

身体の動きをともなう遠隔協調作業における 上半身映像の効果

山下直美^{†1} 葛岡英明^{†2} 平田圭二^{†1}
青柳滋己^{†1} 白井良成^{†1}
梶克彦^{†1} 原田康徳^{†1}

対面協調作業においてテーブル面の共有が有効であることが示されており、遠隔地間においても同様な環境を実現することが望まれている。近年、そのような環境の実現を目指す研究が多数見受けられる一方で、テーブル面に加えて遠隔ユーザの上半身映像も投影すべきか否かの明確な設計指針はほとんど見当たらない。従来研究の共通認識は、音声通信だけでも多種多様な情報伝達が可能であり、上半身映像は基本的にあまり付加価値を生んでいないというものであった。本研究では、テーブル面を共有して行う作業の利点である「多人数（1地点に複数ユーザが参加）による動きをともなう協調作業が可能」という点に注目し、上半身映像の有無がテーブル上の作業にどのような影響を与えるかを実験により調べた。実験データを分析した結果、上半身映像を投影した場合、ユーザ相互間で上半身映像が次の動作や作業への予測や準備に寄与し、作業効率が有意に向上した。

Effects of Showing User's Upper Body in Video-mediated Collaboration

NAOMI YAMASHITA,^{†1} HIDEAKI KUZUOKA,^{†2}
KEIJI HIRATA,^{†1} SHIGEMI AOYAGI,^{†1} YOSHINARI SHIRAI,^{†1}
KATSUHIKO KAJI^{†1} and YASUNORI HARADA^{†1}

Previous studies have significantly demonstrated the values of sharing a tabletop among collocated collaborators. A tabletop interface that realizes a similar environment between distant collaborators is desired. In designing such a system, sharing a tabletop (workspace) among distant users is obviously important. However, it is still unobvious whether the upper body of distant users is worth showing. In fact, many studies have shown that upper body images of distant collaborators do not add much value to their work. In this study, we examined

whether showing the upper body of distant collaborators is valuable for group work that requires body movement during their work. A comparative study with and without upper body images indicated that group work were more efficient when upper body images were shown. We infer that upper body images helped them predict and brace for distant collaborator's subsequent movement.

1. はじめに

対面協調作業においてテーブル上面の作業領域の共有は有効であることが示されており¹⁸⁾、遠隔ビデオコミュニケーションにおいても同様な共有環境を実現することが望ましいと考えられている^{6),10),19)}。テーブル面に加えて、遠隔地間でユーザの上半身映像をお互いに関連する機能を持ったシステムも提案されている^{10),12),23)}。

遠隔ユーザの上半身映像を閲覧できるシステムでは、遠隔ユーザの表情や視線、頭の動きや上半身の体の向きなどの情報を伝達することで、遠隔ユーザの発話意図をより正確に理解できる効果が期待できる。

しかし、上半身映像の投影場所を誤ると、かえって協調作業の効率が落ちることが報告されている³⁾。たとえば、上半身映像とテーブル面に投影された映像との関連性が直感的に理解できない形で表示されると、遠隔ユーザの身体行為（視線やジェスチャなど）の意味を正しく理解することができないという問題が指摘されている⁴⁾。また、複数の映像が投影されると、ユーザが目移りしてしまって協調作業に集中できないという問題も指摘されている¹⁾。

したがって、テーブル面映像と上半身映像の両方を表示する場合には、これらの映像の物理的な位置関係が直感的に理解でき、遠隔ユーザの身体性になるべく損なわれないようにディスプレイを配置することが重要である。うまくディスプレイを配置できれば遠隔ユーザの身体行為を周辺環境と関連付けて理解することが可能になる。Agora¹⁰⁾やWesugiらの空間共有テーブルシステム²³⁾は、遠隔ユーザのジェスチャが連続して見えるように、テーブル面の水平ディスプレイと上半身を投影する垂直ディスプレイを連続するように配置している。

一方で、この連続的なディスプレイ配置は、本来のテーブル共有の利点である複数ユーザ

^{†1} 日本電信電話株式会社コミュニケーション科学基礎研究所
Communication Science Laboratories, NTT Corporation

^{†2} 筑波大学大学院システム情報工学研究科
Graduate School of Systems and Information Engineering, University of Tsukuba

による自由な位置からの作業参加¹⁶⁾や協調活動中の場所移動¹⁸⁾を妨げる(すなわち, 協調作業位置を固定してしまう)という欠点がある. 上半身映像は, 1 地点あたりの参加者が複数人の場合や身体配置に変化が生じるような活動において特に効果的と考えられるが, そうした条件における効果を調査した例は存在しない.

本論文では, 著者らが開発してきた遠隔協調作業支援システム t-Room を用いて, 1 地点に複数名の参加者がいて, 各参加者がテーブルの周囲を自由に移動しながら協調作業をするような状況において, 上半身映像が何に対してどの程度役に立つかを調べる. t-Room は, 遠隔ユーザの身体性の再現と, 身体の移動をともなう協調作業の両立を目指して, テーブル面と上半身投影ディスプレイが連続するように配置されたビデオコミュニケーションシステムである. t-Room を介した遠隔協調作業において, テーブル面と上半身映像の両方を共有した場合とテーブル面のみ共有した場合を比較し, 遠隔地間でテーブル面の共有に加えて上半身映像を投影することの得失を考察する.

本論文の構成は以下のとおりである. 2 章では先行研究に基づいて上半身映像の有無が遠隔協調作業に及ぼす影響について仮説を立てる. 3 章では仮説の検証に用いる遠隔協調作業支援システム “t-Room” を紹介し, 4 章で t-Room を用いた被験者実験の説明をする. 5 章で実験の分析結果を述べ, 6 章で実験より得られた知見の考察を行う. 最後の 7 章でまとめを述べる.

2. 関連研究と仮説

テーブルを共有した対面型協調作業の特徴として, テーブルの周囲に複数の参加者が自由に身体を位置とることができるといこと¹⁶⁾や, テーブルへの参加の仕方が時間とともに動的に変化すること¹⁸⁾が指摘されている. したがって, テーブル共有型の遠隔協調作業を支援するシステムでもこうした特徴が支援できる必要があり, そのようなシステムが効果的に利用されるために必要とされる機能を明らかにすることは重要であると考えられる.

そこで本論文では, テーブル共有型の遠隔協調作業システムにおいて, 1 地点に複数名の参加者がいて, 各参加者がテーブルの周囲を自由に移動しながら協調作業をするような状況を想定する. そしてそのような環境において, 上半身映像の有無が遠隔コラボレーションに及ぼす影響を調べることを目的とする. 具体的には, 上半身映像の有無が遠隔ユーザ間の立ち位置, 呼びかけ行為, 発話効率, そして作業効率に及ぼす影響を調べる. 以下では, これらに関する仮説を, 従来研究に基づいて導出する.

