

無意識的な身ぶりを表出する テレプレゼンスロボットの提案

Telepresence Robot as Media for Expresses Unconscious Gestures

長谷川 孔明¹ 中内 靖¹

Komei HASEGAWA¹, Yasushi NAKAUCHI¹

¹筑波大学大学院 システム情報工学研究科

¹Graduate School of Systems and Information Engineering, University of Tsukuba

Abstract: It is known that robots can interact with humans in non-verbal ways. This aspect is utilized in telepresence robots. Telepresence robots allow the operators to express some of the non-verbal behavior such as eye contact, interpersonal distance, and gestures. And most of all telepresence robots are operated by a controller. It is supposed that humans receive meaning from not only conscious gestures but also unconscious gestures. But almost all gestures made by the controller are considered as operators' conscious gestures. It is difficult to express unconscious gestures via the controller. So we propose the telepresence robot which can express the operator's unconscious gestures by using motion capture.

1. はじめに

ロボットや擬人化エージェントの大きな特徴は、人と人が行う場合と同じように人とコミュニケーションを行うことができることである。パソコンを使用する際には必要となるキーボードやマウスといった専用のインタフェース機器は不要となり、目的のタスクを行わせるための命令や操作手順、エラーメッセージなどをいちいち覚える必要もなくなる。人が人に頼み事をする場合と同じようにロボットやエージェントに頼めば、人が受け答える場合と同様に返答を返して頼み事を行ってくれる。このようにロボットやエージェントが人と同じようにコミュニケーションを行える理由として、それらが身体を有しその身体によりノンバーバル情報を表出することが出来るという点があげられる。

ノンバーバル情報は表情や視線方向、対人距離、身ぶりといった非言語的な情報であり、対面のコミュニケーションにおいて言語的な情報であるバーバル情報よりも重要であるといわれている。そして、ノンバーバル情報が表出できるメディアであるというロボットの特徴を活かした研究としてテレプレゼンスロボットとよばれるものがある。

普段、遠隔地にいる人とコミュニケーションを行う手段として用いられるのは、メールや電話、ビデオチャットといったものである。これらのコミュニケーション手段には相手の顔が見えない、見たと



図 1 テレプレゼンスロボット

してもディスプレイ越しでは存在感が希薄になる、相手がどこを注視しているのか分かりづらいといったノンバーバル情報の欠如という問題点がある。テレプレゼンスロボットは欠如する存在感やノンバーバル情報をロボットの身体と動きを介して相手に伝

えることにより、あたかも遠隔地の人が目の前にいるかのようにコミュニケーションを行えるようにするロボットである。

テレプレゼンスロボットは、遠隔地の人のノンバーバル情報をそっくりそのまま伝えることが理想であると考えられる。しかし、ノンバーバル情報のひとつである身ぶりに着目してみると、既存のテレプレゼンスロボットはコントローラ操作によるものであり、操作者が意識的に表出しようとした身ぶりのみしか伝達できていないと考えられる。人と人との対面コミュニケーションの場面では、話し手が意識的に表出した身ぶりだけではなく、無意識的な身ぶりからも聞き手は情報を読み取っているということが示唆されている。そのため、操作者の無意識的な身ぶりが表出されない既存のテレプレゼンスロボットでは身ぶりの情報を十分に伝えきれていないと考えられる。

本研究では、無意識的な身ぶりを表出可能なテレプレゼンスロボットを提案する（図1参照）。そして、ロボットを介して表出する操作者の無意識的な身ぶりが対話者に与える印象を評価することにより、無意識的な身ぶりの機能を明らかにし、テレプレゼンスロボットが無意識的な身ぶりを表出することの重要性を示す。本研究により得られた知見は、ロボットや擬人化エージェントのより自然な身ぶり設計に役立つものと期待する。

2. 研究背景

既存のテレプレゼンスロボットに関する研究として、Paulosらはロボットに搭載したディスプレイに操作者の顔を表示させるPRoPと呼ばれるロボットを開発し、テレプレゼンスについて奉公している[1]。また、鈴木らは遠隔会議支援ロボットシステムを開発している[2]。これは卓上サイズの移動ロボットにパンチルトカメラを搭載した構成となっており、カメラのパンチルトと移動を組み合わせることで注意喚起能力を高められることを報告している。しかしながら、これらのテレプレゼンスロボットはノンバーバル情報である身体距離や視線方向は表出できるものの、人の腕にあたる機構を持たないため身ぶりを表出することは出来ない。

