

窓越しインタフェース MoPaCo による指示作業への効果検証 Effectiveness of Remote Instructor's Action by Window Interface MoPaCo

石井 亮[†] 小澤 史朗[‡] 川村 春美[‡] 小島 明[‡] 中野 有紀子^{*}
Ryo Ishii Shiro Ozawa Harumi Kawamura Akira Kojima Yukiko Nakano

1. はじめに

我々は、遠隔地間において、お互いの実空間を共有しながら、部屋のレイアウトの相談や、空間中にある物の移動、運搬といった指示作業を円滑に行うことのできるメディアスペースの実現を目指している。そのようなシーンで、メディアスペースに期待されているのは、視線、表情、身振りといった非言語情報の伝達である。非言語情報は、遠隔地での指示作業や協調作業を円滑に行うために重要であることが知られている[1, 2]。しかしながら、既存の TV 電話や TV 会議システムなどの映像メディアでは、非言語情報の伝達が上手いかわからないことが指摘されている。特に、カメラ映像をそのままディスプレイに表示した場合、ユーザがディスプレイの映像中の対象物に視線を向けたり、指差しを行っても対話相手には正しく伝達できない。これは、2 つの空間の幾何学的な大きさや位置関係がディスプレイ上の映像に考慮されていないことが大きな原因である[3]。

これに対して、我々は、ディスプレイが 1 枚の窓ガラスになったかのように、対話相手の空間がディスプレイの奥に再現される、窓越しインタフェース MoPaCo を提案してきた[4]。MoPaCo では、ユーザの観察位置に合わせて運動視差を生成し、大きさや位置関係といった幾何学手整合性を考慮して対話相手の空間がディスプレイに表示されるため、視線、指差し、身体配置といった空間的な方向や位置関係が重要となる非言語情報が自然に伝達されることが期待される。そこで本研究では、MoPaCo により非言語情報がどのように伝達され、会話が円滑に進んでいるかを実験的に検証し、MoPaCo の有効性を示す。

2. 関連研究

空間や物を共有しながらの指示作業において、視線や指差し、身体配置の重要性が社会心理学の分野で明らかにされている。非言語情報と課題遂行との関係を分析したもののとして、鈴木らは、視線、指差しなどの非言語情報が、課題の成功を左右していることを明らかにした[1]。また、松田らは、集団課題において、親和性の高い作業者間で取り交わされるジェスチャなどの非言語行動が課題達成の促進に寄与していることを示した[2]。このように、非言語行動が、タスク課題において重要であることが知られている。

非言語行動が実際にどのように行われているかを分析したもののとして、作業指示において対象物の参照を行う際に、対話相手の体の向きや視線方向から対話相手がどの対象物

に注意を向け、説明や作業を行うかを予期すること (projectability, predelectability) は、対象物の参照を容易にする上で重要であるとされている[5]。これに関連して、Goodwin は、対面状況において、話し手が聞き手に、ある対象物に対する指示を与えるシーンでの非言語行動を分析した[6]。まず聞き手は、話し手が見ている対象物に共同注視しようと身体を適切な方向へ向ける。これによって、聞き手は話し手の話を、積極的に聞き入れる体勢になっていることを示す。一方、話し手は指示を与える際に、対象物と聞き手が両方見えている位置に移動をする。このように、聞き手は話し手の興味対象を理解し、次の指示対象となるものに注意を向ける。これにより、対象物を話し手が参照した時に、聞き手は円滑にその対象物を参照することができるのである。

このような非言語情報が、映像メディアで正しく伝わらないことによる問題も明らかにされている。Health らは、映像コミュニケーションにおいて、カメラ映像をそのままディスプレイに表示した場合、自分が相手の話を聞く準備ができていることを示す身体姿勢や、相手の方向をこれから見るということを示す身体動作、また相手の行動を理解したことを示す身体動作が、対面時に比べて、十分な効果を発揮しなかったことを示しており[3]、すなわち予期することが困難であることが示されている。

このような背景から、遠隔コミュニケーションシステムにおいても視線や指差し、身体情報を伝達可能とするシステムが開発されてきた。例えば、実物体に対する指差しやジェスチャを伝達可能にする試みとして、作業対象物と教示者のジェスチャを合成した映像を学習者に見せられるシステムがある[7, 8]。また葛岡らは、遠隔にいる教示者の指差し、頭部方向 (疑似的な視線方向) および身体的位置関係を代理ロボットにリアルタイムに反映させることのできる代理ロボットを提案している。展示説明代理ロボットとして利用した際の実験では、聴衆がロボットの頭部方向に自然と視線を向け共同注視することや、指差しされた対象物を円滑に参照することが可能であり、非言語情報の提示による円滑な遠隔コミュニケーションの実現可能性を示している[9]。この研究では、人間の代替となる装置や代理ロボットを用いて視線や指差し、身体位置のみに焦点を当て、それらの非言語情報のみを伝達することを目的としていた。しかしながら、非言語行動は、単体で伝達しても上手く情報は伝達できず、複数の非言語情報を複合的に伝達する必要がある。すなわち、視線、指差し、身体動作に加えて、表情、頷きといったようなその他の非言語情報も併せて、提示することが重要である。

この観点から、メディアスペースにおいて、複数の非言語行動を実際の対面状況と同等に伝達するためには、映像中の人物から表出される非言語情報を、実際の対面状況と同様に伝達する方が望ましいと考えられる。

