

直感的ロボットWOZシステムの構築とそれを利用したHRIとHHIの比較

The construction of the robot WOZ system intuitively controllable and the comparison of HRI by using the system with HHI

大本 義正*¹ 大橋 洋輝*¹ 西田 豊明*¹
Yoshimasa Ohmoto Hiroki Ohashi Toyoaki Nishida

*¹京都大学 情報学研究科

Graduate School of Informatics, Kyoto University

In order to observe the natural interaction behaviors in HRI, it is necessary to allow the operator to intuitively understand the situation surrounding the robot, and to intuitively control the robot. This research aims to construct the WOZ system by using immersive environment and capturing and mirroring the operator's pose onto the robot and to evaluate the effectiveness of the system by investigating influences of physical limitations of a robot on interaction behaviors. We made the WOZ system by using multiple range sensors to improve prototype system in capturing the pose of the operator in real-time in narrow and closed immersive environment. We then conducted experiments of HRI and HHI by using the WOZ environment. As a result, we observed some interesting and important differences of interaction behaviors, such as pointing gestures and expressions of agreement, between HRI and HHI. The differences suggested the effectiveness of the WOZ environment.

1. はじめに

近年、家事や介護、情報提示、接客など、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。こうしたロボットには、その分野特有の能力が求められるのと同時に、人間と自然なインタラクションを行う能力が必要になると考えられる。

これまでも自然な Human-Robot Interaction (HRI) を実現するために、多くの研究がなされてきている。代表的なアプローチとしては、Human-Human Interaction (HHI) を観測し、それを分析してルールを記述しロボットに実装するもの ([Mutlu et al. 2009] など)、HHI データから機械学習の手法を用いてルールを自動抽出しようとするもの ([Huang et al. 2010], [Lee et al. 2010] など) 等が挙げられる。これらの先行研究では、研究対象とした特定の機能においては、ある程度うまくいったことが示された。

その一方、人間と、ロボットやエージェントとの身体的・心理的差異が考慮されていない。このような差異は、インタラクション行動の表出方法やインタラクションのプロセスに影響を与えるため、HHI におけるインタラクション行動を完全に再現できたとしても、それがロボットの身体を通して表現されたときに同様の意味を伝達するとは限らない。このような問題を解決するためには、実装するロボットの身体を用いて表現されたインタラクションデータに基づいて、ロボットの行動生成をすることが必要である。

実際の HRI データを取得する方法として、WOZ (Wizard Of Oz) による HRI の分析が行われてきた ([Kanda et al. 2007], [?] など)。WOZ を用いてインタラクションを行う上では、操作者 (Wizard) がある程度自在にロボットなどをコントロールできることが重要である。先行研究では、ロボットやエージェントなどが取り得る行動を予め規定しておき、比較的少ない行動を選択する形でこの条件を満たしてきた。しかし、このような方法は、ロボットが行うべき意思決定が単純で、行動自体もあらかじめ想定できる場合にしか適用できない。

複雑な情報に基づいた意思決定を行い、その結果として現

れる柔軟なインタラクション行動を、WOZ 操作されたロボットによって実現するには、インタラクション対象とロボットとの間でやりとりされる情報を適切に把握し、その上で、状況に合わせた柔軟な身体動作を直感的に表現できるような、新しい WOZ 環境が必要になる。

本研究の目標は、このような直感的な HRI を実現できる WOZ 環境を構築し、その有用性を検討することである。我々は、以前よりこのような WOZ 環境の開発を進めてきた ([Ohmoto et al. 2010])。本研究では、プロトタイプ WOZ 環境を改善したシステムを構築し、WOZ による HRI と同じタスクにおける HHI について、ロボットの制約が身体行動に与える影響を実際に検討することで、システムの評価を行う。

2. WOZ 環境

構築する WOZ 環境は、ユーザがロボット視点の情報を適切に取得でき、かつ、インタラクション行動に直接関係のない操作 (例えば、環境カメラの操作や動作モードの選択など) を極力排除し、ロボットの制御自体の負荷も小さくする必要がある。よって、できる限り操作者の身体にインタラクションの妨げとなるような特殊なデバイスを付けることを避けながら、直感的にロボットの操作ができるようにしなければならない。身体動作によってロボットを操作できる既存のシステム (例えば、[Ghobadi et al. 2008], [Tachi et al. 2008] など) は、身体に装着した操作デバイス自体がインタラクション行動を阻害したり、特定の狭い領域に存在する手や顔などの一部分による制御しかできない、といった問題がある。