2.1 立ち位置に関する仮説

まず, 遠隔テレビ会議システムに関する従来研究により, ユーザは同室ユーザよりも遠隔ユーザの投影されているディスプレイを注視する傾向があり, 遠隔ユーザの投影ディスプレイ位置を変えると, ユーザは遠隔ユーザを見やすい方向に体の向きを変えることが報告されている²⁴⁾.

これまで提案されてきたテーブル面と上半身映像の両方を投影することを目指したシステムでは, 協調作業中にユーザが移動することを想定せず, 遠隔ユーザの投影ディスプレイをテーブル面とつなげて対面位置に固定することが多かった^{10);23)}. しかしもしユーザが協調作業中に自由に動き回ることが可能になれば, ユーザはつねに遠隔ユーザの様子が確認できる位置になるように立ち位置を調整しながら協調作業を進めることが予想される.

したがって, 遠隔ユーザの上半身映像が見えるとき, 作業のしやすさを優先した立ち位置よりも遠隔ユーザの映像位置に合わせて立ち位置を変えることが予想される.

仮説 1. 遠隔ユーザの上半身映像が見える実験条件では, 見えない実験条件と比較して, 遠隔ユーザの様子を確認可能な位置に移動する傾向がある.

2.2 明示的な発話交代に関する仮説

従来研究により, 遠隔テレビ会議システムを用いた会議では, 対面会議と比較して, 遠隔ユーザの名前を呼んでから発話するケース (explicit handover) が多いことが報告されている¹⁵⁾. これは, 通常のテレビ会議システムでは頭や視線の向きが正しく伝わらず, 誰に対する発話かが伝わりにくいためと考えられている²¹⁾. 円滑な発話交代を行ううえで視線情報は特に重要と考えられており, これまでに遠隔地間で視線の向きを正しく伝えることを目的としたシステムが複数提案されている^{17);22)}. しかし, これらのシステムの大半は 1 地点につき 1 人のユーザしか許容せず^{*1}, 会議中や協調作業中にユーザが移動することを想定していない.

ここで, 従来研究の結果をふまえ, 本研究で想定する「1 地点に複数名のユーザが参加し協調作業中に移動できる状況」を考える. まず, 1 地点に複数人のユーザがいる遠隔環境では, モナリザ効果^{*2}により頭や視線などの正確な向きを伝えることが困難になることが予想される. しかしその一方で, 立ち位置の自由度が与えられることによって, ユーザは遠隔ユーザとの距離や立ち位置によって誰に対する発話かを伝達したり推測したりすることが可

*1 Multiview は 1 地点に複数のユーザの参加を許容する¹⁴⁾.

*2 モナリザ効果とは, 人の正面顔の二次元表示画像は, どの角度から見ても閲覧者自身を向いているように見える認知的な現象のことである.

能になる⁷⁾。たとえば、指示者と複数の作業者がいる環境において、指示者がある特定の作業者に十分に近づいてから指示を行えば、指示者がどの作業者に指示を出しているかは明白である。

このように、遠隔ユーザの上半身映像が見える場合、上半身映像が見えない場合と比較して遠隔ユーザとの位置関係や移動にともなう動きが観察しやすいため、誰に対する発話かが伝達しやすくなることが予想される。したがって、遠隔ユーザの上半身映像が見える場合、上半身映像が見えない場合よりも、誰に対する発話かを明示的に伝えずに話し始めるケースが多いことが予想される。

仮説 2. 遠隔ユーザの上半身映像が見える実験条件では、立ち位置を適宜変えることによって誰に対する発話かを明示することができるので、見えない実験条件と比較して、遠隔ユーザの名前を呼んでから発話することが少ない。

2.3 効率に関する仮説

最後に、従来の遠隔作業指示などに関する研究では、テーブル面 (working space) の共有は重要であるが、遠隔ユーザの上半身映像はあまり重要視されていない⁵⁾。これは、遠隔ユーザの上半身映像が見えなくても、声 (発話内容) をやりとりし、テーブル面の映像 (作業者が作業する様子や指示者がテーブル上の物にジェスチャする様子) を観察することによって、遠隔ユーザの様子をおおむね把握できるためである。一方、最近の研究 (Luff や葛岡らの Agora を用いた遠隔協調作業実験) では、上半身映像とテーブル映像をつなげて連続な一面に見えるよう表示することが、上半身映像がテーブル上のジェスチャの予測に寄与したと報告されている¹¹⁾。したがって、上半身映像とテーブル映像の物理的な位置関係を一貫して表示すると、上半身映像が遠隔ユーザのジェスチャ理解度を向上させると期待できる。

本研究で想定する「1 地点に複数名のユーザが参加し協調作業中に移動できる状況」では、ジェスチャを行うための準備動作 (たとえば、少し離れたテーブル上の物を指差すために、テーブルの周りを移動する動作や、テーブルの物体に対してジェスチャする際の腕の動きなど) を観察することによって、どの遠隔ユーザが次にどのような動きをするかの予測が容易になり、遠隔ユーザのジェスチャ理解度がいっそう向上することが期待できる。

以上より、遠隔ユーザの上半身映像が見える場合、上半身映像が見えない場合よりも遠隔ユーザが次にどのような動作をするか予測しやすくなり、その結果として協調作業効率が向上することが予想される。また、遠隔ユーザのジェスチャに対する理解度が向上することで発話効率の向上が予想される。

仮説 3. 遠隔ユーザの上半身映像が見える実験条件では、見えない実験条件と比較して、発話効率と作業効率がともに良い (発話量が少なく、作業時間も短い) 傾向にある。

3. ビデオコミュニケーションシステム t-Room

t-Room は、遠隔ユーザの身体性の再現と身体の移動をともなう協調作業の両立を目指して、テーブル面と上半身投影ディスプレイを配置したビデオコミュニケーションシステムである。t-Room は 1 地点につき複数ユーザを許容し、ユーザが t-Room 内のどこにいても、遠隔ユーザの立ち位置をほぼ正確に伝達できる²⁴⁾。

