

ロボット発話の冗長性制御による 指示物体認識性能向上への寄与

飯尾 尊優^{1,a)} 塩見 昌裕¹ 篠沢 一彦¹ 下原 勝憲² 萩田 紀博¹

受付日 2011年6月26日, 採録日 2012年1月13日

概要: 本稿では, 人間がロボットに対して環境中の物体を繰り返し指示するような状況において, その人間が指示した物体を認識する際の認識性能がロボットの発話の冗長性を制御することにより向上することを示す. 人間とロボットの物体指示対話において, (1) 人間はロボットが確認発話で使用する言葉と同じ言葉を指示発話で使用する傾向がある, (2) ロボットが冗長性のある確認発話を行うと人間は冗長性のない指示発話を行う傾向がある, という知見がある. これらの知見は, ロボットが物体の識別に関係ない情報を含む冗長な発話を行うと, 人間は不要な情報で構成され, かつ冗長性もない指示発話を行うようになり, 指示物体認識性能の低下を招く可能性があることを示唆している. 本研究では, ロボットが指示物体の識別に必要な最小の情報で構成される冗長性のない確認発話を行う場合と, 指示物体の識別に利用可能なすべての情報で構成される冗長性のある確認発話を行う場合で, 指示物体認識性能が変化するかどうか検証した. その結果, 指示物体の認識成功率は, 冗長性のない確認発話を行う場合で 75.6%, 冗長性のある確認発話を行う場合で 54.7%であり, この差は有意であった. つまり, ロボットが発話の冗長性を制御することで, 指示物体認識性能が変化することが示された.

キーワード: ヒューマンロボットインタラクション, 指示物体認識, 発話の冗長性制御

Contribution to Performance of Object Reference Recognition by Redundancy Control of Robot Speech

TAKAMASA IIO^{1,a)} MASAHIRO SHIOMI¹ KAZUHIKO SHINOZAWA¹
KATSUNORI SHIMOHARA² NORIHIRO HAGITA¹

Received: June 26, 2011, Accepted: January 13, 2012

Abstract: This paper reports that the performance of object reference recognition is improved through human-robot interaction. In human-robot interaction, the following two phenomena have been observed; (1) people tend to use the same words as robots use, and (2) people tend to use non-redundant words when robots use too redundant words. These findings suggest that if robots confirm by using redundant words which do not contributed for identification of an object, people may tend to use non-redundant references which consist of the unnecessary information for the identification. Such phenomena would change the performance of object reference recognition with a speech recognition technique. To investigate whether the redundancy have effects for object reference recognition, we conducted an experiment with two conditions. One was a condition that a robot confirmed by using minimum requisite attributes for identification of an object. Another was a condition that a robot confirmed by using all attributes of an object. As a result, the success rates of object recognition of former condition and latter condition are 75.6% and 54.7%, respectively. Moreover, the statistical comparison shows significant difference, i.e., using minimum requisite attributes shows high success rate than the case of using all attributes. Thus, the results suggested that the performance of object reference recognition was improved through redundancy control of robot speech.

Keywords: human-robot interaction, object reference recognition, redundancy control of robot speech

1. はじめに

人間が指示した物体を認識する技術は、日常環境下で人々へのサービス提供を行うロボットにとって、インタラクションの基盤となる技術の1つである。すでにロボットは、科学博物館内での展示案内 [1] や、商業施設内での道案内 [2] など、一般の人々へサービスを提供するために用いられている。また、買い物支援 [3] や物品の運搬 [4] といった、身体性を利用したサービスを提供するロボットの研究開発も進んでいる。このような、実環境下での物理的なサービスを提供する状況においては、ロボットは人間の指示に応じて特定の物体を運搬したり、物体に関する情報を提供したりする必要がある。つまり、人々とインタラクションを行うロボットにとって、人間が指示した物体を精度良く認識する技術が必要不可欠となる。

指示物体を認識する手法に関しては、画像処理によって認識された視線や指差し動作などの非言語情報を利用する手法 [5], [6], [7] や、それらと音声などの言語情報とを組み合わせた手法 [8], [9], レーザポインタを利用する手法 [10] が開発されている。これらの従来研究は指示物体認識システムの性能を向上させるために、新しい機器やアルゴリズムの開発に焦点を当ててきた。たとえば Nickel ら [6] は、実時間で指差し動作を認識し指差し方向を推定するシステムを開発することを目的とし、指差し動作をモデル化するため、頭と手の位置の3次元情報および頭の向きを利用する手法を提案した。Schauerte ら [8] は、高精度な指示物体認識を実現するために、音声認識と画像処理による指差し方向の認識とを組み合わせる手法を提案した。また Kemp ら [10] は、実環境で人間が直感的に3次元的に位置を指し示すことができる新しいインタフェースを開発することを目的とし、市販のレーザポインタを利用する手法を提案した。

さらに近年では、上記のような工学的技術だけでなく、人間とロボットのインタラクションの繰返しを通じて、人間の行動がロボットの行動に同調するようになる、という認知心理的な現象を利用することで、指示物体の認識性能が向上する可能性が示唆されている。三谷ら [11] は、人間とロボットが計算機の画面上でボール遊びを行うというインタラクションにおいて、対話を繰り返すうちに実験参加者がロボットと同一の発話をするようになる傾向があったと述べている。そして、三谷らはこのようなロボットの発話による人間の発話の誘導が人間の意図理解に役立つと述べている。また飯尾ら [12] は、人間がロボットに対し環境

中の物体を繰り返し指示するというインタラクションにおいて、人間は、ロボットが指示された物体を確認する際に使用した言葉と同じ言葉を使用する傾向があることを明らかにした。そして、この現象を利用することで、指示物体の認識性能が向上する可能性について論じている。このように、インタラクションの繰返しを通じて人間の指示発話を変化させることで、指示物体の認識性能の向上に有利な状況を作り出すという手法は、特別な機器を必要とせず、また工学的な手法とは独立に利用できることから、有用性が高いと考えられる。

しかしながら、今までに、インタラクションの繰返しを利用することで実際に指示物体の認識性能の向上を実現した研究結果は報告されていない。そのため、どのようにインタラクションを利用することで、どの程度指示物体認識の性能を向上させることができるのか、という疑問に明確な回答は与えられていない。この疑問を解決し、インタラクションの繰返しを通じて指示物体の認識性能を向上させる手法を実現することは、人間とロボットのインタラクションの応用技術の発展に寄与すると考える。

そこで本稿では、工学的な手法を用いて指示物体認識システムの性能向上を図ってきた従来研究とは異なり、人間とロボットのインタラクションの繰返しにおいて生じる現象を利用して指示物体認識システムの性能が向上させられるかどうかを明らかにすることを目的とする。具体的には飯尾ら [12] が行った人間がロボットに対し環境中の物体を繰り返し指示するというインタラクションにおいて、ロボット発話の冗長性が人間の指示発話を変化させる現象に着目し、人間に指示された物体を認識する際の認識性能がロボット発話の冗長性を制御することで向上する可能性について検証する。以下の章では、まず、人間が人間やコンピュータやロボットと対話するときに生じる現象について説明し、指示物体の認識性能を向上させるためのロボットの確認発話の仕方について述べる。3章では、確認発話の仕方による指示物体の認識性能の違いを検証するために構築した実験システムについて述べる。そして4章では、その実験システムを用いた確認発話の仕方を要因とする対照実験について説明し、5章でその結果を述べる。6章で結果の考察を行い、7章で本研究の結論を述べる。

2. 指示物体認識におけるインタラクションの利用

指示物体の認識性能の向上に関する研究は広く行われているが、それらの研究は認識技術そのものの開発や発展に注目しており、人間とロボットのインタラクションを性能の向上に利用しようと試みるものではなかった。本章では、まず人間同士の対話や人間とロボットの対話において観測されている2つの現象—発話の同調傾向、冗長な発話における同調傾向の低減—について述べる。さらに、これ

¹ ATR 知能ロボティクス研究所
ATR Intelligent Robotics and Communication Laboratories,
Souraku, Kyoto 619-0288, Japan

² 同志社大学
Doshisha University, Kyotanabe, Kyoto 610-0321, Japan

a) iio@atr.jp

らの現象を考慮し、指示物体の認識性能を向上させるためのロボットの確認発話の仕方について述べる。

2.1 発話の同調傾向

人間同士の対話において、互いの発話が同調する現象が知られている。これまでに、話速 [13], 単語 [14], [15], 構文 [16], [17], 表現 [18] など様々なレベルでの発話の同調が観測されている。たとえば, Brennan ら [14] は, 視覚的に隔たれた 2 人が, 互いの机においてある絵の描かれた複数のカードを同じ順番に並べ替えるというタスクを使って実験を行った。その結果, その 2 人が繰り返しカードを参照するにつれて, 同じ語を使用する傾向があることを明らかにした。

人間同士の対話だけでなく, コンピュータやロボットとの対話における相互作用についても研究が行われている。Branigan ら [19] は, 参考文献 [17] で明らかにされた, 人間同士の発話構文の同調が人間とコンピュータの対話でも起こるかどうかを検証した。この実験では 2 つの条件が設定されていた。1 つはコンピュータ対話条件, もう 1 つは人間対話条件である。コンピュータ対話条件では実験参加者はコンピュータとゲームをすると教示され, 人間対話条件では, 実験参加者は人間とゲームをすると教示された。しかし, 実際はどちらの条件でも実験参加者は同じコンピュータプログラムとゲームをしていた。実験の結果, 実験参加者はどちらの条件でも対話者 (コンピュータプログラム) が使用した構文と同じ構文を使用する傾向があった。

人間とロボットのインタラクションにおける相互作用では, 飯尾ら [12] が, 人間とヒューマノイドロボットが周囲におかれた物体を参照するときに, 人間の発話がロボットの確認発話によってどのように変化するかを調べた。その結果, 人間は物体を指示するとき, ロボットが確認発話で使った言葉と同じ言葉やその言葉のカテゴリと同じカテゴリに含まれる言葉を使用する傾向があることを示した。

これらの研究結果から, 人間の発話は, 対話者が人間でない存在であっても, その対話者の発話に同調する傾向があるということが明らかにされた。これは, 人間の発話内容をロボットの認識機能に有用な言葉に同調させることができる可能性を示唆している。すなわち, インタラクションを利用することで指示物体の認識性能は向上する可能性があると考ええる。

2.2 冗長性による同調傾向の低減

発話の同調傾向は, 対話者同士の発話が様々なレベルで類似する傾向にある, という現象である。しかし, 過去の研究では, この現象がつねに成立し続けるわけではないことが示唆されている。たとえば, 大坊 [20] は, 人間同士の 2 者間の対話において個体の発言量は時間が経つにつれて, 負の相関が見られるようになるということを報告してい

る。大坊はこの現象について, 会話における発言量は会話の進行につれて一定の量に近づくため, もし一方の発言が増大すれば, 相手の発言量は自然な会話を進めるために抑制的になるのではないかと考察している。すなわち, 発話が冗長になるにつれて, 同調傾向が抑制される可能性があることを指摘している。

