

# ヒューマノイドロボットを用いた 遠隔コミュニケーションにおける注目伝達

森田 友幸<sup>†</sup> 間瀬 健二<sup>††</sup> 平野 靖<sup>††</sup>  
梶田 将司<sup>††</sup> 岡留 剛<sup>†††</sup>

会議において、人は言語情報の交換だけでなく、様々な非言語的な合図を用いてコミュニケーションを行っている。本稿では、ヒューマノイドロボットを用いた遠隔コミュニケーションシステム、すなわちロボット会議システムを前提として、ロボットによるマルチモーダルな非言語情報伝達の可能性について調べる。ロボットを遠隔ユーザのアバタとして用いると、遠隔ユーザの頭部方向や目の動きなどの身体的な情報を伝達し再現することによって、注目状況を表現し、それにより対話内容に対する興味の度合いを相手に伝達できると考えられる。そこで、本稿では、注目状況の表現と伝達がロボットを使うことで優位に働くかどうかを調べる。遠隔で行う会議における発表とその聴講の場面を想定して、遠隔ユーザが発表聴講に集中して注目しているかどうかを、発表者が非言語情報により推測できるという仮説を立てる。対面での自然なコミュニケーションと比較して、提案するロボット会議システムが、従来のビデオ会議システムに対しどの程度優劣があるかを実験的に検証し、ロボット会議においてビデオ会議より優れて推測できることが示された。

## Distant Attention Communication with Humanoid Robot

TOMOYUKI MORITA,<sup>†</sup> KENJI MASE,<sup>††</sup> YASUSHI HIRANO,<sup>††</sup>  
SHOJI KAJITA<sup>††</sup> and TAKESHI OKADOME<sup>†††</sup>

We propose to use a humanoid robot for remote communication with multimodal information such as head orientation and eye movements. We communicate each other in a meeting with speech information as well as various non-verbal signals. Humanoid robot based remote conferencing system will be useful by its embodied characteristics for smooth non-verbal communication. Gestural behavior such as eye-gazing and attention of remote user can be transmitted and expressed by the robot to the meeting members to transfer the attention and interest of the remote user easily. This paper reports the experimental results of comparing the proposed "robot meeting" system with a "face-to-face meeting" mode and a "video meeting" mode. Our proposing system is shown to be more effective than video conferencing in terms of determining whether a remote user is actively giving attention to the presentation or not.

### 1. はじめに

通信インフラの発展などを背景として、音声会議システムやビデオ会議システムを用いた遠隔コミュニケーションが多く行われるようになってきている。ビデオ会議システムは、音声だけの電話と比較して、相手の表情が見えるという利点がある。しかし、アイコンタ

クトが困難であることや、遠隔ユーザがどこを見て何に注目しているかといったことを把握するのが困難であるといった問題があった。遠隔ユーザの立場から、議論への参加が困難であるという不便があった。これらの問題は、ロボットをコミュニケーションの媒介(メディア)として用いることにより解決可能であると考えられる。そこで、本稿では、特に遠隔ユーザのプレゼンテーションへの注目状況が、ロボットを用いることで効果的に伝達可能であるかどうかを、実験的に検証する。実験を統制するため、発表者とその聴講者という1対1の場面を想定し、ロボット会議、ビデオ会議、対面会議の3条件で、参加する聴講者の注目状態を、プレゼンテーションを行う発表者がどの程度推測可能かを比較する。その結果、ロボット会議がビデ

<sup>†</sup> 名古屋大学大学院情報科学研究科  
Graduate School of Information Science, Nagoya University

<sup>††</sup> 名古屋大学情報連携基盤センター  
Information Technology Center, Nagoya University

<sup>†††</sup> 日本電信電話株式会社コミュニケーション科学研究所  
NTT Communication Science Laboratories, NTT Corporation

才会議より対面会議に近く優れて推測可能であることを示す。

我々人間が対面でコミュニケーションを行う際には、言語情報、周辺言語情報、非言語情報など様々な情報をやりとりしている。身体的特徴・服装、表情、動き・姿勢、視線、距離・相対位置などの非言語情報は、主に視覚的に伝達される情報であり、様々な役割を担っている。動きや姿勢から、相手の興味や疲労状態などを知ったり、視線を用いて相手の注意をひいたり、または、相手の視線方向を見ることで相手が何に興味を持っているかなどを把握できる。我々が会議や打合せを行う場合に、その内容が重大であるほど電話やメールなどではなく直接会おうとするのは、電話やメールでは伝達できない非言語情報に重要な情報が含まれているからである。

一般に対面会議において利用される視覚的な情報には、姿勢、ジェスチャ、表情、視線などがある。ビデオ会議では、表情そのものは伝達可能であるが、その表情の示唆する意味はその表情を向けた対象によって変化するため、その対象が分かりにくい2次元ディスプレイでは、表情による情報が完全に伝達されているとはいえない。また、視線は互いの空間が異なるために伝達が困難であり、相手と視線が合わず違和感を感じるというアイコンタクトの問題<sup>1)</sup>や、その逆のモノリザ効果(カメラ目線をしている画像をどの角度から見ても目線が自分を見ているように感じる現象)など、非常によく知られた問題がある。距離や相対位置関係は、ビデオ対話では画面を隔てるために、「そちら」と「こちら」という一定以上の離れた距離関係に感じることが知られており、伝達が困難であると考えられる。

これらの問題の多くは、ヒューマノイドロボットをコミュニケーションの媒体として用いることにより解決できると考えられる。視線の問題は、遠隔ユーザの視線や頭の向きをロボットの視線や頭の向きとして表現することで、遠隔ユーザの視線を容易に知覚することができるようになり解決できると考えられる。ロボットの眼にカメラが搭載されていれば、遠隔ユーザの視点とロボットの視点が一致するため、アイコンタクトをとることも可能になる。また、ヒューマノイドロボットは人間に近い見かけを持っているため、人間の身代わりと見なしやすいという特徴がある。そのため、ヒューマノイドロボットを遠隔ユーザの意志に合わせて動作させることで、遠隔ユーザがそこにいるような存在感が得られる可能性がある。