3.1 ディスプレイとカメラの配置

t-Room は、ユーザのいる空間を囲むように配置された複数のカメラと縦置きディスプレイ、中央のテーブルとそれに投影するプロジェクタと撮影するビデオカメラからなる (図 1)。t-Room にはマイクとスピーカも設置されているが図示していない。複数地点に同一構成の t-Room を設置しそれらの間で映像・音声を通信し合い、遠隔地にいる t-Room 内のユーザ同士で同じ作業空間を共有する。たとえば t-Room の 1 番目の縦置きディスプレイの前に立っているユーザは遠隔 t-Room の 1 番目の縦置きディスプレイに投影される。テーブル上の物体は遠隔テーブル上の同じ位置に 2D 映像として再現される。こうして、同じ t-Room 内の複数ユーザ同士だけでなく遠隔のユーザ同士でも対面環境と同様にテーブルを囲み自由な立ち位置をとることができる。

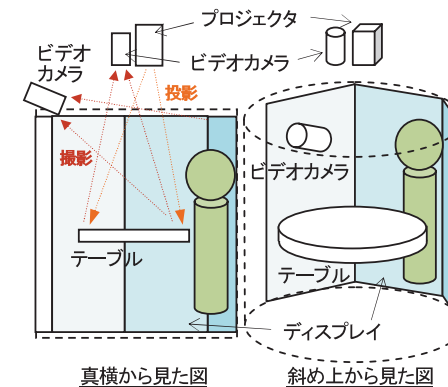


図 1 t-Room の構造とディスプレイ・カメラ配置
Fig. 1 Architecture of t-Room and layout of displays and cameras.

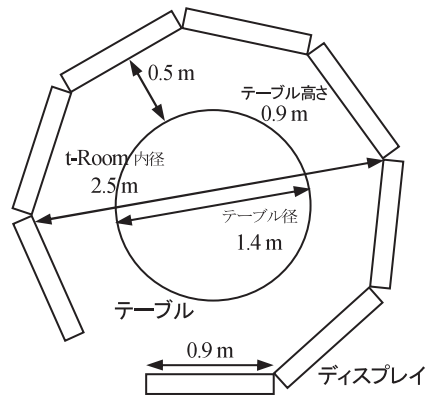


図 2 t-Room の実験環境
Fig.2 t-Room's experimental setup.

Agora や Wesugi らの空間共有テーブルシステムはテーブル用ディスプレイと上半身投影用の縦置きディスプレイを持ち、表示される映像に欠ける部分が生じないようにディスプレイどうしを連続して配置している。これに対し、t-Room ではテーブル用ディスプレイと縦置きディスプレイとが連続して配置されておらず、その代わりに、テーブルと縦置きディスプレイの間のできた空間を人が自由に移動することができる。

3.2 ハードウェア構成

図 2 に t-Room のハードウェア構成を示す。実際の t-Room では、大型液晶ディスプレイ（縦置き 65 インチ、解像度 $1,920 \times 1,080$ ）が 9 角形の 8 辺に配置され、中央に円形の作業用テーブルが置かれている。t-Room の内径は 2.5 メートルである。液晶ディスプレイが配置されていない残り 1 辺は t-Room 内への出入口として利用され、その長さは他の 8 辺より短い。縦置きディスプレイは身長 180 cm のユーザの膝から頭頂部をカバーする程度の高さに置かれる。縦置きディスプレイ 1 面分を、正対する位置のディスプレイ上部に設置した HD ビデオカメラ（解像度 720×480 ）1 台で撮影する。

t-Room 内部中央には円形の作業テーブルを設置した。テーブルの直径は 1.4 メートル、テーブルとディスプレイの間の隙間は 0.5 メートルである。テーブル上面中心の 2 メートル上方には、そのテーブル面を撮影するビデオカメラ 1 台と遠隔テーブル面の映像をそのテーブル面に投影するプロジェクタ 1 台を設置している（図 2）。

縦置きディスプレイを撮影するビデオカメラのレンズ前面には偏光フィルタが取り付けら

れており、液晶ディスプレイの前方に存在する物理的な物体だけを撮影するようになっている。テーブル上方のビデオカメラにも同様の理由で偏光フィルタが取り付けられている。なお、ユーザが液晶ディスプレイ表面に近い位置に立つと、そのユーザの像が遠隔ディスプレイにほぼ等身大の大きさで表示されるようにキャリブレーションしている。実際には、中央テーブルの存在によってユーザの動きはディスプレイ付近に制限されるため、遠隔ユーザはつねにほぼ実寸大で表示される。

ビデオカメラが撮影した映像は IEEE1394 経由で PC に転送され、映像の歪み補正や解像度の変換が行われた後、Motion JPEG にエンコードされ、ディスプレイを制御する PC へ TCP/IP で配信される。音声は各ユーザが装着するワイヤレスマイクで収録され PC に転送される。PC 上で 16 bit、44.1 KHz のリニア PCM にエンコードされ、イヤホン（あるいはスピーカ）を制御する PC に UDP/IP で配信される^{*1}。

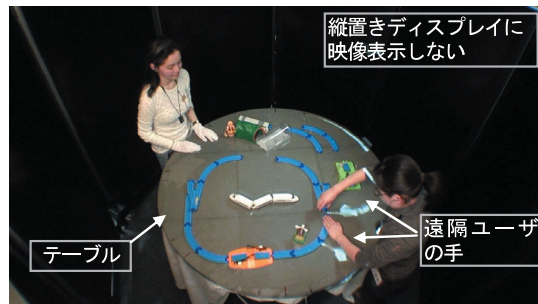
我々は神奈川県と京都府にまったく同じ構造の t-Room を設置し、これら 2 台の t-Room をギガビットネットワークで接続した。映像と音声の遅延は、各々約 0.3~0.4 秒と 0.2~0.3 秒であった。遠隔ユーザの映像位置と音源位置に相違による影響を除くために被験者にイヤホンを装着してもらい、音源から遠隔ユーザの位置を推測できない（遠隔ユーザの位置は映像からのみ把握できる）ようにした。

3.3 遠隔ユーザ間の位置関係や方向性の伝達

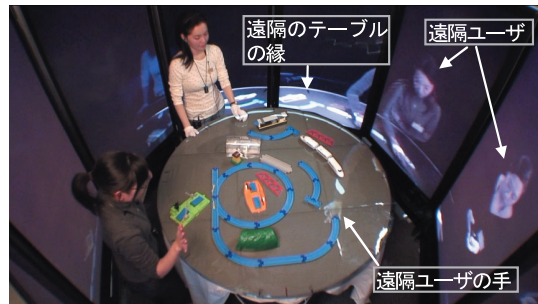
t-Room はディスプレイとカメラによって囲まれた空間になっているため、出入口を除けば、ユーザがどの位置に立っても遠隔 t-Room 内から観測可能である。また、t-Room 内のすべての物体や人が遠隔 t-Room 内の同じ位置に映像として再現されるため、ジェスチャと物体の関係（たとえば、遠隔ユーザが何を指差しているか）を一目で把握することができる。また 2 章では、t-Room の特徴として、複数のユーザが動きを制限されることなく協調作業中にテーブルの周りを移動可能であることを述べたが、移動して遠ざかったり近づいたりする物体や人なども現実世界と同じ動きをするので、把握が容易である。

