

Title	遠隔協同作業環境のための人の状態推定と視線表現の研究
Author(s)	磯, 和之
Citation	
Issue Date	2020-09
Type	Thesis or Dissertation
Text version	ETD
URL	http://hdl.handle.net/10119/16994
Rights	
Description	Supervisor: 由井 蘭 隆也, 知識科学研究科, 博士

博 士 論 文

遠隔協同作業環境のための人の状態推定と
視線表現の研究

磯 和之

主指導教員 由井 菌 隆也

北陸先端科学技術大学院大学
知識科学研究科

2020 年 9 月

目次

第 1 章	序論	1
1.1	研究の背景	2
1.2	遠隔協同作業における円滑なコミュニケーション支援	2
1.3	本論文の貢献	5
1.4	本論文の構成	7
第 2 章	関連研究・関連知識	8
2.1	緒言	9
2.2	協同作業とアウェアネス支援	9
2.3	遠隔協同作業環境におけるコミュニケーションシステム	11
2.4	協同作業を支援するプレゼンス共有に関する研究	13
2.5	リアルタイムコミュニケーションにおける視線表現に関する研究	17
2.6	結言	19
第 3 章	遠隔協同作業環境のための人の状態推定手法と視線表現手法の提案	20
3.1	緒言	21
3.2	円滑なコミュニケーションを支援する遠隔協同作業環境	21
3.3	円滑な遠隔協同作業環境のためのプレゼンス共有システム	23
3.4	円滑な遠隔協同作業環境のためのリアルタイムコミュニケーションシステム	25
3.5	結言	26
第 4 章	人の作業負担を軽減する状態推定器の構築手法の研究	28
4.1	緒言	29
4.2	研究の目的	29
4.3	推定器を構築する 4 つのステップ	31
4.4	実験 1：ラベリング作業の対象となる映像シーンの削減効果の検証	36

4.5	実験1の結果に対する考察	46
4.6	実験2:対話インタフェースを活用した推定器構築手法の検証	47
4.7	実験2の結果に対する考察	58
4.8	結言	59
第5章	二画面積層表示を用いた視線表現手法の研究	61
5.1	緒言	62
5.2	研究の目的	62
5.3	二画面積層表示による視線方向の表示手法	63
5.4	提案手法の表現性能に関する実験	66
5.5	基礎実験の結果に対する考察	73
5.6	プロトタイプシステム	75
5.7	考察	89
5.8	結言	90
第6章	考察	92
6.1	緒言	93
6.2	本研究成果の遠隔協同作業における活用	93
6.3	本研究成果の活用領域の拡大	95
6.4	知識科学への貢献	96
6.5	結言	99
第7章	結論	100
7.1	本論文のまとめ	101
7.2	今後の課題	102
	参考文献	105

表目次

4.1	クラスタリング法の分類精度と計算速度の比較	33
4.2	特徴量を計算するための設定値	41
4.3	各作業環境において収集したデータ量	41
4.4	選択ステップの結果	41
4.5	特徴量を計算するための設定値	53
4.6	実験結果	53
5.1	基礎実験用評価機のディスプレイに関する主な仕様	67
5.2	基礎実験 1 の結果	70
5.3	基礎実験 2 の結果	71
5.4	基礎実験 3 の結果	73
5.5	プロトタイプシステムの端末表示部分の主な仕様	76
5.6	映像生成をしている PC の主な仕様	77
5.7	正面を向いている評価映像による実験結果 (model1)	85
5.8	正面を向いている評価映像による実験結果 (model2)	86
5.9	正面を向いている評価映像による実験結果 (model3)	87
5.10	評価が用映像の視線が左右に変化したときに最も視線が合うと感じる角度 を測定した結果	88

目次

2.1	実空間を接続してコミュニケーションする形態	11
2.2	仮想空間を共有してコミュニケーションする形態	12
2.3	1つの実空間に代理となるロボットなどを置いてコミュニケーションをする形態	12
2.4	センサーの種類から分類する人の状態を推定する技術	15
3.1	自宅で作業している様子をオフィスで表現している事例	22
3.2	自宅で休憩している状態をオフィスの共有スペースで表現している事例	23
3.3	非同期作業状態で協同作業者同士の状態を共有するプレゼンス共有システム	24
3.4	本研究で推定する作業状態の種類	25
3.5	オフィスの空間を共有しながら同期作業が可能なコミュニケーションシステム	26
4.1	作業者が行っていた作業を端末が確認している様子. 端末はあらかじめ分類したデータを使って選択された映像を作業者へ提示する.	30
4.2	教師有り機械学習を用いた推定器構築方法と提案手法の比較	31
4.3	マイクロフォンと距離センサからのデータ収集と特徴量の計算方法	32
4.4	分類結果に基づいた代表シーンの選択方法	35
4.5	プロトタイプシステムの構成 (a) ハードウェア構成. (b) プロトタイプシステム内のデータの流れ	37
4.6	プロトタイプ端末	38
4.7	作業環境1のレイアウトと, プロトタイプ端末とカメラの位置. 部屋および家具の大きさは(横幅 x 奥行き)で表記. 下段の写真は作業環境中のカメラから撮影した作業環境の様子	39

4.8	作業環境 2 のレイアウトと、プロトタイプ端末とカメラの位置、部屋および家具の大きさは（横幅 x 奥行き）で表記 段の写真は作業環境中のカメラから撮影した作業環境の様子	40
4.9	WS1 および WS2 における作業者の状態	43
4.10	WS1 における推定結果（5 時間分）	44
4.11	WS2 の推定結果（6 時間分）	45
4.12	作業者の状態推定とプレゼンス共有が可能なプロトタイプシステムの構成 .	48
4.13	プロトタイプシステムの構成	49
4.14	プロトタイプシステムで使用されるメイン端末とカメラ端末	49
4.15	マイクロフォンと距離センサからのデータ収集と特徴量の計算方法	50
4.16	作業環境 1 のレイアウトと、センサ端末とカメラ端末の位置、部屋および家具の大きさは（横幅 x 奥行き）で表記、下段の写真はカメラ端末から撮影した作業環境の様子	51
4.17	作業環境 2 のレイアウトと、センサ端末とカメラ端末の位置、部屋および家具の大きさは（横幅 x 奥行き）で表記、下段の写真はカメラ端末から撮影した作業環境の様子	52
4.18	チャットを介して作業者とシステムが対話を行う様子、システムから作業 者へ映像と質問を送り、作業者は映像を確認した後返答する	55
4.19	作業環境 1 において推定が正しかったときの例	56
4.20	作業環境 2 において推定が正しかったときの例	57
5.1	将来的なシステムイメージ	63
5.2	二画面積層表示による 3 次元表示の原理	64
5.3	二画面積層表示による顔の向きの表現	65
5.4	二画面積層表示による顔の向きの制御	66
5.5	基礎実験用機器の構成	67
5.6	基礎実験の評価時に使用した画像	68
5.7	基礎実験用の評価機へ実験用画像を表示した時の様子	69
5.8	映像のずらし量と視線が一致する範囲の関係	72
5.9	ずらし量と目があると感じる範囲の関係	75
5.10	プロトタイプシステムの構成	77
5.11	二画面積層表示を用いたプロトタイプシステムの端末	78
5.12	リアルタイム撮像部の構成	79

5.13	リアルタイムに生成した映像	79
5.14	プロトタイプ端末を使って顔の向きおよび視線の向きを表現している様子 .	80
5.15	評価用画像の撮影条件	81
5.16	撮影された評価用画像	82
5.17	実験時のシステムの構成	84
5.18	実験者が映像を観察している様子	84
5.19	実験結果 (Model1)	85
5.20	実験結果 (Model2)	86
5.21	実験結果 (Model3)	87
5.22	プロトタイプシステムによる実験結果	89
6.1	それぞれの空間に応じて情報を表現する形態	95
6.2	生活融合通信:ComAdapater による生活空間の実現イメージ	97
6.3	生活融合通信:ComAdapater のプロトタイプシステムの実験の様子	98

第 1 章

序論

1.1 研究の背景

企業や組織における革新的なサービスや技術の創出には、多くの課題解決が必要である。そうした課題解決では、様々な分野の知識を持った人々の協力が求められる。アレンら [1] は、技術情報のコミュニケーションのあり方が新製品のイノベーションプロセスに決定的な影響を及ぼすと述べている。さらに、組織の内外の人々と幅広く連携したサービス開発や技術開発の事例が増加している [2]。複数の企業や組織に所属する人たちは、互いに離れた場所にいながら協力して作業を行うため、遠隔協同作業をおこなう機会は増加する傾向にある。また、日本国内では、ワークライフバランスの向上、多様な人材活用の観点から、柔軟な働き方としてテレワークが注目され、制度を導入する企業は増えている [3]。

一方で、大規模災害の発生時や感染症への対策としてもテレワークは注目されており、1つの職場に集まることなく、企業の活動を継続出来る働き方として注目される。高速/広帯域ネットワークの普及と、ICT技術の発展に伴い、多地点のテレビ会議システムや、チャットシステムなど、様々なコミュニケーションシステムが普及してきた。このようなシステムを使用し、映像や音声、文字通信、資料共有など複数のメディアを使った会議が行われている。複数の人が協力して遠隔で作業をする機会は、今後も増加していく傾向にある。

アレンら [1] は、技術情報のコミュニケーションを活発にするためには、そこで働く人々が所属する組織構成と、働く場所の空間構成の両方が重要と述べている。組織構成は人同士の関係性を明確にする。空間構成は、関係のある人たちが対面してコミュニケーションする機会をつくる。フォーマルな会議に加えて、人々が偶然に出会い、序列や階層にとらわれることなく率直にコミュニケーションする空間を作ることで、イノベティブなアイデアが生まれると述べている。遠隔で働く人同士では、そうした共有する空間がないため、コミュニケーションの回数が減少する問題や、通信システムを介したコミュニケーションに変化することで、意図通りに情報が伝わらない問題が生じる。

1.2 遠隔協同作業における円滑なコミュニケーション支援

協同作業は、「非同期作業」と「同期作業」の2種類に分類される。「非同期作業」は、資料の作成や、ソフトウェアのプログラミングなど、個人が担当する作業である。「同期作業」は、複数の人が集まって行う会議や、実験など、複数の人が同期して行う作業である。「同期作業」では、作業者同士のコミュニケーションが必要となる。「同期作業」は、時間や場所を調整しておいて開始される会議などがある。また、協同作業者同士が顔を合わせたことに

より、声をかけて偶然に始まる会話も含まれる。このようなインフォーマルに開始されるコミュニケーションも、協同作業において必要なコミュニケーションと考えられている [4, 5]. 遠隔協同作業で円滑なコミュニケーションを支援するためには、協同作業同士が互いの状態を共有し、適切なタイミングで非同期作業から同期作業へ移行できる環境が必要である。また、同期作業をしながらコミュニケーションするときには、あたかも対面で対話しているかのような通信環境が求められる。

遠隔共同作業を支援するコミュニケーションシステムのための様々な方式が研究されている。オフィスにテレプレゼンスロボットを設置し遠隔協同作業を行った研究 [6, 7] では、ロボットが遠隔の作業者の存在を表現し、オフィスにいる作業者と対話しながら作業できることが報告されている。しかし、「非同期作業」、「同期作業」のそれぞれにおいて、円滑なコミュニケーションを実現するための課題がある。

遠隔で「非同期作業」を行う場合、作業者同士が互いに離れた場所でそれぞれの作業をしているため、互いの状態を認識することが困難になる。遠隔にいる人が、どのような状態にあるかわからないため、話しかけて良いタイミングがわからず、会話が始まらないことや、突然電話をして、作業を中断し作業効率を低下させることがある。こうした問題について従来から研究されている [8, 9, 10]. Lee ら [7] の研究において、テレプレゼンスロボットは、オフィス内のロボットがいる位置によって、遠隔にいる作業者の状態を表現している。しかし、この状態を表現するためのロボットの操作は、遠隔にいる作業者にとっては煩わしいことが報告されている。本研究は、遠隔にいる作業者の状態を自動的に推定し、オフィスに設置されたテレプレゼンスロボットを介して、遠隔の作業者へ自然と作業状態を伝達することで、コミュニケーションを始めるきっかけを作り、連帯感を醸成しながら、円滑にコミュニケーションを始められる環境構築を目指す。1.2.1 節にて、新たな状態推定手法の概要について述べる。

次に、「同期作業」では、協同作業者との間でリアルタイムにコミュニケーションする必要がある。同じ空間にいる人同士がリアルタイムにコミュニケーションする場合、言語的な情報に加え、ノンバーバルな情報も同時に伝達され重要な役割をもつ。ノンバーバルな情報とは、身体動作（表情、視線、動向、身振り、姿勢など）、生理的行為、対人接触、対人距離、周辺言語などがある [11]. その中でも、視線は人の発話と関係がよく、話者交替、会話の調整、伝達意図の表明などの機能があると言われている [12]. 遠隔にいる人同士は通信システムを使ってコミュニケーションを行う。テレプレゼンスロボットは、遠隔にいてもオフィスにいる作業者と協同作業を行うシステムとして着目されている [6, 7]. テレプレゼンスロボットは、遠隔にいる作業者の映像を伝達するための二次元ディスプレイが搭載されている [13, 14, 15, 16]. しかし、こうした二次元の映像では、三次元情報を正しく表現できな

いため、視線や顔の向きなどの三次元情報が人の意図した通りに伝わらない問題がある。そこで、本研究では遠隔コミュニケーションにおける視線表現に着目し、新たな視線表現手法を提案する。1.2.2 節にて、手法の概要を述べる。

1.2.1 「非同期作業」でプレゼンス共有するための人の状態推定手法

遠隔協同作業における人の存在や状態に関する情報共有は、プレゼンス共有と呼ばれる。協同作業者同士の連帯感の醸成や、コミュニケーションを始める機会の増加、作業の中断を抑止するための研究が行われている [17, 18]。

プレゼンス共有には、人の状態を自動的に推定する技術が求められる。従来研究では、PC の操作ログからの認識する手法 [19] や、特定の部屋にいることを RFID で識別する手法 [17] が提案されている。近年、機械学習技術の発展に伴い、部屋の中で行われている人の行動を推定し細かく分類する技術が報告されている [20, 21]。これらの機械学習を用いた手法は、センサーによって収集した大量のデータを使い、学習用データを作成する。学習用データを作成する際、手作業で収集したデータへ正解情報（ラベル）を付与する。さらに、遠隔協同作業の場合には、作業環境の形状やそれぞれの作業環境毎の作業内容が異なるため、環境毎に学習データが必要となる。これはシステムを導入する際の負担になる。

そこで、本研究では人とシステムが協調して推定器を構築することで、人が行う作業の負担を軽減する手法を確立した。

1.2.2 「同期作業」のコミュニケーションを円滑にする視線表現

遠隔協同作業において、複数の人が同期して行う作業ではリアルタイムのコミュニケーションが必要となる。現在の二次元ディスプレイを使ったテレプレゼンスロボットやテレビ会議システムでは、人が意図したとおりの方向に視線を表現することができない。二次元ディスプレイに表示されたカメラ視線の映像を、複数の参加者が同じ場所にしながら同時に見ると、映像を見ているすべての会議参加者が視線を合っていると感じる。これは、モノリザ効果と呼ばれる。逆にカメラのレンズを見ない状態で撮影された映像では、誰とも視線を合わせることはできない。

この問題を解決するためには、複数の方向に異なる映像を同時に表示しなければならない。従来方式では、多視点の 3D ディスプレイを使った提案 [22] や、回転するミラーを使った方式 [23]、顔の形状のスクリーンを使う方法 [24] など多くの手法が提案されている。しかし、会議参加者の数に制限があることや、大型の装置が必要なこと、一人一人専用のスクリーンが必要になるなど、それぞれの方式に課題がある。

そこで、本研究では、2枚の映像を積層する3D表示手法を応用し、視線表現する手法を確立した。表示装置は小型化が可能である。テレプレゼンスロボットへ搭載することによって、遠隔にいる作業者は、オフィスにいる作業員へ視線情報を自然と伝達しながら、リアルタイムにコミュニケーションすることができる。

1.3 本論文の貢献

サービス開発や技術開発を遠隔にいる人同士が連携して行う機会が増えている。そうしたイノベティブな活動において、協同作業員間のコミュニケーションが不可欠である。円滑なコミュニケーションを支援するためには、協同作業員同士が互いの状態を共有し、適切なタイミングでコミュニケーションを開始し、あたかも対面で対話しているかのような通信環境が求められる。そこで、遠隔協同作業環境の課題を整理し、円滑なコミュニケーションを支援する解決方法として、作業員の状態を自動的に推定する技術と、人が意図した通りに視線情報を表現できる新たな手法を考案し、その性能を実証した。

協同作業は「非同期作業」と「同期作業」に分類される。非同期作業では、適切なタイミングで円滑に同期作業へ移行する機能が必要である。作業を円滑に移行するためには、遠隔協同作業員同士の状態を共有する機能（プレゼンス共有）が有効である。人の状態を自動的に推定できる従来手法について、遠隔協同作業環境への適用時の課題を整理し、それらの課題を解決する新たな人の状態推定手法を考案した。

また、同期作業は人の周辺にある空間やモノの情報も含めて共有しリアルタイムにコミュニケーションする通信システムが求められる。そうした環境を実現するための通信システムとして、テレプレゼンスロボットの活用が注目されている。しかし、遠隔作業員の意図した通りに視線情報を表現できないため、コミュニケーションが円滑に進まない問題がある。視線情報を表現可能な表示手法について、遠隔協同作業環境への適用時の課題を整理した。その課題を解決するために、2枚の映像を積層して表示する3D映像表示手法を応用し、小型で様々な人の顔を表現できる新たな視線表現手法を考案した。

貢献 1. 遠隔協同作業をしている人の状態を推定するシステムを開発しその性能を実証した。

- a) 既存の人の状態推定手法では教師あり機械学習が用いられ、人の状態を推定し細かく分類できることが示されている。しかし、遠隔協同作業環境へ適用するためには、それぞれの環境毎に大量の学習データを作成する必要がある。学習データの作成の負担軽減が課題になることを明らかにした。

- b) 教師無しクラスタリング手法と、ラベル付与の対象データを分類結果に基づき自動的に選択する手法を使い、少ない作業負担で推定器を構築する手法を考案し、実験用のプロトタイプシステムを開発した。
- c) 作業環境毎のラベルを付与し推定器を構築するため、対話インターフェースを用いて、システムと人が協調して推定器を構築する手法を考案し、実験用のプロトタイプシステムを開発した。
- d) 開発したプロトタイプシステムを使い性能を検証した結果、従来手法と比較し、同等の正解率を維持し、学習データにラベルを付ける作業を 60%~80% 削減できることを実証した。
- e) 対話インターフェースの用いることで、ラベル付与の作業負担を軽減しつつ、各作業環境に応じた状態を定義できることを実証した。

貢献 2. 二画面積層表示による視線表現手法を用いたリアルタイムコミュニケーションシステムを開発し、その性能を実証した。

- a) 視線表現が可能な表示装置の関連研究を調査し、表示装置、撮影装置の大型化、作業者個人専用のスクリーン製作が必要なことが、遠隔協同作業環境へ適用する際の問題となることを明らかにした。
- b) テレプレゼンスロボットに搭載可能な小型の表示装置で、視線方向を表現できる視線表現手法として、2枚の映像を重ね合わせた二画面積層表示を応用することを考案した。
- c) 二画面積層表示による視線の基本的な表現性能を検証するために基礎実験用のシステムを開発した。
- d) 基礎実験の結果、視線が一致する範囲が約 17.5° の範囲にあること、積層する画像を左右にずらすことで、視線が合う範囲を左右に変化させることができることを実証した。
- e) 基礎実験の結果を受け、リアルタイムに映像伝送し、コミュニケーションすることが可能なプロトタイプシステムを開発した。
- f) 基礎実験およびプロトタイプシステムにより、視線の表現性能を検証した結果、向き合った実験者が特定の方向でのみ視線が合っていると感じ、他の方向を向いているときには、自分に視線が合っていないと感じることを実証した。視線が合うと感じられる範囲は、17° ~21° となることを実証した。

1.4 本論文の構成

本論文の構成について述べる。本論文で紹介する手法や、システム、そこから得られた知見については、後述の論文を基にした。

2章では関連知識・関連研究について紹介する。はじめに、協同作業とウェアラブル支援に関する研究について紹介する。次に、遠隔協同作業のためのコミュニケーションシステムの研究についてまとめる。さらに、人の状態の共有を行うプレゼンス共有システムに関する研究と、リアルタイムにコミュニケーションする時の視線表現に関する研究をまとめる。

3章では、円滑なコミュニケーションを支援する遠隔協同作業環境のための人の状態推定手法と、視線表現手法の2つの手法について説明する。最初に、遠隔協同作業環境で求められる円滑なコミュニケーションのための支援について課題を整理し、本研究が目指す遠隔協同作業環境を示す。次に、協同作業を「同期作業」、「非同期作業」に分類し、それぞれの作業に対しての課題とその解決手法を述べる。非同期作業から同期作業への円滑な移行を支援するためのプレゼンス共有システムの実現方法と、円滑な同期作業を支援するためのリアルタイムコミュニケーションシステムの実現方法について述べる。

4章では、プレゼンス共有システムを実現するための人の状態推定手法について述べる。学習用データの作成負担を軽減し推定器を構築する手法について、新たな手法を考案し、その性能について示す。この章は、International Journal of Informatics Societyに掲載の論文[25]と、日本VR学会論文誌に掲載されている論文[26]を基に記述した。

5章では、円滑な同期作業のためのコミュニケーションシステムの映像表示手法について述べる。映像表示装置は小型化が可能で、遠隔にいる人が意図した相手の方向に視線が伝達できる新たな視線表現手法について述べる。この章は、情報処理学会論文誌に掲載された論文[27]および、IEEEのJournal of Display Technologyに掲載された論文[28]を基に記述した。

6章では、4、5章にて得られた知見をまとめ、遠隔協同作業環境での活用方法とその効果について考察する。さらに、活用領域を拡大するための、作業者の状態表現方法について考察する。考察するにあたり、日本VR学会論文誌に掲載されている論文[29]を発展させた。最後に7章にて遠隔協同作業に求められる人の状態推定、視線表現について総括し、今後の課題を述べ、本論文を結ぶ。

第 2 章

関連研究・関連知識

2.1 緒言

本章では、円滑なコミュニケーションを支援する遠隔協同作業環境に関連する従来研究、関連知識について述べる。2.2 節は、遠隔協同作業におけるアウェアネス支援に関する研究を紹介する。2.3 節は、遠隔協同作業におけるコミュニケーションシステムの研究を紹介する。2.4 節は、プレゼンス共有システムと人の状態を推定する技術に関する研究を紹介する。2.5 節は、通信システムを介したリアルタイムコミュニケーションにおける視線表現に関する研究を紹介する。

2.2 協同作業とアウェアネス支援

企業や組織の中では複数の人が協力して作業が日々行われている。松下ら [30] は、協同 (collaboration) を次のように定義している。

複数の人が協力することにより、一人ではなし得ない、あるいはなすことが困難な新しい価値を創造すること

サービス開発や技術開発は企業や組織が発展するための重要な活動であり、協同作業によって推進される。近年は、サービス開発や技術開発を1つの企業や組織だけでなく、複数の企業や組織が連携して行う事例 [2] が報告されており、また、テレワークを導入する企業は増えており、遠隔協同作業の機会は増加している。

岡田 [31] の協同の一般階層モデルでは、「協同」、「共有」、「気づき」、「共在」の4つの階層でモデル化されている。

1. 協同
「個の主張」、「互いの協力」
2. 共有
「意識」、「情報」、「作業」
3. 気づき
「人間」、「場/空間」、「モノ」
4. 共在
「場所」、「時間」

また、最も高いレベルの状態にある協同を、協調とよび次のように定義される。

お互いに自分の意見をはっきり主張すると同時に相手の主張をよく理解して、それぞれの意見を統括し新しい価値を創造するのが協調である。

このモデルによると、最も基本的な要素は「共在」である。「共在」は協同作業同士が互いの存在を認知できている状態と定義され、「場所」、「時間」の2つの次元で考える。同じ「場所」、同じ「時間」で互いの存在を認知できる状態が最も高いレベルの共在とされる。互いが認知できる状態になると、「人間」、「場/空間」、「モノ」に対する様々な気づきが生まれる。「人間」に対しての気づきは、存在、動作、感情、「場/空間」に対しては、状況、雰囲気、「モノ」に対しては、形状、属性などである。こうした気づきから、協同作業同士のコミュニケーションが生まれ、「意識の共有」、「情報の共有」、「作業の共有」が円滑に行われる。これらが共有された状態で、協同作業を行うと、自らの意見をはっきりと主張しながらも、相手の主張も理解でき、互いに協力して、新しい価値を生むことができる。

遠隔協同作業では、離れた「場所」で作業が行われる。同じ「場所」にいるかのような気づき（アウェアネス）を支援するための技術が求められる。

Dourish ら [8] は、アウェアネスを次のように定義する。

awareness is an understanding of the activities of others, which provides a context for your own activity.

