

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

石井健太郎¹

概要： 本論文では、アバタを用いた遠隔コミュニケーションにおいて、実体を伴うロボットをアバタとして用いる場合と投影映像をアバタとして用いる場合の比較に基づき、アバタ操作者の空間認知について議論する。先行研究により、投影映像をアバタに用いる場合に、アバタ操作者が対話相手とアバタの位置関係を無視した発話をしやすいことが示されている。具体的には、日本語の指示語「これ」・「それ」・「あれ」は、本来は発話者と聞き手の位置関係によって用いられる語が変わるが、アバタ操作者はいずれの場合においても「これ」と発話しやすいことが示されている。原因として、アバタ操作者がアバタの位置を把握しにくいことが考えられる。そこで、物理的な身体動作を伴うことにより、空間中の位置を意識しやすいロボットをアバタとして用いることを考える。本研究では、あらたに実体を伴うロボットによる実験を行った。その結果、一部の条件において、投影映像を用いた場合よりもロボットを用いた場合のほうが、多くのアバタ操作者が自然な発話を行ったことが示された。

Spatial Cognition by Avatar Controller in Avatar-Mediated Telecommunication

KENTARO ISHII¹

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)。本研究では、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。

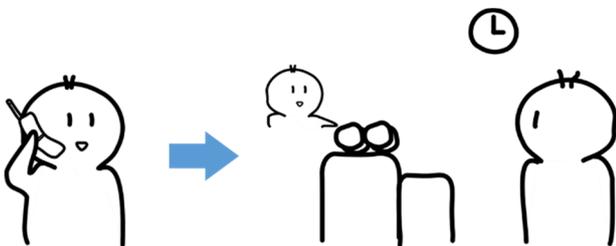


図1 アバタを用いた遠隔コミュニケーション

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。

る。例えば、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、聞き手は話し手に期待するようになる。

Ishii らの研究によると、投影映像をアバタに用いる遠隔コミュニケーションシステムを用いると、上記の指示語の使いわけをアバタとの対話者(以後、単に対話者と呼ぶ。)は自然と期待することが示されていると同時に、遠隔地からアバタを操作する者(以後、単に操作者)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[1]。2節で詳述するように、操作者は対話者とアバタの位置関係を無視して、「これ」という指示語を発話しやすいことが示されている(図2)。

この研究では、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察しており、実際にアバタの位置を強調するユーザインタフェースを導入することにより、位置関係を無視した発話が抑制されることを示している。それでは、アバタの位置を強調する仕組みとして、投影映像ではなく実体を伴うロボットをアバタとして用いた場合にはどうか。投影型アバタのアニメーションではなく、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だろうか。Gesture Man [2]・Geminoid [3]・駅における実験[4]のような、遠隔から操作されるロボットを用いるこ

¹ 専修大学
Senshu University

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[5]やディスプレイが動くこと[6]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。本研究では、あらたに実体を伴うロボットをアバタとして用いるシステムによる実験を行い比較することにより、以上の課題に関する知見を明らかにすることを目的とする(図2)。

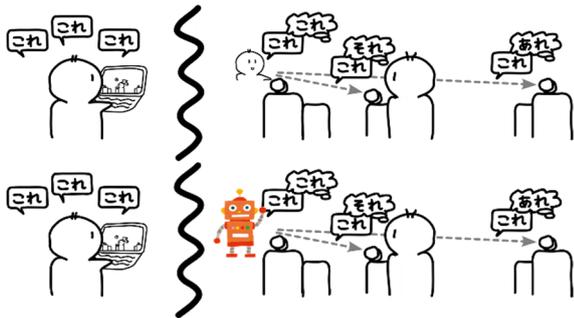


図2 投影映像と実体を伴うロボットとの比較

2. 投影型アバタを用いた実験

先行研究[1]における実験についてまとめる。この研究では、操作者の顔を利用した映像を可動式プロジェクタ[7]を用いて投影する方式により、投影型のアバタを提示するシステムを実現している。また、音声は超音波変調スピーカから射出され[8]、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。指示語の利用は、発話者が採用する視点に密接に関連しているため[9]、指示語の利用を計測することで発話者がどの視点を採用していたかを測ることができる。対話者にとって自然なのは、操作者がアバタの視点を採用することである。先行研究[1]では、操作者が指差し位置をマウス入力すると適切な移動位置を計算するが、操作者側のユーザインタフェースにおいて、指差し位置をマウス入力すると直ちに移動と指差しを行う手法を自動操作手法と呼び、指差し位置を入力するとアバタの移動位置がまず提示され、その位置をもう1度マウス入力すると移動と指差しを行う手法を半自動操作手法と呼び、両手法を比較している。

