

## 科学研究費助成事業（科学研究費補助金）研究成果報告書

平成 25 年 5 月 22 日現在

機関番号：32665

研究種目：基盤研究(C)

研究期間：2010～2012

課題番号：22500179

研究課題名（和文） マルチモーダル言語に基づく生活支援ロボット

研究課題名（英文） Robots for daily life assistance based on a multimodal language

研究代表者

岡 哲資 (OKA TETSUSHI)

日本大学・生産工学部・准教授

研究者番号：00282921

研究成果の概要（和文）：生活空間内における目的地への移動、物体の移動、家電操作を命令できるロボットの評価結果から、音声、タッチスクリーンおよび3次元ジェスチャ入力を用いたマルチモーダル対話を通じて、特別な知識を持たない日本語話者が生活支援ロボットに意図を容易に伝えられることが示された。また、より幅広い支援を受けるためのマルチモーダル言語の設計に有用な知見が得られた。

研究成果の概要（英文）：Some user studies of robots that can be directed to move to goals, move objects, and operate home electrical appliances were conducted. The results show that untrained Japanese speakers can effortlessly communicate with robots for daily life assistance through multimodal interactions using speech, 3D gesture and touch on a screen. Furthermore, they provide new insight into designing multimodal languages for commanding robots to receive wider spectrum of help.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2010年度	2,200,000	660,000	2,860,000
2011年度	500,000	150,000	650,000
2012年度	600,000	180,000	780,000
総計	3,300,000	990,000	4290,000

研究分野：総合領域

科研費の分科・細目：情報学，知覚情報処理・知能ロボティクス

キーワード：ロボット / 生活支援 / インタラクション / 音声 / ジェスチャ / タッチスクリーン / マルチモーダル / 言語

### 1. 研究開始当初の背景

音声認識・言語処理技術、対話処理、計算機ハードウェアの進歩などによって、ユーザと音声言語で対話できる電話応答システム、擬人化エージェント、音声対話インタフェースをもつロボットの研究・開発の報告が数多くなされていた。音声インタフェースは、従来のインタフェースに比べて、習得と利用が容易であると考えられている。しかし、誤認識、雑音や音声入力以外の発話などによる誤検出、自然言語の曖昧さ、自然言語理解の計

算コストなど実用上いくつかの未解決の問題が残されていた。

また、視線、表情、手ぶり、指さしなど言葉以外の（ノンバーバル）伝達様式が利用できる対面コミュニケーションの方が電話応答よりも自然であることから、ロボットのインタフェースにおいてもノンバーバル伝達様式が利用できた方が、利用者の認知的な負荷が小さく、より自然であると考えられている。Perzanowskiらは、音声言語と指さしを組み合わせたロボットのマルチモーダルイ

インタフェースを実現した。また、Ibaらは音声(“go”などの一語命令文)と手ぶりを組み合わせてロボットを制御、プログラムする方法(マルチモーダルプログラミング)を提案した。研究代表者らは、家庭用ロボットの利用者のための人工マルチモーダル(複数の伝達様式による)命令言語RUNAとRUNAによって命令が可能なロボットシステムを複数開発し、100人以上の利用者による評価を行った。RUNAは、日本語に基づく音声言語と音声の語句を代用するノンバーバルイベント群で構成される。音声言語は、家庭用ロボットへの命令を行為の種類(例えば、前進、旋回、物体の持ち上げ、エアコンの温度変更、掃除など)と複数のパラメータ値(速度、移動量、角度、物体の幅、温度、掃除範囲など)を音声で曖昧さなく一意に伝えることを可能にするものであった。

RUNAの音声命令は、例えば、「ゆっくり30cm(少し)前進しろ」「床から幅25cmの箱(中くらいの箱)を持ちあげろ」「周囲1mを丁寧に掃除しろ」などの命令文である。ノンバーバルイベント群は、利用者の手動き、ロボットへの接触、ボタン操作で発生し、行為のパラメータ値を音声の代わりに伝達する。例えば、上の「ゆっくり」「30cm」はボタン操作や手の動きで代用できる。