この定義によれば、アウェアネスは自らの活動に影響を与える他人の活動の理解であり、広い概念である。複数の人が協力して活動する場所では、周囲の人から様々な気づきが得られる。例えば、医療現場における患者からの気づきを共有するシステムが研究されている [32]。また、人の行動に関連する物や空間からの気づきもある。Kanai ら [33] は、介護現場における、しまい忘れや、置き忘れられた物体に対する気づきや、空間の中での危険性への気づきを与えるシステムを研究している。

Dourish ら [34] や、Gutwin ら [35] は、ワークプレイスのアウェアネスの要素として、「自分の周りに誰がいるか」、「何をしているのか」、「誰と誰が話をしているのか」、「何を見ているのか」、「何を見ることが可能か」、「何のモノが作業の対象か」、「どこにいるのか」など、岡田 [31] のモデルと同様に人、モノ、場/空間に関してアウェアネスの要素を挙げている。國藤ら [36] は、こうした気づき（アウェアネス）が、コミュニケーションを始めるきっかけや、協同作業者の連帯感を生むとしている。アレンら [1] は、欧州メーカの役員室の構成や、MIT のスローン・スクール・マネジメントの部屋の構成を例として、イノベティブな活動をする空間の中に、人同士が集まる空間があることに注目し、その空間の中で行われる偶然の出会いや、序列や階層にとらわれない会話から、既成概念にとらわれない率直でイノ

ベータなアイデアが生まれるとしている。

協同作業同士が、同一空間にいれば、高いレベルの共在の状態が成立し、自然と気づきが生まれる。それによって、作業同士が円滑にコミュニケーションを行い、高いレベルの協同作業を実現する。遠隔協同作業では、通信システムにより遠隔協同作業同士の存在を認知させ、様々な気づきの支援を行うことで、円滑なコミュニケーションを支援する必要がある。

2.3 遠隔協同作業環境におけるコミュニケーションシステム

リアルタイムにコミュニケーションが可能な通信システムを3種類に分類し、協同作業環境における課題について述べる。第1は、映像を表示するディスプレイを介して実際の作業空間同士を接続してコミュニケーションする形態(図2.1)である。第2は仮想的に1つの共有空間を作りコミュニケーションする形態(図2.2)、第3は1つの拠点となる実空間共有するコミュニケーション形態(図2.3)である。それぞれの形態のコミュニケーションシステムについて述べる。

第1の形態の代表的なシステムは、テレビ会議システムである[37]。大型ディスプレイを使い拠点同士を接続した状態を作る。映像の中に自らも表示し、あたかも鏡の中で遠隔の人とコミュニケーションをとる方法[38]や、実際の空間の中で、ディスプレイの配置を合わせる、同じ場所にいるかのような感覚でコミュニケーションを行える手法[39]もある。現実の空間を映像でつなぎ、1つの空間のようにコミュニケーションする環境を実現する。人は自分の周囲の空間や、周囲に存在するモノとの関係も映像として表現しながら、コミュニケー

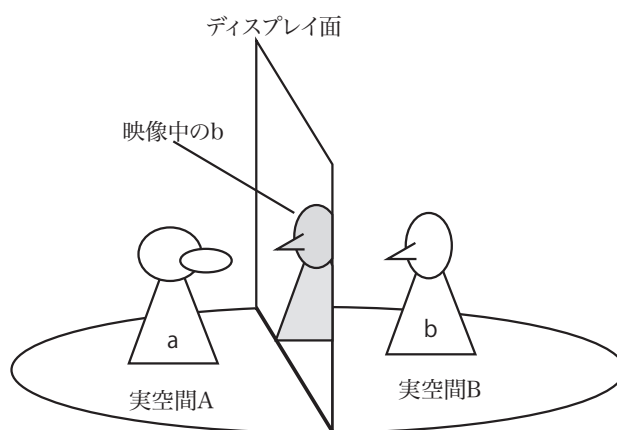


図 2.1 実空間を接続してコミュニケーションする形態

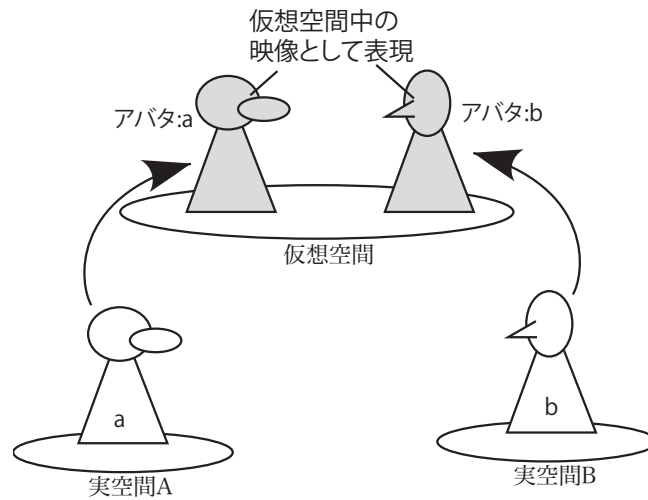


図 2.2 仮想空間を共有してコミュニケーションする形態

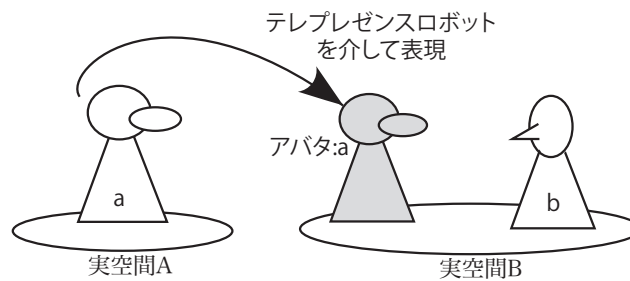


図 2.3 1つの実空間に代理となるロボットなどを置いてコミュニケーションをする形態

ションすることができる。

また、映像通信環境を使い、インフォーマルにコミュニケーションを始めるきっかけを作る研究 [4, 40, 41] がある。これらの研究では、常時、映像・音声のコミュニケーションが可能な状態を作る。例えば、オフィスの中の共用スペース同士をテレビ会議システムで接続する。作業者はその場所に行けば、互いの状態を映像から認知し、必要に応じて、映像中の遠隔の人に声をかけ、会話を始めることができる。状態確認からコミュニケーションへシームレスに遷移することができる。遠隔にいる人の状態を知るために、映像が見える場所への移動する必要がある。自宅などを利用するテレワークの場合では、共有スペースを作ることは難しく、プライバシーの問題 [42, 43] を考慮しなければならない。遠隔にいる人の状態を知りたい時に、短時間だけ映像を送りあい、相手の状態を伺ってコミュニケーションを始める研究 [44, 45, 46] もあるが、常に人の状態を共有し続けることはできない。

第2の形態は、1つの仮想的な空間を共有してコミュニケーションする方法である。コンピュータグラフィックス (CG) で作られた空間を PC に表示し、コミュニケーションする方法が研究されている [47, 48]。大型の没入型ディスプレイを使用し、実際に CG 空間の中に入り込んだような状態でコミュニケーションするシステム [49, 50, 51] がある。1つの仮想的な空間に没入することで、遠隔にいる作業者があたかも同じ空間にいて、対面しながらコミュニケーションしているような環境を実現できる。遠隔協同作業者と共有する空間の中で、作業者を表現する物をアバタと定義する。仮想空間中では、人の形状を模した CG のキャラクターがアバタとして使用される。

本田ら [47] は、仮想空間の中に、インフォーマルにコミュニケーションを行う場所と、集中して作業する場所を分けて作り、仮想空間の中でアバタを表示する場所を変え、コミュニケーションの可否を協同作業者に提示しつつ、すぐに会話が始める環境を実現している。しかし、現在の遠隔協同作業をすべて仮想空間の中で行うことや、実空間で行っている作業の様子を、仮想空間内でアバタの動作として反映することは、現実的には困難となる。

第3の形態は、テレプレゼンスロボット [13, 14, 15, 16] を用いる手法である。テレプレゼンスロボットは、オフィスの中で遠隔にいる人の代理つまり、アバタとして動作する。テレプレゼンスロボットを使うことで、遠隔にいる人の存在を、オフィスにいる人へ認識させ、職場環境の人や場/空間、モノに関する気づきを支援することができる。テレプレゼンスロボットによる遠隔協同作業を2ヶ月～36ヶ月行った研究 [6, 7] では、遠隔からロボットを操作し、会議室へ移動し、会議に参加することや、共用スペースへ移動し、休憩している人たちと会話をしながら、活用されていることが報告されている。円滑な遠隔協同作業環境を実現する形態として、実現性の高いコミュニケーションシステムの形態と考える。

2.4 協同作業を支援するプレゼンス共有に関する研究

2.4.1 協同作業に適用されたプレゼンス共有システムに関する研究

Lee ら [7] は、テレプレゼンスロボットを使って遠隔にいる作業者の状態を表現している事例を報告している。テレプレゼンスロボットを使用している遠隔の作業者は、予めオフィスで実施される会議の予定などが無い場合も、遠隔からロボットへアクセスしオフィスにいる人とのコミュニケーションができる状態を保持している。オフィスにある個室（遠隔作業者がオフィスへ行ったときに使う場所）へ、テレプレゼンスロボットを置くことで、職場側にいる人へ、声をかけても良いことを表現している。ロボットの位置や状態が遠隔にいる作業者の状態を表現している。しかし、このロボットをその場所に移動させる操作は、遠隔作

業者にとっては煩わしい作業であったことが報告されている。

遠隔にいる人の状態を表現するためだけに、ロボットを操作することは遠隔にいる作業員として煩わしい。そこで、Vaidaら [52] は、遠隔で働く人の状態を表現しながら、コミュニケーションの機会を作る方法として非対称なメディアスペースを提案している。オフィスにある個室のドアへ自動的に開閉する装置を付け、遠隔にいる人が業務を始め、協同作業員とのコミュニケーションが可能な状態になったときに、オフィスのドアが自動的に開くようにした。オフィスの個室のドアが開いている場合に、個室の中へ入りテレビ会議システムでコミュニケーションを始められる。作業員の状態を自動的に推定することによって、ドアの開閉を使って、作業員の状態を表現することができる。遠隔にいる人はドアの開閉操作を直接意識しておこなう必要はない。

Kuzuokaら [53] は、テレプレゼンスロボットによる雑談の誘発手法を提案している。テレプレゼンスロボットがオフィスにいる人に対して顔を向ける動作をすることで、テレプレゼンスロボットへ自然と注意を向けさせ、雑談の開始を促している。作業員が雑談可能な状態になったことを推定し、ロボットの顔の動作として表現することで、コミュニケーションを始めるきっかけを作ることができる。

作業員の状態共有の効果として、コミュニケーションの開始のきっかけ作りや、連帯感の醸成に寄与することが報告されている。高橋ら [17] は、RFID タグ情報、会議室の利用状況、パソコンの利用状況（電源断、アイドル、利用中など）を組み合わせ、オフィス内のプレゼンス情報システムを実装し、18ヶ月の実験を行っている。プレゼンスの情報は、連絡が取りたい相手の状態を確認する手段や、確認後にコミュニケーション手段を変更するために利用された。また、相手の状態を確認することによって安心した利用者がいたなど、心理的なつながりの支援に利用されていた。

人の状態共有に求められるもう1つの効果は、共同作業員からの声かけや電話などによる作業中断の抑止である。不意に作業が中断されることによって、個人で行っている作業の効率を低下させることが報告されている [9, 10]。通話の可否や個人の状態を、電話帳やバディリストを用いて共有する研究 [54, 55, 56, 57] がある。現在普及しているコミュニケーションシステム [58, 59] においても、同様の機能がある。通話の可否に関する状態を示すことで、不要な割り込みを抑止し、適切なタイミングでのコミュニケーションの開始を促している。Hashimotoら [19] は、作業員がPCを使っている時の操作ログから状態を推定し、作業の中断を軽減するシステムを構築している。Zügerら [18] は、コミュニケーション可否 (Availability) を協同作業員に明示する装置を使い実験を行った。結果として、作業への割り込みが低減する効果が報告されている。

高橋ら [17] は、オフィスの中の作業員の状態を推定するために、人のいる位置情報などを

活用していた，オフィスでは，事務的な作業を行う居室，会議を行う会議室，実験を行う実験室など，機能毎に部屋を分けることが可能であるが，自宅でテレワークを行う場合には，1つの部屋の中で様々な業務（PCを使った資料作成や，テレビ会議，書類作成など）が行われるため，作業者の位置だけでは作業状態を判断することは難しい．テレワークを含む遠隔協同作業を支援するためには，1つの作業環境で行われる複数の作業（状態）を分類して推定する機能が求められる．人の状態を自動的に推定する手法について，次節で述べる．

2.4.2 人の状態推定に関する研究

作業者の状態を，作業者自身に手動で入力させるのは煩わしい作業となるため，システムが自動的に推定する手法が求められる．人の状態推定技術は，センサー技術の発展や機械学習を用いた手法が研究され，推定時の正解率が向上し，分類できる状態の数も増えている [60]．図 2.4 のように，人の状態を推定するためのセンサーの種類によって，推定手法を 3 種類に分類する．

ウェアラブルセンサを用いて人の行動を推定する研究として，大内ら [61] は，スマートフォンや携帯電話に搭載されたセンサから人の行動を推定している．Muraio ら [62] は，作



図 2.4 センサーの種類から分類する人の状態を推定する技術

業者に加速度センサを装着する手法を提案している。これらの手法は、作業者に装着したセンサからデータを取得し、作業者の状態推定を行っている。センサと作業者個人を結びつけることが可能であり、作業環境内に複数の人がいる場合でも、状態を識別することが可能である。作業者がセンサを装着していれば常に情報を取ることができるため、屋内外にかかわらず情報共有が可能である。ウェアラブルセンサは電池で駆動されるため、定期的な電池交換や充電が必要となり、一時的に推定できないことがある。

作業に使われる道具から情報を取得し作業状態を推定する研究として、Hashimoto ら [19] は、作業者が使用しているパソコンやサーバ等の操作ログから作業者の状態を判別している。パソコンの操作ログから遠隔地にいる作業者の状態を推定し、作業者の状態を共有する手法を提案している。計算機上の作業に限定すると、正確かつ詳細な情報を取得することができる。江木ら [63] は、ペンに加速センサを装着し、使用者の状態を推定している。作業を PC の作業や、ペンを用いた作業に限定するため、推定できる作業を制限される。

環境に設置されたアンビエントセンサを使う手法として、住宅の水道の利用状況を監視し、自宅内の人の活動の識別をする研究が行われている [64], [65]。家全体の様々な活動を推定することはできるが、作業状態を細かく推定することは困難である。

Daniel ら [21] は、コンビニエンスストアの店員の行動やデスクワーク中の作業者の行動を推定するため、RF センサを用いて、作業者がいる位置や姿勢に関する情報を用いている。Laput ら [20] では、屋内の人の行動を推定するために、マイク、赤外線センサ、照度センサなど複数のセンサを組み合わせたセンサを用いている。これらの研究は、部屋へ設置されたセンサによって人の作業を妨げることなく人の行動を推定している。推定時の正解率も高い。ただし、これらの方法では機械学習を用いるため、手作業で正解情報（ラベル）を付与した学習用データが大量に必要である。

学習データの負担を軽減するために、教師無し学習を用いて人の状態推定を行う手法が研究されている。村上ら [66] は、看護業務行動を推定するために、教師無しクラスタリング手法を用いてセンサーデータを分類した後、看護業務に関する業務知識を用いて業務名を決定している。Riboni ら [67] は、センサーを持った家電が配置されたスマートホームにおいて、人の行動と家電の関係を記述したオントロジーを作成し、負担軽減を図っている。教師情報を用いずに状態推定を行うためには、予め蓄積されている外部の情報を用いる必要がある。

遠隔協同作業では、それぞれの作業環境の形状や、物の配置、作業の内容が異なるため、それぞれの環境毎にセンサーデータを収集する必要がある。それぞれの環境の作業が異なるため、共通の業務知識や行動内容を蓄積しておくことが難しい。学習データを作成するために、その場所で作業を行っている作業員から作業内容に関する知識を入力する必要がある。そのため、作業員の負担を軽減し学習データを作成する手法が求められる。

2.5 リアルタイムコミュニケーションにおける視線表現に関する研究

Venolia ら [6] は、テレプレゼンスロボットを介したコミュニケーションにおける問題点として、遠隔作業者の見ている情報が意図した通りに伝わらないことを報告している。遠隔にいる人はテレプレゼンスロボットに搭載されたディスプレイに表示される。そのディスプレイに表示された映像では、遠隔にいる人が何処を向いて会話しているのか、意図した通りには伝わらない。

遠隔協同作業において、視線情報の伝達が重要であり、多くの研究がされている。Ishii ら [68] は、二人の人が対面で画面上の資料を共有しながら作業を行えるシステムを実現した。対面に座っている両者が画面上に描かれた文字や図とともに、相手の注視する対象を共有しながら、あたかも対面作業をしているかのような環境が実現されている。特に視線情報は発話と強い関係があり、協同作業において視線に関する情報が対話の切り替わりの予測や作業の効率化することが報告されている [69, 70, 71]。Okada ら [72] は、スクリーン正面から、プロジェクタで映像を投影しつつ、スクリーン背面からカメラで撮影が可能なスクリーンを使い、3人の協同作業を支援するシステムを実現している。カメラとディスプレイを搭載し参加者間の位置も揃えることで視線一致を実現したシステム [73] や、仮想的に作られた3D空間の中で、2次元の顔映像の向きを変えて視線情報を伝達するシステム [74] も研究されている。これらの研究により、視線情報の伝達が遠隔協同作業において重要な役割を持つことがわかる。

しかし、これらの研究のように、2次元ディスプレイを使用する場合、ディスプレイの前には一人に限定される。2次元ディスプレイにカメラ目線で撮影された映像が表示されると、そのディスプレイの前にはいるすべての人が視線を一致していると感じるため、個別に視線を一致させることや、実際の空間の中で見ている対象などを意図した通りに伝達することができない。これは、モナリザ効果と呼ばれる。レオナルド・ダ・ヴィンチの絵画モナリザの絵画を見ると、どの角度から見ても自分の方に視線を向けられていると感じる。同様に、ニュース番組などのアナウンサーの映像を見ているときも同じように、映像を見ている視聴者は自分の方を向いて話しているように感じる。

そこで、複数の人に対し個別に視線を合わせられる様な映像を表示手法が研究されている。同じ空間にいる人の顔は、見る方向によって顔の角度が変化する。この変化により、視線の方向や顔の向きが伝達される。方向別に異なる映像を提示できるディスプレイを使うこ

とで、同じ空間にいる複数の人それぞれと個別に視線が合わせられるテレビ会議システムが研究されている [22]. このディスプレイの前にいる 2 名の会議参加者は、それぞれ異なる向きからみた対話相手の映像を見ることができると、個別に視線を合わせて会話することができる. Tachi ら [75] は、再帰性反射スクリーンと複数のプロジェクタを利用し、ロボット表面に操作者の映像を表示する. 方向別に映像が切り替わるため、ロボットの前に複数の人が立っていても、会話している時に見ている方向を表現できる. これらの手法は、ディスプレイの構造の制約から、映像をみる人数や、見る位置が制限される.

Jones ら [23] は、回転するミラーとプロジェクタの映像を使い複数の方向の映像を表示するシステムを提案している. Ito ら [76] は、円筒状の装置の中で、回転するディスプレイを使って立体的な映像を表示している. Takaki ら [77] は、二次元ディスプレイとレンチキュラーシートを用いて多視点のディスプレイを構成する. これらの方式では、表示をみる方向や人数の制限は小さくなるが、映像として表示される遠隔作業者は、複数の方向から映像を撮影する必要がある. 撮影環境が大型化すると本来の作業の妨げになり、遠隔協同作業への適用が困難となる.

テレプレゼンスロボットの頭部に、顔の向きを表現するための構造を付与する取り組み [78] もされている. ロボットの胴体部分に遠隔地にいる参加者の顔の映像を表示するディスプレイを設置し、参加者の顔の方向は、ディスプレイとは別にロボットの頭部の回転を使って実現されている. 物理的なロボットの頭部を使うことで、顔の向きを表現する. また、スマートフォンの上に視線を表現するための目玉型のデバイスを使う手法も提案されている [79]. しかし、顔の表情などを伝える映像と顔の向きや視線表現は別の部位で表現されているため、対話をする相手側の人間は、ディスプレイと視線や顔向き表現用のデバイスを交互に見ることとなる. これは、人同士の対話では本来発生しない視線の動きであり、顔の表情などの伝達と、顔の向きは同一の部位で表現されることが望ましい.

Jouppi ら [80] はディスプレイを複数並べたロボットを作成している. しかし、1つ1つのディスプレイは二次元ディスプレイのままであり、モナリザ効果があり、視線を正しく表現することは困難である. Misawa ら [24] は、人の顔の形状に合わせたスクリーンを作成し、顔の向きや視線の方向、顔の表情を伝達する手法を提案している. テレプレゼンスロボットに搭載することも可能な大きさである. しかし、スクリーンは個人個人の顔の形状に併せて製作する必要があるため、テレプレゼンスロボットを複数の人が共用することができない.

