

## 人間型対話ロボットにおける非言語情報の役割

帆足 啓一郎 横山 真男 荒井 大輔 安藤 義範 白井 克彦

早稲田大学理工学部  
〒169 東京都新宿区大久保3-4-1  
hoashi@shirai.info.waseda.ac.jp

あらまし

本研究では人間同士の対話における視線情報の役割の重要性に着目し、これを対話システムに適用して評価実験を行なった。音声対話システムの究極の形の一つに「人間型対話ロボット」、すなわち人間と同様の外見を持ち、人間と同様の対話を行なうことが出来るロボット、が考えられる。しかし、こうしたロボットを実現するためにはさまざまな課題が存在する。本研究ではまず、人間同士の対話における非言語情報、特に発話権移動時における視線情報の役割について検討し、人間型ロボットのシミュレーションシステムにこの視線情報を適用して評価実験を行なった。その結果、視線情報の適用により、対話の進行がより円滑なものになり、非言語情報の有効性を示すことが出来た。

## Nonverbal Information of a Humanoid Robot

Keiichiro HOASHI Masao YOKOYAMA Daisuke ARAI  
Yoshinori ANDO Katsuhiko SHIRAI

School of Science and Engineering, Waseda University  
3-4-1 Okubo Shinjuku-ku Tokyo 169  
hoashi@shirai.info.waseda.ac.jp

### Abstract

In this paper, we evaluated the nonverbal information of a dialogue system. The "Humanoid Robot", i.e. a robot with human-like appearance which can make natural conversation with humans, is one of the ultimate forms of future spoken dialogue systems. However, there are numerous problems to be solved in order to realize such a robot. In this research, we considered the role of nonverbal information, especially the role of gaze in turn-taking, in conversation between humans. We applied this information in a simulation system of the humanoid robot, and evaluated it, proving the effectiveness of gaze information.

## 1 はじめに

現在、世界中の研究期間で数多くの音声対話システムが開発・発表されている。また、簡単なものながら音声認識機能を搭載した製品も実用化されつつある。こうした状況の中、音声の入出力以外のモダリティをも使用することが出来る、いわゆるマルチモーダル対話システムの研究が盛んに行なわれている [1][2][3]。これらのシステムを開発する目的の一つとして人間同士の対話と同様な対話を人間-計算機間で実現させることがあげられる。「人間型対話ロボット」はこうした対話システムの究極の形の一つであるといえよう。

しかし、このようなロボットと人間同士のような対話を実現するには数多くの課題が山積しているのが現状である [4]。音声認識の性能向上、自然な合成音声の生成など、従来から音声対話システムを設計する上での未解決な問題点も数多い。しかしながら、ジェスチャーなどの非言語情報 (nonverbal information) の扱いもロボットと人間との対話を円滑なものにするために解決しなければならない重要な問題である。これまでのマルチモーダルシステムの研究ではあくまで対話システムのインタフェースとして複数のモダリティをどのように統合するか…に焦点が置かれることが多かった。しかし、人間型対話ロボットを実現するには人間同士の対話におけるマルチモーダル性を計算機上でどのように実現するかについて研究する必要がある。具体的な課題としてジェスチャー、表情などの非言語情報の役割の検討などがあげられる。しかしながら、これまでこうした問題を取り上げた研究例は少ない。そこで本研究ではこの非言語情報が対話者の意志疎通に果たす役割について検討する。また、実際の対話システムにこれらの非言語情報を適用し、このシステムとの対話実験を通じて対話インタフェースにおける非言語情報の有効性を証明する。

## 2 非言語情報の役割

人間が普段のコミュニケーションで用いている非言語情報の役割はさまざまであり、重要であることはいうまでもない。一般にいわれる「非言語情報」とは、ジェスチャーなどの身体動作のほか、姿勢、対人距離、服装・装飾等の外見、身体的特徴など、幅広い概念が含まれており、人間のメッセージ伝達

の60%~90%以上までがこの非言語情報によって占められているという説もある。これらの非言語情報全ての役割について検討・評価することは難しい。したがってここではロボットで実現が可能な身体動作という概念にのみ着目する。

人間の身体動作は一般に以下のように分類される [5]。

- 標識 (emblem)  
音声語句に翻訳可能で表象、サインとも呼ばれる。
- 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。
- 情感表示 (affect display)  
情動に伴う表情、身振りなど。
- 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。