人間とロボットの対話においても, 上記のような同調傾向の抑制に関する報告がある。Shinozawa ら [21] は, 人間がロボットに本を取ってくるように依頼するというタスクを設定し, 人間がどのように本を指示するかを調べた。この実験では, ロボットが指示された本を確認する方法として, (1) 本のタイトルを使って確認する, (2) 本のタイトルと指差し動作を使って確認する, (3) 指示語と指差し動作を使って確認する, という 3 種類の確認方法を行った。実験参加者が本を一意に識別できる言葉と指差し動作の両方を使用した回数, すなわち冗長性のある指示発話を行った回数を測定した結果, 2 番目の確認方法において, 最もその数が少なかった。この結果から Shinozawa らは, ロボットが冗長性のある確認発話を使うと, 人間は冗長性のない指示発話を使う傾向がある, と結論している。

これらの研究結果は, 人間の発話が対話者のどのような発話に対しても同調傾向を示すわけではないことを示唆している。すなわち, 対話者の発話の冗長性に応じて, 自身の発話の冗長性を調整している可能性があると考ええる。

2.3 インタラクションを利用したロボットの発話

指示物体認識性能を向上させるという観点から見て望ましい指示発話は, 物体の識別に利用可能な情報が多く含まれる冗長性のある指示発話であろう。なぜなら, 音声認識で発話区間の切り出しを失敗したり, 雑音などの影響で部分的に物体の特徴を誤認識したりする可能性があるためである。冗長性のある指示発話の方が, このような誤認識に対して頑健であると考えられる。

人間にそのような冗長性のある指示発話をしてもらうために, ロボットはどのような発話をすべきだろうか。先に示した物体指示対話における 2 つの知見を要約すると次のようになる。

- 人間はロボットが確認発話で使用する言葉と同じ言葉を指示発話で使用する傾向がある。
- ロボットが冗長性のある確認発話を行うと, 人間は冗長性のない指示発話を行う傾向がある。

最初の同調傾向に関する知見によれば, ロボットが物体の識別に利用可能な情報を含む確認発話を行うことで, 人間も同様に物体の識別に利用可能な情報を含む指示発話を行うようになると考えられる。一方で, 2 つめの冗長性に関する知見によれば, ロボットが物体の識別に利用可能な情報を多く含む確認発話を行うと, 人間は物体の識別に利用可能な情報をあまり含まない指示発話を行う傾向が増

える。

以上の考察は、ロボットが物体の識別に関係ない情報を含む冗長な発話を行うと、人間は不要な情報で構成され、かつ冗長性もない指示発話を行うようになり、指示物体認識性能の低下を招く可能性があることを示唆している。したがって、人間の指示発話から指示物体を認識する場面では、ロボットは指示物体の識別に必要な最小の情報のみで構成される確認発話を行うことが望ましい、と考えられる。ロボットの物体の識別に利用可能な情報を含む確認発話に人間の指示発話を同調させ、かつ、ロボットの提示する情報を最小限にして人間の指示発話の冗長性の低下を抑制することで、指示物体認識性能を向上させられる可能性があると考えられる。

3. 実験システム

ロボット発話の冗長性制御による指示物体の認識性能向上の可能性を検証するための実験システムを構築した。実験システムは指示物体認識と確認発話決定の2つの機能から成り立つ。各機能の設計方針および具体的なシステム構成について説明する。

3.1 インタクションのシナリオ

本実験では、実験参加者とロボットが環境中に配置されている物体を参照するというタスクを設定した。このタスクでは、図1に示す手順に従って実験参加者とロボットが

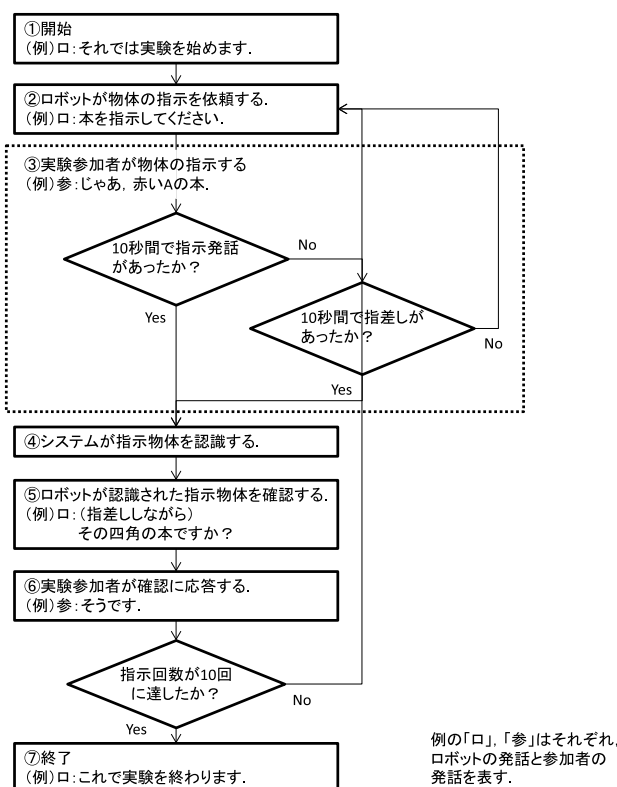


図1 本実験システムで想定されるインタクションのシナリオ
Fig. 1 Scenario of interaction in our experimental system.

インタクションを繰り返した。最初にロボットは実験参加者に物体を1つ指示するように求め、システムは10秒間の指示発話の音声入力待ち状態になる。この間に実験参加者は複数の物体の中から自由に1つを指示した。実験参加者が指示した物体をシステムが認識し、ロボットが認識された指示物体を実験参加者に確認した。仮に、実験参加者が指示する本の選択を迷うなどして、指示発話が10秒間なされなかった場合、ロボットは再び実験参加者に物体の指示を依頼する。また、無言で指差しだけが行われた場合には、10秒間の指示発話の音声入力待ち状態終了後、指差し認識の結果に基づいて指示された物体を認識し、ロボットがその指示物体を実験参加者に確認するようにした。実験参加者はその確認に答えると、確認の正否にかかわらず、ロボットが再び物体の指示を依頼した。このインタクションが10回繰り返された。

3.2 設計方針

3.2.1 指示物体認識

本システムでは、音声認識と指差し認識を組み合わせる指示物体認識を行った。これは、環境内の物体の特徴や位置関係に応じて、言葉のみまたは指差し動作のみの指示発話がありうると思ったためである。たとえば、指示対象が他の物体と離れている場合、その指示対象は指示語と指差し動作だけで指示される可能性がある。一方、指示対象が他の物体と近い距離にある場合、指示語と指差し動作だけで指示対象と他の物体を区別することは難しいため、その指示対象はその名称や色や形などの言語表現を用いて指示される可能性がある。このように、言葉と指差し動作の使用状況が変化する可能性があるため、音声認識と指差し認識とを組み合わせる指示物体認識機能を実現した。

3.2.2 確認発話

本システムでは、人間に指示された物体をロボットが確認するときの発話の冗長性を制御した。確認発話に注目した理由は、物体指示対話における確認行為が人間同士でも頻繁に見られる自然な行為であるためである。たとえば、環境内に複数の雑誌が存在する状況で、「その雑誌を取ってくれ」と指示された場合を考える。複数の雑誌が密集していて、一意に雑誌を特定することが困難であった場合、頼まれた人は「情報処理学会誌の4月号ですか？」などと確認をすると考えられる。また、指示された雑誌が明らかな場合でも、人間は念のために「これですか？」などと簡単な発話を通じて確認をすることもあるだろう。このように物体指示対話において確認発話を行うことは自然な行動であると考えたため、ロボットが発話による確認を行うようにシステムを設計した。

指示物体に対する確認発話の選択方法について、我々は2種類の方法を設計した。1つは、2.3節で提案したロボット発話の方針に従った確認発話、すなわち、指示物体の識

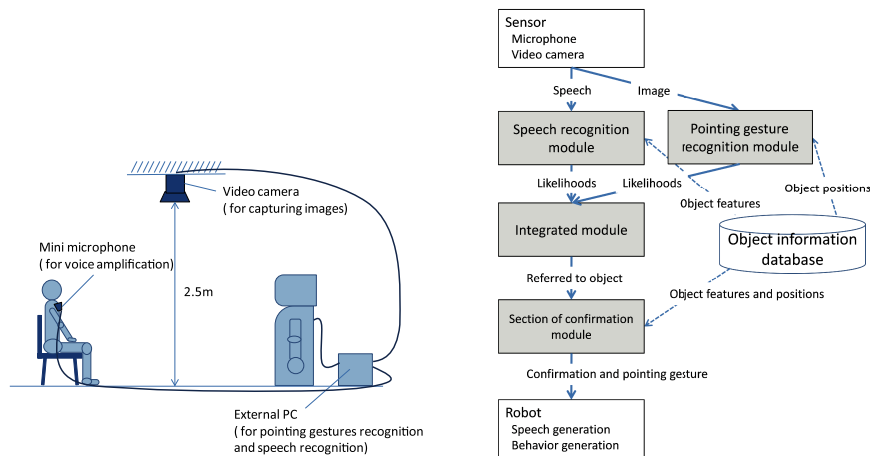


図 2 ハードウェア構成 (左) とソフトウェア構成 (右)

Fig. 2 Hardware configuration (left) and software configuration (right).

別に必要な最小の情報を用いるように確認発話を選択する方法であり、もう1つは、指示物体の識別に関するすべての情報を用いるように確認発話を選択する方法であった。

3.2.3 確認動作

本システムでは、本の確認時にロボットに指差し動作を行わせるようにした。ロボットに指差し動作をさせた理由は2つある。1つめは、実験参加者にロボットがどの本を確認しているのかを分かりやすく伝えるため、2つめは、ロボットの視線と指差し動作が人間の指差し動作を引き込む可能性があるためである [22]。人間の指差し動作は指示物体を認識するうえで有用な情報となるため、実験参加者が指差し動作を使用する割合を増やすように、ロボットは視線と指差し動作を行った。指差し動作は物体の座標を指定することで、その座標に向かって手を伸ばし、顔をその座標に向けることで実現した。

3.3 ハードウェア構成

本システムは、実験参加者と対話するロボット、指示発話に関する情報を取得するセンサ、センサ情報を処理するための外部 PC から構成された。本システムのハードウェア構成を図 2 の左に示す。

実験参加者と対話するロボットとして、Robovie-R ver.2 を使用した。Robovie-R ver.2 は ATR 知能ロボティクス研究所で開発されたヒューマノイドロボットで、人間とのコミュニケーションのために、人間らしい上半身を持つ。頭部は3自由度、両腕はそれぞれ4自由度を持つ。頭部にはスピーカがあり、ここから外部 PC で合成された音声を発話する。この自由度によって物体を指示する指差し動作を再現することが可能である。音声合成には、ATR 音声言語コミュニケーション研究所で開発された音声合成システム XIMERA を使用した [23]。身長 1,100 mm、幅 560 mm、奥行き 500 mm で体重は 57 kg である。我々は、付属のソフトウェアを使い、挨拶モーション、指示依頼モーション

を作成した。これらのモーションはそれぞれ最初の挨拶をするとき、本を指示するように依頼するときに合成された音声とともに再生された。

人間の物体指示に関する情報を取得するセンサとして、ビデオカメラと小型マイクロフォンを使用した。ビデオカメラは実験参加者の指差し動作を検出するために実験室の天井に取り付けられた。天井の高さは約 2.5 m であり、このビデオカメラは床の 2.5 m × 3.5 m の範囲を撮影することができた。小型マイクロフォンは実験参加者の発話音声を取得するために用いられた。

