

ロボットをコミュニケーションメディアとして利用したシステムの 複層性に関する考察

葛岡英明* 上坂純一* 山崎敬一** 山崎晶子*** 菅靖子**

*筑波大学

**埼玉大学

***公立ほこだて未来大学

kuzuoka@esys.tsukuba.ac.jp

BYI06561@nifty.com

RXU04370@nifty.com

kosaka_j@esys.tsukuba.ac.jp

ysuga@post.saitama-u.ac.jp

筆者らは、ロボットをコミュニケーションメディアとして利用することによって、遠隔作業指示を支援する研究をおこなっている。実世界においては、言葉と共に、頭の動きのような身体動作を併用しながら作業指示が与えられる。こうした身体部位によって示される志向を観察することで、人は対話者の行動を予測することができる。同様の考え方をすれば、ロボットと人の協同作業においても、ロボットの頭が、その行動を予測させるように動作しなければならない。このために筆者らは *GestureMan* を開発した。しかしながら、このシステムを利用した実験から、ロボットをメディアとして利用するシステムは、遠隔参加者のインタフェースが作り出す環境と、ローカル参加者のインタフェースが作り出す環境における人間の行動という、2つのエコロジーの整合性を考慮しなければならないことがわかった。

Dual Ecologies of Robot as Communication Media: thoughts on Coordinating Orientations and Projectability

Hideaki Kuzuoka* Jun'ichi Kosaka* Keiichi Yamazaki** Akiko Yamazaki*** Yasuko Suga**

*University of Tsukuba

**Saitama University

***Future University-Hakodate

1-1-1 Tennoudai, Tsukuba,
Ibaraki 305-8573

255 Shimo-ohkubo, Sakura-ku,
Saitama, Saitama 338-8570

116-2 Kamedanakano-cho,
Hakodate, Hokkaido 041-8655

Abstract The aim of our study is to support remote instruction via a mobile robot. In the real environment, instruction is given by words and orientations of body parts such as head movement. These orientations of body parts as well as verbal words make one capable to project other's action. Projectability is an important resource in organizing multiple actions among multiple participants coordinately. From the same viewpoint, it can be said that in the case of robot-human collaboration, we have to design a robot's head so that a co-present may project the robot's (and remote person's) actions. *GestureMan* is a robot that is designed to support projectability. It is arguable that a remote controlled mobile robot, designed as a communication medium, holds dual ecologies: ecology at a remote (robot operator's) site and at a co-present's (robot's) site. In order to design a robot as a viable communication medium, it is essential to consider how these ecologies can be mediated and supported.

1. はじめに

実世界における協同作業を遠隔的に支援しようとする試みは、CSCW において重要な研究課題となっている[11]。これに対して、我々はロボットをコミュニケーションメディアとして利用することを提案し、いくつかのロボットを開発してきた。その中で最も新しいシステムが、*GestureMan* である。我々はすでに *GestureMan* を利用し、指示者が作業員に対して遠隔作業指示をおこなうという実験を重ねてきた。そうして、これらの実験に

よって、システムの有効性を確認することができた。しかしながら、同時に問題点も明らかになってきたのである。特に本論文で議論する問題点は、

参加者がインタラクションをする環境は、参加者、ユーザインタフェース、そしてそれらが置かれた環境によって構成される。このとき、ロボットを遠隔操作する参加者（以後、遠隔参加者）側のインタフェースが構成する環境と、ロボットと対話する参加者（以後、ローカル参加者）側のインタフェースが構成

する環境との間に不整合が生じ、コミュニケーションにおける予期を支援できなくなる。

ということである。類似した問題は、Gaver がメディアスペースのアフォーダンスの問題として議論し[2]、Heath らがメディアスペースの非対称性の問題として議論している[6]。本論文では、これらの問題を、具体的なシステムや実験結果によって、より具体的に論じる。

まず、博物館の来館者の行動を観察することにより、対面的なインタラクションの場面において、人々の志向がどのように予期を可能としているのかということを説明する。次に、GestureMan と名付けたロボットシステムの概要を説明する。最後に、システムを利用した実験結果に基づいて、ロボットを媒介としたコミュニケーションにおける、エコロジーの複層性の問題に関して考察する。

