

簡素な音声言語と触れ合いで指示が可能なロボット

A robot that can be directed in a simplified spoken language and by tactile interaction

正 岡 哲資 (福岡工大) 佐々木 崇 (福岡工大)

Tetsushi OKA, Fukuoka Institute of Technology, oka@fit.ac.jp
Takashi SASAKI, Fukuoka Institute of Technology

This paper demonstrates a prototype test bed for usability studies of a multi-modal language for directing robots, using a Sony AIBO and JULIAN, a grammar-based speech recognition engine. We developed a grammar for spoken commands based on Japanese in order to simplify command understanding and an event-driven system to interpret multi-modal directions. In our pilot studies, we confirmed our robot can properly react to spoken, non-verbal and multi-modal directions by humans. Results from recent experiments involving subjects unfamiliar with robots are also reported and discussed.

Key Words: Robot, Direction, Spoken Language, Tactile Interaction

1. 緒言

ロボットが家庭などで一般ユーザに使われる状況では、ユーザがロボットに容易に意図を伝えられることが重要になる。ロボットのハードウェアは、家事やエンターテインメントといったロボットそのものの機能に加えて、新しいユーザインタフェース機器としての可能性も持っている。例えば、家庭用ロボットは、図 1 に示すように、人間からの命令を受け、物理的なタスクを実行するとともに、家庭内のコンピュータネットワークやインターネットと接続し、情報サービス、家電や機器類の操作、留守番などの様々なタスクを実行するように、多目的化していくと考えられる。このようなロボットの実用化には、人とロボットの円滑なコミュニケーションが極めて重要である。このようなことから、著者らは、誰でも簡単に使えるロボットの実現を目指した研究を行っている。

誰でも使えるロボットの課題の一つとして、特別な訓練なしで、ロボットに直感的に命令を与えられるようにすることがある。音声、ジェスチャ、触れ合いなど、人間同士、あるいは人間と動物のコミュニケーションに近い方法でロボットとコミュニケーションが行えれば、ユーザへの負担が減り、ロボットの実用性が高まる。このような考えから、著者らは、家庭用ロボットを目的としたマルチモーダル命令言語とその処理系、直感的な命令が可能なロボットシステムの開発と評価を行っている[1][2][3]。

本稿では、これまでに開発した音声言語と触れ合いによって指示が行えるロボットについて述べる。ロボットがマルチモーダル命令言語を解釈し、実行するためのシステムアーキテクチャと構成要素、音声命令言語、非言語情報の扱い、システムの評価方法を論じ、これまでの実験結果を示す。

2. ロボットの指示とインタフェース

人とロボットのコミュニケーションの研究には、様々なアプローチが存在する[4][5][6]。例えば、人とロボットとの言語対話の研究では、音声認識と合成、自然言語の理解と生成、対話制御によって、意味のあるコミュニケーションを実現させようとしている。しかし、人間同士のコミュニケーションや言語対話のモデルをそのまま人とロボットのコミュニケーションに応用するには、言語理解を始めとした、多くの課題が残されている。したがって、家庭用ロボットなどの実用化を考えれば、計算コストを考慮した人とロボット固有のコミュニケーションのモデルに基づく、直感的に分かりやすいインタフェースを実現すべきである。さらに、ロボットとのイ

ンタフェースで基本になるのは、人間からロボットの命令であることから、命令を中心にコミュニケーションを考えるのが自然である。

以上の考察から本研究ではユーザがロボットに迅速かつ容易に命令(個々の動作、アクション、タスク、ゴール、任務)を伝えることを最重要課題と考え、実用化のアプローチとして、人間が覚えやすく、ロボット側の情報処理のコストも小さい命令言語を開発する。ここでの命令言語は、自然言語をベースにした音声言語が中心になるが、非言語情報を併用できる“マルチモーダル命令言語”である。また、中心となる音声言語は、命令文を基本として、極力曖昧性を排除し、語彙と文法規則を慎重に選ぶことで、ユーザが学習しやすく、ロボットが命令を特定しやすいものを目指している。