2.6 結言

企業や組織における技術開発やサービス開発は、複数の人が協力して行う協同作業である。協同作業では、人の存在や状態に関する認識や人間の周辺にある場所/空間、モノに関する気づきが、コミュニケーションを始めるきっかけ作りや、円滑な情報の共有、作業の共有、意識の共有を促進し、円滑な協同作業を実現する。

遠隔協同作業のためのコミュニケーションシステムが多く研究されている。その中でも、テレプレゼンスロボットを用いたコミュニケーションシステムは、オフィスにいる複数の作業者と、遠隔にいる作業者ととの間で、実際に協同作業を長期間行った事例が報告されている。オフィスの空間や、オフィス内のモノを共有しながら、オフィスにいる作業者と遠隔にいる作業者がコミュニケーションをとりながら、作業を進める実用的なシステムとして注目される。

遠隔の作業者は、テレプレゼンスロボットを介して、その存在や状態を伝えることができるが、状態を伝えるためにロボットを操作することが煩わしいと報告されている。作業者の状態を自動的に推定し、その推定結果に応じてオフィスの部屋のドアを開閉する研究がある。この手法を応用することによって、作業者の状態はテレプレゼンスロボットを介してオフィスの作業者に伝達することができる。

作業者の状態推定は機械学習を用いることで、高い精度で状態推定を行えるが、学習データの作成の負担が大きく、遠隔協同作業への適用が困難である。人手の作業負担を軽減した推定器構築手法の提案が必要である。

また、テレプレゼンスロボットを使ったコミュニケーションでは、二次元ディスプレイを用いているため、遠隔作業者の視線情報が意図した通りに伝達できず、対話が円滑に進まない問題があった。視線表現が可能な映像表示装置の研究も多くされているが、表示装置の大型化、撮影装置の大型化、見る方向の制限などがあり、遠隔協同作業への適用が困難であった。テレプレゼンスロボットに搭載し視線表現が可能な表示手法として、作業者の顔の凹凸に合わせた個人専用のクリーンを用いる手法が研究されているが、テレプレゼンスロボットは複数の作業者が共有して使う場面もあるため、この手法の適用も困難であった。テレプレゼンスロボットへ搭載可能な小型の表示装置で、複数の作業者の顔映像を表示できる手法が必要である。

第3章

遠隔協同作業環境のための人の状態 推定手法と視線表現手法の提案

3.1 緒言

本章では、円滑なコミュニケーションを支援する遠隔協同作業環境のための人の状態推定手法と、視線表現手法の2つの手法について説明する。3.2節は、遠隔協同作業環境で求められる円滑なコミュニケーションのための支援について課題を整理し、本研究が目指す遠隔協同作業環境を示す。次に、協同作業を「同期作業」、「非同期作業」に分類し、それぞれの作業に対する課題とその解決手法として述べる。3.3節は非同期作業から同期作業への円滑な移行を支援するためのプレゼンス共有システムの実現方法について述べる。3.4節は、円滑な同期作業を支援するためのリアルタイムコミュニケーションシステムの実現方法について述べる。

3.2 円滑なコミュニケーションを支援する遠隔協同作業環境

遠隔協同作業では、自宅やオフィスで集中して行う非同期作業と、同じ時間に作業者がリアルタイムにコミュニケーションをとりながら行う同期作業がある。作業者が作業に集中している状況においては、他の協同作業からの割り込みを抑制し、効率的に作業を進められる環境が求められる。そうした非同期作業の間でも、遠隔にいる協同作業者とコミュニケーションをとることで、解決しやすい問題が発生する。また、作業と作業の合間で、遠隔にいる協同作業者とコミュニケーションをとりやすい時間もある。同じ空間で作業をしている場面では、遠隔協同作業者の状態を確認し、インフォーマルにコミュニケーションをとりながら、問題の解決や、情報交換が図られている。遠隔にいる協同作業者の状況を認識しながら、コミュニケーションを始めるきっかけを作れる協同作業環境が求められる。

さらに、リアルタイムにコミュニケーションを伴う同期作業では、場/空間/周辺のものに関する情報も共有しながら、遠隔作業者と向き合ってコミュニケーションできる環境が求められる。円滑なコミュニケーションを支援する遠隔協同作業環境では、次の2つの機能が求められる。

1. 「非同期作業」から「同期作業」への移行支援機能
「非同期作業」中に互いの存在や状態に共有しつつ、シームレスに「同期作業」へ移行する機能
2. 空間を共有しリアルタイムにコミュニケーション可能な通信機能
場/空間/周辺にあるものの情報を共有し、音声による言語情報に加えノンバーバルな情報も伝達できるコミュニケーションシステム

そこで、本研究では、この2つの機能を実現するために、図 3.1,3.2 の様な遠隔協同作業環境の構築を目指す。

在宅勤務している作業員 a が、集中して資料作成している状態を、システムが自動的に推定し、オフィスにあるテレプレゼンスロボットを、オフィスの中にある作業員 a の座席付近に移動する。この場所で集中して作業していることをテレプレゼンスロボットが表現する（図 3.1）。作業員 a が一時的に休憩を始めると、その状態をシステムが自動的に推定し、職場の共有スペースへテレプレゼンスロボットを移動する。ロボットは作業員 a が休憩している様子を共有スペースで表現する（図 3.2）。共有スペースにいるロボットに作業員 b が声をかけることで、ロボットを介して作業員 a と会話を始めることで、インフォーマルな同期作業を始めることができる。作業員 a は、オフィスのロボットの煩わしい操作を行うことなく、自らの状態を表現することができる。オフィスにいる人とのコミュニケーションが必要なタイミングで、テレプレゼンスロボットを介して始めることができる。

このような遠隔協同作業環境を実現するため、作業員の状態を自動的に推定し共有するプ

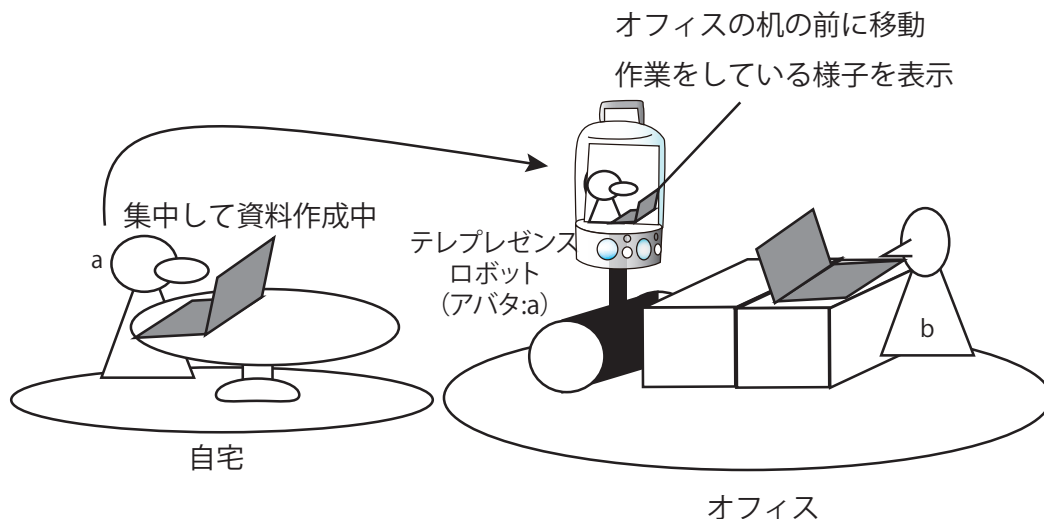


図 3.1 自宅で作業している様子をオフィスで表現している事例

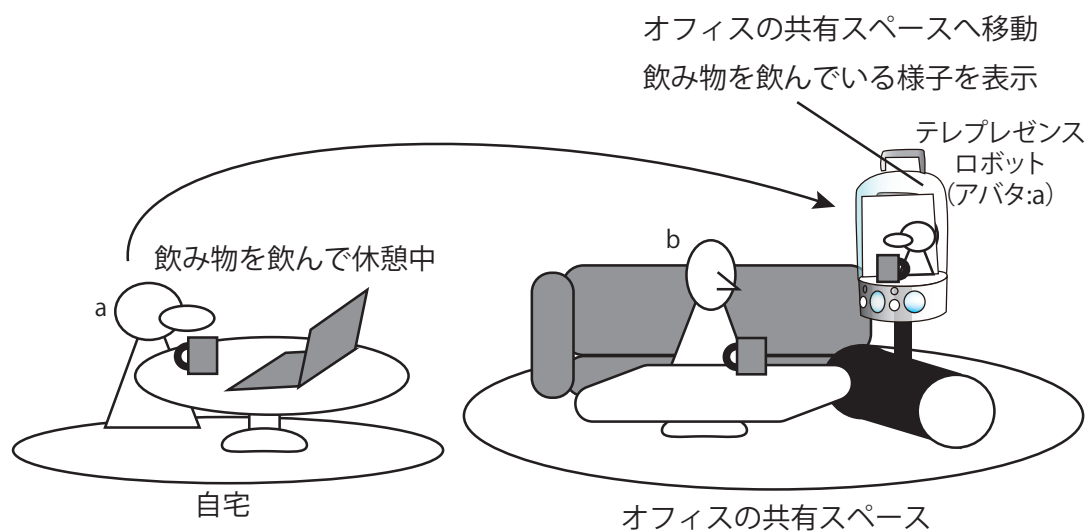


図 3.2 自宅で休憩している状態をオフィスの共有スペースで表現している事例

レゼンス共有システムについて、3.3 節で述べる。また、テレプレゼンスロボットを用いてオフィスの中で対面しているかのようなコミュニケーションを実現するコミュニケーションシステムについて、3.4 節で述べる。

3.3 円滑な遠隔協同作業環境のためのプレゼンス共有システム

プレゼンス共有システムは、(1) 作業を妨げることなく自動的に作業者の状態を推定する機能と、(2) 推定された状態を遠隔にいる協同作業者が自然と認識できるように表現する機能が必要である (図 3.3)。

本研究は、(1) の自動的に作業者の状態を推定する機能の研究に焦点を当てる。(2) の表現機能については、テレプレゼンスロボットは、すでにオフィスの中を移動できるロボットが製造、販売されている。従来研究では、プレゼンスロボットを手動で操作することで、遠隔作業者の状態を表現している研究も行われている [6, 7]。一方で、遠隔協同作業をおこなう作業環境は、それぞれの環境で異なる作業をしており、作業環境の構造も異なる。こうした環境において、作業者の状態を自動的に推定する手法の確立ができていないため、本研究では (1) の機能を実現する技術を提案する。

テレワーク [81] は、在宅勤務、サテライトオフィス勤務、モバイルワークの 3 種類に分

類する。モバイルワークは、移動中や顧客先での作業と定義されており、遠隔にいる作業者とのリアルタイムコミュニケーションが困難な場面も多い。一方、在宅勤務やサテライトオフィス勤務の場合は、室内での作業が中心であり、テレプレゼンスロボットで状態を表現しながら、リアルタイムコミュニケーションへ移行しやすい環境にあると考えた。

そこで、室内で作業を行う作業者の状態推定について検討を行う。在宅勤務やサテライトオフィスの環境は、それぞれの部屋の構造や、家具の配置など、条件が異なる。それぞれの部屋で収集したデータを単純な閾値で分類し、状態推定を行うことは困難である。そこで、それぞれの環境で収集されたセンサーデータを用いて、機械学習を用いて推定を行う。

機械学習は大量の学習データが必要であり、学習データを作成する作業負担の軽減が課題であった。従来研究 [66, 67] では、予め蓄積されている業務記録などを活用し、人の作業負担の軽減している。しかし、サービス開発や技術開発などの作業では、非定型的な作業が多く、作業内容を予め収集した外部知識はない。そこで、本研究は、それぞれの作業環境の作業者とシステムが協調することにより、人手の作業負担を軽減し推定器を構築する手法を考案した。

本研究において、推定する作業者の状態の種類を図 3.4 のように分類した。部屋の中での作業を推定するため、はじめに、作業者の存在の有無を分類する。次に、部屋の中での作業は、同期作業中と、非同期作業中の二つに分類される。同期作業には、会議中、実験中など、遠隔にいる作業者と同期して行っている作業が含まれる。次に、非同期作業は、さらに同期

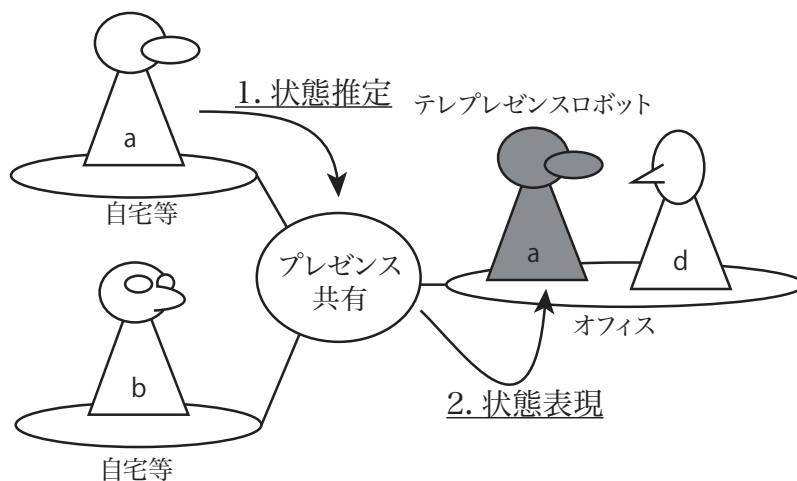


図 3.3 非同期作業状態で協同作業者同士の状態を共有するプレゼンス共有システム

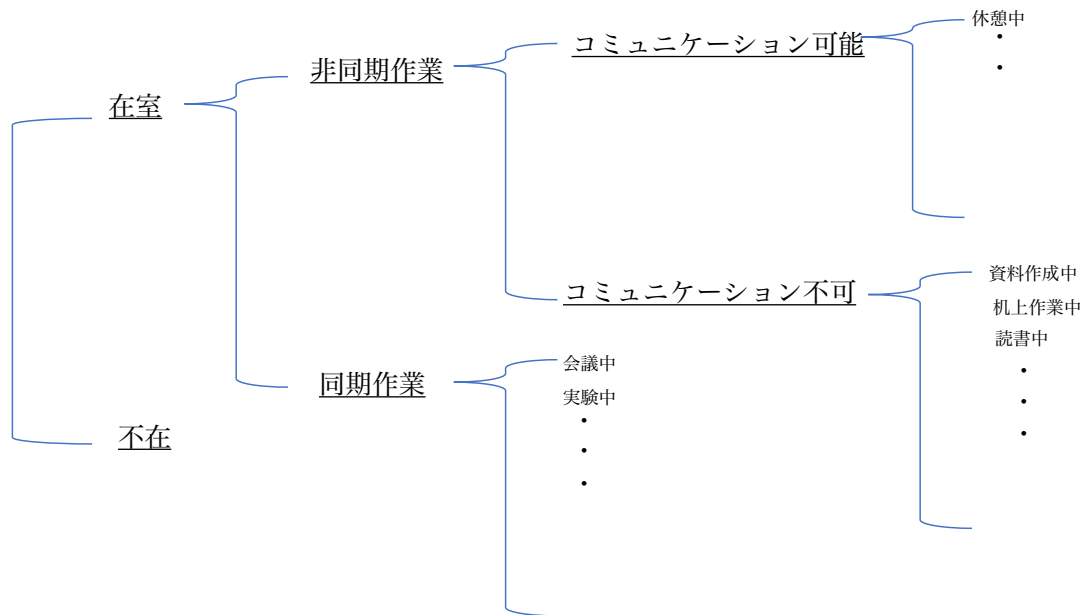


図 3.4 本研究で推定する作業状態の種類

作業中への移行の可否で二つに分類される。遠隔にいる作業者とコミュニケーションが可能な状態として、休憩中などがある。遠隔にいる作業者から、割り込みが許容できない作業として、PCでの資料作成や、机上での書類作成などが含まれる。

作業状態をこうした種類に分けて推定することにより、非同期作業から同期作業へ移行するタイミングを適切に認識することができ、不要な作業の割り込みも抑止できると考えた。

提案手法を用いたプロトタイプシステムの内容と実証結果については、4章で述べる。

3.4 円滑な遠隔協同作業環境のためのリアルタイムコミュニケーションシステム

遠隔協同作業における同期作業を支援するリアルタイムコミュニケーションシステムの構築を目指す。特定のオフィス環境を中心とした遠隔協同作業では、テレプレゼンスロボットを用いたコミュニケーションシステムが活用されている [6, 7]。テレプレゼンスロボットを使うことで、オフィスの空間やオフィス内のモノに関する情報共有が可能となる (図 3.5)。現在、製造・販売されているテレプレゼンスロボット [13, 14, 15, 16] では、遠隔作業者の映像表示に、二次元のディスプレイが利用されており、遠隔にいる作業者の視線情報が意図した通りに表現できない。オフィスにいる作業者と、遠隔にいる作業者の間のコミュニケー

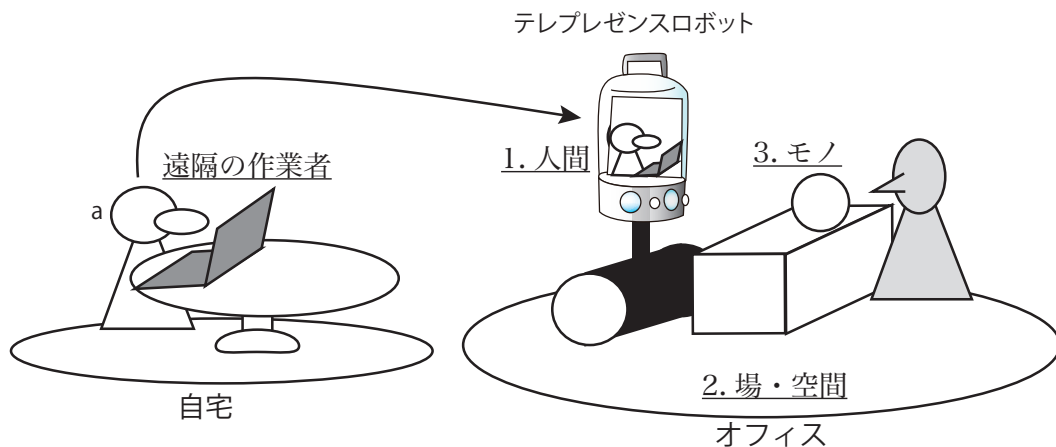


図 3.5 オフィスの空間を共有しながら同期作業が可能なコミュニケーションシステム

ションにおける問題が報告されている [6].

顔の向きや視線の方向を伝達できるディスプレイの研究がされているが、テレプレゼンスロボットは、大型の表示装置 [22, 75] の搭載は困難である。また、多視点の映像表示方式 [77, 76] は、遠隔にいる作業者の映像を複数の方向から撮影が必要であり、撮影装置の大型化が懸念されるため、遠隔協同作業環境には適していない。テレワークの場面では、テレプレゼンスロボットを複数の作業者が共有することもあり [7]、個人個人で専用のスクリーンを使用する手法 [24] を適応させることも困難である。

そこで、本研究は、二枚の映像を積層する DFD 表示方式 [82] を応用し、視線表現手法を行う手法を提案する。映像を二枚重ねる構造のため、奥行きがあるが、テレプレゼンスロボットへ搭載な大きさに小型化が可能である。映像は 3 次元距離カメラと RGB 画像を撮影するカメラを組み合わせることで 1 方向から撮影でき、リアルタイムに伝送できるため、コミュニケーションシステムへの応用が可能である。自宅などで働く作業者の作業を妨げることなく、通常のテレビ会議システムと同様に利用できる表示手法と考えられる。本手法で構築したプロトタイプシステムとその実証結果について、5 章で述べる。

3.5 結言

本章では、遠隔協同作業には、「非同期作業」と「同期作業」がある。非同期作業は集中して行われている作業を阻害することなく、作業者が必要と感じたタイミングで、「同期作

業」へ移行できる機能が必要とされる。同期作業では、あたかも対面対話しているようなコミュニケーション環境の実現が求められる。そこで、協同作業同士が互いに状態を共有しながら、シームレスに同期作業へ移行可能な遠隔協同作業環境を目指す。また、同期作業に移行後は、同じ空間を共有し、対面しているときのように自然と顔の向きや視線が伝達できるコミュニケーションシステムの実現を目指す。

人の状態を共有するためには、自動的に作業者の状態を推定する機能が求められる。機械学習を用いた推定手法は高い推定精度があるが、学習データの作成の負担軽減が課題であった。本研究では、作業者とシステムが協調して推定器を構築することで、人の作業負担を軽減する推定器構築を目指す。

テレプレゼンスロボットを活用することで、オフィス空間を共有するコミュニケーションが可能となるが、ロボットに搭載された二次元ディスプレイでは意図した通りに作業者の視線を表現できず、円滑なコミュニケーションができなかった。そこで、二枚の映像を積層する表示装置を用いて視線表現をする手法の確立を目指す。表示装置は小型化が可能で、ロボットへの搭載が可能である。

第4章

人の作業負担を軽減する状態推定器 の構築手法の研究

4.1 緒言

学習データの作成の負担を軽減するために、人とシステムが協調して推定器を構築する手法を開発した。本章は提案手法について詳しく説明する。4.2 節は、提案手法の目的について述べる。4.3 節は、推定器を構築する手順を4つのステップに分けて説明する。4.4 節は、ラベリング作業の対象となる映像シーンの削減効果の検証について述べる。4.5 節で、その検証結果について考察する。次に、4.6 節は対話インタフェースを活用した推定器の構築に関する実験について述べる。4.7 節で、その実験結果について考察する。

4.2 研究の目的

本章ではシステムと人が協調することで、人の作業負担を軽減し推定器を構築する手法について述べる。提案手法により作業者とシステムが対話し協調して推定器を構築する様子を図 4.1 に示す。本研究は、作業空間に端末を設置するだけで、自動的にデータを収集し、その場所にいる作業者と対話をしながら協調して推定器を構築するシステムの実現を目指す。端末は収集したデータを特徴毎に分類を行い、分類結果に対応した作業状態の映像を作業者に提示する。作業者は提示された映像を確認し、そのときの作業状態を回答する。システムは回答を使って推定器を構築する。

教師あり機械学習を用いた推定器構築方法（従来法 [21, 20]）と、提案手法の比較を図 4.2 に示す。従来法における、システムと人の関係は図 4.2 (a) のようになる。人が大量の学習データを作成し、一方的にシステムへ入力する。このときの推定器の構築ステップを図 4.2 (b) のようになる。データ収集、ラベリング、学習の3つのステップで推定器を構築する。大量にセンサーデータを収集した後、人手で教師情報を付与（ラベリング）し学習データを作成する。システムは、その与えられた学習データを使い機械学習を行って推定器を構築する。

人とシステムが対話をしながら協調して推定器を構築する提案手法の人とシステムの関係は図 4.2 (c) のようになる。提案手法は「データ収集」、「分類」、「選択」、「ラベリング」の4つのステップで構成される（図 4.2 (d)）。システムは自動収集したデータを教師情報が無くてもデータを分類できる教師無しクラスタリング手法で分類する。各クラスタへ作業状態のラベルを付与するために、各クラスタを代表する映像シーンをシステムが自動的に選択し、作業者に提示する。作業者は提示された映像を観察しラベリングすることで推定器が完成する。対話インタフェースを通じて代表シーンを作業者に提示するとともに、そのときの作業



図 4.1 作業者が行っていた作業を端末が確認している様子。端末はあらかじめ分類したデータを使って選択された映像を作業者へ提示する。

状態を質問する。作業者は提示された映像を見つつ、そのときに行っていた作業を回答することにより、推定器は構築される。

システムは自動的に実行される3つのステップ（データ収集、分類、選択）で、ラベリングする対象を減らし、人の作業負担を軽減する。さらに、対話インターフェースを用いることで、ラベリング作業をする際の、システムに専用の学習データ作成用ソフトウェアの使用方法を習得することなく、システムからの質問へ回答するだけでラベリング作業が実施できる環境を構築する。

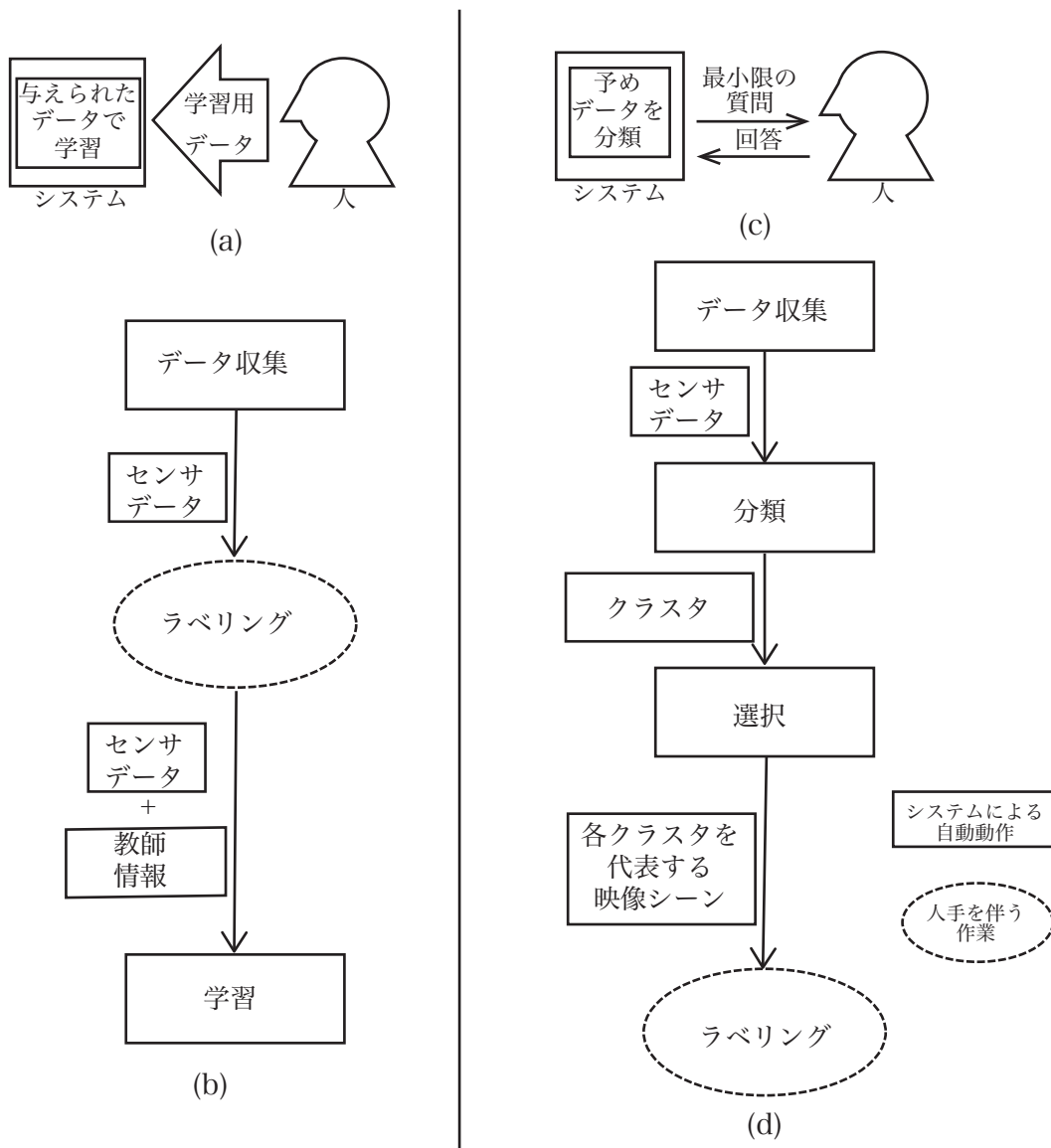


図 4.2 教師有り機械学習を用いた推定器構築方法と提案手法の比較

4.3 推定器を構築する 4 つのステップ

推定器を構築するための 4 つのステップについて詳細を述べる。システムが自動的に行う 3 つのステップは、「データ収集」、「分類」、「選択」である。この 3 つのステップに加え、作業者が手動で行う「ラベリング」のステップを加えて 4 つのステップで推定器を構築する。各ステップの詳細について述べる。