2. コミュニケーションにおける志向と予期

我々は、現実の環境で人々がどのようにインタラクションを行っているかを観察するため、博物館での鑑賞場面を対象としたフィールドワークを行った。その結果、以下のことが観察できた。

- 対象物に対する志向は、頭、目、胴体、腕、指、等の、身体の各部位によって表現されている。
- それらの身体部位の志向と言語を観察することによって、対話者は発話者の次の行動を予測することができる。
- こうした予期可能性は、複数の人々が協調的に行動するための重要な資源となる。

こうした身体的志向の組織化に関する議論は、エスノメソドロジ的な会話分析や相互行為分析において議論されてきた[9, 3, 4, 5, 14, 15]。例えば、Schegloff [14] は'Can I ask a question?'といった発話は単なる質問ではなく、次に質問という行為が来ることを準備させる「行為の projection」となっていることを示した。Goodwin[3]は、考古学教育のための発掘実習において、どのように教示がなされるのかということ进行分析し、その結果、言葉だけではなく、頭部（視線）の動きが教示に対して重要な役割を果たしていることを示した。さらに Streek は、ジェスチャと予期可能性との関係を明らかにした。

本章では、頭の動きの問題に焦点を当てながら、身体の指向性と予期可能性が、いかに参加者同士の相互行為を連携的に組織化しているのかを示す。

図 1は、エレキテルという、ハンドルを回すと火花がでる展示物を親子連れで鑑賞している場面である。父親を M1、男の子を B1、女の子を G1 で示す。（）内は身体的な行動を示す。

- 1 B1: (うなずき、天井を指さす)
こんな明るくちゃ、よく見えない



- 2 M1: あ、そう? (言い終わると同時に左を見、後ろに下がる)



- 3 B1: (M1と同じ方向を見る)
4 G1: (ハンドルを離し、すぐにM1に続く)
5 M1: (歩きながら)お父さんには見える。



図 1. 頭部の志向による予期の例。

2の場面で、父親(M1)が「あ、そう?」と言いながら後ろに下がると、男の子(B1)は何も言わずにM1と同じ方向を見る。そして女の子(G1)もM1と同じ方向を見る。最後に3人は同時に、この展示から離れて歩き出す。

ここでは父親の頭とそれに伴う視線の動きが、この場面を組織化している。頭と視線の志向的な

動きは、次に父親が何を焦点として行為しようとしているかを 2 人の子どもたちに予期可能にしている。これによって、子どもたちも父親と同じ方向を見て、最後に 3 人は同時に、この展示物から歩き出したのである。

この例からわかるように、頭とそれに伴う視線の動きは、次に何を見ようとしているのか、そして何をしようとしているのかということの予期可能としているのである。周りの人たちは、同じ場所に視線を動かすことで、何が焦点となっているかを見ることができる。そしてさらにその焦点を確認することで、次の行為が予期可能になる。

では、こうした頭の動きは、メディアとしてのロボットによって支援されうるのだろうか。次章で GestureMan の概要を示した後、予期の支援という観点からシステムの問題点に関して議論する。

3. GESTUREMAN

筆者らは、遠隔コミュニケーションを支援することを目的として、ロボット開発とその評価を続けてきた[10, 17]。GestureManはそうした継続的な研究によって開発した最も新しいロボットである(図 2)。ロボットの設計に際しては、予期を支援するために、以下の要件を考慮した。

- **相互観察の支援**：参加者はお互いの行為や志向を相互に観察できなければならない。
- **身体的な志向表現の支援**：遠隔参加者の志向がわかるように、ロボットの胴体や頭は志向を明確に表現できなければならない。

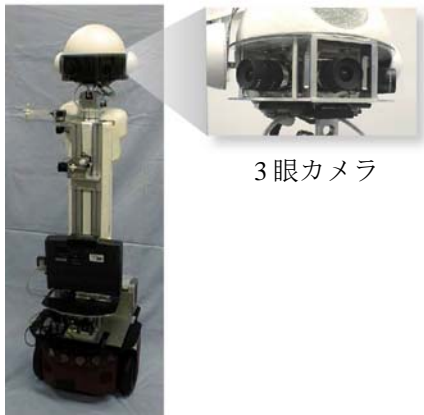


図 2. GestureMan

遠隔参加者がローカル参加者の志向を観察するためには、ロボットに広視野角のカメラを装備することが効果的である。図 2 に示すように、GestureMan は 3 眼カメラを装備している。遠隔