我々の作成していたプロトタイプ ([Ohmoto et al. 2010]) では推定できる姿勢にいくつかの制約が存在したため、それらの制約を緩和して、ロボットに実現可能なほとんどの姿勢推定 (頭部および上半身) が可能になるように手法を改善した。これにより、実際の HRI を観察できる環境が整った。

2.1 アーキテクチャ

ロボット周辺の情報を全方位カメラを用いて取得し、その映像を 7 枚の大型ディスプレイで構成される円筒型ディスプレイに投影する没入型環境を利用することで、ロボットの置かれ

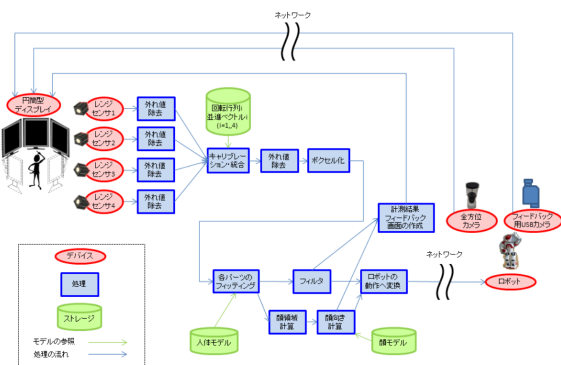


図 1: WOZ システムアーキテクチャ

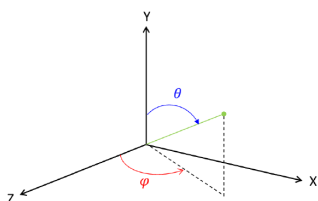


図 2: 3次元角度の定義

ている状況を直感的に把握できるようにする．それとともにロボットの姿を正面から捉えたフィードバック画像と姿勢推定の結果が操作者に表示されてフィードバックされる．操作者は没入型環境に提示された任意の対象とインタラクションを行う．没入型環境内には，操作者の姿勢推定システムが組み込まれており，その結果をロボットの身体にマッピングすることによってロボットの動作は制御される．人の動作推定を行うには，光学式のモーションセンサが存在するが，多数のカメラを様々な角度に設置する必要があるため，円筒型ディスプレイ内での使用には向かない．

本研究で構築する WOZ システムのアーキテクチャを図 1 に示す．インタラクションを行う操作者の行動をなるべく制約せずに，没入型環境内という狭小空間で計測するために，姿勢推定には複数のレンジセンサを使用した．このレンジセンサのキャリブレーションを行い，ボクセル化されたデータに対して人体モデルのフィッティングを行った．その推定結果をロボットの姿勢へマッピングし，同時に，顔と判定された領域における特徴点抽出によって顔向きを計算した．

2.2 複数レンジセンサによる姿勢推定精度

本研究において改善した，姿勢推定とその結果のロボットへのマッピングについて，主に制御をする上腕・前腕部分の評価を行った．図 2 のように θ と ϕ を定義し，初期位置からの回転角 θ の誤差を評価した．光学式モーションキャプチャシステム MAC3D (ナックイメージテクノロジー) による計測から得られる操作者の腕の回転角 (以下「人 (MAC3D)」) を正解データとし，レンジセンサを用いて推定される操作者の腕の回転角 (以下「人 (レンジセンサ)」) ，及び MAC3D による計測から得られるロボットの腕の回転角 (以下「ロボット (MAC3D)」) の誤差を計算した．

約 45 秒間 (233 フレーム) でインタラクション中に生じ得る様々な腕の動きを収録した，MAC3D 及びレンジセンサによる姿勢推定から出力される初期位置からの回転角の変遷は

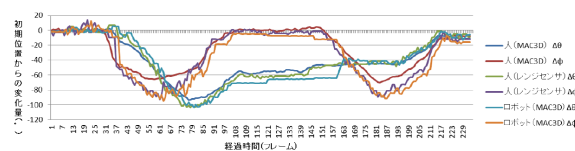


図 3: 上腕の位置変化の変遷

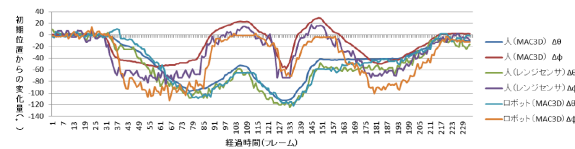


図 4: 前腕の位置変化の変遷

図 3, 図 4 に示すようになった．これを見ると，方向に関しては前腕・上腕ともおよそ 7.5 度以内の精度が達成できていることが分かる．一方，方向に関しては比較的大きな誤差が出ているものの，その誤差は一定方向に偏っていることから，アルゴリズムの改善により，誤差を軽減することが可能になると考えられる．前腕で誤差が大きい，上腕部の誤差が前腕部の推定の際に影響して増幅したためと考えられる．

