音声言語情報処理 29－21
(1999. 12. 20)

[特別招待論文]

ヒューマノイドロボットにおける マルチモーダル会話インタフェース

小林 哲則　白井 克彦
早稲田大学

169-8555 東京都新宿区大久保 3-4-1
koba@tk.elec.waseda.ac.jp

あらまし　ヒューマノイドロボットを用いた音声対話研究について述べる。早稲田大学においては，３０年に亙り，人間型ロボット（ヒューマノイドロボット）における対話の研究を行ってきた。WABOT，WABOT-2に代表される第 1 世代の対話ロボットは，ロボットに対する指令を音声によって伝えることを主な目的としたが，Hadaly，Hadaly2 などの第 2 世代においては，円滑な対話を進めるための，身体表現と言語表現の協調にまで踏み込んで検討が行われた。現在のロボット ROBITA は，第 3 世代にあたり，主に身体表現を活用しながら，グループ会話を実現する手法について検討が進められている。本稿では，これらのロボットの概要について説明しながら，ヒューマノイドロボットを用いた音声対話研究の動向について紹介する。

キーワード　音声対話，会話ロボット，ヒューマノイドロボット，グループ会話

Multi-Modal Conversational Interface for Humanoid Robot

Tetsunori Kobayashi, Katsuhiko Shirai
Waseda University

3-4-1 Okubo, Shinjuku-ku, Tokyo 169-8555
koba@tk.elec.waseda.ac.jp

Abstract　This paper describes a research on spoken conversation using Humanoid robots. We have been studying conversational robots which have human-like bodies (Humanoid robots) for these 30 years. In the first generation (WABOT, WABOT-2), we concentrated to realize spoken commands to the robots. Then, in the second generation (Hadaly, Hadaly2), we studied cooperative use of the language and the body expressions for the smooth communication. Now, in the third generation (ROBITA), we are trying to realize group conversation utilizing body expression. In this paper, we introduce the outlines of these robots, and show the trend of spoken conversation research using Humanoid robots.

key words:　Spoken dialogue, Conversation robot, Humanoid robot, Group conversation

1．はじめに

人間同士の対話において，言語情報のみならず非言語情報の伝達が，円滑なコミュニケーションを実現するために大きな役割を果たしていることが知られている。例えば，身振り・手振りや非言語音などは，言語と補完しあって，標識，例示，情感，調整，適応などに関わる重要な情報を対話相手に伝えている（表１参照）[1]。

表１．会話における非言語情報の役割（Ekmanによる身体動作の分類）

役割	例
標識	ある事物を象徴的に表す身振り，ポーズ，声。（Vサイン，ブーイング，等）
例示子	発話と結びついて，それを補足する身振り。（指示動作，例示動作，等）
情感表示	情動に伴う表情に関わる，身振り，声。（握り拳，笑み，唸り声，等）
調整子	発話権の授受を制御したり，対話の流れを円滑にするみ身振り，声。（うなずき，あいずち，等）
適応子	状況に適応するための動作

ここで，音声による言語情報の伝達が意識的に行われるのに対して，身振り等による非言語情報の多くは意識下で伝えられることに注意を要する。このため，普段我々は，それら非言語情報の役割に気づくことは少ない。しかし，厳然として意識下において，それらの情報は処理され，対話の自然性に役立っている。

さて，これら重要な情報処理が意識下で行われていることを考慮すると，人間－機械の音声対話システムを実現する場合，機械側も人間と同じような手段を用いて同様な情報を送ることが必要とされる。異なる手段では，人間の意識下での情報処理機構に訴えることは期待薄だからである。この意味では，対話用機械は，人間と同じ形を持つことが重要な意味を持つことになる。

このような観点から，我々は人間型ロボット（ヒューマノイドロボット）を用いた対話研究を進めている[2]。本論文では，早稲田大学においてこれまで開発してきた対話ロボットを紹介するとともに，現在の研究課題について述べ，ヒューマノイドロボットを用いたマルチモーダル対話研究の動向を紹介する。

2．WABOTとWABOT2：第１世代の対話ロボット

早稲田大学における対話ロボット研究の歴史は古く，１９７０年初頭まで遡ることができる。当時の対話ロボット研究の主な目的は，音声指令によって，ロボットを動かすことにあった。

