

コミュニケーションエッセンスを有する自律ロボットの 実現に向けたWOZ環境の構築

Constructing the WOZ Environment for Realizing an Autonomous Robot
which Has Essence of Communication

大橋 洋輝 高橋 暁弘 岡田 将吾 大本 義正 西田 豊明
Hiroyuki Ohashi Akihiro Takahashi Shogo Okada Yoshimasa Ohmoto Toyoaki Nishida

京都大学 情報学研究科 知能情報学専攻

Dept. of Intelligence Science and Technology, Graduate School of Informatics, Kyoto University

The goal of this study is to realize an autonomous robot which has essence of communication with human beings. In order to achieve this goal, the robot has to learn the rules of communication. This paper presents the novel way of acquiring the natural interaction data from human-robot communication. The key idea is based on the method of Wizard of Oz(WOZ). We show why such a way is needed, what is important when constructing the environment for acquiring data, how we can realize such environment. We also introduce a result of preliminary experiment which we conducted using the prototype system of the proposed WOZ environment. As a consequence, we concluded that the proposed WOZ environment is very useful to get data of human-robot communication.

1. はじめに

近年,家事ロボットや介護ロボット,案内ロボット,エンターテインメントロボットなど,様々な方面で自律ロボットへの需要が高まってきている.これら特定の分野で活躍するロボットには,その分野特有の役割をこなす能力が求められるのと同時に,人間社会の中で活躍するために,共通して人間とうまくコミュニケーションをとる能力が必要になると考えられる.

本研究の最終目標は,人間と自然なコミュニケーションを行うためのエッセンスを身につけた自律ロボットを実現することである.ここで自然なコミュニケーションを行うためのエッセンスとは,人間ならば日常のコミュニケーションの中で自然に行っているであろう,コミュニケーションを円滑にするような役割のある動作のことを指す.例えば,複数人で話している際に,話している人の方向に顔を向けたり,相手の話に対して適切なタイミングで頷いたり,自分が何か表現をする際に適切なジェスチャーを表出したりする,などがその例として挙げられる.実際にこのような要素がコミュニケーションにおいて重要であることが,例えば [Gratch 07] などで報告されている.この研究では,会話の中身を理解しなくても,適度に頷きを入れることで人間がバーチャルエージェントに同調を感じることが実験によって確かめられている.

2. 研究の枠組み

2.1 先行研究とその限界

コミュニケーションエッセンスを身につけた自律ロボットを実現するためには,ロボットにコミュニケーションのルールを獲得させる必要がある.

例えば, [Cassell 94] では,エージェントの表情や頭・手の動き,発話のイントネーションなどをルールベースで自動生成するようなアプローチが紹介されている.このようなアプローチは対象とするコミュニケーションを限定した場合には,非常に強力な手段である.ところが,コミュニケーションとは本質的に非常に多種多様なものであるため,その全てのルールを列

挙していくことは大変困難であるし,仮にできたとしても大変な手間がかかってしまう.

そこでこの問題を解決するために,しばしばロボットにコミュニケーションルールを学習をさせるというアプローチが用いられる. [Ohashi] では,人間同士のコミュニケーションを模範としてバーチャルエージェントを学習させることにより,コミュニケーション中の人間の行動を模したエージェントを開発した.しかし,人間とロボットでは物理的な制約として,動かせる関節やその方向,可動域などに違いがある.また,心理的な面から考えても,人間を相手にするとロボットを相手にするのはその存在感にも違いを感じると考えられる.このような違いのために,人間同士のコミュニケーションを模範としてロボットを学習させてうまく動作をするようになったとしても,それが人間とロボットとのコミュニケーションにおいてもそのままうまく機能するとは限らない.従って,本研究のような目的でロボットを学習させる際には,人間とロボットのコミュニケーションデータを用いることが必要であると言える.