外部 PC はセンサのデータから指差し認識と音声認識を行うために用いられた。また、配置された物体の情報を格納するデータベースとしても用いられた。

3.4 ソフトウェア構成

指示物体認識と確認発話選択の機能を実現するソフトウェアは、物体情報データベースと、音声認識モジュール、指差し認識モジュール、統合モジュール、確認発話選択モジュールという4つのモジュールから構成される (図 2 右)。まず、センサから取得された音声と画像は、それぞれ音声認識モジュールと指差し認識モジュールに送られる。そして、この2つのモジュールは物体情報データベースを用いて、配置されている各物体の指示尤度を計算する。指示尤度とは、物体が指示された可能性を示す値であり、指示尤度が高いほど、その物体が指示された可能性が高いとする。音声認識モジュールで計算された指示尤度と指差し認識モジュールで計算された指示尤度は、統合モジュールに送られる。統合モジュールは最終的な指示尤度を計算し、指示物体を推定する。確認発話選択モジュールは物体情報データベースを用いて、その場にある物体から指示物体を識別する最小限の情報を算出し、それに基づいて確認発話を選択する。

3.4.1 物体情報データベース

物体情報データベースには、実験環境中に存在する物体の位置および属性が格納される。ここで、実験環境とは、ビデオカメラで撮影される 2.5 m × 3.5 m の範囲を指す。属性とは、色や形や大きさなど物体の識別に利用可能な情報であり、すべての属性の集合を属性集合 F と定義する。物体 i の属性集合を O_i とし、 F の部分集合として定義する。各物体の属性集合と位置は、音声認識モジュールと指差し認識モジュールが指示尤度を計算するとき、確認発話選択モジュールが確認発話を選択するときに用いられる。

なお、配置された物体の属性と位置は実験者によって手動で物体情報データベースに登録された。

3.4.2 音声認識モジュール

音声認識モジュールでは、音声認識で得られた指示発話の文字列を解析し、音声認識に基づく指示尤度を計算する。本システムでは、音声認識に ATRASR を用いた [24]。ATRASR は背景雑音が 70 dBA の環境下において 90% 以上の精度の音声認識を実現できる。ATRASR はマイクロフォンからの音声を入力として受け取り、その音声に適合する文字列を出力する。

音声認識モジュールは、ATRASR から出力された文字列を解析して、その文字列に含まれる属性を抽出し、指示発話の属性集合を生成する。そして、その指示発話の属性集合と物体情報データベースに格納されている属性集合を比較し、共通要素の個数から各物体の音声認識に基づく指示尤度を計算する。ここで、音声認識に基づく物体 i の指示尤度 S_i の計算を次に示す。

R を指示発話の文字列に含まれる属性集合で F の部分集合として定義する。次に、 R と O_i の積集合を計算し、共通要素の個数を求める。実験環境中の指示対象となる全物体の共通要素数を求め、各物体の共通要素数を全物体の共通要素数の総和で割る。ここで、 $|A \cap B|$ を集合 A と集合 B の共通要素数を求める演算と定義すると、物体の数が n 個ある場合、以上の計算は次式で表される。

$$S_i = \frac{|O_i \cap R|}{\sum_{i=1}^n |O_i \cap R|} \quad (1)$$

音声認識に基づく指示尤度計算の具体例を、図 3 に示

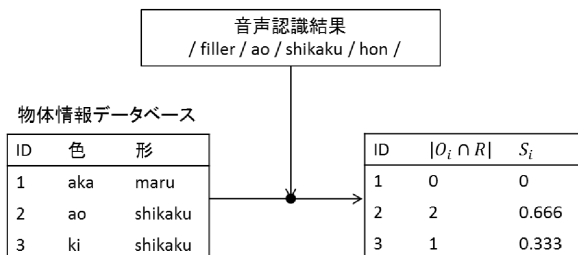


図 3 音声認識に基づく指示尤度の計算の例

Fig. 3 Example of calculation of likelihoods by means of the results of speech recognition.

す。この例では、物体情報データベースに 3 つの物体が登録されている。ここで、「えーっと、青の四角の本」という指示発話が入力されると、正しく音声認識できた場合、ATRASR は「/ filler / ao / shikaku / hon /」という文字列を出力する。なお、「filler」や「hon」など属性集合の要素でない文字列は省かれる。この文字列から抽出された指示発話の属性集合と物体指示データベースに登録されている物体 1, 2, 3 の属性集合の共通要素数は、それぞれ 0, 2, 1 である。したがって、音声認識に基づく指示尤度はそれぞれ 0, 0.666, 0.333 と計算される。正しく音声認識できなかった場合でも、正しく音声認識できた場合と同様に、ATRASR が出力した文字列から指示発話の属性集合を生成し、それを各物体の属性集合と比較して、各物体の音声認識に基づく自由度を計算する。たとえば、先ほどの指示発話を ATRASR が誤認識して「/ filler / filler / shikaku / maru /」という文字列を出力したとする。この場合、物体 1, 2, 3 の属性集合の共通要素数は、それぞれ 1, 1, 1 となり、音声認識に基づく指示尤度は 0.333, 0.333, 0.333 と計算される。

3.4.3 指差し認識モジュール

指差し認識モジュールは、指差し認識で得られた指差し角度に基づいて、指示尤度を計算する。本システムでは、指差し認識のために、肌色抽出を行って指差し角度を出力する画像処理機能プログラムを作成した。このプログラムは、天井に取り付けられたビデオカメラで撮影された画像を入力として受け取り、実験参加者の指差し角度を計算する。指差し角度は、図 4 に示すように、実験参加者が背筋を伸ばして椅子に座ったときの頭の中心位置と指差しをしている手の位置を結ぶ直線から求められる角度と定義する。頭の中心位置は、実験参加者が最初椅子に座ったときに天井カメラの映像から実験者によって手動で計測され、実験中はその計測値が一貫して用いられた。手の位置は、入力画像から肌色抽出によって求められた。この手の位置が実験参加者を中心とする縦横 50 cm の範囲外にあったときに、指差しがあったと見なした。

指差し角度 α_h は、実験参加者の頭の中心位置 (x_c, y_c) と指差し認識で得られた手の位置 (x_h, y_h) から次式で計算される。

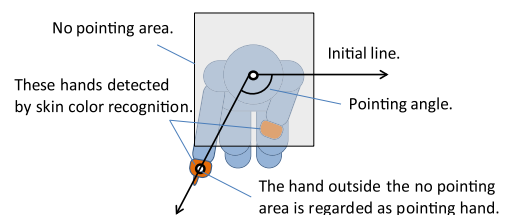


図 4 指差しの角度

Fig. 4 Angle of a pointing gesture.

$$\alpha_h = \begin{cases} \arctan\left(\frac{y_h - y_c}{x_h - x_c}\right), & (x_h - x_c > 0) \\ \frac{\pi}{2}, & (x_h - x_c = 0, y_h - y_c > 0) \\ \frac{3\pi}{2}, & (x_h - x_c = 0, y_h - y_c < 0) \\ \arctan\left(\frac{y_h - y_c}{x_h - x_c}\right) + \pi, & (x_h - x_c < 0) \end{cases} \quad (2)$$

α_h の範囲を 1 回転にするために上記のような場合分けをした．この場合分けにより α_h の値域は $(-2/\pi, 3\pi/2]$ となるが，実験環境の設定上，物体は実験参加者の正面にあるため，実際に得られる α_h の値域は $(0, \pi)$ となる．また，実験参加者から物体 i までの角度 α_i は，物体 i の位置 (x_i, y_i) から次式で計算される．

$$\alpha_i = \begin{cases} \arctan\left(\frac{y_i - y_c}{x_i - x_c}\right), & (x_i - x_c > 0) \\ \frac{\pi}{2}, & (x_i - x_c = 0, y_i - y_c > 0) \\ \frac{3\pi}{2}, & (x_i - x_c = 0, y_i - y_c < 0) \\ \arctan\left(\frac{y_i - y_c}{x_i - x_c}\right) + \pi, & (x_i - x_c < 0) \end{cases} \quad (3)$$

ここで， G_i を物体 i が指示された確率とし， G_i は α_h と α_i の差を確率変数としたときの平均 0 分散 1 の正規分布に従うとモデル化する．指差しした認識に基づく物体 i の指示尤度 P_i は実験環境中の G_i の総和で計算される．

$$G_i = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(\alpha_i - \alpha_h)^2}{2}\right) \quad (4)$$

$$P_i = \frac{G_i}{\sum_{i=1}^n G_i} \quad (5)$$

指差し認識に基づく指示尤度計算の具体例を，図 5 に示す．まず，物体情報データベースに登録されている物体 1, 2, 3 の座標から，実験参加者から各物体までの角度を求めることができる．物体 1, 2, 3 の角度はそれぞれ 2.356, 1.751, 1.326 (rad) となる．ここで，指差し角度が 1.801 であった場合，最終的な物体の指示された尤度はそれぞれ 0.312, 0.363, 0.325 となる．

3.4.4 統合モジュール

統合モジュールは，音声認識に基づく指示尤度と指差し

ID	x_i	y_i	手の位置と指差し角度 $(x_h, y_h) = (80, 10), \alpha_h = 1.801$			
1	40	50				
2	60	120				
3	100	90				
ID	$x_i - x_c$	$y_i - y_c$	α_i	$\alpha_i - \alpha_h$	G_i	L_i
1	-40	40	2.356	0.555	0.342	0.312
2	-20	110	1.751	-0.050	0.398	0.363
3	20	80	1.326	-0.475	0.356	0.325

図 5 指差し認識に基づく指示尤度の計算

Fig. 5 Example of calculation of likelihoods by means of the results of pointing gesture recognition.

認識に基づく指示尤度を統合する．物体 i の最終的な指示尤度 L_i は次のように計算される．

$$L_i = \frac{S_i + P_i}{\sum_{i=1}^n (S_i + P_i)} \quad (6)$$

ここで， S_i と P_i を組み合わせるために和を用いた理由は，実験参加者が指示語しか含まない指示発話をしたり，指差しをともしない指示発話をしたりする場合，つまり S_i が 0 であったり， P_i が 0 であったりする場合があるためである．仮に S_i と P_i の組合せに積を用いると，これらの場合ではすべて物体の L_i が 0 となるため不適切である．

この最終的な指示尤度が最大の物体が指示物体として推定される．

3.4.5 確認発話選択モジュール

確認発話選択モジュールは，物体指示データベースに格納されている物体の位置や属性集合と推定された指示物体の位置や属性集合から，確認発話で使用される属性もしくは属性の組合せを決定する．本システムでは，指示物体の識別に必要な最小の情報をを用いるように確認発話を選択する方法と，指示物体の識別に関するすべての情報をを用いるように確認発話を選択する方法の 2 種類がある．ここで，最小の情報をを用いるように確認発話を選択する方法について説明する．この場合に確認発話で使用される属性の選択は次の流れで行われる．

- (1) 指示物体の近傍にある物体を特定する．
- (2) 指示物体を近傍の物体から識別できるすべての属性組を求める．
- (3) 得られた属性組の中から，最小の要素数を持つ属性組を求める．
- (4) 最小の要素数を持つ属性組の中で，文字列の差異が指示物体と周囲の物体との間で最も大きい属性組を確認発話とする．

ここで，属性組とは属性の組合せを指す．物体が色，形，文字という属性を持つとき，「色」や「色・形」，「色・形・文字」など属性の組合せで表されるものはすべて属性組である．以下，物体 i の確認発話で使用される属性組を決定する処理の詳細を説明する．

3.4.5.1 近傍の物体の特定

物体の識別に必要な最小の情報は指示物体の近傍に物体が存在するかどうかによって変化する．たとえば，指示物体の近傍に物体がない場合，指差し動作によりその指示物体はほぼ一意に特定される．したがって，ロボットは指示物体の属性を 1 つ提示するだけで，人間は確認された物体を理解できると考える．一方，指示物体の近傍に物体がある場合，指差し動作でその指示物体を特定することは困難になる．このような場合，ロボットは指示物体を近傍の物体から識別できる属性組を提示する必要があると考える．

確認発話選択モジュールは，以下の式を用いて，物体 i の近傍に物体が存在するかどうかを検証する．ここで，物

体 i の近傍を、物体 i を中心とする半径 r の円の方程式で定義する.

$$(x - x_i)^2 + (y - y_i)^2 < r^2 \quad (7)$$

この式を満たす位置にある物体は、物体 i の近傍にあると見なされる. なお、本システムでは $r = 700 \text{ mm}$ とした. この式を満たす物体が存在する場合には、各物体の属性の関係を考慮した属性組を、存在しない場合には 1 つの属性を確認発話として選択する.