１９７３年に完成したWABOT（図１）では，移動に関する指令を音声によって行うことができた[3]。マルコフモデルを用いて記述した言語モデルと単語音響モデルとを用い，ベイズ的アプローチによって文章を認識するものであり，今日の音声認識システムの原型を見ることができる。当時は音声認識・合成システムを実現したこと自体が画期的なことであった。この意味で，WABOTは対話ロボットとして

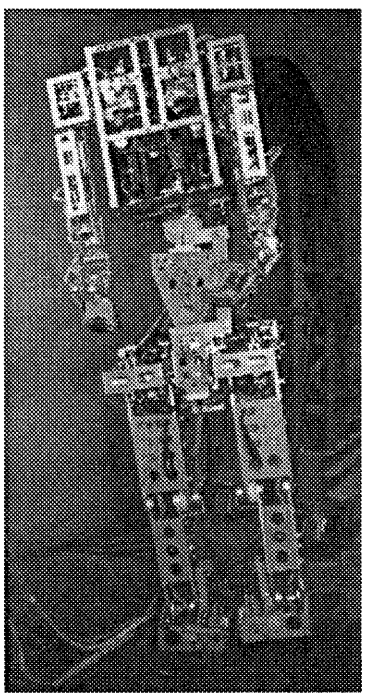

図１．WABOT（1973）

のみならず，音声認識システムとしても記念碑的存在といえる。

これにつづく８０年代は，対話の流れに応じてロボットの振る舞いを決めるための枠組みについて研究が行われた。

１９８４年に開発した WABOT-2（図２）では，鍵盤楽器演奏ロボットに対する演目の依頼を音声によって行ったが，そこでは受理対象文の意味付けを対話の流れに応じて動的に行う枠組みが検討された[4][5]。

これら初期の対話ロボット研究に共通していえることは，ロボットに音声対話システムが組み込まれてはいるものの，対象とした技術課題は，あくまで音声あるいは対話であって，残念ながらロボット対話ではなかったということである。すなわち，対話の主体がロボットであることの意義（目を持ち身体を持つことの価値）についてはまだ目が向けられていなかった。身体はあくまで作業のための身体であって，対話には貢献していなかった。音声システムがロボットの傍らで動作しているという感覚に近く，ロボットと対話しているという雰囲気は，十分には実現できていなかった。

WABOT，WABOT２に代表される，音声指令可能なロボットを実現することを主な目的としたものを第１世代の対話ロボットとして分類することにする。

３．Hadaly と Hadaly2：第２世代の対話ロボット

WABOT-2 の反省から，９０年に入って，ロボットが持つ身体的機能を対話に役立てるための研究が行われた。

１９９５年の Hadaly[6]，９７年の Hadaly２（図３）[7][8][9]では，主に，視線制御や，身振り，まばたきなどの身体表現が持つ，対話の調整的役割に焦点を当て研究が行われた。WABOT-2 が何を話しかけても譜面台を睨みつけて微動だにしなかったのに対し，Hadaly2 は，ユーザの目を見，瞬きをしながらユーザと会話する。こうすることで，ユーザは

