっていくと考えられる。そこで明確に認識判断が行えることを確認するために, 知識の学習回数による確信度の推移を評価する。

確信度は, すべての知識状態を考慮した知識と入力の適合確率である。よって学習を繰り返すことで, 学習が行われた知識とその知識に認識されるべき入力の知識適合確率が上昇し他の知識との知識適合確率が下降したすることが予想される。この場合, 確信度は上昇することから, ある知識に対する認識が明確になっていることを判断することができる。

実験環境として学習を一回づつ済ませた 3 個の知識を用意した。さらに学習用の入力データとして, 用意した知識に対応する 3 種類の画像をそれぞれ 9 枚計 27 枚用意した。図 4 のように, 9 枚それぞれがバリエーションの異なる画像である。

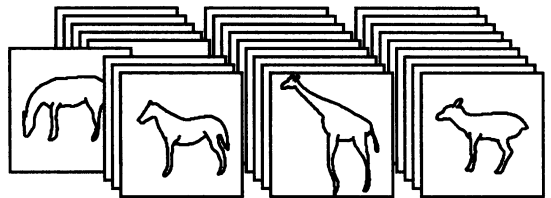


図 4 学習用データ

Fig. 4 Image data for study

学習用データをそれぞれの知識に学習させる度に確信度評価用データの入力を行い, それぞれの学習状態での確信度を確認する。学習初期の段階では, 知識は不明確な認識しかできない状態であり, どの知識も入力と高い知識適合確率を示すと考えられる。学習が繰り返されることによって, 知識が確固たるものに変化することから入力と知識が類似するものとそうでないものの知識適合確率の差が大きく, 学習状況によって判断の変化する認識能力が備わっているかの評価を行う。

確信度評価用の入力は常に同一の画像である。Uma.Kirin,Shika のそれぞれ 1 枚を用意した。また, 前述の初期知識として用意した知識は, これら 3 種の画像に対応する知識である。方法としては, 1)~3) を全ての画像を学習し終わるまで繰り返し行う。

- 1) 学習用の入力をそれぞれの知識に学習させる
- 2) 確信度評価用の入力と知識を比較し, それぞれの入力と全ての知識の知識適合確率を求める
- 3) 知識適合確率から確信度を求める

実験結果として, それぞれの知識と入力の確信度推移を図 5 に示す。

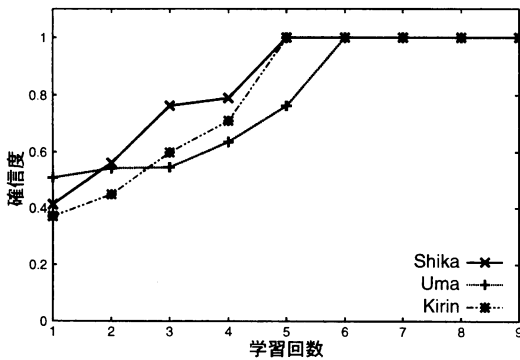


図5 認識の実験結果

Fig. 5 The experimental result of Recognition

確信度の推移は一樣上昇である。学習を繰り返す度に確信度は高くなっており、逐次学習による知識状態の変化が認識に反映されている。想定した認識判断能力を獲得できているといえる。

### 5.2.2 学習評価実験

学習が行われることで、知識は学習した情報の特徴をとらえた確固たる状態に移行していくと考えられる。つまり学習を繰り返すことで、認識判断の条件は厳しくなっていく。知識と入力との差異を敏感に判断できるようになる。そこで、認識判断の指標である知識適合確率を評価し、学習が行っているかを確認する。学習および評価用のデータとして Usagi, Uma, Kirin の3種類、各15バリエーションの画像を用意した。それぞれを逐次学習させていき学習時の入力と知識の知識適合確率を確認する。その際の知識適合確率は、それ以前の学習の状況が反映されている値と言えるので、知識適合確率を観察することで、学習能力の妥当性を評価することができる。実験結果として、それぞれの知識と入力の知識適合確率推移を図6に示す。