4.3.1 データ収集

システムによるデータ収集に使用したセンサは、マイクロフォンと距離センサの2種類である。データの収集から特徴量の計算までの流れを図4.3に示す。

マイクロフォンは、会議や対話中の音声、キーボードの打鍵音、紙・ペンなどを使っている時の音などを収集することが可能である。作業状態によって、音の周波数や音量が変化するため、一定時間毎にマイクロフォンの音声をフーリエ変換し、特徴量とする。距離センサは、作業者が最も利用する頻度の高い場所である机などに設置して利用する。距離センサの値は、作業者が着席している時と、着席していないときで大きく変化する。また、作業者の姿勢によっても、距離センサの値は変化する。作業状態によって、着席の有無、作業者の姿勢は変化するため、距離センサの値の一定時間毎の平均値を特徴量とする。

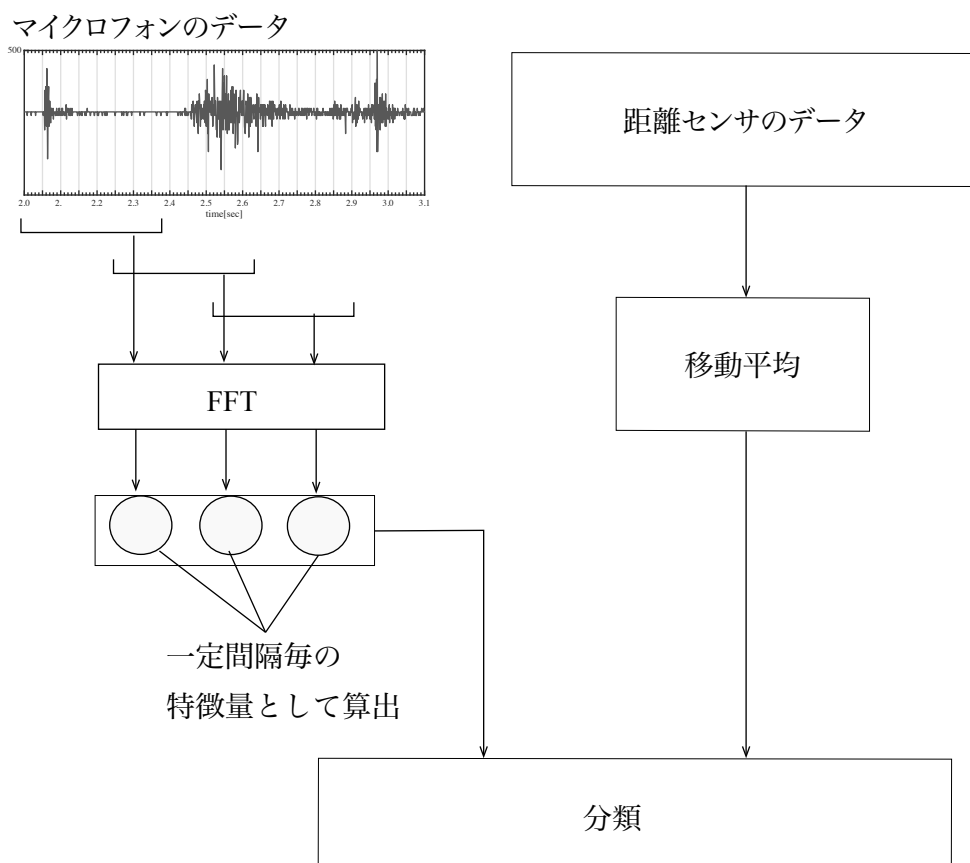


図 4.3 マイクロフォンと距離センサからのデータ収集と特徴量の計算方法

4.3.2 分類

収集したデータをシステムが自動的に分類するために、教師情報を使わないクラスタリング手法を用いる。教師情報を使わないクラスタリング手法は多く提案されている [83, 84, 85]。K-means 法, Gaussian Mixture Model (GMM) 法, Mean-shift 法, Spectral clustering 法の 4 つの方式を候補とし、比較検討を行った。作業環境で収集したデータを使用し分類精度と計算時間を比較した。使用したデータ数は 28,000 で、データ収集時の作業状態は、「離席中」、「PC を使って作業中」、「机上で工作中」、「ビデオ会議中」の 4 種類とした。比較結果を表 4.1 に示す。

クラスタリング手法の計算時間は、最も計算時間が短かった K-means 法を基準として、その比率で表す。分類精度については 4 つの手法の差は小さかった。計算時間については、K-means 法が最も高速で、GMM 法がその 1.3 倍、Mean-shift 法は 21.8 倍、Spectral clustering 法は 18.4 倍となった。本研究は、遠隔協同作業環境へ端末を設置するだけで、推定を行える環境構築を目指している。データ収集、分類、選択の 3 つのステップは、センサと一体化が可能な小型計算機で実行することで、特別な設定や大量のデータの転送無しに推定器の構築が可能となる。分類精度の差は小さいため、最も計算時間が短い K-means 法を採用する。

Lim ら [86] は、K-means 法を用いて画像の輪郭検出のための学習データの作成を行っている。クラスタリングすることにより、特徴の近いデータの集合にまとめてラベルを付与出来るため、ラベル付与の作業回数が軽減されている。しかし、教師情報を用いないクラスタリング手法の結果に、ラベルを付与するためには、複数のセンサデータで構成される各クラスタの作業状態を効率的に確認する手段が必要となる。そこで、本手法は、システムによりクラスタリングされた結果に基づいて、データ収集と同時に録画された作業環境の映像の中

表 4.1 クラスタリング法の分類精度と計算速度の比較

クラスタリング手法	精度 [%]	計算時間
k-means	89.4	1.0
GMM	89.6	1.3
Mean-shift	91.0	21.8
Spectral clustering	91.6	18.4

から、作業状態を確認するための映像シーンを選択する。データの選択手段については、次節で詳しく説明する。

4.3.3 選択

システムが作業員へ提示する動画を選択する方法について述べる。提示する動画は分類された各クラスタに対して、1シーンずつ選択される。元映像はセンサデータの収集と同時に録画された映像である。提示映像を選択する手順を図4.4に示す。分類結果に基づき、センサデータにクラスタ番号を対応させる。センサデータをクラスタ番号毎に時系列に並べた後、一定時間毎にヒストグラムを作る(手順1)。クラスタ番号毎に高い頻度で最も長く継続している時間帯を抽出する(手順2)。抽出された時間帯に対応する映像を切り出す(手順3)。切り出された映像がそのクラスタ番号の代表シーンとして、作業員へ提示される。

4.3.4 ラベリング

作業はシステムによって決定され代表シーンを観察し、そのときの作業状態を決定する。作業員が観察するビデオシーンの数は、システムが分類したクラスタの数と同じである。選択ステップで、代表的な映像シーンに限定されることで、作業員が観察する映像の時間は、すべての映像を順番に観察する場合と比較すると短くなる。作業員のラベリングの負担は軽減される。

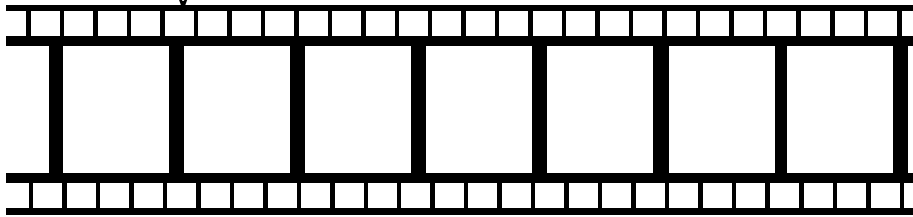
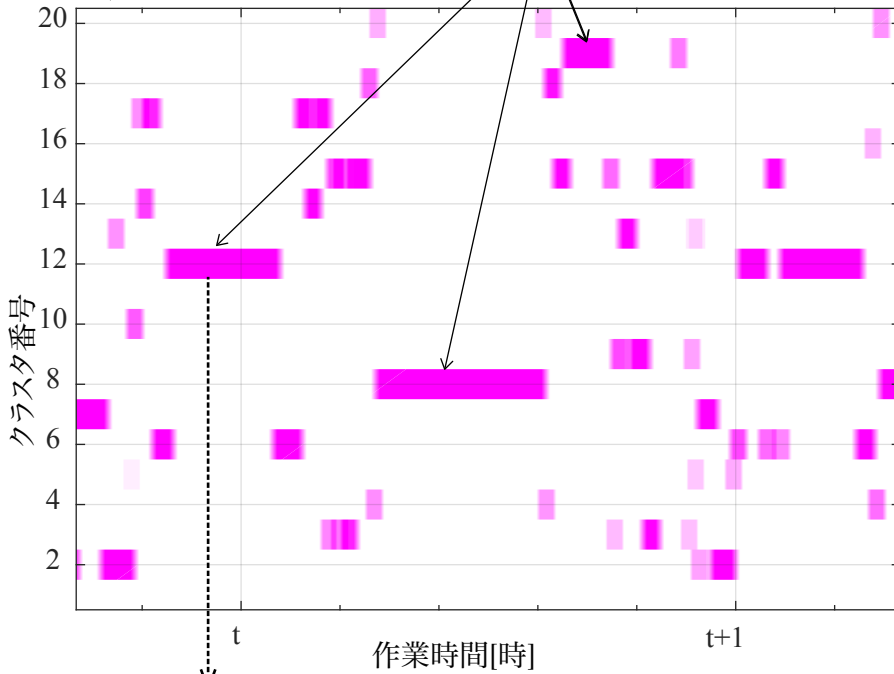
4.3.5 対話インタフェースを活用したラベリング方法

Apple社のスマートフォンに搭載されるSiriやAmazon社のスマートスピーカAmazon Echo, Google社のGoogle Homeに搭載される音声対話によるインタフェースは、人が指示や要求を伝える手軽なインタフェースとして注目されている。チャットシステムとチャットボットの連携が容易になり、複雑なシステムを制御するインタフェースとして、チャットシステムとチャットボットの活用が注目されている[87]。チャットシステムを介して、自然言語でチャットボットと人の対話が可能になる。チャットボットは人から送られてきた指示や要求を解釈し、複雑なシステム制御を行う。チャットボットはあたかもそのシステム制御を専任にしている人の様に振る舞う。システム利用者は複雑なシステム構造や制御に関する知識を保持することなく、自然言語による指示や要望を伝えるだけで、システムを利用することが可能となる。

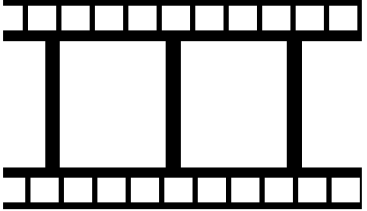
本手法では、複雑な推定器の構築手順をシステムが保持し、システムはデータの収集や分

手順1: ヒストグラムの作成

手順2:
同じクラス番号に分類される
データが継続的に観測される
時間帯を抽出



手順3:
抽出されたデータと同時に
撮影されたシーンを抽出



クラス番号12を
代表する映像シーン

図 4.4 分類結果に基づいた代表シーンの選択方法

類を自動的に行う。作業状態を定義する段階では、人からの情報が必要となるため、システムは作業者と対話する。作業者が特別な知識を習得することなく、システムと人が協調して推定器を構築する環境を実現するために、対話インタフェースは適している。

本手法ではラベリングステップをより効率的に行うため、対話インタフェースを活用した。システムが作業者と対話し作業状態を確認する場面では、システムは、選択した代表シーンを作業者へ提示すると同時に、「この時に何をしていましたか？」の質問を送信する。その質問を受けた作業者は、動画で状態を確認し作業状態を回答する。遠隔にいる共同作業者に対する自身の作業状態を伝えるメッセージとして、システムはこの時の回答を使用する。

従来法の機械学習の場合、すべての学習用データに対応する映像を確認し、教師情報を作成しシステムに入力する必要がある。本手法では、システムが自動的に行う3つのステップによって、作業者が確認する映像の範囲が絞り込まれる。システムによって絞り込まれた映像を、作業者は確認することで推定器は構築される。作業者とシステムが互いに協調することで、人が負担する作業量を軽減する。

4.4 実験1：ラベリング作業の対象となる映像シーンの削減効果の検証

4.4.1 プロトタイプシステム

提案手法の4つのステップで、ラベリングの対象として作業者が閲覧する映像シーンが収集した全映像シーンと比較して削減されるかを検証するために、プロトタイプシステムを開発し、実験を行った。プロトタイプシステムの構成を図4.5に示す。プロトタイプシステムは本手法の性能を評価するために開発した。「データ収集」と「分類」は、プロトタイプ端末(図4.6)で行う。「選択」、「ラベリング」で使用するビデオ映像は、データ収集と同期しつつビデオカメラで撮影した。「選択」、「ラベリング」のステップは、プロトタイプ端末の出力結果と、ビデオカメラの映像をPCに集めておこなった。

プロトタイプ端末は、作業者の机の上に設置する。また、この端末は、マイクロフォンと距離センサの2つのセンサを持ち、小型計算機(Raspberry Pi3 Model B)で同期してデータ収集を行う。マイクロフォンには、Analog Devices社製ADMP441を使用した。距離センサはシャープ社製GP2Y0A710Kを使用した。距離センサはアナログセンサのため、A/Dコンバータを介して計算機に接続した。ビデオカメラには、GoPro社製HERO5 Blackを使用した。プロトタイプ端末とPCは同一のLAN上に接続されている。ビデオカ

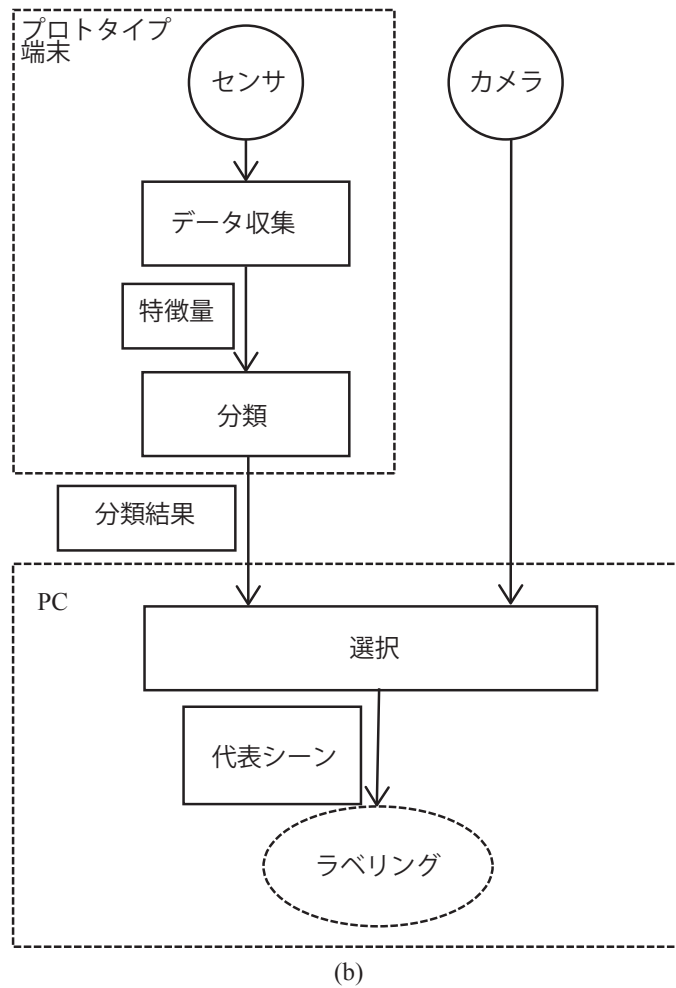
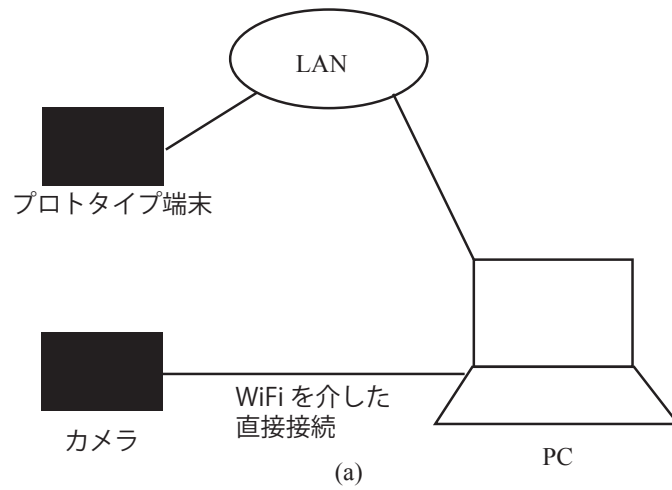


図 4.5 プロトタイプシステムの構成 (a) ハードウェア構成. (b) プロトタイプシステム内のデータの流れ

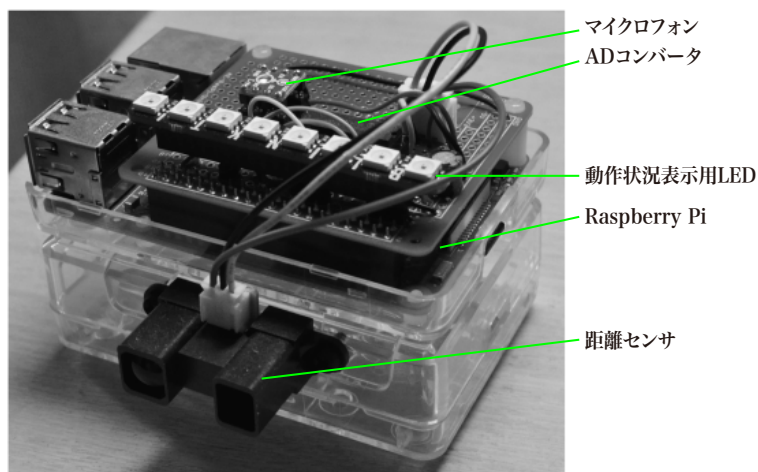


図 4.6 プロトタイプ端末

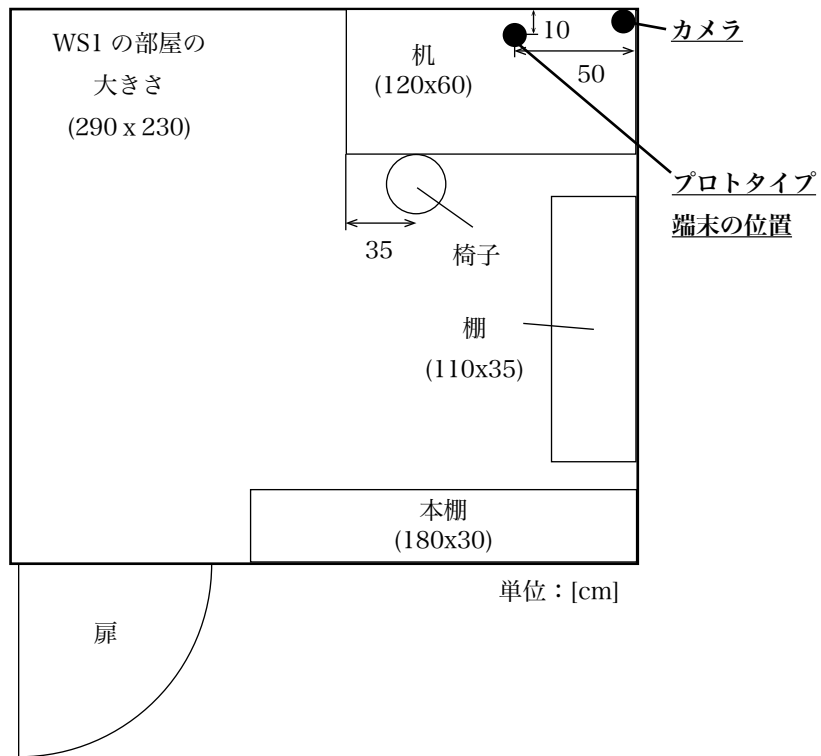
メラの映像は、直接 PC に取り込めるようにした。分類結果は、LAN を経由してプロトタイプ端末から PC へ転送される。選択ステップは、PC 上のプログラムで行い、ビデオ映像から代表シーンを選択した。ラベリングステップは、選択された映像を PC 上で観察して行った。

4.4.2 実験方法

プロトタイプ端末とビデオカメラを2つの部屋に設置し実験を行った。作業環境には、日常の業務でテレワークや遠隔指導に使用している場所を使った。作業環境1と作業環境2の部屋の様子を図4.7,4.8に示す。作業環境1 (WS1) は個人の住宅内の部屋であり、主に一人で使用している。WS1内は週に1日から2日程度、テレワークで作業する場所である。二人以上の作業員で使うことはないため、対面で会議をすることはないが、遠隔にいる共同作業員とPCを使ったテレビ会議や電話で会話することがある。作業環境2 (WS2) は大学内の教員室であり、教員専用の部屋でほぼ毎日この部屋で作業している。PCを使った資料作成などの作業以外に、学生との対話による会議や、PCを使ったテレビ会議、電話で他の人と会話することがある。

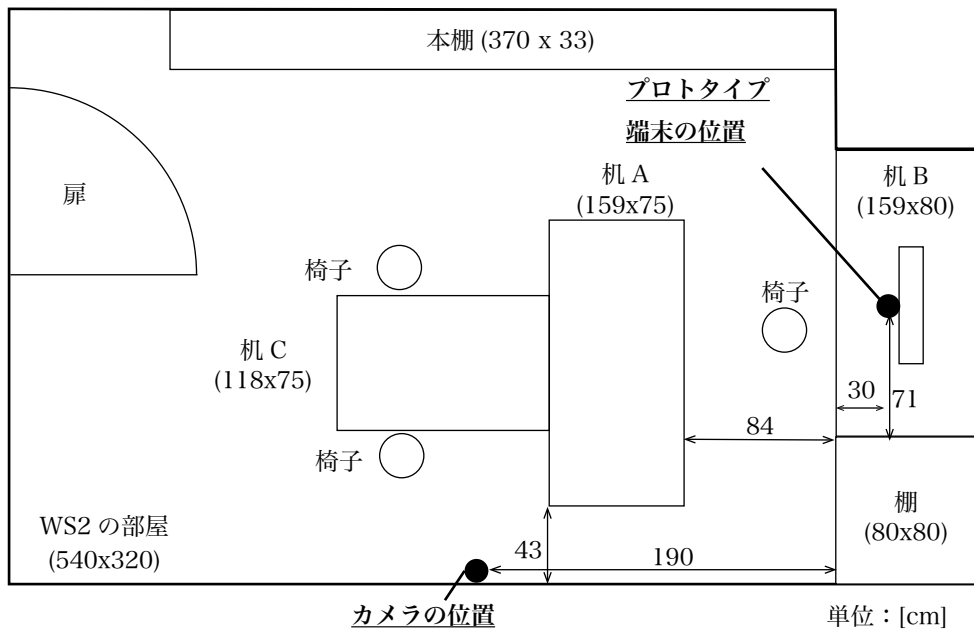
WS1では、2日分のデータを収集し1日分を学習用のデータ、別の1日分を評価用のデータとした。WS2では、3日分のデータを収集し、2日分のデータを学習用とし、1日分のデータを評価用のデータとした。