開発したロボットの利用者評価によって、RUNAの有用性と問題点が明らかになった。

90分以内の短い利用経験でも、意図伝達の成功率は90%以上であり、遠隔操作におけるロボットの誘導、物体の移動、掃除、エアコンの操作などのタスクを殆どの利用者が達成できたことから、RUNAは、日本語話者に利用可能なマルチモーダル言語であると言える。また、ロボットの応答時間は1秒以内と短く、一つの命令に要する時間は発話時間にはほぼ等しい。RUNAのノンバーバルイベントには、「音声命令の語数を減らし、発話のわずらわしさ(利用者の言語的な負荷)と音声認識誤りを避ける」、「発話を容易にし、言い間違いを避ける」、「命令時間を短縮する」、「ロボットへの発話開始の合図となる(音声入力誤検出を減らす)」などの利点があることも分かった。

しかし、RUNAの音声命令には、「大きく」などの定性的な表現では、指定できるパラメータ値が限られている、「25mm」などの定量的な表現は不自然であり、実空間における量を目測で数値化することが難しい(指定したい長さや角度の数値が正確に分からない)、などの問題がある。手の動きやボタン操作などのノンバーバルを用いる場合も、数値化は不要であるが、一般に目測との正確な対応付けは難しい。したがって、利用者の真の意図とロボットの行為には誤差が生じやすい。また、ノンバーバルイベントを用いると、短時

間で複数のパラメータを同時指定できる反面、パラメータ指定が難しいケースや、指定方法が恣意的で覚えにくいケースもみられた。例えば、単眼の手のジェスチャ検出では、物体や動きの大きさ・距離の情報が正確に得られないため、行為パラメータの指定が大きく制限された。また、ロボットへの接触やボタン操作では、接触場所(ボタン)、回数、時間と行為パラメータの覚えやすく、指定しやすい対応付けが難しい。さらに、利用者が「命令の解釈中や実行中に次の命令を行う」、「命令をやり直す」、「独り言を言う」、「周囲と話す」などの行動をとった場合に混乱が生じ、スムーズな情報伝達が行えないことがあった。

上記の問題の解決には、意図伝達をより容易にするマルチモーダル命令言語の設計とロボットから利用者への適切な情報伝達が不可欠であると考えた。利用者からロボットへの情報伝達では、特にノンバーバルの使用方法が重要である。RUNAのように複数のパラメータ値全てを同時に伝達するのではなく、対話的に一つずつ追加できればよいと考えた。すなわち、命令のタイプといくつかのパラメータ値を音声とノンバーバルで同時に伝えたのち、不足する情報を対話的に伝える方法である。また3次元視覚認識を応用すれば、距離、方向、位置などの伝達が容易になると考えた。さらに、ロボットから利用者へもマルチモーダルの情報伝達を行うことで、利用者の意図伝達をスムーズかつ容易なものにできると考えた。このようにして、ロボットと利用者間の双方向情報伝達のためのマルチモーダル言語を適切に設計すれば、生活支援ロボットの実用化に貢献できると考えた。

## 2. 研究の目的

生活支援ロボットの利用者がロボットから支援を受けるための意図伝達言語の設計指針と具体的な仕様を提案すること、言語による意図伝達の容易さ、言語の習得の容易さ、言語の問題点と満たすべき条件などを明らかにすることを目的とした。さらに、生活支援ロボットを開発し、利用者の目的達成の容易さと問題点を検証することを目指した。

## 3. 研究の方法

まず、生活空間内でのロボットの移動、物体の移動、家電操作の意図伝達を行うためのマルチモーダル言語をそれぞれ設計し、それらに言語に基づく生活支援ロボットシステムを開発した。次に、個別のシステムを用いて、複数の利用者による評価を行った。さらに、評価結果を検討し、多様な生活支援における意図伝達のための言語の仕様を検討した。

まず、タッチスクリーンと音声入力によって、ロボットの移動と物体の移動を命令するためのマルチモーダル言語を設計し、シミュレータ上のロボットによる評価を行った。この言語でロボットを目標位置に移動させるためには、「移動しろ（ここに行け、すすめ）」という音声とロボットに搭載されたカメラ画像のタッチを組み合わせる。また、画像のタッチと「旋回しろ（こっちをむけ、まわれ）」の音声によって、ロボットの向きを変えることもできる。さらに、「前進」、「後退」、「右に旋回」、「左に旋回」の音声で動作を開始し、ボタンでの停止と速度変更も行えるようにした。物体を移動するには、「これをつかめ（つかめ、取れ）」、「ここに置け（置け）」の音声と画像のタッチを組み合わせる。関連して、円の内部のタッチによる半径 1m 以内の移動先および旋回方向の伝達、ロボットアームの旋回による物体の方向の伝達についても調査した。