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

[‡] 日本電信電話株式会社 NTT サイバースペース研究所, NTT Cyber Space Laboratories, NTT Corporation

* 成蹊大学 理工学部,

Seikei University, Dept. of Computer and Information Science

そこで我々は、メディアスペースにおいて、ユーザが発する視線や指差しを正確に伝達するために、2つの空間の幾何学的な大きさや位置関係をディスプレイ上の映像に考慮し提示することを目的に、ディスプレイを窓に見立てて、あたかも2つの空間が窓越しに繋がっているかのように、窓越し映像を提示するインタフェース MoPaCo を提案した[4]。MoPaCo では、実際の窓を介して会話をした際と同等に、2つの空間の大きさ、位置関係が再現される。そのため、身体配置を自然に伝達しあえることや、映像中の人物が対話相手の空間にいる人や物に対して注視や指差しを行った際に、その対象を自然に伝達可能であることが期待される。

そこで本研究では、前述したように、指示作業において社会心理学の分野で重要視されている指示対象の参照および指示対象の予期について、MoPaCo 環境下でどのように行われるかを検証し、MoPaCo の有用性を検証する。

3. 窓越しインタフェース MoPaCo

3.1 システム概要

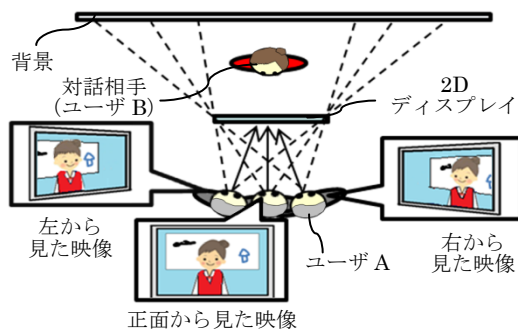


図 1 窓越しインタフェース MoPaCo の映像表現イメージ

窓越しインタフェース MoPaCo[4]が実現する映像表現のイメージを図 1 に示す。図 1 は、ユーザ A が対話相手であるユーザ B の映像を視聴している様子である。このとき、ユーザ A とユーザ B の空間が、ディスプレイを窓に見立ててあたかも窓越しで繋がっているかのように、ユーザ A が

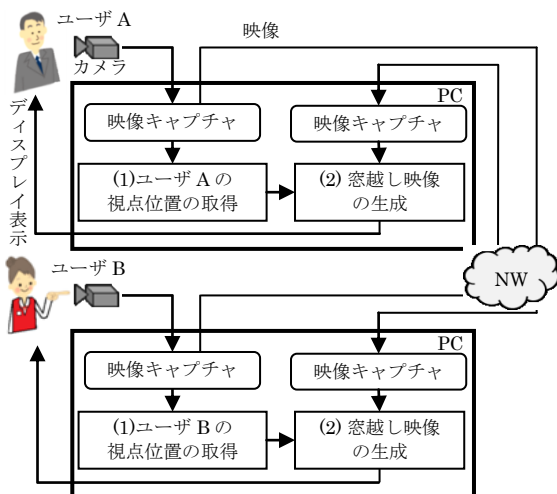


図 2 MoPaCo のシステム構成図

左側、右側に移動することで、窓越しに左側、右側からユーザ B の空間を視聴した風景が 2D ディスプレイ上に再現される。

我々は広く普及することを目指して、一般的な 1 台の固定カメラと 2D ディスプレイのみでこの機能を実現するための手法を提案してきた。次に、MoPaCo のシステム構成を図 2 に示す。

ユーザ A にユーザ B の窓越し映像を提示するためには、ユーザ A、B のカメラ映像を用いて、(1) ユーザの顔の 3 次元位置の測定と、(2) 上下、左右、前後といったユーザ A の視点位置の移動に対応し、ユーザ B の空間の大きさと位置関係を再現した窓越し映像の生成を行う。

3.2 ユーザの視点位置の取得

ユーザの顔の 3 次元位置（両眼の中心の 3 次元位置）の測定手段として、画像中の顔位置を測定し、その位置情報を利用する。顔の検出処理には、状態履歴を参照したパーティクルフィルタによる顔検出・追尾技術[10]を用いた。本技術を用いることで、画像中の顔のパーツ（眼、鼻、口）の 2 次元での位置および顔の大きさがリアルタイムに測定される。図 3 に顔検出結果の様子を示す。位置情報と大きさ情報に加えて、事前に顔の大きさとオートフォーカス機能実行時のデプスフロムドフォーカス機能によって算出されるユーザの位置情報から顔の大きさと位置関係を同定することで、リアルタイムにユーザの 3 次元位置を算出する。

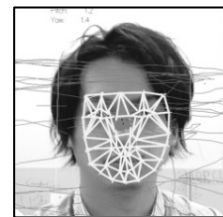


図 3 顔追尾結果の様子

3.3 ユーザの視点位置の測定

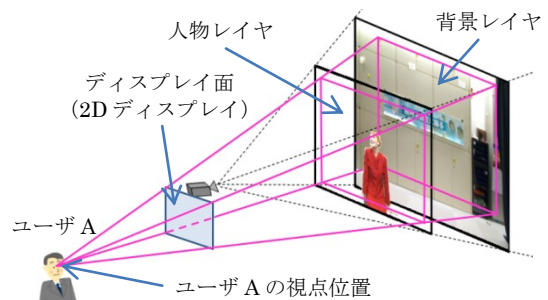


図 4 窓越し映像の生成手法

次に、窓越し映像を生成する手法として、ユーザ B のカメラ映像からユーザ B の空間に対して、大きさや位置関係を考慮した 3 次元モデルを生成し、その 3 次元モデルをユーザの顔位置に合わせてディスプレイ面に透視投影する（図 4 参照）。3 次元モデル生成手法として、まずカメラ映像に対して、事前に取得された背景画像と比較することで、事前に取得した画像には存在しない人物を抽出する背