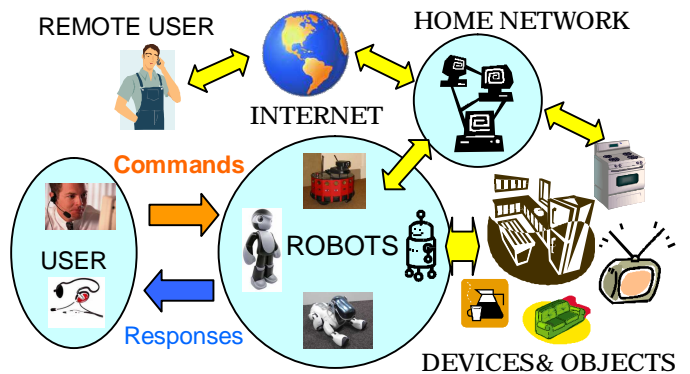


Fig. 1 Multi-modal interaction with home robots

3. ロボットシステムの開発

3.1 システムアーキテクチャ

図 2 にマルチモーダル命令言語の解釈・実行システムのアーキテクチャを示す。ユーザの音声は、音声認識モジュール(ASR)に送られ、音声イベントの検出に用いられる。音声イベントには、音声認識結果など、命令解釈に必要な情報が含まれる。また、視覚や触覚などのセンサ情報に基づいて非言語イベントを検出するモジュール群によって、非同期にジェスチャや触れ合いのイベントが生成される。全てのイベントは、命令解釈モジュールに送られ、イベントキューに格納される。発生時刻の近いイベントは関連付けられ、一つの命令として解釈される。これによって、例えば体の右側を触ってから、「90度、ゆっくり回れ」と命令することでロボットを90度右旋回

させることができるようになる。

命令解釈の結果は、ロボットのアクションの種類と各種パラメータであり、アクション生成モジュールに送られ、アクションが実行される。

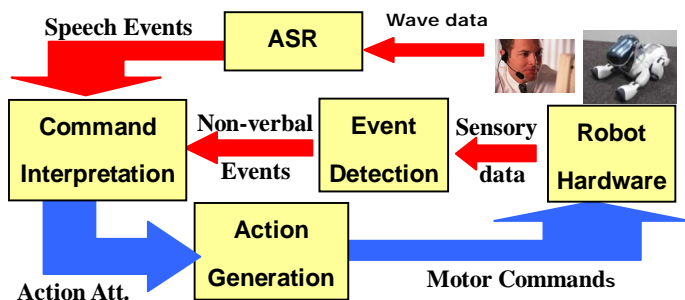


Fig. 2 System architecture for multi-modal directions

3.2 AIBO を用いたプロトタイプシステム

図 2 のアーキテクチャに基づくプロトタイプシステムを Sony AIBO ERS-7 と Windows PC 1 台を無線 LAN 接続したハードウェア上で実現した。音声認識には、PC 上で文法ベースの音声認識エンジン JULIAN (<http://julius.sourceforge.jp/>) を用い、音声命令の文法を記述した。AIBO 上のプログラムは、Open-R 開発環境を用いて C++ で記述した。非言語イベント検出、命令解釈、アクション生成は、全て AIBO 上の CPU で動作している。非言語イベント検出は、AIBO の接触センサなどをモニタしつつ行われ、PC から送られてくる音声イベントと合わせて命令の解釈が行われる。

3.2 音声命令言語 MODLIN-0

ロボットに命令を行うための音声言語のプロトタイプ MODLIN-0 (the Multimodal Language for Instructing Robots) を JULIAN の音声認識用文法として作成した。命令の意味の理解に複雑な推論を必要としないように、曖昧な発話を除外するとともに、発話の長さや命令文の構造を限定し、多目的ロボットの様々な命令に対応できるように文法規則を決定した。プロトタイプ言語の語彙は、移動ロボットの基本動作の命令に対応できるように最小限のものを選定している。文法規則により、「立て」、「進め」などの 1 単語による指示だけでなく、「ゆっくり」、「大きく」、「10cm」、「90 度」などのパラメータの指定、「右を見る」、「手を挙げろ」などの目的語を含む自然な発話も認識可能としている。

MODLIN-0 の文法規則の一部を表 1 に示した。ユーザの音声命令は、アクションの指定 (ACTION) やロボットの実行したアクションの修正 (MODIFIER) などで構成される。アクションは、タイプに分類され、タイプごとに、速度、方向などの異なるパラメータを指定できるようにした。MODLIN-0 の文法規則には、再帰がないため、受け付ける発話の長さは限定されるが、規則を細かく作成することで、「30 度右に回れ」と「右に 30 度回れ」の同じ意味の異なる表現に対応した。表 2 は MODLIN-0 の語彙の一部である。