さらに同様の理由で、t-Room では遠隔地のユーザ同士で大まかに方向感を共有することもできる。遠隔ユーザ間で方向感を共有する従来研究において開発されたシステムでは、モナリザ効果による影響を除去するために 1 地点あたり 1 ユーザの利用を想定していることが多い。これに対し、t-Room では、遠隔とローカルにいるユーザ全員が、環境内にいるほ

*1 スピーカはディスプレイの下部に設置されており、将来的には被験者の動きにともなって音源を移動させる予定であるが、本実験を実施した時点ではこの技術は未完成であった。



(a) メディア条件 1：テーブル面のみ共有した状態



(b) メディア条件 2：テーブル面と上半身映像を共有した状態

図 3 メディア条件

Fig. 3 Media Condition.

ぼすべての人と物体の映像や位置関係を共有できるため、誰が誰（何）を見ているのかをある程度推察することができ、モナリザ効果が軽減されることを確認している²⁴⁾。筆者らの経験では、ローカルユーザ同士の間隔が縦置き 65 インチディスプレイ 1 面分以上空いていれば、遠隔ユーザの視線や頭の向きなどから、誰を見ているのかをおおよそ把握することができる。遠隔ユーザの移動方向に関する認識（時計回りか反時計回りか）は、ユーザの立ち位置にかかわらず正確に伝達できた。

3.4 上半身映像とテーブル面の連続性

Luff らは、Agora を用いた実験を通して、上半身映像とテーブル映像の物理的な位置関係を一貫して表示させることの重要性に加え、上半身映像とテーブル映像のつなぎ目を連続に表示させることの重要性を説いている¹¹⁾。Luff らは、遠隔ユーザの上半身映像を投影

する縦置きディスプレイに、テーブル（ワークスペース）の縁の映像を表示させることで、テーブル面に現れる遠隔ユーザのジェスチャが予測しやすくなる様子が報告されている。

この報告結果を受け、t-Room の側面ディスプレイにもテーブル面の縁の映像を投影した（図 3 (b)）。上半身映像とテーブル面の縁の映像を同じ画面に表示することによって、テーブル面に投影される遠隔ユーザのジェスチャがよりいっそう予測しやすくなり、円滑な協調作業につながることを期待できる。

4. 比較実験

4.1 実験条件

本実験の実験環境に、3.2 節で述べた t-Room を用いた。本実験には、互いに面識のない被験者 32 名（神奈川県に 16 名、京都府に 16 名）が有償で参加した。被験者の移動において性別による影響を排除するために、すべての被験者は女性とした。神奈川県と京都府から各 2 名ずつの合計 4 名で 1 グループとし、合計 8 グループを構成した。各グループは、2 つのメディア条件で 2 回ずつ身体の移動をともなう遠隔作業指示に関するグループ課題に取り組んだ。2 つのメディア条件とは「メディア条件 1：遠隔地間で中央のテーブル映像のみ共有した場合（図 3 (a)）」と、「メディア条件 2：中央テーブルに加えて上半身映像も共有した場合（図 3 (b)）」である。これらのメディア条件を比較することによって、身体の動きをともなう遠隔協調作業において遠隔ユーザの上半身映像を投影することの価値を調べる。実験の設計では、実験計画法に則り、メディア条件の順序と課題の順序の釣り合いをとった。

4.2 グループ課題

本実験の課題には、テーブルを用いた遠隔協調作業でよく題材として用いられる「実物を用いた遠隔作業指示（mentoring collaborative physical task^{1),5),6),9)}」を採用した。この課題は、専門家（指示者）が遠隔地にいる初心者（作業員）に指示を出すことによって目的を達成するというものである。本実験では、異なる専門知識を持つ 2 名の専門家が神奈川県、初心者 2 名が京都府にいる状況を想定し、合計 4 名で遠隔協調作業を実施するものとした。具体的な協調作業内容は「プラレールの組み立て」とした^{*1)}。

実験では、指示者はプラレールの完成図（図 4）を記憶し、記憶を頼りに作業員に指示を行うこととした。指示者と作業員 1 名ずつでペアを組み、レール作成担当と情景部品配置担当に役割分担をした。プラレール部品は作業員のいる t-Room 内のテーブル上に過不足な

*1 プラレールはタカラトミー社の製品で電車、線路、情景部品からなるプラスチック製の玩具である。



図 4 プラレールの組み立て完成図

Fig. 4 The railroad model used in the experiment.

く置いてあり、作業員だけが操作することができる。このように完成図の知識を持っている者と実物を操作する者とを分けることによって、遠隔地間で協力し合わなければ課題が遂行できない状況を作り出した。

被験者らはメディア条件を変えて 2 回課題に取り組んだ。ただし、課題に対する学習効果を軽減するために、プラレールの完成図を 2 つ用意した。

4.3 実験手順

手順 1 (自己紹介): 実験開始前、各被験者は実験前に知己になることを防ぐため別々の部屋で待機した。各被験者は待機部屋で実験に参加するための同意書に署名し、その後 t-Room に案内された。4 名の被験者は t-Room を介して自己紹介を行った。

手順 2 (役割分担): 神奈川県と京都府の作業員 1 名ずつをペアとし、役割分担 (1 ペアはレール作成を担当、もう 1 ペアは情景部品の配置を担当) を決めた。情景部品の配置を担当するペアは、テーブル映像しか見えないときでも手の映像だけで遠隔ユーザの識別を可能にするために、両者とも軍手を着用した。

手順 3 (システム説明と練習): t-Room の環境やテーブルの使い方に慣れるために、実際にテーブル上の物体を扱う練習 (テーブル上の物体を移動させたり回転させたりする指示と作業) を 5 分程度行った。

手順 4 (課題の説明): 被験者らは課題の流れについて説明や注意事項を受けた。注意事項は以下のとおりである。原則として、手順 2 で定めた役割分担に従って指示や作業を行うこととし、自分と関係のない部分については指示や作業を行わないものとした。ただし、記憶

誤りなど、課題の目的を達成するうえで重大な問題が生じた場合には負担を超えたアドバイスや手伝いをしてよいことにした。また、組み立てる順番や方針、t-Room の利用方法は自由とし、相談や話し合いは 4 名で自由に行ってよいこととした。なお、作業時間は最長 20 分としたうえで、可能な限り早く作業を完成させるように伝え、もし 10 分以内に完成できた場合には景品をプレゼントする旨を伝えた。

