2K1-R-11-5

操作者の振る舞いと実世界情報の統合による半自律型テレプレゼンスシステムの構築

Construction of A Semi-automatic Telepresence System Correcting Motion by Integrating A Controller's Behavior With A Remote Sensing Data

大本 義正[*1] 齊賀 弘泰[*1] 西田 豊明[*1]
Yoshimasa Ohmoto Hiroyasu Saiga Toyoaki Nishida

[*1]京都大学 情報学研究科
Graduate School of Informatics, Kyoto University

There are some differences between human beings and robots such as their body structures and recognition of real world. The differences prevent natural and smooth interaction when human controls a robot in a remote location. To realize an intuitive remote robot-human interaction, we developed a new tele-presence system semi-automatically generating motions of remote-controlled robot based on a real world situation and human behavior and intentions. For the purpose, we implemented three semi-automated behavior of a remote-controlled robot; detecting and converting a pointing motion, mutual gaze behavior as a backchannel, and fltering motions according to the situation. We conducted interaction experiments to evaluate the tele-presence system which implemented the semi-automated component. The result shows that a collaborative worker better understand the robot behavior controlled by our semi-automated system than by a normal system. In addition, operators of the system tended to naturally express their interaction behavior by using the robot body.

1. はじめに

近年，出張用の簡易なコミュニケーションロボットから診療，介護ロボットなど遠隔作業を行うロボットに対する需要が高まっている（e.g. AnyBots 社の QB 等）．このようなロボットと人間では，感覚や身体が大きく異なることが多い．これらの違いは，操作者が受信する情報と送信する情報に，認知的あるいは身体的な齟齬を生じさせるため，ロボットのコントロールに重大な影響を与える．

情報の受信における齟齬は，遠隔地から操作者に対し送られてきた情報が実際にその場にいる情報に対し少ないことで発生する．例えば，ロボットのカメラ位置と身体の関係が人間から見た場合と異なっている場合や，光学カメラ以外のセンサによる人間には通常認識できない情報を利用する場合等が典型的な場面としてあげられる．このような問題に対して，立体的な三次元映像の復元を行ったり，特殊なセンサを用いて触覚的フィードバックを得ることよって，操作する人がその場にいるかのように遠隔地での環境を再現する試みは多くなされている（e.g. Tachi ら [5] など）．

情報の送信における齟齬は，操作者が表出する情報がロボットによって表現する際に情報が劣化することで発生する．例えば，関節の可動範囲が狭かったり身体が小さかったりすることで身体表現が一致しない場合や，操作者の行動計測によるノイズが反映されることで身体表現の情報粒度が下がる場合などがあげられる．このような問題に対しては，ロボットや計測手法のパフォーマンス向上というアプローチで解決を目指すことが多い．ロボット自身にセンシング能力を持たせることで操作者の支援を行う研究もいくつかなされている（e.g. Mizumoto ら [3]，Briere ら [1]）が，これらの研究は実世界でのセンシングした情報を表示したり，操作者の意図とは関係なく回避を行うなど操作者の認知面の強化を目的としている．

こうした操作者とロボットの間で送受信される情報の齟齬に起因する問題に対して，我々は，操作者の振る舞い（操作者が受信した情報に従って送信する情報）だけでなく，遠隔地でロボット自身がセンシングした情報を統合して，状況に合わせてロボットの行動を適切に生成することを目指す．本研究では，ロボットによる実世界のセンシング情報と操作者の振る舞いを統合的に解釈し，認知的・身体的な齟齬を低減するように行動を補完する，半自律型テレプレゼンスシステムを構築する．そのために操作者の動きをセンシングする操作基盤と，遠隔地の情報をセンシングする実世界センシング部，そしてそれらを統合し行動を生成する半自律行動部の三つのコンポーネントからなる新しいテレプレゼンスシステムを提案する．