姿勢推定の結果の例を図 5 に示す．コミュニケーション中によく生じるポインティング・首を傾げるといったジェスチャーの他，体の前でもの大きさを表現するようなジェスチャーや腕組みなど，体のパーツ同士の重なりが大きいようなジェスチャーを表出している場合でも正しく姿勢推定ができています．また，万歳のポーズやお辞儀など体が大きく動くような場合でも推定ができています．

3. インタラクション実験

開発した WOZ システムを用いた HRI と，ロボットと同様の制約を課した HHI の観察を行う実験を行った．本実験では，HRI データを取得する上で本研究の開発した WOZ システムが有効に働いているかどうかを，ロボットの持つ身体的制約によってもたらされる，HHI と HRI の間のインタラクション行動の違いにより検討する．

3.1 タスク

タスクの目的は，クリスマスツリーをきれいに飾り付けることであった．実験参加者は，クリスマスツリーの飾りつけを実際に行うデコレータ，もしくは，デコレータに飾り付け方を指示するアドバイザーとして実験に参加した．アドバイザーはアドバイスをするだけで飾り付けはしない．

アドバイザーとデコレータの発話には，あらかじめ定められた限られた種類の発話のみを組み合わせで行うという制約が課された．また，HHI 条件においては，今回使用するロボットの身体性に合わせた制約を課した．また表情や視線情報を伝達してしまうことを避けるために，マスクとサングラスを着用してもらった．逆に，HRI 条件では，人間が制御していることを隠すために，機械的に聞こえるように音声を変換した．

3.2 実験参加者

実験参加者は大学生，男性 19 名，女性 10 名の合計 29 名の参加者を得た．これらの参加者を以下のように振り分けて，合計 17 セットの実験を行った．

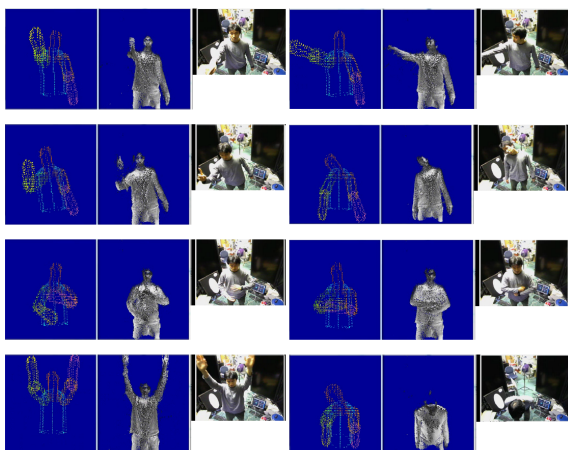


図 5: 姿勢推定結果の例

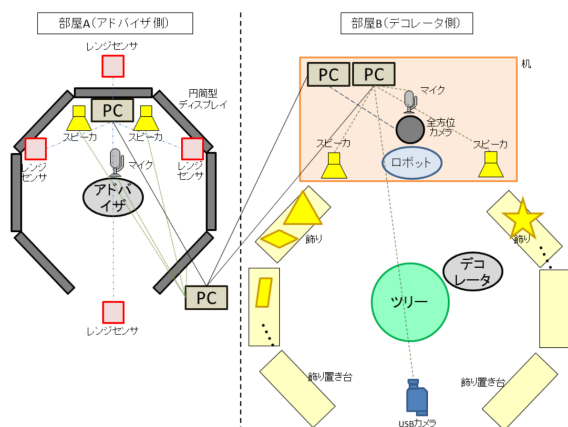


図 6: 実験環境

HRI 条件 合計 13 セット

アドバイザ: 実験者, デコレータ: 参加者, 5 セット
 アドバイザ: 参加者, デコレータ: 参加者, 8 セット

HHI 条件 合計 4 セット

アドバイザ: 参加者, デコレータ: 参加者, 4 セット

3.3 実験環境

実験室の配置図を図 6 に示す。実験の様子は、部屋ごとに配置された 4 つのカメラによって記録された。

HRI 条件では、アドバイザは別の部屋から本研究の WOZ 環境 (部屋 A) を用いてロボットをコントロールしてアドバイスをを行った。HHI 条件では、図の部屋 B だけを使用し、アドバイザはロボットの位置に座って対面で指示を出した。