2.2 本研究のアプローチ

本研究では,人間とロボットとのコミュニケーションのデータを取得するために, Wizard Of Oz(WOZ, [Fraser 91]) の考え方を利用した新しいアプローチを提案する. WOZ とは,従来開発中の知的システムの挙動のテストなどの際に用いられてきたシミュレーション手法である.その基本的な考え方は,知的システムの裏で, Wizard と呼ばれる人間がその機能の一部または全てを肩代わりし,システムの振りをしながらユーザとやりとりするというものである. [岡本 02] では, WOZ 法を利用して対話型エージェントを構築するという試みがなされている.しかし,ロボットにコミュニケーションエッセンスを獲得させるためには,言語行動だけでは不十分である.そこで本研究では,従来行われてきた WOZ の枠組みを拡張し,非言語行動まで含めてシミュレーションができるようにする.ロボットを裏で自由に操作できるような環境(WOZ 環境)を構築し,ロボットの操作者がロボットを介して他の人とコミュニケーションを取ることで,人間とロボットとのコミュニケーションのデータを取得出来るようにする(図 1).

連絡先: 大橋洋輝, 京都大学 情報学研究科 知能情報学専攻,
〒 606-8501 京都市左京区吉田本町(工学部 10 号館 131
号室), 075-753-5867, ohashi@ii.ist.i.kyoto-u.ac.jp



図 1: WOZ のイメージ図 (左) と本研究のアプローチ (右)

3. WOZ 環境の構築

3.1 WOZ 環境に求められる条件

[Tachi 08] では、上で述べたような、ロボットを人間が自由に操作できる環境が紹介されている。この研究はリアルな遠隔地コミュニケーションの実現を趣向したものであるため、特殊な操作デバイスを用いて、遠隔地にいるロボットの触覚を操作者にも伝えて没入感を高めるといった取り組みがなされている。本研究では、人間とロボットとのコミュニケーションのデータを取得することが主な目的であるため、このように特殊なデバイスを身につけて、操作者のコミュニケーション行動に制約を加えることは望ましくない。人間とロボットとのコミュニケーションの適切なデータを取得するためには、ロボットの操作者が、日常行っているようなコミュニケーションをできる限りそのまま行えるような環境を構築しなければならない。例えば人間は誰かが話し始めたとき、何か物音がしたりモノが動いたときなどには、特に強く意識することなくそちらの方向に顔を向ける。あるいは、話している際に何かジェスチャーを表現する、話題のものを指差すなどの動作は、発話や何かそれ以外の動作とともに表出されることが多い。ロボット側の状況を把握するために複数台のカメラ映像を視点切り替えスイッチを制御しながら見たり、リモコンや操縦桿のようなものを用いてロボットを制御するような環境では、状況把握やロボットの操作のために特別な負担がかかってしまい、上述のような動作を取得できない。従って、構築する WOZ 環境では、操作者がロボットの置かれた状況を直感的に把握出来ること、直感的な操作でロボットが制御できることが必要になると考える。

3.2 本研究の提案する環境

本研究では上述の要件を満たすために、図 2 のような環境を構築する。

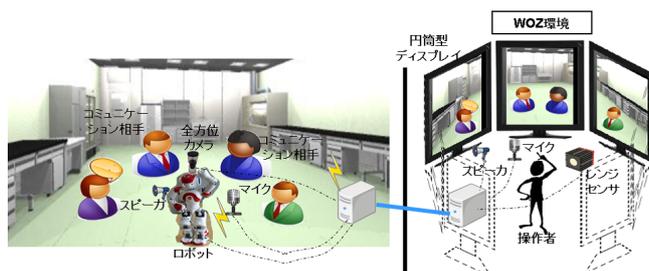


図 2: 本研究で構築する WOZ 環境

まず、全方位カメラという 360 度全方位方向の映像が同時に取得できるカメラをロボット側に設置し、その映像を円筒形ディスプレイという円筒形に配置された没入型大型ディスプレイに映し出す。操作者は円筒形ディスプレイの中でロボットの操作、即ち他の人々とのコミュニケーションを行う。このようにすることで、操作者の眼前にはロボットのいる環境がそのまま再現されることになり、ロボットの置かれた状況を直感的に把握することができるようになる。また、センサを用いて操作者の動作を自動的に認識し、操作者がある動作を取るとそれ

