

人間型ロボットの対話インタフェースにおける 発話交替時の非言語情報の制御

横山 真男^{†,☆} 青山 一美[†] 菊池 英明[†]
帆足 啓一郎^{†,☆☆} 白井 克彦[†]

人間型ロボットのコミュニケーション能力を人間のそれに近付けるために、人間同士のコミュニケーションにおいて重要な役割を持つ視線や手振りなど非言語情報の利用を検討した。本論文では、まず人間同士の対話において、各種非言語情報の出現タイミングについての分析を行う。さらに、ロボット側の非言語情報の出力タイミングによる対話への影響を分析する。分析の結果、非言語情報の種類による発話交替における制約としての強さや自然性の違いが明確になった。また、非言語情報の出力タイミングとして、人間同士と同様に発話開始直後あるいは終了時が自然かつ円滑な対話の実現にとって適切であることが確かめられた。最後に、ロボットへの視線情報制御の適用を行い、インタフェース評価実験の結果より、ロボットの対話インタフェースにおける非言語情報制御の有効性について述べる。

Controlling Non-verbal Information in Speaker-changing for Spoken Dialogue Interface of Humanoid Robot

MASAO YOKOYAMA,^{†,☆} KAZUMI AOYAMA,[†] HIDEAKI KIKUCHI,[†]
KEIICHIRO HOASHI^{†,☆☆} and KATSUHIKO SHIRAI[†]

In this research, we consider the use of non-verbal information in human-robot dialogue to draw the communication ability of robots closer to that of human beings. This paper describes analysis of output timing of non-verbal information in the dialogues between human beings. Moreover, we analyse influences of the output timing by controlling it in the dialogue of a CG robot. As the result, we clarify the strength of constraint and naturalness of various types of non-verbal information. Also, we confirm that appropriate output timing of non-verbal information is during or at the end of utterances, which is the same as in human-human dialogue. At last, we applied non-verbal information to the humanoid robot and made similar experiments. As a result, non-verbal information made speaker-changing more smoothly for the humanoid robot than in the case of the CG robot.

1. はじめに

近年、ヒューマノイドロボット（人間型ロボット）の開発がさかんに行われてきている¹⁾。ヒューマノイドロボットには、産業における生産効率向上のためだけでなく人間の生活環境における活躍も期待され、家事や高齢者・身障者介護のように人間に対する補助や代替作業、共同作業などが求められる。いわば「ひとにやさしいロボット」の登場に期待が寄せられている²⁾。

そのようなロボットに求められる人間とのコミュニケーション能力の課題に対して、人間が日常的に用いている情報伝達手段、たとえば音声インタフェースをロボットに持たせることが考えられる³⁾。近年ヒューマンインタフェースの研究開発がさかんに行われており、音声インタフェースの有効性が様々な観点において示されている^{4),5)}。最近の飛躍的な計算機の性能向上と音声処理技術の進歩により、大語彙連続音声の実時間処理が可能となり⁶⁾、音声認識ソフトも市販されるまでになっている。そのような背景を受けて、人間とロボットの対話における音声インタフェースの研究も行われてきた⁷⁾。

しかし、視覚による情報を共有する実空間でのコミュニケーションにおいては、音声のみの情報のやりとりでは不都合が生じる状況も多い。たとえば、道案

[†] 早稲田大学理工学部
School of Science and Engineering, Waseda University

[☆] 現在、株式会社東芝
Presently with TOSHIBA Co.

^{☆☆} 現在、株式会社 KDD 研究所
Presently with KDD Research and Development Laboratories, Inc.

内タスクなどにおける方向指示のように、「ことば」だけの説明よりも、指差しなどの視覚的な情報を付加した説明の方が理解しやすい。また、周りに複数の人間がいるような状況ではロボットがだれに向かって発声しているかなども、音声情報だけでは分かりにくい。ロボットの顔の向きや視線、ジェスチャなどの「非言語情報」によって自然に人間が判断できるのが良いインタフェースといえる^{8),9)}。

人間同士の普段のコミュニケーションにおける非言語情報¹⁰⁾には、ジェスチャといわれる身体動作の他に、姿勢、対人距離、服装・服飾等の外見、身体的特徴などが含まれており、人間のメッセージ伝達の65%¹¹⁾あるいは93%¹²⁾が非言語情報で占められているともいわれている。人間同士のコミュニケーションにおいて、非言語情報が重要な役割を担っていることはいうまでもない。人間対機械系の対話においても、同様にスムーズなコミュニケーションを図るための重要な要素であることが指摘されており^{13),14)}、ロボットにおいても非言語情報の利用を見逃すことはできない。

本研究では、それらのうちロボットに実装可能な身体動作に注目する。ここで、人間の身体動作の機能については、表1のような分類がなされている¹⁵⁾。これまでに、例示子と標識に分類される身体動作をロボットの応答音声に加えた結果、ユーザの理解度や親和性を向上する効果が得られている⁸⁾。本研究では、さらに「発話の番（以降ここでは発話権と呼ぶ）」の交替をスムーズにする調整子の利用を目指す。

調整子に分類される非言語情報として、これまでに、人間同士の対話の分析により、うなずきなどの頭の動作が発話権と密接に関係していることが確認されている¹⁶⁾。また対話において、聴者が発話権を獲得しようとする場合には相手と視線を合わせ、そうでない場合には視線を合わせない傾向があることが報告されてい

