論　文　　Technical Papers

視聴覚情報の統合化に基づく概念の獲得

Concept Acquisition based on Combining with Visual and Auditory Information

中川　聖一*1　中西　宏文*2　古部　好計*1　板橋　光義*1
Seiichi Nakagawa　Hirobumi Nakanishi　Yoshikazu Kobu　Mitsuyoshi Itahashi

*1 豊橋技術科学大学情報工学系
Dept. of Information and Computer Sci., Toyohashi University of Technology, Toyohashi 441, Japan.
*2 愛知教育大学情報科学コース
Dept. of Information and Computer Sci., Aichi University of Education, Kariya 448, Japan.

1992年8月13日　受理

Keywords: concept acquisition, visual and auditory information, image processing, speech processing.

Summary

There have been many works to embed the learning capability into computers. But these researches have not yet arrived at the goal. Traditional artificial intelligence systems have been responded to only one kind of external stimulus by given knowledges but has not been able to enhance its ability or efficiency fitted to their environment. In contrast, recent language acquisition systems which efficiently learn the vocabulary and it's meanings by using linguistic and non-linguistic information has been studied. But these systems handle non-linguistic information as predicate calculus instead of the natural stimulus (as visual information) and only small categories of non-linguistic input.

From this point of view, we made a concept acquisition system for the purpose of formalizing the method to acquire the concept with two external stimuli, that is, visual scene and auditory sound. Our system acquires concepts without a priori knowledge. This system learns the concepts which contain names, locations, colors and sizes of objects, using visual (image) information and the related auditory (voice) information. The basic operation is to extract a common part or feature from two images or two speech sounds, and is to map the extracted common part of images on the extracted common part of sounds. The correspondence is refined by the generalization and specialization.

Consequently, some concepts are acquired about correspondence of voice features to image features, by sequencially learning from image and voice. We have realized the first stage of human's concept acquisition process on a computer system.

1. はじめに

人工知能の研究は，人間にのみ可能であった知識の保有・処理などを情報処理技術の立場から計算機処理の対象とし，計算機に演繹・推論能力をも持たせようとするもので，将来的には人間の知的機能を模倣し，あるいは，その代行をすることをも目指している[Michalski 87, 大須賀 86, 大須賀 87, Paul 90]．従来の人工知能の研究は，対象に関する概念・意味体系をあらかじめ分析し知識として記憶しておき，そのうえで推論操作などを行う，いわば静的システムの研究にとどまっている．つまり，与えられた知識によって外的刺激に反応するが，外的刺激に対してシステムの知識を変更することはなかった．このため，環境に適応して能力・効率を高めていくことができないばかりでなく，まったく未知の外的刺激に対して何ら合理的な反応を行うことができない．しかも，外的刺激にはさまざまな種類があり相互に関連しているにもかかわらず，ただ1種類の外的刺激を扱うだけのものがほとんどであった．

これらの困難を克服するには，新しい概念や意味体系を外的刺激から獲得する能力，および未知の複数の外的刺激に対する意味づけを行うメカニズムの解明が必要である．このような概念獲得メカニズムが明らかになれば，人間の情報処理方式の解明ばかりでなく，

人工的なシステムに未知の環境での適応能力を与えたり，大規模な知能システムのデータベースを自動作成する場合にも有用である．

Siskindは，初期知識として単語の辞書的なカテゴリーや意味を全く知らない状態で，言語的な情報(文)と，非言語的な情報(文に対応する意味記述)からなるシナリオの列を入力として与えることにより，その二つの相関する入力源から辞書（言語入力における単語のカテゴリーと意味）を推論する方法を提案し，コンピュータプログラムMAIMRAに実現した[Siskind 90]．MAIMRAでは，言語と非言語の相関に対し統語的・意味的・語用論的制約を適用することにより，単語の辞書的な知識を持たずに単語の意味が学習される．

また，Bell研究所のA. L. Gorinらは，「行うことによる学習（learning by doing)」のメカニズムについて提案した．これは，あるタスクを行う過程でそのタスクに必要な言語上の知識を獲得するものである．この研究では，「いくつかの特殊なタスクに対するゴールは入力メッセージを意味のある行為にマッピングすることである」とし，入力メッセージの集合を言語，マッピングを理解と位置づけている．実際のシステムでは，会話文をタイプしたものとそれに関連する行為の対の集合を与え，マッピングの重みを算出しネットワークを構築した．概念獲得実験では，12人の1 000文入力で1 500単語以上の語彙が獲得された．これは入力文中に含まれる知識の99%に相当するものである[Gorin 91 a, Gorin 91 b]．しかし，メカニズムは単純で，単語と行為との相関の統計的学習と位置づけできる．

筆者らは過去に，文とその意味記述を与えて自然言語の文法と意味解析規則を帰納的に学習するシステムを開発しており[中川 88 a, 中川 89]，本研究はこのシステムを，実際の音声とその意味に相当する画像情報を与えるようにし発展させたものである[古部 91]．

本研究では，視覚と聴覚という二つの外的刺激を結びつけて未知の入力に対する概念の獲得方式を定式化する．具体的には，①視聴覚情報による子供の概念獲得の計算機シミュレーションを行うこと，②機械的に効率の良い帰納的・確率的概念獲得アルゴリズムを確立すること，③学習システムによって得られた情報を音声言語理解システムや画像理解・生成へフィードバックする方法の検討，および，④音声－言語－概念－画像の相互関連を整理し，概念獲得という立場から融合化することを最終目的とする．