と同様の動作をロボットが取れるような仕組みで、ロボットの操作を行うようにする。この際、コミュニケーションを行う操作者の行動をなるべく制約しないように、非接触のセンサを使用する。非接触センサを用いて人の動作認識を行うには、光学式のモーションセンサを使うと安定した結果を得やすいと考えられるが、光学式モーションセンサにはマーカの隠れに弱いという問題が存在するため、多数のカメラを設置して様々な角度から計測を行わなければならない。本研究では、円筒型ディスプレイ内での使用を前提とするため、カメラを設置できる範囲には限界があり、光学式モーションセンサを用いることはできない。このため本研究では操作者の動作認識のためのセンサとしてはレンジセンサを用いる。レンジセンサは、一般のカメラと違い距離を含めた 3 次元情報をリアルタイムに取得できるため、円筒型ディスプレイ内で操作者の背後に映る可能性がある他の人の画像との切り分けが簡単にできると期待できる。ロボットについては Aldebaran Robotics 社の NAO を採用する。NAO はプログラム可能で、ワイヤレスでも動作できるロボットである。首、肩、肘、手首、指、股関節、膝、足首を制御することができ、基本的なコミュニケーション動作を行うことができると考えられる。本研究で使用するこれらの主なデバイスを図 3 に示した。



図 3: 本研究で使用する主なデバイス。左から全方位カメラ、円筒形ディスプレイ、レンジセンサ、ロボット NAO

3.3 システムプロトタイプ

上で述べたコンセプトに基づき、実際にシステムのプロトタイプを開発した。このプロトタイプでは、操作者の上半身の姿勢や腕の動き、顔の方向などを自動認識してその結果をロボットの動作に反映させることができるようになってきている。現在のところ、ポインティングなどのコミュニケーション中によく生じる基本的な動作が取得できている(図 4)。本研究では、レンジセンサを用いて距離情報を計測しているため、ステレオカメラのように、取得される情報に変換をかける必要もない上に、その距離情報を用いて背景と操作者の切り分けを k-means クラスタリングを利用した簡素な方法で実現しているため、動作認識をした上でロボットの制御を行っても、十分にリアルタイムで動作をさせることが可能になっている。音声情報については、無料ビデオ通話ソフト skype を用いてネットワーク越しに伝達している。これら言語・非言語の情報を共に送受信できる機能を持たせているため、現時点でもこのプロトタイプを用いればロボットを介したコミュニケーションを行うことが可能である。また、現在は全方位カメラの映像をネットワーク越しに転送する部分が未実装であるため、画角の広いネットワークカメラを用いることで代替しているが、今後この部分の実装を済ませれば図 2 の環境が全て実現されることになる。

4. 評価実験による予備的検討

4.1 実験の設定

本研究で提案する環境の有効性を確認するため、また現在のプロトタイプシステムの持つ課題を明確にするために実験を行った。実験の内容は 3 人の参加者とロボットが日常会話を行



図 4: プロトタイプの動作例．距離情報を用いた動作認識を行っているため，操作者の背後のディスプレイの映像が切り替わっても認識に影響を与えない．右の図のように腕と体に重なりがあるような場合にも，距離情報のクラスタリングを行うことである程度対応ができる．

うというものである．参加者は大学生で，ロボットの操作は筆者自らが行った．6人の参加者を得て，3人1組として2度実験を行った．実験はi) ロボットを含め4人で20分程度の日常会話を行う，ii) その様子を記録した映像を直後再生しながら(図5)，参加者に各シーンでロボットに対しどのような印象を持っていたか，何か行動を起こした時のきっかけは何であったかなどを詳しく話してもらい，iii) アンケート回答，という3つの手順を踏んで行った．実験環境の配置図を図6に示す．