2.1 実験手順

図3に、対話者側の実験環境を示す。実験参加者は、はじめにこの部屋に案内され、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。



図3 投影型アバタを用いた実験の対話者側の実験環境

操作者側の部屋への案内後、実験参加者に自動・半自動のどちらかの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に3つのライトが設置してある様子(図3)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、リモコンでライトの色を変更する。この一連のタスクを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。図4は、3つのライトに対してシステムが計算した投影に適切な位置である。図4(a),(b),(c)はそれぞれ、対話者にとっては「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。

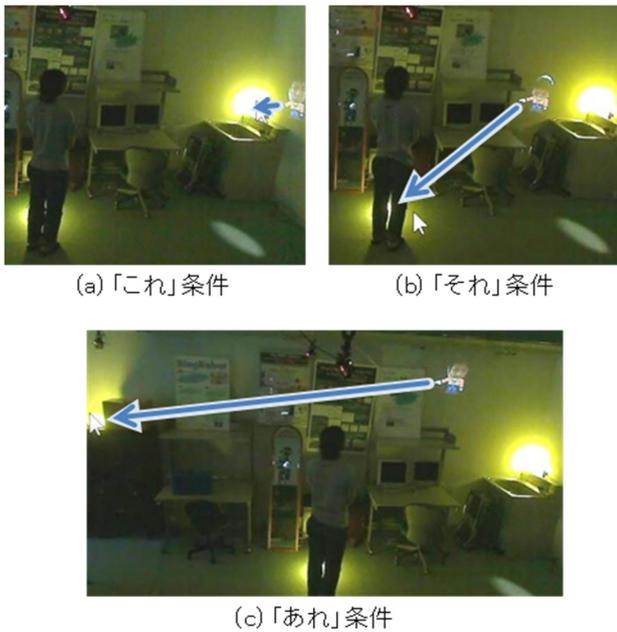


図 4 投影型アバタと対象物体の位置. 図中の青い矢印は, アバタの位置から指し示す物体の位置まで引かれている. 条件ごとに, アバタと指し示す物体の距離・対話者と物体の距離を変えている.

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.

2.2 実験結果

日本人の 36 名を実験参加者として実験を行った. 36 名のうち 18 名ずつを自動操作手法・半自動操作手法の 2 グループに分けて実験を行った. 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 5 上部に, 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発話を示している. 図 5 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, 自動操作手法よりも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
自然な指示語	4	14	6	13
別の指示語	11	4	2	1
合計	3	0	10	4

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

($p < 0.05$)

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

($p < 0.1$)

図 5 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.

2.3 考察

実験の結果は, 操作者が適切な視点を採用することが求められる遠隔コミュニケーションシステムにおいては, アバタ操作の完全な自動化は必ずしも適切ではないことを示している. 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, 先行研究においては, 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からではないかと考察されている.

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た[10]. 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た[11]. 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た[12]. したがって, 操作者の視点をア

バタの視点に誘導することができれば、不自然な発話を抑えることができると考えられる。この実験の結果は、システムの自動処理を一部制限することによって、操作者の視点をアバタの視点に誘導しうることを示唆している。

本研究は、システムの自動処理を制限することなく、操作者の視点をアバタの視点に誘導することができないかを模索するものである。その方策として、物理的な身体を持つロボットを用いて同様の遠隔コミュニケーションを行う場合を考える。ロボットの指差しには物理的な身体動作を伴うため、操作者がアバタの位置を意識して自然とアバタの視点に誘導することを期待している。

3. ロボットを用いた遠隔コミュニケーション

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実装する。対話システムは、操作者側と対話者側のシステムからなる(図6)。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(図7)。ロボット操作ソフトウェアはこの制御用コンピュータで稼働する。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図8の通り画面上で見ることができる。