一方、Adalgerissonらの開発したMeBotと呼ばれるロボットは、卓上サイズで顔を表示するパンチルト可能なディスプレイを搭載し、身ぶりを表出するための腕の機構を有している[3]。しかしながら、MeBotの腕はこの腕と同じ機構を有したコントローラを操作することにより、コントローラと同じ腕の動きをMeBotも表出するようになっている。この操

作方法では、操作者が意識的に表出しようとした身ぶりのみしか表出されず、無意識的な身ぶりは伝達されないものと考えられる。

Cassellらは、人の対面コミュニケーションの場面において話し手が意識的に身ぶりを表出している場合でなくとも、聞き手は常に身ぶりから情報を得ていることを示唆している[4]。すなわち、話し手が無意識的に表出している身ぶりもコミュニケーション上において何かしらの機能を有していると考えられる。そのため、テレプレゼンスロボットも操作者の無意識的な身ぶりを表出する必要があると予想される。

また、対面でのコミュニケーションでは話し手が意識的か無意識的に関わらず、すべての身ぶりが聞き手へと受け取られるため、意識的な身ぶりのみが聞き手へと伝達されるような状況が起こるとは考えにくい。そのため、無意識的な身ぶり情報が欠落するという状況はテレプレゼンスロボットを介したコミュニケーション特有のものであると考えられる。このような特殊な状況下における無意識的な身ぶりの機能を明らかにすることにより、今後のテレプレゼンスロボットやそれ以外のコミュニケーションを目的としたロボット、擬人化エージェントの設計に役立つ知見を得られるものと期待される。

3. システム実装

3.1 無意識的な身ぶりの取得方法

前章で紹介した既存のテレプレゼンスロボットの身ぶり操作方法は、ロボットの腕と同じ機構を有したコントローラを手で操作するため、操作者が表出しようと思った意識的な身ぶりのみしか表出することができない。そのため、本研究では無意識的な身ぶりの取得方法としてモーションキャプチャを用いることとした。モーションキャプチャは人や物体の動きを記録する技術である。人の動作それ自体を取得することが可能なため、取得できるデータは人が意識的に動いたかどうかには依存しない。そのためモーションキャプチャを利用することにより意識的な身ぶりに加えて無意識的な身ぶりの情報も取得することが可能である。

3.2 システム構成

実装したテレプレゼンスロボットのシステム構成を図2に示す。

操作インタフェースにより取得された操作者の音声、表情、身ぶりといった情報は対話者側へと送信され、テレプレゼンスロボットを介して対話者へと表

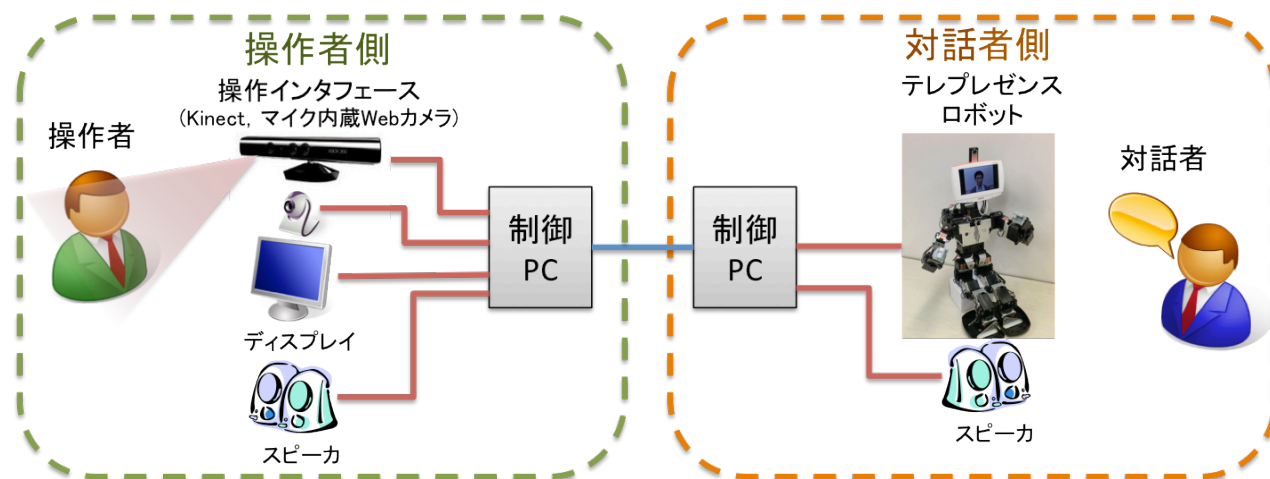


図 2 システム構成

出される。対話者側の情報はテレプレゼンスロボットにて取得されて操作者側へと送信され、ディスプレイ、スピーカから操作者へ表出される。次項において、より詳細な説明を行う。