連絡先: 大本義正，京都大学，京都府京都市左京区吉田本町，ohmoto@i.kyoto-u.ac.jp

2. テレプレゼンスシステムの構築

2.1 本研究において補完する動作

我々はまず，操作者とロボットの間で送受信される情報の齟齬がどのように影響しているか分析するため，研究グループにおいて作成していたプロトタイプ [6] によって予備検討を行った．その結果に基づき，本研究で半自律行動として補完する動作を以下の 3 つとした．

身体動作の抑制 ：姿勢推定の計測結果や人間の身体動作自体にはコンテキストとは無関係のノイズがのる．このようなノイズを低減するため，操作者が身体表現を行っていないと判断したときに，ロボットが余計に動かないようにフィルタをかける必要がある．

バックチャネル ：人間同士の対面コミュニケーションでは，顔向けや頷き、相づちといったバックチャネルを生成する．特に予備検討で問題になった顔方向に対処するため，相手が注視してきた際に顔方向を補正し，相手へ顔向けを行うことを行う．また，相手の発話直後に頷きを生成する．

指さし ：身体構造とオブジェクトの相対的な関係によって，指さしの意味が決定される．テレプレゼンスでは，操作者がモニタに表示されたオブジェクトを指す場合の姿勢と，実世界でのロボットとオブジェクトの関係は，通常は一致しないと考えられる．そのため，遠隔地の情報と操作

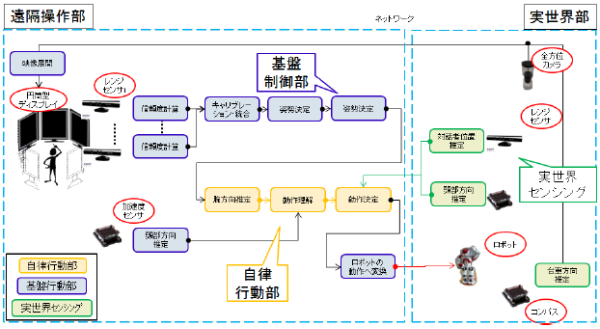

図 1: 半自律テレプレゼンスシステムアーキテクチャ

者の身体動作を統合して指さし行為を補完する必要がある．補完は方向の補正だけでなく，ロボットの体，顔の向きも補正し，より明示的な指さし行為とする．

いずれの補完動作においても，適切な動作生成のためには，実世界の状態と操作者の動作の間の齟齬を，タスクの文脈と操作者の意図に基づいて推測する必要がある．本研究では，あらかじめタスク内容を分析して文脈を固定し，各文脈を人間が WOZ（Wizard of OZ）によって入力することとした．従って，文脈を既知のものとして，操作者の意図を文脈と身体動作に基づいて推測し，適切な動作補完を行うことを目標とした．

2.2 テレプレゼンスシステムの概要

本研究で目標とするシステムを作成するためには，遠隔地での身体表現を行うロボット，操作者の身体動作および遠隔地の実世界情報を取得するセンシング環境，それらを統合して行動を生成する半自律行動生成コンポーネントが必要である．これらを踏まえたシステムのアーキテクチャを図 1 に示す．

身体表現を行うロボットとして，ヒューマノイドロボット NAO（Aldebaran Robotics）を採用した．NAO は首，肩，肘，手首，指，股関節，膝，足首を制御することが可能であり，自由度の高い表現能力を持つ．さらに，実世界内を動くインタラクションを円滑に遂行するため，前後移動・左右平行移動・回転を行うことができる台車に乗せた．これによって基本的なコミュニーケション表現を人間に類似した身体によって実現することができた．操作者にロボット視点での情報を直感的に与えるため，全方位没入型環境 ICIE[4] を使用した．そして，全方位カメラで撮影したロボットの周囲の映像をネットワークを通じて没入型環境に送信し，画像を全周囲に投影した．

システムは，直観的な操作を行うための環境である基盤制御部と実世界センシングを行う実世界センシング部，それらの情報を統合して行動生成を行う半自律行動部の 3 つからなる．