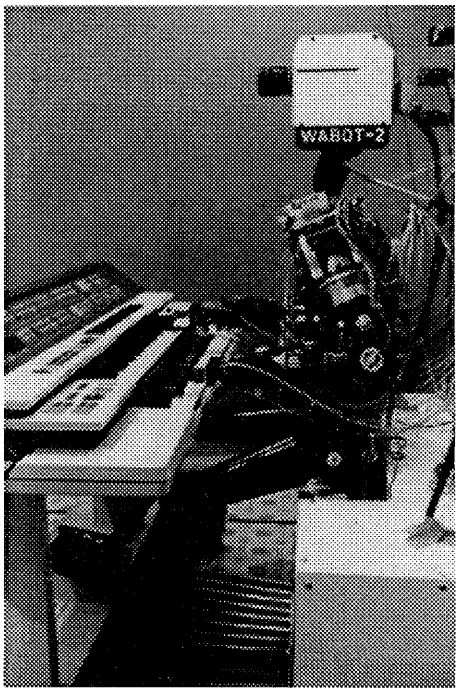

図２．WABOT-2（1984）

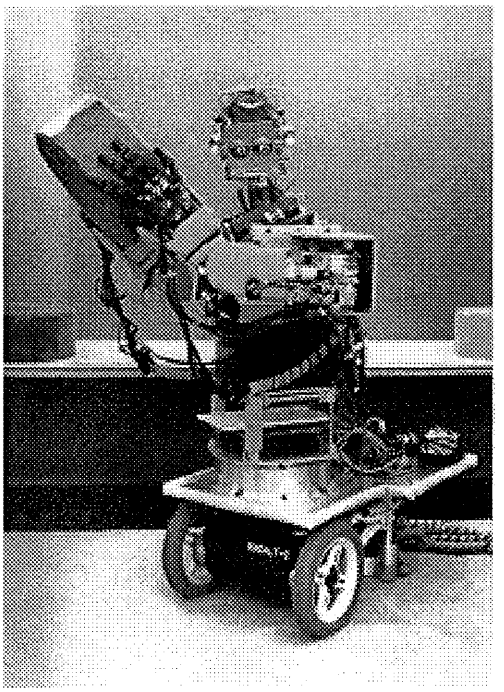
図３．Hadaly２（1997）

発話のタイミングを取りやすくなり，自然な対話が実現されることが実験的に示された[9]。

実際対話をしてみるとわかるのであるが，身体動作を駆使したことによって，Hadaly2 は，第1世代の対話ロボットとは明らかに異なった，ヒューマンフレンドリーな印象を醸し出すのに成功している。この意味で，Hadaly 2 を第 2 世代の対話ロボットと分類することにする。

4．ROBITA：第3世代の対話ロボット

Hadaly 2 の成果を受けて，現在開発中のロボットが ROBITA(Real-world Oriented BI-modal Talking Agent)である（図4）。

ROBITA において，これまでのロボットと最も大きく異なる点は，グループ会話を行うことを課題としたことである。グループ会話とは，共通の話題に対して 3 人以上の参加者が会話を行う状況をいう。機械と人間の1対1の情報交換では，人間が機械に合わせることで音声対話以外にも効果的な手段を考えることができるのに対し，グループ会話に機械が参加する場合，人間と人間とが対話している場面を想定する必要があるため，機械が音声対話機能を持つ必要度は一段と高くなる。グループ会話は，音声会話研究における新しくて重要な研究課題であり，これを扱う ROBITA を第 3 世代の対話ロボットと分類することにする。

グループ会話では，これまでの 1 対 1 の対話では考慮する必要の無かった様々な問題が生じる。例えば，誰が誰に向かって話しているのか，次には誰が話すと予想されるのか，ロボット自身はいつ話すことが求められているのか，などに関するその場の状況を理解しなければならない。また，ロボット自身も意思表示をしながら，積極的にこの状況に関与していかなければならない。

ROBITA は，顔向きの認識機能，音源の認識機能などによって，複雑な発話交代に関する場の状態を察知する機能を実現するとともに，顔向きと視線の制御によって，ROBITA 自身の発話権に関する意思表現機能を実現している[10]。

例えば，対話参加者の多くが見つめている人は，その場において会話の中心的な話題提供者であるから，ROBITA はまずその人に注目する。その人が発話をはじめた場合，その人が話しながら見つめている人が，次に発話する可能性が高い（次発話の候補

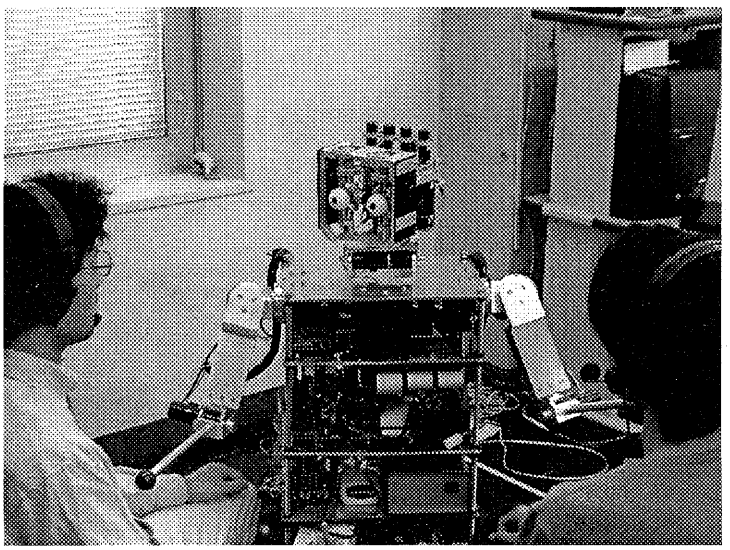

図4．ROBITA（1999）

者となる)。よって発話が終わった段階で，ROBITA は次発話の候補者へ視線を向ける。また，現発話者がロボットを見つめているならば，次には ROBITA 自身が発話を求められていると解釈する。