3.3 操作インターフェイス

操作者の身ぶりを取得するモーションキャプチャ用デバイスとしてMicrosoft社のKinectを用いる。

KinectはRGBカメラ、深度センサー、マルチアレイマイクロフォンを有しており、カラー画像、距離画像、音声を取得することが可能である。また、同社のソフトウェア開発キット Kinect for Windows SDK を用いることにより人物のボーンを認識し、体の動きを各関節や手先位置などの3次元情報として取得することが可能である。これにより操作者の体による身ぶりを取得することができる。

また、KinectとKinect for Windows SDKを用いただけでは頭部の動きを取得することが出来ない。そのため、頭部の動きの取得には別途 Web カメラと Seeing Machines 社の faceAPI を用いる。faceAPI は Web カメラの画像を元に顔認識と3次元顔追跡が可能である。これによりリアルタイムに操作者の頭部の動きを取得する。

3.4 テレプレゼンスロボット

本研究で提案するテレプレゼンスロボットの外観を図3に示す。

ロボットのプラットフォームとして近藤科学株式会社製ヒューマノイドロボットKHR-3HVを使用する。本研究で使用するKHR-3HVはサーボモータの追加を行い、頭部3自由度、腕部各4自由度、腰部1自由度、脚部各4自由度の動作が行えるようになっている。この自由度の追加により、うなずきや首を傾げるとい

った動作、腕を組むなどといったより複雑な操作者の身ぶりを模倣し表出できるようになっている。

テレプレゼンスロボットの頭部には小型 USB ディスプレイ Mimo um-430 を取り付けている。このディスプレイには操作者の顔の映像を映し表情を対話者へと表出する。

音声の伝達はテレプレゼンスロボット制御用 PC のスピーカを用いて行う。

テレプレゼンスロボットの頭部、小型 USB ディスプレイの上部にはマイク内蔵 USB カメラ LOAS MCM-15BK が取り付けられており、これにより操作者側に表示するための対話者側の映像・音声情報の取得を行う。取得された映像、音声は操作者側のディスプレイとスピーカから提示される。

3.5 ソフトウェア

テレプレゼンスロボットを制御するソフトウェアの構成を図4に示す。

操作インターフェイスから取得された情報をもとにしてテレプレゼンスロボットにより身ぶりが表出されるまでの処理の流れを説明する。

まず、操作インターフェイスの1つであるKinectにより操作者の全身を含む距離画像が取得され、Kinect for Windows SDKと組み合わせることにより操作者の体の動きのトラッキングデータを得る。それと同時に、もう1つの操作インターフェイスであるWebカメラから操作者の顔を含む画像を取得し、faceAPIにより顔の動きのトラッキングデータを得る。得られたトラッキングデータはSkypeAPIを利用して対話者側の制御PCへと送信される。

対話者側の制御PCでは、SkypeAPIによりトラッキングデータを受信し、そのデータをKHR制御プログ



図 4 ロボットのハード構成

ラムへと入力する。KHR 制御プログラムでは、3次元の関節位置座標であるトラッキングデータをもとに、身ぶりがトラッキングデータと一致するようにKHR-3HVの各サーボモータの回転目標角度を計算する。その後、計算した目標角度をサーボモータ動作命令としてシリアル通信によりKHR-3HVの制御マイコンへと送信する。こうして、操作インタフェースを介して得られた操作者の身ぶり情報にあわせてテレプレゼンスロボットが身ぶりを表出する。

また、操作者と対話者間の音声と映像のやり取りはSkypeのビデオチャットを利用して行う。その際、テレプレゼンスロボットの操作者の映像はロボットの頭部に取り付けた小型ディスプレイに表示され、対話者の映像は操作者側のディスプレイに表示される。

4. 仮説と実験方法

この章では、本研究で提案するテレプレゼンスロボットの表出する無意識的な身ぶりが持つコミュニケーション上の機能に対する仮説と、それを確認するために予定している実験方法についての説明を行う。