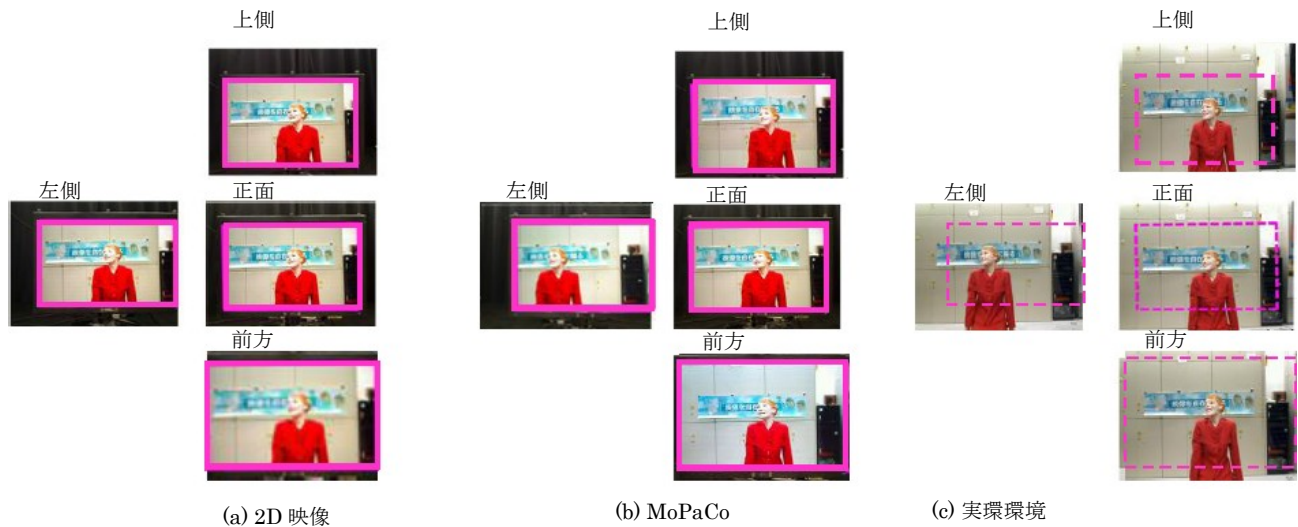


図6 2D映像, MoPaCo, 実環境における対話相手の映像の見え方

景差分処理を行い、人物画像と背景画像を抽出した。それらを平面のレイヤ（以後、それぞれ人物レイヤ、背景レイヤと呼ぶ）とし表現しユーザBの空間を再現する。人物画像と背景画像を用いて、実寸サイズとカメラからの距離関係を持つ2層のレイヤの生成方法は図5のように、各レイヤ*i*の実寸サイズ（横 w_i ×縦 h_i ）を、取得した距離情報 d_i 、カメラ画角（横 θ_w 、縦 θ_h ）から、式(1)を用いて算出する。以上より、実寸のサイズと位置情報を持つ3次元空間が構成される。

$$w_i = 2 * d_i * \tan \theta_w / 2, \quad h_i = 2 * d_i * \tan \theta_h / 2 \quad \dots (1)$$

距離情報 d_i は、背景レイヤの距離情報には、背景差分用の背景画像取得前に、カメラのAF機能を利用し、AF実行時のデプスフロムドフォーカス機能によって算出される値を利用する。人物レイヤについては、3.2節の手法で取得されるユーザBの視点位置情報を用いる。

最後に描画処理として、2次元ディスプレイをディスプレイ面として、生成したユーザBの空間の3次元モデルをユーザの視点位置に応じて、透視投影変換を行うことで視点位置に応じた映像を生成する（図4参照）。このとき、カメラとディスプレイの位置関係を考慮して、座標系を統一する。

3.4 実装結果

MoPaCoを実装した際の、実装環境および性能を表1に示す。表1中の“ユーザ動作からの遅延”は、ユーザAの視点位置が動いてから映像に運動視差が反映されるまでの

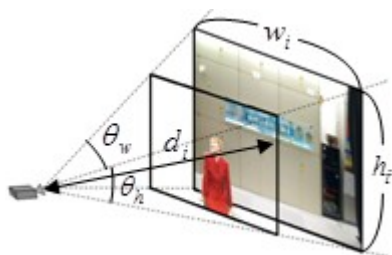


図5 2層表現による空間の3次元モデルの生成

時間，“カメラ画像の遅延”はユーザBの映像が撮影されてからディスプレイに提示されるまでの時間であり、一般的な映像コミュニケーションシステムで言われる映像遅延にあたる。ユーザAの映像処理による運動視差の反映と、ユーザBの映像処理はそれぞれ別々のスレッドで処理するため、遅延時間は異なる。また、映像コミュニケーションシステムでは、“カメラ画像の遅延”に通常、ネットワーク遅延が発生するが、地点間のデータはネットワーク経由ではなく直結をしたため、“カメラ画像の遅延”は映像のネットワーク伝送による遅延を含まない。すなわち、“カメラ画像の遅延”は映像処理時間である。

表1 実装環境および性能

実装環境	CPU	Intel Core i7 Extreme 980X
	メモリ	12GB
	グラフィックボード	NVIDIA GeForce GTX480
	入力画像サイズ	1920×1080 (HD)
性能	描画速度	30fps
	頭部動作からの遅延	約 300ms
	カメラ映像の遅延	約 200ms