また，ROBITA は顔を向けるという行為によって，

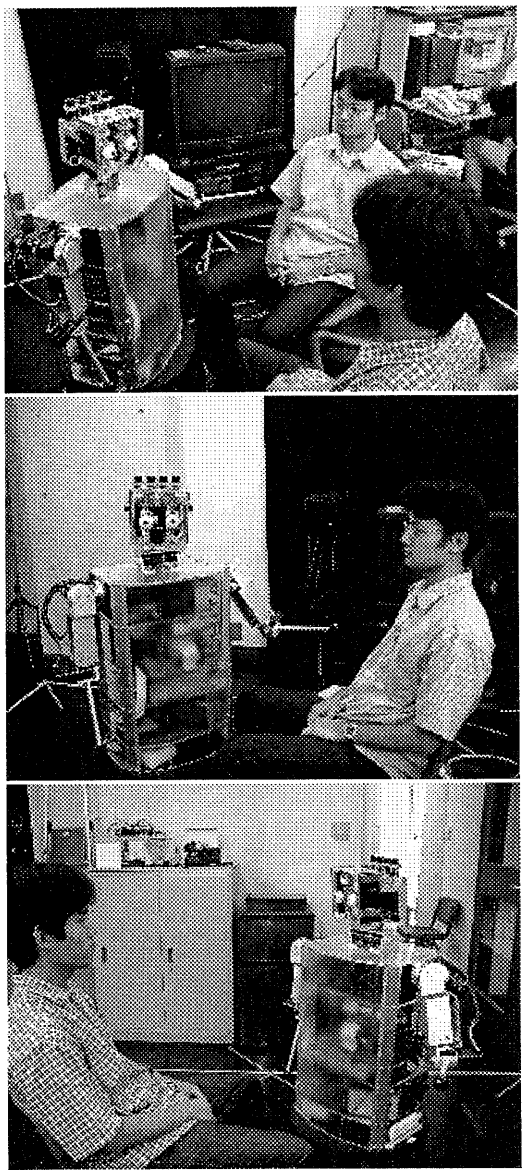

図５．伝達情報の視点依存性。上図：3 者対話図。右からユーザ A，ユーザ B，ROBITA。中図：ユーザ A の視線。下図：ユーザ B の視線。

「あなたに注目している」ことを表現する。また目を向ける行為によって，「あなたに話している」ことを表現する。さらに，これらを組合せて，ユーザ A に顔を向けた上で，視線を B に向け，「ちょっと待ってください」ということによって，A と話すために B に待つように言ったことを伝える。このとき，３次元空間に ROBITA が実在することによって，対話相手の位置に依存した情報を伝えることができる（伝達情報の視点依存性：例えば図５においては，ユーザ A に対しては，「彼（B）を指しながら，あなた（A）に向かって話しかけている」ことを伝え，ユーザ B に対しては「あなた（B）を指しながら，彼（A）に向かって話しかけている」ことを伝えている)。2 次元の擬人化エージェントでは，ユーザ個々を意識した空間的表現はできない（例えば，カメラ正面を見て撮った映像は，どこから見ても自分を見つめているように感じる（モナリザ効果))。ROBITA では，このような実空間に存在することの利点を生かしながら，視線・顔向きを制御し，また認識しながら，対話場の状況に関する自然な制御と理解を可能にすることを検討している。

ROBITA における研究テーマは，グループ会話を扱うことだけではない。ROBITA は，第 2 世代の対話ロボットと同様に，身体動作によって対話における例示，調整機能を実現しているが，それらの機能についても Hadaly 2 と比べ幾つかの改良がなされている。例えば，指示動作においては，単に指差しを行うのではなく，ロボットの可動性を利用して，ユーザに対しより分かりやすい位置に立って対象物を指示する機能を実現している[12]。また，Hadaly にはなかった，ユーザの発話に合わせたうなずきも実現している[13]。うなずきは，相槌と同じく，バックチャネルフィードバックとして，「情報は伝わっていますよ」ということを対話相手に伝えるのであるが，相槌が音でフィードバックを与えるために，タイミングを誤ると対話のリズムをかえって乱すのに対し，動作でのフィードバックは，少々タイミングがずれても悪影響を及ぼさないという利点がある。また，簡単な言語の獲得機能も実現している[14]。

