知能と複雑系 113－9
(1998. 7. 24)

画像・音声情報により表現される知識の整合性保持を目的とした人間–計算機間対話モデル

伊与部洋，荒井秀一

武蔵工業大学工学部電子情報工学科

東京都世田谷区玉堤 1-28-1

人間は対話を用いることで互いの持つ知識の擦り合わせを行なっている．そこで，人間–計算機間の対話を通じて計算機に人間の持つ概念を構築することを考えた．本研究では，画像・音声刺激を用いた，認識対象に依存しない認識・理解を行なっており，全知識の獲得過程にまで至る，学習・認識・理解の過程をモデル化している．本モデルの特徴は，複数の異なるメディアからの情報を刺激の時間同時性から統合していくことである．これによりシンボリックな処理を用いずに言葉の持つ意味概念を獲得することを可能にしている．この概念獲得を行なうため，人間対計算機の対話をモデル化し，対話を利用した言葉の概念獲得方法を提案している．

画像理解, 音声理解, 概念獲得, 対話

Man-machine Communication for Retaining Consistency in Knowledge with Image and Voice Media

Hiroshi IYOBE , Shuichi ARAI

Musashi Institute of Technology

Tamazutsumi 1-28-1, Setagaya, Tokyo

Human adjust his knownledge to other's with communication. Therefore, a computer can acquire the concept which human understands with Man-Machine communication. This paper describes the knowledge representation method to manage the different kinds of information, and that enables acquiring the knowledge by learning. This system can handle two different kinds of information: image and sound. In this system, these informations are related using the rule that the many observed stimulus in same time have same semantic information. This relationship means the semantic concept of linguistic symbol.

Moreover, this study proposes the simple Man-machine communication model, to realize the flexible understanding system that is independent of recognition target.

image understanding, speech understanding, knowledge acquisition, communication

1 はじめに

我々の社会には様々な目的の対話が存在する．知識を出し合い，物事の是非をただすことで新たな知識を獲得する討論のような対話，互いの知識を確認するだけの取り留めのない雑談のような対話や，自分の持つ情報を一方的に出すことで相手に知識の確認をしてもらうカウンセリングのような対話などが考えられる．また対話は目的に応じて様々な形態をとる．参加話者で見ると，2者間対話や多者間対話，情報媒体で見ると，音声による対話，文字による筆談や身振り手振りによるジェスチャ，情報の流れで見ると，一方的な教示のような対話や双方向的な対話などが考えられる．また対話の動機を見ると，人間の対話動機の一つとして，人間には自分の存在を示し，確認したいという欲求があると考えた．この欲求から生まれる対話は目的がなく，発言にも情報量のほとんどない対話である．またもう一つの動機として，不安解消がある．人間は自分の持つ知識で理解出来ないような事物に出会うと不安になり，この不安を解消するために，他人に意見を求めて質問すると考えた．この知識を求める対話はやりとりされる発言の情報量が多いため，有益な対話と言える．以上様々な対話について述べた．これら全ての形態の対話を単一のモデルで表現することは難しいため，対話の目的に応じたモデルを作る必要がある．

ところで，現在までに我々は画像・音声刺激間の関係から概念を計算機上に構築する枠組を提案してきた．[2]–[7] 一般的に概念とは，経験される多くの事物に共通の内容を取り出し (抽象)，個々の事物にのみ属する偶然的な性質を捨てる (捨象) ことにより生成されるものである [1] ので，本研究では，人間が意図を持って提示する音声情報，画像情報から学習により何らかのクラスタが形成されれば，その部分空間を概念と呼ぶことにする．すなわち，本研究では明示的には言語の存在を前提としない概念が存在することになるが，実際には人間の持つ概念は言語によって規定されたものが多く，人間により提示可能な音声や画像情報は暗示的には言語により支配されていると考える．[2]