子供の概念の獲得過程は認知心理学の分野などで活発に研究が行われ，部分的にはそのメカニズムが明らかにされている．しかし，例えばチョムスキーの提唱する不偏文法ですら異議を唱える学説もあり[安西 92]，概念獲得過程の解明は難しく，ましてやこれを工学的にシミュレートすることはできない．

そこで今回は，子供が概念を獲得しているという事実にヒントを得て，上述の②に関して検討と実験を行い，良好な結果を得たので報告する．この研究の発展によって工学モデルから，逆に認知過程への解明の糸口や未知語の獲得，誤った概念表現の検出など工学的応用が見い出されると期待される．

2. 幼児期における言語および概念の獲得

2・1 概念獲得法の分類

[安西 92, 村田 81, 大久保 67]

人間の幼児が言葉を使えるようになるのは，個人差があるが，およそ1歳前後といわれており，この頃の言葉は単語ではあるが文の機能を持っているので，一語文といわれている．次の段階としては，1歳半前後において，これまで使用している一語文の一語が，述語部分になったり，述語の対象(目的語)の部分になったりして，二語結合文が使用されるようになる．場面がなくても言葉を聞いただけで，要求であるか叙述であるかが聞き手に判断できるようになる始まりである．しかし，これも最初は二語の羅列で，2歳前後になって，関係を表す助詞「が」，「も」，「に」などが使えるようになり，三語文・四語文も出てくる．そして，3歳くらいの幼児になるとかなりの種類のものについて大人とほぼ同程度の分類，命名ができる．

こういった言葉やそれに対応する概念の獲得の仕方は，その概念の種類や個人によりさまざまであるが，以下の(1)～(5)の方法のいずれか，あるいはそのいくつかの組合せを通じて獲得してきたと考えられている[小林 92, 坂本 83]．

(1) 定義による概念獲得
　直接概念の定義を教える方法
(2) アフォーダンスによる概念獲得
　概念にそれに関係した動作などを併せて学習させる方法
(3) 事例による概念獲得
　個々の概念の具体例を示す方法
(4) 言語文脈による概念獲得
　概念の用いられる言語文脈から理解・誤解する方法
(5) 概念の名による複合概念の推察
　複合語のもとの語から推測して得られる概念

今回の研究ではこれらのうち，(1)，(2)を考慮しながら(3)の事例による概念獲得について研究・実験を行った．そこで(1)〜(3)についてもう少し詳しく説明する．

2・2 定義による概念獲得

古典的な語彙獲得に関する考え方では，次のような能力が子供に備わっていると仮定している．

① 物体を数多くの属性（色，形，材質など）に分解する能力
② カテゴリー形成の属性リストを抽出，検証，修正，廃棄，保持する能力
③ 新しい事例がすでに形成しているカテゴリーメンバとなり得るかを検証する能力

これに対しピアジェの理論では，高度な仮説検証は小学生くらいで初めて可能になるとされる．また，幼児は1日約9語の速さで語彙を獲得するともいわれていることから，マークマンは古典的考え方が仮定する高度な仮説検証能力を3歳くらいの子供が持つという考えを否定した．そして，子供は語の意味の仮説においてバイアスを持っており，あらゆる可能な仮説を考慮するのではなく，ほんの1〜2個の仮説しか考慮しないと考えた．このバイアスを仮定と呼ぶ．さらにこれらは，事物の命名に関しては，母親がまずある事物を指して命名し，子供は与えられた言葉がその事物の何を意味するのか推測を行う，といった前提のもとに考えられていると思われる．つまり，辞書や教科書，父母や教師などに直接概念の定義を与えられて獲得する．

2・3 アフォーダンスによる概念獲得

定義による概念獲得について，さらに深く考えてみると，実際の命名場面では単に指さして命名するだけではなくその事物を使用したアクションを子供に見せている．このアクションのことをギブソンはアフォーダンスと呼んだ．

子供はまず多くのものに幅広く適用できるような「さわる」，「つかむ」，「なめる」などのアクションから，物に対する特殊的なアクション（「それを握って食べ物を食べる」，「バウンドさせる」など）に収束していくと考えられる．これについては，すべてのものについて子供がものの名称を覚える以前にものに特殊的なアクションを行うことが，実験により確認されている．さらに，ものの名称が獲得されるまでの間に，特殊的なアクションを言語的に表現する段階があることが観察されている．つまり，子どもはものの名称を学ぶ際，ものの全体の形と名称を即座に結びつけるのではなく，まずその物のアフォーダンスを学び，それからものの名称を学ぶと考える学説である．

2・4 事例による概念獲得

「かわいいワンワンだね」，「これはワンワンじゃないの，シカさんていうんだよ」などと，「ワンワン」の概念は個々の事例を通じて獲得される．個々のワンワンをワンワンの正事例（positive instance），シカさんのようなワンワンでない事例を負事例（negative instance）という．そして，正事例と負事例とに分ける基準の学習の過程およびそれに作用する変数の研究は，弁別学習実験の手続きを用いて行われてきている．この手続きのいくつかの基本的な特徴を指摘しておく．