参加者側には 3 台の液晶ディスプレイ (LCD) を横に並べて設置し、ここに 3 眼カメラの映像を表示する(図 3)。各カメラのレンズは 60 度の視野角を持つため、遠隔参加者は、合わせて 180 度の視野角を得ることになる。



図 3. 遠隔参加者のための 3 画面ディスプレイ



図 4. 科学技術館における遠隔ガイド実験

我々はこれまで、GestureMan を利用した実験をいくつか行ってきた。図 4 は、ロボットを科学技術館のガイドとして利用した実験例である。この実験では、遠隔参加者がロボットを操作し、博物館の来館者に対して展示物の解説をおこなった。これらの実験を通して、広視野角映像の支援が、以下の理由により有効であることを確認した。

- 図 4 に見られるように、ローカル参加者はロボットの横方向に立つことが多く、この場合左右のディスプレイに表示されることになる。従って、遠隔参加者がローカル参加者を観察するためには広視野角が有効である。
- 遠隔参加者はローカル参加者側サイトにある対象物をディスプレイで確認する必要がある。当然、この目的には広視野角が有効である。

ローカル参加者が遠隔参加者の志向を観察できるようにするためには、これを提示する資源をロボットに装備する必要がある。このために、3眼カメラは白いヘルメット内に納め、頭のように見えるようにした(図2)。また、耳とバイザーを取り付けることによって、中央のカメラが向いている方向がわかりやすくなるようにした。頭部はモータ駆動によりパン・チルトが可能である。これによって、ロボットは胴体の志向や頭部の志向によって示すことができる。実際、科学技術館における実験では、遠隔参加者がどの対象物に対して説明しようとしているのかと言うことを、ロボットの身体の向きから予測している例が見られた。

こうした観察から、GestureManは、遠隔参加者がその頭部を適切に操作しさえすれば、予期の支援に有効であるということを確認した。しかし、同時に別の問題点も明らかになってきた。遠隔参加者側のサイトとローカル参加者側のサイトの環境の不整合によって、遠隔参加者の志向が正しくローカル参加者に伝わらないのである。次章では、コミュニケーションメディアとしてのロボットには避けられない、環境の複層性に関して議論する。

4. DUAL ECOLOGIES

遠隔参加者は3画面ディスプレイによる広い視野角によって、ロボットの頭をあまり操作することなく、次に説明すべき対象物を探したり、進むべき方向を確認したりできる。このとき、遠隔参加者は左右のディスプレイを見るために首を左右に振ることになる。

2章で示したように、対面的なコミュニケーションでは、対話者のこうした志向の変化を観察することによって、次の行為を予測することができた。しかし、本システムでは、ディスプレイの前での首振りにはGestureManに反映されないのである。遠隔参加者は、そのインタフェースが作り出す環境において首振りをしているにもかかわらず、その動作がローカル参加者側のインタフェースが作り出す環境に反映されないのである。このようなエコロジーの不整合により、ローカル参加者による予期が支援されなくなってしまうのである。

それでは、遠隔参加者側のディスプレイを1画面にすればこうした問題は解決するだろうか。この疑問を明らかにするために、視野角の広さが、ロボットをメディアとしたコミュニケーションシステムに与える影響を調べるための実験を行った。

4.1. 実験設定

GestureManを用いて、遠隔参加者(指示者)が、ローカル参加者(作業員)側サイトの空間内にある様々な対象物について説明を行い、その対象物に作業員が触れるという実験を行った。作業員側の空間には機械、人形、ケーブル、時計、本など、計60個の指示対象物が混在した。実験環境を図5に示す。3画面ディスプレイを指示者の前に設置し、ロボットの3眼カメラからの映像が表示された。指示者はジョイスティックを利用してロボットの頭のパン・チルトのみを操作した。ロボットの頭の操作、指示の仕方は指示者の自由とした。ただし、指示者の志向に対する作業員の反応に注目したいため、指示者には指示対象物をモニターで発見した後で指示を開始するようにしてもらった。また、作業員にはロボットの正面に立ってもらい、指示対象物に触れるとき以外には、その位置を動かさないようにしてもらった。