る¹⁷⁾。

うなずきや視線移動だけに限らず、調整子に分類される種々の非言語情報をロボットの対話インタフェースにおいて有効に利用するために、人間同士の対話における各非言語情報の個々の特徴を明らかにするとともに、それらの変化が人間とロボットのコミュニケーションに与える影響を調べる必要がある。なぜなら、人間同士の対話モデルを人間とロボットの対話に適用することが有効かどうか分からないうえ、非言語情報の制御の仕方によって、場合によってはかえって不快に受けとられたり、コミュニケーションの効率を落とすこともありうるからである。コミュニケーションへの影響を調べる際には、すべての身体動作をロボットに実装して比較するのは困難であるため、コンピュータグラフィクス(CG)によるシミュレーションロボットを用いる。さらに、非言語情報のうち、視線制御がコミュニケーションに与える影響について、CGロボットと実ロボットの比較を行う。

以下、2章では、人間同士の対話における非言語情報の出現タイミングの分析について述べる。3章では、CGロボットと人間の対話実験より非言語情報の種類・出力タイミング・発話内容など、それらの違いによるユーザへの影響やコミュニケーション場への影響についての分析について述べ、2章での結果と比較し考察を述べる。4章では、実際の対話の流れの中で視線制御が対話に与える影響について、CGロボットと実際のロボットの比較を行う。最後に5章において、これらの実験結果をふまえて人間型ロボットの対話インタフェースにおける非言語情報の出力制御方法を検討する。

2. 人間同士の対話の発話交替時における非言語情報の出現タイミング

本章では、人間の身体動作をともなった発話の様子を収録し、人間が言語情報にともなってどのように非言語情報を相手に提示しているかを分析する。

2.1 対面対話データ収録

前述した調整子の役割である円滑な発話交替の実現に関係する状況として、システムがユーザに発話を促す(発話権譲渡)状況を設定した。その際、問いかけ→応答、といった対話の一部分を抜き出した形の発話対¹⁸⁾単位を設定し、表2に示すような5種類10文を用意した。被験者(大学生10人)をシステム役とし、ユーザ役の実験者に発話を促すという設定で、被験者には、表2に例示するような発話文を憶えてもらい、自分で自然と思うような身体動作(たとえば、手

表1 Ekmanによる身体動作の分類
Table 1 Classification of body-actions by Ekman.

標識 (emblem)	音声語句に翻訳可能で表象、サインとも呼ばれる。Vサイン、お金、など。
例示子 (illustrator)	発話の内容や流れと結び付き、発話内容を強調、精緻化、補足する。思考動作、指示動作など。
情感表示 (affect display)	情動にともなう表情や身振り。握り拳など。
調整子 (regulator)	発話権の授受を制御したり対話の流れを円滑にする動作。
適応子 (adaptor)	状況に適応するための動作。頭をかく、貧乏ゆすり(癖)、など。

表2 相手に発話を促す発話対の例
Table 2 Examples of pair utterances.

発話対の種類	例
挨拶-挨拶	「どうも、こんにちは」-「こんにちは」
依頼-承諾/拒否	「えーと手を挙げて下さい」-「はい」/「いいえ」
呼びかけ-応答	「それでは、始めましょう」-「はい、始めましょう」
yes/no 質問-応答	「あなたは日本人ですか」-「いいえ」
属性質問-属性伝達	「あなたのお住いはどちらですか」-「***です」

表3 ユーザに発話を促す合い図
Table 3 Signs of turn-giving from system to user.

種類		定義			
非言語情報	人間にない	ビーブ音	聴覚	特定の音を短時間鳴らす	
	人間にある	図表示	視覚	文字の表示や口のイラスト	
		視線	視線	視覚	視線を対話相手に向ける
			まばたき	視覚	まぶたの開閉
パラ言語	うなずき	うなずき	視覚	頭を下に振る	
		手振り	視覚	手を差し出す	
	韻律	聴覚	イントネーション, ポーズ, 発話速度の強調		
言語情報	発話内容	聴覚	疑問文, 確認文, あいづち		

表4 各非言語情報の出現タイミングの分布

Table 4 Distribution of output timing of non-verbal information.

	視線	まばたき	うなずき	手振り
標本数	83	119	59	20
発話開始からの相対時刻による分布				
平均	-0.059	-0.244	0.683	0.415
標準偏差	0.704	0.874	0.742	0.532
キーワード開始からの相対時刻による分布				
平均	-0.556	-0.766	0.242	-0.064
標準偏差	0.725	0.938	0.739	0.464

「こんにちは」のようにキーワードとその他に分けられない発話に関しては除外。

の平を上に向け前に出す動作など)を加えてシステム役に発話をしてもらい、その様子をVTRに収録した。なお、ユーザ役の実験者は適当な返答をして被験者による発話権譲渡を成立させた。

それぞれの被験者にカメラを向けVTRに録画した後、音声・画像データをワークステーションに毎秒30フレームで取り込み、対話分析ツールを用いて、言語情報および非言語情報出現の時間関係を分析した。なお、ツールは独自に開発したものであり、時間軸上に音声波形を表示し、音声および画像データを参照しながら非言語情報が出現した時刻をマークして記録することを可能にしている。非言語情報としては、調整子として機能すると考えられる視線、まばたき、うなずき、手振りに着目した。これらの非言語情報の分類・定義を表3に示す。これらの非言語情報において、た

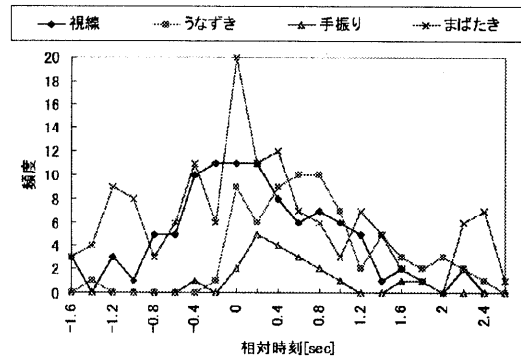


図1 発話開始時に対する非言語情報の出現頻度

Fig. 1 Frequency of non-verbal information against start of utterance.