以下、まず3章で実験のために試作したロボットを用いた遠隔コミュニケーションシステムの概要を示

す。4章で実験の内容について示し、5章と6章で実験の結果を示し考察する。

## 2. 関連研究

ビデオ会議におけるアイコンタクトの問題に関しては、多くの研究がなされている。その解決にロボットを使う方法として、Talking Head<sup>2)</sup>は、非常に早い時期にこの問題に注目し、頭部形状ディスプレイに顔映像を伝送投影し、ディスプレイ回転と表情画像伝達により視線一致の会議が可能となるコンセプトプロトタイプを提案している。それに対しEye-ballロボット<sup>3)</sup>は、ロボットによる遠隔コミュニケーションのコンセプトを提案しつつ、2自由度を持つ目玉のある頭部ロボットをアバタとするカメラとHMDによる遠隔コミュニケーションシステムを試作し、その同調効果を確認している。KuzuokaらはGestureMan<sup>4)</sup>において、レーザポインタによる対象指示機構を持つロボットによる遠隔コミュニケーションを提案している。特に文献5)では、指示動作において遠隔ユーザの頭部動作とロボットの連動が有効であることを示している。視線や頭部動作の表現の重要性・有効性については、本稿も同じ立場をとる。しかしこれらの研究では、遠隔ユーザがコミュニケーションに参加していることを前提としているが、本稿では、ロボットコミュニケーションにおいて、出席している遠隔ユーザが会議に注目して参加している状態かどうかの推定に貢献する情報が何か、に着目している。コミュニケーションチャンネルの確立にどのモダリティがいかに貢献するかを個別に調査していくことはこの分野の重要な課題である。MAJIC<sup>6)</sup>では遠隔ユーザが投影される半透過スクリーンの裏側にカメラを設置し、遠隔ユーザの顔が投影される位置から撮影することにより視線一致を実現している。このシステムでは、専用のハードウェアと、ユーザの目の位置とカメラの光軸を合わせるための細かい設定が必要となる。また、このシステムは複数の参加者が1つの場所にいる場合には利用できない。

PROP<sup>7)</sup>やJouppi<sup>8)</sup>は、移動ロボットにカメラとディスプレイを設置したアバタを使った遠隔コミュニケーションやテレプレゼンスについて報告している。遠隔の身体性の重要性について指摘している点は、本稿と同じ立場であるが、ディスプレイを用いているため、表情の情報の貢献が混在してしまっていた。同様の問題に対して、画像処理を用いた研究もある<sup>9),10)</sup>。これらの研究では、ユーザの視線を画像処理により認識し、3次元の合成画像により画面上で表現する。1つの画面の表示可能なのは1人のユーザの仮想的な視

界のみであるため、このシステムも複数の参加者が1つの場所にいる場合には利用できない。

まったく別のアプローチとして、Morikawaらは2人の遠隔ユーザが1つの画像に投影される“Hyper-Mirror”<sup>11)</sup>と呼ばれるシステムを開発した。この研究では、視線一致を実現しなくてもユーザは不自然さを感じないという結果が示された。しかし、このシステムを用いたコミュニケーションは新しい形のものであり、会議といった既存の形式のコミュニケーションには適用しにくいと考えられる。平田ら<sup>12)</sup>はt-Roomという多人数遠隔会議システムを試作して、空間の共有を目指している。多数のカメラと平面ディスプレイを配置して、相互の位置関係を保存したまま遠隔地の空間を重ね合わせて、遠隔会議の距離の克服を試みている。

石黒らは、人間にきわめて近い外見と表情・動作制御機構をもったアンドロイドロボットであるGeminoid<sup>13)</sup>を開発し、自分の分身ともいえるロボットで遠隔地と対話し、自分がそこに存在するかのような感覚を与えようというアプローチをとっている。坂本ら<sup>14)</sup>は、Geminoidを存在感メディアとして用いて、存在感の伝達に効果があることを被験者実験をもとに報告しているが、その存在感が何に起因しているかの分析に至っていない。存在感は、共存の場の形成が重要であるとの立場から影を用いた遠隔コミュニケーション<sup>15)</sup>によっても検討されている。本稿は、表情を表出できないロボットでも、頭の動作の伝達が注目を伝達し、それが存在感の一要因となりうることを示す。

遠隔コミュニケーションのアパタとして3次元CGキャラクタや、3次元CGとカメラ画像の合成などを用いた研究も行われている<sup>16)~18)</sup>。3次元CGを用いる利点は、複数の遠隔地に分散している参加者を1つの3次元空間に描画することにより、同一の空間にいるという感覚を与えることができる点である。また、各ユーザがそれぞれディスプレイの前にいるという前提であればキャラクタとの視線一致により、相手との視線の一致を表現できるという利点もある。しかし、ある拠点に複数の参加者がいるような場合、ビデオ会議と同様のモナリザ効果が起こり、誰に対して視線を送っているかを伝えることは困難となる。また、Kiddらの研究<sup>19)</sup>により、コミュニケーションにおいて画面のキャラクタよりも、実体を持ったロボットの方がユーザの知覚に良い影響を与えることが示されている。

### 3. ヒューマノイドロボットを用いた遠隔コミュニケーション

我々は、図1のようなロボットを用いた遠隔コミュニケーションシステム(ロボット会議システム)を、ヒューマノイドロボットを媒体として用いて実現することを提案する<sup>3)</sup>。ヒューマノイドロボットは人間に似た存在と見なすことができる対象であり、遠隔ユーザのアパタとして適していると考えられる。遠隔ユーザの注意や注視をロボットの頭部方向や視線方向で表現し、ロボットの目カメラで撮影した映像を遠隔ユーザに送り、会議に参加する<sup>5)</sup>。また、遠隔ユーザのジェスチャや動きもロボットのジェスチャや動きで伝達することができる。