手順 5 (課題の遂行): 以下の内容の課題 (手順 5.1 から手順 5.3) を、遠隔ユーザの上半身映像が見える状態と見えない状態で課題の完成図を変えて合計 2 回実施した:

手順 5-1: 2 名の指示者は完成図 (図 4) を 2 分間で記憶した^{*1}。2 分後に完成図を印刷した用紙は回収された。

手順 5-2: 被験者らは再び t-Room 内に戻り、組み立て作業を行った。

手順 5-3: 組み立て完成後 (あるいは作業開始から 20 分後)、協調作業のしやすさに関するアンケート I に回答した。

手順 6 (アンケート II): メディア条件を変えて 2 回課題に取り組んだ後、被験者はメディア条件の違いに関するアンケート II に回答した。

手順 7 (インタビュー): 京都府で実験に参加した被験者は、遠隔指示者の上半身映像の有無による違いや指示者との立ち位置のとり方、指示内容の理解のしやすさなどに関するインタビューを受けた。

5. 分析結果

我々はメディア条件の違いが協調作業に及ぼす影響を詳しく調べ、2 章で立てた仮説を検証するために、両 t-Room の出入り口の頭上付近および出入り口と反対の位置にビデオカメラ (合計 4 台) を設置し、実験の様子を撮影した。以下に、これらのビデオ映像を分析した結果を述べる。

5.1 遠隔ユーザ間の立ち位置

我々は、テーブル一周を 8 等分する印をあらかじめテーブルの縁に付けておき、この尺度を用いてユーザ間の距離を計測した。ここで、距離 4 は最も遠い対面の位置関係 (180 度) を指し、距離 2 は斜め向かい (90 度) の位置関係を指し、距離 1 は約 1 歩離れた状態を指す。

*1 本実験では、指示者にだけ完成図の情報が伝わるように (t-Room を介して作業員に完成図が見えることを避けるため)、完成図の配布はせず、記憶してもらったことにした。

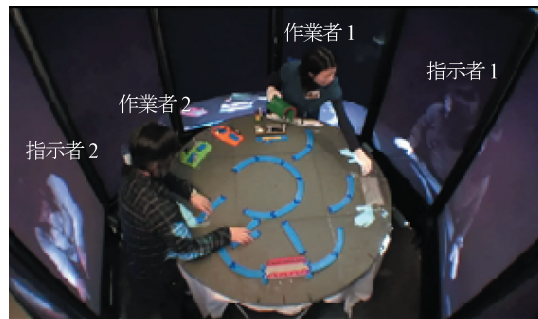


図 5 指示者と作業員の距離が 1 以上 2 未満の例
Fig. 5 Scene of helper-worker distance between 1 and 2.

表 1 各ペアにおける指示者と作業員間の平均距離
Table 1 Average distance between instructor and worker pairs.

	距離 1 未満	距離 1 以上 2 未満	距離 2 以上 3 未満	距離 3 以上 4 未満
テーブル映像のみ	33.4%* ¹ (15.7%)	20.9% (10.3%)	26.1% (13.7%)	21.7% (16.7%)
上半身映像あり	22.9% (16.9%)	30.8%* ² (11.1%)	25.1% (9.4%)	20.7% (9.6%)

図 5 は、ペアの距離が 1 以上 2 未満の場合と 1 未満の場面をキャプチャしたものである。情景部品担当のペア間（指示者 1 と作業員 1）の距離は 1 以上 2 未満であり、作業員は指示者の上半身映像を参考に指示内容の把握に努めている。一方、レール担当のペア間（指示者 2 と作業員 2）の距離は 1 未満であり、作業員はテーブル面に投影された指示者の指示を見ながら作業を行っている。

仮説 1 では、上半身映像が見える場合に、ユーザが遠隔ユーザの様子を認識しやすい位置に移動することを予想した。課題の作業はペアで進められることが多かったため、我々は各ペアにおける指示者と作業員の立ち位置のとり方に注目した。各ペアにおける指示者と作業員の距離を 10 秒置きに計測し、その分布を調べた（表 1）。表 1 の括弧内の値は各セルに対応する標準偏差の値を示している。

表 1 より、上半身映像の有無にかかわらず、指示者と作業員の距離は状況に応じて様々

に変化していたことが分かる。しかし詳細に比較すると、メディア条件 1（テーブル映像のみの場合）では、メディア条件 2（上半身映像がある場合）と比較して、指示者と作業員の距離が 1 未満の場合が多く（表中 *1）、逆に距離が 1 以上 2 未満の場合が少ないことが分かる（表中 *2）。メディア条件と課題の 2 要因による繰返し分散分析（repeated measures ANOVA）を行ったところ、メディア条件による主効果が検出され、条件間でこれらの割合の違いに有意差が検出された（指示者と作業員の距離が 1 未満の場合： $F(1, 29) = 3.23$, $p = .08$ ；指示者と作業員の距離が 1 以上 2 未満の場合： $F(1, 29) = 6.63$, $p < .05$ ）。なお、課題による主効果は検出されなかった。

ビデオ映像からも、この結果を支持する場面が多数観察された。たとえば、いずれのメディア条件においても指示者が作業員に歩み寄って作業員の傍でお手本を示すように指示を行う場面が多く見られたが⁵⁾、上半身映像が見えるメディア条件 2 の場合には、作業員がジェスチャの見えやすい位置に身体配置をずらしたり、作業が一段落した段階で作業員が指示者の上半身映像の見えやすい位置に 1, 2 歩移動したりする場面が多かった。協調作業が一段落した合間に遠隔ユーザの様子を観察する動作が多いことは、文献 20) でも報告されている。なお、指示者と作業員の距離がある程度離れており（距離 2 以上）、遠隔ユーザの上半身映像が見やすい位置にいる場合には、立ち位置の調整はほとんど観察されなかった。一方、テーブル面のみ共有するメディア条件 1 の場合には、指示者と作業員の距離や作業のタイミングにかかわらず、互いの立ち位置を調整する行動はほとんど観察されなかった。

以上より、遠隔ユーザの上半身映像が見える環境では、ユーザは互いの立ち位置を調整しながら作業を行っていたことが分かる。課題終了時に実施したアンケート調査からも、被験者らがこのような立ち位置の調整を多少意識して行っていたことが分かった。具体的には、被験者にテーブル周りを移動した理由について尋ねたところ、(両メディア条件で「部品を移動させるため」といった課題の性質に依存する回答が大半を占める中)上半身映像の見えるメディア条件 2 では、「相手（遠隔指示者）に近づいた方がいいと思った（5 名）」や「相手（遠隔作業員）に見えやすい位置に移動した方が良かった（5 名）」といった遠隔被験者との立ち位置を意識した回答が複数寄せられた。テーブル面のみ共有したメディア条件 1 で課題を実施したときには、このような回答はほとんど得られなかった（1 名）。