たとえば視線移動をともなったうなずきのように、同時に起こる組合せがあるが、今回は簡単のため独立して扱うことにした。なお、手振り以外の身体動作については開始時刻と終了時刻がほぼ同時であり、その差は計測困難であった。したがって、ここでは非言語情報の出現タイミングとして開始時刻を扱うことにする。

2.2 非言語情報の出現タイミング

非言語情報の出現タイミングの傾向を調べるために、各非言語情報の発話開始、発話文中のキーワード発声開始からの相対的な出現時刻の分布を調べた。両者の分布の平均および標準偏差を表4に、発話開始からの相対的な出現時刻の分布を図1に示す。なお、キー

ワードとは、たとえば、「あなたは日本人ですか」という発話文における「日本人」のように、内容を伝えるうえで最も重要な単語を意味し、あらかじめ人手で与えるものとする。

図1において、各非言語情報の出現タイミングとして、発話開始直後のピークが特徴的であることが分かる。さらに詳細に見ると、視線・まばたきに関しては発話開始と同時に出現することが最も多いが、うなずき・手振りに関しては発話開始時よりも発話開始からわずかに遅れて出現することが多いことが分かる。ここで、表4に示すように、うなずきと手振りについては、発話開始からの相対時刻よりもキーワード開始からの相対時刻の方が平均および分散の値が小さく、これらの非言語情報は、キーワード開始付近で出現することが多いと考えられる。

本章では、人間同士の対話における非言語情報の出現タイミングの分析を行った。しかし、人間対ロボットの対話において、人間同士の対話に現れるモデルの適用がロボットの対話インタフェース構築に有効であるかどうかは不明である。次章では、実際に被験者にロボットと向きあって対話してもらい、その際に非言語情報の出力方法を変化することによってどのような出力方法なら自然でスムーズに対話ができるかを検証する。

その結果として適切な制御方法を検討するとともに、本章で示した人間同士の対話分析結果を考察する。

3. CG ロボットにおける非言語情報の制御とその効果

まず、対話実験を行うにあたり、前章と同様に、発話権譲渡の状況を設定する。発話権譲渡の合図として、表3に示した情報が考えられるが、ここではこのうち6種類の非言語情報について比較評価を行う。

3.1 発話交替における制約

ここで、システムに対してユーザが自由なタイミングで発話できるということは対話インタフェースとして重要なことである。この発話タイミングが制限されると、ユーザの思考の妨げやストレスなどを引き起こす¹⁹⁾。逆に、ユーザへの負担を減らすあまり、制限を弱くしすぎると、円滑な発話交替がかえって行われなくなる危険も考えられる。ここではユーザに発話を促す場合の、ユーザの発話タイミングの自由度を「発話交替における制約」の強さと呼ぶ。つまり、発話交替における制約が弱いインタフェースでは、ユーザの自由度が高くユーザへの負担も少なく、逆に制約が強いインタフェースでは、ユーザは発話のタイミングを限

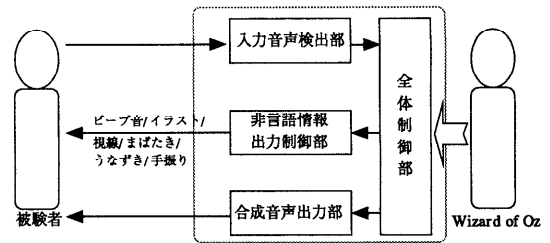


図2 実験システム構成
Fig. 2 System for experiment.

定されるということであり、状況に応じて制約の強弱を利用すべきと考える。

以下に述べる実験では、非言語情報制御の効果として対話の円滑さを評価するために、システムの発話に同期して非言語情報を出力した後のポーズ長（先行するシステム発話終了から後続のユーザ発話開始までの時間）を用いる。一概に、ポーズ長が短い方が良いとはいえないが、ポーズ長が長いということは、ユーザにとって発話すべきタイミングが分からず不必要に長い無音区間が生じた結果であると考えられ、コミュニケーションにおいて非効率であるといえる。実験において非言語情報の出力がある場合とない場合を混ぜた場合、制約が強い非言語情報ほどユーザはその出力に影響を受けやすく、そのため非言語情報の出力がない場合のポーズ長が長くなると予想できる。したがって、以下では、このポーズ長の長短によって上述した発話交替における制約の強弱については発話タイミングの自由度を判断することにする。

3.2 非言語情報の制御方法

次に、実験システムと実験手順について説明する。実際のロボットに様々な制御方法をすべて適用して比較評価を行うのは困難であるため、ここではCGによるシミュレーションロボット（DoraeMan）を用いて実験を行った。

