

客観視点によるアバタ操作者の空間認知

Spatial Cognition by Avatar Controller with an Objective Viewpoint

ネットワーク情報学部 石井健太郎

School of Network and Information Kentaro ISHII

Keywords: telecommunication, avatar, viewpoint

Abstract

This paper discusses the findings of experiments on spatial cognition by an avatar-controlling user of telecommunication systems. With the presented telecommunication systems, the avatar-controlling user has an objective viewpoint to see the remote environment. However, sometimes, the words of an avatar-controlling user may not be intuitive for an avatar-viewing user, because the avatar-controlling user does not necessarily share the viewpoint of the avatar. We conducted a series of experiments to investigate how user interface or avatar implementation can mitigate this mismatch of the avatar-viewing and avatar-controlling users. The results showed that a refined user interface and a physical robot brought more intuitive utterances for the avatar-viewing user than an original projection-based system.

1. はじめに

遠隔地にいる他者と対話する際に、相手の環境に自分の分身となるアバタを配置することによって、対話主体が明確になることやキャラクターの表出がしやすくなることにより、対話のきっかけをつかみやすい効果が期待できる。また、アバタの身体表現を用いることによって、相手の環境内の物体や位置を指し示すことができる。このように、電話に代表される音声対話に加えて、遠隔コミュニケーションにさらなる効果をもたらすことが期待できる(図1)。本研究は、このような遠隔の話者が自分の分身となるアバタを介して対話する環境について扱う。

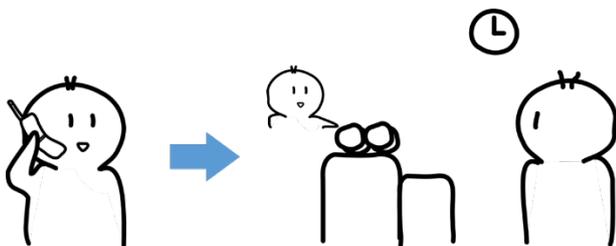


図1 アバタを用いた遠隔コミュニケーション

一方、対話主体が明確になることにより、対話相手は部分的に対面での対話のルールを暗黙的に期待するようになる。例えば、先行研究[1]のように、投影映像の位置から音

声が聞こえるアバタ遠隔コミュニケーションシステムを用いた場合、話し手と聞き手と対象物体との位置関係によって指示語を使いわけられることを、アバタとの対話者(以後、単に対話者と呼ぶ。)は自然と期待するようになる(図2)。言い換えると、遠隔地からアバタを操作する者(以後、単に操作者と呼ぶ。)にとって、アバタの視点を採用して発話することを期待するようになる。

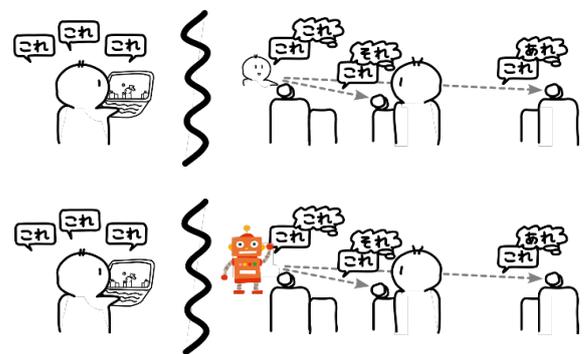


図2 アバタとの対話者が期待する指示語とアバタの操作者が使用しがちな指示語の不一致. 3節では投影型アバタを用いた解決を試み、4節では人間型ロボットを用いた解決を試みる。

しかし、2節で述べるように、操作者にとって、アバタの視点を採用して発話することが難しいことが示されてい

る。操作者は対話者とアバタの位置関係を見捨て、「これ」という指示語を発話しやすい(図2)。これは、操作者が自身のアバタの位置を把握しにくいことが一因であると考えられる。本論文では、3節にてアバタの位置を強調するユーザインタフェースによる方法、4節にて実体を伴うロボットをアバタとして用いる方法をそれぞれ導入し、位置関係を見捨てた発話をしやすい問題の解決を試みる。

2. 対話者と操作者の視点の不一致

本節では、本論文の提案手法が必要とされる状況を示す遠隔コミュニケーションの実験について述べる。この実験を、ベースライン実験と呼ぶこととする。ベースライン実験では、図2に示したとおり、対話者が期待する発話と操作者が実際に行う発話が一致しないことが明らかにされている[1]。

