論　文

サービスロボットのための対話システム*

滝澤　正夫†・槙原　靖†・白井　良明†・島田　伸敬†・三浦　純†

Dialog System for Service Robot*

Masao Takizawa†, Yasushi Makihara†, Yoshiaki Shirai†,
Nobutaka Shimada† and Jun Miura†

In this research, we develop a dialog system for a service robot which brings user-specified objects such as cans, bottles, and PET bottles from a refrigerator to a physically handicapped user. This paper describes two of the functions that are necessary for the system: one is the function to recognize objects in the image by using a dialog with a user; the other is the one to estimate a meaning of a word which is not registered. When the system cannot recognize specified objects automatically, it obtains necessary information by interacting with the user by speech, and tries to recognize the target object. When the system detects a word which is not registered, the system estimates its meaning by using a model of probability and registers it. We show the validity of these functions by experiments with real refrigerator scenes.

1. はじめに

現在，高齢化社会の到来により，サービスロボットの研究が数多く行われている [1,2]．サービスロボットの一つとして，ユーザの指定した物体を取ってくるロボットが考えられるが，このようなロボットは全自動では実現不可能である．ロボットが自動で行動できないときには何らかの情報を用いる必要がある．

指定された物体を取ってくる機能にとって必要な要素の一つとして，画像中から指定した物体を認識することが挙げられる．ユーザの指定するものとしては，飲み物，薬，本など様々であるが，これらは一般的にどこかに収納してあり，複雑背景下で物体認識を行う必要がある．画像を用いた物体認識については様々な研究が行われているが，複雑背景下では画像処理のみを用いた自動的な認識は困難な場合がある．このような場合には，画像情報以外の情報を用いて物体認識を行う必要がある．またサービスロボットには，認識結果を情報として与える能力と，認識に必要な情報を得る能力が必要であり，情報の流れが一方通行であるものはサービスロボットには適していない．

渡辺ら [3] は，植物図鑑の花や果物を，図に添えてある説明文を用いて認識するシステムを提案している．これは情報が一方通行であり，サービスロボットには適していない．Ahlrichs ら [2]，Wachsmuth ら [4] はユーザの助言を認識に用いている．これらは指定された物体の特徴をユーザの助言から得ることで，指定された物体の検出を可能にしている．しかし，これらの方法は認識が失敗した時に対処できない．また高橋ら [5] は，りんごや本等の物体を人との対話やジャスチャを用いて認識する研究を行っている．しかし，対象とするシーンはそれ程複雑ではないので，物体の切り出しができるものとし，その中で物体を選択するために人との対話を用いている．

また，ロボットは物体認識をした結果をユーザに伝える必要がある．藤井ら [6]，岩田ら [7] は，認識結果に基づいて説明文を生成する研究を行っている．これらは情報の流れが一方通行であり，例えば認識を失敗したときなどには対応していない．

本論文では，病院や家庭で，冷蔵庫から指定した飲み物（缶，瓶，ペットボトル）を取ってくるサービスロボットについて述べる．対象とする冷蔵庫のシーン（Fig. 1）は，背景色に似ている物体や，別の物体によって隠蔽されている物体があるため，画像処理のみを用いた自動的

* 原稿受付　2002 年 6 月 26 日
† 大阪大学　大学院　工学研究科　電子制御機械工学専攻
Dept. of Computer-Controlled Mechanical Systems, Osaka University; 2-1, Yamadaoka, Suita city, Osaka 565-0871, JAPAN
Key Words: spoken dialog, estimation of unknown words, learning of unknown words, object recognition, service robot.

Fig. 1 Refrigerator scene

な認識が困難な場合がある．このような場合に，認識に必要な情報をユーザとの対話によって取得し，物体認識を可能にすることを目指す．本論文では，対話システムについて述べる．物体認識については [11] を参照のこと．

本論文では，ユーザとして身体的に障害を持ち，話ができる人を対象としているので，インタフェイスには特別な装置を用いず，音声を利用する．ユーザが物体認識の結果を見ることができない場合は対話が非常に難しくなるので，手元のディスプレイで画像を見られることを前提とする．

手元にディスプレイがある場合，インタフェイスとしてタッチパネルが有効な手段として考えられる．しかし，対象とするユーザが手を動かせなくて細かく場所を指定できない場合や，寝たきりでディスプレイを触れない場合はタッチパネルは使えない．小さなディスプレイではうまくポインティングできないことも欠点として挙げられる．またタッチパネルは，音声デバイスより一般的でなく，コストも高いので，本研究では音声を利用する．

インタフェイスに音声を用いることの問題点として，音声認識の認識率の低さがある．システムが受け付ける単語を限定すると，この問題はある程度解決できるが，ユーザが受け付ける言葉を覚える必要があるので，ユーザの負担になる．ユーザの負担を減らすため，本論文では未知語（誤認識された単語や登録語の同義語）が検出されたときに，その単語の意味を推定し，学習する方法を提案する．