3.4.5.2 識別に利用可能な属性組の導出

確認発話選択モジュールは、物体 i の近傍に他の物体が存在する場合には、その近傍の物体の属性を考慮して、物体 i を一意に識別できる属性組を求める. 物体 i を近傍の物体の集合 Q から識別できる属性組の集合 D_i は、以下のように定義できる.

$$D_i = \text{Power}(O_i) - \sum_{j \in Q} (\text{Power}(O_i) \cap \text{Power}(O_j)) \quad (8)$$

ここで、 $\text{Power}(O_i)$ は O_i のべき集合である. 第 1 項は、物体 i の属性集合がとりうるすべての属性組の集合を表す. 第 2 項は物体 i および近傍の物体のとりうるすべての属性組の共通する属性組の集合を表す. いい換えれば、物体 i を近傍の物体から識別できない属性組の集合を表す. したがって、第 1 項と第 2 項の差を求めることで、物体 i を近傍の物体から識別できる属性組の集合を得ることができる. たとえば、図 6 において、物体 1 を近傍の物体 2, 3, 4 から識別できる属性組の集合は、次のように表される.

$$(D_1) = \{\{\text{aka, shikaku}\}, \{\text{aka, ei}\}, \{\text{aka, shikaku, ei}\}\} \quad (9)$$

3.4.5.3 最小の要素数を持つ属性組の導出

次に、指示物体を近傍の物体から識別できる最小の要素数を持つ属性組を求める. 図 6 における物体 1 の確認発話の候補となる属性の組合せは次のように表される.

$$\min(D_1) = \{\{\text{aka, shikaku}\}, \{\text{aka, ei}\}\} \quad (10)$$

ここで、 $\min$ は属性組の集合の中から要素数が最小の属性組を取得する演算子と定義する. 図 6 の例では、物体 1 を近傍の物体から識別可能な属性組のうち、最小の要素数を持つ属性組として、2 つの要素を持つ属性組が 2 種類得られた.

物体情報データベース

ID	x_i	y_i	色	図形	文字
1	80	70	aka	shikaku	ei
2	93	67	ki	shikaku	bii
3	76	60	ao	shikaku	ei
4	89	57	aka	sankaku	bii

図 6 色、図形、文字の属性を持つ物体の物体情報データベース
Fig. 6 Object information database that contains the objects with the attributes of color, symbols and characters.

3.4.5.4 文字列の差異の大きな属性組の導出

最小の要素数を持つ属性組が複数ある場合、確認発話での属性組を用いるかを決定する必要がある. この決定方法に関して、本研究では音声認識において誤認識が発生しやすい状況に着目した. ある物体に対する指示発話に含まれる属性組と他の物体に対する指示発話に含まれる属性組とが類似している場合、音声認識で 1 つの属性を誤認識しただけでも、指示物体を誤って推定する可能性が高くなる. したがって、それらの属性組は互いに異なっていることが望ましい. 同調傾向に関する知見から、ロボットが近傍の物体の属性組と類似性が低い属性組を含む確認発話を使用すれば、人間も同様の属性組を含む指示発話をする可能性が増えると考えられる.

確認発話における指示物体の属性組と近傍の物体の対応する属性組との類似性を計算するために、我々はレーベンシュタイン距離を利用した. 対応する属性組とは、同じ種類の属性で構成される属性組であり、レーベンシュタイン距離は 2 つの文字列の異なり具合を示す数値である. レーベンシュタイン距離が大きいほど、その 2 つの文字列は大きく異なっているということ意味する. レーベンシュタイン距離は文字の挿入・削除・置換の 3 種類の操作によって、1 つの文字列を別の文字列に変形するのに必要な手順の最小回数として定義できる. 2 つの文字列の k 番目と l 番目までのレーベンシュタイン距離は、次のように再帰的に表される.

$$\begin{aligned} LD(str1_k, str2_l) \\ = \arg\text{-min}(LD(str1_{k-1}, str2_l) + 1, LD(str1_k, str2_{l-1}) + 1, \\ LD(str1_{k-1}, str2_{l-1}) + c) \end{aligned} \quad (11)$$

$$c = \begin{cases} 0(\text{char}(str1, k) = \text{char}(str2, l)) \\ 2(\text{char}(str1, k) \neq \text{char}(str2, l)) \end{cases} \quad (12)$$

なお、空の文字列のレーベンシュタイン距離は以下のよう定義される.

$$LD(0, 0) = 0, LD(str1_k, 0) = k, LD(0, str2_l) = l \quad (13)$$

ここで、 c は置換のコスト、 char は文字列の i 番目の文字を表す関数とする. 置換は挿入と削除の操作で実現されるため、ここでは置換のコストを 2 とした. 挿入・削除・置換の 3 種類の操作によって計算されるコストのうち、最小のコストがレーベンシュタイン距離として採用される.

以下に、「おはよう」「こんにちは」「こんばんは」の文字列を事例に、レーベンシュタイン距離がどのように計算されるかを説明する. まず、「おはよう」から「こんにちは」への変形は 2 回の削除と 3 回の挿入、1 回の置換で実現できるため、この 2 つの文字列のレーベンシュタイン距離は 7 と計算される. 一方で、「こんにちは」と「こんばんは」のレーベンシュタイン距離は 4 であることから、「こんにちは」は「おはよう」に比べて「こんばんは」に類似して

いるといえる。

指示物体の識別に必要な最小の属性組が複数存在する場合、指示物体と近傍の物体とのレーベンシュタイン距離の和を計算し、その値が最大になる属性組を確認発話に採用する。例として、図 6 の状況を考える。物体 1 の識別に必要な最小の属性組は {赤, 四角} もしくは {赤, A} である。物体 1 の属性組 {赤, 四角} と物体 2 の対応する属性組 {黄, 四角}, 物体 3 の対応する属性組 {青, 四角}, 物体 4 の対応する属性組 {赤, 三角} のレーベンシュタイン距離はそれぞれ 5, 4, 4 であり、その総和は 13 である。同様に、物体 1 の属性組 {赤, A} と物体の対応する属性組 {黄, B}, 物体 3 の対応する属性組 {青, A}, 物体 4 の対応する属性組 {赤, B} のレーベンシュタイン距離はそれぞれ 6, 4, 1 であり、その総和は 11 である。これは物体 1 を物体 2, 3, 4 から識別するとき、属性組 {赤, 四角} の方が属性組 {赤, A} より文字列の差異が大きいことを示している。したがって、図 6 の状況では属性組 {赤, 四角} が確認発話で用いられる。

4. 実験

我々は、3 章で述べた指示物体認識と確認発話選択の機能を持つ実験システムを用いて、確認発話の冗長性を要因とする対照実験を行う。まず、この実験で検証する仮説について述べた後、実験の詳細について説明する。

4.1 仮説

本研究の目的は、ロボットの発話の冗長性を制御することによって指示物体認識性能が変化するかどうかを検証することである。そのために、人間とロボットの物体指示発話における同調傾向と発話の冗長性に関する知見から、次の仮説をたてた。

仮説：ロボットが指示物体の識別に関するすべての情報を用いるように確認発話を選択する場合より、ロボットが指示物体の識別に必要な最小の情報を用いるように確認発話を選択する場合に、指示物体の認識性能が高くなる。

4.2 実験計画

本実験では上記の仮説を検証するためにロボットの確認発話を操作した。また、指示物体の認識性能がロボットの発話の冗長性だけでなく物体配置によっても変化する可能性があることを考慮し、実験参加者が最初に物体を並べるときの配置も操作した。たとえば、物体どうしが離れて疎に配置されている場合、指差し動作と指示語を使った指示発話が増え、物体の属性を使った発話は減る傾向にあると考えられる。また、物体どうしが密に配置されている場合、指差し動作で物体を識別することは困難になるため、指差し動作を使うよりも物体の属性を多く使った発話が増える傾向にあると考えられる。つまり、物体配置によりロ

ボットの確認発話による指示物体認識性能への効果が変化する可能性がある。そこで本研究では、(1) ばらばらの配置、(2) 2 つか 3 つの集団ができる配置、(3) 1 つの集団ができる配置、という 3 種類の物体配置を想定し、これらの物体配置において指示物体認識性能が変化するかどうかを検証した。

本実験における要因と水準は以下のとおりである。

- 確認発話要因（実験参加者間要因）
 - 最小属性確認水準
 - 全属性確認水準
- 物体配置要因（実験参加者内要因）
 - ばらばら配置水準
 - 2, 3 の集団配置水準
 - 1 つの集団配置水準

4.2.1 確認発話要因

確認発話要因は実験参加者間要因であり、各実験参加者はロボットの確認発話の冗長性に関する 2 つの水準のうち一方に割り当てられた。具体的に、最小属性確認水準では、ロボットは、指示物体の識別に必要な最小の属性を用いた確認発話を行った。この確認発話は「[[図形] と [文字] の [色] 本ですか?」という形式であり、必要のない属性部分の発話は省略された。たとえば、図 6 の矢印で指示された本の確認を行う場合には、ロボットは「四角の赤色の本ですか?」と発話した。全属性確認水準では、ロボットは、指示物体の識別に関するすべての属性を用いた確認発話を行った。この確認発話は「[[図形] と [文字] の [色] 本ですか?」という形式であった。たとえば、図 6 の矢印で指示された本の確認を行う場合には、ロボットは四角と A の赤い本ですか?と発話した。各水準での実験参加者とロボットの会話例を表 1 に示す。

4.2.2 物体配置要因

物体配置要因は実験参加者内要因であり、各実験参加者は実験者からの物体配置に関する指示に関する 3 つの水準のすべてに割り当てられた。具体的には、実験者が実験参加者に物体を床上に並べるように指示するとき、ばらばら配置水準では、「物体どうしが離れるようにばらばらに配置してください」、2, 3 の集団配置水準では、「2 つか 3 つ

表 1 確認発話要因の各水準での会話例

Table 1 Example of experimental conversations at each level in a confirmation factor.