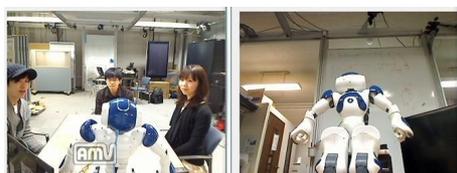


図 5: 直後再生ビデオのスクリーンショット．

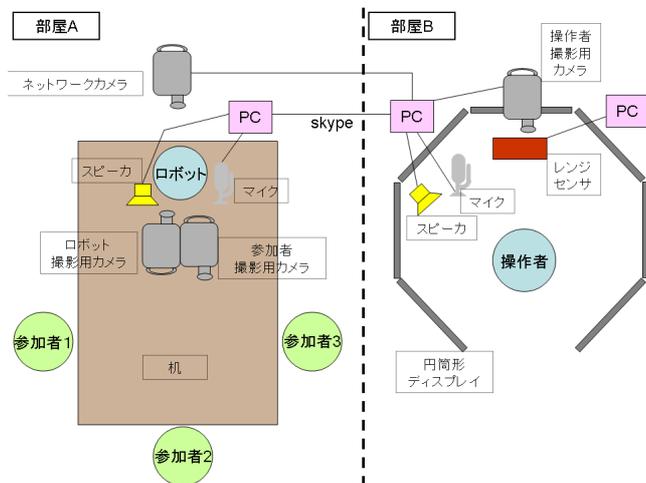


図 6: 実験の配置図．円筒形ディスプレイにはネットワークカメラからの映像が展開されている．参加者からは操作者の様子は見えていない．互いの音声は skype を通して伝達する．

4.2 実験結果と考察

実験で行ったアンケートの結果を表1にまとめる．アンケートは7段階のリカート尺度で計20項目行った．各項目の括弧内の左側の回答を+3，右側の回答を-3として+3~-3の7段階で集計し，同様のカテゴリーごとに結果をまとめて，6人の参加者の回答の平均値を算出した．

この結果を見ると，項目1, 3から，ロボットの動作に関しては，まだまだ不自然であると感じられている部分，遅延の影響が見て取れる部分などが存在することが分かる．このような

表 1: アンケート結果のまとめ．

	アンケート項目	平均
1	ロボットの動きは (自然だ 不自然だ)	-1.23
2	ロボットの動きは (意味がある 意味ない)	+0.50
3	ロボットの動きのタイミングは (適切だ 遅い)	-0.75
4	ロボットをコミュニケーション相手として (捉えた 捉えなかった)	+0.92
5	ロボットとのコミュニケーションは (円滑に行えた 噛みあわなかった)	+0.17

問題が開発の初期の段階で確認できたことは，はじめからロボットの身体を持った人間と，他の人とのコミュニケーションが実現できる本研究の環境ならではの利点であると言える．

アンケートの他の項目に目を向けると，2,4,5からは，この環境を用いることによって，ある程度人間とロボットとのコミュニケーションがうまく行えることが分かる．実験ビデオの直後再生時にも参加者から「ロボットの動作の意図を理解することができ，それに応じた反応をした」「話をしている時ロボットに顔を向けられていたことで，自分が話し続けていいことが分かった」といった感想が報告された．実験中には図7に示すように，ロボットの動作に従って参加者が何らかの行動を取るといった例も見られ，ロボットとのコミュニケーションは十分に可能であったと言える．従って，上であげているような動作のぎこちなさや遅延の問題を今後解決していけば*1，より自然な形で人間とロボットとのコミュニケーションが観測できると期待できる．