基盤制御部では操作者の姿勢推定を行う．没入型環境では，光学式のモーションキャプチャに必要なカメラを安定的に計測するために必要なだけ設置することが難しかった．そのため，本システムでは Microsoft Kinect によって姿勢推定を行うこととした．しかし，Kinect による姿勢推定はカメラが正面にあるときは安定しているものの，横を向くなどすると計測エラーが生じる．没入型環境では頻繁に体の向きが変わるため，本システムでは複数台の Kinect を円周上に配置し，それらの情報を統合することで頑健な姿勢推定を行った．頭部方向はインタラクションに重要な情報を持っており，画像を用いて推定する手法がいくつか提案されている．しかし，没入型環境では身体の回転や腕の動きによってオクルージョンが頻繁に発生するため，画像による推定手法では安定的に取得できないと考えられた．そこで本システムでは加速度，ジャイロ，地磁気を組み合わせたセンサ（IMU-Z）で頭部方向を計測した．

実世界センシング部では，半自律行動の生成のために必要となる，ロボットが存在する実世界側の情報を取得する．本研究で目標とする補正は指さしや相手の注視に対する頭部方向の補正であるため，実世界側で必要となる情報は対話者との相対的な位置関係，向きといった情報を計測した．特に作業者とのインタラクションにおいて相手の顔を注視するバックチャネルを行うためには，現在の相手との相対位置を取得する必要がある．そのためロボットにレンジセンサを取り付け，対象物，対話者の相対位置を取得した．頭部方向は，加速度，地磁気，ジャイロセンサを統合したセンサによって取得した．

半自律行動部では，指さし，バックチャネル，動きの抑制の 3 つを生成する．これを生成するために，実世界センシング部で取得された情報を統合する．本システムでは，非言語表現を行おうとしている場合に指さしかどうかを判定し，指さしであれば指さし行動への変換を行う．次にそれ以外の時では相手がロボットを見ているかどうかを調べ，見ていれば相手の顔を見返すバックチャネル行為を生成する．最後に，なにも非言語表現を行っておらず，かつ対話者が見ていない状態であれば余計な行動を行わないよう身体の動きの抑制を行うこととした．

3. 評価実験

本実験では，実世界情報と操作者の動きを統合した半自律コンポーネントを持つシステムと，操作者の動きに追従するテレプレゼンスシステムの間で，ユーザのインタラクションの円滑さがどのように変化するのか，また，インタラクション中に行われる身体表現にどのような変化が起きるのかを調べることを目的とする．そのためにクリスマスツリーの飾り付けおよび，作業風景の記録というタスクを行った．分析では，タスクにおける非言語行動の表出の変化，ロボットとインタラクションを行う作業者二人のロボットへの注視度を記録したビデオ，および，実験後に記入してもらったアンケートに基づいて行った．

3.1 タスク

本実験では，子供向けのクリスマスツリーの飾り付けの動画教材を，3 人で協力して作成することを目的とした．飾り付け作業は 9 フェーズに分かれており，各フェーズには道具のつけ方や配置の仕方など 2 つ程度のポイントが含まれていた．撮影はそのポイントに応じてどこを撮影してほしいかを指示した．

3 人のうち 1 人が指示者として，遠隔地の没入型環境から指示を出した．指示者はあらかじめ作成したマニュアルに記載されている飾り付け，および，撮影におけるポイントを暗記してもらった．子供向けに言葉は簡潔で分かりやすくし，身振り手振りを使って子供が興味を持ちやすくなるよう教示した．

残りの 2 人のうち 1 人はロボットの指示にしたがって飾り付けを行った．飾り付け道具は一つのテーブルにまとめられていた．残りの 1 人は動画教材の撮影者として飾り付けの様子をカメラで撮影した．この際，動画の様子は大型のディスプレイに表示され，指示者が確認できるようにした．

3.2 実験参加者

実験参加者は大学生である．参加者の内訳は，男性 15 人，女性 12 人の計 27 人であった．1 回の実験には同性 3 人ずつが参加し，計 9 セッション（半自律行動あり：5 セッション，半自律行動なし：4 セッション）行った．全員ロボットとのコミュニケーションは初めてであった．