一般に，言葉の概念獲得にはシンボリックな処理が用いられることが多い．言語シンボルは抽象的なものであるため多くの情報を含み，これを利用することで複雑な概念を容易に獲得できるが，シンボルの意味する具体的な事象を生成することが難しい．そこで，抽象的な概念を様々な具体的事象から築くことを考えた．本研究では人間の幼児が具体的な事象を抽象化することで概念を構築するように，画像・音声刺激を抽象化し，その関係をリンクで表現することでシンボリックな処理を一切用いずに言葉の概念を獲得させることができると考えた．言葉に関する意味は外界との対話により初めて具象化され伝搬されるので，計算機が言葉の意味を獲得するためには，言葉の意味を理解している人間との対話を行なう必要がある．[2] そこで，本研究では言葉の意味する概念獲得のための人間–計算機間対話について考えていくことにする．本研究では，計算機が持つ知識を具象化する方法と，文法などの言語知識を必要としない単純な対話モデルを提案し，これにより，人間の乳幼児期のような，言語に関する知識を全く持たない状態からの言葉の意味概念学習を行なう．

2 知識の整合性保持を目的とした対話

我々が普段，家族や親しい友人と対話をしているとき，互いの発言は良く理解でき，意志疎通に支障はない．これは互いの知識の整合性がとれているからと捉えることが出来る．人間が知識の整合性をとる一手段として，相互作用的に情報をやり取りする対話が挙げられる．例えば，外部刺激と知識との整合性がとれないとき，人間は不安になり，不安解消のために外部から情報を得ようとする．その場に他者がいる場合，人間は質問をすることで他者に働きかけ，情報を引きだそうとする．また，このとき質問された他者は，質問者の反応を見ながら，必要であると考える情報を与えることができる．

人間が二人以上集まり，互いの知識の整合性をとると，そこにコミュニティが発生する．人間は同じコミュニティに属する他者との対話や共通な体験等を通じて，自分の属するコミュニティの持つ共通概念と自分の知識との整合性をとっていると考えられる．

知識の整合性をとる一般的な手段として対話を挙げたが，対話を情報量の流れで見ると，情報の流通が一方向的な対話は教師から生徒への教示と見ることができ，また情報のやり取りが均衡している場合は同程度の知識を持った者同士の対等な対話と捉えられる．このように教示と対話の差異は話者間で流れる情報量の差でしかないと本研究では考えることとし，教示を含んだより一般的な対話のモデル化を提案する．

3 対話モデル

本研究の目的は言葉で示せる物体概念獲得のための対話であり，その動機が不安な状態の解消にあることは前章までで述べた．

一般に対話については自然言語理解の分野などで数多く研究されているが，言語解釈・理解は人間にとっても高等な知的行為に当たり，非常に複雑で大量の知識を前提とする．本研究の目的が知識獲得にあるため，対話を行なうために複雑な知識を前もって与えることはこの目的に矛盾すると考え，対話のための知識は最低限の先天的知識のみを与えることにした．

話者間の知識の整合性が取れないために話者は不安になる．このとき，対話のきっかけとなった外部刺激を“話題”と呼び，“話題”を認識・理解した結果を画像・音声刺激を用いてお互いに伝え合うことを本研究では対話と定義する．

以上の対話を実現するためには，

(1) 刺激を抽象化して認識・理解

(2) 刺激を学習

(3) 対話相手の発言の意図の判読

(4) 知識を具象化して対話相手に発言

の4つの機能が必要である．このうち(1)～(2)については．6章で詳しく述べる．(3)の機能の実現は，画像・音声といった獲得刺激だけでは難しい．そこで，人間の意志伝達手段について考えた．人間の意志伝達の手段としてもっとも単純なものに，表情が挙げられる．表情は人間が先天的に持つ意志表示手段であり，知識獲得という本研究の目的に矛盾しない．本研究では人間と同様に表情を用いて意志伝達を行なうことで対話を実現する．表情については，近年マルチモーダルインターフェースといった分野で多数研究[11] [12]が行なわれているが，本研究では表情認識・理解までは研究の対象とはせずに，常に理解できるものとした．(4)の機能を実現するためには，既に本枠組で提案されている刺激の抽象化と対を為す機能が必要となる．これについては8.3節で詳しく述べる．話者は表情を用いることで自分の意志を相手に伝えることができ，その結果として刺激の判断・知識獲得に必要な情報を得ることが出来る．