さらに，実験の様子を収録したビデオを分析したところ，実験中操作者は日常のコミュニケーションのように，話している人の方向を見たり，次に話を振る人に顔を向けたりと特に意識はせずとも頻繁に顔を動かし，それによってうまくコミュニケーションを取ろうとしていたことが分かった．さらに，顔を動かしながら同時に腕でも何らかのジェスチャーを行うような場面や，会話をしながら何か特定の動作をするような場面も多数見られた(図7)．実際にこのような動作を多数観測できたことは，ほとんど意識することなく，ただ日常行っているようなコミュニケーションをそのまま行うだけでこのような動作を実現することができるという，本研究の提案する環境の有効性を示していると言える．

4.3 現在の到達点と将来課題

本節では，予備検討の結果を受けて，研究の現在の到達点と今度どのような実験を行うべきかについて議論をする．

実験中参加者はロボットを無視して自分たちだけで話をするようなことはなく，一貫してロボットの呼び掛けや意味のある動作に積極的に応え，話題がある時には参加者からロボットに話しかけていた．実際に表1のアンケート項目4にまとめた項目を詳細に見ても，参加者がロボットと積極的にコミュニケーションを取ろうとしていたことが見て取れる(表2)．このことから，今回開発したプロトタイプは，人間とロボットのコミュニケーションが成立するための前提となる条件を満たしていたことが分かる．一方で，現状ではロボットの動作は不自然に感じられたことがアンケートの結果から分かったが，直後再生時の参加者の報告から，この印象にはロボットの動作に遅延があ

*1 [Shiwa 08] では，ロボットと人間のインタラクションにおいて，人が一番好ましいと感じる遅延時間は1秒程度であり，それが2秒に達すると好ましくないと感じる事が報告されている

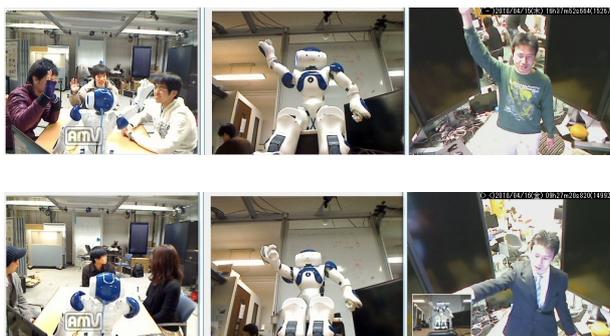


図 7: ロボットの動作を受けて参加者が何らかの行動をしている例。上段:「〇〇な人」と参加者に呼びかけ、参加者もロボットと同様の手を上げるというジェスチャーを行っている。下段:「あのディスプレイが」と発話をしながら、ロボットがある方向を指し示している。それにつられて参加者も指した方向を向いている。

表 2: 詳細なアンケート項目の一部抜粋。

アンケート項目	平均
ロボットに促されたとき (反応しようと思った 反応する気にならなかった)	+2.00
話題がある時、ロボットに (話しかけようと思った 話しかけようとは思わなかった)	+1.17

ること、動作が常に一定速度でなされることの二つの要因が大きく影響していたことが確認できた。従って、今後、人間とロボットの自然なコミュニケーションのデータを取得するという、本 WOZ 環境構築の目的を達成するためには、主にロボットの動作遅延へ対処と、動作に抑揚をつけるという部分的な問題に注力していけば良いと考えられる。このためには、操作者が動作を始めた段階でその動作を予測し、先行してロボットを制御するような動作予測のアルゴリズムの検討が必要となる。

また直後再生時に、参加者はロボットの顔の向きや腕の指す方向によく注意を払ったということが報告された。今回のようなロボットを含め 4 人で向い合って座するという実験設定では、指定する方向がどの参加者からも理解しやすく、そのような方向指定が有効に行えるため、この実験設定は妥当であったと考え、今後の実験に際してもこの設定を引き継いでいく。