伊藤ら [8] は，文法のみを用いて未知語の属性を推定している．あいまい性があるときはユーザに選択を求めるが，あいまいな属性が多数存在するときは効率的な対話を行えない．高橋ら [9] は，対話の状況を考慮し，実例を用いて属性を推定しているが，発話中のどの単語が未知語であるかわかっているものとしている．Damnati ら [10] は，対話システムのボキャブラリに追加する単語のクラスを推定しているが，対話による確認がないため，間違った推定をすることがある．本論文では，発話中から未知語を検出し，対話の状況，前後の単語，認識された文字列を考慮して，属性だけでなく，意味 (どの登録語と等しいか) も推定する方法を提案する．推定結果はユーザに確認した後，登録語に追加する．

またこのシステムでは，認識した物体を操作することが必要であるが，これについては [12] を参照のこと．

2. 対話システムの概要

サービスロボットに必要な機能として，目的地までの移動，物体認識，物体操作などが考えられるが，本論文ではユーザの対話による支援を最も必要とする物体認識について述べる．システムはユーザとの対話を利用して物体認識を行うが，なるべく自然な対話を目指し，ユーザに煩わしさを感じさせないようにする必要がある．これには，以下のことに気を付ける必要がある．

- なるべく自動で物体を絞り込む
  例えば，同じ物体を二つ検出したときに,「どちらにしますか？」と言ってユーザに選んでもらうよりは,一つ選んで「こちらにしますか？」と質問した方が良い．
- ユーザから必要な情報を得やすいような質問をする
  例えば物体を検出できなかったときは，ただ単に「見つかりませんでした」と答えるよりは,「どのへんにありますか？」と質問した方がユーザは情報を与えやすい．

システムはこのようなコンセプトに基づいて対話をする．実際に様々な状況に対してどのような対話が行われるかは，**3.** で述べる．

次に，音声処理の概要を Fig. 2 に示す．音声認識には，IBM 社の ViaVoice を用いる．本研究で用いる ViaVoice の音声認識エンジンは以下の二つである．

- 文脈自由文法をサポートしたエンジン
  あらかじめ定義しておいた文法に一致した場合のみ認識可能．語句を限定するため，認識率は高い．
- ディクテーションをサポートしたエンジン
  入力された音声をすべて文字列に変換する．新聞などの文章の変換には強いが，話し言葉の認識率は低い．

まず，文脈自由文法をサポートしたエンジンで認識を行う．これで認識できた場合（文法に一致した場合）は，その情報を画像処理部に渡して画像処理を行う．ユーザがあらかじめ定義してある文法以外の言い方をした場合や認識がうまくいかなかった場合には，このエンジンでは認識できない．この場合にはディクテーションをサポートしたエンジンにより音声を文字列に変換する．システムはこの文字列と確率モデルを用いて，未知語の意味の推定を行い，ユーザに推定が正しいことを確認した後で学習を行う．

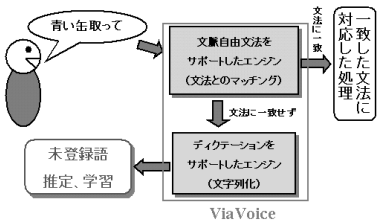


Fig. 2 Speech processing

3. 対話を利用した物体認識

まず物体をどの方向から見ても認識できるよう，モデル（物体の大きさ，代表的な色，代表的な色以外の特徴など）を登録する．認識時には，代表的な色から物体の位置を絞り込み，その他の特徴を用いてモデルとの照合を行う．物体認識の機能としては，指定物体を認識する機能と冷蔵庫内の全物体を認識する機能とがある．詳しい処理については [11] を参照のこと．

このような自動の物体認識機能と対話を利用することで，ユーザの欲しい物体を検出することができる．ここでは，物体認識に対話をどのように用いるかについて述べる．まずユーザはシステムに指示を与えるが，指示の方法としては，(1) 指定物体の色と種類を与える場合（「赤い缶」など），(2) 指定物体の名前をシステムに与える場合，(3) 冷蔵庫に何があるかを聞いてから指定する場合が考えられる．(2) については **3.1** で，(3) については **3.2** で詳しく述べる．(1) の場合には，対応する色の領域を抽出してから，種類の照合をする．対話については **3.1** を参照のこと．また，**3.3** で実験結果を示す．

3.1 指定物体の名前を与える場合

指定物体が登録されている場合は，システムはその物体を認識することができる．システムはその物体名に相当する物体（例えば,「なっちゃん」と指定された場合には,「オレンジ」「アップル」「グレープフルーツ」の缶，ペットボトル）の検出を試みる．登録されていない場合は認識することができないので，システムは「何色の物体ですか？」と質問し，指定物体の色に関する情報を得ることで認識をする．認識結果として，以下の 4 通りが考えられる．