話者	最小属性確認水準	全属性確認水準
ロボット	本を指示してください。	本を指示してください。
実験参加者	じゃあ、赤い A の本。	じゃあ、赤い A の本
ロボット	その四角の赤色の本ですか?	その四角と A の赤色の本ですか?
実験参加者	そうです。	そうです。
ロボット	本を指示してください。	本を指示してください。
実験参加者	うーん、その黄色い本	うーん、その黄色い本
ロボット	その三角の本ですか?	その三角と B の黄色の本ですか?
実験参加者	そうです。	そうです。
ロボット	本を指示してください。	本を指示してください。
本の指示が 10 回繰り返されるまで続く。		

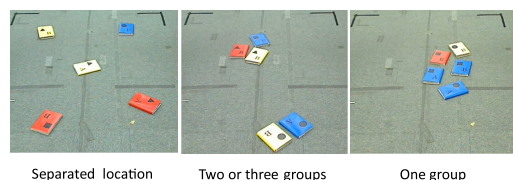


図 7 3 種類の本の配置

Fig. 7 Three ways of arrangement of books.

の集団ができるように物体を配置してください」, 1つの集団配置水準では, 「物体どうしが近くであり1つの集団ができるように配置してください」という3種類の指示を行った. これらの指示に従った配置の例を図7に示す.

4.3 手順

この実験では, 物体識別に有効な複数の属性を持つ物体が複数個存在する環境において, 実験参加者が複数回物体を指示するという状況を設定した. 指示される物体として, B6判の本を用いた. これらの本は, それぞれ異なる色と図形と文字の組合せが印刷された紙で包まれている. この色と図形と文字がこの実験での物体の属性である. 色は赤と青と黄の3種類, 図形は丸と三角と四角の3種類, 文字はAとBの2種類であった. 本は属性のすべての組合せを網羅するように18冊用意された. 以下に, 実験の手順を述べる.

最初に, 実験参加者は実験の目的と内容の簡単な説明を受け, その後, 実験同意書を読み, 署名するように求められた. そして実験参加者は実験室に移動し, そこで実験の具体的な手順について教えられた.

実験の指示として, 我々は実験参加者に「私たちは人間が指示した物体を認識できるロボットを開発しています. 今日はそのロボットの評価実験に参加してもらいます」と伝えた. また, ロボットへの指示の方法について「このロボットは人の発話と身振り手振りを認識することができますので, 普段人相手に行うように指示してください」と指示した. そして, 実験参加者は確認発話要因の2つの水準のうち一方に, 男女比が均等になるように, 12名ずつ無作為に割り当てられた.

実験では, 最初に実験参加者が18種類の本の中からの自由に5冊を選んだ. そして, 実験者の指示に従いその本を配置した. なお, 配置に関する指示の順番は, 確認発話要因の水準間でカウンタバランスをとるように決定された.

実験参加者は本を決められた領域内に配置した後, 領域内の椅子に座るように指示された. 実験参加者が椅子に座り, 実験者が実験を開始することを告げてから, ロボットは「本を指示してください」と発話した. 実験参加者は, ロボットの発話を聞き, 自由に本を1冊指示した. 実験参加者が本を指示した後, ロボットは指示物体認識機能によって推定された本を, 確認発話選択機能によって選ばれた確



図 8 実験の様子

Fig. 8 Experimental scene.

認発話を用いて確認した. 実験参加者には, ロボットの確認に対して, 「そうです」か「違います」と答えるように指示した. 実験参加者は, 物体指示を10回行い, 異なる指示に従って配置を変更し, 再び10回の物体指示を行った. これを, すべての配置に対して行ったため, 最終的に合計30回の物体指示を行った. 実験時の実験参加者とロボットの様子を図8に示す. ロボットの確認発話の正否にかかわらず, ロボットは実験参加者に次の本の指示を促した. 実験参加者は5冊の本を10回指示したので, 2度か3度繰り返して指示される本が存在した.

4.4 実験参加者

本実験には男性12名, 女性12名が参加した. 全属性確認水準には男性6名(平均年齢21.17歳), 女性6名(20.33歳), そして最小属性確認水準には男性6名(20.17歳), 女性6名(20.67歳)がそれぞれ割り当てられた. なお, 実験参加者はすべて日本語を母国語とする大学生であった. 実験参加者には, 1時間あたり1,000円の謝金が支払われた.

5. 実験結果

5.1 仮説の検証

4.1節で述べた仮説を検証するために, 各条件で指示物体の認識成功率を測定した. ロボットが実験参加者の指示した本と同じ本を確認した場合を指示物体の認識成功とし, 異なる本を確認した場合を認識失敗とした. すなわち, 実験参加者がロボットの確認を聞いて「そうです」と応答した場合を指示物体の認識成功とし, 「違います」と応答した場合を指示物体の認識失敗とした. なお, ロボットが実験参加者の指示した本を正しく確認したのに, 実験参加者が「違います」と答えたり, 逆にロボットが実験参加者の指示した本を間違えて確認したのに, 実験参加者が「そうです」と答えたりする事例は, 実験者が後からビデオで確認したところ, 存在しなかった. 認識成功率は各配置における10回の指示発話あたりの認識成功回数で計算された. その結果を表2に示す.

認識成功率に関する各要因の主効果と交互作用を検証するために, 1要因対応あり(配置)1要因対応なし(確認発話)の混合計画で2要因分散分析を行った. その分析結果

を表 3 に示す。その結果、確認発話要因の主効果が認められた ($F(1,22) = 6.171, p = 0.021$)。この結果から、最小属性確認水準における認識成功率は、全属性確認水準における認識成功率よりも高く、その差は有意であることが示された。この結果は、指示物体の認識性能は、ロボットが指示物体の識別に関するすべての情報を用いるように確認発話を選択する場合より、ロボットが指示物体の識別に必要な最小の情報を用いるように確認発話を選択する場合に高くなるという仮説の妥当性を支持している。

一方、配置要因では有意な主効果が認められず ($F(2,44) = 0.312, p = 0.733$)、また確認発話要因と配置要因との間にも有意な交互作用は認められなかった ($F(2,44) = 1.396, p = 0.258$)。この結果は、配置によって認識成功率に差があったとはいえないということ、そして、確認発話要因の主効果が配置要因の水準によって変化したとはいえないことを示している。したがって、物体がばらばらに配置された場合、2つか3つの集団ができるように配置された場合、1つの集団ができるように配置された場合で、指示物体認識性能が変化したとはいえなかった。

表 2 指示発話の認識成功率の平均と標準偏差 (単位は%)

Table 2 Mean and standard deviation of success rates of recognition of each participant.

	ばらばら	2, 3 の集団	1 つの集団
最小属性確認水準	74.2 (25.3)	78.3 (16.2)	74.2 (19.8)
全属性確認水準	54.2 (22.5)	50.0 (26.5)	60.0 (24.8)

表 3 混合計画の 2 要因分散分析による認識成功率の分析結果

Table 3 Results of two-factor repeated ANOVA (mixture design) of the success rates.

要因	平方和	自由度	平均平方	F	有意確率
主効果：確認発話	0.781	1	0.781	6.171	0.021 *
誤差：S(確認発話)	2.785	22	0.127		
主効果：配置	0.014	2	0.007	0.312	0.733
交互作用：確認発話×配置	0.061	2	0.030	1.396	0.258
誤差：配置×S(確認発話)	0.959	44	0.022		
全体	4.560	71			

*: $p < 0.05$

5.2 認識性能向上の理由

本実験で示された確認発話要因での認識性能の向上の理由について、指示発話の冗長性と認識性能の相関、そして各要因での指示発話の冗長性の変化という観点から考察する。

5.2.1 指示発話の冗長性と認識性能

2.3 節で述べたように、本研究では、指示物体の認識性能を向上させるために、人間は物体の識別に必要な情報を多く含む指示発話、すなわち冗長性のある指示発話を行うことが望ましい、という考えのもとで、インタラクションの利用による指示物体認識性能の変化について検証した。本実験の結果がこの考えに沿うものであるかどうかを検証するために、実験参加者の指示発話の冗長性の平均と、その認識成功率の平均の相関を求めた。ここで、指示発話の冗長性を、指示発話に含まれる属性数と言葉のみで指示物体を識別するのに必要な最小属性数の差と定義した。実験参加者の 1 回の指示発話に含まれる属性数は 0 から 3 の値であり、言葉のみによる識別に必要な最小属性数は 1 から 3 の値となる。したがって、1 回の指示発話の冗長性は -3 から 2 の値で表される。具体例として、全属性確認水準と最小属性確認水準において、ある実験参加者の指示発話とその冗長性および、それに対するロボットの確認発話を表 4 と表 5 にそれぞれ示す。これらの指示発話はどちらも 2, 3 の集団の配置で行われたものである。表 4 に示される指示発話の冗長性の平均は -0.3 で認識成功率は 50%であるのに対し、表 5 に示される指示発話の冗長性の平均は 0.7 で認識成功率は 80%となっている。

各実験参加者の指示発話の認識成功率の平均と冗長性の平均の関係を見るために図 9 にそれらの散布図を示す。確認発話要因の水準ごとにその相関係数を計算したところ、最小属性確認水準では 0.767、全属性確認水準では 0.803 であった。すなわち、両水準において、各実験参加者の指示発話の認識成功率の平均と冗長性の平均との間に強い正の相関が認められた。この結果は、冗長性のある指示発話に対する指示物体認識の成功率は、冗長性のない指示発話に対する指示物体認識の成功率よりも高い傾向にあるという

表 4 全属性確認水準における指示発話とロボットの確認発話の例

Table 4 Example of references and confirmations in the maximum condition.

全属性確認条件(配置された本:赤・三角・A, 赤・四角・B, 青・四角・B, 青・丸・B, 黄・丸・B)					
実験参加者の指示発話	指示された本	冗長性	ロボットの確認発話	確認された本	認識
黄色い本を取って下さい	黄・丸・B	0	その三角とAの赤色の本ですか?	赤・三角・A	失敗
三角の本を取って下さい	赤・三角・A	0	その三角とAの赤色の本ですか?	赤・三角・A	成功
その四角い本を取って下さい	青・四角・B	-1	その四角とBの青色の本ですか?	青・四角・B	成功
赤色とBの本を取って下さい	赤・四角・B	0	その四角とBの青色の本ですか?	青・四角・B	失敗
その黄色い本を取って下さい	黄・丸・B	0	その丸とBの黄色の本ですか?	黄・丸・B	成功
その丸の本を取って下さい	青・丸・B	-1	その丸とBの黄色の本ですか?	黄・丸・B	失敗
その四角の本を取って下さい	赤・四角・B	-1	その四角とBの青色の本ですか?	青・四角・B	失敗
赤いAの本を取って下さい	赤・三角・A	1	その三角とAの赤色の本ですか?	赤・三角・A	成功
その丸の本を取って下さい	黄・丸・B	0	その丸とBの青色の本ですか?	青・丸・B	失敗
その四角い本を取って下さい	青・四角・B	-1	その四角とBの青色の本ですか?	青・四角・B	成功

表 5 最小属性確認水準における指示発話とロボットの確認発話の例

Table 5 Example of references and confirmations in the minimum condition.