2.1. アバタコミュニケーションの装置

実験には、図3に示す遠隔コミュニケーションシステム PROT AVATAR を用いる。PROT AVATAR は、カメラでキャプチャした自分の顔に、身体アニメーションを付与して作成したアバタを投影するシステムである。また、音声は超音波変調スピーカから射出され、投影されているアバタの位置から聞こえてくるように感じることができる[2]。したがって、図1のような本研究が対象とする遠隔コミュニケーションを実現できる。さらに、PROT AVATAR は、対話者にとって適切ではない平坦ではない背景や単色ではない背景を、アバタの投影場所としては回避して移動する機能を備えている。

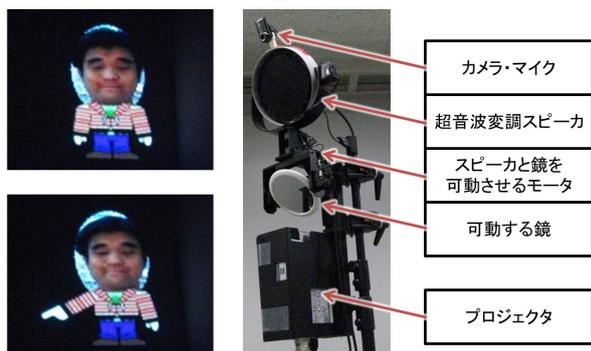


図3 PROT AVATAR. プロジェクタ・可動する鏡・超音波変調スピーカを利用した遠隔コミュニケーションシステム。

PROT AVATAR は、操作者側と対話者側のシステムからなる(図4)。操作者側のハードウェアは、制御用コンピュータとカメラ・マイク・スピーカによって構成されている(図5)。カメラ画像から操作者の顔を認識して切り出し、

対話者側に投影されるアバタを作成するするとともに、マイクにより操作者の音声をキャプチャし、スピーカにより対話者の音声を出力する。一方、対話者側のハードウェアは、制御用コンピュータと PROT・カメラ・マイクによって構成される。PROT は、可動式の鏡を取り付けたプロジェクタと可動式の超音波変調スピーカからなる映像・音声の投射システムである[3]。カメラは対話者側の部屋を撮影するものであり、その映像を操作者は画面上で見ることができる。また、対話者の音声をマイクでキャプチャし、操作者側のシステムに送信して出力する。

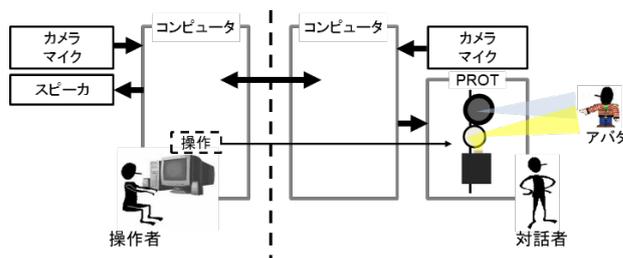


図4 投影型アバタシステムの構成。操作者側と対話者側のシステムはネットワーク接続されている。操作者と対話者は互いに対話することができ、操作者は遠隔地に投影されているアバタを操作することができる。

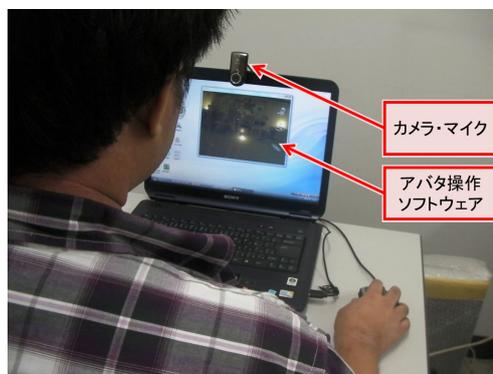


図5 操作者側のハードウェア

操作者側システムからモータ制御情報が対話者側システムに送られると、対話者側システムは PROT のモータを動かして、アバタ映像と音声の投射位置を変更する。また、同様に操作者側システムから、アバタの体部分の画像を変化させ、指差しのジェスチャを行うことができる。本研究では、基本的な操作の方法として、操作者が部屋の映像上の物体をクリックすることによりアバタを操作することを考える。これは、操作者が物体を指し示す場合に通常先に意識するのは、アバタをどう移動させるかではなく、どの物体を指し示すかであるためである。