3.1.1 物体が一つ検出された場合

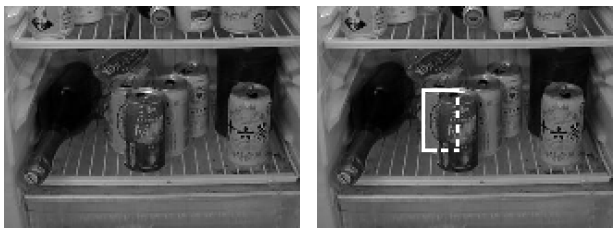
システムは「一つ見つかりました．これでいいですか？」と言って結果を示す．ユーザが「はい」と言えば，認識した物体は正しい．間違っている場合には，(1) 指定物体の位置の情報を与える，か，(2) 必要に応じてシステムに認識が間違っているという情報を与える．

(1) 位置の情報を与える場合

ユーザが，絶対的な位置情報（例えば「上の棚の左の方にあります」）や，相対的な位置情報（例えば「物体 A の右にあります」）を与えた場合には，システムは与えられた位置情報をもとにして指定物体の認識を試みる．ユーザが隠蔽情報（例えば「物体 A の後ろです」）を与えた場合には，システムは A の周囲を探すことで指定物体を認識する．Fig. 3 に，異なる色の物体に隠蔽された物体（同色物体の場合は **3.1.3**(2) 参照）の認識結果を示す．Fig. 3(b) の破線は隠蔽されているエッジを表す．

(2) 認識が間違っているという情報を与える場合

ユーザが認識結果を訂正した場合（例えば「いいえ，それは物体 B です」），システムは今後同じ間違いをしないよう，物体のモデルを修正する．



(a) Automatic recognition result　(b) Recognition result after dialog

Fig. 3 Recognition of the object hidden by another object with the different color

3.1.2 二つ以上の物体が検出された場合

システムは検出した n 個の物体すべてを表示してユーザに質問する．すべて同一の物体である場合には，そのうちの一つを選んで「n 個見つかりました．これでいいですか？」と質問する．ユーザが「はい」と答えた場合，選んだ物体は正しい．ユーザが位置や種類（缶，ペットボトルなど）に関する情報を与えた場合には，他の物体を選ぶ．この場合，システムは選んだ物体が間違っていると推定し，指定物体の認識が終わってから「ところで，先ほどの物体は何ですか？」と質問する．隠蔽情報や間違っている情報を与えた場合は **3.1.1** と同様の処理をする．

3.1.3 候補となる領域は検出されたが指定物体は検出されなかった場合

指定物体のモデルに完全に一致はしないが，それに近い物体が検出された場合は，ユーザに注意深く見てもらうため「自信がないですが，これで正しいですか？」と質問する．このような認識結果が得られる場合として，次の 3 通りが考えられる．

(1) 指定物体が検出されている場合

システムが指定物体と認識できなかった原因は，照明条件の変化などにより，すべての特徴を抽出できなかったからである．ユーザから認識が正しいという情報を与えられると（「はい」と答える），システムは今後同じ間違いをしないよう物体のモデルを修正する．

(2) 同色物体に隠蔽されている場合

システムは二つの物体を，間違って一つの物体と認識している．ユーザが，二つの物体が重なっているという情報（例えば「二つ重なっています」）を与えた場合，システムは検出した領域を二つに分割することで指定物体を認識することができる．同色物体に隠蔽された物体の認識結果を Fig. 4 に示す．

(3) 全く別の領域（背景あるいは未登録の物体）が検出された場合

ユーザは **3.1.1**(1) と同様に真の物体の位置の情報を与える．

(a) Automatic recognition result (b) Recognition result after dialog

Fig. 4 Recognition of two overlapping objects of the same color

3.1.4 候補となる領域も指定物体も検出されなかった場合

システムは「見つかりませんでした．どのへんにありますか？」と言って，ユーザから位置の情報を得ようとする．ユーザの返事を待つ間，システムは冷蔵庫内の全物体の検出を試みる．全物体の検出が終わるまでにユーザから位置の情報を得た場合には，その情報をもとに認識をする．ユーザの返事より先に全物体の検出が終わった場合には，ユーザが指定しやすいよう，検出された物体を表示する．システムから「物体 A と物体 B が見つかりました」という情報を与えられると，ユーザは「その A の後ろの缶を取って」などと情報を与えやすいからである．

3.2 冷蔵庫に何があるかを聞く場合

システムは冷蔵庫内の全物体の検出を試みる．ユーザの欲しい物体が検出されれば，ユーザはそれを選ぶことができる．欲しい物体が検出されなかった場合や認識結果が間違っている場合には，**3.1** と同様の対話をする．

3.3 実験結果

研究室内の被験者 5 人に対し，実際の冷蔵庫画像を用いて（登録物体は 40 種類）実験した．自動では認識できない場合（異なる色の物体に隠蔽されている場合，照明条件の変化により検出を失敗する場合，間違って認識した場合，候補が複数見つかる場合，同色物体に隠蔽されている場合）について，対話を利用することで物体認識を成功に導くことができた．以下に，異なる色の物体に隠蔽されている場合（Fig. 5(a)) の対話を示す．

ユーザ：白い缶（ダカラ）とって
システム：見つかりませんでした．どのへんにありますか？
ユーザ：青い缶（アクエリアス）の後ろ
システム：1 個見つかりました．これを取りますか？ (Fig. 5(b))
ユーザ：はい
システム：取りに行きますので，しばらくお待ちください