最小属性確認条件(配置された本:赤・三角・A, 赤・四角・A, 赤・四角・B, 青・丸・A, 黄・丸・B)					
実験参加者の指示発話	指示された本	冗長性	ロボットの確認発話	確認された本	認識
赤い四角の本で	赤・四角・A	0	その四角とAの本ですか?	赤・四角・A	成功
黄色い本で	黄・丸・B	0	その青色の本ですか?	青・丸・A	失敗
黄色い丸い本で	黄・丸・B	1	その黄色の本ですか?	黄・丸・B	成功
青い丸の本で	青・丸・A	1	その青色の本ですか?	青・丸・A	成功
赤い三角の本で	赤・三角・A	1	その三角の本ですか?	赤・三角・A	成功
赤い四角の本で	赤・四角・B	0	その四角とAの本ですか?	赤・四角・A	失敗
赤い四角のBの本で	赤・四角・B	1	その四角とBの本ですか?	赤・四角・B	成功
黄色の丸の本で	黄・丸・B	1	その黄色の本ですか?	黄・丸・B	成功
Aの赤い四角の本で	赤・四角・A	1	その四角とAの本ですか?	赤・四角・A	成功
青いAの本で	青・丸・A	1	その青色の本ですか?	青・丸・A	成功

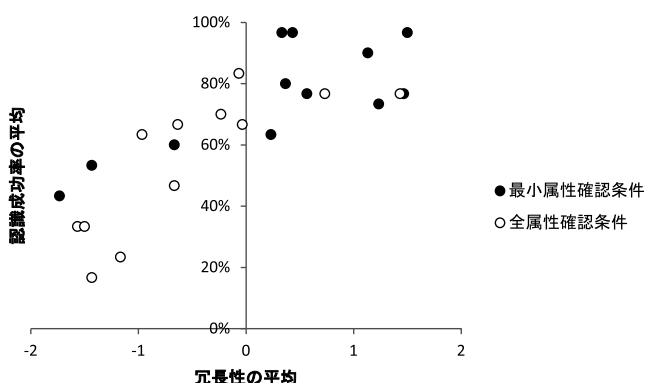


図 9 各実験参加者の指示発話の冗長性の平均と認識成功率の平均の相関

Fig. 9 Correlation between redundancies and success rates of each participant.

表 6 指示発話の冗長性の平均と標準偏差

Table 6 Mean and standard deviation of sum of redundancies of each participant.

	ばらばら	2, 3 の集団	1 つの集団
最小属性確認水準	0.117 (0.986)	0.417 (1.113)	0.325 (1.029)
全属性確認水準	-0.717 (0.941)	-0.517 (0.896)	-0.292 (0.947)

ことを示唆している。

5.2.2 ロボットの確認発話による人間の指示発話の変化

次に、実際にロボットの確認発話によって人間の指示発話の冗長性がどのように変化したかということを検証する。

各条件での指示発話の冗長性の平均を表 6 に示す。指示発話の冗長性に関する各要因の主効果を検証するために、1 要因対応あり(配置) 1 要因対応なし(確認発話)の混合計画で 2 要因分散分析を行った。その分析結果を表 7 に示す。その結果、確認発話要因の主効果に有意傾向が認められた ($F(1,22) = 3.789, p = 0.064$)。人間が冗長性のある指示発話を行う割合の差は条件間で有意ではなかったが、ロボットが冗長性のある確認発話を行う場合より、ロボットが冗長性のない確認発話を行った場合において、人間は冗長性のある指示発話をする傾向がある可能性は示唆された。

また、配置要因に主効果が認められた ($F(2,44) = 6.822$,

表 7 混合計画の 2 要因分散分析による冗長性の分析結果

Table 7 Results of two-factor repeated ANOVA (mixture design) of redundancies.

要因	平方和	自由度	平均平方	F	有意確率
主効果: 確認発話	11.361	1	11.361	3.789	0.064
誤差: S(確認発話)	65.964	22	2.998		
主効果: 配置	1.338	2	0.669	6.822	0.003
交互作用: 確認発話×配置	0.314	2	0.157	1.603	0.213
誤差: 配置×S(確認発話)	4.314	44	0.098		
全体	83.291	71			

+p < 0.1, *p < 0.05

表 8 配置による使用属性数の差の多重比較結果

Table 8 Results of multiple comparison by means of Scheffe's procedure.

	平均値の差	有意確率
ばらばら 2, 3 の集団	-0.250	0.029 *
1 つの集団	-0.317	0.004 *
2, 3 の集団 ばらばら	0.250	0.029 *
1 つの集団	-0.067	0.763
ばらばら 1 つの集団	0.317	0.004 *
2, 3 の集団	0.067	0.763

*p < 0.05

$p = 0.003$)。これは、配置によって指示発話の冗長性に差があったことを示している。配置による冗長性の差を分析するために、Scheffe 法による多重比較を行った。その結果を表 8 に示す。その結果によれば、ばらばらの配置と 2, 3 の集団の配置の平均値の差 ($p = 0.029$)、そしてばらばらの配置と 1 つの集団の配置の平均値の差 ($p = 0.004$) が 5%水準で有意であった。すなわち、ばらばらの配置よりも 2, 3 の集団の配置と 1 つの集団の配置のほうが冗長な指示発話を使用したことを示している。

5.3 結果のまとめ

この実験で得られた結果を以下にまとめる。

- 指示物体の認識性能は、ロボットの確認発話の仕方によって変化した。その認識性能は、すべての情報を用いる確認発話よりも最小限の情報を用いる確認発話において、高かった。
- 指示物体の認識性能と実験参加者の指示発話の冗長性

には高い正の相関があった。

- 実験参加者の指示発話の冗長性は、物体の配置の仕方によって変化した。その冗長性は、物体の配置がばらばらなときより、1つの集団や、2, 3の集団であるときに、高かった。
- 実験参加者の指示発話の冗長性はロボットの確認発話の仕方によって変化する可能性が示唆された。その冗長性は、ロボットがすべての情報を用いる確認発話よりも最小限の情報を用いる確認発話において、高くなる傾向があった。
- 以上のことから、物体配置によって人間の指示発話の冗長性は変化するが、どの物体配置においても、ロボットの確認発話の冗長性を減らすことで人間の指示発話の冗長性が高くなり、結果的に認識性能が向上する可能性が示唆された。

6. 考察

6.1 本知見の適用範囲

6.1.1 既存システムへの適用

本研究の新規性と有用性は、ロボットが行う確認発話のみによって、指示物体認識の性能が変化することを示した点にある。特に、ロボットが指示された物体を特定できる最小限の情報を含む確認発話をすれば、指示物体認識の性能が向上する、という知見は有用性が高いといえる。なぜなら、確認発話は物体指示対話における自然な行為であり、ロボットの対話戦略の中に組み込みやすい。また、確認発話という行為はセンサや認識アルゴリズムなどの工学的な制約に影響を受けないため、既存の指示物体認識システムを利用したロボットにも適用が容易である。

本研究の知見を利用するためには、ロボットが確認発話を行うための、環境内の物体に関する特徴を動的に認識するシステムが必要となる。実環境では、様々な属性を持った物体が、様々な位置関係で存在する。本システムでは、各物体の属性と位置関係を実験開始前にデータベースに登録していたが、物体に無線タグや画像マーカを設置することで、環境内の物体位置を自動的に取得できるようになると考えられる。たとえば、家庭内に設置されたセンサ群と連動して、人々とインタラクションを行うエージェント [25] やロボットへの実装は、比較的容易であると考えられる。

本研究の提案手法が従来研究で提案されているような指示物体認識システムにも有効であるかどうかを検証するためには、従来研究と実験設定や認識システムを同一にしたうえで、ロボット発話を独立変数とし、認識性能の違いを測る必要があると考える。しかしながら、あらゆる指示物体認識手法に対して、提案手法の有効性を確認することは困難である。そのために本研究では、既存の音声認識技術と比較的単純な画像認識技術を用いて指示物体認識システムを構築し、実験を行った。提案手法の有効性が既存技術

を用いて構築される指示物体認識システムを用いて検証されたことは、その他の指示物体認識システムでも提案手法が有効であることを示唆していると考えられる。

6.1.2 他の種類の本と、本以外の物体への適用

本研究では、色と図形と文字という3種類の属性を持つ本を指示するという状況のみにおいて、人間とロボットのインタラクションを利用することで指示物体の認識性能が向上することを示した。しかしながら、他の本や本以外の物体に関する指示対話でも、本研究で得られた知見は適用可能であると考えられる。

たとえば、市販されている本にはタイトル、著者、色、大きさ、分類など様々な属性が付加されている。本研究で得られた知見によれば、ロボットは指示された本を確認するために本に関するすべての属性を使う必要はなく、指示された本を周りの物体から識別するのに必要な最小の情報をいればよい。この意味において、タイトルは本を識別するための有用な情報であると考えられる。しかしながら、同じようなタイトルの本が他にも存在する場合などには、色や大きさなどの属性を使って確認発話をするすることで、人間からより理解しやすい指示発話を引き出せるようになると考えられる。

また、本以外の物体には、タイトルのように物体の識別に役立つ明確なラベルはないことが多い。このとき、色や大きさや形など物体の情報を用いて、その物体を他の物体から識別する必要がある。本研究で得られた知見によれば、このような場合、その物体を他の物体から識別できる最小の情報で構成される確認発話を行うことで、人間から冗長性のある指示発話を引き出せるようになると考えられる。

6.1.3 インタラクションの繰返しの効果

本実験では、実験参加者の物体指示とロボットの指示物体の確認のインタラクションが30回繰り返されていた。実験参加者の指示発話の冗長性の変化はこのインタラクションの繰返しによって引き起こされたものと考えられる。過去の人間とロボットとの物体指示対話における語彙の引き込み現象に関する研究 [12] では、5回の物体指示で言葉のカテゴリの引き込みが見られたと報告されている。そこで本実験結果から、最小属性確認水準と全属性確認水準とで、実験参加者の各セッションの前半5回と後半5回の指示発話の冗長性の平均を求めた。その結果を表9に示す。各水準で前半と後半の指示発話の冗長性の平均について対応の

表9 セッションの前半5回と後半5回の指示発話の冗長性の平均と標準偏差

Table 9 Mean and standard deviation of redundancies at anterior and posterior five references.