次に、距離画像センサを搭載した実ロボットシステムを実現し、音声とジェスチャでロボットの移動と旋回の命令について評価を行った。「こっちをむけ」・「こっちにこい」と手を上げるジェスチャ、「これだけこい」と手を手前に引くジェスチャ、「これだけさがれ」と手を前に出すジェスチャを組み合わせるものである。

最後に、実ロボットシステムを用いてテレビ、ラジオ、冷房、暖房の操作について評価を行った。電源のオンオフ、音量、チャンネル、温度、風量の変更を「～をつける（消せ）」、「～の（音量・温度・風量）を上げる（下げる）」、「～のチャンネルを変えろ」という音声と右手、左手のジェスチャで命令する。右手ジェスチャは、「つける」、「上げろ」、「（正の方向に）変えろ」、左手ジェスチャは、「消せ」、「下げろ」、「（負の方向に）変えろ」と対応させた。音量などの値の設定は、画面表示を確認しながらジェスチャで行う。

利用者による評価は、個々の意図伝達と課題達成の二種類を実施し、質問紙、理解度テスト、記憶テストも行った。

以上の伝達方法の評価結果に基づいて、多様な生活支援のための意図伝達を可能にするマルチモーダル言語の仕様を検討した。

#### 4. 研究成果

音声、タッチスクリーンおよび 3 次元ジェスチャ入力を用いたマルチモーダル対話を通じて、特別な知識を持たない日本語話者が生活支援ロボットに意図を容易に伝えられることが示された。また、より幅広い支援を受けるためのマルチモーダル言語の設計に有用な知見が得られた。

以下に利用者評価結果から得られた知見を示す。

・評価したロボットの移動、物体移動、家電操作の音声入力の語彙と文法の範囲では、初心者でもスムーズに発話できる。

・意図伝達内容を限定した音声認識文法を用いることで初心者に対しても高い音声認識率が実現できる。

・音声入力の例文の記憶は容易である。ジェスチャと合わせて 3 回ずつ家電操作命令を行った利用者は翌日でも殆どの文を想起できた。

・初心者は、まれに例示されていない文を生成するが、例文のパターンや意味を大きくはずれることはない。

・音声と画面タッチやジェスチャを同時に組み合わせることで、初心者でも意図の概要を容易に伝えることができる。また、このときのノンバーバル入力は音声の誤検出と誤認識の問題を軽減できる。

・画像内の物体や目標をタッチして意図を伝達することは容易である。ただし、床・テーブルの上の位置などの伝達は難しい。

・画像のタッチによりロボットの視野内の方向を伝達することは容易である。

・画面内の円のタッチによる移動目標の伝達は、実空間と画面の対応の負荷があるものの、短距離の移動と位置の微調整に有効である。

・初心者でも画面上に表示したボタンによって、停止、速度などの変更の意図を伝達できる。画面表示でのロボットの誘導によって、利用者に要求される知識を減らすことができる。

・安価な距離画像センサを用い、1.5-3m の距離で、3 次元的なジェスチャの高い認識率が実現できる。まれにジェスチャの見落としがあるものの、誤検出や誤認識は殆どない。したがって、3 次元ジェスチャは、生活支援ロボットへの意図伝達に有効である。

・手の前後への移動によって、ロボットの移動量を伝達する方法は容易であり、間違いにくい。また、ロボットの移動量を一定の精度で伝えることができる。

・「こっちにこい」の命令でも手を手前に引く初心者がいた。移動方向の対応付けは自然であることが分かる。

・家電の設定変更の伝達に手の上げ下ろしが実用上利用可能である。

・ロボットが画面表示により利用者による命令の可否を伝えることで、初心者も混乱なく意図伝達が行える。

利用者の目的達成をより確実かつ容易にするための技術課題としては、多様な生活支援に対応した音声認識文法での音声認識、生活環境における離れたマイクでの音声認識、

より確実な利用者の発見と追跡 (1m 以下の近距離を含む)、ロボットの物理的な作業による目的達成を容易にする対話方式 (遠隔操作技術の応用など) などが挙げられる。