2.2. 実験手順

実験参加者は、はじめに対話者側の部屋に案内され、実験者が自身のアバタを操作して行われるシステムの紹介デ

モを体験する。紹介デモの際に、実験者からアバタを介して実験参加者へいくつか質問をたずねる。このことにより、実験参加者は対話者としてアバタコミュニケーションシステムによる対話を体験する。その後、実験参加者は別の実験者によって操作者側の部屋に案内される。

操作者側の部屋への案内後、実験参加者にアバタ操作手法についての説明がなされる。実験参加者は対話者側の部屋の様子を画面上で観察できるようになっており、操作手法の説明終了後、しばらくアバタ操作の練習を行う。十分な練習のあと、実験者は実験参加者に、以下の実験で行うタスクを説明する。実験で行うタスクは、「これ」・「それ」・「あれ」のいずれかの指示語を用いながら説明を受けた手法でアバタを操作して、対話者側の部屋にいる実験者にライトの色の変更指示を出すというものである。対話者側の実験者は実験参加者の指示にしたがい、リモコンで部屋にあるライトの色を変更する。部屋には3つのライトが配置されており、色の変更のタスクを3つのライトすべてに対して行う。

日本語では、発話者の近くにあるものを示すときに「これ」、発話相手の近くにあるものを示すときに「それ」、両者から遠くにあるものを示すときに「あれ」という指示語が用いられる。対象となるライトとアバタの位置関係は、図6(a),(b),(c)に示すとおり、対話者にとってはそれぞれ「これ」・「それ」・「あれ」を用いるのが自然と考えられる位置関係になっており、以降それぞれを「これ」条件・「それ」条件・「あれ」条件と呼ぶ。

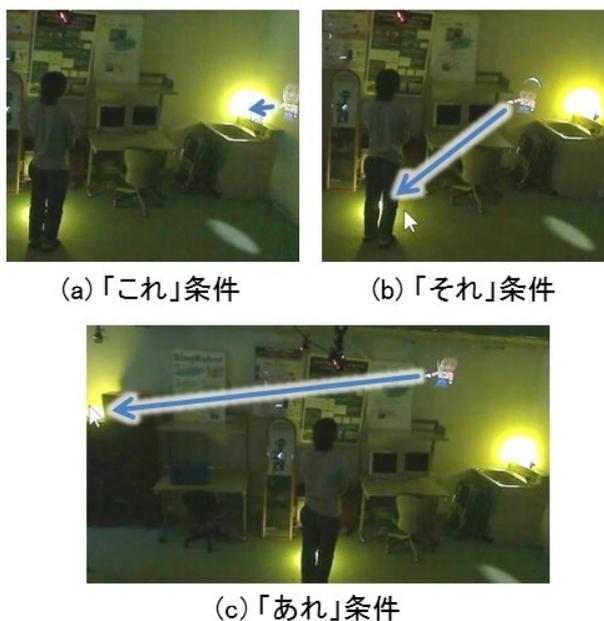


図6 投影型アバタと対象物体の位置。図中の矢印は、アバタの位置から指し示す物体の位置まで引かれている。条件ごとに、アバタと指し示す物体の距離・対話者と物体の距離を変えている。

操作者自身の視点では、3つのどの条件においても「これ」という指示語を使うのが最も自然である。なぜならば、画面上のどの場所も操作者からは十分に近い位置であるためである。そのため、実験参加者が「これ」という指示語を使うか、各条件に応じて対話者にとって自然な指示語を選んで使うかを調べることによって、どの視点を採用しているかを解析することを意図している。具体的には、「これ」条件で「これ」という指示語を用いたかその他の指示語を用いたか、「あれ」条件で「あれ」という指示語を用いたかその他の指示語を用いたか、「それ」条件で「それ」という指示語を用いたかその他の指示語を用いたかを調べる。