MoPaCoによって、生成された映像を図6に示す。図6では、MoPaCoによる映像の見え方に加えて、比較対象として、カメラで撮影した映像を2Dディスプレイにそのまま表示した2D映像を観察したものと、実環境でカメラを実際に移動させて対話相手を撮影した際の見え方を示す。なお、2D映像はユーザがディスプレイの正面にいるときに、実環境と同様の映像が表示されるようにカメラ位置と画角を調節した。図6を見ると、正面から上側、左側、前方に移動した際に、MoPaCo映像では実環境と同じように対話相手と背景の位置関係と大きさが変化していることが確認できる。

4. 評価実験

4.1 実験手続き

遠隔地間の指示作業において、MoPaCoを用いることで、

指差しの参照や会話が円滑になされるかを検証した。MoPaCo では、窓越し映像を表示するため、ユーザが対話相手の空間上の対象物に対して注視や指示を行った際に、対話相手はその対象物を正しく理解できると考えられる。そこで我々は、以下の仮説を建てた。

- ・**仮説**：MoPaCo では、指示者が対話相手の空間の物を指示した際に対話相手は円滑に参照ができる。

また、視線や指示動作の伝達が円滑に行われることで、円滑な会話や意思の疎通が促進されることが考えられる。そこで、会話の円滑さや会話への印象についても評価を行った。

仮説の検証方法として、被験者 2 人にお互いの部屋にある各具のレイアウトを決めるタスクが課された会話において、指示対象の参照がどれくらいの時間で行われるかを検証した。また、会話の円滑さや印象には、質問紙を用いてリッカート法による 6 段階 (1~6 点) の主観評価を行った。主観評価の評価項目を表 2 に示す。また主観的な評価と合わせて、会話 (タスク) への印象の強さや積極的に参加 (engage) していたかを測る指標として、会話 (家具) で使用した家具の記憶力を測り検証を行った。

表 2 主観評価項目

- ・**会話の円滑さ**：会話は円滑に進みましたか？
- ・**意思の疎通**：意思の疎通ができましたか？
- ・**窓越し感**：窓越しに会っているように感じましたか？
- ・**楽しさ**：会話をしていて楽しかったですか？
- ・**親近感**：対話相手に親近感が沸きましたか？

実験条件として、一般的な 2D 映像と実際の窓を介して会話をを行った際と比較検証を行うために、下記のように MoPaCo 条件に加えて、2D 条件、およびガラス窓越しに指示動作を行う REAL 条件の 3 条件を設定した。

- ・**2D 条件**：カメラで撮影した映像を 2D ディスプレイにそのまま表示された映像を介して会話 (このとき、指示者が指示対象となる領域が全て観察できるように、ディスプレイ全体に指示対象が収まるようにカメラの画角を調節した。)

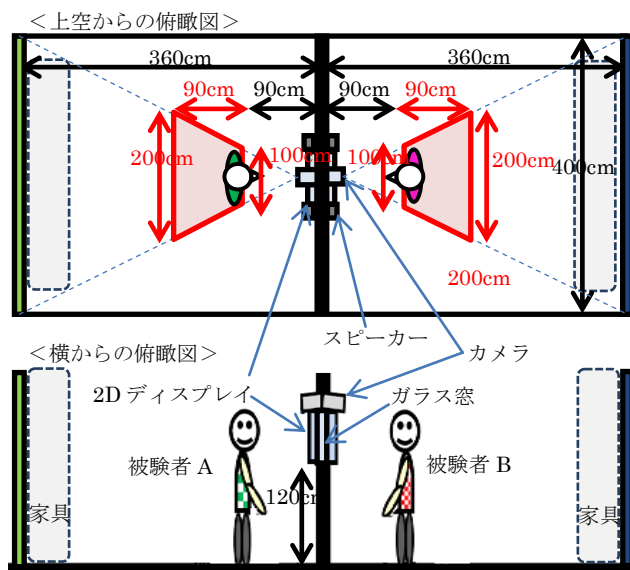


図 7 実験環境



図 8 家具の配置例

- ・**MoPaCo 条件**：MoPaCo による窓越し映像を介して会話。
- ・**窓条件**：隣り合う 2 つの部屋の壁にガラス窓を設置し、ガラス窓越しに会話。

実験環境を図 7 に示す。2 人の被験者は隣り合う別々の部屋 (360cm x 400cm) に入り、2 つの部屋の中央にある壁から 90cm 離れた位置にある移動可能スペース (縦 90cm, 上辺 100cm, 下辺 200cm の台形型) に立った。被験者は、この移動可能スペースの中を自由に移動することが許された。2 つの部屋の中央にある壁には、地面から 120cm の位置に縦 49.8cm x 横 88.4cm のガラス窓を設置した。窓条件では、この窓ガラス越しに会話をを行う。2D 条件、MoPaCo 条件では、ガラス窓のすぐ前にガラス窓と同じ大きさの 40 インチ (縦 49.8cm x 横 88.4cm, 解像度 1920 x 1080) の 2D ディスプレイに設置をし、映像を観察しながら会話をを行う。映像遅延は 2D 条件、MoPaCo 条件共に、300ms であった。カメラは窓のすぐ上に、視線一致するように設置[11]した。音声は、ディスプレイ前に置かれたマイクによって集音され、対話相手の窓のすぐ脇に置かれたスピーカーから出力された。2D 条件、MoPaCo 条件では、遅延発生器を用いて映像に合わせてリップシンクするように 300ms 遅延をさせた。各部屋の壁際にはポスター、テーブル、テレビなどの 14 個の家具が配置された。図 8 に家具の配置の例を示す。