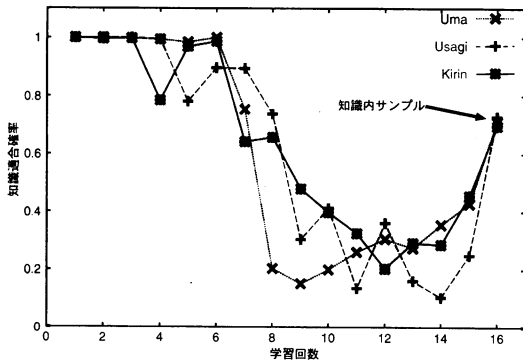


図6 学習の実験結果

Fig. 6 The experimental result of Study

全体的には、知識適合確率が学習回数を重ねる度に徐々に低下している。これは認識判断が厳しくなっている振舞といえる。入力は、一貫してそれぞれ同じ概念を表現するための画像を与えているため、学習によって確固たるものとなっていく知識と入力を比較した特徴の差を敏感に判断しているといえる。ただ

し、単純に知識適合確率が低下しているといった振舞ではなく、図6の学習回数12回目におけるUsagiの知識適合確率が上昇する振舞がみられる。この事より学習によって得た知識が保持する特徴情報に類似した入力を与えられたときには、特徴的な差を的確に判断した知識適合確率が得られていることがわかる。学習済みのデータを入力した場合の知識適合確率を確認するために16回目の入力として、学習済みの知識内サンプルを再入力したところ、高い知識適合確率を示した。このことより、以前の入力を適切に学習できていることがわかる。以上より、学習能力を獲得できているといえる。

### 5.3 形状を示す概念の獲得

本実験では、学習により獲得した知識から形状を示す概念が獲得できるかを評価する。まず、形状を示す概念である“形状概念”を定義する。

ある程度学習を進めていくと、ある音声知識  $s$  と関連付けられている画像領域集合  $R_s$  は音声知識が指すクラスを形成していると考えることができ、形状を示す概念であるとするれば、 $s$  が保持する形状概念を保持していると考えられる。しかし、人間が行う物体の形状表現は、表2のように必ず何らかの基準に対しての評価である。このことから、例え同一形状概念であっても基準が変化することによって異なった形状を表現することが考えられる。そこで形状概念は、 $R_s$  のみでは決定できず、必ず基準となる概念の領域集合  $R_{B_{osc}}$  との比較によって評価されるべきであると考えられる。

表2 形状概念とその基準  
Table 2 Shape Concept and Standard

基準	形状表現
一般的に	長い
耳としては	長い
ウサギの耳としては	長い

そこで、この形状概念  $C_s$  と  $R_s$  の関係を  $R_{B_{osc}}$  を踏まえて記述すると式5のように記述することが可能である。

$$C_s = \text{func}(R_s, R_{B_{osc}}) \quad (5)$$

以上のことより、音声知識の表現する概念情報は基準を規定しない限り、一意に求めることは困難である。

そこで、基準概念  $C_{B_{osc}}$  が全ての知識からなる分布で表現されるとし、全知識の重心を基準に評価を行うこととした。ここで、式5における実際の  $\text{func}()$  の処理内容は、本知識表現法では形状の概念を表現するであろう特徴量を分布の形で保持していることから、領域集合  $R_s$  の分布を基準概念の領域集合  $R_{B_{osc}}$  で正規化することである。この正規化した分布は、基準領域形状に対しての、音声知識  $s$  の形状的特徴の特異性分布を表現する。そこで、本研究ではこの形状の特異性を表現する分布を音声知識  $s$  の形状概念を表現する分布であると定義する。

また、全ての音声知識中においてどの音声知識が形状を表現するものなのかが不明な場合、形状的特異性分布と基準形状を用いることによって音声知識が表現する概念の形状依存度が定義できると考えた。

そこで、形状概念の領域集合  $R_s$  の分布の重心と、基準概念の領域集合  $R_{Base}$  の分布のマハラノビス距離を用い、その音声知識が表現する概念の形状依存度として定義する。この形状依存度を定義することにより、知識内の全ての音声知識内でどの音声知識が形状概念を表現するものなのかを帰納的に求めることが可能となる。

実験には、図7に示すような5種類の画像物体で、それぞれ姿勢の異なる画像3枚ずつ合計15枚用意し、それぞれの各部位に対しカオ、ハナ、クビ、マエアシ、ウシロアシ、シッポ、ミミ、ドウタイ、ミジカイ、ナガイ、フトイ、ホソイ、マルイという13種類の音声刺激と共に学習させた。そして学習データを知識としたとき、各音声知識が表現している形状概念を構築させる実験を行った。