3.1 表情の利用

人間は様々な感情を表情で他人に伝えることができる．しかし，本研究では人間が先天的に持っている基本的な表情は，幼児が示す事のできる安心・不安の感情を伝える笑顔と泣き顔に集約されると考えた．対象とする知識獲得の対話において，上記の表情に至る心理を次のような場合であると考えた．

安心 外部刺激が自分の知識のみで明確に判断できている場合

不安 外部刺激が自分の知識のみで明確に判断できない場合

この表情を伴う発言は

(1) 話者が判断した結果を不安な表情とともに発言する場合

(2) 話者が判断した結果を安心した表情とともに発言する場合

(3) 話者が安心／不安の表情のみで発言する場合

がある．(1)，(2)は話題に関する自分の意見であり，(1)の場合は不安そうな「～でしょうか」という質問の意味の発言と捉えることができる．また，(2)の場合は自信を持って「～です」という断定の意味の発言と捉えることができる．(3)は発言した話者に相手から返された反応で，理解できたか否かの返答と捉えることができる．

3.2 話者の性格

一般に対話を行なう話者には様々な性格が想定できる．本研究では，知識獲得における知識の整合性保持という目的から対象とする話者は，

- 自分の知識のみで刺激を明確に判断できれば安心している．
- 自分の知識のみでは刺激を明確に判断できないと不安になる．
- 相手から質問されると自分の判断に疑問を感じ不安になる．

という性格をしているものとし，またその性格に付随する行動として

- 不安になると誰かに質問をして安心するための情報を得ようとする．
- 刺激を受ける度にその刺激を判断し，自分が安心・不安のどちらの状態にあるかを表情を示す．

を考えた．このような性格の話者が対話を行なう状況を次節のようにモデル化した．

3.3 対話モデル

本研究で対象としている話者は，自分の受けた刺激を明確に判断できないと不安になり，その不安を解消するために外部に情報を求めて発言する．このような話者の対話は，

- 自分が明確に判断出来なかったために不安となった話題について，質問することで対話が開始する．
- 相手に質問を受けて不安になった話者は，相手の不安を解消することで自身も安心しようと対話に参加する．
- 全ての話者が話題を明確に判断できると安心し対話は終了する．

- 全ての話者が意見を出し尽くすと，不安を解消することが出来なくなるので対話は終了する．
- 話題についての相手の質問や意見が明確に判断できないとき，そのままでは不安を解消できないのでその発言を一時的な派生話題として明確にしようとする．
- 最初の話題と無関係な一時話題は最初の話題を明確にするための情報を持たないのでそこまでは派生しない．

というように不安解消モデルとして定義できる．

3.4 対話からの学習

本対話モデルは不安解消モデルであるから，不安を解消することが学習条件となる．知識を全く持たない場合には，判断結果は唯一の知識であるので，不安になることは無く，判断結果を知識として獲得することが可能である．知識がすでに存在する場合には，現在の知識で話題を判断した場合と，刺激を学習して話題を判断した場合のどちらの場合が，より刺激を明確に判断でき安心できるかを調べ，刺激を学習した場合のほうがより安心できるならば学習する．

4 対話のための枠組

本研究では画像・音声刺激からの概念獲得を目的とした対話を提案している．概念獲得のためには，

- 新しい知識となる情報を外部から得る機能
- 得た情報から新しい知識を抽出し，保持するための機能

が必要であり，刺激からの知識獲得には，刺激の認識と理解，学習などの処理が必要であるといえる．また対話を行なっていくために自分の保持している知識から具体的な情報を生成する具象化の機能や，対話相手との情報のやり取りを管理するための機能が必要となる．

これを実現するには次の 4 つの機能が必要であると考えた．

(1) 画像・音声情報の認識・理解
(2) 画像・音声間の関係付け
(3) 刺激の表現形態変換 (抽象化・具象化)
(4) 対話の管理

このうち (1)～(2) についてはは次章から述べる概念獲得の枠組で述べ，(4) については 9章で述べる．

5 概念獲得の枠組

本研究では画像刺激・音声刺激から画像概念・音声概念を構築し，さらにその間の関係を作成する意味付けによって意味概念を獲得することにより、知識獲得を行なってう．その過程を次のように分類して考える．

stage 1 受けた刺激の認識・理解
stage 2 刺激の表現形態変換
stage 3 画像・音声に関する学習
stage 4 画像・音声間関係からの意味概念の学習

stage 1 は受けた刺激がどの物体概念を支持しているのかを判断することであり，どのような物体を指示しているのか判断することに相当する．

stage 2 は具体的な表現形態で示される刺激から抽象的な表現に変換することである．

stage 3，4 で実際に知識を記述する．個々の刺激に関する学習が stage 3 であり，異種メディア間の関係付け，すなわち画像・音声概念の意味付けを行なうのが stage 4 である．