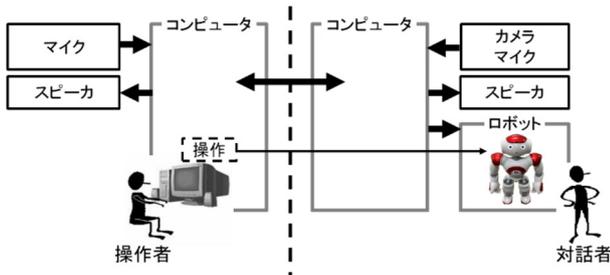


図6 システム構成。操作者側と対話者側のシステムはネットワーク接続されている。操作者と対話者は互いに対話することができ、操作者は遠隔地のアバタであるロボットを操作することができる。

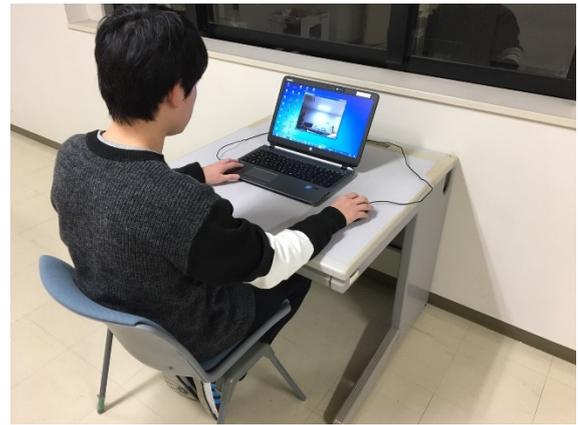


図7 対話者側のハードウェア

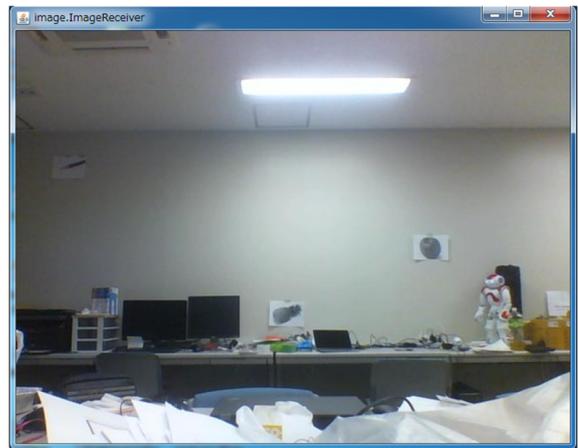


図8 ロボット操作ソフトウェアのスクリーンショット。ロボット操作ソフトウェアの画面には、対話者側の部屋の映像が表示される。操作者は、この映像をクリックすることによりロボットを操作する。

先行研究の設計と同様に、本研究では、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、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達成され、システムの自動処理を制限することなく動作させることとしている。

4. ロボットを用いた実験

実験は、投影型アバタを用いた実験の手順にしたがい行った。まず、実験参加者は対話者として当該システムの遠隔コミュニケーションを体験する。その後、ロボットの操作者となり対話者となった実験者に対して、周辺物体に対

する発話をするように指示される。このとき使用された指示語を調査する。

4.1 実験手順

図9に、対話者側の実験環境を示す。実験参加者は、はじめにこの部屋に案内され、実験者がロボット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、実験者からロボット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操作者との対話を体験する。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。



図9 ロボットを用いた実験の対話者側の実験環境

操作者側の部屋への案内後、実験参加者にロボットの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に3つの写真が貼り付けてある様子(図9)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、しばらくロボット操作の練習を行う。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ロボットを操作して、対話者側の部屋にいる実験者に写真に写っているものを答えてもらうというものである。対話者側の実験者は実験参加者の指示にしたがい回答する。この一連のタスクを3つの写真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。図9の写真はそれぞれ、対話者にとっては「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。

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。

4.2 実験結果

18名を実験参加者として実験を行った。指示対象となる写真の順序は、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。

発話	「これ」条件		「それ」条件		「あれ」条件	
	ロボット	映像	ロボット	映像	ロボット	映像
これ	14	18	11	14	4	13
それ	2	0	5	4	1	1
あれ	2	0	2	0	13	4

「それ」条件		「あれ」条件	
ロボット	映像	ロボット	映像
11	14	4	13
5	4	1	1
2	0	13	4

	「それ」条件		「あれ」条件	
	ロボット	映像	ロボット	映像
自然な指示語	5	4	13	4
別の指示語	13	14	5	14

(n.s.) (p < 0.05)