特徴量を計算するためにシステムに設定した値を表4.2に示す。音声と距離センサは同じ



カメラから撮影した WS1

図 4.7 作業環境 1 のレイアウトと、プロトタイプ端末とカメラの位置。部屋および家具の大きさは（横幅 x 奥行き）で表記。下段の写真は作業環境中のカメラから撮影した作業環境の様子



カメラから撮影した WS2

図 4.8 作業環境 2 のレイアウトと、プロトタイプ端末とカメラの位置、部屋および家具の大きさは（横幅 x 奥行き）で表記 段の写真は作業環境中のカメラから撮影した作業環境の様子

表 4.2 特徴量を計算するための設定値

項目	値
サンプリングレート	16KHz
窓関数	ハン窓
窓関数の幅	1 秒
平均値の計算区間	1 秒

サンプリングレートで収集している。音声データの FFT に用いる窓関数の幅および、距離センサの平均値、標準偏差を計算には 1 秒間 (16,000 個) のデータから計算した。これらの設定値は類似の実験 [88, 89, 90] を参考に決定した。

4.4.3 実験結果

表 4.3 にプロトタイプ端末が収集したデータを示す。システムが選択ステップで選択した代表シーンの合計時間を表 4.4 に示す。システムが選択した代表シーンの映像の長さは WS1 で 26.1% , WS2 では 22.5% である。今回の実験では WS1 のクラスタの数は 20, WS2 のクラスタの数は 30 であり、クラスタ数と同じ数の代表シーンが選択された。

各作業環境における作業者の状態を 4 つに定義し、選択された結果の映像に対してラベリ

表 4.3 各作業環境において収集したデータ量

作業環境	収集時間	総データ数
WS1	12 時間 53 分	44,654
WS2	16 時間 11 分	59,105

表 4.4 選択ステップの結果

作業環境	選択された映像の時間	選択されたデータ数	出力されたビデオ映像のファイル数
WS1	3 時間 21 分	12,097	20
WS2	3 時間 38 分	13,095	30

ングを行った。WS1 および WS2 の作業者の状態を図 4.9 に示す。

WS1 における作業者の状態の定義:

S1-1: 離席中

作業者は席から離れていない、もしくは他の部屋へ移動使用としている。

S1-2: PC を使って作業中

PC を使いながら、ドキュメントの作成や、Web ブラウジングなどを行っている。

S1-3: 机上で工作中

作業者は PC は使わずに作業をしている。今回は、電子回路の作成を行っている。

S1-4: ビデオ会議中

作業者は他の作業者と PC を使ってビデオ会議を行っている。作業者は遠隔にいる作業者と会話をしている。

WS2 における作業者の状態の定義:

S2-1: 離席中

作業者は席から離れていない、もしくは他の部屋へ移動使用としている。

S2-2: 作業者が壁側の PC で作業中

壁側の PC を使いながら、ドキュメントの作成や、Web ブラウジングなどを行っている。

S2-3: 中央の机で作業中

作業者は PC を使わずに、中央の机で書類の読み書きをしている。集中して静かに作業をしている。

S2-4: 会議中

作業者は、部屋の中で会話をしたり、ビデオ会議をして会話をしたりしている。

分類結果とラベリングの結果から推定器を構築し、推定器の正解率を評価した。正解率は、推定器から出力された全推定データの時間に対して、推定結果が正しいデータの時間の割合から算出した。推定器の正解率は WS1 で 93.2%, WS2 で 76.0% であった。各作業空間における推定結果の推移の例として、WS1 の様子を図 4.10 に、WS2 の様子を図 4.11 に示す。従来手法 [21] ではオフィス内の作業者の状態推定を行っており正解率は約 90% と報告されている。本手法も、WS1 において同等の正解率を示している。



S1-1. 離席中



S1-2. PCを使って作業中



S1-3. 机上で工作中



S1-4. ビデオ会議中
(左下のモニタに協同作業者を表示)



S2-1. 離席中



S2-2. 作業者が壁側のPCで作業中



S2-3. 中央の机で作業中



S2-4. 会議中

図 4.9 WS1 および WS2 における作業者の状態

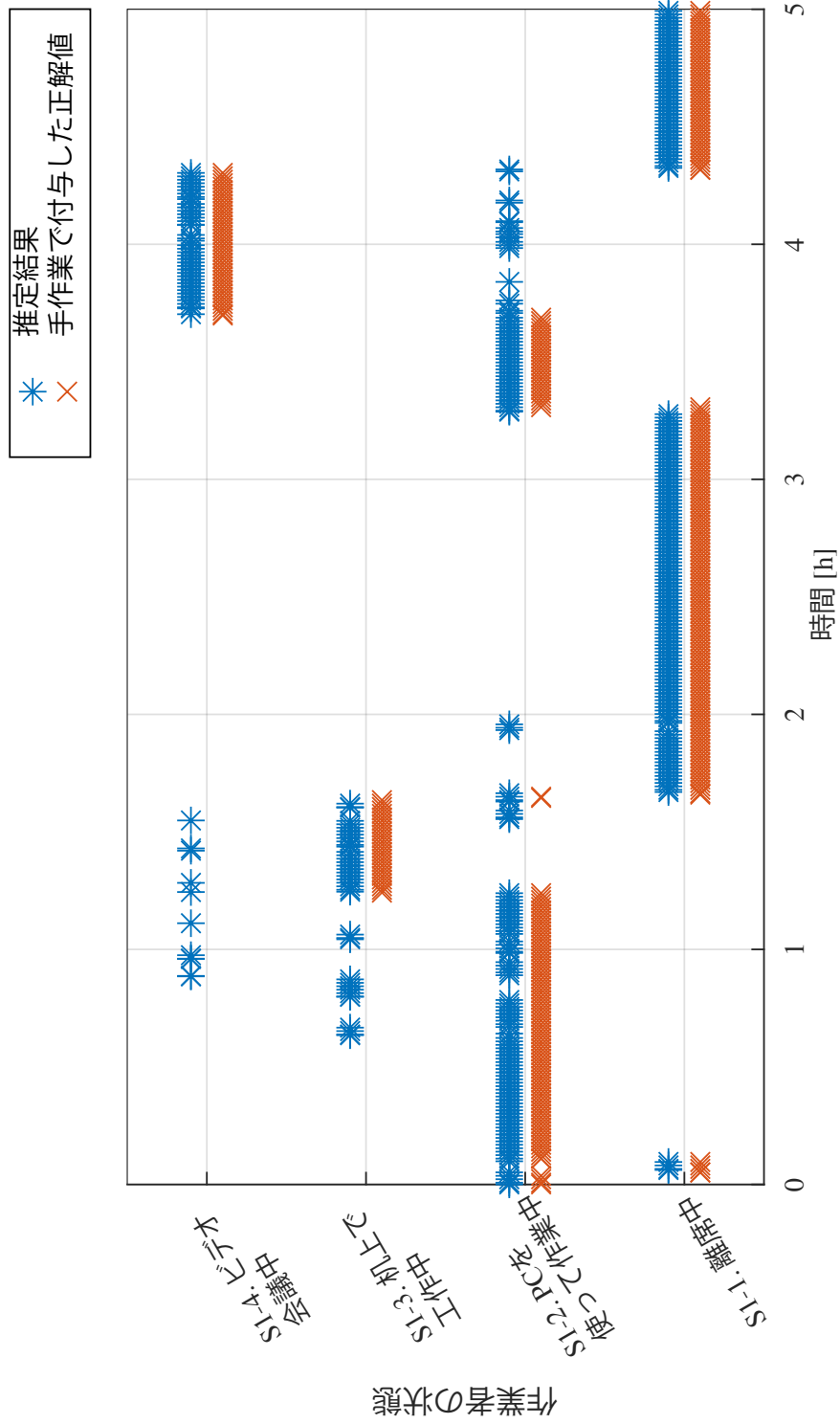


図 4.10 WS1 における推定結果 (5 時間分)

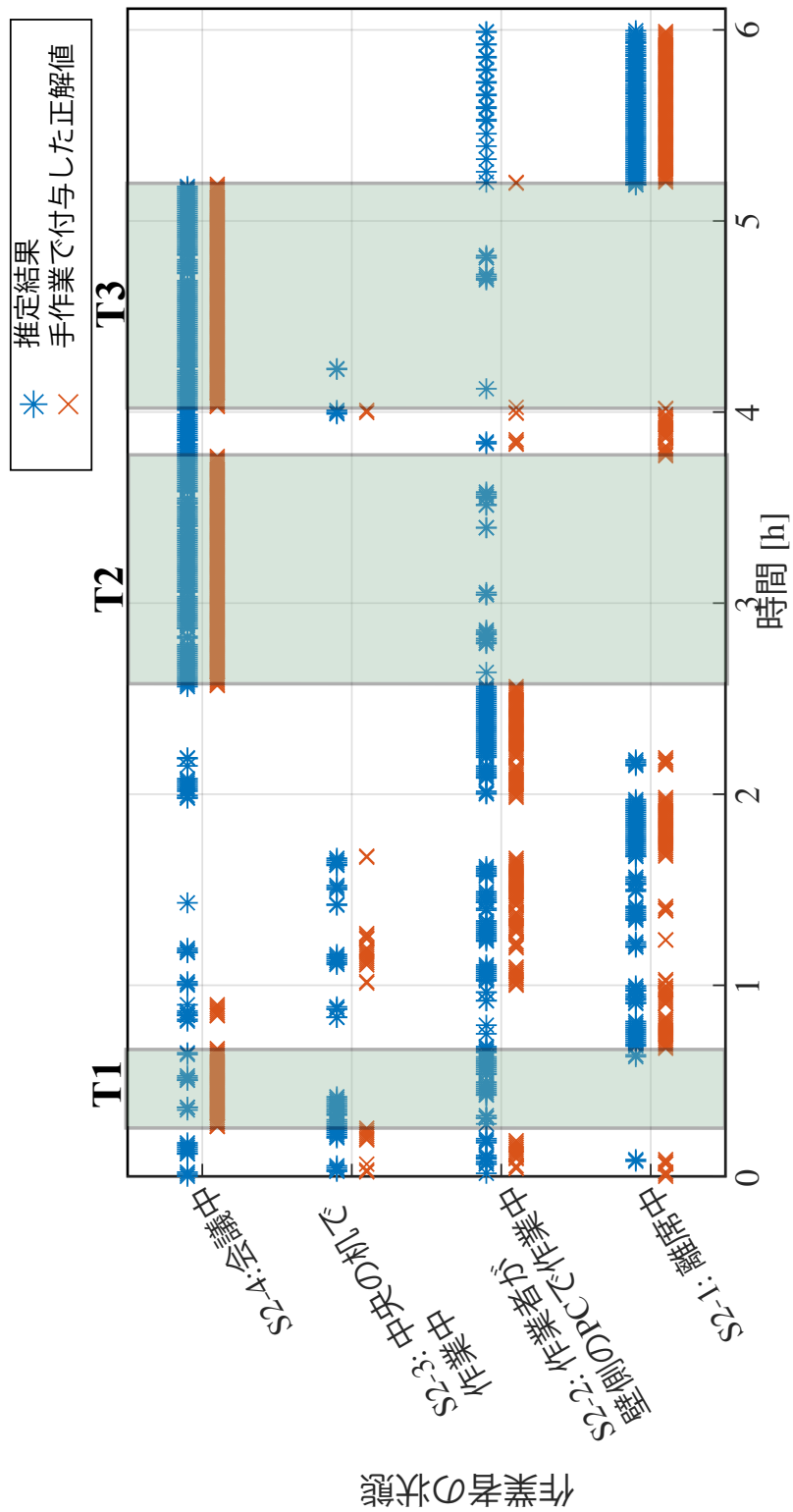


図 4.11 WS2 の推定結果 (6 時間分)

4.5 実験 1 の結果に対する考察

4.5.1 作業量の低減

ラベリングに使用するビデオの映像の長さは、全ビデオの録画時間に対して WS1 で 26.1%、WS2 では 22.5% であり、70 %以上の時間が削減されている。提案手法が作業者の負担軽減に寄与することを確認した。従来法において、学習データを作成する際、学習用に収集したデータすべてに、教師情報を付与する必要があるため、作業状態の確認用映像の時間は、データ取得時間と同じ長さである。映像が短縮されることで、作業者が映像を見る時間は減り、量的な負担の軽減が確認できた。

また、システムが選択した代表シーンは、同じ特徴のセンサデータが連続する時間帯の映像である。その映像中の作業状態は、同じような状態を継続していた。従来法ですべての映像を確認する場合は、動画の最初から順に状態を確認し、状態の変化するタイミングも正確に記録する必要があるが、本手法では、映像中の状態の変化を作業者が注意深く観察する必要はなく、状態を判断することが出来るため、作業の質的な面においても負担を軽減できた。

4.5.2 正解率の低下と特徴量の計算の改良

WS2 の正解率は WS1 と比較すると低い。2つの作業環境を比較し、WS2 の正解率の低下の原因を考える。2つの作業環境の違いは、同室にいる作業者との会話の有無である。図 4.11 の区間 T1 は同じ部屋の中で、作業者同士が会話している区間である。WS2 における T2,T3 の区間は遠隔にいる作業者とビデオ会議をしている区間である。T1 の区間は推定を誤る時間が、T2,T3 と比較すると多い。T1 の区間では、作業者は 2 名いるがドキュメントの読み合わせや本棚から資料を探す動作において、会話がなく、静かな時間が多い。一方、遠隔会議中では互いに会話をしている時間がながく静かな時間はほとんどなかった。また、距離センサーから入力される情報として、机の前の椅子の背もたれの角度によって、あたかもその場所に作業者がいるかのように誤って推定している可能性がある。結果として、S2-4 と推定すべき区間で、S2-2 や S2-3 と推定を誤る時間が増加した物と考えられる。端末の近くにあるものが、作業者なのか、椅子の背もたれなのかを区別するためにセンサーデータから計算する特徴量の改良が必要である。

4.6 実験 2：対話インタフェースを活用した推定器構築手法の検証

4.6.1 プロトタイプシステムの構成

対話インタフェースを介してシステムと作業者が協調して推定器を構築する効果を検証するために、プロトタイプシステムを構築した。プロトタイプシステムの構成を図 4.12, 4.13 に示す。プロトタイプシステムを構成するメイン端末およびカメラ端末を図 4.14 に示す。自動的に行う 3 つのステップ（データ収集、分類、選択）を担うメイン端末と、データ収集と同時に作業状態を録画するカメラ端末がある。次に、チャットシステムに作業者がアクセスするための端末として PC がある。メイン端末、カメラ端末、PC はそれぞれの作業環境に設置される。

メイン端末は、実験 1 のプロタイプ端末とハードウェアの構成は同様である。市販のビデオカメラの代わりに、映像を撮影するためのカメラ端末も作成した。カメラ端末は Raspberry Pi3 Model B へ専用のカメラモジュールを接続して作成した。メイン端末のデータ収集と、カメラ端末の録画は同期して行う。カメラ端末が録画した映像は録画終了後にメイン端末に転送する。メイン端末はデータ収集、分類、選択のステップ後、チャットシステムを使って、映像の提示と質問をその作業環境の作業者へ提示し回答を得る。

メイン端末は机の上に設置し、カメラ端末は、作業者が作業内容を確認しやすい映像を撮れる場所に設置する。作業環境全体を 1 つのカメラで撮影するため、広角のカメラモジュールを使用した。メイン端末がデータを収集する時に、カメラ端末へ信号を出すことで、センサデータと映像は同期して記録される。

プロトタイプシステムでは、チャットシステムに Slack[59]、チャットボット用のプラットフォームとして Hubot[91] を使用した。Hubot はチャットボットを構築するためのプラットフォームであり、それぞれの環境のメイン端末内で実行している。メイン端末は Hubot に実装した対話用のルールを使ってチャットボットとして動作する。クラスタ毎に切り出された動画の提示と同時に「この時に何をしていましたか?」という質問を作業者へ送り、作業者からの回答を受信すると該当のクラスタの作業状態として設定する。

また、実験 1 において正解率が低下した問題を改善するため、距離センサーの特徴量の計算方法を改良した。距離センサーの近傍に作業者がいるのか、椅子等の物体なのかを区別するために特徴量を追加した。データの収集から特徴量の計算までの流れを図 4.15 に示す。距離センサの値は、作業者が着席している時と、着席していないときで大きく変化する。し

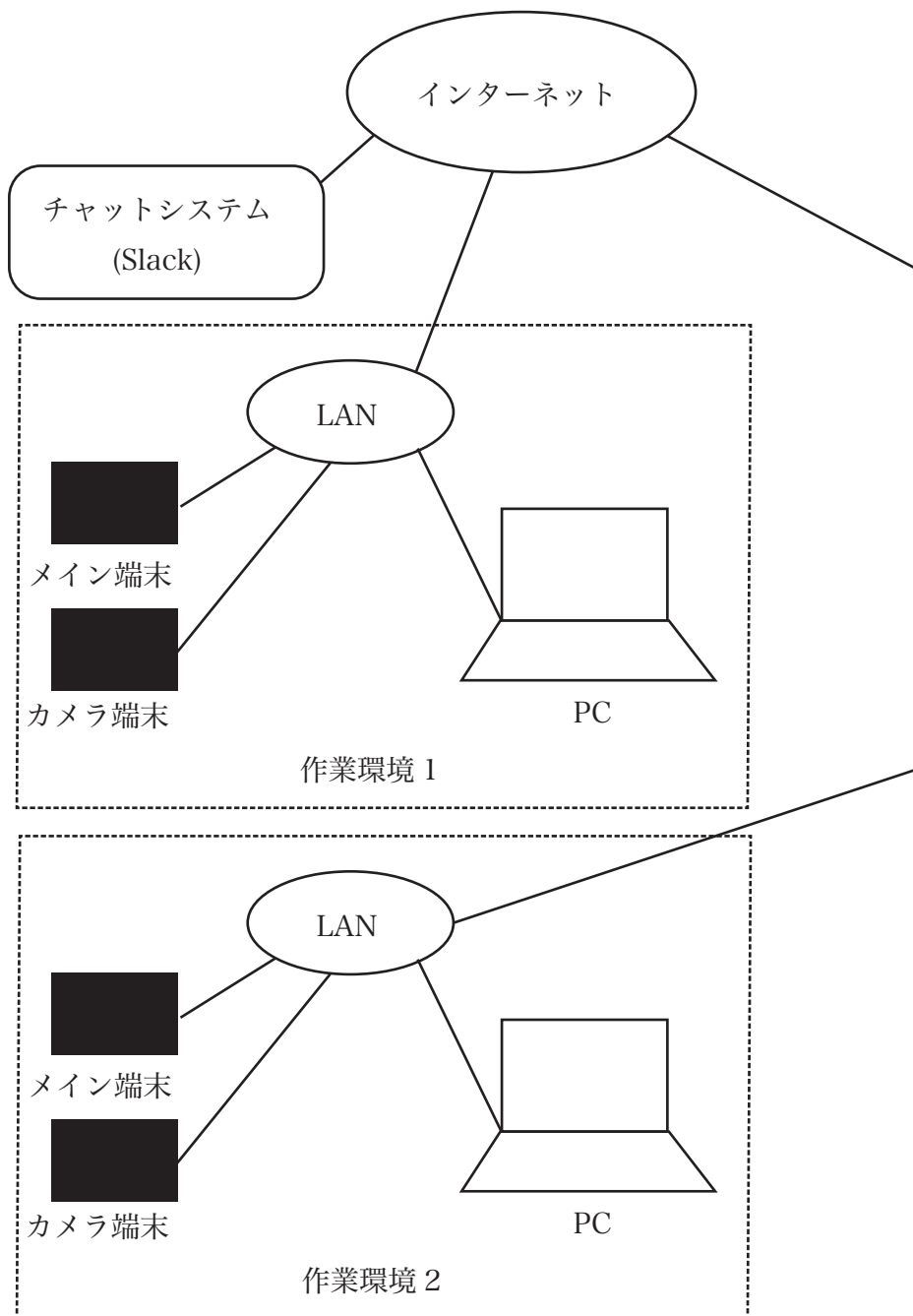


図 4.12 作業者の状態推定とプレゼンス共有が可能なプロトタイプシステムの構成

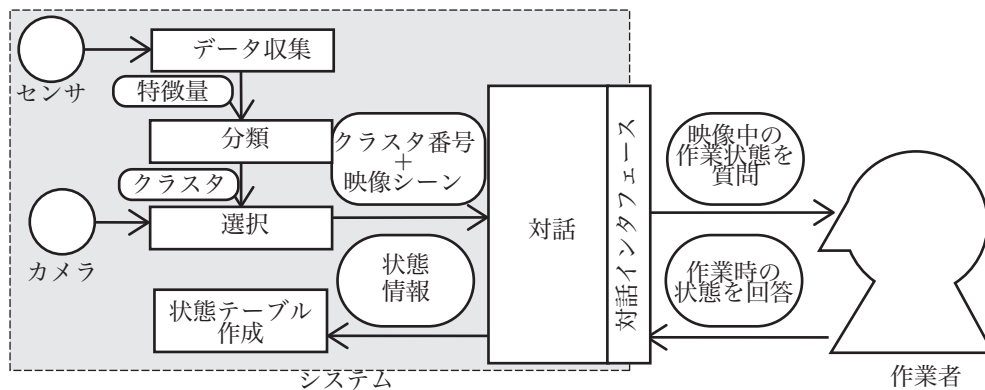


図 4.13 プロトタイプシステムの構成

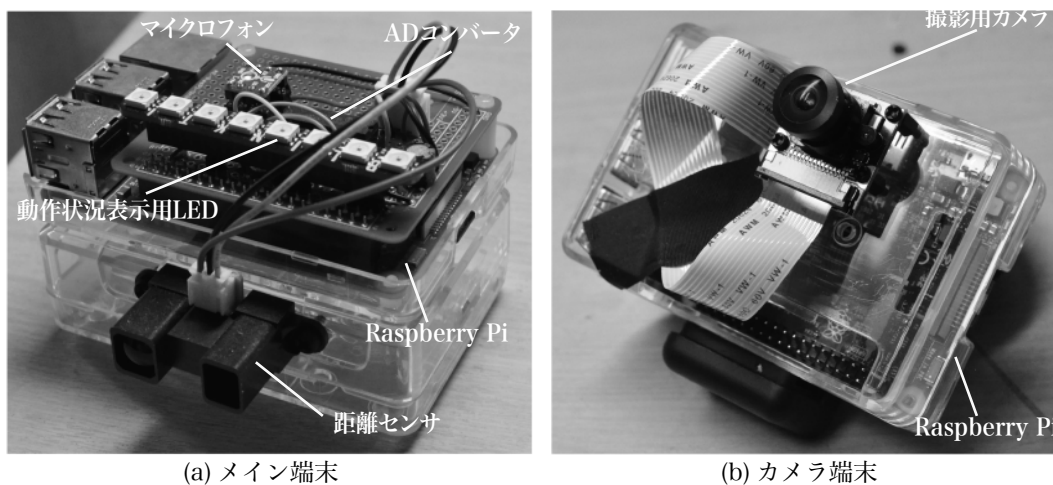


図 4.14 プロトタイプシステムで使用されるメイン端末とカメラ端末

かし、元々の作業者がいた距離と近い位置に椅子の背もたれなどがあると、作業者の有無の区別が困難である。作業者がいる場合、身体は完全に静止することではなく、距離センサの値は細かく変化する。一方で、椅子などの物体の場合には、距離センサーの値は一定である。そこで、距離センサーのデータの一定時間毎の標準偏差を特徴量に加えることで、作業者の有無を区別しやすくなる。距離センサの値から一定時間毎の平均値と標準偏差を算出して特徴量とした。