(a) 指示者サイト (b) 作業員サイト

図5. 実験風景

実験の具体的な手順は以下のとおりである。

1. 指示者は指示対象物の写真(図2)を見て、その指示対象物をディスプレイで探す。
2. 指示者は指示対象物を見つけた後、作業員に指示を行う。
3. 作業員が正しい指示対象物に触れるまで、指示者は指示を行う。

手順1から3までを1回の作業として、それを30回繰り返して1回のタスクとした。1回のタスクにおいて、30回すべて指示対象物は異なり、指示対象物の順番はランダムに決定した。

実験条件として、指示者のLCD3台全てを表示する場合(3画面条件)と、中央の1画面のみ表示する場合(1画面条件)を設定した。

実験では、男女の大学生16名を指示者(8名)、作業員(8名)として採用した。被験者内配置法により2条件、2種類のタスク(それぞれ30個の対象物)を同じペアに割り当て、指示条件の順序、

タスクの順序を変えることで持ち越し効果、順序効果の相殺を意図した。ここで、2種類のタスクの難易度は同程度と仮定した。

4.2. 作業指示時間

実験中に録画したビデオによって、指示者と作業者の行動を観察した。また、実験データを計測する際に、以下に定義する時点を基準とした。

- ・ 作業開始：指示者が指示対象物の写真に目を落とした後、それをモニターで探すために顔を上げた瞬間。
- ・ 指示開始：指示者が指示における最初の言葉を発した瞬間。
- ・ 作業終了：作業者が指示対象物に触れた瞬間。

1回のタスクにおいて、作業開始から作業終了までの時間の平均を作業時間とした。また、作業時間のうち、作業開始から指示開始までジョイスティックを操作していた時間を「指示前操作時間」、指示開始後の残りの時間を「作業指示時間」と定義した。

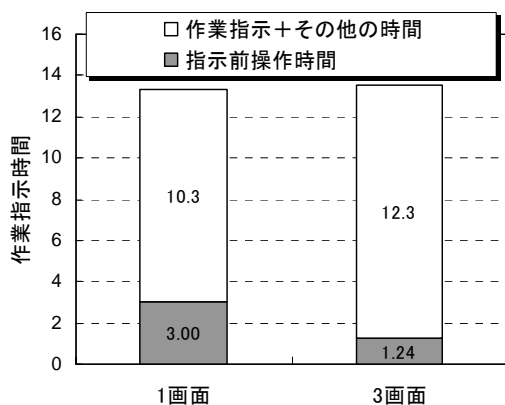


図6. 平均作業時間

図6に結果を示す。指示前操作時間において、ウィルコクソンの符号順位検定を行ったところ、1画面条件の方が3画面条件よりも指示者は指示開始前にジョイスティックを長く操作していた ($p < .01$)。これは、3画面条件の場合は、視野角が広いので、ロボットの頭を操作しなくても指示対象物を発見できたのに対して、1画面条件では、ロボットの頭を操作しないと指示対象物を発見できないことが多かったためと考えられる。

しかしながら、平均作業時間は両条件でほとんど差がなかった。これは、指示開始後に作業者が対象物を見つけて、それに触れるまでの時間に関

して、3画面条件の方が多くの時間を要したことを意味し、この点では1画面条件の方が有利であったといえる。

そこで、もし1画面条件と3画面条件の利点のみを統合することができれば、作業時間は約11.5秒 ($= 10.3 + 1.24$)にまで短くできるかもしれない。そこで、どうすればそのような作業指示が可能になるかということを検討するために、実験結果をより詳細に観察してみよう。

4.3. 予測反応

1画面条件の場合、指示者が画面上で対象物を確認するためには、3眼カメラの中央のカメラで対象物を撮影しなければならないため、ロボットの頭がその対象物の方向を向くように、自然にコントロールされることになった。作業者はこの頭の動きを観察することによって、指示者が言葉で説明を始める前に、その対象物の方向へ頭を向けたのである (