我々はまず、このようなロボット会議を実現するシステムを試作し、それを用いてロボットの頭部動作によるユーザの注目状況の伝達について検討した(図中のcommandからrepresentationまで)。

#### 3.1 ロボット会議プロトタイプシステム

遠隔ユーザは、図2に示すように6自由度のモーショントラッカ(Humanoid Motion Tracker IS-900)のタグを設置した帽子、ヘッドマウントディスプレイ(HMD)(Icuiti V920、視野角32度、2.7m距離で62インチスクリーン相当)およびイヤホンを装着する。モーショントラッ

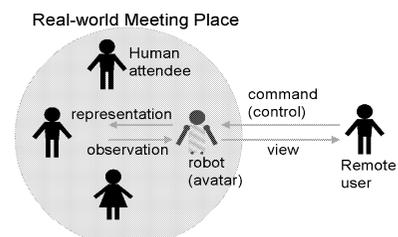


図1 ロボット会議システム

Fig. 1 Robot tele-conferencing system.

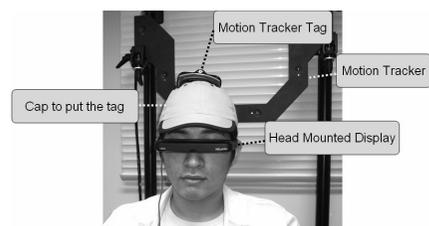


図2 ロボット会議プロトタイプシステム(遠隔ユーザ側)

Fig. 2 Prototype of robot tele-conferencing system (remote user).

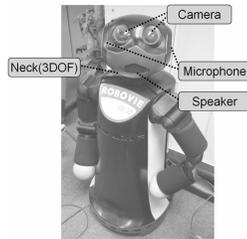


図 3 Robovie R-2  
Fig. 3 Robovie R-2.

力により、システムはユーザの頭部の位置および 3 軸の周りの回転角を検出することができる。この動作データはネットワークを通じてロボットに転送され、ロボットの頭部がそのデータに基づいて動作する。

ロボットの目の部分に搭載されているカメラ (Watec 社製 Wat230A モデル, 標準レンズ f3.8mm, 視野角 51 度) から取得された映像と、耳の部分についているマイクから取得された音声、ネットワークを介して遠隔ユーザに転送される。映像は HMD に表示され、音声はイヤホンで再生される。このシステムで HMD を用いるのは、通常のディスプレイを用いた場合ユーザが横を見た際に、ディスプレイが見えなくなってしまうからである。この問題に対しては、ユーザの見る範囲をすべて網羅するような、非常に大きなディスプレイを用いるという方法も考えられる<sup>5)</sup>。

本実験では、ヒューマノイドロボットとして Robovie R-2<sup>20)</sup> (図 3) を用いた。Robovie は 2 つの目 (パンチルトカメラ) と 2 つの耳 (マイク) を搭載している。首はロール、ピッチ、ヨーの 3 自由度で動作し、人間の頭の動きを模倣することが可能である。今回の実験では、目のカメラは頭に対して正面の角度で固定し、方向は変化させなかった。

#### 4. 注目状況伝達の被験者実験

##### 4.1 概要

遠隔コミュニケーションにヒューマノイドロボットを用いることで、遠隔ユーザの注目状況を伝達できるという有効性を評価するために被験者実験を行った。この実験では、通常での対面でのコミュニケーションを想定した「対面会議」条件、提案システムを用いてロボットを介した「ロボット会議」条件、および従来のビデオ会議システムの使用を想定した「ビデオ会議」条件の 3 条件を比較した。

我々はまず、ロボットの頭部動作を用いることで、プレゼンテーションを聞いているか聞いていないかといった遠隔ユーザの注目状態の把握が、ビデオ会議シ

ステムと比較して容易になると考えた。すなわち、ビデオ会議システムでは、遠隔ユーザの視線はディスプレイの範囲に固定してしまい、どこに注目しているかを認識するのが困難である一方、ロボットを用いた場合、遠隔ユーザの注意はロボットの頭部動作で表現され、ローカルユーザは遠隔ユーザの注目状況を容易に把握できると考えられるからである。

##### 4.2 手順

本実験では、被験者が同じ条件で実験に臨む被験者内実験を以下のようにデザインした。まず、各実験では発表者および聴講者の 2 人の被験者が参加した。発表者には各条件に対して用意された原稿とスライドを用いて 6 回の異なるプレゼンテーションを行うよう指示した。聴講者には各条件のもとで発表者の発表を聞くよう指示した。ただし、各条件の 6 回のプレゼンテーションのうち 3 回は画面に“間違い探しゲーム”が表示され、表示された場合には聴講者には発表を聞かずゲームに集中するよう指示した (ゲームは間違いを探すのみでユーザの入力をいっさい必要としない)。すなわち、6 回中の 3 回はプレゼンテーションを“聞いている”状態で残りの 3 回は“聞いていない”状態である。各回数はともに 3 回の固定であるが、その順序は無作為に決定した。

発表者と聴講者の被験者ペアは、後述のとおり経験者グループおよび未経験者グループ各 4 組、合計 8 組の被験者ペアのそれぞれに対し「対面会議」「ロボット会議」「ビデオ会議」3 条件それぞれについて、6 回ずつの試行 (プレゼンテーション) を行った。したがって試行の総数は 8 組 × 3 条件 × 6 試行 = 144 回となる。