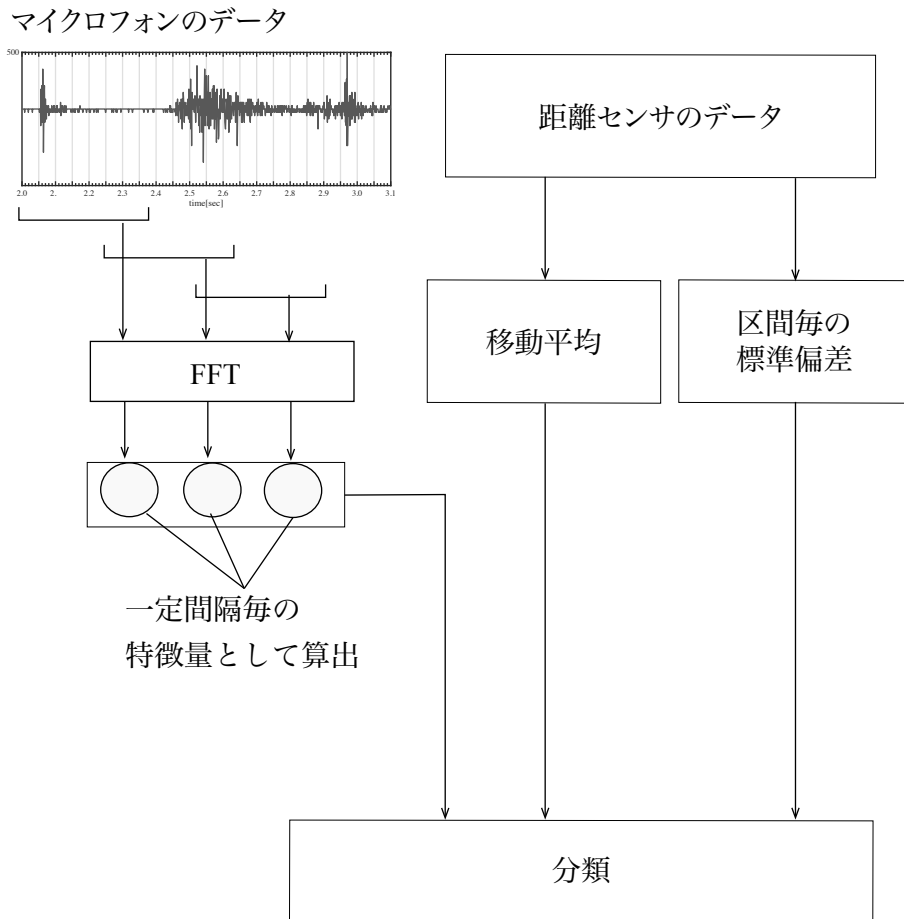
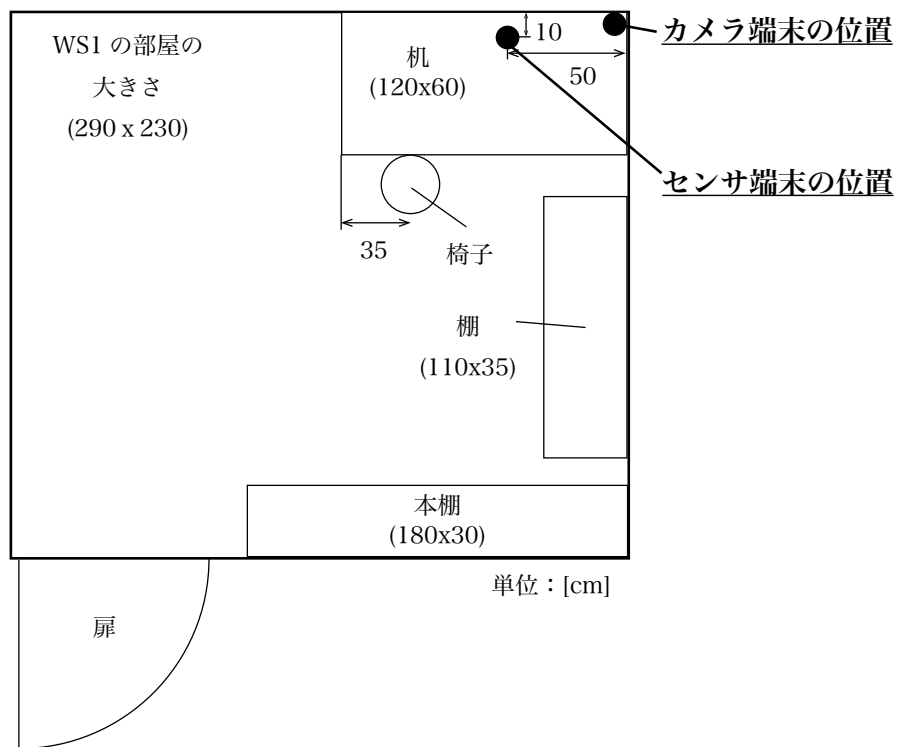


図 4.15 マイクロフォンと距離センサからのデータ収集と特徴量の計算方法

4.6.2 実験方法

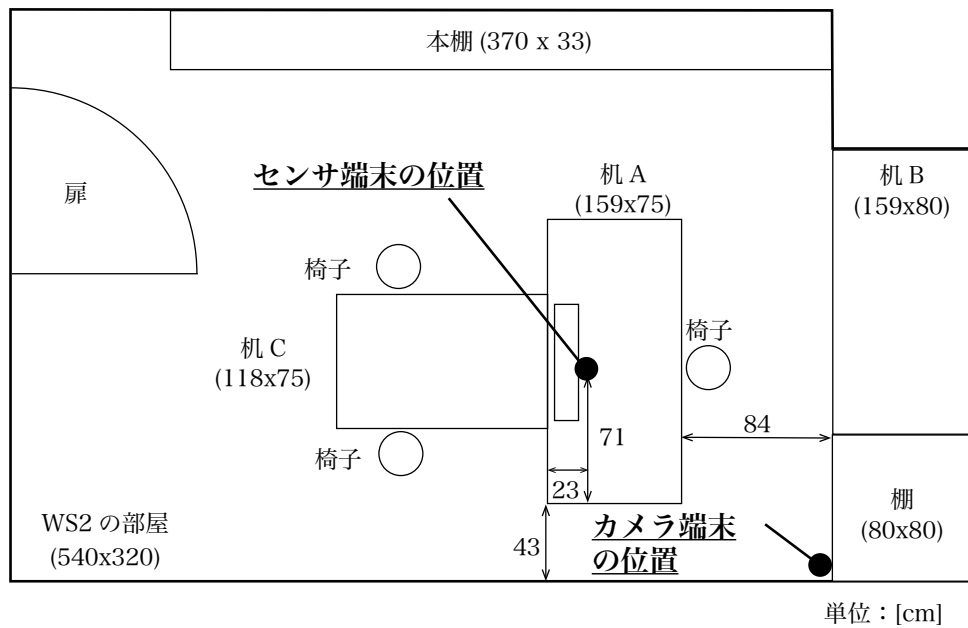
プロトタイプシステムによる性能評価を行うため、メイン端末とカメラ端末を2つの作業環境に設置し再度実験を行った。実験に利用した作業環境のレイアウトや、端末の設置位置を図 4.16,4.17 に示す。

特徴量を計算するためにシステムに設定した値を表 4.5 に示す。距離データの標準偏差を計算する区間も平均値と同じ1秒とした。



カメラ端末から撮影した WS1

図 4.16 作業環境 1 のレイアウトと、センサ端末とカメラ端末の位置。部屋および家具の大きさは (横幅 x 奥行き) で表記, 下段の写真はカメラ端末から撮影した作業環境の様子



カメラ端末から撮影した WS2

図 4.17 作業環境 2 のレイアウトと、センサ端末とカメラ端末の位置、部屋および家具の大きさは（横幅 x 奥行き）で表記、下段の写真はカメラ端末から撮影した作業環境の様子

表 4.5 特徴量を計算するための設定値

項目	値
サンプリングレート	16KHz
窓関数	ハン窓
窓関数の幅	1 秒
平均値および標準偏差の計算区間	1 秒

4.6.3 実験結果

実験結果を表 4.6 に示す。録画された映像の総時間と、システムが選択した代表シーンの合計時間も併せて記載する。WS1 は 8 時間分のデータから推定器を作成し、WS2 は約 38 時間のデータから推定器を作成した。WS1 で作業が行われる日には、PC を使った作業、机上での書類作成、テレビ会議が頻繁に行われる。作業が行われる日には、毎回これらの作業が行われるため、1 日 8 時間のデータを収集し推定器を構築した。WS2 においては、作業者は部屋の中で対面会議や、テレビ会議を行うことがあるが、これらの作業を毎日実施しない。頻度の高いすべての作業状態に対応したデータを収集するために、WS1 よりも長くデータを収集し、推定器を構築した。

推定器の構築に用いた特徴量のデータ数は、WS1 が 28,783、WS2 が 136,286 である。K-means 法のクラスタ数は 50 とした。正解率を評価するため、WS1 において 16 時間分、WS2 において 48 時間分のデータを収集した。これらのデータは推定器構築用のデータとは別に収集したデータである。

作業者がチャットシステムを使用しチャットボットと作業者が対話を行う様子を図 4.18 に示す。チャットボットは、作業者へ動画を提示すると同時に、作業状態を確認する質問

表 4.6 実験結果

部屋	総録画時間	選択された代表シーンの合計時間	正解率 [%]
WS1	8h 00min	3h 8min	90.5
WS2	37h 52min	6h 19min	93.5

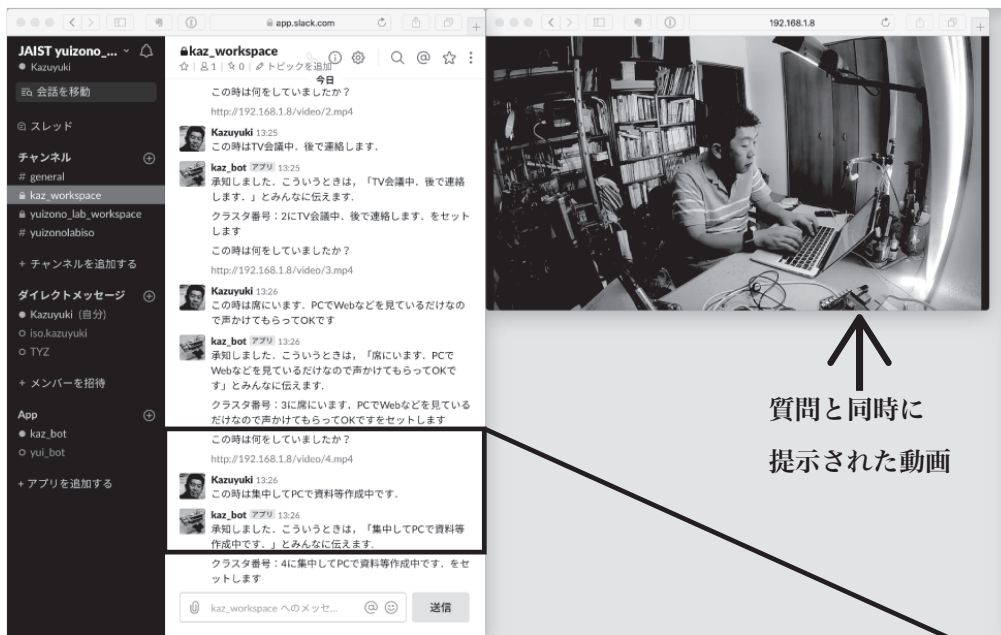
を送っている。作業者は動画で作業状態を確認した後、返答している。

WS1 と WS2 それぞれの作業環境において、システムから作業員への質問の数は 50 である。システムからの質問は分類したクラスタの数と同じとなる。50 個の作業員の回答は各クラスタに割り当てられることにより、推定器が構築される。構築された推定器の正解率は表 4.6 の通りである。

作業員からの回答を確認した結果、類似の回答が複数あったため、類似の回答を集約した。WS1 の回答は 7 種類、WS2 の回答は 9 種類であった。WS1,WS2 における回答を次に示す。

- 作業環境 WS1 における回答:
 - 集中して PC で資料等作成中
 - 席で集中して作業中。後で連絡します。
 - 席にいます。PC で Web などを見ているだけなので声かけてもらって OK です
 - 席にいます。特別な作業はしていないので声かけてもらって OK です
 - TV 会議中。後で連絡します
 - 席にはいないけど近くにいるので連絡可能
 - 部屋に不在です
- 作業環境 WS2 における回答:
 - PC で集中して資料作成中
 - 席にいて書類作成中
 - 部屋の中で打ち合わせ中
 - TV 会議中
 - 席にはいます
 - PC で映像閲覧中
 - 部屋の中で片付けなどをしています
 - 部屋の中を動きながら作業中
 - 外出中

各作業環境において、正しく推定された時の画像を図 4.19,4.20 に示す。図中の画像は、カメラ端末で撮影された画像から、作業状態を確認出来る箇所を切り抜いたものである。推定器構築時と評価時で作業員の位置や、姿勢の状態が類似していることが確認出来る。また、同じ机に着席していても、姿勢の違いや、音声の有無などから、作業状態が分類されていることを確認した。



質問と同時に
提示された動画

この時は何をしていましたか？
<http://192.168.1.8/video/4.mp4>

Kazuyuki 13:26
この時は集中してPCで資料等作成中です。

kaz_bot アプリ 13:26
承知しました。こういうときは、「集中してPCで資料等作成中です。」とみんなに伝えます。
 クラスタ番号：4に集中してPCで資料等作成中です。をセットします

作業者と端末の対話の様子

図 4.18 チャットを介して作業者とシステムが対話を行う様子。システムから作業者へ映像と質問を送り、作業者は映像を確認した後返答する

	推定器作成時	推定時
集中して PC で 資料等作成中		
席にいます。PC で Web などを見ている だけなので声かけて もらって OK です		
席にいます。特別な 作業はしていないの で何かあれば声をか けて OK です		
TV 会議中		

図 4.19 作業環境 1 において推定が正しかったときの例

	推定器作成時	推定時
PCで集中して資料作成中		
席にはいます。		
部屋の中で打ち合わせ中		
TV会議中		

図 4.20 作業環境 2 において推定が正しかったときの例

4.7 実験 2 の結果に対する考察

4.7.1 提案手法による作業の削減効果

実験の結果、本手法によって作業者が確認すべき映像は、WS1 で約 60%、WS2 で約 80% 短縮された。実験 1 と同様に提案手法が作業者の負担軽減に寄与することを確認した。Avrahami らの研究 [21] では、オフィスでデスクワークをしている作業者の 6 つの状態を、85%~95% の正解率で推定している。今回の実験では、WS1 で 7 種類、WS2 で 9 種類の状態を同等の正解率で推定した。

対話インタフェースとしてチャットシステムを使うことにより、作業者とシステムは自然言語で作業状態の確認を行っており、機械学習のための学習データ作成という特別な手順の習得はない。あたかも遠隔にいる共同作業者と対話をするかのように、チャットボットからの質問に回答するだけで推定器を構築出来るため、システムの利用に関して習熟する負担も小さくなった。

メイン端末は、推定器の構築に用いるすべてのセンサーデータ収集した後に、分類、選択の処理を行うため、作業者への代表シーンの提示は、これらの処理の後である。代表シーンを撮影した後、作業者へ提示するまでの時間がかかると、そのとき行っていて作業状態を作業者が忘れてしまうことがある。作業状態を想起しやすい映像を撮影するとともに、データ収集から時間を経過せずに、代表シーンを作業者へ提示する必要がある。センサーデータの収集後、代表シーンの提示までの時間の影響については、検証を行う予定である。

4.7.2 作業環境に適応した状態の定義

評価実験において、システムによって選択された 50 個の代表シーンの中で、WS1 は 47、WS2 も 47 シーンについて、同一の状態が継続されている様子が観察された。これらのシーンについては、作業状態に応じたデータの分類と代表シーンの選択が行われており、作業者に対して回答しやすい映像を提示することが出来た。

一方で、それぞれの作業環境の残り 3 シーンについて考察する。WS1 における 3 シーンは、作業者が部屋の中を移動するシーンであった。「不在状態から作業者が入ってきて着席」が 1 シーン、「不在状態から作業者が入ってきて机に物を置きました部屋を出る」が 2 シーンあった。作業者が WS1 の付近にいるため「席にはいないけど近くにいるので連絡可能」と定義された。また、WS2 における 3 つのシーンについては、「作業者が部屋に入った後、机 B の付近で鞆を整理し部屋から出る」、「机 B 付近で鞆を整理した後、部屋から出る」、「本棚

の付近を移動しながら本棚から資料を取り出した後、部屋から出る」の3つのシーンであった。これらの状態については、「部屋の中で片付けなどを行っています」もしくは、「部屋の中を動きながら作業中」と定義された。これらのシーンにおいて作業者は、作業を一時的に中断し次の作業へ移行する状態である。システムが作業員へ送られた質問文「この時に何をしていましたか？」に対し、次の作業へ移行するための移動や準備をしている状態を、1つの作業状態として定義することは難しかった。しかし、これらのシーンは集中した作業は行われておらず、他者との会話もない状態である。遠隔地にいる共同作業員から連絡可能な状態として、定義することは可能である。今後は、システムから作業状態を確認する時の質問文を変更し、これらのシーンの状態の定義も容易に行えるように改良する。

次に WS1, WS2 の回答を比較すると、それぞれの環境に応じた作業状態が定義されている。「PC で集中して資料作成中」や「TV 会議中」は、両方の作業環境で共通の定義であるが、「部屋の中で打ち合わせ中」は、WS2 の教員室のみで定義された。作業環境毎に作業の種類は変わるが、本手法はそれぞれの作業環境に応じた状態を定義可能で、それらの状態を高い正解率で推定出来る。テレワークを行うそれぞれの環境に適応した状態推定器を構築する手法として有用な特徴である。

さらに、WS1 の「席にいます。PC で Web などを見ているだけなので声かけてもらって OK です」や「席にいます。特別な作業はしていないので声かけてもらって OK です」は、遠隔にいる作業員に向けて、電話やチャット等による割り込みを許容する状態であることを伝えている。自身の作業スタイルに応じて、共同作業員とコミュニケーションする機会を作ることや、集中して作業をしている時に割り込みを抑止するメッセージを発信することが可能であり、遠隔の協同作業を円滑に進めるために重要である。今後、推定された情報を遠隔協同作業の場で活用し、コミュニケーション機会の創出と、作業中断の抑止効果について評価する予定である。

4.8 結言

本章は、学習データの作成の負担を軽減するために、教師無しクラスタリング法や対話インタフェースを活用して人の状態を推定する推定器の構築手法を提案した。システムが自動的に行う3つのステップは、「データ収集」、「分類」、「選択」であり、この3つのステップに加え、作業員が手動で行う「ラベリング」のステップを加えて4つのステップで推定器を構築する。提案手法を評価するためにプロトタイプシステムを構築し、2つの作業環境に端末を設置し、2回の実験を行った。推定器の構築のために、作業員が閲覧する必要のある映像の時間は、収集した全映像と比較すると、約60%~80%少ない時間を閲覧するだけで、推

定器を構築できることが確認された。システムによって同様の特徴の映像が選択されることで、作業状態の確認も容易になり、質的な面においても負担を軽減している。推定器の正解率は、90%以上となった。提案手法により、人の作業負担を軽減しながら、従来法と同等の性能を実現することを確認した。

また、対話インターフェースを用いてラベリングする手法とすることで、システムの利用に関して習熟する負担の軽減にもつながった。また、作業者が対話をしながら、それぞれの作業環境毎に適した状態を定義することができ、コミュニケーションの可否をメッセージとして伝えることも可能となった。

今後は、評価実験を行うテレワーク環境を増やししながら、分類に必要なクラス数の調査を行い、さらなる負担軽減に向けたシステムの改良を行う予定である。また、遠隔協同作業中にプレゼンス情報の共有を行う実験を実施し、作業時のコミュニケーション機会の創出、作業中断の抑止効果について検証を行う。

第5章

二画面積層表示を用いた視線表現手法の研究

5.1 緒言

本章は作業者の状態を推定するための推定器をその空間で作用をしている作業者とシステムが協調して構築する手法について説明する。5.2 節は、研究の目的を、5.3 節は、二画面積層表示を使い視線の向きを表現する手法の詳細を述べる。5.4 節は、本手法の性能を評価するための基礎的な実験について述べ、5.5 節で、基礎実験に関する考察について述べる。次に、5.6 節は本手法を用いた対話用のプロトタイプシステムを開発したので、その詳細について説明し、そのプロトタイプシステムを使用した実験について述べる。5.7 節で実験結果に関して考察する。

5.2 研究の目的

遠隔協同作業において、あたかも同じ場所にいるかのように、顔の向きや視線の情報が伝達できるコミュニケーション環境（図 5.1）を目指している。この会議中にいる参加者であれば、遠隔からの参加者の顔がどちらを向いていて、誰と話をしているのかを、会話をしている二者だけでなく、その他の参加者からも認識できる。本章では、このような映像の表現を実現するための映像表示技法について述べる。視線や顔の向きの伝達のためには、次の 3 つの要素を実現する必要があると考えている。例えば、N:1 型の遠隔協同作業の場合、遠隔作業者の代理となる端末の前には複数の協同作業者がいることが想定される。遠隔の作業者が見ている作業者と視線が一致することに加え、他の参加者とは視線は一致しないようにしなければならない（要件 1）。また、遠隔作業者の視線の方向を、第三者の視点から見て認識できる機能も必要となる（要件 2）。遠隔の作業者が、複数の人の広い拠点の中で様々な方向を見るときに、顔の向きや視線の方向を外部から制御できる特徴は、実際の遠隔作業環境で利用するときに必要な（要件 3）

1. 指向性のある映像表示：遠隔から参加している参加者と、特定の方向にいる参加者のみ目があると感じられること。
2. 第三者的な視点における指向性の認識：遠隔から参加している参加者の視線や、顔の向きは、会議室にいるすべての人間が認識できること。
3. 方向の制御：遠隔から参加している参加者の視線、顔の向きは制御可能であること。

この 3 つの要素を実現するために、二画面積層表示を用いることを提案する。



図 5.1 将来的なシステムイメージ

5.3 二画面積層表示による視線方向の表示手法

二枚の映像を使って立体的な映像を表示する手法として、Suyama ら [82] は DFD 表示方式を提案している。この表示方式は、裸眼で立体的な映像を表示可能で、見る角度によって映像の見え方に変化を生じる。基本的な原理を図 5.2 に示す。

2枚のディスプレイの映像を一定の間隔を空け積層して表示する。表示する対象の奥行きに比例して輝度を配分して前面と、後面の映像を表示する。例えば、手前にある物体は前面に輝度を高く表示する。奥にある物体は後面に輝度を高く表示する。前面と後面の間にある物体は、その奥行きに応じて輝度配分して表示する。このように、奥行きに応じて輝度配分を変えることで、映像を見ている人は物体の奥行きを感じる。

顔画像をこのディスプレイで表示する場合、顔の凹凸に比例して前後二枚の映像の輝度を

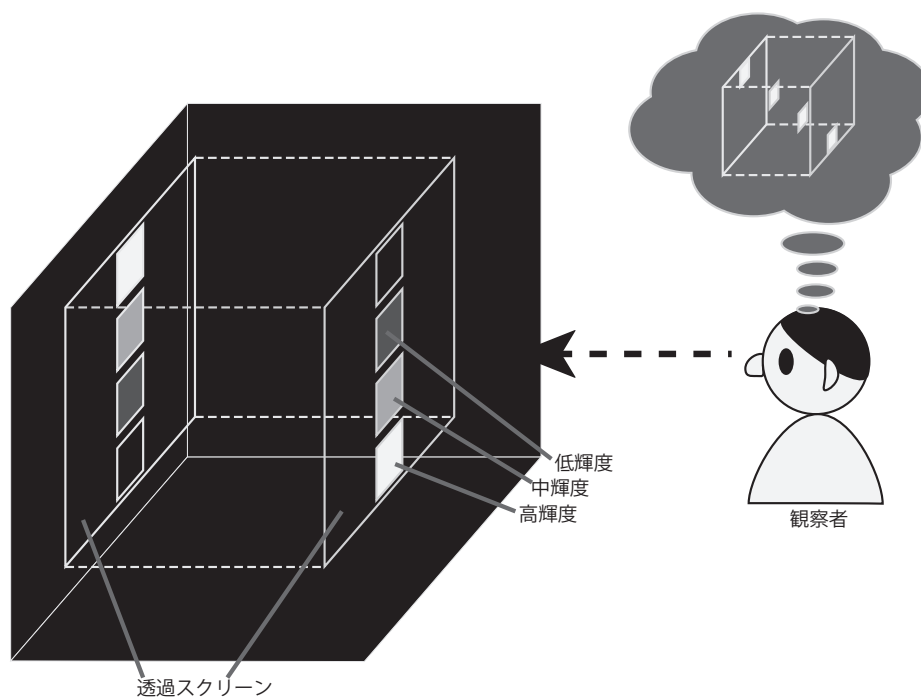


図 5.2 二画面積層表示による 3 次元表示の原理

配分することにより、前後の映像がなめらかにつながり立体的な顔画像を表示できる。

この立体的な顔映像を用いて、視線および顔の向き表現方法について説明する。正面方向を見ている顔の画像を、図 5.3 のように、前面と後面を重ねて表示する。ユーザ A の位置では、前面と後面の積層された映像がきちんと重なる方向となるため、ユーザ A からは、自分の方を向いているように感じられる。