以上のように抽象化した刺激間の関係として意味を記述するため，明示的には言語シンボルを使用すること無く知識の獲得・利用を行なっている．そのため，次のような特徴を持つ．

- 画像・音声といった異種メディアを統合する際の知識記述レベルを考慮する必要がない
- 言語シンボルを利用する際に必要となる言語シンボルが示す事象に関する知識が必要無い．これにより，予め与える知識を最小限にすることが可能である
- 前提として必要な知識が少なく，刺激を直接的に用いて知識を構築するため，学習初期段階における基礎的知識を獲得可能である
- 獲得した知識と予め記述された知識の区別が明確である
- 知識記述の対象を限定する必要がなく，様々な分野への応用が可能である

本研究ではマルチエージェントを採用し、必要な機能を個別に実装する事で、このような知識獲得過程を実現する事にした．まず全体の構成を図 1 に示す．

図 1 中に，エージェントを楕円で示してある．外界からの刺激を画像・音声個別に認識し，表現形態変換までを行なう．その後，画像・音声双方の認識結果を利用して学習する．上記過程の stage 1，2 は画像・音声でそれぞれ独立して行ない，結果を総合して学習の対象を決定する．stage 3 も学習そのものは画像・音声それぞれ独立して行なうが，これは学習の前に認識結果を総合的に判断し，学習の可否を決定する．その後，stage 4 の画像・音声間関係学習を行なう．

この画像・音声の認識，知識の保持・個別の学習を実行する機能を持つエージェントが OA (Object

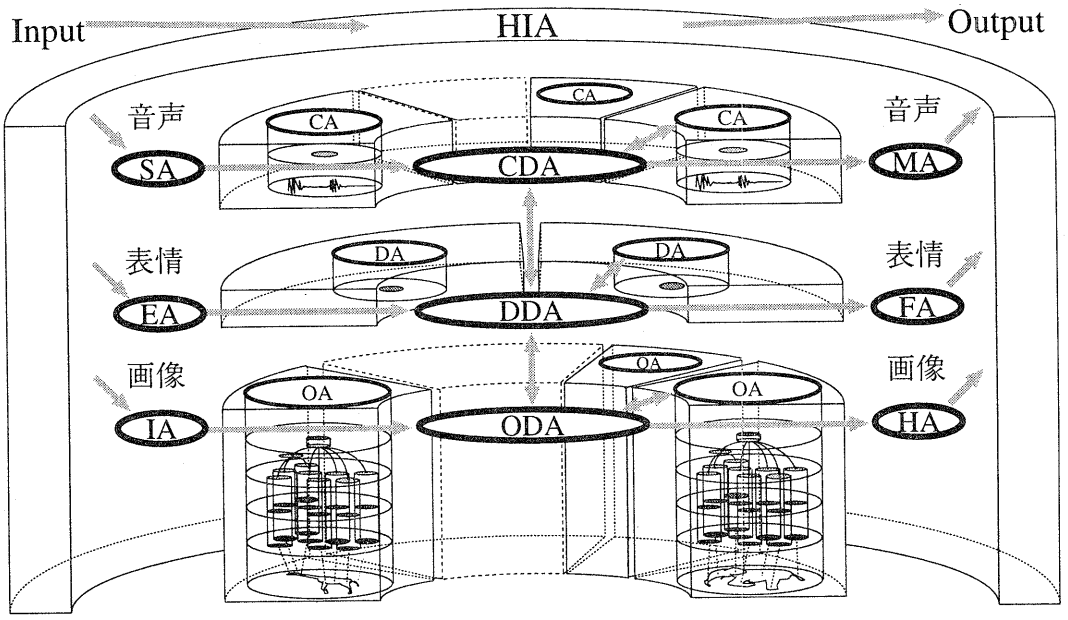

図 1: マルチエージェントによる枠組

Agent), CA (Concept Agent), ODA (Object Domain Agent), CDA (Concept Domain Agent) である．次に，これらの各エージェントの機能・動作について説明する．