以上の文法規則と語彙で、「ゆっくり下がれ」、「右を向け」などの自然な音声命令を網羅している。図 3 は、「8cm 後ろに下がれ」という音声命令を MODLIN-0 を用いて解析した構文解析木である。この文法では、再帰規則を用いた文法よりも非終端記号が増え、木がやや複雑になり、語の細かい分類も必要になる。しかし、このような文法は、発話の長さの制限や曖昧性解消などによって、意味解析のコストにおいて有利で

あると考えている。

Table 1 Partial list of grammar rules of MODLIN-0

文法規則	説明	レベル
$S \rightarrow \text{ACTION}$	アクションの指定	0
$S \rightarrow \text{MODIFIER}$	アクションの修正	0
$\text{ACTION} \rightarrow \text{ACTION_T1}$	アクションのタイプ	1
$\text{MODIFIER} \rightarrow \text{MORE SPEED}$	速度の修正	1
$\text{ACTION_T1} \rightarrow \text{P1 A1}$	パラメータの指定	2
$\text{P1} \rightarrow \text{SPEED}$	タイプ 1 の速度	3

Table 2 Part of MODLIN-0's lexicon

非終端記号	終端記号	日本語
MORE	md_1_2_mochotto	もうちょっと
SPEED	speed_0_0_hayaku	はやく
A4	at_4_1_susume	すすめ
LUNIT	lunit_1_1_cm	センチ(cm)
DIGIT	number_6_1_6	ろく(6)
WO	joshi_1_1_wo	を

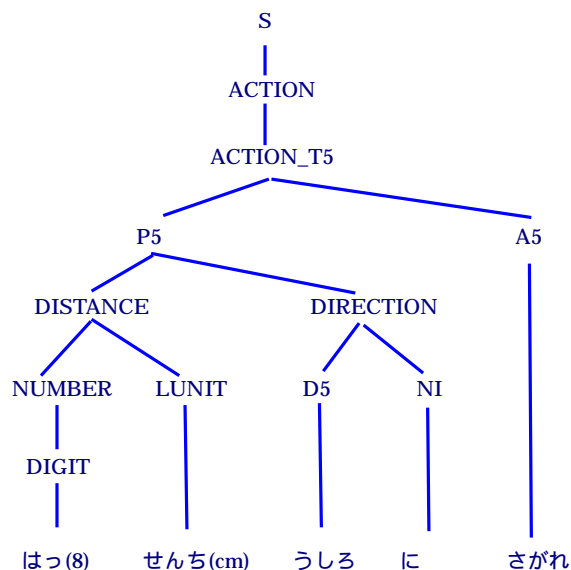


Fig. 3 Parse tree of MODLIN-0 for “Go 8cm back!!”

3.3 非言語情報の利用

コミュニケーションの非言語情報としては、触覚、距離センサ、視覚、聴覚（音声以外）情報などが効果的である。プロトタイプシステムでは、例えば、AIBO の頭、あご、背中、接触センサの出力を定期的に調べ、触られた場所、時刻、長さ(時間)などの情報を含む触覚イベントが生成される。また AIBO のピンクボール追跡によって、ボールの移動方向などで分類された視覚イベントが生成される。イベントの処理のタイミングとしては、全てのイベントの発生時、音声イベントの発生時、タイマによる定期的なイベントの発生時などの複数の方法が試されている(図 4)。マルチモーダル言語の設計では、非言語イベントとしてどのようなものを用いるかを明確に定義する必要があり、重要な研究課題となっている。

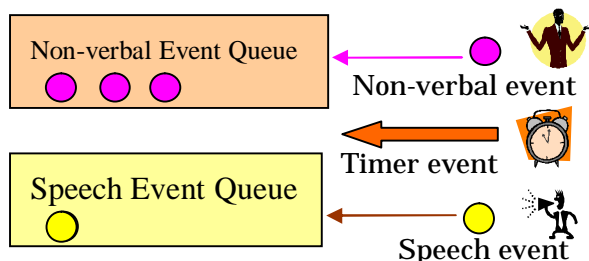


Fig. 4 Event handling for command interpretation

3.3 アクションの実行

命令解釈モジュールの作成する命令の意味表現は、ロボットが実行可能なアクションの一つを特定するための情報を含むものになっている。例えば、「ゆっくり回れ」という音声命令のみが与えられた場合は、アクションの種類と速度のパラメータが特定され、その他のパラメータは、デフォルト値が用いられる。現在のプロトタイプシステムの実装では、実行可能なアクション(立つ、座る、前進、後退、旋回の各バージョン)の中から1つが選択される(図5)。最終的に、各アクションのためのモータへの指令の時系列が生成される。