① 正事例と負事例を分類できることは概念の獲得の重要な指標であり，共通反応を要求する事例と他の事例とを弁別させる手続きは，概念獲得の課題であるといえる．
② この手続きは，個々の事例についての情報から分類の基準を学習することを要請する．
③ 事例の構成の仕方には種々のものが考えられるが，新しい事例にも適用可能な概念を達成させるために，事例の何らかの性質に基づいて分類できるような概念が設定される．
④ 事例の組の構成を工夫することによって，人工的ではあるが被験者に対して一種の世界（環境）を提供し，その構造を各様に変化させることができる．
⑤ 事例を継時的に与えることによって，思考過程を時間的に引き延ばし，それぞれの時点における被験者の反応を分析して，情報処理過程を明らかにする可能性を開く．
⑥ 言語理解などに関しては，正事例だけからなされるとされている．負事例を与えることにより，効率的な学習が達成され得るが，多くの負事例から有効な事例を選択することは難しい．

3．本システムの概要

本研究では，具体的に人間の幼児がどのような情報によって概念形成を行っているのか考察し，これをヒントに工学的に概念形成メカニズムを計算機上で実現することを最終目標にしている．今回開発したシステム全体の構成を図1に示す．

人間の場合，いくつかの感覚器を単独であるいは組み合わせて使用し，外部からの刺激を感じてそれらの

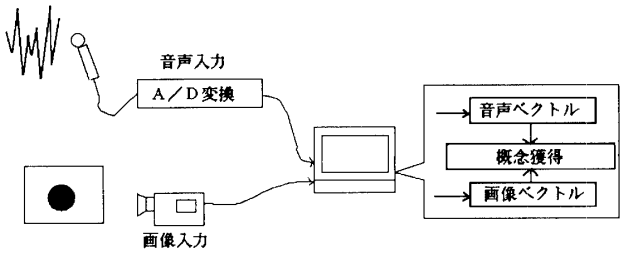


図1 概念獲得システムの構成図

情報が脳に伝えられさまざまな概念を獲得していると考えられる．そのなかでも特に，事物の名前などを学習するためには視覚と聴覚が最も重要な役割を果たしているであろうことは容易に想像がつく．そこでこの視覚と聴覚によって得られる情報，つまり音声と画像の情報を用いて計算機にものの名前や位置などの概念を学習させるシステムを作成した．つまり，あるものを表現する画像があったとするとその画像を説明する文を音声によって与えることにより，逐次画像上の形状，色，大きさ，位置といった概念に対応する音声言語を獲得すること，逆にいえば，ある「音」に対応する形状の概念を獲得することが本研究の目標である．ただし，ものの名前や位置などの概念を単語として与えるのではなく，簡単な文の音声データとそれに対応する画像データとを用いて，形状，大きさ，位置，色などの概念を形成することとした．ただし，今回は，大きさよりも形状のほうが早く獲得されるという子供の概念形成過程のモデル化は意図せず，これらの概念は互いに同質と考えて，同一のメカニズムで獲得する方法を考案する．このことから，画像どうしの類似性の自動判定・音声どうしの類似性の自動判定・画像と音声の対応づけなどの機能が基本操作となる．

4．画像からの図形データの抽出

カメラで取り込んだ図形画像から，直接概念を形成することは難しい．そこで，概念形成に必要とされるパラメータを図形画像から抽出しなければならない．

今回の実験で形成する概念は，図形の存在性，図形の位置，図形の大きさ，図形の色，図形の形状の五つの概念グループに分類される．以下ではこれらの概念およびその概念を形成するために求めるパラメータについて述べる[古部 91]．このプリミティブな機構は先天的に人間に備わっている（あるいは，すでに概念獲得されている）と仮定している．

（1） 図形の存在性

図形の存在性の概念には，入力画像中に図形がある，図形を画像中に出す，あるいは図形を取り去る，といった概念が含まれる．これらの概念とマッチングを取るためのパラメータとして，画像中の図形の数を求める．

画像中に図形が存在するかどうかを判断する方法として，その画像に，2値化に適当なしきい値が存在するかどうかを調べる．具体的には，判別分析法で求めたしきい値の良さ（クラス分離度）[大津 80]が，ある値より大きければ，図形が存在するものとする．図形が存在すれば，求めたしきい値により2値化処理を行った後，ラスタースキャンにより図形を探す．

（2） 図形の位置

この概念グループには，図形の絶対的な位置の概念，すなわち，画像中の右，左，あるいは上や下と，相対的な位置の概念，つまりある図形の右だとかいう概念が含まれる．この概念とマッチングを取るために，2値化した図形画像より，図形の重心座標，すなわち図形を形成する画素の座標の平均値 (μ_x, μ_y) と，その画素の分散 (σ_x, σ_y) をパラメータとして求める．

二つの図形間の位置に関するマッチングは，各図形の画素が，x 方向に (μ_x, σ_x)，y 方向に (μ_y, σ_y) で正規分布しているとみなし，それぞれの方向での，二つの図形の正規分布の重なりを求めることにより行う．ある程度以上重なっていれば同じ位置にあるとし，同じ位置にない場合の相対的な位置関係は，二つの図形の重心座標の差により求めることができる．

しかし今回の実験では，絶対的な位置としての，右，中，左，あるいは上，下という概念のみの形成を行ったので，重心位置を適当なしきい値で区切り，横方向に3段階，縦方向に2段階の計5段階に量子化した値を用いる．