実験システムは、3次元CGロボット画像表示と音声入出力を統合したマルチモーダルインタフェースを備えている。システム構成図を図2に示す。

実験では、CGロボットが出力する非言語情報の種類（図3）と、合成音声に対する出力タイミング（図4）の制御方法を自由に選択できるようにし、大学生20人の被験者が、非言語情報の種類6通りと出力タイミング5通りの組合せ計30通りを比較した。

対話収録にあたって、どのような種類の非言語情報が自然であり、どのタイミングで出力するのが適切であるかを1つずつ明確に調べるために、発話権譲渡の状況を設定し発話対として表2に示した5種類10文

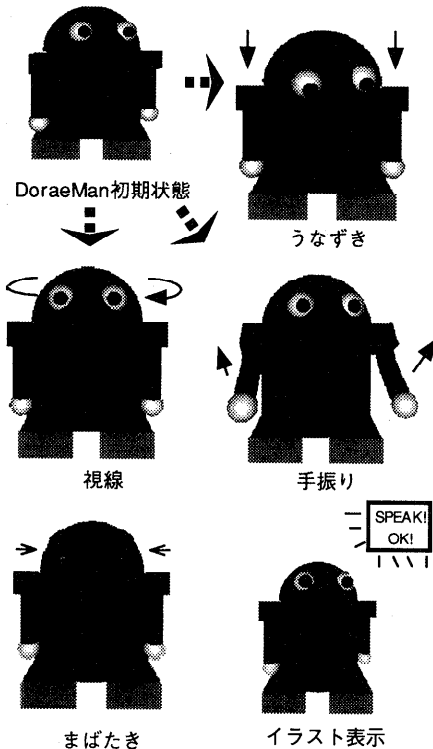


図3 表現する非言語情報の種類

Fig. 3 Types of non-verbal information expressed by CG robot.

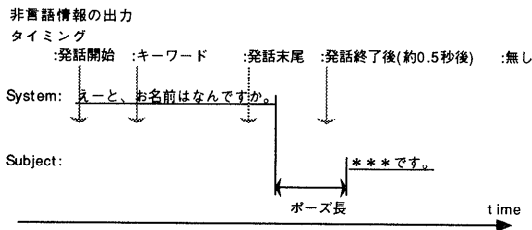


図4 発話と非言語情報の出力タイミング

Fig. 4 Output timing of utterance and non-verbal information.

についてそれぞれ均等に1回ずつ対話を収録した。

さらに、被験者にはアンケートにより発話交替の自然さや返答のしやすさなどを評価してもらった。1種類の非言語情報につき5パターンのタイミングをランダムで出力し(1種類につき5回行い、これを1セットとする)、1セット終了後、その種類についての評価をしてもらった。なお、セットの順序もランダムとした。そして最後に全体を通しての評価についてインタビューを行った。また、種類とタイミングを変えて出力をしていることだけは被験者に伝えているが、次のどのタイミングで出力されるかは知らせていない。ま

た、実験は Wizard of Oz 方式で行ったが、被験者にそのことは知らせていない。なお、ロボットの問いかけに対する被験者の返答内容は自由にした。

3.3 実験結果および考察

3.3.1 ポーズ長による制約の強さに関する比較

非言語情報の種類および出力タイミングによるポーズ長の平均値の変化を図5および表5に示す。非言語情報なしの場合、全体的にポーズ長が増加する傾向にあり、不自然な沈黙が生じやすくなっていた。これは、3.1節に示したとおり、被験者は同じような発話対のやりとりを繰り返すことによる慣れから、いつもシステムに何らかの動作があると思いそれを待ったためと考えられる。特に、ビープ音・イラストにおいて、出力有りの場合に対する出力なしのポーズ長の比率が大きい(最もポーズ長が小さいキーワード開始時の場合に対してそれぞれ197.7%、263.6%)ことから、これら人間にない非言語情報であるビープ音・イラストは、システムの動作に人間が合わせその合図に影響させられる度が高く、制約が強い非言語情報といえる。逆に、視線など人間にある非言語情報の場合は、その割合が小さいことから制約が弱い非言語情報と考えられる。表5に見られるように、まばたきとうなずきについては、出力なしの場合と出力ありのうちの最小の場合の分布において有意な差が見られ、人間にある非言語情報の中でもこれらは比較的制約が強いといえる。

なお、各非言語情報を出力した際の発話交替の自然性に対する5段階評価の平均をとった結果(図6)においても、ビープ音・イラストが他に比べ劣るという結果が得られ、それ以外は、まばたき、視線、手振り、うなずきという順に評価が高かった。

これらの結果から、対話において発話交替の円滑さを追求する場合、発話交替を強制的に行うには人間にない非言語情報を出力し、それ以外は状況に相応しい制約の強さを持つ非言語情報を出力するような制御方法が考えられる。また、発話交替の自然さを追求する場合には、人間にある非言語情報を出力するのが望ましいといえる。

3.3.2 各非言語情報の出力タイミング

図5において、人間にある非言語情報に限ると、おおむね発話開始時かキーワード開始時のどちらかにおいて最もポーズ長が短くなっている(視線のみ、発話末尾時の方が発話開始時より若干ポーズ長が短い)。視線・まばたきは、キーワード開始時よりも発話開始時の方がポーズ長が短く、一方、手振り・うなずきはその逆となっている。なお、2.2節で示したように、人間同士の対話においても、視線・まばたきは発話開始

表5 非言語情報の種類によるポーズ長の分布
Table 5 Distribution of pause length against types of non-verbal information.