5．むすび

早稲田大学において進められている，ヒューマノイドロボットを用いたマルチモーダル対話研究について述べた。現在の ROBITA は身振りや視線の切り替えうなずき等を交えながら，自然なグループ会話を実現するところまで成長しているが，簡略化した顔しか持たないため，表情に乏しいという欠点を持っている。優れたインタフェースは，内部状態の適切なフィードバックを必要とし[15]，表情はこのための有用な道具となりうる。当大学では，ロボットの顔表情に関する研究も進められている[16]ので，今後，この成果を組み込むことにより，より自然な次世代の対話ロボットを構成したい。

さて，ここで述べたロボット対話の研究は，ロボットをヒューマンフレンドリーにするために音声対話を用いる，といった音声応用的な視点で見られがちである。しかし我々は，むしろ対話の本質を明らかにするためにヒューマノイドロボットを道具として用いるという基礎研究的な立場がより重要と考えている。制御困難な人間と人間との対話の観察によって，対話の本質に関する知見を得る代わりに，一方を制御可能な機械に置くことによって，対話の自然性あるいは効率に影響を与える要因を，より系統的に探ろうとする立場である。現状のロボットを，この目的に使うにはまだ若干の無理があるが，近い将来より高いレベルのヒューマノイドロボットを実現し，これを道具として音声対話の本質に迫ってみたい。

文献

[1] P.Ekman, W.V.Friesen, "The repertoire of nonverbal behavior", Semiotica, Vol.1, pp.49-98 (1969).

[2] 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.956-969(1997).

[3] 白井克彦，藤澤浩道，"ミニコンを用いた音声入出力システム"，電気学会論文誌，Vol.94-C，No.7，pp.149-155(1974).

[4] 白井克彦，小林哲則，岩田和則，深沢克夫，"ロボット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，Vo.3，No.4，pp.104-113（1985）.

[5] T.Kobayashi, K.Shirai, "A network model dealing with focus of conversation for speech understanding system", IEEE Proc. ICASSP86, pp.1589-1592 (1986).

[6] S. Hashimoto, S, Narita, H. Kasahara, A. Takanishi, S. Sugano, K. Shirai, T. Kobayashi, H.Takanobu, T.Kurata,K.Fujiwara, T.Matsuno, T.Kawasaki and K.Hoashi, "Humanoid Robot-Development of an Information Assistant Robot Hadaly-", 6th IEEE International Workshop on Robot and Communication (1997).

[7] 橋本周司,成田誠之助,白井克彦,高西淳夫,笠原博徳,小林哲則,菅野重樹,"人間共存ヒューマノイドロボット：Hadaly-2," 第 15 回日本ロボット学会学術講演会, 3C33, pp.761-762 (1997).

[8] H.Kikuchi, M.Yokoyama, K.Hoashi, Y.Hidaki, T.Kobayashi, K.Shirai, "Controlling Gaze of Humanoid in Communication with Human,", Proc. IROS, pp.255-260, Oct. 1998.

[9] 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"，情報処理学会論文誌，Vol.40，No.2，pp.487-496（1999）.

[10] 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）.

[11] Y.Matsusaka, T.Tojo, S.Kubota, K.Furukawa, D.Tamiya, S.Fujie, T.Koabyashi, "Multi-person Conversation via Multi-modal Interface —A Robot who Communicates with Multi-users", Proc. Eurospeech 99, pp.1723-1726, Sep. 1999.

[12] 松坂要佐,東條剛史,古川賢司,藤江真也,小林哲則，"ジェスチャの表現・理解機能を有するロボットによる空間情報共有型対話の実現"，日本音響学会秋季研究発表会講演論文集, pp.111-112, (1999).

[13] 東條剛史，松坂要佐，小林哲則，"身体動作による対話調整機能をもつ対話ロボット"，情報処理学会研究報告，SLP-30，（2000.2 発表予定）.

[14] 藤江真也,小林哲則,"自律型ロボットの行動を介した言語獲得",人工知能学会第 13 回全国大会，pp.223-224(1999).

[15] 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，Vol.J79-DII，No.12，pp.2176-2183（1996）.

[16] A.Takanishi, H.Takanobu, I.Kato, T.Umetsu , "Development of the Anthropomorphic Head-Eye Robot WE-3RII with an Autonomous Facial Expression Mechanism", Proc. 1999 IEEE International Conference on Robotics and Automation, pp. 3255-3260 (1999).