これらの中で本研究では特に調整子 (regulator) の役割に着目する。調整子は、話す順番を決定したり、発話権のやりとりを制御したり、会話の流れを円滑にする機能を持つ動作である。相手話者にもっと大きな声で発話して欲しいときに耳に手をあてる動作、あるいは相手の発話を促す相づちなどは調整子の例である。例示子が発話内容と結び付いているのに対して、調整子の方は対話の流れと密接に関係する。調整子は標識や例示子ほどには意図的に発信されないが、意識には残ることが多い。円滑なコミュニケーションを維持するためには話者間の発話権の受渡しを円滑にする必要があるが、調整子はその役割を果たす重要な情報であるといえる。

最近の研究報告 [6] で人間同士の対話を収録したデータを分析した結果、人間は発話権を獲得するような発話をする時は相手の目を見て発話し、そうでない発話の場合は視線を相手から外して話をするということが報告された (図1)。

この視線情報が特に有効であると思われる状況にユーザ割込み状態と認識処理中状態が考えられる。

従来のシステム発話に対する割込み処理の手法の多くはシステム発話を一旦停止してユーザの発話の

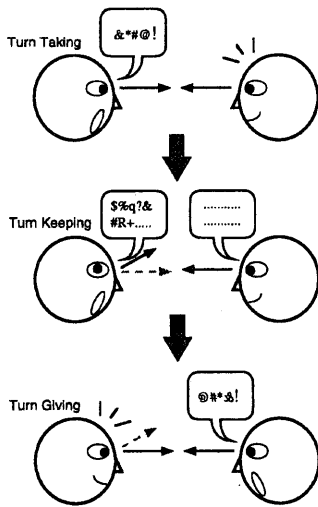


図 1: 発話権の移動と視線の関係

認識を行なうといった形式のものである [1][7]. しかし、この手法ではシステム発話が一時的に中断されることにより更に不自然さが増すものと思われる. 特にロボットの場合は割込み発話によりロボットの動作が一時停止するため、割込みのたびにロボットに対する不安がつのる危険性がある. こうした手法の代わりに視線によって発話権を譲渡することにより、対話の円滑化が図られるものと考えられる.

認識処理中も視線をユーザから外すことにより、システムが処理中であるという状態をユーザに提示することができると思われる. この手法により、認識処理中のユーザ発話を抑えることができることが期待される.

本研究の目的は代表的な調整子の一つであるこれらの視線情報を実際に対話システムに適用し、その情報の有効性を証明することである.

### 3 シミュレーションシステム DoraeMan

この節では実験に使用したシステム DoraeMan について解説する.

### 3.1 ヒューマノイドロボット Hadaly II

早稲田大学ヒューマノイドプロジェクトでは 1995 年に構内案内ロボット Hadaly を製作した [8]. Hadaly は理工学部構内の数箇所の研究室までの道順をユーザに音声とジェスチャーを用いて案内する人間型のロボットである.

しかし、Hadaly は見た目こそ人間型ロボットであったが、実際は既存の情報検索型の音声対話システムのインタフェースとして機械が使用されただけのものに過ぎなかった. そこでより高度なコミュニケーションを行なうことが出来るロボットを目指し、本年度から人間と協力して積木でオブジェを製作するロボット Hadaly II (図 2) の製作を開始した. 本研究の評価に用いたシステム DoraeMan はこの Hadaly II の動作を計算機上でシミュレートしたシステムである.

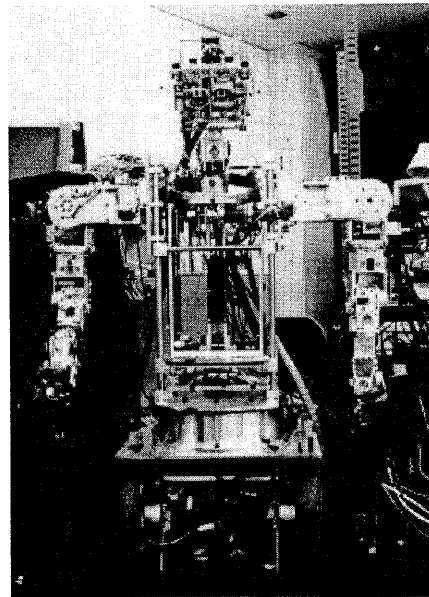


図 2: Hadaly II (開発中)

### 3.2 DoraeMan システム構成

DoraeMan のシステムはシステム制御部、対話制御部、音声出力部、音声認識部、割込み検知部、および CG シミュレーション生成部の 6 つのサブシステムにより構成されている (図 3).