4. 音声処理

ここでは音声認識の結果をどのように処理するかを述べる．まず，文脈自由文法を用いた音声認識に，どのような文法，登録語を用いるかについて述べる．次に，ユーザの発話が文法と一致しなかった場合に，どのようにして未知語の意味を推定し，学習するかについて述べる．

(a) Original image (b) Recognition result

Fig. 5 Example of object recognition using dialog

4.1 文脈自由文法を用いた音声認識

文法を定義するため，本論文では，文法に含める単語（受け付ける単語）を Table 1 のように，カテゴリーに分けて登録しておく．登録語は，発音類似度（**4.2.5** 参照）を計算するため，発音も合わせて登録しておく（例えば,「缶」の場合は「かん」）．

これらのカテゴリーの発話される順番を決めておくことで，文法を定義する．例えば「＜物体名＞の＜相対位置＞の＜種類＞を＜動詞＞」という文法に対しては,「ダカラの左の缶を取って」や「コーラの後ろのペットボトルを取って」などの認識が可能である．助詞に関しては「助詞」というカテゴリーを文法に記述するのではなく，それぞれの登録語を直接記述しておく．

4.2 未知語の推定と学習

ユーザの発話に未知語が含まれている場合は文法に一致しないので，ディクテーションで文字列に変換する．変換された文字列を用いて未知語の意味を推定して登録することで，音声認識の認識率を改善することやユーザ特有の言い回しを学習することができる．

4.2.1 登録語の検出

ディクテーションで音声認識をすると，Table 2 のように，第一位の認識結果以外に，それぞれの単語の代替解釈を取得することができる．第一位の認識結果と代替解釈から登録語を検出し，どちらからも登録語が検出されないものを未登録語とする．Table 2 の場合,「五」「本」「茶」が未登録語である.「五」に関しては代替解釈の「の」が登録語であるが，助詞が最初に発話されることはないので，未登録語として扱う．

Table 1 登録語とカテゴリーの例

カテゴリー	登録語
物体名	缶，瓶などの商品名（ダカラ，コーラ）
種類	缶，瓶，ペットボトル
動詞	取って，開けて，見て，入れて
相対位置	左，上，後ろ
助詞	を，の，は，が，に

Table 2 「のほほん茶の左です」に対する認識結果（下線部は登録語，括弧内は発音（認識と同じ場合は省略），—は代替解釈が存在しないことを表す）

認識	代替解釈
五（ご）	の（），同（どう），後（ご）
本（ほん）	今（こん），黄金（おうごん）
茶（ちゃ）	地（ち），著（ちょ），所（じょ）
の（）	—
左（ひだり）	—
でした（）	です（），上下（うえした）

Table 3 未知語の種類

種類	対策
登録語の誤認識語	確率モデルと発音類似度を利用
登録語の同義語	確率モデルを利用
雑音	確率モデルを利用

4.2.2 未知語の種類と推定条件

本論文では，未知語として Table 3 のような 3 種類を想定している．登録語の誤認識語を学習すれば，音声認識の認識率を向上させることができ，登録語の同義語を学習することでユーザ特有の言い回しを学習することができる．また，雑音が認識された場合は，雑音であると推定する必要がある．

次に，未知語を推定するための条件について述べる．未知語は，Table 4 のように複数の単語に分割されることがあるため，例えば未登録語が 3 語連続で検出された場合，実際に発話されたのは 1 語か 2 語か 3 語かはわからない．本論文では 1 語の場合について扱う．Table 4 の認識結果 1 のように未登録語が 3 語までなら一つの未知語と考えて推定し，認識結果 2 のように 4 語以上未登録語が連続する場合には二つ以上の未知語と考えて推定は行わない．つまり，未知語を推定するための条件は，未登録語が 4 語以上連続しないことである．

4.2.3 確率モデル

未知語の推定には，Table 5 のような確率モデルを用いる．

ここで，C,C_1,C_2 はカテゴリーを，W,W_1,W_2 は登録語を，S は状況を表す．本論文では，未知語の推定には前後の単語を用いているため，認識された文字列の一番最初，あるいは一番最後の単語が未知語の場合には，その単語の前，あるいは次に登録語が存在せず，推定できない．これを避けるため,「発話の最初」と「発話の終わり」も登録語に含めておく．

状況 S とは画像処理の結果に対して以下の項目をチェックし，あてはまれば 1，あてはまらなければ 0 として，それを順に並べたものである．つまり，画像処理の進行状況を表す．

- 候補が一つに絞れているか

Table 4 未知語の例（下線部は登録語を表す）

ユーザの発話	のほほん茶取って
認識結果 1	五　本　茶　取って
認識結果 2	五　本　茶　おって

Table 5 確率モデル

モデル	意味
$P_{c-pre}(C_1\|C_2)$	C_2 の前に C_1 が発話される確率
$P_{c-next}(C_1\|C_2)$	C_2 の次に C_1 が発話される確率
$P_c(C\|S)$	S のもとで C が発話される確率
$P_{w-pre}(W_1\|W_2)$	W_2 の前に W_1 が発話される確率
$P_{w-next}(W_1\|W_2)$	W_2 の次に W_1 が発話される確率
$P_w(W\|S)$	S のもとで W が発話される確率