2.3. 結果と考察

日本人の18名を実験参加者として実験を行った。指示対象となるライトや色の変更順は、学習効果による影響を抑えるためにカウンターバランスをとって実験を行った。

図7に、実験の結果として指示語の発話人数を示す。それぞれの列は条件ごとに18名中何名がその指示語を発話したかを表している。影付きのセルは対話者にとって最も自然な指示語を示している。実験参加者は、「あれ」条件・「それ」条件において、自然な指示語ではなく「これ」という指示語を選択しやすい傾向がうかがえる。

発話	「これ」条件	「あれ」条件	「それ」条件
これ	18	14	13
それ	0	4	1
あれ	0	0	4

図7 ベースライン実験の結果

この結果は、ベースライン実験の実験参加者が、アバタの視点に立って発話できていないことを示唆するものであり、次節以降では、このことについての改善を試みる。人間は、対話の際に視点を自由に变化できることがこれまでも示されている。例えば、Ullmer-Ehrichの実験では、発話者が自分の部屋の構成を説明するときに、ドアから部屋の内側に向かう視点を採用することが示されている[4]。Kleinの実験では、発話者が対話相手に目的地までの道順を教える際に、その経路にそって視点を变化させていくことが示されている[5]。Imaiらの実験では、人間がロボットを持ち上げて移動させる際に、人間がロボットの視点を採用していることが示されている[6]。このような視点の変化を引き起こすことができれば、不自然な発話は軽減されていくものと考えられる。

3. 投影型アバタを用いた実験

不自然な発話を軽減する第1の方策は、ユーザインタフ

エースの変更によるものである。ベースライン実験と同様の装置と手順を用いて、操作者側のアバタ操作ソフトウェアのユーザインタフェースを変更した場合に、結果が異なるかを検証する[1]。

3.1. アバタコミュニケーションの装置

ハードウェアはベースライン実験と同一のものを用いる。PROT AVATARには、対話者にとって適切ではない平坦ではない背景や単色ではない背景を、アバタの投影場所としては回避して移動する位置を計算する機能を備えているが、この移動位置の取り扱いについて、アバタ操作ソフトウェアには変更を加えた。ベースライン実験では、操作者が指し示したい画面内の物体をクリックすると、システムは直ちに自動的にアバタが移動すべき位置までアバタを移動させる。これを、自動操作手法と呼ぶこととする。本実験では、操作者が指し示したい画面内の物体をクリックすると、システムはまずアバタが移動すべき位置を画面上に赤丸で提示し、次に操作者が提示された赤丸内をもう1度クリックするとアバタはその場所に移動させる。これを、半自動操作手法と呼ぶこととする。図8に、以上の変更を示す。



図8 アバタ操作手法の変更。ベースライン実験の自動操作手法では、操作者が対話者の足元にあるライトをクリックするとアバタは移動すべき位置として計算された壁面に移動する。本実験の半自動操作手法では、操作者が対話者の足元にあるライトをクリックするとシステムはアバタが移動すべき位置として計算された壁面に赤丸を提示し、操作者が提示された赤丸内をさらにクリックするとアバタはその場所に移動する。

3.2. 実験手順

基本的な実験手順はベースライン実験と同一である。操作者側の部屋への案内後のアバタ操作手法の説明のみ半自動操作手法の説明に置き替えて実施する。

3.3. 結果と考察

日本人の18名を実験参加者として実験を行った。指示対象となるライトや色の変更順は、学習効果による影響を抑えるためにカウンターバランスをとって実験を行った。

図9上部に、本実験の半自動操作手法における指示語の発話人数を、ベースライン実験の自動操作手法における結果を転記したものと示す。それぞれの列は条件ごとに18名中何名がその指示語を発話したかを表している。影付きのセルは対話者にとって最も自然な発話を示している。図9下部は、対話者にとって自然な指示語を用いたかその他の指示語を用いたかの集計であり、「それ」条件・「あれ」条件において、ベースライン実験の自動操作手法よりも本実験の半自動操作手法を用いた場合に、対話者にとって自然な指示語を用いた実験参加者の人数が多くなった。

発話	「これ」条件		「それ」条件		「あれ」条件	
	半自動	自動	半自動	自動	半自動	自動
これ	17	18	4	14	6	13
それ	1	0	11	4	2	1
あれ	0	0	3	0	10	4