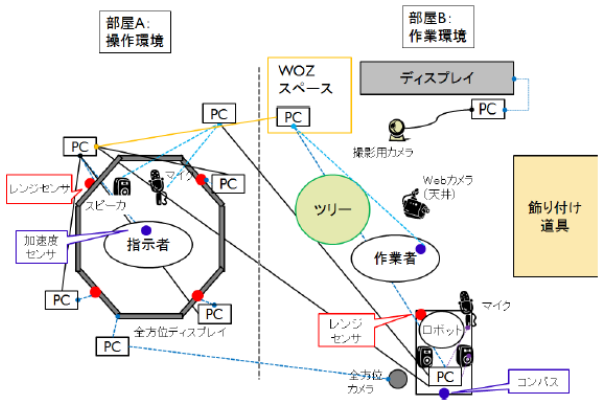

図 2: 評価実験環境

3.3 実験環境

実験の機材の配置を図 2 に示す．没入型操作環境はレンジセンサの処理を行う PC4 台の他，全方位映像を映す PC，音声通話用の PC で構成される．音声通話用 PC にはマイクとスピーカが接続されており，音声による双方向の会話を可能としている．なお作業者に周囲の雑音が聞こえないよう，操作者にはヘッドセットをつけてもらった．作業環境では，ロボットの正面に撮影者のカメラの動画を表示した．右側に飾り付け道具を置き，左側にはツリーを設置した．ロボットは台車に乗っており，そこに音声会話用のマイクおよび会話用の PC，さらにセンシング用のレンジセンサを配置した．

本実験時点では，作業者二人の識別および追跡位置の初期化を自動化できなかったため WOZ 操作で行った．また，システム概要で説明したように，作業の文脈を WOZ で入力した．本実験において，作業の文脈とは，タスクの各フェーズと，フェーズの最初に作業を説明する以降期と，作業の経過確認などを行う確認期の 2 つのタームが対応した．これらの WOZ 操作のためにパーティションで仕切った空間にスペースを確保し，実験者はここで操作を行った．

3.4 結果と考察

3.4.1 非言語行動の表出の違い

指さし動作が何回行われたかについて表 1 にまとめた．ただし動画がうまく撮影できなかったため session2 は分析から除外した．指さしの総数では，全体的に半自律行動を行う場合で回数が多いという結果が得られた．ただし有意差はなかった．これは，セッション 5 において指さし回数がかなり少なかったためであると考えられる．サンプル数を増やせば差が出ることが期待される．

表 1: 指さし動作の種類ごとの回数

	半自律（有）				半自律（無）			
No.	1	3	8	9	4	5	6	7
指さし回数	8	9	7	5	17	2	17	6
同じ側の指さし回数	3	3	0	2	7	0	9	3
反対側の指さし回数	0	2	0	0	7	0	9	3

同じ側の指さし回数とは，例えば，ロボットから見て左側にあるものを左手で何回指をさしたかということである．それに対し反対側の腕の使用数とは，操作者から見て左側にあるものを右手で指さすことである．半自律行動を行っている場合，反対側の腕を用いて指さすことがセッション 5 以外で見られた．それに対し，半自律行動を行わない場合セッション 3 のみ反対側の腕を使って指さした．反対側の腕を使用する回数に対し Mann-Whitney の U 検定を行ったところ，p=0.090 となり有意傾向が見られた．これは，半自律行動が無い場合，指さしをわかりやすくするために，操作者は左側にあるものは左腕で指さし，右側は右腕で指すためではないかと考えた．これにより，半自律行動によって人間が日常的に行う直感的な操作により，意図の伝達が可能になったといえる．