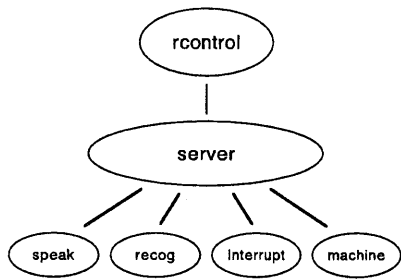


図 3: DoraeMan システム構成

- システム制御部 **server**  
システム全体の通信を制御する。
- 対話制御部 **rcontrol**  
対話の制御を行なう。各部への命令は原則的にここから発信される。
- 音声出力部 **speak**  
テキスト文から音声を合成し、出力する。音声合成には HP735/125 に接続された日本語テキスト変換ボード TSS-MW1 (日本電素工業) を使用。
- 音声認識部 **recog**
- 割り込み検知部 **interrupt**  
今回の実験では **recog**, **interrupt** は実験者によって操作された (Wizard of Oz 方式)。
- CG シミュレーション生成部 **machine**  
Hadaly II の動作を CG によりシミュレートしている。CG 描画・出力には Indigo2 R10000 Maximum Impact (Silicon Graphics) を使用。

## 4 対話システムへの視線情報の適用

### 4.1 評価実験

#### 目的

本実験の目的は人間同士の対話の発話権譲渡時における視線の情報を対話システムに適用することによって人間-システム間の対話が受ける影響を調べ、人間型対話システムの視線情報の有効性を示すことである。

## タスク

実験のタスクは与えられたオブジェを DoraeMan に音声で指示をすることにより作成することである。課題のオブジェは4つある積木のうち、3つを使用した簡単なものである。被験者にはシステムが Wizard of Oz 方式で動いていることは知らせていない。また、システムが誤認識した結果、DoraeMan が自分の意図と反した動作をする可能性を示唆し、そのようなことが生じた際には割込みの発話により DoraeMan の動きを止めることができることを教示した。

## 被験者

大学生 10 名。

## 方法

各被験者は DoraeMan と「視線情報あり」(条件 A)、および「視線情報なし」(条件 B) の対話を1つずつ、つまり被験者 1 人あたり 2 対話を行なった。順序効果を考慮し、被験者を表 1 のように 2 つのグループに分けた。

表 1: 実験条件

実験順	Group 1 (5 名)	Group 2 (5 名)
1 回目	視線情報なし	視線情報あり
2 回目	視線情報あり	視線情報なし

「視線情報あり」の対話では以下の視線情報を DoraeMan に付加した。

- 発話権譲渡時の視線情報  
DoraeMan が被験者に何かを尋ねる時など、被験者に発話権を譲る時に被験者の方を向かせる。
- 割込み時の視線情報  
DoraeMan が誤動作した際、被験者は任意の発話により、DoraeMan の動作に割込みをかけることができる。この割込みがかけられたときに、ユーザの方向を向かせる。
- 認識処理時の視線情報  
被験者の命令が DoraeMan に受理されるまで 3~5 秒の認識処理時間を設定した (WOZ 方

式で動作しているため、仮の処理時間)。この処理時間中、すなわち被験者の発話終了から Doraeman の次動作まで、積木を見渡すなど、視線を移動させる。

「視線情報なし」の対話では以上の視線情報は一切適用されてない(図4, 5参照)。

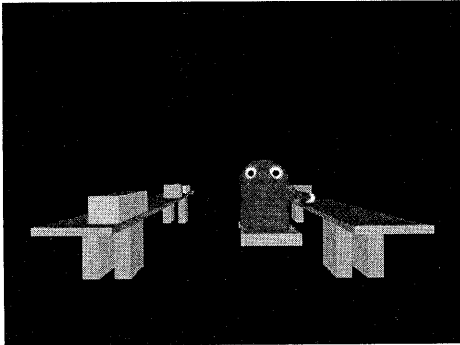


図 4: 視線情報ありの画面例

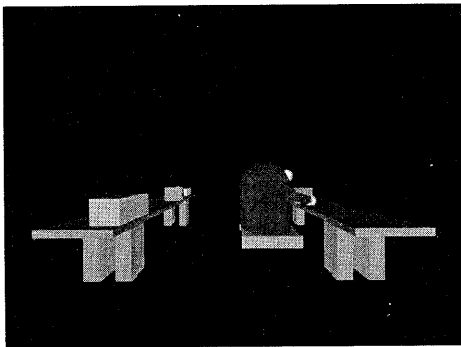


図 5: 視線情報なしの画面例

被験者の対話(途中中断された対話を含め、合計23個)をビデオに収録(システム実行画面と被験者の様子の2場面)し、分析を行なった。また、被験者には各対話終了後、アンケートに回答してもらった。アンケートは以下の項目について5段階評価をつけるものである。