一方、ユーザ B の位置からは、図 5.3 中で示したように前面映像と後面映像の中心が少しずれて重なって見える。人の顔の画像を表示している場合には、目や鼻の位置がやや左の方へずれ、右のほほの部分の面積などがやや広く、左のほほの部分の面積などはやや狭く見える。その結果、ユーザ B の位置から見える顔の画像は、やや左の方を向いているように感じられる。

次に、視線や顔の向いている方向の制御について述べる。図 5.4 のように、後面に表示されている映像の表示位置をずらし、前面と後面の中心軸を結んだ線が、ユーザ B に向くようにする。この時、ユーザ B の位置からは、立体的かつ正面を向いた顔として見ることがで

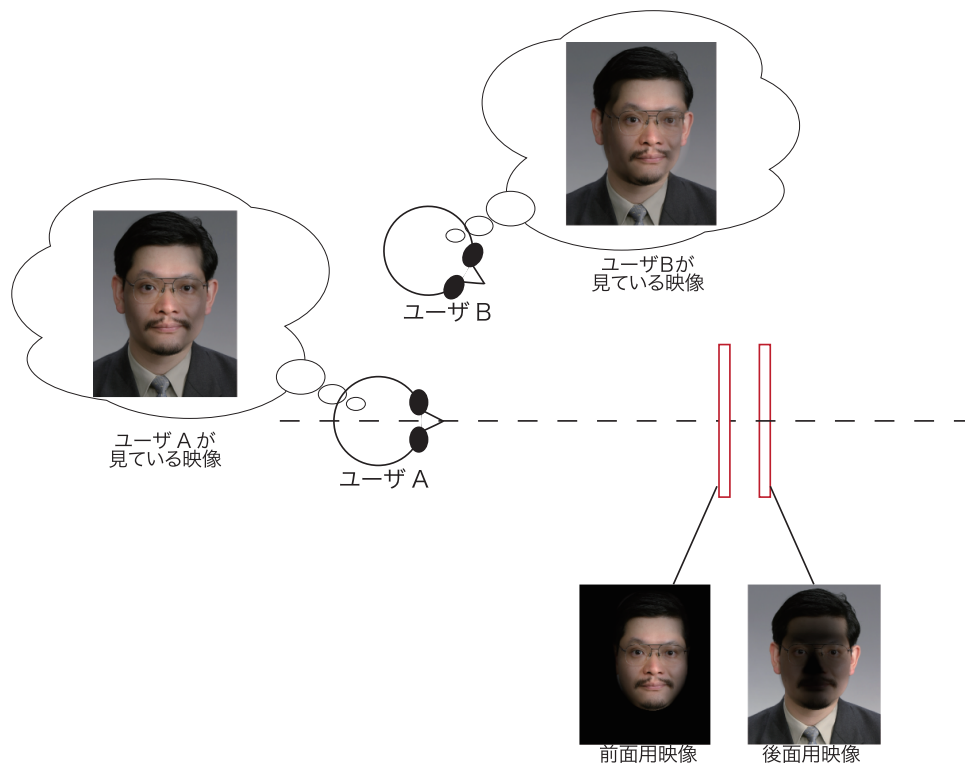


図 5.3 二画面積層表示による顔の向き表現

き、ユーザー B は自分の方を向いているように感じる。一方、ユーザー A の位置から見ると、目や鼻の位置がずれ、左右のほぼの部分の面積などが変化するため、表示されている顔がやや右の方を向いているように感じられる。つまり、映像の表示位置をずらすという操作によって、表示されている顔画像の視線や顔の向いている方向の感じ方を制御することが可能となる。

二画面積層表示の基本的な特性の評価を行うため、次の3つの基礎実験を行った。実験参加者は、20才代～50才代の男女12名である。

- 視線が合うと感じる方向は制限されるか、また視線が合うと感じる範囲はどの程度かを確認するための実験（基礎実験1）
- 後面の表示位置をずらすことによって、視線が合うと感じる方向が制御可能かを確認するための実験（基礎実験2）
- 後面の表示位置をずらすことによって、左右異なる方向を見せられた場合に、その視線方向が認識可能かを確認するための実験（基礎実験3）

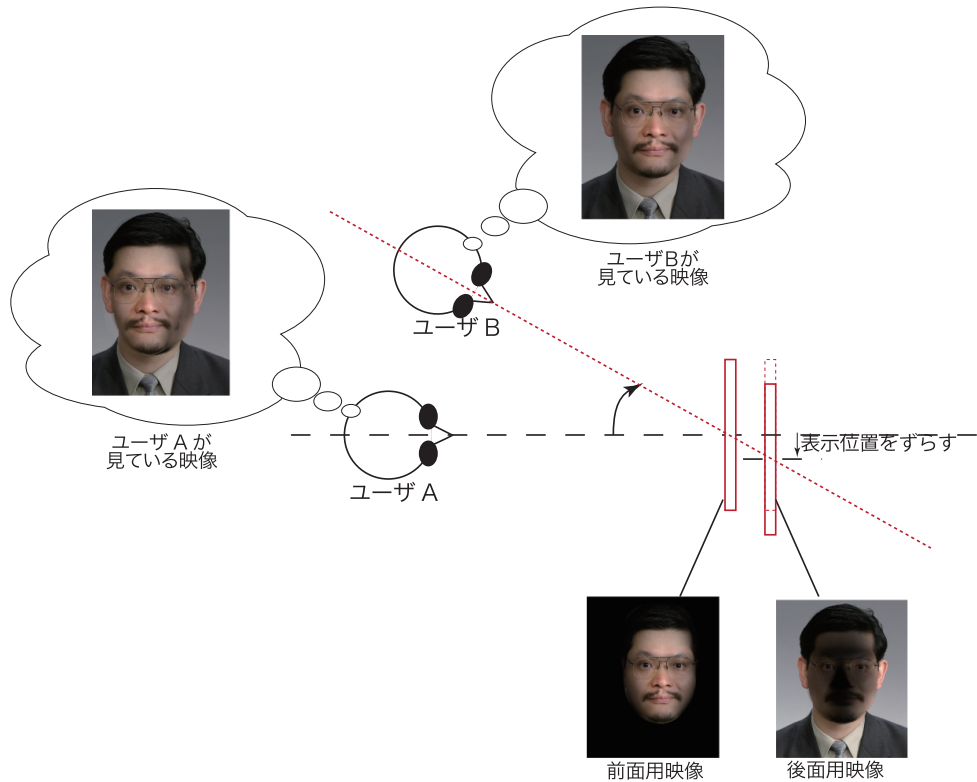


図 5.4 二画面積層表示による顔の向き制御

5.4 提案手法の表現性能に関する実験

5.4.1 基礎実験の環境

基礎実験用環境の表示部は、図 5.5 のように二つのディスプレイを配置し、その間にハーフミラーを置いている。ハーフミラーは、前面用 LCD の映像を反射しつつ、後面用 LCD の映像は透過する。その結果、前面用 LCD の映像の鏡像と、後面用 LCD の映像を重ねて見ることができる。二つのディスプレイとハーフミラーは、自動回転ステージの上に設置されており、回転速度や回転角度を PC によって制御する。二つのディスプレイには、図 5.6 の静止画を表示する。二画面積層ディスプレイに関する仕様は、表 5.1 の通りである。評価用の画像を基礎実験用の評価器に表示している様子を図 5.7 に示す。実験参加者は、ディスプレイから 1,200mm 離れた場所に設置したあご台の上に顔を固定し、この画像を観察する。

表 5.1 基礎実験用評価機のディスプレイに関する主な仕様

ディスプレイサイズ	17型 (縦方向で使用)
表示方式	液晶
表示領域 [mm]	270.3 x 337.9
解像度 [pixel]	1024 x 1280
ドットピッチ [mm/pixel]	0.48
二画面合成方式	ハーフミラー
面間 [mm]	35

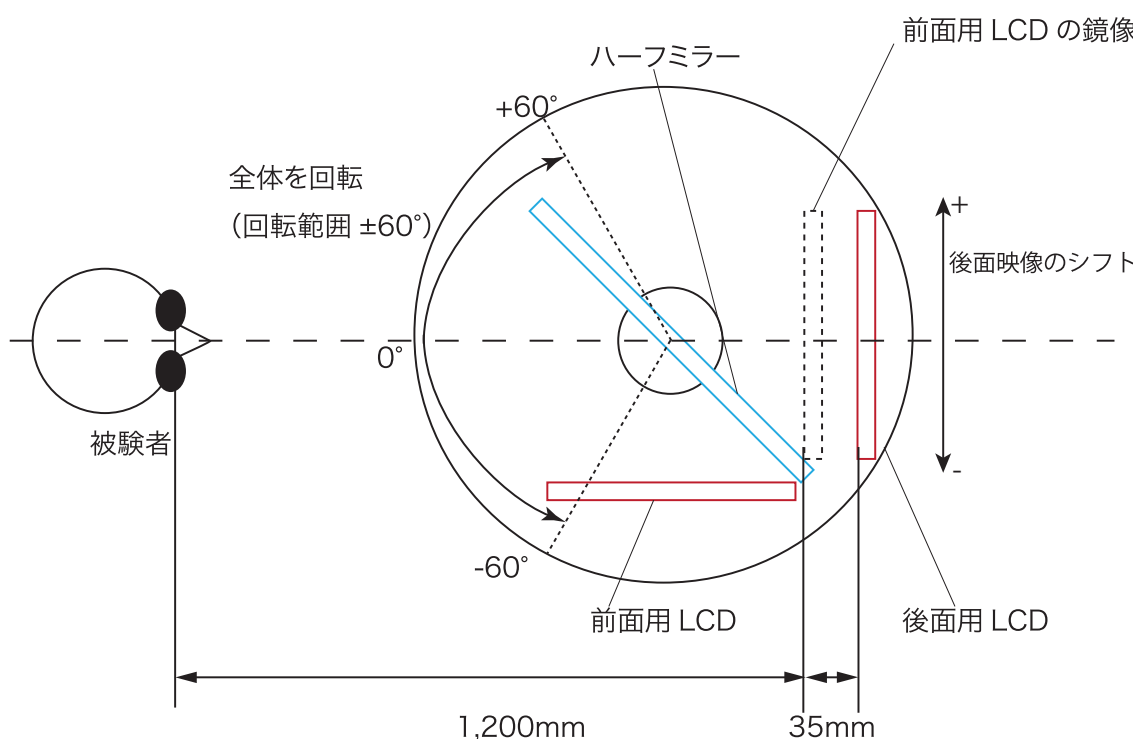


図 5.5 基礎実験用機器の構成

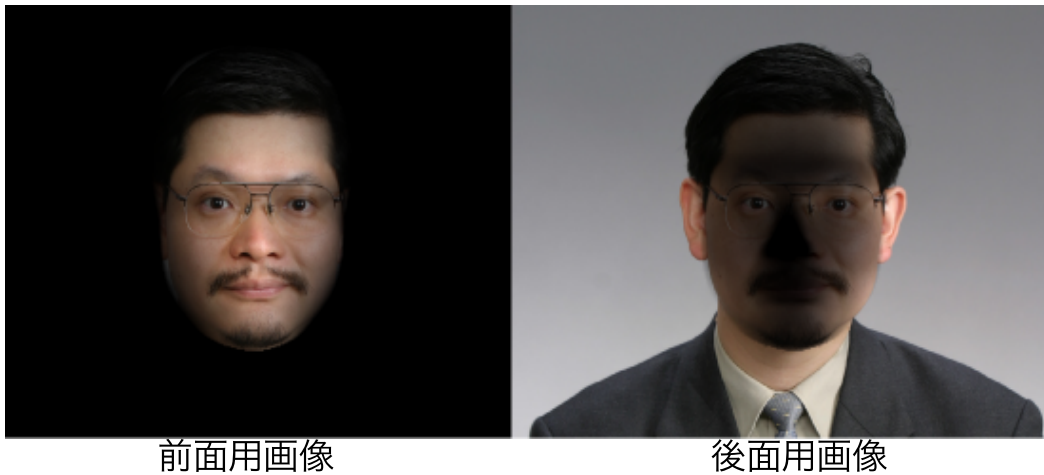


図 5.6 基礎実験の評価時に使用した画像

5.4.2 実験方法および実験結果

基礎実験 1

この実験では、目が合っていると感じる範囲を測定した。画像の表示されているディスプレイを左右に回転させ、実験参加者には目が合っていると感じる間だけマウスボタンを押すという動作で、その範囲を回答してもらった。具体的な実験方法を示す。図 5.6 が表示されたディスプレイを、図 5.5 の $+60^\circ$ の方向から、 -60° の方向へ一定の速度で回転させる。回転速度は、 $3.8^\circ / \text{sec}$ である。実験参加者は、マウスを握った状態で、あご台に顔を固定する。実験参加者に対しては、「視線が合っていると感じている間は、マウスの右ボタンを押しつづけてください」と指示した。実験では、マウスを押している区間の、ディスプレイの回転角度を自動回転ステージから取得した。次に、 -60° から $+60^\circ$ の方向へ逆回転させ同様の測定を行った。一人の実験参加者に対し、ディスプレイを二往復させ、合計 4 回測定を行った。

各実験参加者に対する実験で測定された右端、左端の値より、中央値および、目があっていると感じた範囲を計算した。中央値と目があっていると感じた範囲の計算式は次の通りである。



図 5.7 基礎実験用の評価機へ実験用画像を表示した時の様子

表 5.2 基礎実験 1 の結果

	中央値 [°]	左端 [°]	右端 [°]	範囲 [°]
平均値	+3.0	-5.8	+11.7	17.5
標準偏差	2.0	4.4	6.0	9.7

$$\text{中央値} = \frac{\text{右端} + \text{左端}}{2}$$

$$\text{範囲} = \text{右端} - \text{左端}$$

全実験参加者の平均と、標準偏差を表 5.2 に示す。中央値は、平均値で +3°，標準偏差は 2° であり、ばらつきは小さい。すべての実験参加者はディスプレイがほぼ正面方向を向いている時を中心にして、目があると感じたと答えている。また、目が合っていると感じられる範囲については、平均値として、17.5° という幅に制限されているが、ばらつきを示す標準偏差は 9.7° であり、中央値のばらつきと比較すると大きく、目が合うと感じる範囲については、個人差が大きいことが確認された。

基礎実験 2

この実験では、後面の表示位置を変えることで、目が合うと感じられる方向が変化することを確認した。後面に表示されている画像の表示位置を、+10, -10, +20, -20pixel ずらした、4つの条件で実験 1 と同様の測定を行った。表 5.1 に示したドットピッチと、前面と後面の面間の値の関係から、後面の位置を 10pixel 左右にずらすと、前面と後面の画像の中心を結んだ線は、約 7.8° 左右にずれる。20pixel ずらした場合には、左右に 15.3° ずれたこととなる。前面と後面の中心を結んだ線の向いている方向を、表示されている顔の振れ角と考える。

実験の結果は、表 5.3, 図 5.8 の通りである。はじめに、中央値の平均値に注目する。後面の表示位置を +10pixel ずらした場合、ディスプレイに表示されている顔の振れ角は +7.8° となるのに対し、実験参加者は +6.4° を中心に目があうと感じている。同様に、振れ角が -7.8° に対しては、-1.2° が中央値となり、顔の触れ角と中央値の変化する方向は一致している。このことは、後面の表示位置を変える操作によって、ディスプレイ上の顔画像の向きを方向については、正しく制御ができていていることを示している。その一方で、振れ角 +15.3° に対して、実験参加者は目があうと感じる方向は +10.2° が中央値となり、同様

表 5.3 基礎実験 2 の結果

ずらし量 [pixel]	顔の振れ角 [°]		中央値 [°]	左端 [°]	右端 [°]	範囲 [°]
+20	+15.3	平均値	+10.2	+3.3	+18.5	15.2
		標準偏差	2.2	7.2	4.0	10.8
+10	+7.8	平均値	+6.4	-1.9	+14.7	16.6
		標準偏差	1.7	4.1	4.9	8.3
-10	-7.8	平均値	-1.2	-10.2	+7.7	17.7
		標準偏差	2.3	4.6	6.5	10.3
-20	-15.3	平均値	-4.3	-12.5	+4.0	16.5
		標準偏差	2.8	5.5	6.7	10.9

に、振れ角を -15.3° に対しては、 -4.3° が中央値という結果となった。

つまり、顔画像の向きが変わっていることは認識されているが、その顔画像の振れ角の大きさについては、それほど大きく振れているようには感じないことを示している。

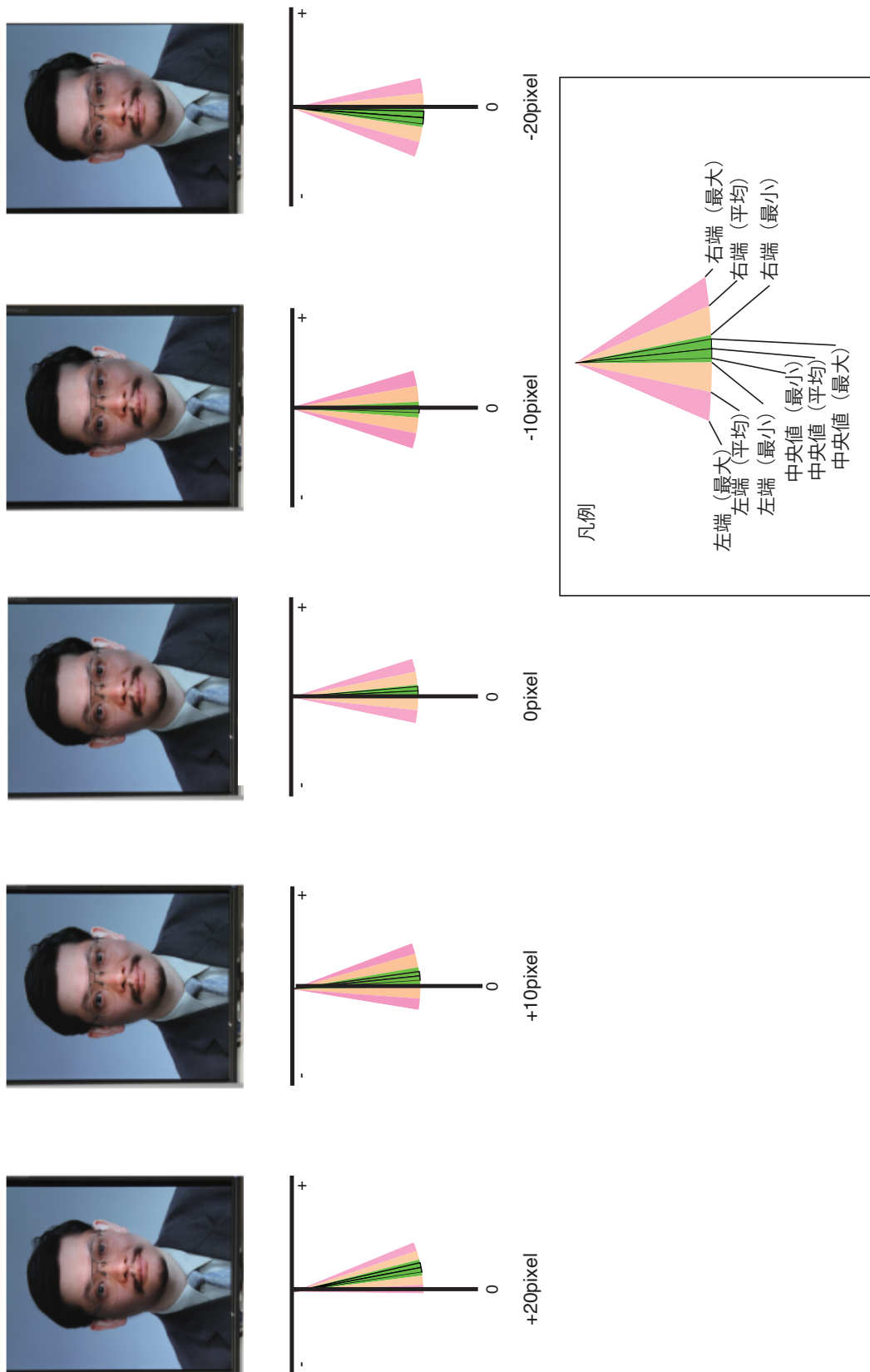


図 5.8 映像のずらし量と視線が一致する範囲の関係

表 5.4 基礎実験 3 の結果

ずらし量 [pixel]	-20	-10	+10	+20
制御した顔の振れ角 [°]	-15.3	-7.8	+7.8	+15.3
実験参加者が認識している顔画像の振れ角 (平均値) [°]	-6.2	-2.9	+7.9	+8.2

基礎実験 3

この実験では、表示されている顔の振れ角が、第三者の視点で見た際どのように認識されるかを確認した。実験参加者は、後面の映像の表示位置を左右に変化させたディスプレイを、正面から見た状態で、表示されている顔画像がどの方向を向いていると感じるか回答した。

実験参加者に対しては、「ディスプレイの顔がどこを向いているのか指をさしてください。」と指示をした。実験参加者は、アゴ台に顔を固定させた状態のまま、表示されている顔がどちらの方向を向いていると感じるか、自分の指で示した。実験参加者の額の中心位置と、指の位置の間の距離から、実験参加者が認識している顔画像の振れ角を計算するため、あご台の額の部分に、予め約 1m の金属の棒を水平に設置した。実験参加者は、その金属の棒上で指を置くように、顔画像が向いていると感じる方向を指し示した。額の中心から、指までの距離を測定した結果と、その距離から計算した実験参加者が認識している顔画像の振れ角を表 5.4 に示す。

はじめに、制御した顔の振れ角と、実験参加者が認識している顔画像の振れ角の方向について注目する。例えば、制御した振れ角が $+7.8^\circ$ に対して、実験参加者が認識した顔画像の振れ角は、 $+7.9^\circ$ であり、制御した顔の振れ角が -7.8 に対して、認識されている顔画像の振れ角は -2.9° ある。画像は右を向くように制御すれば、実験参加者からも右の方を向いているように感じられ、左に制御すれば左を向いたように感じられるように制御が可能であると考えている。次に、その大きさに注目すると、実験 2 と同様に振れ角を大きくしても、実験参加者はそれほど顔が大きく振れているようには感じなかったことを示している。

5.5 基礎実験の結果に対する考察

基礎実験 1 の結果より、二画面積層表示を用いることで、すべての実験参加者が一定の方向でのみ目が合っていると確認された。目があっていると感じる範囲は、実験参加者によって個人差があるものの、平均すると約 17° の範囲に制限されている。仮

に実験環境と同じ 1.2m の距離から観察した場合には、約 36cm の範囲で目が合っていると感じることに相当する。会議室の中で特定の人間に向けて、視線や顔を向けるためには十分な指向性が得られていると考えられる。

また、基礎実験 1, 2 の結果より、視線が合っていると感じる範囲の中央値の比較を行った。図 5.9 に後面のずらし量に対する中央値の平均と標準偏差を示す。各線の中央の目盛りは、中央値の平均を示している。このグラフの通り、表示位置をずらすことによって、目が合っていると感じる方向を制御できることが確認された。また、実験 3 の結果から、表示位置によって顔の画像を右に制御すれば、人間は顔が右に振れているように感じ、顔画像を左に制御すれば人間は顔画像が左に振れたように感じられており、顔画像がどちらの方を向いているのかについては正しく認識されたことが確認された。つまり、第三者的な視点から見た場合においても、その方向に顔の向きが変わっているように認識できることを示している。視線の方向を制限することは、二次元的な映像だけでは不可能なことであり、2 枚の映像の重なる位置をずらすという簡易な方法で、このような表現をできるということは、テレビ会議システムに実装する上で有利な特徴となる。

一方で、図 5.9 より、表示位置のずらし量は、左右対称に変化させたとしても、実験参加者が感じる視線の方向には偏りがあることがわかった。この原因の 1 つとして、顔の側面の陰影が影響していると考えている。今回の実験では、図 5.6 のように、顔の右側面が陰になった画像を使用している。後面の映像は、中央部分が暗くなっている。後面の映像をずらした際、右側面の陰とつながって、陰の部分の面積が広く見える場合と、後面の暗い部分が隠れ、陰の部分の面積が狭くなる場合がある。結果として、後面のずらす方向によって陰の部分の面積が異なるため、顔画像の向いている方向の感じ方が、左右に振ったときに異なったものと考えている。