条件ごとの実施順序は「対面会議」条件に対する比較をみるために「対面会議」条件を最初にして「ビデオ会議」条件と「ロボット会議」条件の順序をランダムに組む方針をとって順序効果をおさえるようにした。その結果「対面 - ビデオ - ロボット」の順が 4 組、「対面 - ロボット - ビデオ」3 組、「ビデオ - 対面 - ロボット」が 1 組の割合となった。

発表用題材は、ニュース記事をもとにした 1~2 分程度の発表用題材を全部で 36 個用意し、その中からそれぞれランダムに 18 個 (3 条件 × 6 試行分) を選んだ 4 グループの題材 (m1~m4) を構成した。各被験者の組はそれらのうち 1 つの題材を用いて発表を行った。

各試行における発表者のタスクは聴講者が「聞いて

後述の分析では 3 条件を対等にして検定を行っており、順序効果の影響を排除するには 6 通りの組合せを作るように実験をデザインするのが本来望ましい。

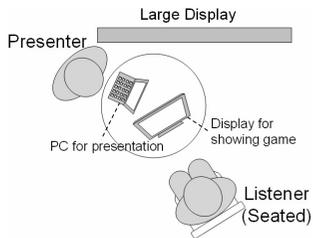


図4 「対面会議」条件における設置  
Fig. 4 Layout for “face-to-face” mode experiment.

いる”状態であるか“聞いていない”状態であるかを推測することである。発表者には、各試行後に推測結果を回答させた。聴講者（遠隔ユーザ）の状態表出が顕著で、かつ発表者による状態把握が容易であれば、この推測の精度は良くなると考えられる。発表者にはゲームの画面は見えずその順序も知らされないため、聴講者を観察する以外の方法で推測することはできない。

なお、聴講者にはすべての条件で発話しないよう指示した。また、各参加者には互いに手を振るといった、明示的に会話チャンネルを確認するような会話制御コミュニケーションをとらないよう指示した。これらの指示は条件の統制をとるためであるが、通常の会議でも、後者のような会話制御コミュニケーションを行うことは少ないためである。

発表用大型ディスプレイは対角 50 インチのサイズで、ディスプレイと、ロボット・人・カメラ・モニタの距離は約 1.7m とした。

#### 4.3 比較条件

##### 4.3.1 「対面会議」条件

「対面会議」条件では、聴講者は発表者の前に座り、発表を直接聴講する（図4）。ゲームは聴講者の前に置かれた 17 インチの液晶ディスプレイに表示される。このディスプレイは、画面が発表者に見えず、また聴講者から見て発表スライドを表示する大型ディスプレイを隠さない位置に設置した。

##### 4.3.2 「ロボット会議」条件

「ロボット会議」条件では、聴講者は 3.1 節で述べたシステムを用いて、発表者が直接見えない場所で発表を聞く（図5）。聴講者は HMD に表示される画像を通して、発表者および発表スライドが表示される大型ディスプレイを見ることができ、イヤホンを通して発表者の声を聞くことができる。一方、発表者は聴講者の様子を直接見ることはできないがロボットの動作により頭の動きを知ることができる。聴講者は発話できない（聴講者側の音声は発表者側に転送されない）。

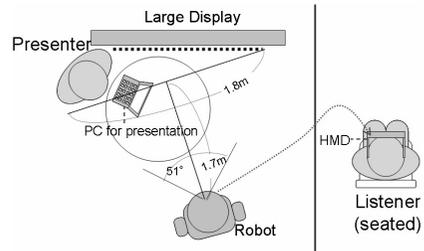


図5 「ロボット会議」条件。点線は映像データの流れを表す  
Fig. 5 Layout for “robot tele-conferencing” mode experiment. Dotted line shows the flow of image data.

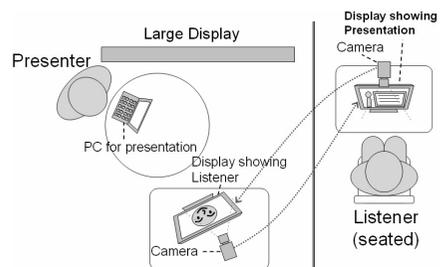


図6 「ビデオ会議」条件。点線は映像データの流れを表す  
Fig. 6 Layout for “Video tele-conferencing” mode experiment. Dotted line shows the flow of image data.

ゲームは HMD の画面に表示され、その際は発表者側の映像はゲームの画面で完全に隠れるため見ることはできない。一方、音声は聞こえたままである。これは、対面条件との整合のためである。

HMD の視野に入る画像はロボットのカメラ視野角（51度）で決まる。ロボットカメラから大型ディスプレイまでの距離（約 1.7m）の位置では、ロボットのカメラは約 1.8m 幅をカバーするので、発表者から大型ディスプレイの表示エリアの両端を含めた 1.7m 程度の幅は、あるカメラ向きにおいて、両者はロボットカメラの視野中に十分収まる。しかしながら、実験中の観察によれば、聴講者は、対象（発表者または大型ディスプレイ）を視野の中心に置くためと思われる首振りの動作が見られた。またそれより頻度は少ないが、ゲームタスク時にも頭を動かす動作が観察された。

##### 4.3.3 「ビデオ会議」条件

「ビデオ会議」条件では、聴講者はビデオチャットシステムを通じて発表を聴講する（図6）。聴講者は発表者側に設置されたカメラからの映像により、発表者および発表スライドが表示される大型ディスプレイを、同時に画面内に見ることができ、発表者の声を聞くことができる。一方、発表者は聴講者側に設置されたカメラからの映像により聴講者の顔を見ることができ、発表者側の画面に映る聴講者の顔の大きさが人

表 1 被験者組 (A-H) と発表題材 (m1-m4) の対応

Table 1 Presentation materials assigned to each subjects.