今後自律ロボットの実現を目指すためには、参加者のどのような情報をロボットが収集すべきかを検討していかなければならない。そのために再度予備実験を行い、今度は実験協力者にロボットの操作を依頼し、今回の実験と同様にアンケートや直後再生によってこれらを調査していく予定である。今回の実験を観測する限りでは、参加者の顔の向きや発話の有無、表情や視線、腕の動きなどが重要な要素になると予想される。

最終的に人間とロボットのコミュニケーションデータが取得でき、自律ロボットを開発した際には、統計的な評価ができるように大人数の参加者を募って評価実験を行う予定である。その際には、アンケート評価に加え、[Ohashi] でも行った [Nosek 01] で紹介されている GNAT という潜在的な印象が計測できる心理学的手法や、我々の研究室で研究を進めた生理指標を用いる評価手法 ([大本]) を統合して、複数の観点から客観的な評価を行い、その結果を報告する予定である。

5. まとめ

本論文では、人間と自然なコミュニケーションが行える自律ロボットを実現することを最終目標とし、そのために必要とな

る人間とロボットとのコミュニケーションデータを容易に取得するための新しいアプローチを提案した。実際にその環境のプロトタイプシステムを構築し、実験を行って提案環境の持つ優位性を確認した。その一方で、実験を通して、ロボットの動作のぎこちなさや動作遅延の問題などが自然なコミュニケーションの障害になっていることが明らかになった。今後は、動作予測のアルゴリズムを実装することによって、遅延の解消と抑揚のある自然な動作の実現に取り組んでいきたいと考えている。さらに後には、実際にこの環境を用いて人間とロボットとのコミュニケーションのデータを取得し、その時のロボット操作者の動きをもとにロボットを学習させ、本研究の最終目標であった自律ロボットの実現へと研究を進めていきたいと考えている。

参考文献

- [Cassell 94] Cassell, J., Pelachaud, C., Badler, N., and Steedman, M.: ANIMATED CONVERSATION: Rule-based Generation of Facial Expression, Gesture & Spoken Intonation for Multiple Conversational Agents, *Proceedings of the 21st annual conference on Computer graphics and interactive techniques*, pp. 413–420 (1994)
- [Fraser 91] Fraser, N. M. and Gilbert, G. N.: Simulating Speech Systems, *Computer Speech and Language*, Vol. 5, No. 1, pp. 81–99 (1991)
- [Gratch 07] Gratch, J., Wang, N., Gerten, J., Fast, E., and Duffy, R.: Creating rapport with virtual agents, *Lecture Notes in Computer Science*, Vol. 4722, p. 125 (2007)
- [Nosek 01] Nosek, B. A. and Banaji, M. R.: THE GO/NO-GO ASSOCIATION TASK, in *Social Cognition*, Vol. 19, pp. 625–664 (2001)
- [Ohashi] Ohashi, H., HUANG, H.-H., and Nishida, T.: *Attentive Conversational Agent with Internal State Transition for Multiple Users*, Vol. 1 of *Smart Innovation, Systems and Technologies, SIST*, Springer: in press.
- [Shiwa 08] Shiwa, T., Kanda, T., Imai, M., Ishiguro, H., and Hagita, N.: How quickly should communication robots respond?, *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pp. 153–160 (2008)
- [Tachi 08] Tachi, S., Kawakami, N., Nii, H., Watanabe, K., and Minamizawa, K.: TELEsarPHONE: Mutual Telexistence Master Slave Communication System based on Retro-reflective Projection Technolog, *SICE Journal of Control, Measurement, and System Integration*, Vol. 1, pp. 1–10 (2008)
- [岡本 02] 岡本 昌之, 山中 信敏: Wizard of Oz 法を用いた対話型 Web エージェントの構築, *人工知能学会論文誌*, Vol. 17, No. 3, pp. 293–300 (2002)
- [大本] 大本 義正, 三宅 峰, 西田 豊明: 複数ユーザインタラクションにおける外発的な盛り上がりの雰囲気判定方法と影響の検討, *電子情報通信学会*, Vol. J93-D, No. 6 : in press.