3.4 実験手順

実験時間は、アドバイザは約 2 時間、デコレータは約 1 時間 30 分であった。飾り付けに関する学習を行うため、アドバイザの方が実験時間が長くなった。実験の手順は以下の通り。

アドバイザ側

1. 概要説明
2. WOZ 環境における計測のための前準備
3. 飾り付けに関する学習, タスク内容の確認

4. WOZ 環境の使い方説明
5. 飾り付けタスクの実行
6. アンケート記入

デコレータ側

1. 概要説明
2. タスク内容の確認
3. 飾り付けタスクの実行
4. アンケート記入

3.5 結果と考察

WOZ 環境を用いた HRI と HHI の間にどのような共通点や差異があるかを調べるために、実験ビデオの分析を行った。以下にビデオ分析で見られた特徴的な傾向を挙げる。

1. HRI では方向を指し示すポインティングをする際に両腕を使うことがあった (HRI : 4/8 人, HHI : 0/4 人)
2. 「もう少し上に」などのアドバイスをすると、HHI では小刻みに手を動かすジェスチャを使ったが、HRI では移動する方向自体を指し示すことがあった (HRI : 5/8 人, HHI : 0/4 人)
3. HRI では方向指示に、ジェスチャと発話を統合的に用いることが多かった (HRI : 60.1%(101/168), HHI : 37.6%(35/93))
4. Yes/No の意思表示は、HHI では細かく頷くことが多いが、HRI では言語を伴うことが多かった
5. HHI では発話の際に文末が省略されることが多かったが、HRI では最後まで発話されることが多かった
6. HHI ではロボットの身体性の制約を超えた動作を行ってしまうことが頻繁に見られた。HRI でも同様の場面が見られたが、ロボットに表現できないことを確認して、表現方法を変えるという場面が見られた (HRI : 6/8 人)
7. HRI, HHI ともに言語と非言語を統合的に利用していた (HRI : 64.2% (292/455), HHI : 56.5% (134/237))
8. HRI, HHI ともに両方の手を使ってアドバイスを行うことがあった (HRI : 7/8 人, HHI : 3/4 人)

1~5 の項目には、HHI と HRI におけるインタラクション行動の違いが現れているものと考えられる。全体的に、HHI の場合よりも HRI の方が表現が大きさになる傾向が見られたと言える。これは、HHI 条件の場合には、制約しきれない微妙な仕草などで相手に意図が伝わることがあるのに対し、HRI 条件の場合では、ロボットの身体的制約が厳密に適用されるため、何度か意図がうまく伝わらない試行を経て、相手に伝わりやすいように表現が変化したと考えられる。こうした HHI と HRI の違いが実際に観測されたことで、HRI における最適な行動の実装のために、本研究で提案しているような WOZ 環境の必要性が認められると考える。

また、上記のような微妙な仕草に加えて 6 のような現象が見られることから、ロボットの身体的制約を厳密に人間に適用できないことが分かる。こうした制約しきれなかった要素によって意図が伝わってしまった場合、ロボットがどのような動作を行えば相手に同様の意味を伝達できるかは自明ではない。本研究の WOZ 環境を用いると、得られるデータにはロボットに表出不可能なものは含まれ得ない。さらに、ロボットの身

体で表現できないことに操作者が気づき、その場でロボットに表出可能な表現方法に変えて相手に伝える様子も見られた。こういったことから、HRIにおける最適な実装を探る上で、提案しているWOZ環境の有効性が確認できる。

7や8のような行動は、WOZを直感的な身体動作によって行うことの有用性を示唆していると考えられる。仮にキーボードインタフェースなどを用いてロボット操作をしていたら、これらのように言語と非言語を統合的に用いたり、複数の非言語行動を同時に表出するように制御することは容易ではない。

4. 議論

クリスマスツリーの飾りつけという同一のタスクの中でも、HRIとHHIの間には様々な違いが見られた。これは人間とロボットの差異が表れた結果であると考えられる。また、本研究のWOZシステムを用いて実際にロボットの身体で表出するのに適した表現方法が獲得されていく様子も観測することができた。これらによって、本研究のようなWOZ環境の重要性は確認できたと考えられる。