Subject pair	Inexperienced	A	B	C	D
	Experienced	E	F	G	H
Material		m1	m2	m3	m4

間の顔の大きさとほぼ等しくなるよう設定した。聴講者側の音声は発表者側に転送されない。ゲームは聴講者の前のディスプレイに表示され、その際は発表者側の映像はゲームの画面で完全に隠れるため見ることができない。一方、音声は聞こえたままである。これは、対面条件との整合のためである。

#### 4.4 被験者および実験条件

「ビデオ会議」条件と「ロボット会議」条件の比較において、システムを使用した経験による“慣れ”の影響を観察するため、使用経験者 4 組 (A, B, C, D) と未経験者 4 組 (E, F, G, H) の計 8 組の被験者を準備した。経験者は事前に同様の実験を行っており、各条件での発表/聴講を経験している。各被験者の用いた発表題材 (materials) は表 1 のとおりである。実験の被験者は 15 名が男性で 1 名が女性であった。被験者は、情報科学もしくは機械工学を専攻する大学生または大学院生で、年齢は 20 歳代前半である。発表者として参加する被験者はある程度発表に慣れていることが望ましいため全員大学院生とした。

## 5. 結果

### 5.1 聴講推測の正解率

各条件および経験者、非経験者のグループにおける「聞いている/いない」のアンケートによる聴講推測の正解率は、図 7 に示すとおりであった。それぞれのグループに対して、各条件の正答率の比率の差の検定を対応ある多重比較に適したテューキー (Tukey) の HSD 検定により、検定を行った (以下、各条件を「対面」、「ロボット」、「ビデオ」と短縮表記する)。

その結果、全体では、対面 > ロボット ( $p < 0.05$ ), 対面 > ビデオ ( $p < 0.01$ ), ロボット > ビデオ ( $p < 0.01$ ) で有意差が認められた。経験者と非経験者のグループでの差異を調べると、経験者では、ロボット > ビデオ ( $p < 0.01$ ), 対面 > ビデオ ( $p < 0.01$ ) の有意差 (有意水準 1%) が認められたが、対面 > ロボット ( $p < 0.07$ ) では有意差が認められなかった。非経験者では対面 > ビデオ ( $p < 0.05$ ) で、有意差 (有意水準 5%) が認められるが、そのほかは有意差が

等分散性は、クラスカル-ウォリス検定で 1% 有意差あり。分散のない水準があり、参考値として示す。

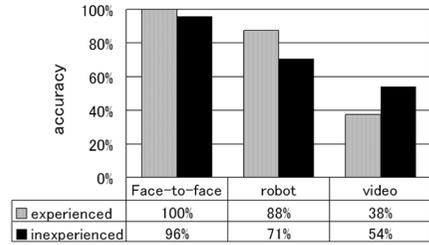


図 7 条件ごとの聴講推測の正解率 (灰: 使用経験者, 黒: 未経験者)

Fig. 7 Estimation accuracy of listening/non-listening for each condition.

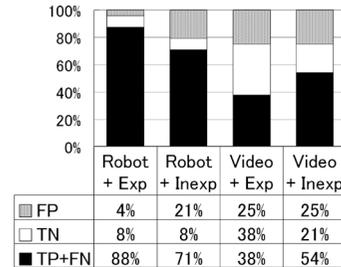


図 8 ロボット条件とビデオ条件における使用経験者と未経験者の推測誤りの傾向 (TP: true-positive, FN: false-negative, TN: true-negative, FP: false-positive)

Fig. 8 Estimation error tendency.

認められなかった。

図 8 には、ロボット条件とビデオ条件における推測誤りの状況をグラフで示す。これによるとビデオ条件について、経験者も非経験者も、True-Negative (TN, 聴講を非聴講と誤った) と False-Positive (FP, 非聴講を聴講と誤った) がほぼ同数となり同じ程度の間違いをしていると考えられる。一方、ロボット条件では、経験者と非経験者に多少差があり、非経験者では FP が多少大きい傾向がある。

### 5.2 発表者の注視頻度

各システム使用時のユーザ行動を分析するため、発表者が聴講者を注視する回数を計測した。計測は、実験を撮影したビデオを用い、手作業により発表者が聴講者の方向に視線を送った回数を数えた。各被験者に対し、各条件における 6 回のプレゼンにおいて注視頻度を計測し、条件ごとに平均および標準偏差を計算したものを図 9 に示す。また、各被験者ごとに注視頻度を正規化し、その値を、各条件ごとに平均化したものを同図右に示す。この正規化された頻度について、条件間の注視頻度の差をテューキー (Tukey) の HSD 検定により有意水準 1% 以下で行った。その結果、ロボット < ビデオ ( $p < 0.01$ ), 対面 < ビデオ ( $p < 0.01$ ) で、有意差が認められたが、対面とロボットでは有意

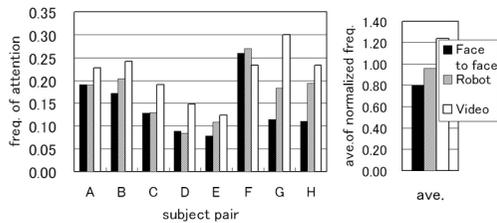


図9 被験者ごとの聴講者に対する注視頻度(回/秒)(左)と正規化した頻度の平均(右)

Fig. 9 Frequency of eye-gaze given to listener (freq./sec, left) and its normalized average (right).

差は認められなかった ( $p = 0.11$ ).