- 物体の代表的な色が絞れているか
- 種類が絞れているか
- 上下の棚が絞れているか
- 同じ棚に二つ以上物体が存在しないか

また，この確率モデルは初期値を適当に与えておき，ユーザごとに更新していく．こうすることで，ユーザに応じた推定をすることができる．

4.2.4 カテゴリーの推定

まず，システムは未知語がどのカテゴリーであるかを推定する．Fig. 6 のような認識結果が得られているとすると，$P(C_1,C,C_2|S)$（状況 S のもとで，カテゴリ C_1,C,C_2 がこの順で発話される確率）を最大にするような C を推定結果とする．ここで，$P(C_1,C,C_2|S)$ を以下のように変形する．

$$
\begin{aligned}
&P(C_1,C,C_2|S)\\
&=P(C_1,C|S)P(C_2|C_1,C,S)\\
&=P_c(C|S)P_{c-pre}(C_1|C,S)P_{c-next}(C_2|C_1,C,S)\\
&\simeq P_c(C|S)P_{c-pre}(C_1|C)P_{c-next}(C_2|C) \qquad (1)
\end{aligned}
$$

ここで，$P_{c-pre}(C_1|C,S)$ は S のもとで C が発話され，その前に C_1 が発話される確率，$P_{c-next}(C_2|C_1,C,S)$ は S のもとで C_1,C と連続して発話され，その次に C_2 が発話される確率を表す．4 行目では，次のような近似をしている．

- C の前に C_1 が発話される確率は S に依存しないと仮定して，
  $P_{c-pre}(C_1|C,S)\simeq P_{c-pre}(C_1|C)$
- C_1,C と連続して発話され，その次に C_2 が発話される確率は C_1,S に依存しないと仮定して，
  $P_{c-next}(C_2|C_1,C,S)\simeq P_{c-next}(C_2|C)$

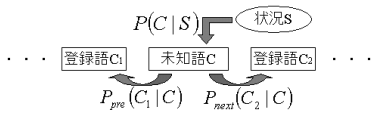


Fig. 6 Estimation of category

このような近似をすることで，$P_c(C|S)$, $P_{c-pre}(C_1|C)$, $P_{c-next}(C_2|C)$ の三つの確率モデルと (1) 式を用いてカテゴリーの推定をすることができる．最大の $P(C_1,C,C_2|S)$ の値が閾値以下であれば，未知語は雑音であると判断して単語の推定は行わない．この場合は C_1,C_2 という語順で解釈する．

4.2.5 単語の推定

次に，未知語がどの登録語と意味が等しいかを推定する．まず，未知語が Table 3 の「登録語の誤認識語」であると仮定して，推定したカテゴリーの中から未知語と最も発音が似ている登録語を検索する．本論文では,「発音類似度」を次のようにして求める．

(1) 未知語の発音を，母音，促音（「っ」），撥音（「ん」）の 7 種類の基礎音に変換する（登録語についてはあらかじめ変換しておく）．
(2) Fig. 7 のようにして一致基礎音数を求める．これは，同じ位置に基礎音がある場合を 1，一つ隣の位置にある場合を 0.5，その他を 0 とした合計値である．基礎音数が異なる場合（Fig. 7(b)）はずらしながら順に計算していき，最も高いものを選ぶ．
(3) 次のようにして発音類似度 R を計算する．

$$R=(1-\frac{\text{基礎音数の差}}{\text{未知語基礎音数}})\times\frac{\text{一致基礎音数}}{\text{登録語基礎音数}}$$

例えば Fig. 7(a) の発音類似度は次のようになる．

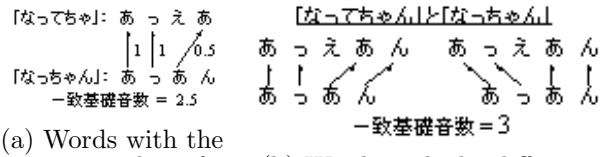
$$R=(1-\frac{0}{4})\times\frac{2.5}{4}=0.625$$

R が 0.6 以上であれば二つの単語は似ているとする．R が 0.6 以上で最大値をとる登録語を推定結果とする．

もし推定したカテゴリーの中の登録語に対する発音類似度がすべて 0.6 未満であった場合には，未知語の前後に助詞がないかをチェックする．助詞の認識結果は他のカテゴリーに比べて信頼度が低いので，前後に助詞が存在する場合には助詞の代替解釈を採用し，再度カテゴリー推定を行う．これについては **4.2.6** で詳しく述べる．

前後に助詞が存在しない場合は，推定したカテゴリーが物体名かどうかチェックする．物体名でなければ，未知語を Table 3 の「登録語の同義語」であると仮定して，確率モデルを用いた単語推定を行う．これは，カテゴリー推定と同様に $P(W_1,W,W_2|S)$（W は未知語，W_1 は W の前の単語，W_2 は W の次の単語）を最大にする W を求める．$P(W_1,W,W_2|S)$ は，**4.2.4** の $P(C_1,C,C_2)$ と同様の変形，近似をする．