非言語情報		ビープ音	イラスト	視線	まばたき	うなずき	手振り
出力なし	平均 μ_1	1.96	2.82	1.60	1.64	1.63	1.14
	標準偏差 σ_1	1.89	2.58	0.98	1.13	1.00	0.58
出力あり のうちの	タイミング	キーワード	キーワード	発話末尾	発話開始	キーワード	キーワード
	平均 μ_2	0.99	1.07	1.19	1.02	1.01	1.00
最小の場合	標準偏差 σ_2	0.53	0.72	0.73	0.70	0.58	0.66
	μ_1/μ_2 [%]	197.7	236.6	134.1	161.3	161.5	113.4
有意差		あり	あり	なし	あり	あり	なし

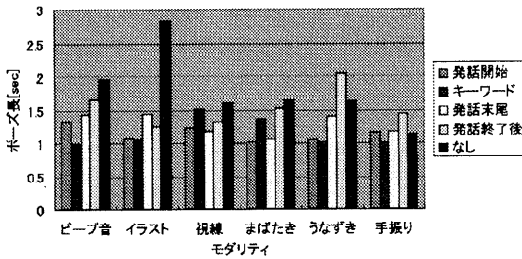


図5 非言語情報の種類と出力タイミングによるポーズ長の変化
Fig. 5 Length of pause and controlling non-verbal information.

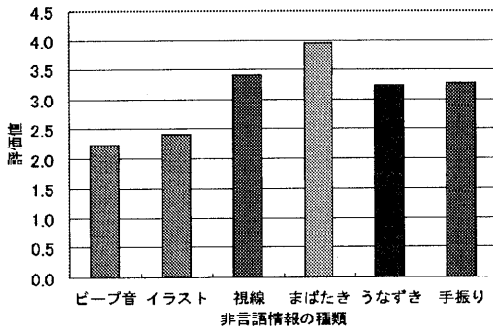


図6 発話交替の自然性に関するアンケート結果
Fig. 6 Result of questionnaire on naturalness of speaker-change.

時での出現が多いのに対して、手振り・うなずきは発話開始からやや遅れて（キーワード開始時あたりに）出現することが多い。このように、非言語情報のタイミングについて、人間同士の対話での傾向と、人間対ロボットの対話で効率の良いタイミングの傾向が類似している点から、人間の動作モデルをロボットに適用することがある程度有効であることが示唆される。

ここで、アンケート結果・インタビューから、各々の非言語情報の評価のまとめを表6に示す。視線など人間の動作にある非言語情報については発話開始からキーワードあたりでの早い段階で出力した場合の評価が良かった。また、システムの発話後の動作は発話の

表6 各非言語情報のアンケートおよびインタビューによる評価のまとめ

Table 6 Summary of evaluation on non-verbal information.

非言語情報	制約	評価
ビープ音	強い	聴覚 出力タイミングは発話後が良い システム発話と重なると聞きにくい（聴覚どうして組合せが悪い）
イラスト	弱い	視 出力タイミングは発話後が良い イラストに気をとられる どちらを見たらいいかわからない
視線など人間にある非言語情報		覚 出力タイミングは発話開始もしくはキーワード時が良い 自然、人間らしいなど

妨げになるという意見が寄せられた。

逆に、人間の動作にない非言語情報であるイラスト・ビープ音に関しては、ポーズ長の結果と異なるが発話後に出力する場合は良いという意見が多かった。理由として、日常的によく用いられているシステムがそのような発話権譲渡の方法をとっており（たとえば留守番電話の発信音など）、被験者がこれに慣れているためと考えられる。また、システムの発話中の生成に対しては、イラストとCGロボットのどちらを見たらいいかわからない、ビープ音と音声为重なると聞きにくい、などの意見があった。

以上の結果から、非言語情報の出力タイミングについて、人間にある非言語情報はいずれの場合も発話開始直後が適切であるのに対し、人間にない非言語情報は、対話において発話交替の円滑さを追求する場合はキーワード開始時、発話交替の自然さを追求する場合は発話後が適切であるといえる。

これまでの実験により非言語情報の種類による発話交替における制約の強さや自然性の違いが明確になった。さらに、人間の動作にある非言語情報の出力タイミングとして、発話開始あるいはキーワード開始時が自然かつ円滑な対話の実現に有効であることを確認した。この結果をふまえて、次章ではCGシミュレーションでの評価と実際のロボットでの評価の比較を行い、ロボットにおける非言語情報制御の有効性を検証

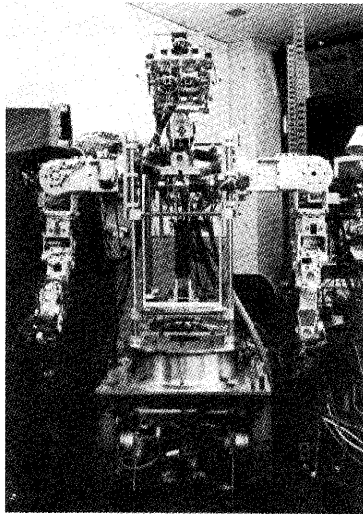


図7 Hadaly-2 概観
Fig. 7 Hadaly-2.