以上の知見により、ロボットから幅広い支援を受けるための意図伝達言語の設計指針が明確になり、具体的な仕様を提示する準備が整った。以下の方法で対話的な意図伝達を行えば、ロボットの利用者は、時間をかけずに確実な意図伝達を行い、生活支援の目的を達成できると考える。

- (1) ロボットは、意図伝達が可能な状態にあることを画面表示、ジェスチャ、視線等で利用者に知らせる。
- (2) 利用者は、音声とノンバーバル (3次元ジェスチャ、画面タッチ、ボディタッチ) を同時に組み合わせたマルチモーダル入力で、意図の概要をロボットに伝える。
- (3) 意図の詳細は、ロボットの誘導にしたがって画面表示等を利用しつつ、ノンバーバル入力に対話的に伝達する。

(2)の音声入力 (言語) には、本課題で開発した各文法をベースに、多様な意図伝達に対応し、一貫性のある文法を開発する。簡潔かつ自然な表現にすることによって、日本語話者が自然かつ容易に発話できると考える。例文から文法規則と伝達可能な意味を把握できること、使用語彙に一貫性をもたせることが重要である。ノンバーバル入力は、音声認識の弱点をカバーするための冗長入力としての役割を果たす。(3)の対話方式 (言語) は、利用者が知識なしで直感的に入力が行えることが重要である。画面表示等を利用し、音声対話よりも高速な意図伝達を実現する。

上記の方法は、音声・タッチスクリーン・ジェスチャのみの意図伝達の弱点を互いに補う。従来のマルチモーダル命令よりも時間はかかるが、確実な伝達が行え、利用者が目的を達成しやすいと考えられる。

タッチスクリーンは、音声入力の後、対話的に意図を伝える際に特に有効であり、ロボットや物体の移動などの物理的な操作も容易にする。3次元ジェスチャによる伝達は、手元にタッチスクリーンがない場合や、タッチスクリーンを用いるよりも自然な場合 (呼び寄せ、物の受け渡し、向き合っている状況での意図伝達など) に有効である。

生活支援ロボットによる目標達成に関しては、ロボットの位置座標や移動量、ラジオの音量、冷房の設定温度など、利用者が明確な目標値をもたない場合があることに注意が必要である。漠然とした欲求や目標のイメージから目的を達成するには、画面上で良い

目標値を決定する方法と、結果を見ながら値を微調整する方法が重要になると考える。

## 5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計 3 件)

- ① 岡 哲資、松本大賢、木林龍一、A multimodal language to communicate with life-supporting robots through a touch screen and a speech interface、Artificial Life and Robotics、16、292-296、2011、査読有り DOI: 10.1007/s10015-011-0924-x
- ② 岡 哲資、阿倍豊和、杉田薫、横田将生、User study of a life-supporting humanoid directed in a multimodal language、Artificial Life and Robotics、16、224-228、2011、査読有り DOI: 10.1007/s10015-011-0923-y
- ③ 岡 哲資、阿倍豊和、杉田薫、横田将生、Commanding a humanoid to move objects in a multimodal language、Artificial Life and Robotics、15、17-20、2010、査読有り DOI: 10.1007/s10015-010-0758-y

[学会発表] (計 3 件)

- ① 岡 哲資、島津正次郎、中村喜宏、音声と手のジェスチャーによる家庭用ロボットへの家電操作命令、情報処理学会第75回全国大会、2013年 3月8日、東北大学(宮城県仙台市)
- ② 岡 哲資、木林龍一、松本大賢、Gesture detection based on 3D tracking for multimodal communication with a life-supporting robot、17<sup>th</sup> International Symposium on Artificial Life and Robotics (AROB17)、2012年 1月21日、ビーコンプラザ(大分県別府市)
- ③ 岡 哲資、松本大賢、木林龍一、音声とタッチパネルによるロボットの物体移動の命令方法、日本機械学会ロボティクス・メカトロニクス講演会11、2011年5月28日、岡山コンベンションセンター (岡山県岡山市)

## 6. 研究組織

(1) 研究代表者

岡 哲資 (OKA TETSUSHI)  
日本大学・生産工学部・准教授  
研究者番号: 00282921