5.2 遠隔ユーザに対する呼びかけ

課題の進め方は、いずれのメディア条件の場合も、おおむねレールに沿ってレール担当と情景部品担当が交替で作業するというものであった。時折、指示者同士が互いの記憶を確認し合ったり、指示者がペアでない作業員に助言したりする場面も見られた。このようなと

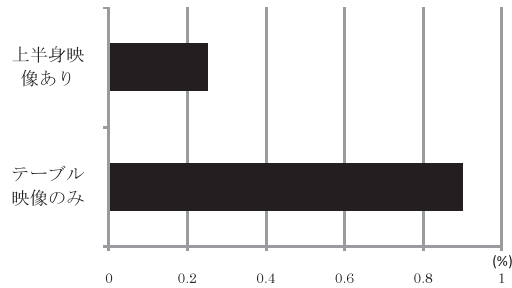


図 6 発話相手の名前を呼んでから発話した割合
Fig. 6 Ratio of explicit handovers per utterance.

き、もし上半身の映像が次の指示への予測につながっているなら、遠隔ユーザの名前をわざわざ呼ばなくても、立ち位置や体躯や頭部の向きなどから、誰に対する発話であるかを推測できるはずである（仮説 2）。

そこで我々は、メディア条件ごとに、被験者が遠隔ユーザの名前を呼んでから発話した割合を調べた。具体的には、メディア条件ごとに、各グループで課題を解くにあたって必要となった総発話数を求め、それらの発話の中で「被験者が遠隔ユーザの名前を呼んでから発話した回数」の割合を調べた。

図 6 より、仮説 2 で予想したとおり、被験者は遠隔ユーザの上半身映像が見える場合（メディア条件 2）、見えない場合（メディア条件 1）と比較して、遠隔ユーザの名前を呼んでから発話する割合が少ないことが分かる。分散分析により、メディア条件による主効果が検出され、条件間の発話割合に有意差が検出された ($F(1, 13) = 4.67, p = .05$)。課題による主効果は検出されなかった。

テーブル映像しか共有していない場合（メディア条件 1）、互いの状況を把握することが難しく、これから発話しようとしている相手が準備状態にあるかないかを察知することが難しいことが予想される。このように遠隔ユーザの名前を呼ぶ行為は、発話相手が聞く体制をとっていないと思われる場合や、仮に遠隔ユーザが準備態勢をとっていたとしてもその状況を（遠隔地から）把握することが難しい場合に使われることが多い。このことについてより詳しく調べるため、我々は課題終了時のアンケート調査において、被験者らが自分の状況を遠隔ユーザに知らせるために工夫していた事柄について尋ねた。その結果、メディア条件 1 では、多くの被験者が自分の状況や立ち位置を伝えるために、両手をテーブルの縁に置くように心がけていたことが分かった（8 名）。なお、実験後のインタビューでも、メディア条

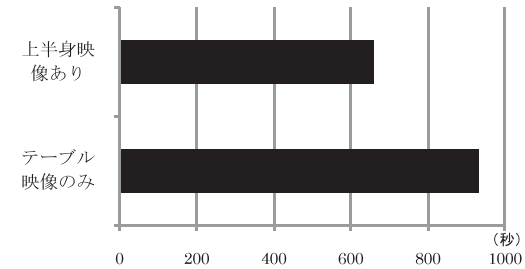


図 7 メディア条件ごとの平均作業時間
Fig. 7 Average time spent in completing the task.

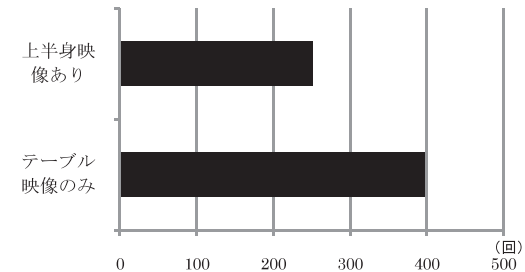


図 8 メディア条件ごとの平均発話数
Fig. 8 Average number of utterances needed to complete the task.

件 1 において、「（遠隔作業者の）手が机に映っていないとき、ペアの人が何をしているのか把握できず焦った」や「他のペアが話していて（自分に）指示がないとき、相手（遠隔指示者）の姿が見えないと落ち着かなかった」といった意見が聞かれ、作業内容とは直接的に関係なくても、互いの状況を気にしている様子がうかがえた。

以上より、テーブル映像のみ共有する場合は、被験者らは遠隔地間で互いの状況を伝達・把握し合うことが困難であり、そのため遠隔ユーザの名前を呼んでから発話するケースが多かったと考えられる。

5.3 発話効率と作業効率

仮説 3 を検証するために、各グループが課題を終えるまでにかかった時間（完成できなかった場合は 20 分とした）と総発話数を計測した。

図 7 と図 8 は、各メディア条件において被験者らが課題を終えるまでにかかった平均

時間および平均発話数を示している。図7と図8より、上半身映像が見える場合（メディア条件2）は、見えない場合（メディア条件1）と比較して、少ない発話数で早く課題を終える傾向にあったことが分かる。分散分析により、メディア条件による主効果が検出され、条件間の違いに有意差が検出された（作業時間： $F(1, 13) = 9.91, p < .01$ ；総発話数： $F(1, 13) = 12.10, p < .01$ ）。課題による主効果は検出されなかった。

実験の最後に、2つのメディア条件を比較するアンケートを実施し、上半身映像の効果について尋ねた。上半身映像が遠隔ユーザの次の行動予測につながったことを直接示唆するような回答はほとんど得られなかったが、逆にテーブル面しか共有しなかった場合の問題点について尋ねたところ、遠隔ユーザの発話意図を汲んだり予測したりすることが難しかったという意見が多く見られた。たとえば、作業者からは、「相手（指示者）が私のすぐ前を指していても気づかないときがあった」、「相手の手を探したり追ったりするのに必死でした」、「次にどのような指示があるか分からないから、落ち着かなかった」といった回答が寄せられた。同様に、指示者側も「前後左右など自分の立ち位置や組み立てる物を基準に位置を示しても伝わらなかった」、「言葉で的確に説明しなければ伝わらなかった」といった意見が寄せられた。

以上より、被験者らは上半身映像を通して互いの行動をモニタし、指示や作業を行っていない場面でも互いの立ち位置を調整しながら次の指示の予測や準備を行っていたと考えられる。これら個々の活動の積み重ねが発話効率や作業効率の向上につながった可能性がある。