なお、被験者ごとの様子を見ると、図9左のように、多くの場合、対面  $\leq$  ロボット  $<$  ビデオの傾向がみられるが、被験者ペアによって、注視頻度の逆転が起きているケースもあることが観察されている。

### 5.3 主観的評価

実験後、発表者に対してアンケートをとった。「A: プレゼンテーションのしやすさ」、「B: 聴講者が聞いているように感じられたか」、「C: 聴講者のプレゼンへの興味を感じ取ることができたか」という3つの質問に対して7段階評価(1~7)で回答してもらった。その結果の中央値を図10に示す。アンケートは各被験者に対してすべての試行が終了してから実施した。

各項目に対し、条件間の差に関して、Steel-Dwassの方法でノンパラメトリックな多重比較による検定(有意水準5%以下)を行った。「A: プレゼンテーションのしやすさ」に対しては、対面  $>$  ロボット ( $p < 0.05$ )、ロボット  $>$  ビデオ ( $p < 0.05$ )、および対面  $>$  ビデオ ( $p < 0.01$ )の優位差が得られた。一方、「B: 聞いているように感じられたか」に対しては、対面  $>$  ロボット ( $p < 0.05$ )、ロボット  $>$  ビデオ ( $p < 0.05$ )、対面  $>$  ビデオ ( $p < 0.01$ )の優位差が得られた。しかし、「C: 興味を感じ取ることができたか」に対しては、対面  $>$  ビデオ間での有意差 ( $p < 0.05$ )のみ得られ、ビデオとロボット条件間での差異はみられなかった。

アンケートの自由記述によると、ビデオ条件では視線の読み取りが困難であったのに対し、ロボット条件では発表への反応(動き)がみられるので話しやすいというたぐいのコメントが8名中5名あった。ビデオ条件よりも視線に関わる頭部動き情報を視認しやすく、プレゼンテーションがしやすかったと考えられる。

### 5.4 ビデオ観察

実験を聴講者やロボットの後ろから記録したビデオを観察して、ロボット会議条件における、ロボットの目立つ動きを調べたところ、聴講しているときと非聴講(別タスクをしているとき)では、聴講しているとき

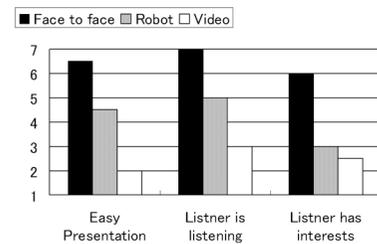


図10 アンケート結果の集計: 7段階評価の中央値(左: プレゼンテーションのしやすさ, 中: 聴講者が聞いているように感じたか, 右: 聴講者の興味を感じ取ることができたか)

Fig. 10 Summary of questionnaire: median.

きのほうが頻度が大きい傾向がみられた。非聴講の場合でも、動きはあったが、少ない傾向であった。また、対面会議条件においては、予想に反して、発表中に頭をほとんど動かさない状況も観察された。むしろ、リラックスした姿勢で発表を聞いているか、ゲームに熱中している姿勢かにより、後ろからでも発表に注目しているか、あるいは別タスクを実行しているかの区別はある程度推測できる状況であった。

## 6. 考 察

### 6.1 聴講推測の正解率

まず、推測の正解率の結果に関して考察する。全体で得られた有意差検定の結果から、対面、ロボット、ビデオの順に推測の正解率が高く、提案システムがビデオ会議と比較して有効であることが示された。しかし、ロボット条件の正解率が対面条件に比較して低いことから、対面と同等の情報が伝達されているとはいえない。これには、提案システムでは聴講者の表情や目の動き、あるいは姿勢を伝達できないことが要因であると考えられる。また、推測誤りにおいて、TNとFPの割合は全体としては大きな差異はなく、ランダムな要因により発生していると考えられる。未経験者のロボット条件においては、FPが少し多めの傾向であり、任意の頭の動きを手がかりとして、聴講と誤判定している可能性があるが、経験者ではFPとTNはほぼ同じとなり、頭の動かし方の差異から何らかの情報を得て推定している可能性がある。このように、ロボット条件においては、経験者のグループで推測率が高くなる傾向があり使用経験による学習効果を期待できる可能性が示唆される。

### 6.2 発表者の注視頻度

次に、発表者の注視頻度の計測結果に関して考察する。本実験では、発表者に対して聴講者を観察し聞いているかどうかを推測するよう指示をしたため、発表者の聴講者に対する注視は、聴講者の状態を把握する

ために行われたと考えられる。よって、注視の回数が多ければ多いほど、相手の状態を把握するために多くの注視が必要であったと考えられ、状態の把握がより困難であったと考えられる。

5.2節の結果から、注視頻度について全体として、ビデオ > 対面、ロボット > 対面、ビデオ > ロボットについて有意差が得られたことから、ビデオ > ロボット > 対面の順で注視頻度が高い傾向があるといえる。特に、図9のグラフから分かるように、8組中7組でビデオ条件の時の注視頻度が最も高く、注視頻度が高くなる傾向が顕著であるといえる。逆にいえば、ビデオより、ロボットの方が状態の把握が容易であったと推測できる。

### 6.3 主観的評価

アンケートAの結果から、対面、ロボット、ビデオの順でプレゼンがしやすかったことが分かる。「プレゼンのしやすさ」はユーザがその条件においていかに話しやすかったかを表している。被験者は自然な条件である対面を基準に他の条件を比較すると考えられ、現に対面では平均が満点の7に近い値となっている。ロボット条件はビデオ条件との比較では有意に高い評価が得られ、ビデオ対話と比較して提案システムが有効であることが確かめられた。しかし、ロボット条件と対面条件との比較でも有意差が認められ、提案システムはまだ対面での対話には及ばないことが分かる。

アンケートBの結果から、対面、ロボット、ビデオの順で聴講者がより聞いているように発表者が感じたことが分かる。この結果でも、対面では平均スコアが満点の7に近い値になっており、被験者は対面の時を基準に評価していると考えられる。ロボット条件はビデオ条件との比較で有意に高いと認められ、ビデオ対話と比較して提案システムでは、発表者は聴講者がより聞いているように感じられることが確かめられた。これには、次の2つの要因が考えられる。第1に、ロボット条件では発表者はロボットと目を合わせることができるが、ビデオ条件では目を合わせられず、また視線が画面に固定されてしまいどこに集中しているのか分かりにくい。第2に、ロボット条件ではロボットがオブジェクトとして発表者の近くに存在しているため、相手が聞いてくれているという感覚が得やすい。