実験手順は、被験者は移動可能スペースの中央に起立し、会話開始の合図と共に対話相手の映像と音声が出力された後、10 分間の対話を行った。対話内容として、無造作に配置されたお互いの家具の配置をどこに配置するかを考え、それが終わった後、互いの部屋に相手の部屋から一つ家具を選択して、自身の部屋の適切な位置に配置するタスクを課した。実験条件の試行順序は、順序効果を考慮して、3 条件をランダムに実施した。なお、家具は全部で 84 個あり、その中からランダムに選択し配置した。なおその際、同じ被験者の組に同じ家具は使用しなかった。各条件施行後に、質問紙を用いて主観評価と家具の記憶力を測る評価に回答をさせた。

被験者は、8 ペア 16 人であり、20~40 代の男性 10 名、女性 6 名、ペアはいずれも友人や家族などの親しい知人関

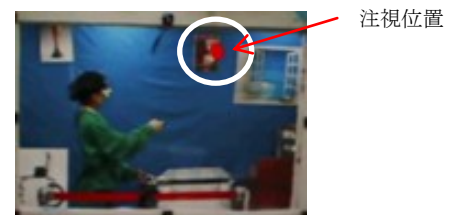


図 9 Tobii glass の計測結果例

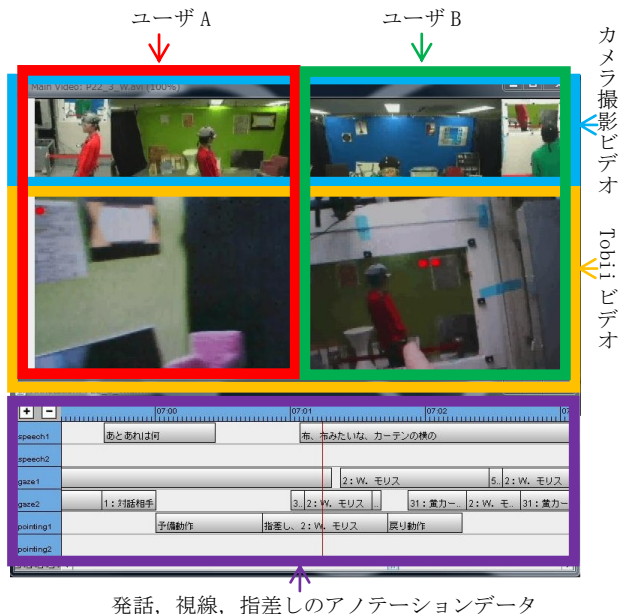


図 10 Anvil による会話コーパスデータ

係であった。

4.2 収集した会話コーパスデータ

分析に必要な被験者の発話、視線行動、指差しジェスチャを以下の方法で取得した。

- ・**発話**：音声を録音し、書き起こしを行った。
- ・**注視対象**：被験者の注視行動を計測するため、装着型の Tobii glass[12]を用いた。Tobii glass は透明なメガネを装着するのみで計測が可能であるため、ユーザへの負担が少なく、目を覆うことによって視線方向の参照を妨げることは無かった。Tobii glass は、装着した被験者の視野映像上の注視位置の 2 次元座標を 30fps で出力する。実際に計測されたビデオデータの例を図 9 に示す。図 9 中の赤い点が注視位置であり、壁に掛けられたポスターを注視している。このビデオ映像から、アノテーションツール anvil[13]を用いて、注視対象物のアノテーションを行った。ラベリングされた注視対象物は、2 つの部屋の家具計 28 個と対話相手である。
- ・**指差しジェスチャ**：被験者の動作をビデオ撮影し、anvil を用いて、アノテーションを行った。指差しジェスチャについては、腕を上げ始めてから指差しを行うまでの“予備動作”、指示対象物を差している“指示中”、指示を終えて腕を戻すまでの“戻し動作”の 3 段階の動作を定義し、アノテーションを行った。

これら 3 つのデータの時間的同期を取った後、撮影ビデオとアノテーションデータを一つの anvil データに統合し、会話コーパスデータを作成した。図 10 にその一例を示す。データの総数は、被験者 8 組 3 条件の全 24 会話、各 5 分（各会話 10 分間の内、後半 5 分のみを取扱いデータを使用）で、総計 120 分のコーパスデータを構築した。

4.3 指示対象の参照の円滑さに関する結果

どちらかの被験者が、対話相手の空間にある物に対して、「それ」、「これ」といった指示語や、対象物の名称を指す発話や指差しによって、物を指示した際に、対話相手が

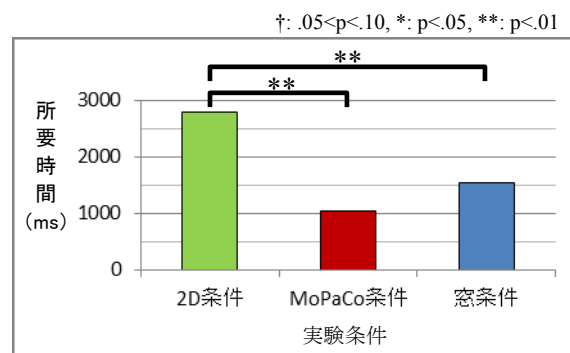


図 11 指示対象の参照の所要時間

その対象物を参照するまでの時間を測定した。測定時間は、指示者が指差しを伴って指示語や名称を示す発話を開始してから、聞き手の参照行動として、対話相手が対象物を注視開始するまでの応答時間とした。なお、2D 条件と MoPaCo 条件では、映像および音声遅延が 300ms あるため、指示者の声を実際の発話よりも 300ms 遅れてスピーカーから再生された時刻を開始時間とした。全会話での参照所要時間の平均を図 11 に示す。図 11 を見ると、2D 条件で所要時間が平均約 2.8 秒と最も時間がかかっており、MoPaCo 条件、窓条件に比べて分散分析の結果、有意差が認められた ($F(83,107)=3.94, p<.01$ (2D 条件と MoPaCo 条件), $F(83,102)=3.81, p<.01$ (2D 条件と窓条件))。MoPaCo 条件は約 1 秒程度、窓条件で約 1.5 秒であり、指示がされてすぐに対象物を参照可能であったと言える。以上より、仮説が正しいことが示唆された。