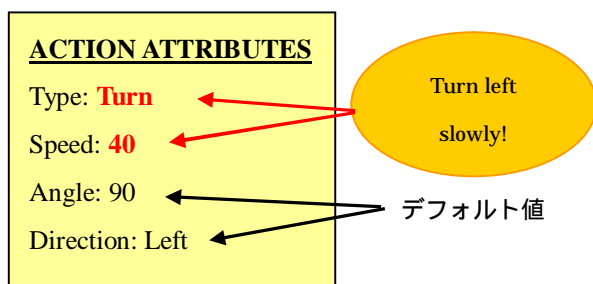


Fig. 5 Action attributes

4. ロボットシステムの評価

4.1 評価方法

マルチモーダル命令言語とロボットシステムの評価においては、システムと人のインタラクションによる実験が不可欠である。システムの内部を良く知っている開発者による操作実験では、チェック項目を決め、使いやすいシステムの条件が満たされているか、ロボットの反応時間、計算時間、イベント発生記録が行われ、分析がなされる。また、このような研究においては、システムについての知識のない一般ユーザ、特にロボットやコンピュータに慣れていない被験者による実験は、特に重要である。被験者にタスクを与えて、ロボットとのインタラクションをビデオ撮影などで記録する。タスクの達成率、タスク完了時間、学習に要する時間などのデータに加えて、アンケートによるユーザの評価も行う。

4.2 結果の分析

実験で得られたデータ、ビデオ、アンケート結果を整理、分析することにより、ロボットの命令しやすさを評価し、命令言語およびシステムの問題点、改良点を明らかにする。客観的なデータとしては、被験者による実験で用いられた単語や構文、非言語メッセージの種類、頻度、タスクの完了時間、実際に検出された非言語イベントなどが挙げられ、詳細かつ綿密な分析が可能になる。さらに、アンケートにより、ユー

ザの主観的評価を確認し、条件の異なる実験の比較や個別意見に基づいて改良点を発見する。

4.3 これまでの実験結果と考察

これまで、実現したロボットシステムを用いた実験を行ってきた。そのうち、被験者による実験では、以下の3つのタスクを与えた。

1. 立ち上がらせ、すぐに座らせる。
2. 180度向きを変えさせる。
3. 指定した目標地点に移動させる。

これらのタスクに対して、ロボットの行えるアクション、認識する言葉などについての予備知識を与えずに、音声のみ、触れ合いのみ、音声と触れ合いの組み合わせで実験を行った。被験者は、AIBOに近づいて触ったり、PCに接続されたマイクロフォンで声をかけたりすることができる。

実験後に行ったアンケートの設問と結果を表3に示す。回答は、5(非常にあてはまる)から1(全くあてはまらない)までの5段階で、各被験者の平均値を示している。この他、全体の印象、良かった点、悪かった点・改善点を自由に記述してもらった。自由記述からは、触れ合いや音声の命令は、予備知識なしでも比較的容易に覚えられ、良い印象を与えたものの、旋回の動きが分かりにくいという意見が多くみられた。また、命令以外の声や音への反応が気になることや、言い回しや単語により自由度が欲しいという意見も見られた。後者は、言い回しとともに命令の種類にも関係するものと思われる。

タスクの完了時間(分・秒)と達成率(%)を表4に示す。表4から、この実験では音声のみの場合が最もタスク完了時間が短く、確実に達成されていることが分かる。表3のアンケート結果からも、音声の命令が容易に行えたことがうかがえる。実験では、一切の予備知識を与えていないが、触れ合いのみでも、ロボットの反応を一度理解すれば、スムーズに命令ができることが表4の結果やアンケート結果、ビデオから見て取れた。音声と触れ合いの組み合わせは、現時点のシステムでは、時間の短縮や使いやすさには結びついていないことも分かる。音声を補うための非言語情報の利用方法、複数のモダリティを効果的に組み合わせた命令方法の探究が今後重要となる。

全体的にみると、どの場合もほぼタスクを完了しており、予備知識の与え方や、非言語情報の用い方などシステムの改良によっては、音声と触れ合いの組み合わせで、より良い結果を得る可能性が残されている。触れ合いのみの場合、左右方向を示すのに適切な触覚センサがERS-7にないため、触った長さで指示するようになっており、予備知識なしでは、タスク2がうまくいっていない。ロボットを軽く数回たたいたり、撫でたりする被験者もあり、「触れる圧力が難しかった」など意見があった。これは、触覚イベントの種類と検出条件が重要であることを示している。