図7)。そこで、このような反応を「予測反応」と定義し、それぞれの実験条件における予測反応の発生回数を比較した (図8)。その結果、1画面条件の方が、予測反応回数が有意に多いことがわかった ($p < .01$)。



図7. 予測反応の例 (左：遠隔、右：ローカル)

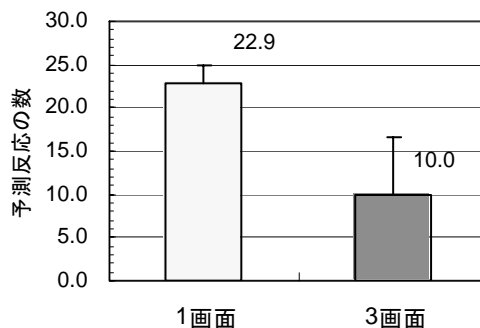


図8. 予測反応の数

4.4. 誤予測反応

3画面条件において、対象物が右、又は左側のディスプレイに表示されている場合、ロボットの頭はその対象物の方向とは全く異なった方向を向いていることになる。このような場合、指示開始の段階では、作業者が対象物とは反対の方向を向いてしまうという行動が観察された。例えば

図9では、指示者が左側のディスプレイ内に対象物を発見したため、彼はロボットの頭を左側に向けず、単に少し下向きに操作しただけであった。そして、言葉で説明を開始したと同時に、作業者は頭を（彼女から見て）左側へ、すなわちロボットの頭が向いている方へ向けたのである。ここではこのような動作を「誤予測反応」と呼ぶことにする。図10に誤予測反応の回数の比較を示す。この図からわかるように、3画面条件における誤予測反応の数は、1画面条件よりも多く、ウィルコクソンの符号順位検定により有意な差が認められた($p < .01$)。



図9. 誤予測反応の例

アンケート調査からは、7名の指示者が、3画面条件の方が指示しやすかったと答えた。一方、作業者に関しては、7名が1画面条件の方が対象物を発見しやすかったと答えた。

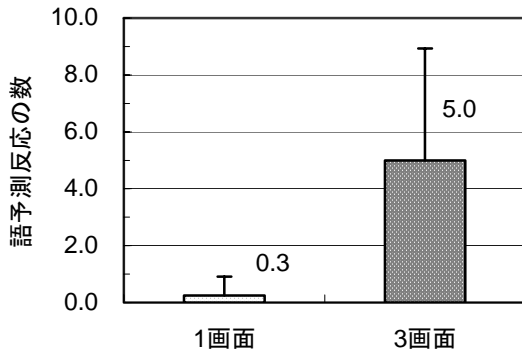


図10. 誤予測反応の数

これらの結果から、指示者は、対象物の発見しやすさから3画面条件の方を好むことがわかった。これに対して作業者は、ロボットの頭の向きから指示対象物を予測できるので、1画面条件の方を好むのである。

5. 考察

本論文では、指示対象物を探すための頭部の志向を変化させるという予備的な動作が、予期の重要な資源となっていることを示した。ロボットを媒体としたコミュニケーションでは、志向を提示する機構が適切に操作されさえすれば、予備的な動作が作業者から観察され、予期が支援されることになった。しかし、前章の実験結果から、遠隔参加者に提供するインタフェースによっては、ロボットの志向提示機構が操作されず、従って予期が支援されなくなってしまう場合があることがわかった。

ここで、遠隔参加者とそのユーザインタフェースによるエコロジーと、ローカル参加者とそのユーザインタフェースによるエコロジーという、2つのエコロジーの整合性という観点から、この問題を考えてみよう。

3画面条件の場合には、遠隔参加者は、そのインタフェースが作り出す作業環境に対応して、頭部を左右へ向けると言う自然な動作をおこなっている。しかし、こうした動作はロボットの頭部の動作には反映されない。ローカル参加者側では、その空間内に存在する対象物と、そこに配置されたロボットが作り出す環境に対応した動作をおこなうことになる。ここで、ロボットの頭が、遠隔参加者の実際の志向とは異なる志向を示している場合には、ローカル参加者はロボットの志向を信用して、それに対応した反応をおこなってしまうために、指示対象物の発見が遅くなる。