4.4 主観評価の結果

被験者の主観評価の平均値を図 12 に示す。図 12 を見ると、会話の円滑さや意思の疎通の項目で、2D 条件に比べて MoPaCo 条件と窓条件の評価値が高く、対応のある t 検定の結果、有意差が認められた（会話の円滑さ： $t(15)=2.82, p<.05$ (2D 条件と MoPaCo 条件), $t(15)=2.06, p<.05$ (2D 条件と窓条件), 意思の疎通： $t(15)=2.03, .05<p<.10$ (2D 条件と MoPaCo 条件), $t(15)=1.91, p<.05$ (2D 条件と窓条件))。また、MoPaCo 条件と窓条件で有意な差は見られなかったことから、MoPaCo 条件で窓条件と同等に、意思

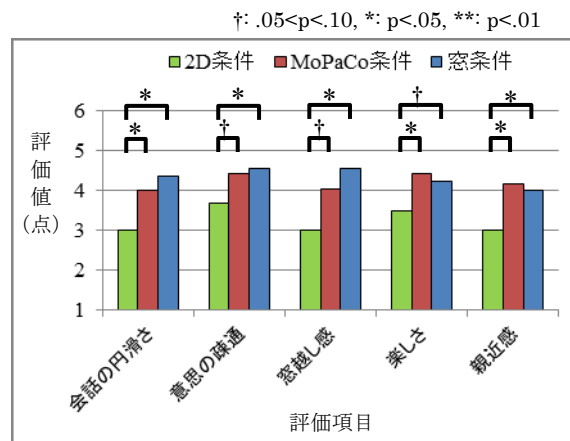


図 12 主観評価の結果

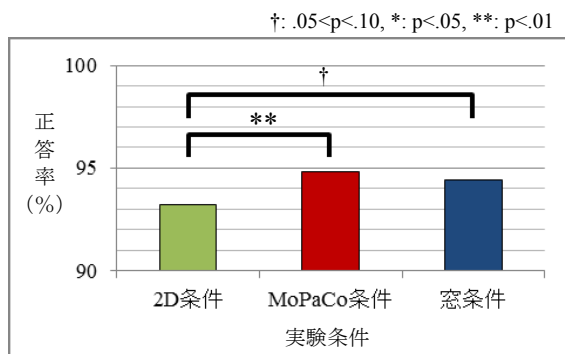


図 13 家具の記憶力の結果

の伝達および会話が円滑に行われたと考えられる。さらに、窓越し感、楽しさ、親近感の項目においても、2D 条件よりも MoPaCo 条件と窓条件の評価値が高く、対応のある t 検定の結果、有意差および有意傾向が認められた（窓越し感： $t(15)=2.76$, $.05<p<.10$ （2D 条件と MoPaCo 条件）、 $t(15)=4.14$, $p<.05$ （2D 条件と窓条件）、楽しさ： $t(15)=2.52$, $p<.05$ （2D 条件と MoPaCo 条件）、 $t(15)=2.28$, $p<.05$ （2D 条件と窓条件）、親近感： $t(15)=3.89$, $p<.05$ （2D 条件と MoPaCo 条件）、 $t(15)=2.57$, $p<.05$ （2D 条件と窓条件））。

次に、対話相手の部屋に会った家具についての回答結果を図 13 に示す。図 13 の結果は写真で示された 84 個の家具についての回答の正誤を集計し、被験者 16 人の正答率の平均を算出したものである。2D 条件に比べて、MoPaCo 条件と窓条件で正答率が高く、対応のある t 検定の結果、それぞれ有意差および有意傾向が見られた（2D 条件と MoPaCo 条件： $t(15)=3.37$, $p<.01$ 、2D 条件と窓条件： $t(15)=1.18$, $.10<p<.05$ ）。2D 条件よりも、MoPaCo 条件および窓条件において、家具の記憶力が向上したことが確認された。

5. 考察

指示者が対話相手の空間にある物を指示し、対話相手が参照するまでの所要時間が、MoPaCo 条件で窓条件と同等に短かったこと、主観評価にて、会話や意思の疎通の円滑さが同等の評価であったことから、MoPaCo を利用することで実際の窓を介した状況と同様に、指示対象の伝達や会話が円滑に行われたと言える。すなわち、メディアスペースにおいて 2 つの空間が窓越しに繋がったかのように、大きさと位置関係を考慮した窓越し映像を提示することにより、指示動作の参照が円滑になることが示唆された。この理由として、実験における被験者の行動から 2 つの理由が考えられた。