6. 考察とまとめ

遠隔地間でテーブル面を共有することによって遠隔協調作業を支援する研究がさかんに行われている。しかし、テーブル面に加えて遠隔ユーザの上半身映像も投影すべきか否かの明確なデザイン指針は今のところ見当たらない。上半身映像に関する従来研究の共通認識は、一部の例外を除けば²⁰⁾、あまり付加価値を生んでいないというものであった^{13), 15)}。音声通信だけでも多種多様な情報伝達が可能であることを考えると、この結果はさほど驚くべきことではないだろう。

本研究では、テーブルを共有して行う作業の利点である「多人数（1地点に複数ユーザが参加）による動きをともなう協調作業が可能」という点に注目し、上半身映像の有無が作業にどのように影響するかを調べた。先行研究では、上半身映像とテーブル映像が連続して見えるように表示させると、上半身映像がテーブル上のジェスチャの予測に寄与することが報告されている¹¹⁾。しかし、その例では、参加者は各地点に1名ずつで、参加者の立ち位

置は固定されるという制約された条件であった。また、上半身映像の有無は比較されておらず、分析も質的な手法に限られていた。これに対して本実験では、各地点に複数の参加者がいて、動的に立ち位置を変更できるという環境において、2つのメディア条件（遠隔地間で「テーブル面のみ共有する場合」と「テーブル面と上半身映像を共有する場合」）を比較し、定量的な分析を行うことによって、上半身映像の特質を客観的に示すことができた。

我々は、上半身映像の有効性を示す以下の3つの分析結果を得た：

(1) 被験者らは遠隔被験者の上半身映像を頻繁にモニタし、互いの立ち位置や距離を柔軟に変えながら作業をしていた。特に、被験者らは遠隔被験者の様子が認識しやすい位置に移動する傾向があることが分かった。その結果、遠隔地間でお互いのジェスチャ、体の向きや動きなどをモニタしやすくなることで、作業中だけでなく作業をしていない間も互いの映像を見ながら立ち位置を調整（たとえば指示者の動きに合わせて作業者が近寄るなど）していた。このことから、被験者らは、上半身映像は次の動きや作業を予測したり準備したりしていたと考えられる。

(2) 上半身映像が表示されていると、被験者らが遠隔被験者の名前をわざわざ呼んでから発話を開始する場合は少なかった。これは、お互いの立ち位置の近接関係が分かりやすかったために、誰に対する発話かが自明であったことや、上半身映像を見ることによって対話者の次の動きを予測し、それに対する反応を準備していたため、わざわざ発話相手の名前を呼んで注意を喚起する必要が少なかったためと考えられる。

(3) 上半身映像が表示されていると、協調作業にかかる時間や発話数が大幅に減少した。テーブル映像のみの場合でも、ペアの判別がしやすいように片方のペアには手袋をさせるといった工夫をしていたため、共同作業に必要な情報は十分に伝達できていたと思われる。それにもかかわらず上半身映像が表示されている方が効率向上したのは、遠隔地間でユーザ同士がお互いモニタすることによって、作業の予測が可能になったり、相手ユーザが自分の指示をどの程度理解しているのか（理解度）を把握できたりしたこと起因するのではないかと考えられる。

以上、本研究により、上半身映像が、多人数対話における円滑な話者交代や遠隔ユーザのジェスチャ理解や理解度の把握などに役立ったことは容易に推察できる。多人数の動きをともなう協調作業において、上半身映像を表示している方が有効であることが示された。ただし、本実験だけでは、具体的に上半身映像のどの部分を表示する効果が大きいのかを示すことはできない。協調作業のパフォーマンスを向上させるうえで、どの要素（表情、視線、体の向き、腕の動きなど）がどの程度役立ったかを調べるのが今後の課題である。

最後に、本論文では上半身映像を表示することによる利点を強調してきたが、実験での協調作業の様子を詳しく観察したところ、新たな課題も見えてきた。実験では、作業者が遠隔指示者の上半身映像を注視している間に遠隔指示者がテーブル上の物体にジェスチャを行い、作業者が遠隔指示者の腕の動きに追従してテーブルを見たときには（テーブル上の）ジェスチャは終わっていて見逃してしまう光景が頻りに観察された。このような遠隔指示者のジェスチャの見逃しは、上半身映像の有無にかかわらず、特に指示者と作業者の立ち位置が近いときに頻りに観察された。実物の腕や指と比較して、2次元表示された腕や指が気付かれにくいことも²⁾、見落としや見逃しが多い要因の1つではないかと推察する。テーブル面を共有する遠隔協調作業システムでは、これらの問題の対策法も考えていく必要があろう。

謝辞 本研究は科学研究費基盤研究(C)課題番号20500122の補助を受け実施したものである。t-Roomプロジェクトメンバ諸氏には実験の計画、準備、実施にあたり多大な協力をいただいた。ここに深謝の意を表す。