（3） 図形の大きさ

このグループには，図形が大きい，普通，小さいという概念が含まれる．そこで，パラメータとして，図形の絶対的な大きさ，つまり，2値化画像での図形の占める画素数(面積)を用いる．図形間のマッチングは，2図形の面積の比率をとることにより行う．

しかし，例えば人間には四角形と三角形が視覚的にほぼ同じ大きさに思えても，絶対的な面積ではかなり異なるような場合が考えられるが，今回の実験では考慮していない．この概念を正確に獲得するには，各図形に対し，[大きい]，[普通]，[小さい]といった概念を，音声とそれに対応する画像で与え，複数個の学習サンプルを提示することにより，初めは「大きい三角形」と「普通の四角形」の区別が困難であっても，学習に伴い，それぞれの形状に応じて，概念の違いに対応する面積のしきい値を求めることができる．

しかし，この大きさについても，今回はあらかじめ，絶対的な大きさ（大，中，小）の概念形成の実験のみを行ったので，図形の相違による混乱の生じないデータを用い，面積を3段階に量子化した値を用いる．

（4） 図形の色（グレーレベル）

図形が黒い，または，灰色だ，という概念グループである．このグループに対しては，図形のグレーレベルの平均値と，その分散をパラメータとする．

これらの値は，判別分析法で，2値化しきい値を決定するときに求めた値を流用する．

図形間の色情報のマッチングは，位置のマッチングと同様に，正規分布の重なりを求めることにより行う．しかし，この場合も（2），（3）と同様に絶対的な色情報のみ(黒，灰色，白)の概念形成を行ったので，グレーレベルの平均値をしきい値で区切って3段階に量子化した値を用いる．

（5） 図形の形状

図形の形を表す概念グループであり，今回は丸，三角形，四角形が含まれる．図形の形状を表すパラメータとして最適なものはまだ明らかでない．図形の形状を捉える方法として，図形の輪郭線を用いるものと，図形の面積的な要素から形状を表すものの二つが考えられる．

今回は，丸，三角形，四角形といった単純な図形のみを対象としているので，パラメータとしてもあまり複雑なものを用いる必要はないと考えられる．そこで，これら3種類の図形を識別するため，図形の輪郭線から屈曲点を抽出し，屈曲点の数をパラメータとすることにした．以下に屈曲点の求め方を記す．

1° 図形の輪郭線を追跡し，輪郭線を構成する画素の位置を得る．

2° 輪郭線を構成する画素の一つ一つにつき，その画素の位置での輪郭線の曲率を求める．

3° 曲率が極大となる点を探す．そして，その極大となる曲率が，あるしきい値以上であれば，それを屈曲点とする．

今回は，音声との照合に屈曲点数をパラメータとしたが，より複雑な図形の概念獲得を行う場合には，屈曲点数のほかにより良いパラメータを抽出する必要がある．

以上の処理により，図2に示される画像データより次のような量子化された4次元特徴量を得る．

(y_1, y_2, y_3, y_4)

ここに，y_1：形状（屈曲点数＋1）

y_2：大きさ（1：中，2：大，3：小）

y_3：色（1：白，2：黒，3：灰色）

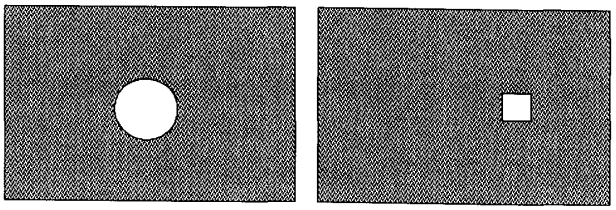


図2 画像データの処理結果例

y_4：位置（1：中心，2：右，3：左，4：上，5：下）

今回，使用した画像特徴量は，すべて原画像から得られるデータを，あらかじめ定めたしきい値により量子化して用いた．しかし，原画像から得られる画像情報のある特徴に着目し，音声データと対応がとれるようにクラスタリングすることによって，量子化の際のしきい値を自動学習させることも可能である．

5．音声データからの音声情報の抽出

画像に対応する音声から対応する音声情報を抽出する．ここではその基本操作であるDPマッチングによる二つの音声どうしの類似区間の抽出法[兵後 89]と，ある「音声」が入力音声中に含まれているかどうか検出する方法[中川 84]について述べる．

前者の方法としては，二つの音声（文あるいは文の一部区間）をそれぞれ10 msごとにLPC分析し，求めたLPCメルケプストラム係数とその回帰係数の時系列データどうしのDPマッチングを行い，それによって算出された最適照合パスおよび照合距離により，類似区間を抽出する．

ここで，1番目の音声の時系列データを

$$A = a_1\ a_2\ \cdots\ a_I,$$

2番目の音声の時系列データを

$$B = b_1\ b_2\ \cdots\ b_J,$$

a_i と b_j のフレーム間距離を $d(i,j)$ とし，対称形のDPパスを用いるとAとBの累積照合距離 $D(I,J)$ は次の漸化式で求められる[中川 88 b]．

$$D(i,j)=\min\begin{cases}D(i-1,j)+d(i,j)\\ D(i-1,j-1)+2*d(i,j)\\ D(i,j-1)+d(i,j)\end{cases}$$