図7 入力画像の例

Fig. 7 The example of input image

各部位の教示例を図8に示す。

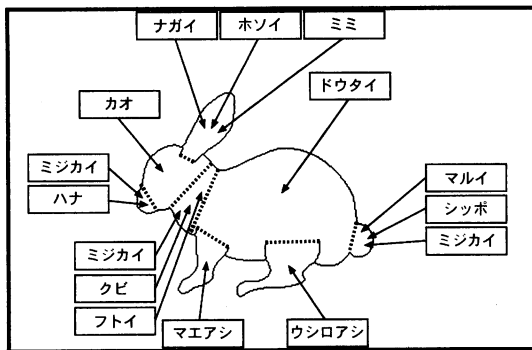


図8 教示した画像・音声の例

Fig. 8 The example of the taught image and speech

ここでナガイ、ミジカイ、フトイ、ホソイ、マルイといった形状を表現する入力刺激を与えているが、この基準は実際の尺度を考慮したものでなく、画像中で教示者が主観的に感じた部分に対して教示したものである。

実験結果として、構築された各音声表現する形状概念の重心位置と、形状依存度を示すマハラノビスの距離を各形状特徴量毎に表3に示す。

形状依存度のマハラノビスの距離はその値が大きいくほど、知識内の平均形状と比べて特異な形状を持つことを示している。結果について検討すると、マルイ、フトイ、ミジカイからなる形状を表現すべき概念の音声知識の形状依存度は大きく、カオ、ミミ、クビなど部位を表現する概念の音声知識の形状依存度は小さくなっている。

表3 形状概念の構築結果

Table 3 The experimental result of shape concept

音声知識	面積	長さ	幅	最外殻距離	周囲長	形状依存度
ハナ	-1.379	0.316	0.263	0.666	0.120	2.990
ミミ	-0.909	-0.285	-0.672	-0.373	0.487	4.221
カオ	-0.362	0.400	-0.560	-0.230	-0.199	2.772
クビ	-0.664	1.540	1.636	-0.421	-1.019	4.420
マエアシ	-0.721	-0.626	-0.262	-0.391	-0.038	3.428
ウシロアシ	-0.127	-0.316	-0.472	-0.448	-0.642	7.323
ドウタイ	0.410	1.008	-0.824	-0.607	0.540	23.157
シッポ	-0.596	-0.399	-0.031	-0.275	-0.220	2.078
ナガイ	-0.170	-0.967	0.172	-0.805	0.068	4.769
マルイ	-0.808	0.762	-0.729	-0.172	-0.432	21.805
フトイ	-0.134	1.483	0.544	-0.587	-0.492	24.668
ホソイ	-0.618	-0.9092	0.711	-0.686	-0.048	3.343
ミジカイ	-0.921	1.731	1.479	-0.256	-1.359	21.887

また形状概念を示すナガイとホソイに関しては、今回の実験で与えた画像からはそれらの概念を明確に示す形状の特徴を与えられなかったことから、形状依存度が他の形状概念の形状依存度に比べて小さくなったと考えられる。

さらに、ドウタイに関しては形状を表す概念ではないが、形状依存度が形状を表す概念と同等の大きさを示している。これは、ドウタイとして教示した部位が図8に示すように、一貫して大きい領域を持つ領域であったため、そういった形状を示す概念として構築されたと考えられる。

以上より、学習によって知識を獲得し、獲得した知識を認識に再利用するといった一連の処理を実験により確認した。さらに獲得した知識を認識以外に再利用し、形状概念を獲得することが可能であることを確認し、モデルの妥当性を示した。

## 6. あとがき

本研究は、異なった尺度で得られる画像・音声の知識適合度を統計的な手法を用いて確率値として得る手法を提案した。さらにその確率値を用いたモデル化を行った。これにより、メディア毎に特別な共通化ルールを用いずに画像・音声入力の知識適合度の比較が可能となった。さらに、学習および認識について評価を行い、提案手法に基づいたモデルが認識・学習能力を獲得できていることを示した。また、学習によって獲得した知識から、形状概念が獲得できることを示し、知識の再利用性を示した。

## 文 献

- [1] 亀井: "画像・音声メディアによる対話からの物体概念獲得に関する研究", 千葉工業大学大学院工学研究科, 情報工学専攻修士課程学位論文 (1998-2)
- [2] 黒田, 荒井, 石川: "画像・音声メディアを用いた人間-エージェント間対話による言語シンボル概念の獲得", 電子情報通信学会, SAA2000 講演論文集, pp.297-304, (2000-11)
- [3] 小松: "エージェントの確信度を用いた画像・音声情報からの物体概念獲得", 情報通信学会, 信学技報 A196-38(1997-01)
- [4] 奥野忠一: "多変量解析法 (改訂版)", 日科技連出版社 (1984-09)
- [5] 新村: "広辞苑 第五版", 岩波書店