4.1 無意識的な身ぶりの持つ機能の仮説

喜多によるとコミュニケーションにおける身ぶりの機能は大きく以下の3つに分類できる[5]。

- A. 伝達内容の表現
- B. 会話ターンの管理
- C. 親近感の形成

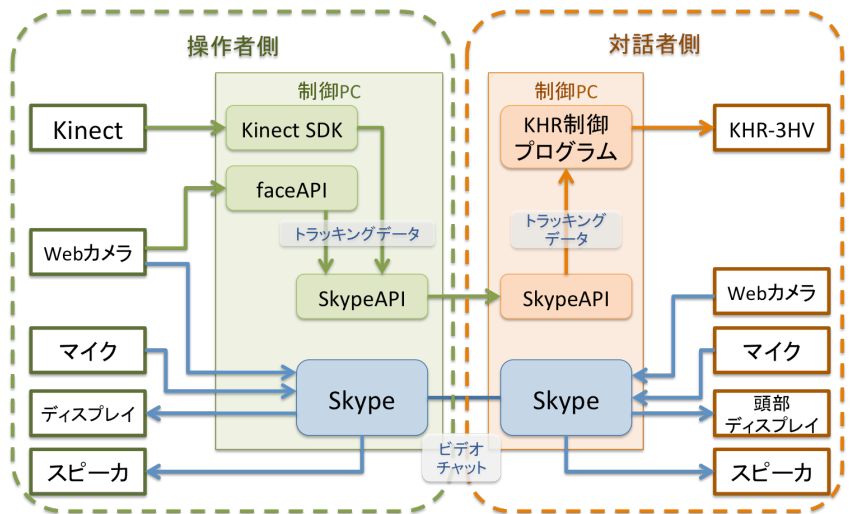


図 3 ソフトウェアの構成

まず、Aの伝達内容の表現は、相手に伝えたい内容そのものを表す機能のある身ぶりである。人差し指と親指で丸を作るOKサインや、親指だけを立てるthumbs upなどの形と意味が社会的な取り決めにより決められている「エンブレム」と呼ばれる身ぶりがAに含まれる。その他には、物体の動きや大きさを表す「描写的ジェスチャー」、指差しに代表される時空間的隣接性を表す「直示的ジェスチャー」が含まれる。

Bの会話ターンの管理は、相手に話す順番を渡すつもりがあるかないかを伝える機能を持つ身ぶりである。例えば、言い淀んでいる時に身ぶりで伝えたいことを表し相手の助けを促したり、発話をする前に行われる準備動作から新たな発話が行われるかどうかを先読みしたりするものなどがBに含まれる。

Cの親近感の形成は、コミュニケーションの参加者と一体感のようなものを生む機能を持つ身ぶりである。話し手と聞き手が同時にうなずいたり、同じ姿勢や身ぶりを行ったりする同調現象がCに含まれる。

会話を行う際に、話し手と聞き手の意識は主に会話内容に向けられると考えられる。そして、話し手と聞き手の意識が向きやすい身ぶりとは、会話内容と関連の深い身ぶりであることが期待される。上で機能別に分けた身ぶりのうち、発話内容に最も関連深いのはA伝達内容の表現である。このことから、本研究では無意識的な身ぶりの持つ機能として以下の仮説が成り立つものと期待する。

仮説：無意識的な身ぶりは主に会話ターンの管理と親近感の形成の機能を持つ

4.2 実験方法

前項で立てた無意識的身ぶりの持つ機能の仮説を検証するために予定している実験についての説明を行う。

本研究で提案するテレプレゼンスロボットとの比較対象として、MeBotを参考にして構成したロボットを用意する[3]。

この比較対象のロボットは操作インタフェースをモーションキャプチャからコントローラに変更したものである。コントローラには、テレプレゼンスロボットのロボットプラットフォームとして用いたKHR-3HVと同じ自由度追加を行ったもう一台のKHR-3HVを用いる。コントローラとして用いる方にはサーボモータにトルクをかけず手で簡単に動かせる状態にしておく。操作者の手により動かされたコントローラの各サーボモータの回転角度情報を取得し、遠隔地にあるテレプレゼンスロボットへと送る。ロボットは受け取った角度同じ角度へとサーボモータを動作させることにより、コントローラで入力した身ぶりをそのままテレプレゼンスロボット上で表出することができる。

実験に際し、操作インタフェースとしてモーションキャプチャを利用したテレプレゼンスロボットシステムをモーションキャプチャ操作条件とし、比較対象として用意した操作インタフェースにコントローラを利用するシステムをコントローラ操作条件とする。コントローラ操作条件は意識的身ぶりのみが表出可能な状態であり、モーションキャプチャ操作条件は意識的身ぶりに加え無意識的身ぶりも表出可能な状態であると仮定できる。この2つの条件それぞれを用いて対話タスクを行わせる実験を行うことにより、前項で立てた仮説が支持されるかどうかの検証を行っていく。