図10 ロボットを用いた実験の結果

図10 上部に、ロボットを用いた実験における指示語の発話人数を、投影型アバタを用いた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10 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、「あれ」条件において、投影型アバタよりもロボットを用いた場合に、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。

「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、「あれ」条件において有意差(p<0.05)が見受けられた。「これ」条件・「それ」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、ロボットをアバタとして用いたほうが対話者

にとって自然な指示語を用いることを示唆している。

5. 考察と議論

「あれ」条件における実験の結果は、ロボットが物理的な身体動作を伴って動作している場合は、システムの自動処理に制限しなくても、操作者が適切な視点を採用しやすいことを示している。これは、先行研究[1]におけるアバタの位置を意識させることを支持するものである。一方で、「それ」条件における実験の結果は、これまでの仮説とは一致しない。1つの可能性として考えられるのは、「それ」条件の指差し先の写真が、聞き手の実験者から遠い位置にあると認知されたために、位置関係を把握していたにもかかわらず、自然な発話を行わなかったのではないかと考えている。つまり、実験環境の不備があったことになるが、さらなる実験で精緻化していく予定である。

話し手・聞き手から近い・遠いというのは、主観的な感覚に基づくとともに、話し手・聞き手の大きさにも依存する。本研究で用いたロボットは、投影型のアバタと大きさがほぼ同様となるようなロボットを採用したが、理想的には見た目も同じアバタを使用することが望ましい。この点についても、引き続き実験を重ねて明かにしていく予定である。

先行研究においては、システムの自動処理を制限することでコミュニケーションが円滑になることが示されたが、本研究では、システムの自動処理を制限することなくコミュニケーションが円滑になることが示唆された。システム的设计において、対話者が自然と期待するコミュニケーションの作法がシステムに組み込まれた形で実現することが求められるが、同様に、操作者が意識すべきポイントをシステムが自然と誘導することができれば、誘導することができれば、よりよいシステムとなることを示している。

6. まとめ

本研究では、物理的な身体動作を伴うロボットを用いて、遠隔コミュニケーションシステムの操作者が聞き手と参照物体との位置関係を考慮した発話を行うことができるかについて検証した。検証のため、人型ロボット nao を採用し、音声に加えてロボットの身体動作により対話を行うシステムを開発した。開発したシステムを用いて実験を行ったところ、一部の条件においては、投影型のアバタを用いるよりもロボットを用いるほうが位置関係を考慮した発話を行うことが示され、操作者がアバタの位置を意識することが示唆された。

今後、実際に位置関係を考慮していたのか・どのような要因が位置関係を考慮するに必要なことであるかを明らかにすべく、実験を重ねていく予定である。また、適切な発

話を行った実験参加者とそうでない実験参加者にかけて、追加の実験により実験参加者に生じている主観的な感覚について調査することも検討している。

参考文献

- [1] Ishii, K., Taniguchi, Y., Osawa, H., Nakadai, K., and Imai, M.: Merging Viewpoints of User and Avatar in Automatic Control of Avatar-Mediated Communication, International Conference on Human-Agent Interaction, 1-2-3., (2013)
- [2] 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).
- [3] Ishiguro, H., and Nishio, S.: Building artificial humans to understand humans, Journal of Artificial Organs, Vol.10, No.3, pp. 133-142, (2007).
- [4] 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).
- [5] 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).
- [6] 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).
- [7] Fujimura, R., Nakadai, K., Imai, M., and Ohmura, R.: PROT - An Embodied Agent for Intelligible and User-Friendly Human-Robot Interaction, International Conference on Intelligent Robots and Systems, pp.3860-3867, (2010).
- [8] Ishii, K., Yamamoto, Y., Imai, M., Nakadai, K.: A Navigation System Using Ultrasonic Directional Speaker with Rotating Base, International Conference on Human-Computer Interaction, Lecture Notes in Computer Science, Vol.4558, pp.526-535, (2007).
- [9] Sugiyama, O., Kanda, T., Imai, M., Ishiguro, H., and Hagita, N.: Human-Like Conversation with Gestures and Verbal Cues based on a Three-Layer Attention-Drawing Model, Connection Science, Vol.18, No.4, pp.379-402, (2006).
- [10] Ullmer-Ehrich, V.: The structure of living space descriptions, Speech, Place and Action, pp.219-250, (1982).
- [11] Klein, W.: Local Deixis in Route Directions, Speech, Place and Action, pp.161-182, (1982).
- [12] 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).