この方法で最適照合パスが求められると，音声の共通区間どうしのパターンは区間長もほぼ同じで時間軸

の伸縮があまりなく，しかも照合距離が比較的小さいと期待できるので，以下の3種類の基準により共通区間の抽出を行った．

(1) 最適DPパスの線形性
(2) 平均距離
(3) 距離の最大値

ここで，「線形性」については，DPパスが i 方向または j 方向に連続して4フレーム以上進むならば，非線形とみなし共通部分でないとした．「平均距離」および「距離の最大値」はこれらが一定のしきい値を越えるものは共通部分でないとした．

このようにして得られた共通区間を2文間の共通区間として用いる．

また，入力音声中に登録辞書と共通の音声区間があるかどうかを検出するために，DPマッチングによるワードスポッティングアルゴリズム[中川 84]を用いた．これは，連続音声中から音韻や音節，単語などの候補を抽出するもので，辞書と入力文をマッチングした際に，辞書のパターンに対し入力文の各フレームで終端するDPパスの最適な照合位置と照合距離を求め，照合距離を正規化したものがあるしきい値以下であれば，入力文中に辞書と同じパターンが存在すると判断する方法である．この方法により，入力文と辞書との共通区間を抽出した．

6. 概念の獲得アルゴリズム

6・1 アルゴリズム

ここでは**4章，5章**で述べた画像データと音声データから概念を獲得するためのアルゴリズムについて述べる[古部 91]．ここでいう概念とは図3のようにある対象物に対する音と画像特徴を意味する（音声波形のような物理的な量は[ ]で表現している）．

今回提案する方法は，画像と音声の対応づけができたものを辞書部として登録し，逐次学習が行われるものである．この様子を図4に示す．

以下，二つの文音声をマッチングし差異と共通部分を抽出した場合，その差異のあった部分を特徴区間，共通部分を共通区間と呼ぶ．また，画像と音声の対応づけができたものを辞書部として登録していくものとする．

ただし，最初の2文間に共通区間がない場合には，そのまま辞書に登録し，以後登録辞書とのみワードスポッティング法で照合する方法となる．

・記号の定義

R_{sen}：参照文・画像の番号

図3 概念の内容

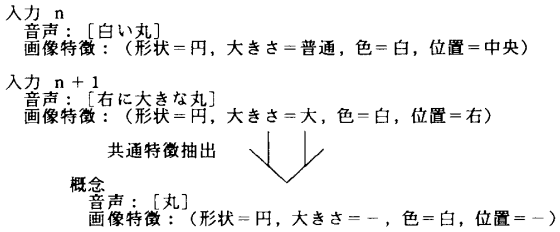


図4 概念獲得

I_{sen}：入力文・画像の番号

R_{num}：辞書に登録されているデータ数

$R_e(L)$：辞書中の L 番目のデータの出現度数

Word：辞書中の参照データ番号

1° 初期設定：参照文 $I_{sen}=2$，$R_{num}=0$ とする．

2° $R_{sen}=I_{sen}-1$ とし，R_{sen} と I_{sen}（ある文とその一つ前の文）の音声データをマッチングし，特徴区間・共通区間を抽出する．また，画像データを各概念を表す次元ごとにマッチングし音声の特徴区間に対応する次元（y_i）を求め，それ以外の次元（$y_j, j\neq i$）をクリアしたものを，その音声の特徴区間と対応するデータとして割り当て，残りの次元（$y_j, j=i$）を共通区間とする．もし，共通部がまったく抽出されない場合，$I_{sen}=I_{sen}+1$ とし，このステップを繰り返す．

特徴区間・共通区間をそれぞれ対応する画像データとともに登録し，R_{num}，R_e(Word) をインクリメントする．

3° $I_{sen}=I_{sen}+1$ とし，辞書に登録された音声部と，画像パラメータの有効次元数の少ないものから順に，I_{sen} の音声部とマッチングする．マッチングが取れたら，画像パラメータの対応する次元をクリアし，I_{sen} の入力音声からマッチングが取れた音声データを削除する．ただし，この際，画像パラメータの有効次元のデータ内容が一致せず削除できないときは，辞書中の登録データを，一致しない次元のパラメータは冗長なものとして，クリアし訂正したうえで削除する．これをすべての辞書と照合するか入力音声データがそれ以上削除できなくなるまで続ける．一定時間長以上の音声区間がまったく削除されなかった場合は，2°へ行く．残った音声部・画像パラメータの次元を対応させて，辞書中に登録し，R_{num} をインクリメントし，出現度数をカウントする．新たな入力データがなくな

るまで，3°を繰り返す．

6・2 処 理 過 程

上記のアルゴリズムに従って処理される過程を例示する．ただし，入力音声文は以下のとおりとする．

1. [まるがあります]
2. [さんかくけいがあります]
3. [しかくけいがあります]
4. [おおきいまるがあります]
5. [おおきいさんかくけいがある]
6. [ちいさいさんかくけいがある]
7. [みぎにまるがあります]
8. [ひだりにさんかくけいがある]
9. [うえにしかくけいがあります]
10. [したにまるがあります]
11. [くろいまるがある]
12. [しろいまるがある]
13. [ひだりにしろいまるがある]
14. [みぎにちいさいさんかくけいがあります]
15. [うえにくろいしかくけいがある]
16. [したにおおきいまるがあります]
17. [くろいおおきいまるがあります]
18. [うえにちいさいしかくけいがある]
19. [しろいちいさいまるがあります]
20. [ひだりにおおきいさんかくけいがある]