推定したカテゴリーが物体名である場合には，単語推定は行わない．例えばユーザが「コーラ取って」と言ったときにコーラが認識されず，推定結果が「ダカラ」（カテゴリーは物体名）であったとすると,「ダカラを取りますか？」より「何を取りますか？」とユーザに質問した方がユーザにとって煩わしくないからである．



(a) Words with the same number of basic phoneme　(b) Words with the different number of basic phoneme

Fig. 7 The number of matched basic phoneme

4.2.6 助詞が誤認識された場合の処理

4.2.5 で述べたように，助詞は誤認識されやすく，未知語の前後に助詞が発話されている場合には，このことが原因で推定を誤ることがある．これを避けるため，未知語の前後に助詞が存在するときには，以下のような処理を行う．

(1) 第一位の認識結果を採用する．
(2) 採用した助詞をもとに未知語のカテゴリーの推定を行う．単語推定の段階で発音類似度が閾値を越えない場合は採用した助詞の認識が間違っているものとし，次に信頼度の高い代替解釈を採用する．
(3) 発音類似度が閾値を越えるまで (2) を繰り返す．すべての代替解釈に対して発音類似度が閾値を越えない場合は第一位の認識結果を採用し，確率モデルを用いた単語推定を行う．

助詞の誤認識の例を Table 6 に示す．この例では未知語は「のって」だけであり，**4.2.4** で述べた方法でカテゴリーを推定すると「相対位置」となり，誤った推定結果となる．この誤推定は，直前の「を」が誤認識されて「の」になったために起こる．また,「相対位置」の中には「のって」に発音が似た単語は存在しない．このような場合には，直前（あるいは直後）の助詞が誤認識されているものとして代替解釈を採用する．Table 6 の例では,「の」の次の代替解釈として「を」が存在する．これを採用して再び「のって」のカテゴリー推定を行うと,「動詞」となる．続いて「動詞」の中から発音の似ているものを検索すると,「のって」は「とって」であるという推定結果が得られる．このようにして，代替解釈をうまく利用することで誤推定を避けることができる．

4.2.7 ユーザへの確認

推定結果は正しいとは限らないので，登録語に追加する前にユーザに推定結果が正しいかどうかを確認する必要がある．確認文として適切なものは，以下の条件を満たしていると考えられる．

(1) なるべく短い
(2) システムが理解している情報も含まれている

Table 6 「コーラを取って」に対する認識結果（下線部は登録語，括弧内は発音（認識と同じ場合は省略），—は代替解釈が存在しないことを表す）

認識	代替解釈
コーラ（こーら）	—
の（）	を（），は（），が（が），＝（わ）
のって（）	—

例えば,「青い缶の後ろの缶を取って」というユーザの発話に対して,「後ろ」が未知語として認識されたとする.「後ろですか？」と確認をした場合，おそらくユーザは「はい」と答える．しかし，ユーザにはどの単語が未知語かはわからないので，実は「缶」を推定した結果「後ろ」になり，それを確認しているのと区別がつかない．ユーザが気付かないうちに「缶」を「後ろ」と学習してしまう恐れがあるため，このような確認文は不適である．しかし「青い缶の後ろの缶を取りますか？」と質問をすると，どの部分を推定したのかがわかりにくい上，システムの発話が長いのはユーザにとっては煩わしい．

そこで本論文では，あらかじめ各カテゴリーがどのカテゴリーとの結びつきが強いかを定義しておき，結びつきの強い単語だけを含めて確認文を生成する．上の例では,「後ろ」すなわち「相対位置」が未知語となっているが，これは直前の「種類」と結びつきが強い．よって，確認文は「缶の後ろですか？」となる．これは上の二つの条件を満たしている．

4.2.8 未知語の追加

ユーザに確認を得られた未知語は，登録語に追加することで学習することができる．この場合，正しい発音を追加することが望ましいが，正しい発音を得ることは困難であるので，認識結果を追加する．また，実際の認識結果を追加する方が，それ以降の認識率は向上する．しかし Table 2 のように，未知語の認識結果は信頼性が低いため，第一位の認識結果だけを登録語に追加しても，認識率は向上しない．このため未知語を追加する際には，候補を複数作り，それらすべてを登録語に追加する．候補は次のようにして生成する．

(1) 各単語の第一位の認識結果，第一位の代替解釈，第二位の代替解釈，と順に調べ，発音が異なるものを三つ検出する（三つ存在しない場合は一つ，あるいは二つでよい）.
(2) 未知語が複数の未登録語に分割されている場合は，語順を変えずにそれぞれの発音をつなげる.

例えば Table 7 のような認識結果が得られたときには，候補は「ちょうだい」「ちょうだいり」「ちょうだいい」の三つとなる.

4.2.9 未知語の推定結果

研究室内の被験者 5 人に対し，冷蔵庫画像を見せてシステムと対話してもらい，未知語の推定実験を行った.