指さし以外のジェスチャ回数について，表 2 に結果をまとめた．こちらも半自律行動を行うセッションの方がジェスチャが多いという結果が得られた．また肘先だけではなく，上腕も動かす大きなジェスチャに絞って回数を調べたものが，(大) 数という項目である．(大) 割合は，表現したジェスチャ中の割合である．ジェスチャの総数，大きいジェスチャ表出数に対し Mann-Whitney の U 検定をかけたが有意差は見られなかった．しかし大きいジェスチャの表出割合に対する検定結果は p=0.091 であり，有意傾向を得ることができた．これは，半自律行動を行う場合，指さしや顔向きを適切に行い，それ以外の場面では動作のノイズを抑制することで，意味のある非言語に注意を向けることができたため，あまり大げさな動きをする必要が無かったのではないかと考えられる．このことから，半自律行動は本研究の目標とする操作者の動作意図とロボットが表出する意図の間の齟齬の解消に貢献していると考えられる．

表 2: 指さし以外のジェスチャの回数

	半自律（有）				半自律（無）			
No.	1	3	8	9	4	5	6	7
総数	13	11	30	7	31	9	26	21
(大) 数	3	4	12	0	6	1	5	0
(大) 割合	0.23	0.36	0.4	0	0.19	0.11	0.19	0

3.4.2 アンケート

遠隔地で作業を行った作業者と撮影者のアンケートでは，どの程度ロボットの非言語行動に注意を払ったか，作業の円滑さなどがどう変わったかについて調べた．

このアンケート結果に対し Mann-Whitney の U 検定を行ったところ，「ロボットの非言語によるポインティング指示（ここ，あそこ）を理解できたか」という項目で有意差が生じた（p ¡ 0.05）．また，「ロボットの範囲の指示（この辺り，このへん）を理解することができたか」という項目では有意傾向が見られた（p = 0.068）．この理由は，半自律行動を行った場合の指さし動作がうまく補完されており，作業者が直感的に相手の指さしの対象を理解できたためであると考えられる．また，半自律行動を行わなかったセッションでは，作業者（飾り付けを行った人）3 名が「そのような指示はなかった」とチェックをつけていた．実際にこれらセッションを確認したところ，指示者は両腕をつかった指さしを行うなどかなりの数の指さしを行っていたが，方向のずれが原因となり，意味のある行動として認識されていないことが示唆された．

「初めに比べ指示者の存在をどの程度気にかけたか」という質問では，有意な傾向は見られなかったものの（p = 0.102），半自律行動無のセッションに比べ半自律行動を行う場合の方が

気にかけたと回答をする人が多く，ポインティングや余計な動きの抑制などが好印象であったのではないかと考えられる．

4. 議論

4.1 指さしの表出

指さしに関しては，方向の補正に加え明示的な指さし行為を表出することで，何を指さしているかを示すことを行った．その結果，共同作業した人のアンケートからわかるように，半自律行動によって正しい意図の表出を補助し，直感的な操作することが可能になったといえる．また，有意差は出なかったものの，指示者のアンケートにおいても，ロボットの操作を直感的に行えたかという質問に対し，半自律行動を生成したセッションの方が平均点が高いという結果が得られた．これらの結果から，半自律行動による動作補完は，インタラクションを円滑に行う上で有効な手段であったと考えられる．

ただし，誤認識や認識の遅れから，かえって円滑なコミュニケーションができなかったと感じた操作者もいた．指さし動作の認識率は Recall-rate，Precision ともに低いという問題があり，指さしの認識精度の向上は重要な課題である．特に，操作者側の身体動作として情報が大きく欠落するのが手形状で有り，手形状認識が可能であれば高精度化やほかのジェスチャの認識に大きく役立つと考えられる．ただし，画像情報を用いる場合背景画像に影響を多大に受けるので，深度画像を用いるなどの工夫が必要であり，今後の主要な課題と考えられる．

4.2 身体行動の抑制と他のジェスチャの表出

実験結果では，腕の動きの意図など，指さし以外のハンドジェスチャに関するアンケート項目で有意差は出なかった．しかし，実際に表出されたジェスチャの総数では半自律行動を行う方が多く，また大きなジェスチャの割合が少なかった．さらに，操作者の身体動作による指示があったにもかかわらず，存在を認識できなかったセッションが 3 つあった．これらの結果は，余計な身体行動を抑制することにより，意味のある動作を認識しやすくなるという仮説を支持していると考えられる．今後は様々な身体動作や実世界の認識を統合し，意味のある動作をあらかじめ特定していなくても，人間から見てノイズを分離できるような動作生成手法を実装する必要があると思われる．