	「それ」条件		「あれ」条件	
	半自動	自動	半自動	自動
	4	14	6	13
	11	4	2	1
	3	0	10	4

	「それ」条件		「あれ」条件	
	半自動	自動	半自動	自動
自然な指示語	11	4	10	4
別の指示語	7	14	8	14

($p < 0.05$) ($p < 0.1$)

図9 投影型アバタを用いた実験の結果

「これ」条件・「それ」条件・「あれ」条件において、対話者にとって自然な指示語を用いた人数の割合をフィッシャーの正確確率検定により比較したところ、「それ」条件において有意差 ($p < 0.05$)・「あれ」条件において有意傾向 ($p < 0.1$)が見受けられた。「これ」条件においては、統計学的差異は見受けられなかった。この結果は、半自動操作手法を用いた実験参加者のほうが対話者にとって自然な指示語を用いることを示している。

半自動操作手法において操作者がアバタの視点を採用しやすいこと理由は、今後実験を重ねて明らかにしなければならない。1つの可能性は、半自動操作手法でアバタの移動位置をクリックさせることにより、明示的にアバタの位置を意識することである。

操作者にとっては、半自動操作手法は自動操作手法に比べて1クリック多く操作を必要とする手法である。ただし、

どの位置に投影すると対話者に提示しやすいかはシステムが計算するため、手動で位置を決定するよりも手間は少ない。

4. 人間型ロボットを用いた実験

不自然な発話を軽減する第2の方策は、アバタ装置を変更して人間型ロボットを用いるものである。ロボットの指差しには物理的な身体動作を伴うため、自然と操作者がアバタの位置を考慮することにはならないかというのが、本設で述べる実験の問いである。GestureMan [7]・Geminoid [8]・駅における実験[9]のような、遠隔から操作されるロボットを用いることにより、遠隔地における方向を指し示したり対話相手に物理的な存在感を感じさせたりすることが研究されている。また、Nakanishiらは、遠隔地にいる対話者の移動に応じて、カメラが動くこと[10]やディスプレイが動くこと[11]によって、対話者の存在感が増すことを示した。これらのような効果があらわれれば、ロボットを用いることで操作者が適切な視点を採用することが期待できる[12]。

4.1. アバタコミュニケーションの装置

以上の問いについて調査するため、ロボットを用いた対話システムを用意する。基本的には、ベースライン実験で使用していたPROT AVATARを人間型ロボット nao に置き替えた構成である(図10)。操作者側のハードウェアは、制御用のラップトップコンピュータとラップトップコンピュータに搭載されているマイク・スピーカによって構成されている。操作者側のシステムは、マイクにより操作者の音声をキャプチャし、スピーカにより対話者の音声を出力する。一方、対話者側のハードウェアは、制御用コンピュータと人間型ロボット nao・カメラ・マイク・スピーカによって構成される。マイク・スピーカはロボット後部にとりつけてあり、ロボットの位置の音声をキャプチャし、ロボットの位置より操作者の音声を出力する。カメラは対話者側の部屋を撮影するものであり、その映像を操作者は画面上で見ることができる。

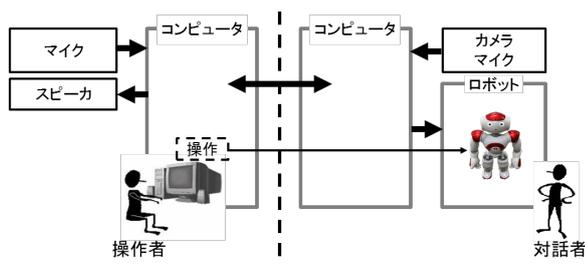


図10 人間型ロボットシステムの構成。操作者側と対話者側のシステムはネットワーク接続されている。操作者と対話者は互いに対話することができる。操作者は遠隔地のロボットを操作することができる。

ベースライン実験の設計と同様に、本実験では、操作者が部屋の映像上の物体をクリックすることによりロボットを操作する。これは、操作者が物体を指し示す場合に通常先に意識するのは、ロボットをどう動かすかではなく、どの物体を指し示すかであるためである。操作者側システムから指差し位置が対話者側システムに送られると、対話者側システムはロボット座標系の指差し位置を計算し、指差しを行うと同時に視線誘導を行う。つまり、ロボットはどの方向に指を向けるべきかという計算はシステムによって自動的に達成される。