Table 7 「ちょうだい」に対する認識結果
（括弧内は発音を表す）

認識	代替解釈
釣（ちょう）	朝（ちょう），長（ちょう），超（ちょう），鳥（ちょう），肇（ちょう），…
大（だい）	代（だい），第（だい），代理（だいり），代位（だいい），…

実験には，登録語 104 語，カテゴリー 16 個，文法 245 個を用い，不特定話者で音声認識をした．その結果，以下のような未知語を獲得できた（未知語は複数の候補のうち，代表的なもののみで，括弧内は推定登録語を表す）.

登録語の誤認識語：こうろ（コーラ），なかちゃん（なっちゃん），かくえりあ（アクエリアス），いだい（左），にいえ（いいえ）

登録語の同義語：おく（後ろ），ちょうだい（取って），そうです（はい）

実際の対話の成功例，失敗例を Tables 8~12 に示す. Tables 8, 9 に関しては，正しい推定結果が得られている.

Table 10 は，二つの登録語が三つの未登録語として認識された例である．システムは，未知語が一つであるものとして推定している（カテゴリー推定が「物体名」

Table 8 成功例 1
（下線部は登録語を表す）

システム	見つかりませんでした. どのへんにありますか？
ユーザ	のほほん茶の左
認識	五 本 茶 の 左
カテゴリー推定	物体名
単語推定	のほほんちゃ
システム	のほほん茶の左ですか？

Table 9 成功例 2
（下線部は登録語を表す）

システム	二つ見つかりました. どちらにしますか？
ユーザ	赤いキャップがついてる方
認識	破壊 キャップ が ついてる 方
カテゴリー推定	色
単語推定	あかい
システム	赤いキャップですか？

Table 10 失敗例 1

ユーザ	コーラ取って
認識	王 だ と
カテゴリー推定	物体名
単語推定	（しない）
システム	何ですか？

Table 11 失敗例 2

ユーザ	ダカラ取って
認識	だからといって
カテゴリー推定	物体名
単語推定	（しない）
システム	何ですか？

Table 12 失敗例 3

システム	二つ見つかりました． こちらでいいですか？
ユーザ	（発話せず）
認識	ううう
カテゴリー推定	相対位置
単語推定	うえ
システム	上ですか？

であるため，単語推定は行わない）．Table 11 は，二つの登録語が一つの未登録語として認識された例である．Table 10 の例と同様に，未知語が一つであるものとして推定している．Table 12 に関しては，ユーザの息，あるいは周囲の雑音を誤って認識されたケースである．ユーザの発話の間に雑音が入った場合は，カテゴリー推定の段階で $P(C_1,C,C_2|S)$ が閾値を越えずに雑音と判断されるが，Table 12 の例のように，雑音だけが認識された場合については 1 語文（「はい」など）と区別がつかないので，現在のところ対応できていない．

5. まとめと今後の課題

本論文では，サービスロボットに必要な対話システムについて述べた．サービスロボットには物体を認識する機能が必要であるが，自動で認識できないときには，ユーザとの音声による対話から必要な情報を得ることで認識を可能にした．対話からは，物体の位置に関する情報や物体の色に関する情報などが得られるが，本論文ではこれらの情報を，認識が失敗したときに対処できるように用いている．

また本論文では，音声認識の認識率を向上させるため受け付ける単語を限定している．しかし，ユーザは受け付ける単語を覚える必要があり，これはユーザにとって負担である．そこで，登録語以外の単語が発話されたときには，登録語の誤認識語，登録語以外の単語，雑音の 3 種類を想定して，確率モデルと発音類似度を用いて意味を推定し，学習する．このようにすることで，ユーザの負担を減らし，音声認識の認識率を向上させることができる．

今後の課題としては，以下の三つが挙げられる．

(1) 未知語が連続する場合の処理
Tables 10, 11 で示したように，二つ以上の連続した登録語が複数の未登録語として認識されたり一つの未登録語として認識されることがある．このような場合に，現在は未知語が一つであるとして推定しているが，未知語が連続することを考慮に入れた推定を行う必要がある．

(2) 雑音に対する処理
Table 12 で示したように，雑音のみが認識された場合に雑音であると推定することができない．このような場合に対しても，雑音であるという推定を行う必要がある．

(3) 追加した未知語の候補の削除
4.2.8 で述べたように，未知語を追加する際には候補を複数追加する．現在のところ，未知語の追加による認識率の低下は見られないが，今後追加数が増えると悪影響を及ぼすことが考えられる．これを避けるためには，不要な候補を削除する必要がある．

参 考 文 献

[1] Y. Takahashi, T. Komeda, T. Uchida, M. Miyagi, H. Koyama: Development of the mobile robot system to aid the daily life for physically handicapped; *Proc. of ICMA2000*, pp. 549-554 (2000)

[2] U. Ahlrichs, J. Fischer, J. Denzler, C. Drexler, H. Niemann, E. Noth, D. Paulus: Knowledge based image and speech analysis for service robots; *Workshop on Integration of Speech and Image Understanding* (1999)