1画面条件の場合には、遠隔参加者が対象物を探すためには、ロボットの頭部を操作せざるを得ない環境を作っていることになる。これによって、遠隔参加者の志向とロボットの頭部が示す志向とが一致し、ローカル参加者は正しく予期をおこなうことができる。しかしながら、遠隔参加者にとって、このインタフェース環境が使いづらいことは明らかである。

もう一度3画面条件について考えてみよう。ディスプレイが3つある場合には、遠隔参加者は頭を動かして志向を変化させている。しかしながら、作業指示者の頭部動作は、ユーザインタフェース

である 3 画面ディスプレイに対して向けられるに過ぎないため、それがロボットの動作に反映されないのである。すなわち、遠隔側とローカル側の 2 つの環境の不整合性のために、予期が支援されなくなってしまうのである。

システムに慣れるに従って、そうした矛盾に順応するということはあり得るが、ほとんどの予備動作は無意識でおこなわれる。従って、多くのユーザにとって、そうした無意識の行為をロボットに表現させるために、意識的にロボットをコントロールすることは難しい。すなわち、無意識的な予備動作を検出することができるユーザインタフェースの開発が必要となるのである。

例えば本研究の例で言えば、遠隔参加者側のインタフェースは、システムが検出可能な程度の頭部の動きを遠隔参加者に起こさせるように設計されなければならない。ローカル参加者側では、遠隔参加者の志向を示すことができる何らかの提示装置が、指示者の動作に呼応して制御されなければならない。これは、当然のことのように思われるかもしれないが、遠隔操作型のロボットにおいて、これを考慮して設計された物はほとんど無く、本論文で紹介した我々のインタフェースと類似の方式を採用しているに過ぎない。従って、同様の問題を抱えていることになる。

この問題を解決する 1 つの方法は、指示者のインタフェースとして、広視野角の頭部搭載型ディスプレイ(HMD)と、3次元位置センサを利用することである。しかし、広視野角 HMD は大型で重くなりがちであり、ユーザにとって負担が大きいのが現状である。

そこで、本研究では図 11に示すシステムを開発中である。このシステムの特徴は、カメラが頭ではなく、胴体に取り付けられているということである。指示者は 3 画面ディスプレイの前に座り、頭に 3 次元位置センサを装着する。3 画面ディスプレイは左右に広く配置されるため、このディスプレイが作り出した空間を眺めるとき、指示者は自然に左右に首を振ることになる。そこで、この頭の動きを検出して、それによってロボットの頭部を動作させるのである。この方法の特徴は、ロボットの頭は単に指示者の志向を表現するためだけに実装されていて、これによってカメラの向きは変化しないという点である。図 12は試作中のシステムであるが、3 眼カメラの向きとは独立に頭部が動作することがわかる。