参 考 文 献

- 1) Fussell, S.R., Setlock, L. and Kraut, R.E.: Effects of Head-Mounted and Scene-Oriented Video Systems on Remote Collaboration on Physical Tasks, *Proc. CHI'03*, pp.513-520, ACM Press (2003).
- 2) Gaver, W.: The affordances of media spaces for collaboration, *Proc. CSCW'92*, pp.17-24, ACM Press (1992).
- 3) Gaver, W., Sellen, A., Heath, C. and Luff, P.: One is not enough: Multiple views in a media space, *Proc. Interchi'93*, pp.335-341, ACM Press (1993).
- 4) Heath, C. and Luff, P.: Disembodied Conduct: Communication Through Video in a Multi-media Office Environment, *Proc. CHI'91*, pp.99-103, ACM Press (1991).
- 5) Kirk, D., Crabtree, A. and Rodden, T.: Ways of the hands, *Proc. ECSCW'05*, pp.1-21, Springer (2005).
- 6) Kirk, D.S. and Fraser, D.: Comparing Remote Gesture Technologies for Supporting Collaborative Physical Tasks, *Proc. CHI'06*, pp.1191-1200, ACM Press (2006).
- 7) Kendon, A.: Spatial Organization in Social Encounters: The F-formation System, *Conducting Interaction: Patterns of Behavior in Focused Encounters*, Kendon, A. (Ed.), pp.209-237, Cambridge University Press (1990).
- 8) Kraut, R.E., Gergle, D. and Fussell, S.R.: The Use of Visual Information in Shared Visual Spaces: Informing the Development of Virtual Co-Presence, *Proc. CSCW'02*, pp.31-40, ACM Press (2002).
- 9) Kraut, R.E., Miller, M.D. and Siegel, J.: Collaboration in performance of physical tasks: Effects on outcomes and communication, *Proc. CSCW'96*, pp.57-66, ACM Press (1996).
- 10) Kuzuoka, H., Yamashita, J., Yamazaki, K. and Yamazaki, A.: Agora: A Remote Collaboration System that Enables Mutual Monitoring, *Extended Abstracts of CHI'99*, pp.190-191, ACM Press (1999).
- 11) Luff, P., Heath, C., Kuzuoka, H., Yamazaki, K. and Yamashita, J.: Handling documents and discriminating objects in hybrid spaces, *Proc. CHI'06*, pp.561-570, ACM Press (2006).
- 12) Minatani, S., Kitahara, I., Kameda, Y. and Ohta, Y.: Face-to-Face Tabletop Remote Collaboration in Mixed Reality, *Proc. ISMAR'07*, pp.43-46, IEEE Computer Society, (2007).
- 13) Nardi, B., Schwarz, H., Kuchinsky, A., Lechner, R., Whittaker, S. and Scabassi, R.: Turning away from talking heads: The use of video-as-data in neurosurgery, *Proc. Interchi'93*, pp.327-334, ACM Press (1993).
- 14) Nguyen, D. and Canny, J.: MultiView: Spatially faithful group video conferencing, *Proc. CHI'05*, pp.799-808, ACM Press (2005).
- 15) O'Connell, B. and Whittaker, S.: Characterizing, Predicting and Measuring Video-Mediated Communication: A Conversational Approach, *Video-Mediated Communication*, Finn, K., Sellen, A. and Wilbur, S. (Eds.), pp.107-131, Hillsdale NJ: Lawrence Erlbaum Associates, (1997).
- 16) Scott, S.D., Grant, K.D. and Mandryk, R.L.: System guidelines for co-located, collaborative work on a tabletop display, *Proc. ECSCW 2003*, pp.159-178, Kluwer Academic Publishers (2003).
- 17) Sellen, A., Buxton, B. and Arnott, J.: Using spatial cues to improve videoconferencing, *Proc. CHI'92*, pp.651-652, ACM Press (1992).
- 18) Taylor, S., Izadi, S., Kirk, D., Harper, R. and Mendoza, A.G.: Turning the Tables: An Interactive Surface for VJing, *Proc. CHI'09*, pp.1251-1254, ACM Press (2009).
- 19) Tuddenham, P.: Remote Review Meetings on a Tabletop Interface, *Proc. CSCW'06*, pp.17-24, ACM Press (2006).
- 20) Veinott, E., Olson, J., Olson, G. and Fu, X.: Video helps remote work: Speakers who need to negotiate common ground benefit from seeing each other, *Proc. CHI'99*, pp.302-309, ACM Press (1999).
- 21) Vertgaal, R., Van der Veer, G.C. and Vons, H.: Effects of Gaze on Multiparty Mediated Communication, *Proc. Graphic Interface 2000*, pp.95-102, Morgan Kaufmann Publishers, (2000).
- 22) Vertegaal, R., Weevers, I., Sohn, C. and Cheung, C.: GAZE-2: Conveying eye contact in group video conferencing using eye-controlled camera direction, *Proc. CHI'03*, pp.521-528, ACM Press (2003).
- 23) Wesugi, S., Ishikawa, K. and Miwa, Y.: Intimate virtual communication place

supported with networked “lazy Susan”, *Human-Computer Interaction Theory and Practice, HCI'03*, Vol.2, pp.563–567 (2003).

- 24) Yamashita, N., Hirata, K., Aoyagi, S., Kuzuoka, H. and Harada, Y.: Impact of Seating Positions on Group Video Communication, *Proc. CSCW'08*, pp.177–186, ACM Press (2008).

(平成 21 年 7 月 8 日受付)

(平成 21 年 12 月 17 日採録)



山下 直美 (正会員)

1999 年京都大学工学部情報工学科卒業。2001 年京都大学大学院情報学研究科数理工学専攻修士課程修了。同年日本電信電話(株)コミュニケーション科学基礎研究所入所。博士(情報学)。CSCW, HCI の研究に従事。



葛岡 英明 (正会員)

1992 年東京大学大学院工学系研究科情報工学専攻博士課程修了。博士(工学)。同年筑波大学構造工学系講師。現在、筑波大学大学院システム情報工学研究科教授。CSCW, HRI, HCI の研究に従事。



平田 圭二 (正会員)

1987 年東京大学大学院工学系研究科情報工学専門課程博士課程修了。工学博士。同年 NTT 基礎研究所入所。1990~1993 年(財)新世代コンピュータ技術開発機構(ICOT)。1999 年より NTT コミュニケーション科学基礎研究所。2001 年度論文賞, 2003 年度山下記念研究賞, 2005~2007 年本会理事。現在、日本ソフトウェア科学会理事, t-Room プロジェクトリーダー。音楽情報処理に興味を持つ。



青柳 滋己 (正会員)

1965 年生。1988 年東京工業大学理学部情報科学科卒業。1990 年 3 月同大学大学院理工学研究科情報科学専攻修士課程修了。同年日本電信電話株式会社入社。現在、同社コミュニケーション科学基礎研究所に所属。複合メディア情報処理の研究に従事。電子情報通信学会, 日本ソフトウェア科学会各会員。



白井 良成 (正会員)

1998 年慶應義塾大学環境情報学部卒業。2000 年同大学大学院政策・メディア研究科修士課程修了。同年日本電信電話株式会社 NTT コミュニケーション科学研究所入所, 現在に至る。インタラクティブデザインに興味を持つ。ヒューマンインタフェース学会会員。



梶 克彦 (正会員)

2007 年名古屋大学大学院情報科学研究科博士課程修了。博士(情報科学)。2007 年 NTT コミュニケーション科学基礎研究所リサーチアソシエイト。人工知能学会, 日本ソフトウェア科学会, ヒューマンインタフェース学会各会員。Web コンテンツ処理, 音楽情報処理, 実世界コミュニケーションに関する研究に従事。



原田 康徳

計算機科学者。ワークショップデザイナー。1963 年北海道生まれ。1992 年北海道大学大学院情報工学専攻博士後期課程修了。同年日本電信電話株式会社 NTT 基礎研究所 1998~2001 年 JST さきがけ研究員。2004~2006 年 IPA 未踏ソフトウェアプロジェクトマネージャ兼務。現, NTT コミュニケーション科学基礎研究所。変わったインタフェース, プログラミング言語の研究を行い, 2003 年にその集大成である Viscuit (ビスケット) を発明する。博士(工学)。