[3] 渡辺, 長尾, 岡田：画像の内容を説明するテキストを利用した画像解析；画像の認識・理解シンポジウム（MIRU'96）, Vol. 2, pp. 271-276 (1996)

[4] S. Wachsmuth, G. Sagarer: Connecting concepts from vision and speech processing; *Workshop on Integration of Speech and Image Understanding* (1999)

[5] 高橋，中西, 久野，白井：音声とジェスチャによる対話に基づくヒューマンロボットインターフェース；インタラクション '98 論文集, pp. 161-168 (1998)

[6] 藤井，杉山：歩行者ナビゲーション支援のための場所案内文生成手法；電子情報通信学会論文誌, Vol. J82-D-II, No. 11, pp. 2026-2034 (1999)

[7] 岩田，鬼沢：絵のつながりを考慮した絵情報の言語的表現；電子情報通信学会論文誌, Vol. J84-D-II, No. 2, pp. 337-350 (2001)

[8] 伊藤，速水，田中：音声対話システムにおける未知語の扱い；人工知能学会研究会資料, SIG-SLUD-9201-1, pp. 1-9 (1992)

[9] 高橋，堂坂，相川：音声対話における実例に基づく未知語属性推定；電子情報通信学会技術報告, NLC2001-35, pp. 101-106 (2001)

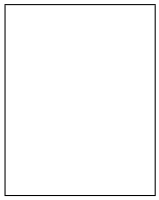
[10] G. Damnati, F. Panaget: Adding new words in a spoken dialogue system vocabulary using conceptual information and derived class-based LM; *Proc. of Workshop on Automatic Speech Recognition and Understanding* (1999)

[11] Y. Makihara, M. Takizawa, Y. Shirai, J. Miura, N. Shimada: Object recognition supported by user interaction for service robots; *Proc. of 5th ACCV*, Vol. 2, pp. 719-724 (2002)

[12] 槙原，滝澤，二ノ方，白井，三浦，島田：必要に応じてユーザと対話しながら行動するロボット -対話を利用した物体の認識と操作-；ロボティックス・メカトロニクス講演会 2001 論文集 (2001)

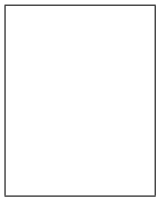
著 者 略 歴

滝澤 正夫 (非会員)



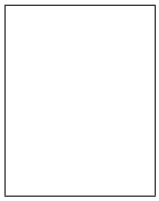
1978年8月31日生．2001年大阪大学工学部応用理工学科卒業，同年4月同大学院工学研究科電子制御機械工学専攻修士課程に進学し，現在に至る．人との対話を用いたサービスロボットのための音声対話に関する研究に従事．日本ロボット学会会員．

槙原 靖 (非会員)



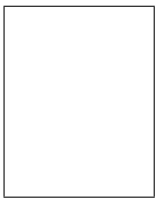
1978年6月1日生．2001年大阪大学工学部応用理工学科卒業，2002年9月同大学院工学研究科電子制御機械工学専攻修士課程修了．同年10月同大学院博士課程に進学し，現在に至る．人との対話を用いたサービスロボットのための物体認識に関する研究に従事．2001年度日本機械学会ロボティックス・メカトロニクス部門ベストプレゼンテーション表彰受賞．日本ロボット学会，日本機械学会の会員．

白井 良明 (非会員)



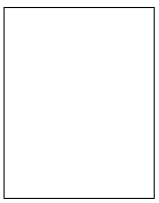
1941年8月3日生．1969年東京大学大学院工学系機械工学専攻博士課程修了．同年4月電子技術総合研究所研究官，研究室長，部長となる．その間，1971年8月より1年間米国ＭＩＴの客員研究員．1988年大阪大学工学部教授となり，現在に至る．その間，1996年4月より3年間，東京大学大学院工学研究科教授併任．知能ロボットに関する研究に従事．工学博士．1975年 Pattern Recognition Society Award，1983年，1994年電子通信学会論文賞．情報処理学会，電子情報通信学会，人工知能学会などの会員．

島田 伸敬 (非会員)



1969年生．1992大阪大学工学部電子制御機械工学科卒，1997同大大学院工学研究科電子制御機械工学専攻博士後期課程了．博士 (工学)．同年同専攻助手，2001年同研究科研究連携推進室情報ネットワーク部門講師．コンピュータビジョン，ジェスチャ認識，ヒューマンインターフェース，インターネットソリューションの研究に従事．情報処理学会，電子情報通信学会，IEEE各会員．

三浦 純 (正会員)



1984年東京大学工学部機械工学科卒業．1989年同大学院工学系研究科情報工学専攻博士課程修了，工学博士．同年大阪大学助手．現在同大学院工学研究科電子制御機械工学専攻助教授．知能ロボット，人工知能，コンピュータビジョンの研究に従事．1994年～1995年，CMU客員研究員．1997年日本ロボット学会論文賞受賞．日本ロボット学会，人工知能学会，電子情報通信学会，情報処理学会，日本機械学会，IEEE，AAAI各会員．