その他の興味深い結果は、AIBO ERS-7が犬に似た外見をしていることから、「お座り」のような音声命令を発したり、「おい」「こっち向け」などと注意をうながす声をかけたりする被験者がいたことである。後者は、ジェスチャ認識、顔発見、音源定位システムとの統合によって、より使いやすいインタフェースが実現できることを示唆している。

また、旋回などのアクションのスムーズさや、細かい動きなど、ロボットの動きのレパートリーも使いやすさに大きな影響を与えることが様々な結果から確認された。

5. 今後の研究方針

現在 AIBO のプロトタイプシステムを用いた評価実験を継続しており、ジェスチャを含めた指示、予備知識の与え方などによって、様々な有益な実験結果が得られることが期待される。今後、車輪移動ロボットや小型ヒューマノイドロボットによるシステムの実装と評価実験を行い、家庭用をターゲットに幅広い機構や機能のロボットの命令のためのマルチモーダル言語と基盤システムの開発を進めていく予定である。これまで述べたように、音声命令の文法と語彙の問題と非言語イベントの用い方は特に重要であり、今後、様々な実験を繰り返すことによって、より良いマルチモーダル言語とロボットシステムを目指していく。

Table 3 Questionnaire and results

アンケート設問	P	S	SP
操作性はよかった	2.8	4	2.6
反応速度は適切だった	3.8	4	3.6
言ったとおりに動いた	-	4	3.2
さわったとおりに動いた	4	-	3.6
思ったとおりに声に反応した	-	3.8	3.2
思ったとおりに触って反応した	3.6	-	3.2
実験を楽しめた	4	4.4	4.4
声と触れ合いで動かせるのは良い	-	-	3.8
声で動かせるのは良いと思う	-	4.8	4.4
触って動かせるのは良いと思う	4.6	-	3.8
思い通りに動かずイライラした	3.8	2.6	3.2
実験をスムーズに行えた	2.4	3.8	2.6
ロボットを自由に動かせた	3.2	3.6	2.4
このような手段で動くロボットを欲しい	3.6	4.2	2.6
このような手段で扱えたら便利になる	4.2	4.8	2.8
自由に命令できるロボットがあれば使用してみたい	3.8	3.8	3.8
簡単に命令できた	2.8	4	2.8
命令方法が難しかった	4	3	3.8

6. 結言

使いやすいロボットを実現するための音声と触れ合いによる命令方法を探求するために、著者らが開発しているロボットのシステムアーキテクチャ、構成要素、音声命令言語、非言語情報の扱い、システムの評価方法、これまでの実験結果について述べた。音声と触れ合いによって、ロボットの移動のための基本動作を指示することが可能になっており、予備知識のないユーザでも訓練を行わずにロボットを動かせることが確認された。今後は、学習時間の検証、ある程度の予備知識を与えた場合の実験、音声と触れ合いの効果的な組み合わせ方法の探究、より複雑なタスクへのシステムの適用などで、現在のシステムの評価と改良を行うことが課題となる。さらには、同じアーキテクチャと命令言語を異なるロボットとタスクに適用していきたい。

文 献

- [1] 佐々木崇, "音声と触れ合いによって自由に命令できるロボット", 福岡工業大学情報システム工学科平成 18 年度卒業論文, 2007.
- [2] 松尾俊哉, "視覚と触覚により命令を理解するロボット", 福岡工業大学情報システム工学科平成 18 年度卒業論文, 2007.
- [3] Oka, T., and Yokota, M., "Designing a multimodal direction language for multipurpose home robots," *Proc. of 12th International Symposium on Artificial Life and Robotics (AROB12)*, Jan. 2007.
- [4] 石黒浩 神田崇行 宮下敬宏 "コミュニケーションロボット" オーム社 2005
- [5] 佐藤知正編著 "人と共存するコンピュータ・ロボット学", 第 1 章～第 2 章ヒューマンロボティクス pp.1--61 オーム社 2004
- [6] Prasad, R, Saruwatari, H. and Shikano, K. "Robots that can hear, understand and talk," *Advanced Robotics*, Vol.18 pp.533—564, 2004.

Table 4 Time to complete tasks (success rate)

	タスク 1	タスク 2	タスク 3
触れ合い	1'08" (100)	7'25" (60)	5'49" (100)
音声	0'59" (100)	2'41" (100)	4'37" (100)
組み合わせ	2'12" (80)	3'56" (100)	7'23" (100)

P: 触れ合い S: 音声 SP: 音声と触れ合い