4.2. 実験手順

基本的な実験手順はベースライン実験と同一である。操作者側の部屋への案内後のアバタ操作手法の説明はロボットアバタの操作手法の説明に置き替えて実施する。ただし、クリックの手順をはじめ、ユーザインタフェースはベースライン実験と同一である。

また、本実験は、同一の実施環境が用意できなかった都合で、ベースライン実験とは異なる部屋で実施し、対話者が行うタスクをライトの色の変更指示から写真に写っているものを答えてもらう課題に変更している。対象となる写真とロボットの位置関係は、図11(a),(b),(c)に示すとおり、対話者にとってはそれぞれ「これ」・「それ」・「あれ」を用いるのが自然と考えられる位置関係になっており、以降それぞれを「これ」条件・「それ」条件・「あれ」条件と呼ぶ。

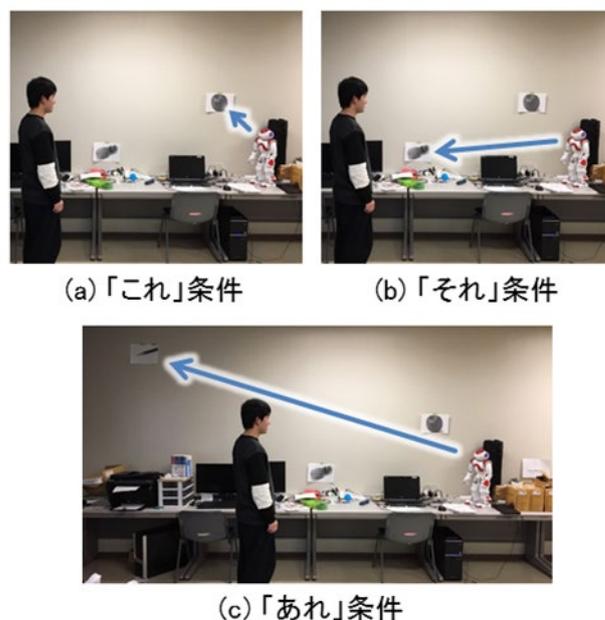


図11 人間型ロボットと対象写真の位置。図中の矢印は、ロボットの位置から指し示す写真の位置まで引かれている。条件ごとに、ロボットと指し示す写真の距離・対話者と写真の距離を変えている。

- [5] Klein, W.: Local Deixis in Route Directions, Speech, Place and Action, pp.161-182, (1982).
- [6] Imai, M., Hiraki, K., Miyasato, T., Nakatsu, R., and Anzai, Y.: Interaction With Robots: Physical Constraints on the Interpretation of Demonstrative Pronouns, *International Journal of Human-Computer Interaction*, Vol.16, No.2, pp.367-384, (2003).
- [7] Kuzuoka, H., Oyama, S., Yamazaki, K., Suzuki, K., and Mitsuishi, M.: GestureMan: A Mobile Robot that Embodies a Remote Instructor's Actions, *ACM Conference on Computer Supported Cooperative Work*, pp.155-162, (2000).
- [8] Ishiguro, H., and Nishio, S.: Building artificial humans to understand humans, *Journal of Artificial Organs*, Vol.10, No.3, pp.133-142, (2007).
- [9] Shiomi, M., Sakamoto, D., Kanda, T., Ishi, C.T., Ishiguro, H., and Hagita, N.: A Semi-autonomous Communication Robot: A Field Trial at a Train Station, *ACM/IEEE International Conference on Human-Robot Interaction*, pp.303-310, (2008).
- [10] Nakanishi, H., Murakami, Y., Kato, K.: Movable Cameras Enhance Social Telepresence in Media Spaces. *ACM SIGCHI Conference on Human Factors in Computing Systems*, pp.433-442, (2009).
- [11] Nakanishi, H., Kato, K., Ishiguro, H.: Zoom Cameras and Movable Displays Enhance Social Telepresence, *ACM SIGCHI Conference on Human Factors in Computing Systems*, pp.63-72, (2011).
- [12] 石井健太郎: アバタを用いた遠隔コミュニケーションにおけるアバタ操作者の空間認知, *マルチメディア, 分散, 協調とモバイルシンポジウム 2018 論文集*, pp.1795-1800, (2018).