1. 楽しく対話ができた。
2. ロボットの発話は自然だった。
3. ロボットの動きは自然だった。
4. ロボットに話しかけるタイミングはわかりやすかった。
5. ロボットは自分の話を聞いていた。
6. ロボットは思い通り動作してくれた。

また、2回目の実験終了後のアンケートには1回目の実験と比較してシステムの良かった点、および改良すべき点を記述させた。

## 4.2 実験結果

### 4.2.1 対話データからの結果

収録した対話は被験者10人×2対話に途中中断された3対話を含めた23対話。その他の対話データに関する情報は表2に記す。

表 2: 収録対話データ概要

対話数	23 (中断対話3含む)
被験者発話数	361
割込み発話数	37
割込み発生数	22
認識処理中発話数	43

割込み発話とは Doraeman の動作に被験者が割込みをかけようとして行なわれた発話であり、その際 Doraeman が割込み動作を行なった状態を割込み発生とする。また、認識処理中発話とは Doraeman の認識処理時間の間に発声された発話とする。

発話権の移動がスムーズに行なわれている基準として、Doraeman の発話および動作終了から被験者の次発話までのポーズを採用し、認識処理中発話を除くすべての発話についてこのポーズを測定した。

まず、Doraeman が被験者に発話権を譲渡した後のポーズ(すなわち、割込み後のポーズを除いたもの)の条件別の平均時間[sec]を測定した。その結果を図6に示す。この図から明らかなように、1回

目の実験(1.90)と2回目の実験(1.66), および条件A(1.80)と条件B(1.76)を比較しても有意な差を得ることはできなかった。これは被験者がDoraemanとの対話を進めるうち、システムの発話文終了のタイミングを把握してきたこと、また、こうした発話権譲渡のタイミングが非言語情報なしでもスムーズに行なわれやすいこと、などが原因として考えられる。

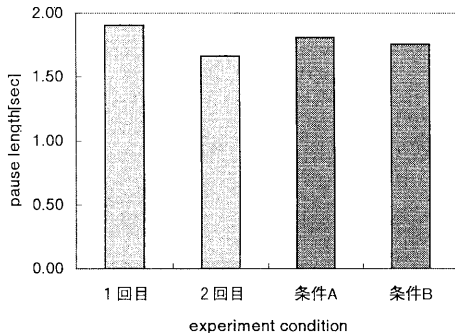


図 6: 発話権譲渡後のポーズ長

次にDoraemanの割込み動作終了後のポーズについても同様に測定した。その結果を図7に示す。1回目の実験(8.50)と2回目の実験(8.87)の間には有意差はなかったが、条件A(4.44)と条件B(11.76)の間には有意差が生じた( $p < 0.05$ )。このことから視線情報を使用した発話権の譲渡の有効性が示された。

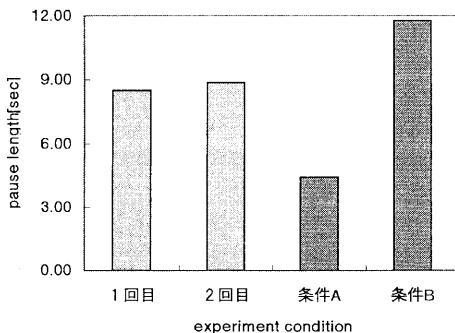


図 7: 割込み発生後のポーズ長

また、認識処理中に被験者が発話した回数を測定

した。その結果を図8に示す<sup>1</sup>。条件Aでは7回しか生じなかった認識処理中発話が条件Bでは14回生じている。このことから認識処理中の視線情報の有効性が示された。

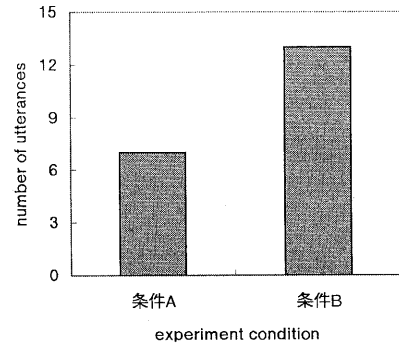


図 8: 認識処理中発話数

#### 4.2.2 アンケート結果

実験後の被験者に対するシステムの5段階評価アンケートからは特にこれといった結果を得ることができなかった(図9, 10参照)。これは前述したように、本来調整子という非言語情報が意図的に発せられることの少ない情報であるため、今回適用した視線情報が被験者の記憶に残りにくい情報であることが原因と思われる。このことは1回目の実験と比較の記述回答からも明らかである。このときの回答は認識処理時間や認識率に関するものがほとんどで、視線情報に着目した回答はなかった。