1 点目は、MoPaCo ではお互いの空間の大きさ、位置関係が幾何学的に整合性を保って表現されるため、対話相手の空間の物への指差しや注視の対象が対話相手に正しく伝達できる点である。2D 条件では、聞き手が指示者が実際に指示した方向とは逆の方向を誤って参照し、指示者が対象の名称や具体的な位置を発話で補うシーンが頻繁に見られた。その一例を図 14 に示す。まず、指示者である被験者 A が、聞き手である被験者 B の右後ろにあるカチンコに注視および指差しをしながら、「じゃあ、そのカチッてするやつあるじゃん。カチンコって言うんだっけ？」と指示を送った（会話開始から 7 分 7 秒 947）。その際、被験者

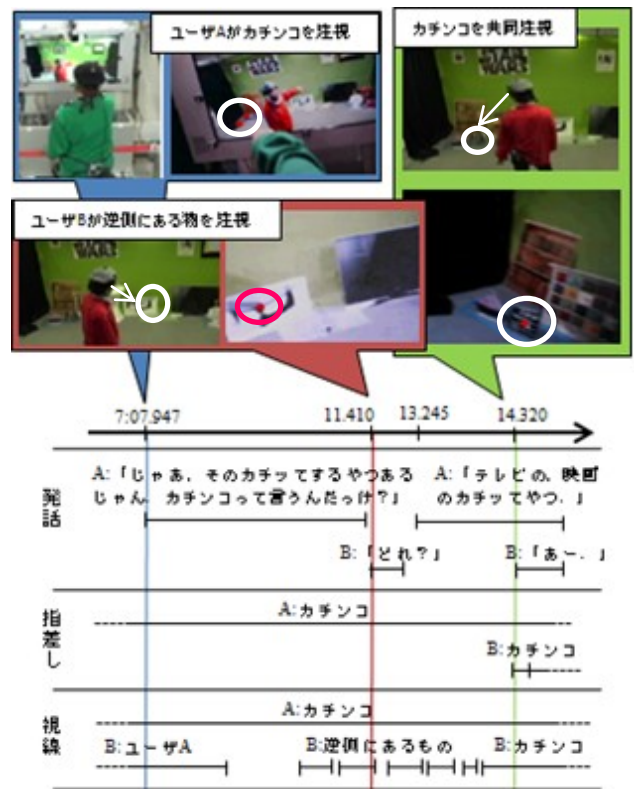


図 14 2D 条件での指示動作と参照の様子

B は被験者 A を注視しており、すぐに対象物を参照しようとするが、指示方向とは逆の左側に振り返り、視線を向けてしまった。その後、被験者 B は被験者 A に「どれ？」と問いかけを行った（7 分 11 秒 410）。それに対して、被験者 A は、「テレビの、映画のカチッてやつ。」とより詳細なカチンコの説明を行った（7 分 13 秒 245）。その後、ようやく被験者 B はカチンコに視線を向け参照をし、指差しをしながら、「あー。」という発話を行った（7 分 13 秒 840）。このように、2D 条件では指示者の視線や指差し方向が正確に伝達されないため、指示対象の参照のための確認作業が多く見られた。これに対して、MoPaCo 条件、窓条件では、このような確認作業はほとんど見られなかった。すなわち、MoPaCo により、空間の大きさ位置関係を窓越しに再現されることで、視線、指差し方向の伝達が正確に行え、指示対象の伝達が実際の窓と同様に円滑に行えたものと考えられる。

2 点目は、窓メタファによって、物を注視する際にユーザの身体移動を伴うことにより、体の位置や向きの変化がより顕著になり、対話相手の興味対象を予期でき易くなっていたことが考えられる。2D 条件では、部屋全体がディスプレイに表示されるため、物を注視する際に頭部の位置や向きを変えずに視線のみを移動させていた。注視方向を映像中の対話相手の眼球運動のみから読み取ることは難しく、そのため、被験者は対話相手の注視方向に合わせて、体を同じ方向に向けたり、共同注視を行うことがほとんど見られなかった。これに対して、MoPaCo 条件および窓条件では、対話相手が物を注視する際に、多くの場合、身体位置や方向の変化を伴うため、それに合わせて、被験者が