	前半5回	後半5回
最小属性確認水準	0.233 (1.174)	0.339 (1.216)
全属性確認水準	-0.517 (1.147)	-0.500 (1.239)

ある t 検定を行った結果、最小確認水準において有意差が認められた ($t(11) = -2.843$, $p = 0.016$)。一方、全属性確認水準では有意差が認められなかった ($t(11) = -0.155$, $p = 0.880$)。この結果は、ロボットが物体を識別できる最小の属性の組合せを使い物体を確認するというインタラクションの繰返しが実験参加者の指示発話の冗長性を変化させたという可能性を示唆している。

ここで、インタラクションの繰返しが人間の行動を変容させるという本研究で提案した手法に関する利点と欠点について考察する。まず、欠点について、提案手法は人間の指示発話の冗長性が変化するまでにある程度の回数のインタラクションが必要になるため、すぐに相手の発話の冗長性を変化させたいという状況には適用できないであろう。したがって、本研究で得られた知見を適応できるのは、物体指示対話が繰返し行われる場合に限定される。

しかしながら、物体指示対話は基本的に繰返し行われる対話である。たとえば、家やオフィスでロボットが物を運搬する場面を考えてとき、1日でどれだけの頻度の物体指示が行われるかは具体的な状況によるが、物体指示はほぼ毎日行われるであろう。また、実社会で活動する様々なロボットが提案手法に基づいた発話を行うことで、人間はロボット全体に対して冗長性のある発話をするようになるかもしれない。つまり、物体指示対話は日々繰返し行われる対話であるため、本知見は幅広く適応されうと考える。

以上の議論が本当に正しいかどうかを検証するためには、具体的にどの程度のインタラクションの繰返しで指示発話の冗長性の変化が現れはじめるのか、インタラクションによる人間の指示発話の変化はどの程度の間持続するのか、また、対話の相手となるロボットが変わっても持続するのか、といったことを検証する必要がある。これらの検証は、今後の課題である。

6.2 分析対象外の発話

実験参加者の指示発話の中には、物体を参照する言葉として、色と図形と文字以外に場所に関わる言葉が含まれる発話がわずかにあった。具体的には、全 720 の指示発話のうち、「端」を含む発話は 3 つ、「右」を含む発話は 2 つ、「左」を含む発話は 1 つ、「真ん中」を含む発話は 7 つ、「手前」を含む発話は 1 つであった。本実験では、ロボットは色と図形と文字に関する言葉のみを使用したため、語彙の引き込み [12] の効果によって、実験参加者も色・図形・文字に関する言葉以外の言葉を多用しなかったものと考えられる。

6.3 本の選ばれ方が認識性能に与える影響

実験環境中に配置された本の認識のしやすさは、本の配置だけでなく、どのような組合せで本が選ばれたかという

表 10 状況 1：選ばれた本が類似している状況

Table 10 Case 1: The case where selected books are similar to each other.

	色	図形	文字
本 1	赤	丸	A
本 2	赤	丸	B
本 3	青	丸	A

表 11 状況 2：選ばれた本が類似していない状況

Table 11 Case 2: The case where selected books are not similar to each other.

	色	図形	文字
本 1	赤	丸	A
本 4	青	三角	B
本 5	黄	四角	B

ことにも影響を受ける。本実験では、最初の実験参加者が 18 種類の本の中から自由に 5 冊を選んでいた。そこで、確認発話要因と配置要因の各水準間での認識性能の差に本の選ばれ方が影響していた可能性について考察する。

配置された本の各属性が類似している場合は、そうでない場合に比べて、本の識別が難しくなるため、本の選ばれ方は認識成功率に影響を与える可能性がある。選ばれた本が類似している状況 (表 10)、そして選ばれた本が類似していない状況 (表 11) の 2 つの状況を例に説明する。状況 1 の場合、本 1 を一意に識別できる属性の組合せは、{赤・A, 赤・丸・A} となる。一方、状況 2 の場合、本 1 を一意に識別できる組合せは、{赤, 丸, A, 赤・丸, 赤・A, 丸・A, 赤・丸・A} である。ここで、実験参加者が本 1 を「赤・丸・A」という属性の組合せを使って参照したとする。状況 1 では、「赤」か「A」のどちらかの属性を認識しそこなった場合、参照された物体が本 1 だと特定できなくなる。一方、状況 2 では、すべての属性を認識しそこなわない限り、参照された物体が本 1 だと特定できる。つまり、状況 1 の本 1 の認識は状況 2 の本 1 の認識に比べて難しいといえる。このように選ばれた本の互いの類似の度合いによって、参照された物体を正しく認識できる割合は変化する可能性がある。したがって、選ばれた本の類似の度合いに水準間で差がないことを示せば、本の選ばれ方によって認識成功率が変化したわけではないことを示すことができる。

類似の度合いは本文中で述べたレーベンシュタイン距離で定量的に示すことができる。2 つの本の各属性間のレーベンシュタイン距離の合計が小さければ、その 2 つの本は類似の度合いが高いといえる。たとえば、先に述べた状況 1 の本 1 と本 2 の各属性のレーベンシュタイン距離は、色属性が 0、図形属性が 0、文字属性が 1 であるため、その総和は 1 となる。また、本 1 と本 3 の各属性のレーベンシュタイン距離の総和は 2、本 2 と本 3 では 3 となる。よって、状況 1 におけるそれぞれの本の異なり具合はそれぞれの合計から、6 となる。一方、状況 2 におけるそれぞれの本の

異なり具合は 20 となる。つまり、状況 1 の方が状況 2 よりも異なり具合が小さいことを意味している、

本実験において、実験参加者が選んだ 5 冊の本のレーベンシュタイン距離の総和の平均と標準偏差を配置ごとに表 12 に示す。2 要因分散分析を行った結果、確認発話要因 ($F(1,22) = 0.915, p = 0.349$)、配置要因 ($F(2,44) = 0.298, p = 0.744$) および、それらの交互作用 ($F(2,44) = 0.121, p = 0.886$) にも有意な主効果は認められなかった。この結果は、水準間で類似の度合いに差があったとはいえないことを示しており、本の選ばれ方が水準間での認識性能の差に影響した可能性は低いことを示唆している。

実験開始前の本の選び方だけでなく、実験参加者の指示する本の選び方も認識成功率に影響を与える可能性がある。この影響の可能性を検証するために、実験参加者による各本の選択回数の分散を計算した。各水準での本の選択回数の分散の平均と標準偏差を表 13 に示す。各本の選択回数の分散が小さいほど、それぞれの本は同程度指示されたということを示し（分散が 0 なら各本は 2 回ずつ指示されたことになる）、分散が大きいほど指示された本が偏っていたということを示している。2 要因分散分析を行った

表 12 配置された 5 冊の本のレーベンシュタイン距離の総和の平均と標準偏差

Table 12 Mean and standard deviation of sum of Levenshtein distance of five books that each participant placed.

	ばらばら	2, 3 の集団	1 つの集団
最小属性確認水準	49.5 (21.18)	47.67 (100.24)	49.5 (30.64)
全属性確認水準	51.17 (39.61)	50.17 (37.42)	50.17 (39.61)

表 13 各本が参照された回数の分散の平均と標準偏差

Table 13 Mean and standard deviation of variance of the number of references to each book.

	ばらばら	2, 3 の集団	1 つの集団
最小属性確認水準	0.267 (0.249)	0.633 (0.553)	0.433 (0.382)
全属性確認水準	0.733 (0.830)	0.58 (0.500)	0.533 (0.298)

結果、確認発話要因、配置要因およびそれらの交互作用に有意な主効果は認められなかった ($F(1,22) = 0.915, p = 0.302$; $F(2,44) = 0.563, p = 0.744$; $F(2,44) = 0.121, p = 0.110$)。この結果は、各水準間で本の選択回数の分散に差があったとはいえないことを示しており、指示する本の選ばれ方が各水準間での認識性能の差に影響した可能性は低いと考える。

6.4 音声認識と指差し認識の性能に関する考察

本実験で実装した指示物体認識システムでは、音声認識と指差し認識を組み合わせていた。ここで、この手法の性能について考察する。まず、実験で得られた音声認識の結果のみから指示物体を推定したときの正解率を表 14 に示す。指示発話の冗長性が 0 以上である場合、すなわち物体を一意に特定できる発話をしている場合は、比較的高い精度で指示物体を推定できている。一方、指示発話の冗長性が 0 未満の場合、すなわち物体を一意に特定できない指示発話の場合、指示物体の推定の精度は低くなっている。これは、仮に音声認識で正しい単語を認識できていたとしても、その単語だけでは物体を一意に特定できないためであった。したがって、人間の指示発話の冗長性を変化させることで、指示物体の認識率を向上させるという本研究の提案は妥当であったといえる。なお、全属性確認水準の 2, 3 の集団で正解率が 0% となっているが、これはもともとこの条件で冗長性 2 の指示発話が 3 回しか行われず、その 3 回とも誤認識をしたためであった。

次に、指差し認識について考察する。本実験では、「このロボットは人の発話と身振り手振りを認識することができまますので、普段人相手に行うように指示してください」と教示したため、指差し動作をともなった指示発話とそうでない指示発話があった。なお、指示発話をともなわない、すなわち無言で行われる指差し動作はなかった。指差し動作をともなった指示発話のうち、指差し認識の結果のみか

表 14 音声認識に基づく指示尤度のみを用いて指示物体を推定したときの正解率（括弧の中は音声認識と指差し認識の両方の尤度から推定したときの正解率）

Table 14 The rate of references in which objects were correctly estimated by the likelihoods using the only results of speech recognition (In parentheses, the case of using the both results of speech recognition and pointing gesture recognition).

	冗長性				
	-2	-1	0	1	2
最小属性確認水準					
ばらばら	6% (28%)	0% (45%)	80% (80%)	83% (90%)	90% (100%)
2, 3 の集団	8% (42%)	0% (62%)	84% (86%)	86% (83%)	90% (86%)
1 つの集団	0% (53%)	0% (40%)	79% (72%)	83% (85%)	92% (85%)
全属性確認水準					
ばらばら	3% (44%)	17% (34%)	77% (81%)	46% (62%)	86% (100%)
2, 3 の集団	0% (30%)	14% (17%)	89% (76%)	67% (81%)	0% (33%)
1 つの集団	4% (24%)	11% (44%)	83% (83%)	77% (77%)	60% (70%)

ら指示物体を推定したときの正解率を表 15 に示す.

指差し認識に関する従来研究 [5], [6], [7], [8] では, 指差し動作の認識や指差し方向の推定を目的としているため, 実験参加者は指差し動作をするように明示的に教示されていた. このため, 従来研究では, ほとんどの実験参加者が求められた姿勢で肘を伸ばし, まっすぐに物体を指示していたと考えられる. 一方, 本実験では前述の教示したため, 実験参加者は前かがみになりながら指差ししたり, 肘を曲げて大雑把に指差ししたりすることがあった. 本実験では指差し動作が自由に行われる状況であったため, 指差し認識の精度が結果的に低くなったと考える.

しかしながら, 最終的に指示物体を推定するうえで指差し認識は重要であった. ここで, 表 16 に指差し動作をともなった指示発話のうち, 音声認識の結果から推定された物体の正解率を, 表 17 に音声認識と指差し認識の両方の結果から推定された物体の正解率を示す. 本実験のように, 自由に実験参加者が物体を指示する場合, しばしば物体を一意に特定できない, 冗長性が 0 未満の指示発話がなされることがある. このような指示発話は表 14 に示したとおり, 音声認識だけでは正しく指示物体を推定できない. よって, 音声認識の尤度のみから推定される指示物体が正しい確率は平均的に低くなる (表 16). 指差し認識は, こ

表 15 指差し動作をともなった指示発話のうち, 指差し認識の尤度のみから指示物体を推定したときの正解率

Table 15 The rate of references with pointing gestures in which objects were correctly estimated by the likelihoods using the only results of pointing gesture recognition.

	ばらばら	2, 3 の集団	1 つの集団
最小属性確認水準	41%	36%	36%
全属性確認水準	42%	30%	32%

表 16 指差し動作をともなった指示発話のうち, 音声認識の尤度のみから推定された物体の正解率

Table 16 The rate of references with pointing gestures in which objects were correctly estimated by the likelihoods using the only results of speech recognition.

	ばらばら	2, 3 の集団	1 つの集団
最小属性確認水準	47%	48%	50%
全属性確認水準	24%	26%	31%

表 17 指差し動作をともなった指示発話のうち, 音声認識と指差し認識の両方の尤度から推定された物体の正解率

Table 17 The rate of references with pointing gestures in which objects were correctly estimated by the likelihoods using the both results of speech recognition and pointing gesture recognition.