1° 初期設定

2° 文1と文2の音声データをマッチングし，特徴区間・共通区間を抽出する．

[丸があります] (1,1,1,1)
[三角形があります] (4,1,1,1)

特徴区間として，音声データの[丸]・[三角形]が，共通区間として[があります]に相当する部分が得られたとする．また，画像データとの対応づけで

[丸] (1,-,-,-)
[三角形] (4,-,-,-)
[があります] (-,1,1,1)

となる．この3個のデータが辞書に登録される．

3° $I_{sen}=2$：処理なし

3° $I_{sen}=3$で，すでに辞書中に3個のデータが登録されているので，これらの音声部と文3の音声をマッチングすると，[があります]に相当する部分で照合がとれる．この部分を削除すると

[四角形] (5,-,-,-)

が残りとして得られるのでこれを登録する．

3° $I_{sen}=4$で，[大きい丸があります]のデータ(1,2,1,1)から，辞書部の音声と[丸]の部分が照合され，[大きい……があります]のデータとして，(-,2,1,1)が残る．次に，登録データの[があります]の部分と残りの音声の一部と照合がとれるが，登録データの画像パラメータ(-,1,1,1)と比較するとy_2が一致しないので[があります]の登録データからこの次元を冗長なものとしてクリアし，[があります](-,-,1,1)と修正したうえで削除すると[大きい](-,2,-,-)が残り，辞書に登録される．

このように繰り返していくと，1から12までの文と画像パラメータを上記のアルゴリズムで処理することによって，音声のマッチングが正しく行われた場合，以下の13種類の音と対応する画像パラメータが得られる．文13以下は，辞書に登録された概念の信頼性を検証するために使用される．

[まる] (1,-,-,-)
[さんかくけい] (4,-,-,-)
[しかくけい] (5,-,-,-)
[おおきい] (-,2,-,-)
[ちいさい] (-,3,-,-)
[みぎに] (-,-,-,2)
[ひだりに] (-,-,-,3)
[うえに] (-,-,-,4)
[したに] (-,-,-,5)
[くろい] (-,-,2,-)
[しろい] (-,-,1,-)
[があります] (-,-,-,-)
[がある] (-,-,-,-)

なお，本アルゴリズムは画像データの抽出ミスに対しては考慮していないため，画像データ抽出にあたってはしきい値を最適化し，できる限り希望する画像の特徴抽出データが得られるようにした．また，音声区間の類似区間の抽出ミスに対しては，すべての概念が複数回出現するようにしたため，一度辞書に正しく登録された音声部に対しては，たとえそれ以降，本来あるべき共通区間が検出されなくても，問題は生じない．しかし，学習の初期の段階で検出ミスが生じると希望する概念がうまく獲得できない場合が生じるが，子供は最初一語文のような抽出ミスの起こらない概念から獲得していることを考えると，むしろ自然であるといえる．

7. 概念獲得実験

7・1 実 験 方 法

実験は，音声資料として成人男性話者2名(話者A，

B) が**6・2節**で述べた20文を各1回ずつ文節ごとに区切って発声したものを,画像資料としては20文に対応する20個の画像をビデオカメラから入力したものを,それぞれ**4章**,**5章**で述べた方法で音響,画像処理して特徴量のベクトル系列に変換したものを用いた.

結果は,アルゴリズムの2°ならびに3°において概念が獲得される過程と最終的に辞書に登録されたもの(PCM録音データを再生して聴取りにより文字化した)を確認する方法を用いた.画像の特徴パラメータ抽出に関しては,**6章**で述べたようにしきい値を最適化したのですべて正しく抽出できた.また,音声データに関しても,共通区間検出の際のしきい値を最適化した.

概念獲得実験として,**6章**で述べたアルゴリズムと入力文を用いて次の2種類の実験を行った.

[**実験1**] 初めの2文間で共通区間,特徴区間を抽出し辞書に登録し,以下ワードスポッティング法を用いて辞書と入力文を比較する方法.

[**実験2**] 最初の2文間の比較を行わず,ワードスポッティング法のみを用い概念獲得を行う方法.この方法では,初めの数文は入力文がそのまま辞書に登録される.

7・2 実験結果と考察

話者Aに対する実験1の結果を表1(a)に示す.この表では,辞書番号,画像データ,出現頻度,対応する音声,付加,脱落の順で実験結果を示す.例えば4番目の辞書(辞書03)には,画像特徴として(5 0 0 0)が獲得されており,出現回数3回,対応する音声特徴は[四角形]であり,その音声特徴は,付加(本来存在しないのに抽出した)誤り0,脱落(本来存在するのに抽出できなかった)誤り1個であったことを示している.