実験のための対話タスクについては、どちらか片方の人が話し続ける一方向タスクではなく、どちらも話をする機会が与えられ話し手と聞き手が頻繁に入れ替わる双方向タスクを選択することにより、無意識的身ぶりが会話ターンの管理を行う機能を持つかどうか確認しやすくなると考える。

また、親近感の形成の機能を持つかどうかの確認が行いやすくなるよう、双方が異なる意見を言い合い、最終的に相手の意見を尊重する場面や自分の意見を貫く場面などが現れる対話タスクが向いていると考える。

この点を考慮し、実験には砂漠生き残り問題を対話タスクとして利用する予定である。砂漠生き残り問題は、砂漠という過酷な環境に取り残された状況を想定し、リストアップされた道具の中から救助が

来るまで生き残り続けるために必要性の高い物の順に優先順位をつけていくというものである。これは双方がその道具を選んだ理由を相手へと伝える必要がある双方向タスクであり、何度も話し手と聞き手が入れ替わる。これにより会話ターンの管理が身ぶりにより行われているか確認することができる。また、道具に優先順位を付ける際に、相手の意見を信頼するか、尊重するかなどの観点から相手への親近感が形成されているかどうか確認できると考える。

また、提案するテレプレゼンスロボットを実装し実際に身ぶりを表出させた際に、操作者が表出する身ぶりとロボットが表出する身ぶりとの明らかな遅延時間が確認された。これは、モーションキャプチャの処理時間、ネットワーク越しのデータの送信時間、サーボモータが命令を受け取り目標角度まで達するまでの時間それぞれが加算された遅延であると考えられる。実験の際には、なるべく遅延が短くなるようシステムの構成を工夫つもりであるが、考察の際にはこの遅延時間が対話に与える影響も考慮する必要があると考える。

5. まとめと今後の展望

本研究では、モーションキャプチャの利用による無意識的身ぶり表出が可能なテレプレゼンスロボットを提案した。

既存のテレプレゼンスロボットの操作方法であるコントローラ操作では、操作者が意識的に表出しようとした身ぶりのみしか伝達できないと予想される。しかしながら、人と人との対面コミュニケーションの場面では、話し手が意識的に表出した身ぶりだけではなく、無意識的身ぶりからも聞き手は情報を読み取っているということが示唆されている。すなわち、話し手が無意識的に表出している身ぶりもコミュニケーション上において何かしらの機能を有していると考えられる。そのため、操作者の無意識的身ぶりが表出されない既存のテレプレゼンスロボットでは身ぶりの情報を十分に伝えきれておらず、十分な情報を伝えるためには操作者の無意識的身ぶりを表出が可能な操作方法を採用する必要があると考える。

また、対面でのコミュニケーションでは話し手が意識的か無意識的に関わらず、すべての身ぶりが聞き手へと受け取られる。そのため、対面でのコミュニケーションにおいては意識的身ぶりのみが聞き手へと伝達されるような状況が起こるとは考えにくく、無意識的身ぶりの情報が欠落するという状況はテレプレゼンスロボットを介したコミュニケーシ

ョン特有のものであると考えられる。そこで、意識的身ぶりのみを表出可能であるコントローラ操作条件と、無意識的身ぶりも加えて表出可能であるモーションキャプチャ操作条件の2種類の操作インタフェースのテレプレゼンスロボットを用いて比較実験を行う予定である。これにより無意識的身ぶりの持つコミュニケーション上の機能を明らかにすることができ、その知見が今後のテレプレゼンスロボットやそれ以外のコミュニケーションを目的としたロボット、擬人化エージェントの設計に役立てられることを期待している。

参考文献

- [1] E. Paulos, J. Canny: Social Tele-embodiment: Understanding Presence, Autonomous Robots, Vol. 11, No. 1, pp. 87-95, (2001)
- [2] 鈴木 雄介, 福島 寛之, 深澤 伸一, 竹内 晃一: 遠隔会議支援ロボットシステムの注意喚起能力評価, 情報処理学会論文誌, Vol. 51, No. 1, pp. 25-35,(2010)
- [3] S. O. Adalgerisson, C. Breazeal: MeBot: A Robotic Platform for Socially Embodied Telepresence, HRI2010, pp. 15-22, (2010)
- [4] J. Cassell, D. McNeill, K. E. McCullough: Speech-gesture mismatches: Evidence for one underlying representation of linguistic and nonlinguistic information, Pragmatics and Cognition, Vol. 7, No. 1, pp.1-33, (1999)
- [5] 齋藤 洋典, 喜多 壮太郎 編著: 認知科学の探究ジェスチャー・行為・意味, 共立出版, (2002)