6 認識・学習に関するエージェント

刺激認識とは，提示された刺激が既知物体のどれに相当するかを判断することである．これは，知識を用いて刺激を識別することで行なう．

この識別機能と，画像・音声個々の学習機能を実現するためのエージェントが，OA，CA であり，これらのエージェントについては 6.1，6.2 節で述べる．

また，識別結果から最も類似度の高い物体を判断し，未知物体である場合には知識の新規作成を行なうエージェントが ODA，CDA である．これらについては 6.3，6.4 節で述べる．

各エージェントには，表 1 に示すような知識を記述するための知識を先天的に与えてある．これらの先天的知識は対話からの知識獲得を実現するために必要な最低限と考えられるものである．

表 1 は本章で述べる認識・学習に関するエージェント以外のエージェントも含んでいるが，これに関しては 7，9，8 章で詳説する．

6.1 OA (Object Agent)

画像刺激の認知処理を行なうためのエージェントである OA は，1 画像に関する抽象的な表現形態である画像概念を保持し，画像概念と刺激が同一物体を示すと仮定し刺激を抽象化しようとする．この抽象化の過程において矛盾が生じるかどうかを検証することにより画像刺激と保持されている画像概念が同一の概念を指すか識別を行う．

表 1: エージェントの先天的知識

OA	刺激と画像概念の比較法 獲得した知識の記述・保持法
CA	刺激と音声概念の比較法 獲得した知識の記述・保持法
ODA, CDA	確信度計算方法
DDA	認識結果の競合検出に関する知識 対話時の意見生成法 学習の可否を判断する方法
IA	画像のセグメント化方法
SA	音声からの特徴量算出法
EA	相手の表情を読み取るための知識
HA	画像概念の具象化法
MA	音声概念の具象化法
FA	確信度→表情変換法

6.2 CA (Concept Agent)

1 単語音声に関する知識を保持し，音声刺激を識別するために知識との比較を行なうエージェントである CA は．逐次学習を実現するために刺激から抽出した情報を用いた音声概念の学習機能を持っている．

6.3 ODA (Object Domain Agent)

ODA は OA 群から受けた識別結果に基づいて認識結果を判断し，確信度を計算する [5]．確信度とは認識結果の確からしさを事後確率を用いて示す値であり，確信度が高い場合，認識結果が確かであると考える．OA は複数 (既知の概念だけ) 存在するため，

確信度のリストが作成される．この時，知識との比較結果として得られる類似度全てが低かった場合は，未知の物体であると考えられるため，新しくOAを生成する機能も保持している．

6.4 Concept Domain Agent (CDA)

ODAが画像に関するOA群を管理しているのに対して，音声に関するCA群を管理するのがCDAである．その機能はODAに対応したものであり，音声についての識別結果判断および，CAを新規作成する機能を保持している．

7 エージェント間の関係による概念の意味付け

本研究では画像・音声の関係獲得によって概念の意味付けを行う．そこで，画像・音声間の関係を知識として保持する必要がある．この関係は画像概念・音声概念それぞれが保持する必要があるため，OAが保持する画像領域毎，またCAが保持する単語音声毎にIDを付与し，それぞれ関係する概念，すなわち意味へのIDを持つことによって画像・音声間の関係を表現する．[6] また，このような異種メディア間の関係は概念の意味付けである．そこで，本研究では関係を表現するリンクを意味リンクと呼ぶことにした．

また，画像・音声間の関係を学習によって獲得するために，次の仮定を行なった．

- 同時に与えられた異種刺激間には関係がある

この仮定が常に真であると定義することで，同時の画像・音声の関係が刺激から得られ，学習することが可能となる．学習の際には意味リンクを作成することが必要である．また，概念の意味を用いた認識結果は意味リンクの存在を考慮することで検証することができる．そこで，このような処理を行うエージェントとしてDDA (Decision Domain Agent) を設計した [7].