これらの傾向は、次のアンケートの自由記述からも支持される。ビデオ条件では「表情（または視線）の読み取りが困難であった」という意見が複数みられ、ロボット条件では「プレゼンテーションへの反応（動き）がみられるので話しやすい」という類のコメントが8名中5名あった。また、ロボット条件でも「表情

が見えないため聞いているかどうか判別しにくい」というものがあり、対面条件に対して、ロボット条件にも不足する要因がある。

しかし推測の正解率の結果から、表情の見えるビデオ条件においても推測が困難であることが分かっており、ユーザは顔の向きの付加的な情報として表情を解釈しているとも考えられる。すなわち、表情だけでは相手の状態を把握するのは困難であるが、顔の向きや注視対象が分かる状況では表情の情報が意味を持つと考えるとつじつまが合う。この点については今後の検証が必要である。

アンケートCについては、ビデオとロボットで有意差がなく、興味を持って聞いているという意味的な情報の伝達には、ロボットもビデオもまだ不十分であるといえる。

## 7. ま と め

本稿では、ロボット会議において、遠隔ユーザの興味などの非言語情報をロボットを介して表現できるかどうかを、実験システムを構築し、被験者実験によって調べた。聴講者が発表者のプレゼンテーションに聞き入っているかどうかを、頭の動作で推測できるという仮説を立てて、ロボット会議、ビデオ会議、対面会議の3条件で比較して実験した。実験の結果、聴講状態の推測精度と注視頻度の分析から、ロボット条件がビデオ条件よりも有効に遠隔参加者の聴講状態を伝達可能であることを示した。しかし、ロボット条件と対面条件との比較から、ロボット条件は対面条件には及んでいないことも示された。これには、ロボット条件では姿勢、表情、視線などの他のモダリティが伝達できていないことが主な原因と考えられる。ロボット条件と対面条件では、頭の動きは同等であるという前提で実験を進めたが、意外にも対面条件で顕著な頭の動きはみられないにもかかわらず、ロボット条件ではディスプレイと発表者を注視するために頭が動くことが観察された。視野が自由な対面では、リラックスして頭を動かさずにタスクをこなすのに対し、視野が制限されたロボット条件では、頭を動かす動作がかえって強調されて表出されたため、結局それを認識して発表者が聴講者の状態を比較的正確に推定できたと考えられる。

聴講状態の推測には、姿勢、頭部のジェスチャ（向きとタイミング）、表情、視線などの複数の要因があり、発表者は、原稿の読み上げの合間に中心視や周辺視によって、それらの状況を把握・統合しているモデルを前提にできる。実験では、ロボット条件と対面条

件では実体があることが有利に働いたため、姿勢や頭部ジェスチャを周辺視で確認すれば、注視の頻度が小さくても正しく聴講を推測できたと考えられる。一方、ビデオ条件では、ロボット条件に比べれば表情や視線情報の伝達が容易にもかかわらず、2次元平面に押し込まれた表情や視線の情報は、周辺視では聴講の推測に十分でなく、注視頻度が増えたものと考えられる。

プレゼンテーションの効果としては、聴講者側の意見も重要である。事後アンケートによれば、視線が合うなどの好意的な意見もあれば、下記の遅延に起因する使いにくさの指摘もあった。また、ロボットが遠隔の参加者の存在感を伝達しているかどうか、興味深いテーマであり、今後研究していく必要がある。アンケートからはロボットの人格性に連関して、「人を相手にしている感じがしない」という否定的な意見や「うなずいてくれる」「反応してくれる」などの擬人的人格を認めるような表現を使った意見があり、今後の研究の手がかりになると思われる。

また、実験で用いたロボットには、頭部の慣性モーメントとモータ制御のために動作遅延が平均 1.1 秒あることが分かっている。そのため遠隔ユーザは、HMDに呈示される映像が遅れるのを補償するため、頭の動きを抑制している傾向が観察されている。動作遅延をゼロにすることは困難であるが、何らかの動作遅延補償をシステムに組み込むことで、動きの抑制が解消できる。動作遅延補償については、全方位カメラ画像を組み合わせた手法が検討されており<sup>21)</sup>。今後は、この手法により頭の動きの抑制を解消したシステムを用いて、再実験をして今回観察されたことの検証をする必要がある。

謝辞 本稿をまとめるにあたって、統計的検定についてアドバイスをいただいた(株)リコー研究開発本部 杉原敏昭氏に感謝する。また、討論していただいた名古屋大学間瀬研究室、NTT コミュニケーション科学研究所の皆様にも感謝する。本研究の一部は、科学研究費補助金(課題番号:18300048)によった。

## 参 考 文 献

- 1) Vertegaal, R.: The GAZE Groupware System: Mediating Joint Attention in Multiparty Communication and Collaboration, *ACM CHI'99 Conference on Human Factors in Computing Systems*, pp.294-301 (1999).
- 2) Brand, S.: *The Media Lab: Inventing the Future at MIT*, Penguin Book (1988).
- 3) Ishibiki, C., Toh, I. and Miwa, Y.: Development of Eye-ball Robot Serving as a Bodily Me-

dia and its Applicability as a Communication Means, *2001 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Vol.2, pp.1177-1182 (2001).