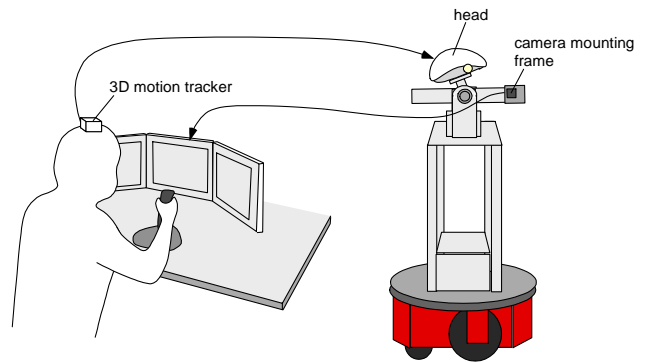


図 11. エコロジーの複層性を考慮したシステムの概念図

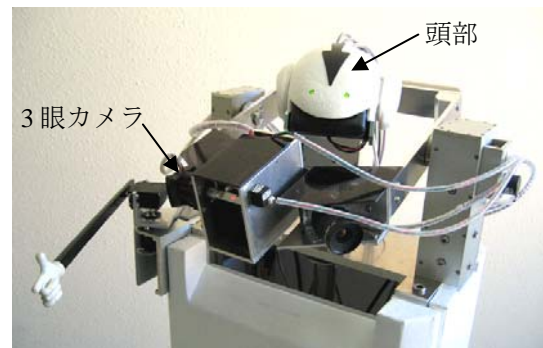


図 12. 試作中のシステム。3 眼カメラと頭部の向きが異なっている。

The Mutually-Immersive Mobile Telepresence [8] も似通った手法を採用しているが、遠隔からの参加者の頭部が、ロボットの頭部に搭載された液晶ディスプレイに表示されている。この手法も興味深いものではあるが、モナリサ効果のために、斜め方向からこのディスプレイを見たときに、正しく志向が伝わらない可能性がある。

6. おわりに

本研究では、遠隔参加者とローカル参加者それぞれの環境の不整合性によって生じるエコロジーの複層性に関して考察した。複層性の存在を考慮せずにシステムを設計すると、遠隔参加者の予備的な動作がローカル参加者に伝わらず、予期の支援ができなくなってしまうのである。

7. REFERENCES

1. Dourish, P., Adler, A., Bellotti, V., Henderson, A. Your place or mine? Learning from long-term use of audio-video communication. Kluwer Academic Publishers Norwell, MA, USA (1996), 33-62.

2. Gaver, W. The Affordances of Media Spaces for Collaboration. *Proc. CSCW'92*, ACM Press (1992), 17-24.
3. Goodwin, C. Professional vision. *American Anthropologist* 96 (1994), 606-33.
4. Goodwin, C. Transparent vision. In: Ochs, E., E.A. Schegloff, S.A. Thompson, eds., *Interaction and Grammar*. Cambridge, Cambridge University Press (1996), 370-404.
5. Goodwin, C. Action and embodiment within situated human interaction. *Journal of Pragmatics*, 32 (2000), 1489-1522.
6. Heath, C, Luff, P. Media space and communicative asymmetries: preliminary observations of video mediated interaction, *Human-Computer Interaction* 7 (1992), 315-346.
7. Heath, C., Luff, P., Kuzuoka, H., Yamazaki, K. Creating coherent environments for collaboration, *Proc. ECSCW2001* (2001), 119-128.
8. Jouppe, N. First steps towards mutually-immersive mobile telepresence, *Proc. CSCW2002*, ACM Press (2002), 354-363.
9. Kendon, A. *Conducting Interaction: Patterns of Behavior in Focused Interaction*, Cambridge, Cambridge University Press, 1990
10. Kuzuoka, H., Oyama, S., Yamazaki, K., Suzuki, K., Mitsuishi, M. GestureMan: A Mobile Robot that Embodies a Remote Instructor's Actions, *Proc. CSCW2000*, ACM Press (2000), 155-162.
11. Nardi, B., Schwarz, H., Kuchinsky, A., Leichner, R., Whittaker, S., and Scلابassi, R. Turning away from talking heads: the use of video-as-data in neurosurgery, *Proc. CHI'93*, ACM Press (1993), 327 – 334.
12. Norman P. Jouppe, First Steps Towards Mutually-Immersive Mobile Telepresence, *Proc. CSCW'02*, ACM Press (2002), 354-363.
13. Paulos, E. and Canny, J. ProP: Personal Roving Presence, *Proc. CHI' 98*, ACM Press (1998), 296-303.
14. Schegloff, E. A. Preliminaries to preliminaries: "Can I ask you a question?", *Sociological Inquiry* 50 (1980), 104-52.
15. Schegloff, E. A.: 'Body torque', *Social Research*, 65 (1998), 535-86.
16. Streeck, J. On Projection. In: E. Goody, ed., *Interaction and Social Intelligence*, Cambridge, Cambridge University Press (1995), 84-110.
17. Yamazaki, K., Yamazaki, A., Kuzuoka, H., Oyama, S., Kato, H., Suzuki, H., and Miki, H. GestureLaser and Gesturelaser Car: development of an Embodied Space to Support Remote Instruction, *Proc. of ECSCW'99* (1999), 239-258.