この実験では,[四角形],[がある],[小さい],[左に]の四つの音声区間に関し,本来共通区間として抽出されるべきところが抽出漏れ(脱落)が生じた.話者Bの音声に対する実験では,[三角形]3回,[丸]1回の抽出漏れが生じた.しかし,本アルゴリズムでは入力音声に対して,共通部分が検出されなかった場合概念の獲得を行わない,つまり負事例による学習は行わないため問題は生じなかった.ここではこれらの単語は他の入力音声により適切な学習が行われ,二人の話者の実験共に最終的に目的とする概念が正しく獲得でき

表1 概念獲得実験結果(話者A)

(a) 最初の2文間で比較し,3文目以後ワードスポッティング法使用

参照辞書	画像ベクトル $y_1y_2y_3y_4$	出現頻度	最初の2文比較,以後ワードスポッティング法		
			概念	付加 *	脱落 *
辞書00	0 0 0 0	11	[があります]	0	0
辞書01	4 0 0 0	6	[三角形]	0	0
辞書02	1 0 0 0	10	[丸]	0	0
辞書03	5 0 0 0	3	[四角形]	0	1
辞書04	0 2 0 0	5	[大きい]	0	0
辞書05	0 0 0 0	7	[がある]	0	1
辞書06	0 3 0 0	3	[小さい]	0	1
辞書07	0 0 0 2	2	[右に]	0	0
辞書08	0 1 0 3	2	[左に]	0	1
辞書09	5 1 0 4	1	[上に四角形]	0	0
辞書10	0 0 0 5	2	[下に]	0	0
辞書11	0 0 3 0	3	[黒い]	0	0
辞書12	0 0 1 0	3	[白い]	0	0
辞書13	0 0 0 0	6	[がある]	0	0
辞書14	0 3 0 0	1	[小さい]	0	0
辞書15	0 0 0 4	2	[上に]	0	0
辞書16	0 0 1 3	1	[左に]	0	0

(b) ワードスポッティング法のみ使用

参照辞書	画像ベクトル $y_1y_2y_3y_4$	出現頻度	ワードスポッティング法のみ使用		
			概念	付加 *	脱落 *
辞書00	1 0 1 0	6	[丸があります]	0	1
辞書01	0 0 1 0	3	[三角形があります]	1	0
辞書02	5 0 0 0	2	[四角形があります]	0	0
辞書03	0 2 0 0	5	[大きい]	0	0
辞書04	0 0 0 0	6	[三角形がある]	2	0
辞書05	0 3 0 0	4	[小さい]	0	0
辞書06	0 0 0 2	2	[右に]	0	0
辞書07	0 0 0 3	3	[左に]	0	0
辞書08	0 0 0 4	3	[上に]	0	0
辞書09	0 0 0 5	2	[下に]	0	0
辞書10	1 0 0 0	3	[黒い丸がある]	2	0
辞書11	5 0 0 0	2	[黒い四角形がある]	1	0
辞書12	0 0 0 1	2	[丸があります]	0	0
辞書13	0 0 0 0	1	[白い]	0	0

*
付加:対象中に存在しないに、存在すると判断し抽出してしまう誤り
脱落:対象中に存在するのに、存在しないとして抽出しない誤り

た．逆に，本来共通区間がない部分で音声が共通区間として抽出（付加誤り）されると，本アルゴリズムでは以降の学習で希望する概念が獲得できない場合も生じる．

また，実験2の結果を表1(b)に示す．この方法は初めの2文間に必ずしも共通区間があることを必要としないが，実験結果からもわかるように，初めの文に含まれる概念はうまく分離できていない．これは入力文の例や量に依存する．しかし，子供は簡単な文(一語文)から学習することを考えると必ずしも欠点ではない．

7・3 問題点と今後の検討課題

今回は画像のデータをパラメータ化する際に，形状を表すパラメータとして屈曲点の個数や形の大小のしきい値の情報を使ったが，これには「屈曲点」などという概念を人間がすでに獲得していることを前提としてしまっているために，ゼロからの概念獲得という点で問題が残った．屈曲点の検出は初期視覚として先天的に獲得できているかもしれないが，形の大小や色の量子化のしきい値の付与などの点で問題がある．このため子供の概念形成のシミュレーションとしては，画像の形状をパラメータ化する何らかの別の方法が必要であると思われるし，一つの概念が二つ以上のパラメータに対応する場合（例えば，色と青，緑，赤の強度の関係）の学習法なども必要である．

また，今回は学習させる文の順序をあらかじめ注意深く決定したために，最終的には期待どおりの概念が獲得できたが，別の予備実験から初めに文2と3をもってきて実験を始めると，[三角形]，[四角形]の概念が[三]，[四]，[角形があります]となり，獲得できないことがわかっており，この点への対処も必要となり，現在，この改良を進めている[中西 93]．

実際の音声と画像を使って概念獲得実験を行ったが，音声データのマッチングにおいて共通区間の抽出精度があまり良いとはいえず，この点の改良も必要となる．

さらに，途中で期待どおりの概念が獲得されなかった場合のことを考え，それぞれ最低2回以上学習するようにしたが，このようなことは必須条件であると思われる．

8. ま と め

本研究では，視覚と聴覚という二つの外的刺激を結びつけて未知の入力に対する概念の獲得方法について検討し実験を行った．これは計算機が環境から長期の自動学習を行う初期の段階をシミュレートする実験である．

その結果，逐次，画像と音声を対応づけながら学習を行うことにより，ある概念に対応する音と画像上の特徴とが対応づけられ，人間の概念獲得手法の初期段階が計算機上で実現できた．最近，文字言語に対して，未知語の獲得の帰納的学習法[荒木 92]も提案されているが，本研究はさらに一歩進んだ，一部未知語を含んだ音声言語（文字言語であればメカニズムはより簡単になる）から，その概念を獲得する場合にも適用でき，工学的意義は大きいと考えられる．