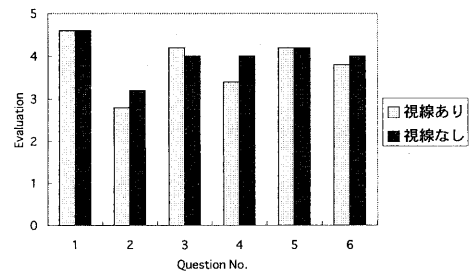


図 9: アンケート結果 (Group 1)

<sup>1</sup>1対話中、22回の認識処理中発話が生じた特異な対話を除く対話データに基づく評価のため、表2の数値と異なる。

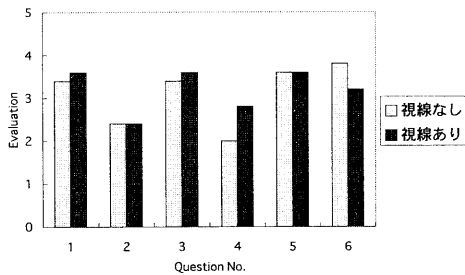


図 10: アンケート結果 (Group 2)

### 4.3 考察

アンケート結果より、システムの視線情報は被験者の記憶に残りにくい情報であることが明らかになった。これは調整子という情報の性質上、ある程度予想された結果であるといえる。

しかし、対話データから導出した結果からは視線情報が全く無意味なものではないことが示された。システムからの質問による発話権譲渡では視線情報は発話権移動を円滑にさせているという結果は導かれなかったが、被験者の割込み発話による発話権譲渡や認識処理中の発話権維持という、音声情報が使用されていない状況下での視線情報の有効性を示すことが出来た。

## 5 まとめ

本研究では人間のコミュニケーションにおける非言語情報、特に視線情報の役割について検討した。特に人間同士の対話で発話権の授受を制御したり対話の流れを円滑にする身体動作である調整子の一つ、視線情報を人間型ロボットに適用し、その有効性をロボットをシミュレートしたシステムとの対話実験により評価した。実験で収録した対話を分析した結果、ロボットにおいても視線情報は人間のそれと同様の効果があり、ロボットにおいての視線情報の有効性を示すことが出来た。

## 6 今後の課題

今回の研究ではロボットのシミュレーションシステムによる評価実験に留まったが、本来このような人間的な情報伝達手段の評価は実際に動くロボット

のような、人間らしいシステムに搭載し、評価を行なわねばならないだろう。冒頭で述べたようなロボットを実現するためにはこのような評価実験を繰り返す必要がある。また、こうしたロボットを組み上げるためには音声処理をはじめとして、機械系、視覚系、制御系など、異なる分野の研究者が各々の枠組を越え、協力できる環境を作り上げることが不可欠である。今後のヒューマノイドロボット開発に期待したい。

## 参考文献

- [1] 竹林, 永田, 瀬戸, 新地, 橋本: 音声自由対話システム TOSBURG II-マルチモーダル応答と音声応答キャンセルの応用-, 情報処理学会研究会資料, HI-45-13, 1992.
- [2] 永田, 竹内: 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.
- [3] 安藤, 畑岡: マルチモーダルなエージェント型ユーザインタフェースの評価と対話制御の検討, 情報処理学会研究会資料, 95-SLP-7, pp 91-96, 1995.
- [4] 山本: ロボットとの対話, 人工知能学会研究会資料, SIG-J-9302-3, pp 17-24, 1993.
- [5] 黒川: “ノンバーバルインタフェース”, オーム社, 1994.
- [6] Novick, D., Hansen, B., Ward, K.: “Coordinating turn-taking with Gaze”, Proceedings of ICSLP96, pp 1888-1891, 1996.
- [7] 菊池, 工藤, 小林, 白井: “音声対話インタフェースにおける発話権管理による割込みへの対処”, 信学論, J77-D-II, No.8, pp 1502-1511, 1994.
- [8] 帆足, 田中, 中里, 白井: “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”, 日本音響学会平成8年度春季講演論文集, 1-P-15, pp 191-192, 1996.