ビデオ分析に加えて、提案するWOZ環境の、操作者の意図を伝達する能力を確かめるために、アドバイスの理解に関するアンケートの結果を分析した。実験者の行ったHRIと参加者同士で行ったHHIの間で、Mann-WhitneyのU検定を行ったところ、飾り付け方を非言語で指示した場合において、HRI条件の方が有意に理解しやすかったという結果が得られた。このことから、本研究の開発したWOZ環境によるロボット操作において、特定のタスクに熟練した人(実験者)が、熟練していない人(参加者)よりも適切に、知識やノウハウなどを相手に伝えることができることを示唆していると考えられる。また、他の項目(言語表現、方向指示、身体動作など)における理解度には大きな違いはなかった。これは、本WOZ環境によるロボット操作において一定以上の意図が円滑に伝達可能であることを示唆しているといえる。

一方で、実験参加者のアンケートの回答からは、動作の正確性や定常性などにおいて、本研究のWOZ環境が十分な能力を持っていないと感じられる部分があることが伺えた。例えば、ロボットの動作存在する遅延について、インタラクションに支障はなかったものの、スムーズにインタラクションができたとは感じられていなかった。このような問題に対処するためには、操作者の動きの時系列データから、次の動きの予測を行うというような改良が考えられるだろう。そのため、計測値にカルマンフィルタを応用して姿勢を推定する手法を開発するなど、さらなるシステムの改善が必要である。

5. 結論

本研究では、複雑な情報に基づいた意思決定を行い、その結果として現れる柔軟なインタラクション行動を、WOZ操作されたロボットによって実現する環境を構築することを目標として、没入型環境における直感的WOZシステムの構築に取り組んだ。プロトタイプを進展させ、実際のHRIが観察可能なWOZ環境の実現に必要な機能を開発した。

開発したWOZ環境を用いて行なったインタラクション実験では、本WOZ環境にHRIを実現する能力があることを確認することができた。また、実際にHHIとHRIとの間に、ロボットの最適な行動を実装する上で重大な問題となるような差異が確認できた。これらによって開発したWOZ環境の有用性を示すことができたと考えられる。

一方で、ロボットの操作をした参加者が思い通りの動作表出ができなかった部分も見られた。今後は、実験で明らかになった改善点を修正した上で、本WOZシステムを用いて収集したHRIのデータを自律ロボットの行動学習に適用する予定である。

参考文献

- [Ghobadi et al. 2008] Ghobadi, S. E., Loeprich, O. E., Ahmadov, F., Bernshausen, J., Hartmann, K. and Lof-feld, O.: Real Time Hand Based Robot Control Using 2D/3D Images, *Advances in Visual Computing*, Vol. 5359/2008, pp. 307-316 (2008).
- [Huang et al. 2010] Huang, L., Morency, L.-P. and Gratch, J.: Parasocial Consensus Sampling: Combining Multiple Perspectives to Learn Virtual Human Behavior, *Proceedings of 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, pp. 1265-1272 (2010).
- [Kanda et al. 2007] Kanda, T., Kamasima, M., Imai, M., Ono, T., Sakamoto, D., Ishiguro, H. and Anzai, Y.: A humanoid robot that pretends to listen to route guidance from a human, *Autonomous Robots*, Vol. 22, pp. 87-100 (2007).
- [Lee et al. 2010] Lee, J., Wang, Z. and Marsella, S.: Evaluating Models of Speaker Head Nods for Virtual Agents, *Proceedings of 9th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2010)*, pp. 1257-1264 (2010).
- [Mutlu et al. 2009] Mutlu, B., Shiwa, T., Kanda, T., Ishiguro, H. and Hagita, N.: Footing in human-robot conversations: how robots might shape participant roles using gaze cues, *Proceedings of the 4th ACM/IEEE international conference on Human robot interaction*, ACM, pp. 61-68 (2009).
- [Ohmoto et al. 2010] Ohmoto, Y., Ohashi, H., Takahashi, A. and Nishida, T.: Capture and Express Behavior Environment (CEBE) as WOZ system to realize effective human-agent interaction, *AAMAS 2010 International Workshop on Interacting with ECAs as Virtual Characters*, pp. 63-69 (2010).
- [Okamoto et al. 2001] Okamoto, M., Yang, Y. and Ishida, T.: Wizard of Oz Method for Learning Dialog Agents, *Proceedings of the 5th International Workshop on Cooperative Information Agents V*, *Lecture Notes In Computer Science(LNCS)*, Vol. 2182, pp. 20-25 (2001).
- [Tachi et al. 2008] Tachi, S., Kawakami, N., Nii, H., Watanabe, K. and Minamizawa, K.: TELEsar-PHONE: Mutual Telexistence Master Slave Communication System based on Retro-reflective Projection Technolog, *SICE Journal of Control, Measurement, and System Integration*, Vol. 1, No. 5, pp. 335-344 (2008).