7.1 DDA (Decision Domain Agent)

DDAではODA，CDAの認識結果を統合して判断を行なう．ODAからは画像確信度リスト，CDAからは音声確信度リストをそれぞれ受け取り，認識確信度と意味リンクの組合せから総合的に結果を判断する．この総合の確信度が低いときは，認識結果がよくわからない，“不安”な状態といえる．つまり，外部からの情報と自分の知識の整合性がとれないため，外部に更なる情報を求めて対話を起こすきっかけとなる．対話の管理に付いては後述の9.1で詳しく述べる．

また，このときの画像・音声間の関係学習から新しい意味の獲得も同時に行う．

新しく意味リンクを作成する条件として次の3つが考えられる．

- 画像・音声がともに新規に作成されたとき
- 画像・音声いずれかが新規に作成されたとき
- 画像・音声各々の確信度が非常に高いとき

DDAはこれらの処理を行なったあとで，判断結果と新しい意味リンク情報，すなわち関係する概念のIDをODA，CDAを通してOA，CAに通知する．

8 刺激の入出力に関するエージェントの設計

前章までで，刺激からの知識獲得に関して述べたが，この刺激は外部からのものであり，また，対話の際には情報を対話相手へ伝達することが必要である．つまり，対話を通じた概念獲得には刺激の入出力機能が不可欠である．

この入出力機能に関してもエージェントによって実現することにした．そこで，本章では入出力に関連したエージェントについて述べる．

8.1 HIA (Human Interface Agent)

HIAは外界との入出力を行う部分を一つにまとめたエージェントである．

8.2 刺激の入力に関するエージェント

刺激入力時には関係付けのための同時性確保と刺激の抽象化が必要になる．画像・音声・表情の情報を扱うため，これら各々に関してエージェントを設計した．

IA (Image Agent)：
IAは画像の入力に関するエージェントである．これにより，刺激として入力された線画像はセグメント画像へ変換される．

SA (Sound Agent)：
SAは音声から特徴量抽出を行なうエージェントである．

EA (Expression Agent)：
EAは表情の入力に関するエージェントである．HIAから相手の表情刺激を受信し，それを表情確信度に数値化してDDAに送り出す．

8.3 刺激の出力に関するエージェント

刺激の具象化は8.2節で述べた抽象化と対をなす機能である．

HA (Hand Agent) :
HAとは，人間の手の役割をするエージェントであり，OAに保持された抽象的な画像知識から，具体的画像を生成する．

MA (Mouth Agent) :
MAは人間の口の役割をするエージェントであり，CAが保持する抽象的な音声情報を具象化する機能を持つ．

FA (Face Agent) :
表情確信度で自分の心的状態をDDAから受信し，それを表情刺激に変換してHIAに送り出す．

9 対話学習に関するエージェントの設計

3章で述べた対話の管理機能について詳しく説明する．機能を具体的に挙げると

- 対話の流れを管理する機能
- 対話中で為された発言を記憶する機能

である．

9.1 対話の管理

対話の流れを管理することは全エージェントを管理することに等しいことから，中心に位置するDDAにこの機能を持たせることにした．また，対話を管理するために対話で交わされた発言を記憶する必要が生じてくる．しかし，発言は話題ごとに独立して記憶する必要があるが，一時的な話題が出現する可能性があるため，DDA自身で記憶するのではなく，DA (Decision Agent) という話題に関する発言を記憶するエージェントを生成させることにした．

DDAに追加した機能について9.1.1項で，DAについては9.1.2項で述べる．

9.1.1 DDAの持つ対話機能

DDAに，7.1節で述べた機能に加えて対話管理のために次の3つの機能を追加した．

話題を管理する機能 :
新しい話題が発生する度にDAを生成し，話題が明確になると消去する．話題が発生したか否かの判断は，相手の表情確信度と自分の判断結果から行なう．

相手の表情を理解，自分の表情を決定する機能 :
相手の表情確信度から相手の心理状態を察する．また，自分の心理として自分の認識結果の確信度を表情を用いて相手に伝える．