	ばらばら	2, 3 の集団	1 つの集団
最小属性確認水準	64%	67%	66%
全属性確認水準	50%	39%	51%

の物体を一意に特定できない指示発話でも指示物体を推定することで, 指示物体認識システムの精度向上に寄与したと考える (表 17).

7. おわりに

本稿では, ロボット発話の冗長性を制御することで, 人間が指示した物体を認識する際の認識性能が向上することを示した. 過去の研究では, ロボットの発話によって指示物体の認識性能が向上する可能性が示唆されていたものの, 実際に認識性能が向上することを示した研究はなかった. そこで我々は, 過去の研究の知見をもとに, 人間に物体の識別に利用可能な情報が多く含まれる指示発話を行ってもらうために, ロボットが指示物体の識別に必要な最小の情報で構成される確認発話を行う手法を考えた. 確認発話の仕方によって指示物体の認識性能が向上することを示すため, 音声認識と指差し認識の結果に基づいて指示物体を認識するシステムを構築し, 指示された物体の識別に必要な最小の情報で構成される確認発話を行う最小属性確認水準と, 指示された物体の識別に利用可能なすべての情報で構成される確認発話を行う全属性確認水準で, その認識成功率が変化するかどうか検証した.

実験の結果, 最小属性確認水準での指示物体の認識成功率は 75.6%, また, 全属性確認水準での認識成功率は 54.7%であり, この差は有意であった. すなわち, 指示物体の認識性能は, ロボットが指示物体の識別に関するすべての情報を用いるように確認発話を選択する場合より, ロボットが指示物体の識別に必要な最小の情報を用いるように確認発話を選択する場合に高くなるということが示された. さらに, 性能向上の要因を検討するために, 指示発話の冗長性と認識性能の相関を調べた. その結果, 指示発話の冗長性と認識性能の間には強い正の相関が認められた. また, 各水準における指示発話の冗長性の変化を検証したところ, 確認要因の主効果に有意傾向が認められた. これらの結果は, 冗長性のない指示発話よりも冗長性のある指示発話において指示物体の認識成功率が高く, ロボットが冗長性のない確認発話を行う場合に, 人間は冗長性のある指示発話をする傾向があったことを示唆している.

本研究の特徴的な貢献は, ロボットの振舞いを利用して, 指示物体の認識性能を向上させる点である. そのため, 工学的な手法による指示物体の認識技術と並行して利用することが可能であり, 人々とインタラクションを通じて支援を行うエージェントへの応用も容易であることから, その応用可能性は高いといえる.

謝辞 本研究は, 総務省の研究委託により実施したものである.

参考文献

- [1] Shiomi, M., Kanda, T., Ishiguro, H. and Hagita, N.: Interactive Humanoid Robots for a Science Museum, *IEEE Intelligent Systems*, Vol.22, pp.25–32 (2007).
- [2] Gross, H.-M., Boehme, H., Schroeter, Ch., Mueller, S., Koenig, A., Einhorn, E., Martin, Ch., Merten, M. and Bley, A.: TOOMAS: Interactive Shopping Guide Robots in Everyday Use – Final Implementation and Experiences from Long-term Field Trials, *IEEE/RSJ Int. Conf. Intelligent Robots and Systems*, pp.2005–2012 (2009).
- [3] Iwamura, Y., Shiomi, M., Kanda, T., Ishiguro, H. and Hagita, N.: Do Elderly People Prefer a Conversational Humanoid as a Shopping Assistant Partner in Supermarkets?, *6th ACM/IEEE International Conference on Human-Robot Interaction (HRI2011)* (2011).
- [4] Mutlu, B. and Forlizzi, J.: Robots in Organizations: The Role of Workflow, Social, and Environmental Factors in Human-Robot Interaction, *ACM/IEEE Int. Conf. Human Robot Interaction*, pp.287–942 (2008).
- [5] Richarz, J., Plotz, T. and Fink, G.A.: Real-time Detection and Interpretation of 3D Deictic Gestures for Interaction with an Intelligent Environment, *Proc. Int. Conf. Pat. Rec.*, pp.1–4 (2008).
- [6] Nickel, K. and Stiefelhofen, R.: Visual Recognition of Pointing Gestures for Human-Robot Interaction, *Image Vis. Comp.*, Vol.25, No.12, pp.1875–1884 (2007).
- [7] 間下以大, 岩井儀雄, 谷内田正彦: 人物の指示動作による指示方向のモデル化と推定, *ヒューマンインターフェース学会誌*, Vol.6, No.1, pp.41–48 (2004).
- [8] Schauerte, B. and Fink, G.A.: Focusing Computational Visual Attention in Multi-Modal Human-Robot Interaction, *Proc. 12th International Conference on Multimodal Interfaces and 7th Workshop on Machine Learning for Multimodal Interaction*, pp.8–12 (2010).
- [9] Sugiyama, O., Kanda, T., Imai, M., Ishiguro, H. and Hagita, N.: Humanlike Conversation with Gestures and Verbal Cues Based on a Three-Layer Attention-Drawing Model, *Connection Science*, Vol.18, No.4, pp.379–402 (2006).
- [10] Kemp, C.C., Anderson, C.D., Nguyen, H., Trevor, A.J. and Xu, Z.: A Point-and-Click Interface for the Real World: Laser Designation of Objects for Mobile Manipulation, *Presented at HRI '08, 3rd ACM/IEEE Conference on Human-Robot Interaction*, March 12–15, 2008, Amsterdam, The Netherlands (2008).
- [11] 三谷 岳, 岡 夏樹: ロボットの発話による人の発話の誘導: 意図理解の精度向上の試み, 第 36 回知能システムシンポジウム資料, pp.91–96 (2009).
- [12] 飯尾尊優, 塩見昌裕, 篠沢一彦, 宮下敬宏, 秋本高明, 下原勝憲, 萩田紀博: 語彙の引き込み: ロボットは人間の語彙を引き込めるか?, *情報処理学会論文誌*, Vol.51, No.2, pp.277–289 (2010).
- [13] Giles, H., Coupland, N. and Coupland, J.: Accommodation Theory: Communication, Context, and Consequence, *Contexts of Accommodation: Developments in Applied Sociolinguistics*, Giles, H., Coupland, N. and Coupland, J. (Eds.), pp.1–68, Cambridge, Cambridge University Press (1991).
- [14] Brennan, S.E. and Clark, H.H.: Lexical choice and conceptual pacts in conversation, *Journal of Experimental Psychology: Learning, Memory, and Cognition*, Vol.22, pp.1482–1493 (1996).
- [15] Cleland, A.A. and Pickering, M.J.: The Use of Lexical and Syntactic Information in Language Production: Evidence from the Priming of Noun-phrase Structure, *Journal of Memory and Language*, Vol.49, pp.214–230 (2003).
- [16] Levelt, W.J.M. and Kelter, S.: Surface Form and Memory in Question Answering, *Cognitive Psychology*, Vol.14, pp.78–106 (1982).
- [17] Branigan, H.P., Pickering, M.J. and Cleland, A.A.: Syntactic coordination in dialogue, *Cognition*, Vol.75, B13–B25 (2000).
- [18] Garrod, S. and Anderson, A.: Saying what you mean in dialog: A study in conceptual and semantic coordination, *Cognition*, Vol.27, pp.181–218 (1987).
- [19] Branigan, H.P., Pickering, M.J., Pearson, J., McLean, J.F. and Nass, C.I.: Syntactic alignment between computers and people: The role of belief about mental states, *Proc. 25th Annual Conference of the Cognitive Science Society*, Alterman, R. and Kirsh, D. (Eds.), pp.186–191, Boston, MA, Psychology Press (2003).
- [20] 大坊郁夫: 2 人間コミュニケーションにおける言語活動性の構造, *実験社会心理学研究*, Vol.17, pp.1–13 (1977).
- [21] Shinozawa, K., Miyashita, T., Kakio, M. and Hagita, M.: User Specification Method and Humanoid Confirmation Behavior, *IEEE International Conference on Humanoid Robots* (2007).
- [22] 河井 恒, 戸田智基, 山岸順一, 平井俊男, 倪 晋富, 西澤信行, 津崎 実, 徳田恵一: 大規模コーパスを用いた音声合成システム XIMERA, *電子情報通信学会論文誌*, Vol.J89-D-II, No.12, pp.2688–2698 (2006).
- [23] Iio, T., Shiomi, M., Shinozawa, K., Akimoto, T., Shimohara, K. and Hagita, N.: An entrainment of pointing gestures via robot motion, *The International Conference on Social Robotics (ICSR2010)*, pp.372–381 (Nov. 2010).
- [24] 伊藤 玄, 葦荊 豊, 實廣貴敏, 中村 哲: 音声認識統合環境 ATRASR の概要と評価報告, *日本音響学会 2004 年秋季研究発表会講演論文集*, Vol.1, pp.221–222 (2004).
- [25] 小林亮博, 上田博唯, 佐竹純二, 近岡正樹, 木戸出正継: 家庭内ユビキタス環境における対話ロボットの実稼動実験と対話戦略の評価, *情報処理学会論文誌*, Vol.48, No.5, pp.2023–2031 (2007).



飯尾 尊優

2007 年同志社大学工学部知識工学科卒業。2009 年同大学大学院工学研究科情報工学専攻博士前期課程修了。現在, 同大学院工学研究科情報工学専攻博士課程後期。人工知能, ヒューマンロボットインタラクション, 社会情報

学に興味を持つ。



塩見 昌裕 (正会員)

2004 年大阪大学大学院工学研究科知能・機能創成工学専攻博士前期課程修了。2007 年同専攻博士後期課程修了。博士（工学）。現在，ATR 知能ロボティクス研究所研究員としてコミュニケーションロボットの研究に従事。

ネットワークロボット，コミュニケーションロボット，集団とロボットの相互作用に興味を持つ。



篠沢 一彦

1988 年慶應義塾大学工学部電気工学科卒業。1990 年同大学大学院修士課程修了。同年 NTT 入社。以来，NTT ヒューマンインターフェース研究所において，ニューラルネットワークに関する研究に従事。1998 年より NTT コ

ミュニケーション科学基礎研究所にて，コミュニケーションロボットの研究に従事。現在，国際電気通信基礎技術研究所知能ロボティクス研究所室長。博士（情報学）。



下原 勝憲

1978 年九州大学大学院工学研究科修士課程修了。同年 4 月電信電話公社横須賀電気通信研究所入所。NTT コミュニケーション科学基礎研究所研究部長，ATR ネットワーク情報学研究所長等を経て，2006 年より同志社大

学工学部情報システムデザイン学科教授。2008 年より同大学大学院工学研究科情報工学専攻教授兼任。博士（工学）。関係性のデザイン，社会情報学の研究に従事。著書に『人工生命と進化するコンピュータ』（工学調査会，1998）等。



萩田 紀博 (正会員)

1978 年慶應義塾大学大学院工学研究科電気工学専攻修士課程修了。同年日本電信電話公社（現，NTT）武蔵野電気通信研究所に入所。文字認識や画像認識等の研究に従事。NTT 基礎研究所等を経て，現在，ATR 知能ロボティ

クス研究所所長。工学博士。