する。

4. 実際のロボットにおける非言語情報の制御とその効果

この章では、CG ロボットと実際のロボットにおける、実際の対話の流れの中でのインタフェース評価比較について述べる。ここでは、すべての非言語情報の実装が困難であったため、発話交替を円滑にする調整子のうち、視線の制御をロボットに適用した。

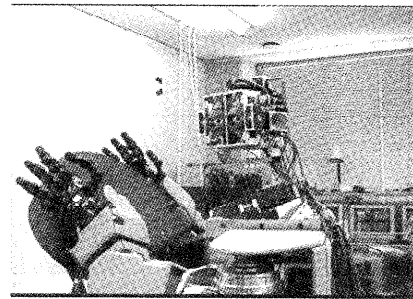
4.1 ヒューノイドロボット「Hadaly-2」

早稲田大学ヒューノイドプロジェクトでは、ヒューノイドロボット「Hadaly-2」(図7)を開発し、ジェスチャや視線移動などの非言語情報制御を可能にした¹⁾。

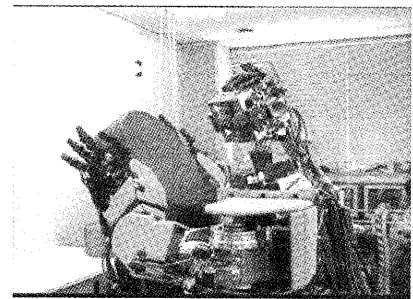
Hadaly-2は頭部、胴体(マニピュレータ、ハンド)といった人間の上半身の機能と、台車による移動機構を備えており、人間の音声指示によって、両手を使って積木の運搬作業をすることができる。人間とのインタラクションとして、握手をしたり積木の手渡しをはじめ、挨拶や方向の指示、ユーザへの問いかけなどのロボットの発話に、胴・両腕・頭部などを制御しジェスチャを加え、さらに視線制御として、両目のCCDカメラおよび頭部をユーザの方に向けることを可能にした(図8)。入力に音声・画像・触覚センサを、出力には合成音声・ジェスチャ・視線移動を用いたマルチモーダルシステムになっている。図9にシステム構成の概略図を示す。

4.2 評価実験

CG ロボットの場合の、ディスプレイ上の仮想空間



視線制御あり



視線制御なし

図8 Hadaly-2における視線の制御
Fig. 8 Controlling gaze of Hadaly-2.

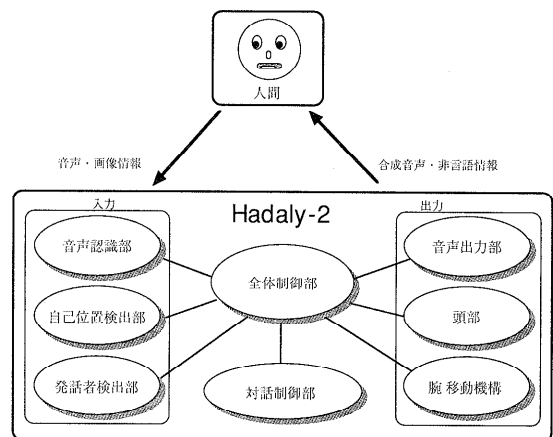


図9 Hadaly-2 構成図
Fig. 9 System composition of Hadaly-2.

における視線制御の効果と、実際に人間と空間を共有するロボットの場合におけるその効果は異なる可能性がある。ここでは、実際のロボットにおいて、発話開始時に人間に視線を向ける視線制御を適用した効果を調べ、CG ロボットの場合との比較を行う。

CG ロボットの実験では10人、実際のロボットの実験では6人の大学生を被験者とした。被験者には実験のタスクとして、対話によりロボットに積木運搬をさせ、その後インタフェースの評価を行うアンケート

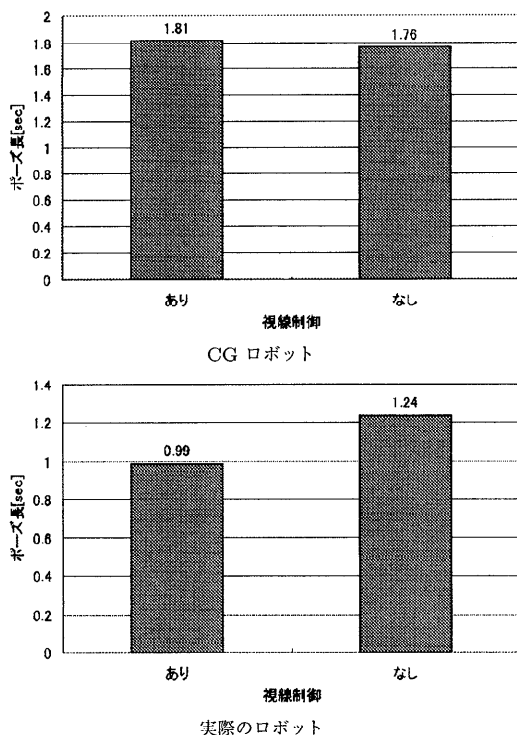


図 10 ロボットの視線制御によるポーズ長への影響

Fig. 10 Influence on length of pause by controlling gaze of robot.