自分の意見を決定し，一時的に記憶しておく機能 :
話題を認識判断した結果を自分の意見とする．このとき物体そのものの他に物体の部分情報も意見として選ぶ．また，同じ発言を避けるために一度発言した情報はDAに記憶させておく．

9.1.2 DA (Decision Agent)

DAは発生された話題一つに付き一つ生成される．そして，その話題に関する自分の発言内容をDDAから受信し記憶している．またDDAから過去の意見の履歴を要求されるとそれを返す．最後に話題が明確になると消去される．

10 対話実験

本対話モデルによる知識獲得を実験により確認する．

実験目的 本実験は対話によって話者の疑問が解消され，またそのために必要十分な知識だけが得られていることを確認する．

実験方法 本研究で提案した対話の枠組をSunワークステーション上にC++言語を用いて実装した．総ステップ数は10万ステップ強である．これを話者Aとし，人間である話者Bとの間で対話による知識獲得実験を行った．話者Aの計算機には予備知識として画像刺激「りす」，音声刺激「Squirrel」とその関係に加え，他に3種類の物体とその関係を事前に学習させた．この事前学習についても本対話モデルによって行なった．

また話者B(人間)にも話者Aと同様の知識があるものとする．ここで話者B(人間)にだけ画像「りす」の詳細の部分情報とその意味リンク先である音声刺激を学習させた．

今回の実験では話者A(計算機)に事前に与えたものとは異なる画像「りす」と音声「Squirrel」の組を話題として二人に提示した．

図2には対話の前後による話者Aの知識の変化と二人に与えた話題を示す．表2は2者間対話のうち話者Aの話題の判断結果について示した．また図3に対話時に2者間で交換された意見を示す．本研究では“意見を受けたら，意見を返す”動作をPhaseと定義し，対話における時間的推移の単位として採用した．

表2より，話者A (計算機) は最初の話題判断 ($P_{A(1)}$) で音声刺激の判断は明確であるが，画像刺激の判断に迷っている．そのために話題全体を明確に判断することが出来ず不安になった．この不安を解消する為に話者B (人間) に質問をしたことから対話が始まった．

話者Bは自分の話題判断が明確であったため誰かに聞く必要はなかったが，話者Aが不安そうに質問をしてきたために自分も不安になり対話に加わった．

最終的に話者Aは ($P_{A(5)}$) で話題判断が明確になり安心した．これにより話者Bへの話者Aからの質

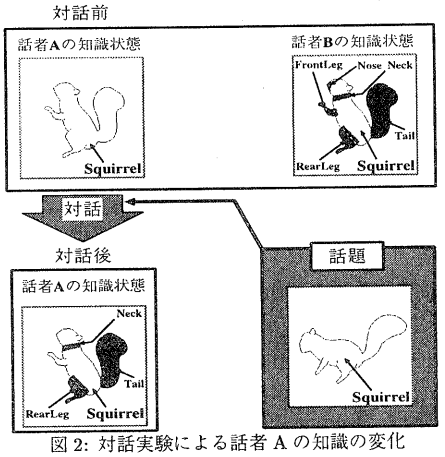

図 2: 対話実験による話者 A の知識の変化

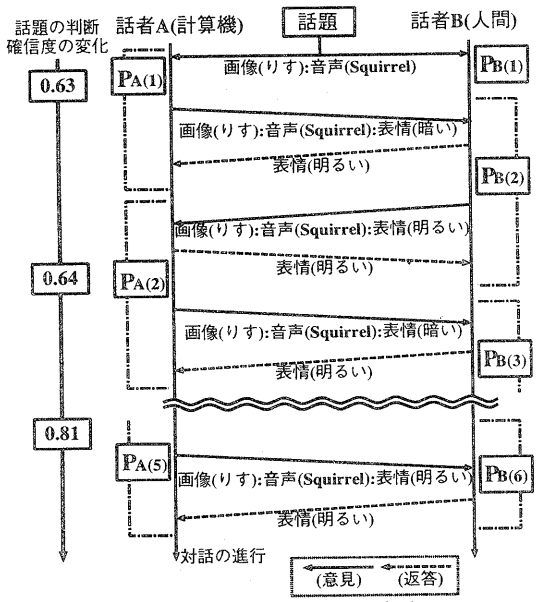

図 3: 対話実験における刺激の伝達

問がなくなったため，この対話は終了した．図 2 の対話後の知識状態より話者 A は話者 B から安心するのに必要な分だけの知識を得ていることがわかる．

この実験は話者 B から話者 A への情報量の流れが多いことから教示型の対話であったことが分る．この教示は対話の内容こそ教師側が管理しているが，話題と対話の流れを生徒側が管理することで生徒に必要十分な知識を獲得させた．