4.3 頭部ジェスチャ

アンケート結果からは，頭部の動きに関しては，半自律行動の有無で有意差が見られなかった．これは，操作者の頭部方向の認識や，実世界のオブジェクト認識の精度が低かった点が原因であると考えられる．アンケート自由記述欄において，半自律行動が無かった参加者からは，ロボットの顔の動きがあったことに気付かなかったという意見があった一方で，半自律行動が有った参加者からは，頭部の動きに意味がありそうに感じたという意見をもらった．そのため，ある程度は動作補完が行われていたものの，タスクの文脈やほかの身体動作との関連を見いだすことができなかったのではないかと考えられる．

頭部ジェスチャには同意などの意味を示すうなずきや否定の首振りなど様々な機能を持つ [2] ため，対象への注視というバックチャネルの表出だけでなく，様々な種類の頭部ジェスチャを認識・表出させる必要があると考えられる．うなずきのような細かい動きは，ロボットによってそのまま再現することが難しいと考えられるため，今後はこういった動きも半自律的に補完していく必要がある．

5. まとめ

本研究では，ロボットによる実世界のセンシング情報と操作者の振る舞いを統合的に解釈し，認知的・身体的な齟齬を低減するように行動を補完する，半自律型テレプレゼンスシステムを構築することを目的とした．半自律行動として指さしと頭部方向の補正，バックチャネルの生成，状況に応じた身体表現の抑制という 3 つの動作に着目した．開発したシステムは直観的な認知，ロボットの操作を行うための操作環境である基盤制御部と実世界センシングを行う実世界センシング部，それらの情報を統合して行動生成を行う半自律行動部の 3 つからなる．

構築したシステムを用いてインタラクション実験を行い評価したところ，半自律型のシステムでは通常のシステムに対し，非言語行動，特に指さし動作の意図が理解しやすくなるという評価を得た．さらに操作者はロボットの可動範囲を気にせずに操作を行ったり，ロボットに合わせて大げさな動きを行わないという傾向が見られた．これは本研究のシステムが操作者とロボットの間に生じる認知的・身体的な齟齬を低減することができたことを示唆している．

今後の研究課題としては，まず，各センシング手法の精度向上があげられる．さらに，広範な身体動作を補完するためには，操作者の身体動作および実世界の情報の両面において，タスクを遂行する上で重要な情報を自動的に取捨選択するような枠組みが必要とされると考えられる．

参考文献

[1] Briere, S., Boissy, P. and Michaud, F.: In-home telehealth clinical interaction using a robot, Proceedings of the 4th ACM/IEEE international conference on Human robot interaction, HRI '09, pp. 225–226 (2009).

[2] Maynard・K・泉子: 会話分析, くろしお出版 (1993).

[3] Mizumoto, T., Nakadai, K., Yoshida, T., Takeda, R., Otsuka, T., Takahashi, T. and Okuno, H. G.: Design and implementation of selectable sound separation on the Texai telepresence system using HARK, Robotics and Automation (ICRA), 2011 IEEE International Conference on, pp. 2130–2137 (2011).

[4] Ohmoto, Y., Ohashi, H., Lala, D., Mori, S., Sakamoto, K., Kinoshita, K. and Nishida, T.: ICIE: immersive environment for social interaction based on socio-spacial information, Proceedings in Technologies and Applications of Artificial Intelligence (TAAI), 2011, pp 119–125 (2011).

[5] Tachi, S., Watanabe, K., Takeshita, K., Minamizawa, K., Yoshida, T. and Sato, K.: Mutual telexistence surrogate system: TELESAR4 - telexistence in real environments using autostereoscopic immersive display -, Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on, pp. 157–162 (2011).

[6] 大本義正, 大橋洋輝, 西田豊明: 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, 人工知能学会全国大会 (第 25 回), Vol. 1E1-4 (2011).