- 4) Kuzuoka, H., Oyama, S., Yamazaki, K., Suzuki, K. and Mitsuishi, M.: GestureMan: A Mobile Robot that Embodies a Remote Instructor's Actions, *CSCW'00: Proc. 2000 ACM conference on Computer Supported Cooperative Work*, New York, NY, USA, pp.155-162, ACM Press (2000).
- 5) 葛岡英明, 山崎敬一, 上坂純一: ロボットを介した遠隔コミュニケーションシステムにおけるエコロジーの二重性の解決: 頭部運動と遠隔インタフェースの評価, *情報処理学会論文誌*, Vol.46, No.1, pp.187-196 (2005).
- 6) 岡田謙一, 松下 温: 臨場感のある多地点テレビ会議システム: MAJIC, *情報処理学会論文誌*, Vol.36, No.3, pp.775-783 (1995).
- 7) Paulos, E. and Canny, J.F.: PRoP: Personal Roving Presence, *CHI'98*, pp.296-303 (1998).
- 8) Jouppe, N.P.: First Steps Towards Mutually-immersive Mobile Telepresence Immersion in the World, *Proc. ACM CSCW'02 Conference on Computer-Supported Cooperative Work 2002*, pp.354-363 (2002).
- 9) Gemmell, J., Toyama, K., Zitnick, C.L., Kang, T. and Seitz, S.: Gaze Awareness for Videoconferencing: A Software Approach, *IEEE Multimedia*, Vol.7, No.4, pp.26-35 (2000).
- 10) Vertegaal, R., Weevers, I., Sohn, C. and Cheung, C.: GAZE-2: Conveying Eye Contact in Group Video Conferencing Using Eye-Controlled Camera Direction, *CHI 2003 Conference on Human Factors in Computing Systems*, pp.521-528 (2003).
- 11) Morikawa, O. and Maesako, T.: HyperMirror: Toward Pleasant-to-use Video Mediated Communication System, *CSCW'98*, pp.149-158 (1998).
- 12) 平田圭二, 原田康徳, 高田敏弘, 青柳滋己, 白井良成, 山下直美, 大和淳司: 遠隔コラボレーションシステム t-Room の中央テーブル, *ヒューマンインタフェース学会誌*, Vol.9, No.1, p.83 (2007).
- 13) Matsui, D., Minato, T., MacDorman, K.F. and Ishiguro, H.: Generating Natural Motion in an Android by Mapping Human Motion, *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp.1089-1096 (2005).
- 14) 坂本大介, 神田崇行, 小野哲雄, 石黒 浩, 萩田紀博: 遠隔存在感メディアとしてのアンドロイド・ロボットの可能性, *インタラクシオン 2007*, 情報処理学会 (2007).

- 15) 三輪敬之, 石引 力: 場の創出に影を活用する共存コミュニケーションシステムの開発, *インタラクシオン* 2004, pp.255-262 (2004).
- 16) Leigh, J., Johnson, A.E., Vasilakis, C.A. and DeFanti, T.A.: Multi-perspective Collaborative Design in Persistent Networked Virtual Environments, *Virtual Reality Annual International Symposium 1996*, pp.253-260 (1996).
- 17) Kurashima, C.S., Yang, R. and Lastra, A.: Combining Approximate Geometry with View-Dependent Texture Mapping – A Hybrid Approach to 3D Teleconferencing, *Brazilian Symposium on Computer Graphics and Image Processing 2002*, pp.112-119 (2002).
- 18) 大崎浩司, 渡辺富夫, 山本倫也: InterActor に手指動作入力を併用した身体的インタラクシオンシステムの開発, *ヒューマンインタフェース学会論文誌*, Vol.7, No.3, pp.89-98 (2005).
- 19) Kidd, C.D. and Breazeal, C.: Effect of a Robot on User Perceptions, *IEEE/RSJ International Conf. on Intelligent Robots and Systems*, pp.3559-3564 (2004).
- 20) Robovie-R ver.2. <http://vstone.co.jp/top/products/robot/Robovie R.html>
- 21) 森田友幸, 平野 靖, 梶田将司, 間瀬健二: ロボットを用いた遠隔コミュニケーションのための視覚システム: 広視野視覚合成と遅延補償提示, 第3回電子情報通信学会ネットワークロボット時限研究会 (2007).

(平成 19 年 4 月 9 日受付)

(平成 19 年 9 月 3 日採録)



森田 友幸

2007 年名古屋大学大学院情報科学研究科社会システム情報学専攻修士課程修了。同年 KDDI 入社。



間瀬 健二 (正会員)

1981 年名古屋大学大学院工学研究科情報工学専攻修士課程修了。同年 NTT 入社。1995~2002 年(株)国際電気通信基礎技術研究所研究室長。2002 年より名古屋大学情報連携基盤センター教授。コンピュータによるコミュニケーション支援の研究を推進している。IEEE, ACM, 電子情報通信学会, 日本 VR 学会, 画像電子学会各会員。博士(工学)。



平野 靖

1999 年名古屋大学大学院情報工学専攻修了。博士(工学)。同年日本学術振興会特別研究員(PD), 2000 年名古屋大学大学院情報工学専攻助手, 2002 年同大学情報連携基盤センター助手, 2004 年同センター助教授, 現在に至る。画像処理とその医用への応用, および大学間認証基盤に関する研究に従事。1998 年日本医用画像工学会論文賞, 2002 年日本医用画像工学会奨励賞。電子情報通信学会, 日本医用画像工学会, コンピュータ支援画像診断学会, 日本生体医工学会, IEEE 各会員。



梶田 将司 (正会員)

1990 年名古屋大学工学部情報工学科卒業。1995 年同大学大学院工学研究科情報工学専攻博士課程満了, 博士(工学)。2002 年名古屋大学情報連携基盤センター助教授, 現在に至る。大学における教育・研究活動での IT 活用に関する研究開発に従事。1998 年日本音響学会第 15 回粟屋潔学術奨励賞, 2001 年電子情報通信学会第 56 回論文賞。電子情報通信学会, 日本音響学会, 日本教育工学会, IEEE 各会員。



岡留 剛 (正会員)

1988 年東京大学大学院情報科学科博士課程修了。センサネットワークを利用した意味処理技術・実世界のモノとコトのコンテンツ化等の研究に従事。博士(理学)。現在, NTT コミュニケーション科学基礎研究所主幹研究員。ACM, 日本認知科学会各会員。