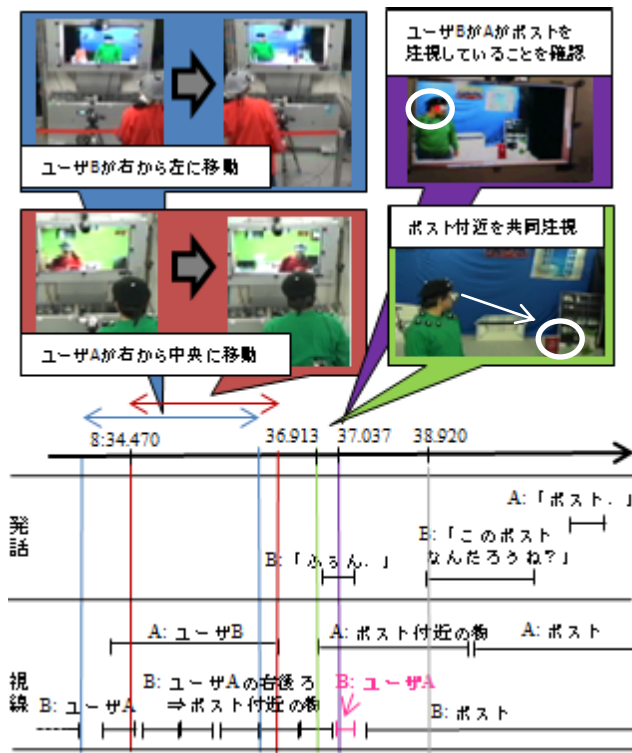


図 15 MoPaCo 条件での共同注視の様子

立ち位置を移動したり、同じ方向に体を向けることが多く見られた。MoPaCo 条件におけるこのような行動の一例を図 15 に示す。まず、被験者 B が、被験者 A の空間にある物を右から左に移動しながら観察をしていた。ユーザ A はユーザ B を注視し、その移動に気づくとすぐに、ユーザ B の位置から常に自分が見えるように、ユーザ B の移動に合わせて右から左に移動を始めた（会話開始から 8 分 34 秒 470）。ユーザ B が移動を終えると、ユーザ A も移動を止め、ユーザ B の注視方向へ体を向け共同注視を行った（8 分 36 秒 913）。これは、ユーザ B の次の指示や説明を予期する行動に当たると考えられる。次に、ユーザ B はユーザ A に視線を向け、ユーザ A がユーザ B 自身の注視対象に共同注視していることを認識し（8 分 37 秒 037）、共同注視対象であるポストについて「このポストなんだろうね？」という発話を伴う指示を行った（8 分 38 秒 920）。このとき、ポストを参照するに当たって、「この」という指示語をユーザ B は使用したことは、ユーザ B が、ユーザ A が興味の対象を共有し、次の説明や指示を予期している状態にあることを理解していたと考えられる。このようなインタラクションが MoPaCo 条件、窓条件でのみ多く見られた。MoPaCo や実際の窓越し状況で、このような行動が見られた要因として、窓をメタファとして物を覗きこむ行為が、ユーザの興味対象（方向）をより強固に表出した可能性がある。すなわち、ユーザが覗き込むことで通常よりも身体の方角や頭部方向の変化が増幅され、対話相手がよりユーザの興味対象を理解できたと考えられる（図 16 の例では、ユーザ A はユーザ B の移動に反応して移動を行い、共同注視を行った）。MoPaCo は窓越し表現をするため、本研究での実験状況のようにある程度の広さを持つ空間を観察する際に、全ての物を一度に観察することができず、

あえて移動をしなければならない。しかしながら、単純に部屋の全てを撮影し、俯瞰可能な 2D 映像を表示するよりも、非言語行動の伝達が円滑になることが確認された。よって、指示動作を伴う協調作業をより円滑に行う目的であれば、全体が常に俯瞰できなくても、より自然に視線や指差しといった非言語情報が伝達できることが重要であり、その観点から窓越しインタフェース MoPaCo は有効であると考えられる。

また、主観評価にて楽しさや親近感の向上が見られたのは、このような円滑な非言語コミュニケーションによる、副次的効果であったと考えられる。さらに、記憶力が向上したことは、MoPaCo がより映像として印象的なものであり、より会話に積極的に参加（engage）させる効果もある可能性が考えられる。

6. まとめ

ディスプレイを窓に見立てて 2 つの空間が窓越しに繋がっているかのような映像を提示可能な、窓越しインタフェース MoPaCo において、遠隔指示作業において重要な指示動作の参照が円滑に行われるかを実験的に検証した。実験の結果、MoPaCo では、実際の窓を介した対面状況と同等に円滑にユーザが対象物を参照できること、また対話相手の次の説明対象を予期することが可能であることが示唆された。また、主観評価から、MoPaCo は円滑な会話や意思の疎通を促進すること、会話の記憶力が向上することが分かった。

参考文献

- [1] Suzuki, N. et al, "Nonverbal behaviors in cooperative work: a case study of successful and unsuccessful team", Proc. of CogSci2007, pp.1527-1532 (2007).
- [2] 松田 昌史, 松下 光範, 苗村 健, "社会的分散認知環境における集団課題達成の促進要因: 集団成員間の親密さの影響", 電子情報通信学会論文誌, J90, D (4), pp.1043-1054 (2007).
- [3] C. Heath and P. Luff, "Disembodied Conduct: Communication through Video in Multi-Media Office Environment", in Proc. CHI'91, pp.99-103 (1991).
- [4] Ryo Ishii, Shiro Ozawa, Harumi Kawamura, Akira Kojima, "MoPaCo: High telepresence video communication system using motion parallax with monocular camera", Proc. of ICCV Workshops, pp. 463-464 (2011).
- [5] P.Auer, "Projection in Interaction and Projection in Grammar", Text, vol.25, pp.7-36 (2005).
- [6] Goodwin, C., "Conversational Organization: Interaction between speakers and hears", Academic Press, New York (1981).
- [7] H. Kuzuoka, "Spatial workspace collaboration: a SharedView video support system for remote collaboration capability", in Proc. CHI '92, pp.533-540 (1992).
- [8] D. Kirk, A. Crabtree, and T. Rodden, "Ways of the Hands", in Proc. ECSCW'05, pp.1-22 (2005).
- [9] H. Kuzuoka, J. Kosaka, K. Yamazaki, Y. Suga, A. Yamazaki, P. Luff, and C. Heath, "Mediating Dual. Ecologies," in Proc. CSCW2004, pp.477-486 (2004).
- [10] 三上 弾, 大塚 和弘, 大和 淳司, "メモリベースパーティクルフィルタ: 状態履歴に基づく事前分布予測を用いた頑健な対象追跡", 電子情報通信学会論文誌, Vol.J93-D, No.8 (2010).
- [11] S. M. Anstis, J.W.M., and T. Morley, "The perception of where a face or television portrait is looking", American Journal of Psychology, Vol. 82(4), pp. 474-489 (1969).
- [12] Tobii glass, <http://www.tobiiglasses.com/scientificresearch/>
- [13] Kipp, M., Anvil - A Generic Annotation Tool for Multimodal Dialogue. in the 7th European Conference on Speech Communication and Technology, pp.1367-1370 (2001).