今後，図形から抽出する特徴パラメータ，および複数図形の取扱い，またその際の相対位置の問題やアクションの概念獲得などを解決していく必要がある．

◇ 参 考 文 献 ◇

[安西 92] 安西祐一郎ほか 編著：認知科学ハンドブック，共立出版（1992）．

[荒木 92] 荒木健治，栃内香次：帰納的学習による言語の獲得および確実性を用いた語の認識，信学論，Vol. 75D-II, No. 7, pp. 1213-1221（1992）．

[Cohen 90] Cohen, P. R. and Feigenbaum, E. A.（田中・淵監訳）：人工知能ハンドブック，III，共立出版（1990）．

[Gorin 91 a] Gorin, A. L., Levinson, S. E., Gertner, A. N. and Goldman, E.: Adaptive acquisition of language, *Computer Speech and Language*, Vol. 5, pp. 101-132（1991）．

[Gorin 91 b] Gorin, A. L., Levinson, S. E., Miller, L. G. and Gertner, A. N.: On adaptive acquisition of spoken language, *Proc. 1991 IEEE Workshop on Neural Networks for Signal Processing*, pp. 422-431（1991）．

[兵後 89] 兵後裕子，中川聖一：連続発声された二発話文間のDPマッチングによる共通部分の抽出，信学全大，A-22（1989-03）．

[小林 92] 小林春美：アフォーダンスが支える語彙獲得，言語，Vol. 21, No. 4, pp. 37-45（1992）．

[古部 91] 古部好計，中西宏文，辰巳昭治，中川聖一：音声と画像の対応付けに基づく概念の獲得，人工知能学会全大，2-10（1991）．

[Michalski 87] Michalski, R. H.（電総研 訳）：*Machine Learning*（邦訳：知識獲得と学習シリーズ，1, 4, 5），共立出版（1987）．

[村田 81] 村田孝次：言語発達研究，培風館（1981）．

[中川 84] 中川聖一：拡張連続DP法による連続音声認識アルゴリズム，信学論，Vol. 67-D, No. 10, pp. 1242-1249（1984）．

[中川 88 a] 中川聖一，若原一彰：自然言語の構文・意味解析

規則の主観的確率を用いた帰納的学習システム，人工知能学会誌，Vol. 3, No. 6, pp. 773-782 (1988).
[中川 88 b] 中川聖一：確率モデルによる音声認識，電子情報通信学会 編，コロナ社 (1988).
[中川 89] 中川聖一，山本幹雄，若原一彰：自然言語の文法と意味解析規則の帰納的学習システム，情処学論，Vol. 30, No. 1, pp. 72-80 (1989).
[中西 93] 中西宏文，板橋光義，古部好計，中川聖一：音声情報と画像情報の統合による概念獲得，信学全大，D-203 (1993-03).
[大久保 67] 大久保 愛：幼児言語の研究，あゆみ出版 (1967).
[大須賀 86] 大須賀節雄：知識情報処理，オーム社 (1986).
[大須賀 87] 大須賀節雄，佐伯 胖：知識の獲得と学習，オーム社 (1987).
[大津 80] 大津展之：判別および最小2乗基準に基づく自動しきい値選定法，信学論，Vol. 63-D, No. 4, pp. 349-356 (1980).
[坂本 83] 坂本 昂：現代基礎心理学 7，東京大学出版会 (1993).
[Siskind 90] Siskind, J. M.: Acquiring core meanings of words, *Proc. 28th Annual Meeting of the Association for Computational Linguistics*, pp. 143-156 (1990).

〔担当編集委員・査読者：堀 浩一〕

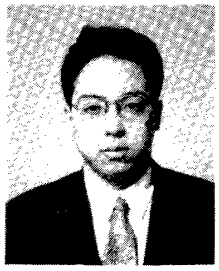
著者紹介

中川 聖一（正会員）

1976年京都大学大学院博士課程修了．工学博士．同年，京都大学情報学科助手．1980年豊橋技術科学大学情報工学系講師，1983年助教授，1990年教授．1985～86年カーネギーメロン大学客員研究員．音声情報処理，自然言語処理，人工知能の研究に従事．1977年電子通信学会論文賞．1988年度IETE最優秀論文賞．電子情報通信学会，日本音響学会，情報処理学会，計量国語学会，神経回路学会，IEEE，ESCA各会員．

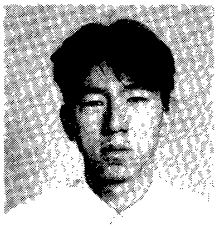
中西 宏文

1984年豊橋技術科学大学情報工学系卒業．1986年同大大学院修士課程情報工学専攻修了．1988年同博士後期課程退学．同年，愛知教育大学情報科学コース助手．1993年助教授．音声情報処理，知識処理の研究に従事．1988年度IETE最優秀論文賞．電子情報通信学会，日本音響学会各会員．



古部 好計

1990年豊橋技術科学大学情報工学系卒業．1992年同大大学院修士課程情報工学専攻修了．現在，NTT関西情報システムセンター勤務．在学中は，人工知能，画像処理の研究に従事．



板橋 光義

1992年豊橋技術科学大学情報工学系卒業．現在，同大大学院修士課程在籍．