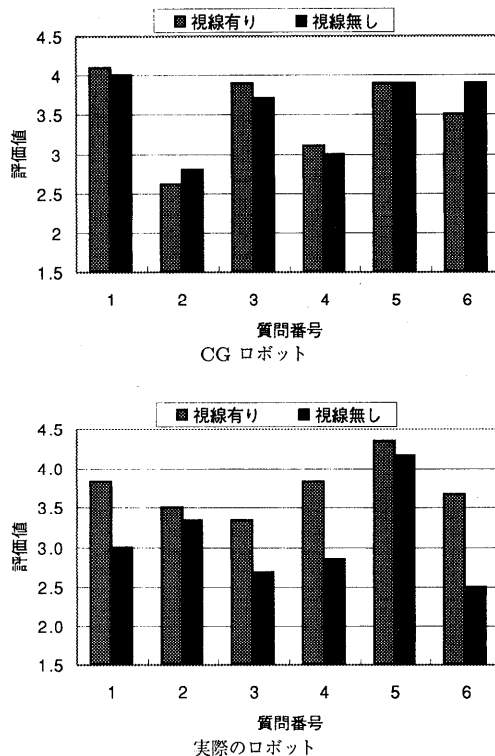
に答えてもらった。なお、ロボットがどのような動作が可能かをあらかじめ教示したが、音声認識可能な発話内容等は教示しなかった。

被験者は視線の制御を行った場合と行わなかった場合各1回ずつロボットとの対話をして、そのつどアンケートに5段階評価(5:良い~1:悪い)で答えてもらった。順序効果による影響をなくすため、それぞれの実験において被験者を2つのグループに分け、実験条件の順序を入れ換えて行った。

4.3 実験結果および考察

視線を制御した場合とそうでない場合の各々について、4章の実験と同様なロボットから被験者への問いかけといった発話権譲渡時におけるポーズ長の平均の比較を行った結果を図10に示す。この図に示すように、CGロボットにおいては視線制御の効果が認められなかったが、実際のロボットでは平均0.25sec(約24%)のポーズ長の短縮が認められた($p < 0.05$)。

図11に本比較実験におけるアンケート結果を示す。実際のロボットに対しては全項目で視線制御適用後の評価が良くなっており、目の前に実在するロボットに対して視線の利用による効果が大きいことが分かった。



1. 楽しく会話できた
2. ロボットの発話が自然だった
3. 動きが自然だった
4. 話しかけるタイミングが分かりやすかった
5. 自分の話しを聞いてくれた
6. 思いどおり動いてくれた

図 11 アンケート結果の比較

Fig. 11 Result of questionnaire.

5. おわりに

本論文では、ロボットと人間の自然な対話の実現を目指し、非言語情報の利用、特に発話交替のスムーズさに大きく影響を与えるとされる調整子という身体動作に注目した。ロボットの動作における非言語情報の制御に向けて、人間同士の対話分析から非言語情報の出現タイミングの傾向を調べたうえで、CGロボットにおける非言語情報の出力方法の違いによるユーザーへの影響についての分析実験、実際の人間型ロボットを用いたインタフェースの評価実験について述べた。

分析結果より非言語情報各々の特徴が明らかになり、ビープ音・イラストなど人間が用いない非言語情報は発話交替における制約が強く、視線など人間にある非言語情報は制約が弱いことがいえ、制約の弱い非言語情報のほうが好まれることなどが分かった。この結果から、自然性をそこなっても発話交替を強制的に行う

べき場合には、人間にない非言語情報を出力し、そうでない場合には状況に応じた制約の強さを持つ非言語情報を出力する制御方法が望ましいと考える。

また、非言語情報の出力タイミングについては、人間にない非言語情報は、対話において発話交替の円滑さを追求する場合はキーワード開始時、発話交替の自然さを追求する場合は発話後が適切であるのに対し、人間にある非言語情報はいずれの場合も発話開始直後が適切であるという結果が得られた。

さらに、人間型ロボットに視線制御の適用を行いインタフェース評価実験を行った結果、ロボットに対する親和性の向上に対して有効であることを実証した。また、CG シミュレーションと実際のロボットに対して、視線付加によるユーザへの影響の比較を行った結果、実空間を共有するロボットの方がより効果が大きかった。

今回行った実験では、非言語情報の制御について、主に発話との同期における出力タイミングについて分析・考察したが、今後、方法を十分に検討したうえで発話内容や意図との関係を分析し、これらを考慮に入れた制御方法の検討を行う必要がある。

謝辞 対話分析ツールを作成、提供してくれた早稲田大学理工学部白井研究室の大平茂輝君、その他実験に協力してくれた白井研究室の学生に感謝する。

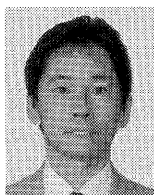
なお、本研究の成果の一部は早稲田大学理工学部総合研究センタープロジェクト研究「ヒューマノイド」および「マルチモーダルヒューマンインタフェース」の一環として行われたものである。