11 おわりに

本研究は，人間・計算機間における，画像・音声刺激を用いた対話によって計算機に画像・音声概念とその意味を獲得させることを可能とした．

表 2: 話者 A(計算機) の話題判断結果とその確信度

Phase	話者 B の意見		話題の判断確信度			
	音声	画像	音声		画像	
$P_{A(1)}$	Squirrel (話題)	りす (話題)	Squirrel	1.00	りす	0.27
			総合 0.63 (暗い)*			
$P_{A(2)}$	Squirrel	りす	Squirrel	1.00	りす	0.29
			総合 0.64 (暗い)			
$P_{A(3)}$	Tail	しっぽ	Squirrel	1.00	りす	0.51
			総合 0.75 (暗い)			
$P_{A(4)}$	RearLeg	後ろ足	Squirrel	1.00	りす	0.57
			総合 0.78 (暗い)			
$P_{A(5)}$	Neck	くび	Squirrel	1.00	りす	0.62
			総合 0.81 (明るい)			

*表情

また，従来の知識獲得に関する研究とは異なり，一方的な教示だけではなく表情を用いて人間，計算機ともに意志を伝達しあえる学習法を用いた．この学習法によって，人間は計算機の学習段階を考慮して学習を進めることが可能である．

またこの表情による意志伝達により相手の反応が分るので，互いに知識が安定するまで対話を続けられ，知識の整合性を保つことが出来る．

さらに，計算機が，従来用いられていた教示のみではなく対話からも知識獲得が行なえるようになることで，計算機は獲得刺激に対する信頼性が取得でき，獲得知識が知識を壊す可能性が無いか確認しながら学習することが可能となった．

そして，対話の過程において獲得した刺激が自身の知識を安定するものか判断し，常に知識が安定する方向へ学習していくなど，心理モデルに沿った知識獲得が行なえるだけでなく，知識獲得に関する研究で常に問題となる知識の発散も抑えることができるようになった．

参考文献

[1] 新村: “広辞苑 第四版”, 岩波書店

[2] 荒井: “画像・音声メディアによる人間対計算機間の対話からの物体概念獲得”, 信学技報, NLC97-69, PRMU97-271, (1998-03)

[3] 荒井: “形状・構造に関する言語シンボルの概念を学習により獲得するための画像－言語知識表現法”, 信学論, D-II Vol.J80-D-II No.9,(1997-09)

[4] 小松, 荒井: “計算機間対話による物体概念学習のための HMM を用いた知識表現法”, 信学技報, AI97-19, (1997-07)

[5] 小松, 荒井: “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”, 信学技報, AI96-32, (1997-01)

[6] 亀井, 小松, 荒井: “マルチエージェントの協調性を用いた刺激の伝搬による画像・言語知識獲得”, 信学技報, AI96-18, (1996-11)

[7] 亀井, 小松, 荒井: “画像・音声刺激から柔軟に知識獲得を行なう為の表情を用いた対話モデル”, 信学技報, AI97-25, (1997-11)

[8] 中川, 鶴見: “最大事後確率推定法と認識結果を用いた連続出力型 HMM の教師なし話者適応化”, 信学論, Vol.J78-D-II No.2,(1995-2)

[9] 樋口, 戸田: “コミュニケーションの認知科学”, 人工知能学会誌, vol.7, no.5, (1992-9)

[10] 小島, 岡本: “CSCW の対話における発言意図の推定に関する研究”, 信学技報, AI95-53, (1996-01)

[11] 長谷川, 森島, 金子: “「顔」の情報処理”, 信学論, vol.J80-D-II No.8, (1997-8)

[12] 新田, 長谷川, 他 “論争支援マルチモーダル実験システム MrBengo”, 信学論, vol.J80-D-II No.8, (1997-8)