参 考 文 献

- 橋本周司, 成田誠之助, 白井克彦, 小林哲則, 高西淳夫, 菅野重樹, 笠原博徳: ヒューマノイド—人間型高度情報処理ロボット, 情報処理, Vol.38, No.11, pp.959-969 (1997).
- 加藤一郎: 人間ロボット論, 日本ロボット学会誌, Vol.10, No.1, pp.76-79 (1992).
- 白井克彦, 小林哲則, 岩田和彦, 深沢克夫: ロボットとの柔軟な対話を目的とした音声入出力システム WABOT-2における会話系, 日本ロボット学会誌, Vol.3, No.4, pp.104-113 (1985).
- 田中修一, 中里 収, 帆足啓一郎, 白井克彦: 復号作業下における音声インタフェースの設計と評価, 電子情報通信学会論文誌, Vol.J79-D-II, No.12, pp.2163-2169 (1996).
- 西本卓也, 志田修利, 小林哲則, 白井克彦: マルチモーダル入力環境下における音声の協調的利用—音声作図システム S-tgif の設計と評価, 電子情報通信学会論文誌, Vol.J79-D-II, No.12, pp.2176-2183 (1996).
- 白井克彦, 小林哲則, 工藤育男: 花開く音声処理技術, 情報処理, Vol.38, No.11, pp.971-975 (1997).
- 今井倫太, 開 一夫, 安西祐一郎: 注意機構を利用したヒューマンロボットインタフェース, 電子情報通信学会論文誌, Vol.J77-D-II, No.8, pp.1447-1456 (1994).
- 帆足啓一郎, 田中修一, 中里 収, 白井克彦: 構内案内ロボットにおける音声とジェスチャーの統合に関する検討, 日本音響学会春季講演論文集, 1-P-15, pp.191-192 (1996).
- 肥田木康明, 益満 健, 山岸則明, 中野裕一郎, 小林紀彦, 春山 智, 小林哲則, 高西淳夫: アイコンタクト機能を有する複数ユーザとの対話ロボット, 情報処理学会研究報告, SLP-17, Vol.97, No.66, pp.1-6 (1997).
- 黒川隆夫: ノンバーバルインタフェース, オーム社 (1994).
- Birdwhistell, R.L.: *Kinesics and Context*, Univ. of Pennsylvania Press (1970).
- Mehrabian, A. and Williams, N.: Non-verbal concomitants of perceived and intended persuasiveness, *J. of Personality and Social Psychol.*, 13, pp.37-58 (1969).
- 竹林洋一: 音声自由対話システム TOSBURG II—ユーザ中心のマルチモーダルインタフェースの実現に向けて, 電子情報通信学会論文誌, Vol.J77-D-II, No.8, pp.1417-1428 (1994).
- 長尾 確, 竹内彰一: コンピュータとの自然な対話のための新しいモダリティ—表情つき音声対話システムの試作と実験, 人工知能学会研究会資料, SIG-SLUD-9204-2, pp.9-16 (1992).
- Ekman, P. and Friesen, W.V.: The repertoire of nonverbal behavior semiotica, 1, pp.49-98 (1969).
- Iwano, Y., Sugita, Y., Kasahara, Y., Nakazato, S. and Shirai, K.: Difference in visual information between face to face and telephone dialogues, *Proc. ICASSP97*, Vol.2, pp.1499-1502 (1997).
- Novick, D., Hansen, B. and Ward, K.: Coordinating turn-taking with Gaze, *Proc. ICSLP96*, pp.1888-1891 (1996).
- 竹下 敦: プラン認識に影響を与える対話現象, 第4回人工知能学会全国大会論文集, pp.367-370 (1990).
- 菊池英明, 工藤育男, 小林哲則, 白井克彦: 音声対話インタフェースにおける発話権管理による割り込みへの対処, 電子情報通信学会論文誌, Vol.J77-D-II, No.8, pp.1502-1511 (1994).

(平成10年5月29日受付)

(平成10年12月7日採録)



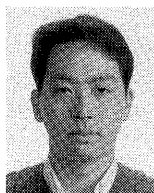
横山 真男 (正会員)

平成 8 年早稲田大学理工学部情報工学科卒業。平成 10 年同大大学院修士課程修了。早稲田大学ヒューマノイドプロジェクトに携わり、ヒューマンインタフェース、音声対話の研究に従事。平成 10 年 (株) 東芝入社。人工知能学会、日本音響学会各会員。



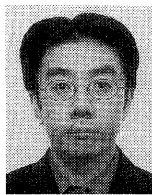
青山 一美

平成 10 年早稲田大学理工学部情報工学科。ヒューマンインタフェース、音声対話の研究に従事。人工知能学会、日本音響学会各会員。



菊池 英明 (学生会員)

平成 3 年早稲田大学理工学部情報工学科卒業。平成 5 年同大大学院修士課程修了。同年 (株) 日立製作所入社。平成 9 年同社退社。同年早稲田大学大学院博士課程入学。平成 10 同大学理工学総合研究センター助手。ヒューマンインタフェース、音声対話の研究に従事。情報処理学会平成 8 後期大会優秀賞受賞。人工知能学会、電子情報通信学会、日本音響学会各会員。



帆足啓一郎 (正会員)

平成 7 年早稲田大学理工学部情報工学科卒業。平成 9 年同大大学院理工学研究科情報科学専攻修了。同年、国際電信電話 (株) 入社。現在、(株) KDD 研究所にて情報検索等の研究に従事。



白井 克彦 (正会員)

昭和 38 年早稲田大学理工学部電気工学科卒業。昭和 43 年同大大学院博士課程了。工学博士。同年、同大学理工学部講師。昭和 45 年同助教授。昭和 50 年同教授 (電気工学科)。昭和 57~平成 2 年同大学情報システムセンター所長。平成 3 年同教授 (情報学科)。音声認識・合成技術、自然言語処理、信号処理向けアーキテクチャ設計、CAI 等を中心にヒューマンインタフェースの研究に従事。1998 人工知能学会会長。日本音響学会、IEEE 等会員。