今回の実験では、実験用画像として図 5.6 の画像一種類だけを使用しているため、表示する顔の個人差の影響については、十分な検証ができていない。人間の顔は、個人によって、顔の輪郭や目や鼻の位置、眼鏡やひげの有無など様々な違いがある。個人差の影響については、特に、基礎実験 3 のような顔が左右のどちらを向いているのかの認識に影響を与える可能性が高い。第三者的な視点から顔の方向を認識させる場合、目や鼻の位置のズレや左右のほぼバランスなどが、向きの認識の手がかりになっていると考えている。人間の顔は必ずしも左右対称ではないことを考慮すると、後面映像をずらしたときの変化が認識されにくい可能性はある。今後、次章で述べるプロトタイプシステムを用いることで、リアルタイムに立体的な映像の撮像と表示が可能となる。今回実験に用いた以外の画像でも実験を容易に行えるため、顔の個人差への影響についても検証ができると考えている。

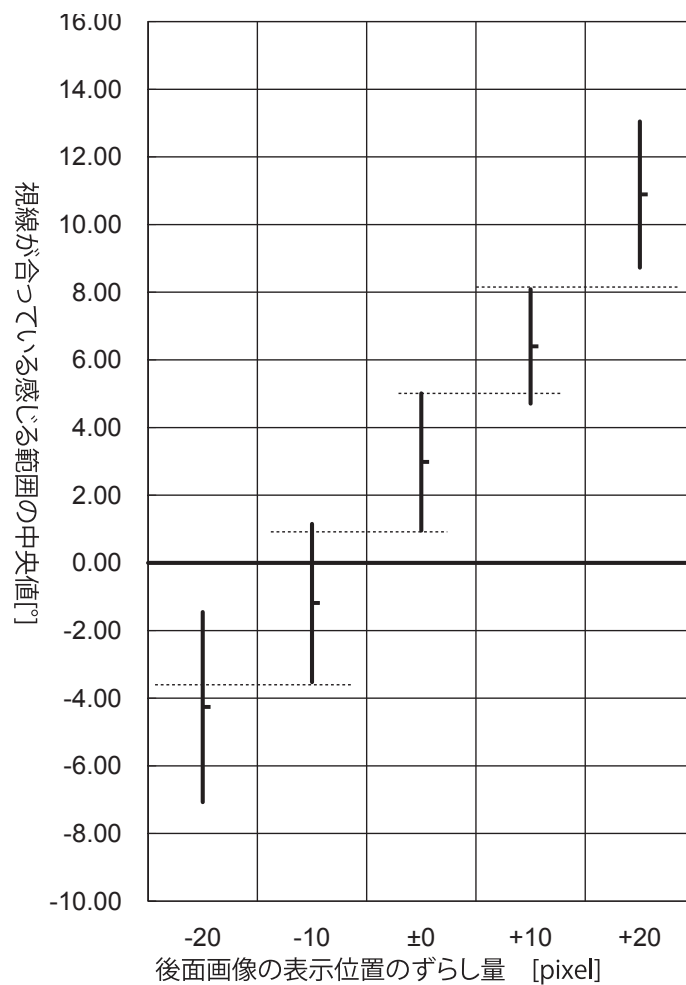


図 5.9 ずらし量と目があっていると感じる範囲の関係

5.6 プロトタイプシステム

5.6.1 プロトタイプシステムの構成

離れた場所から会議に参加する実験を行うため、実験結果に基づき二画面積層表示を用いたプロトタイプシステムの設計を行った。一般的に会議室は、議論される内容や会議の形態に応じて様々な大きさの部屋があり、机や椅子の配置なども変更されるため、あらゆる会議室での利用を想定したシステムを最初から設計していくことは難しい。そこで、今回のプロトタイプ設計においては、小規模の会議を想定して設計を行った。小規模な会議であれば、参加者同士の距離が比較的近く、視線や顔の向きなどの情報が、議論に与える影響が大

表 5.5 プロトタイプシステムの端末表示部分の主な仕様

ディスプレイサイズ	11v 型
表示方式	有機 E L
表示領域 [mm]	251 x 141
パネル解像度 [pixel]	960 x 540
ドットピッチ [mm/pixel]	0.20
二画面合成方式	ハーフミラー
面間 [mm]	20

大きく、今回の二画面積層表示の効果を検証しやすいと考えたためである。実際に準備をした会議室の構成は次の通りである。

1. 会議室の参加人数は最大で 10 人程度で机の周囲に着席する。
2. 会議室の中央には机が配置され、机の大きさは縦 4,000mm x 横 1,300mm である。
3. 机の周りには、プロジェクタ用のスクリーンや、ホワイトボードが設置される。

この会議室の中で、端末がすべての参加者に視線を向けられるようにするためには、二画面積層表示されている顔映像を水平方向に少なくとも 180° 程度回転できるようにする必要がある。そこで、視線の方向や顔の向きの変現については、前面、後面の表示位置をずらす方法と、自動回転ステージによってディスプレイの向きを機械的に変える 2 つの方法の組み合わせを検討した。表示位置を変更する方法は、表現できる視線や顔の向きの範囲は狭いが、機械的な動作を必要とせず、高速かつ細かな変化を表現することが可能である。一方、ディスプレイを機械的に回転させる方法は、広い範囲に視線や顔を向けることが可能となるが、機械的な動作を伴う場合、ディスプレイの表示位置の制御と比較すると応答速度が遅くなり、細かな角度の制御も難しい。それぞれの特徴は、補完する関係にあるため、それぞれの特徴を組み合わせることで、広範囲に、繊細な視線の方向や顔の向きを表現できるものと考えている。

構築したプロトタイプシステムの全体構成図を図 5.10 に示す。また、会議室側に設置する端末の外観を図 5.11 に示す。表示部分の主な仕様は表 5.5 の通りである。

基礎実験用評価器と主な構造の違いとして、会議室側に設置する端末では、ディスプレイとミラーの設置位置を変更し左右対称の構造に変更している。また、小型の自動回転ステージに変更したため、表示部を小型化、軽量化が必要となり、搭載するディスプレイを有機 E

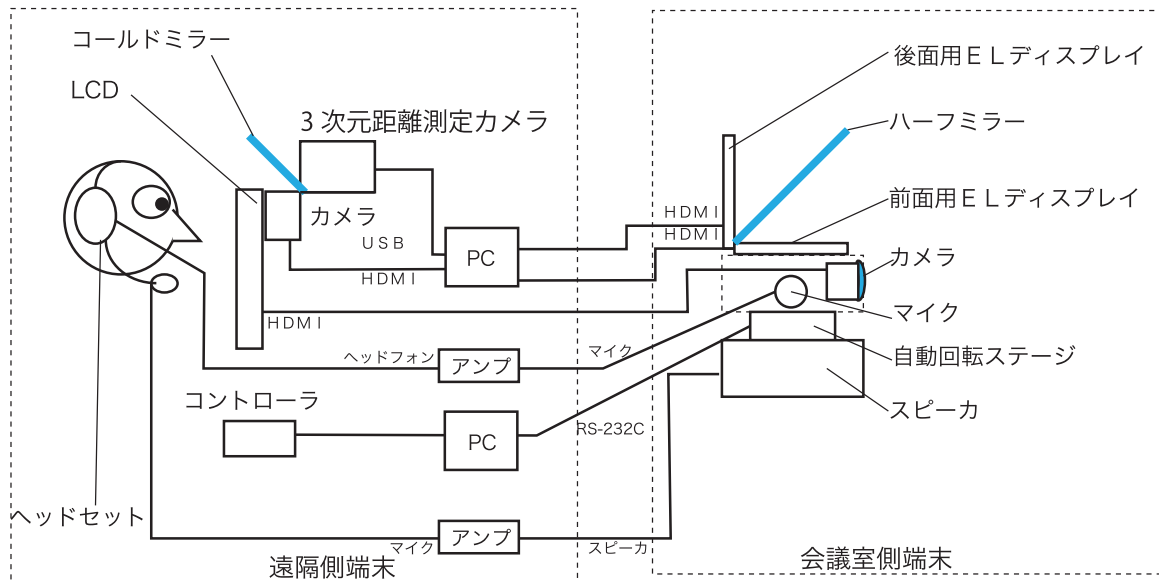


図 5.10 プロトタイプシステムの構成

表 5.6 映像生成をしている PC の主な仕様

CPU	Core i7 950 3.06GHz
Memory	3GB
OS	Windows XP SP3(32bit)

Lの薄型ディスプレイへ変更している。ハーフミラーとディスプレイの相対的な配置は、実験用評価機と同等である。次に、遠隔作業員として立体的な映像が必要となる。そこで、リアルタイムに前後二面の映像を生成するために、遠隔側作業員側の撮影装置には、スイス MESA 社製の 3次元距離測定カメラと、ハイビジョンの小型カメラを組み合わせた撮像装置 (図 5.12) を作成した。この 2つのカメラの映像から、二画面積層表示用の顔映像を生成している [92]。

プロトタイプシステムでは、3次元距離測定カメラから取得した奥行き情報に基づき、前後の映像の輝度分配を計算し、前後二面の映像をリアルタイムに生成している。映像は、1秒間に 20~25 フレーム生成している。映像生成に用いた PC の主なスペックを表 5.6 に示す。リアルタイムに生成した映像は、図 5.13 のようになる。顔の奥行きに応じて、前面用、後面用の映像を生成している。顔の奥行きを計測する範囲については、手動で設定してお



図 5.11 二画面積層表示を用いたプロトタイプシステムの端末

り、撮影する人の鼻の先から耳のあたりまでを計測範囲として設定した。

また、自動回転ステージの回転を操作するインターフェースとして、マウス、キーボード、磁気センサの中から1つを選択して使用することが可能である。ディスプレイと、カメラは連動して回転するようになっており、遠隔からカメラを操作することで、それに連動して二画面積層のディスプレイの向きが変わる構造となっている。プロトタイプシステム表示端末で、顔の向きと視線の向きを表現している様子を図 5.14 に示す。

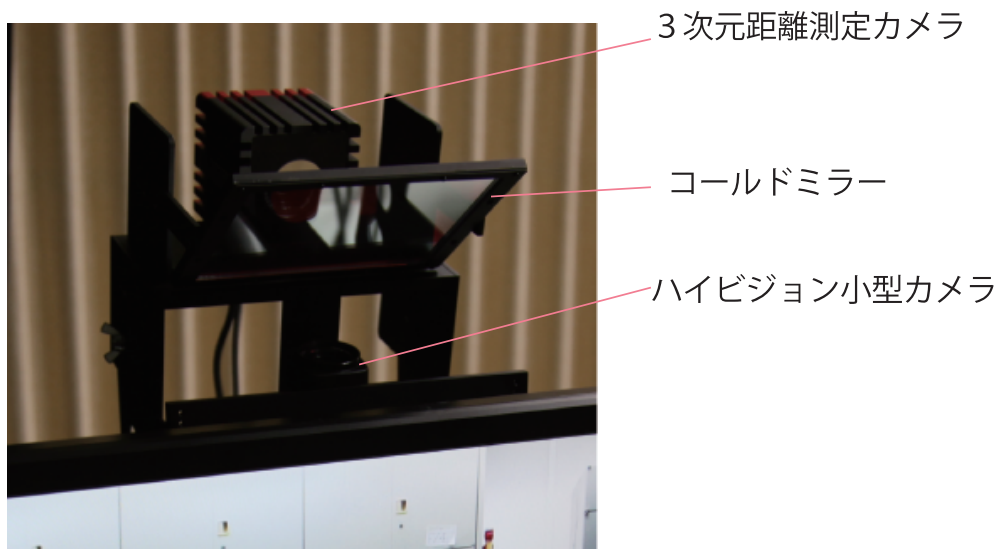


図 5.12 リアルタイム撮像部の構成



図 5.13 リアルタイムに生成した映像

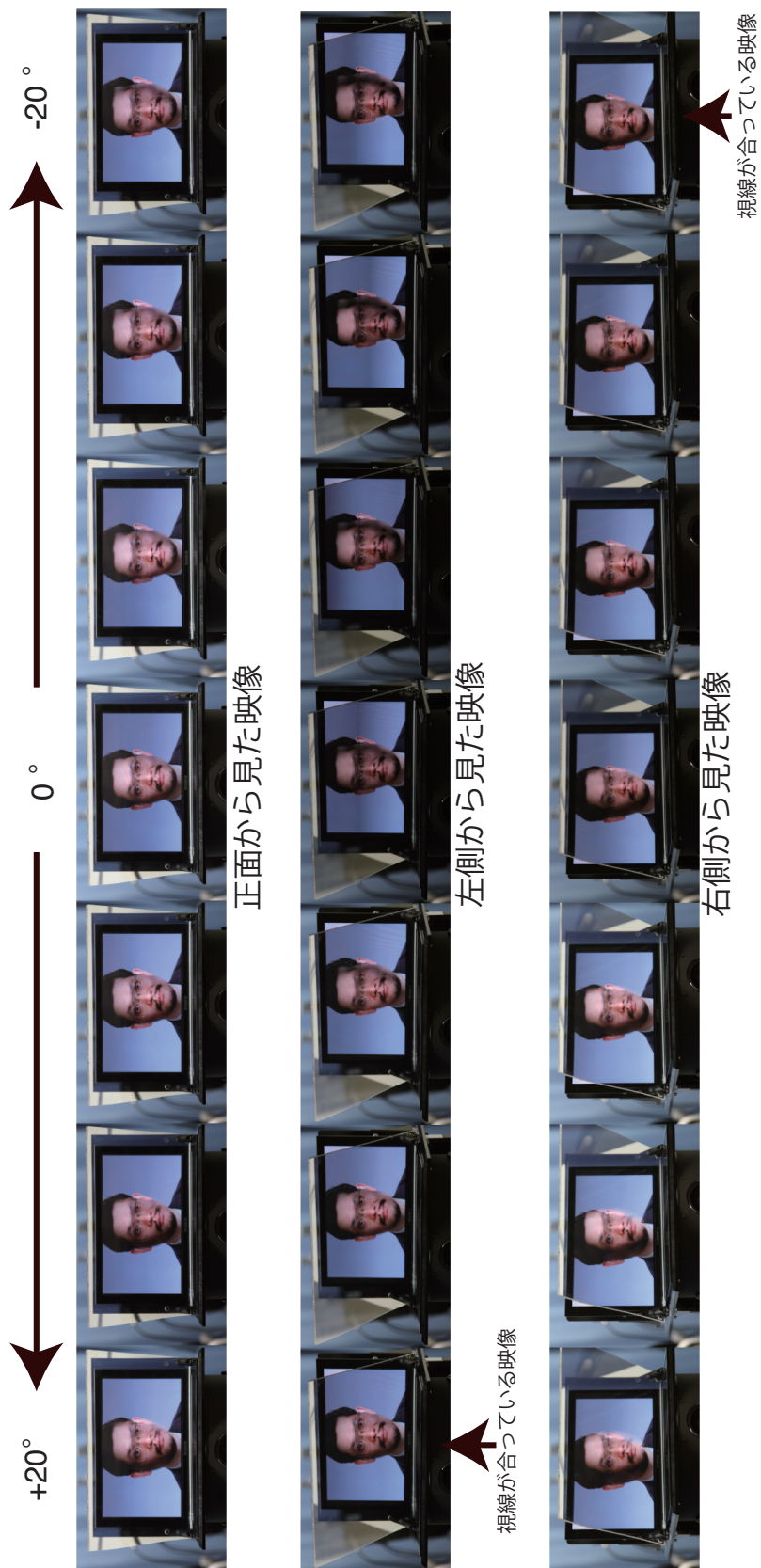


図 5.14 プロトタイプ端末を使って顔の向きおよび視線の向きを表現している様子

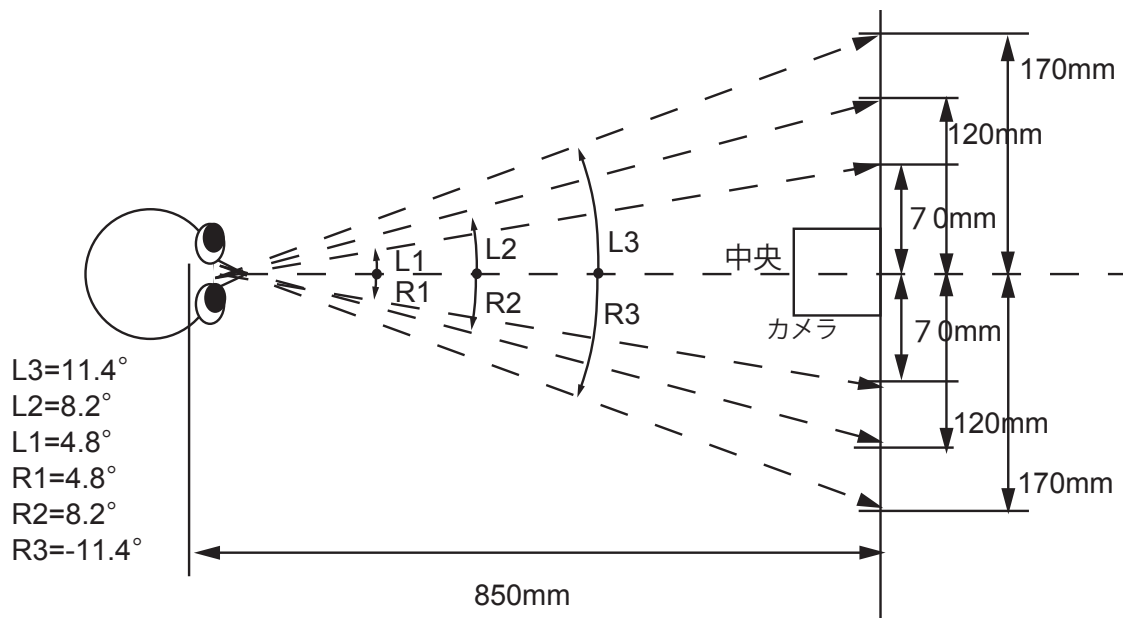


図 5.15 評価用画像の撮影条件

5.6.2 プロトタイプシステムを使った立体映像の撮影

遠隔側端末の撮像装置を使って顔映像の撮影を行った。二人の男性と一人の女性の映像をプロタイプシステムの撮影機能を使って撮影した。撮影時の条件を図 5.15 に示す。カメラから頭部の位置が 850mm になるように着座し撮影を行った。撮影した映像は 1 名あたり 7 種類である。カメラの中心をみて撮影した映像が 1 枚。次に、頭部の方向はそのままの状態とし、視線だけを左右に向けた映像を 6 種類撮影した。カメラの左右に目印を置き、視線の方向だけずらして撮影を行った。撮影された画像は図 5.16 の通りである。

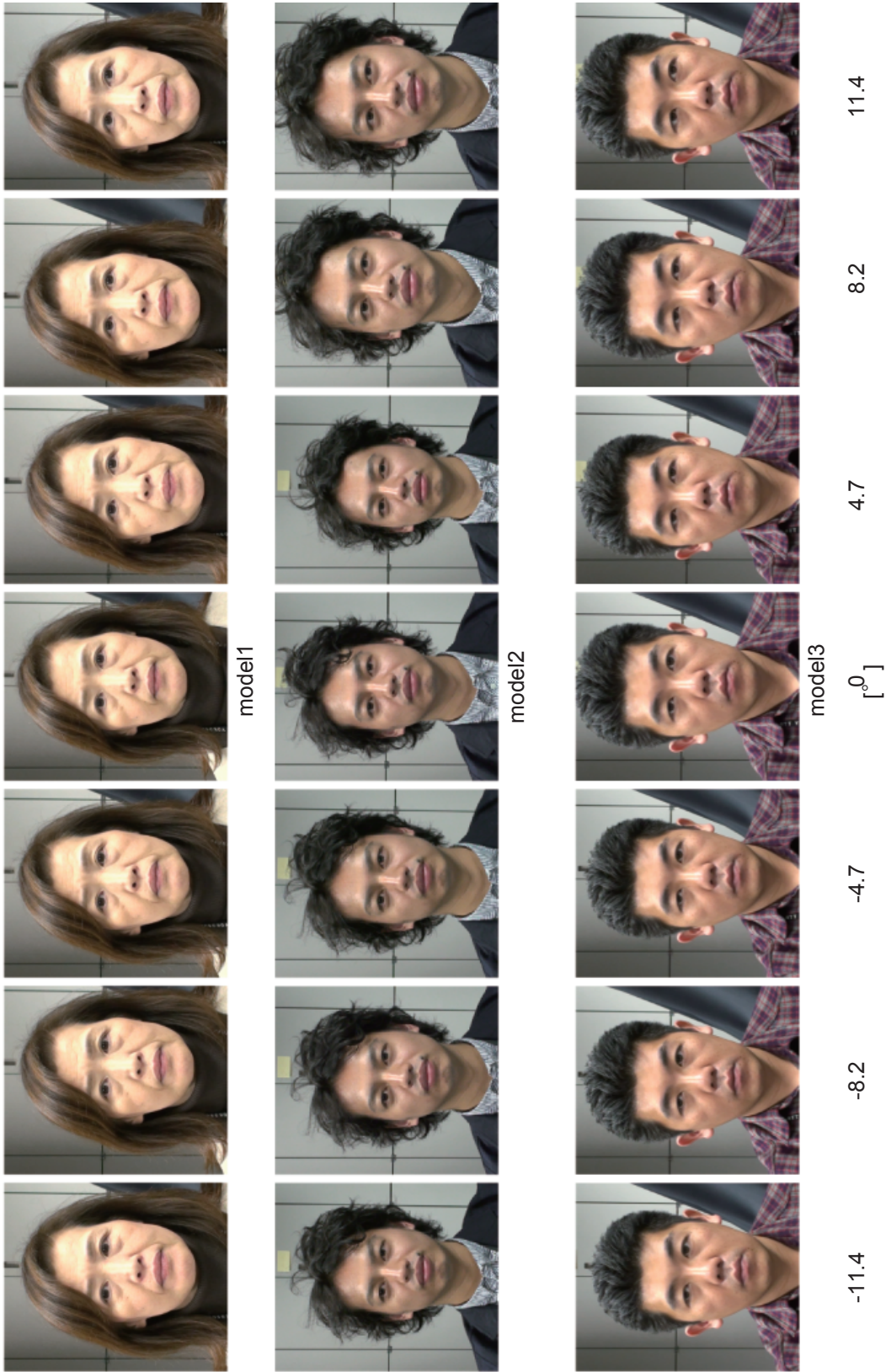


図 5.16 撮影された評価用画像

5.6.3 実験方法

プロトタイプシステムで撮影された映像を使用し、視線表現の性能を評価するため、以下の2つの実験を行った。実験1では、プロタイプシステムで生成した映像が、静止画で行った基礎実験と同じ特性を示すかを検証した。実験2では、顔の向きを変えずに視線のみが変化した場合の視線一致の性能を検証した。

- 実験1：プロトタイプ端末に表示された正面の顔の映像と視線が一致する範囲を測定する
- 実験2：プロトタイプ端末に表示された映像の顔の向きが正面のまま、視線方向だけ変化した場合の視線が合う範囲を測定する。

実験参加者は、年齢が20代～50代の男性7人、女性3人の合計10人である。実験環境を図5.17に、実験時の様子を図5.18に示す。実験参加者は映像から1200mmの位置で頭部を固定し画像を観察する。会議室側端末のディスプレイはマウスを左右に動かすとそれにあわせて回転するようになっている。実験参加者は視線が自分の方を向いていると感じたときに、マウスの移動を止め、クリックする。クリックしたときのディスプレイの角度を記録した。右端、左端および最もよく視線が合うと感じる3点でマウスのボタンをクリックするように指示した。

5.6.4 実験結果

実験1

評価用映像の顔(model1～3)が正面を向いているときの結果を表5.7～5.9、図5.19～5.21に示す。範囲の計算方法は、基礎実験の時と同じである。表5.7～5.9によれば、最もよく視線が合うと感じる角度の平均は、 $-0.29 \sim 1.83^\circ$ であり、標準偏差は $1.44 \sim 3.06^\circ$ であった。平均的に視線が合うと感じている範囲は、 $17.9 \sim 21.1^\circ$ であった。

実験2

評価用映像のmodel1～3の視線が左右に向いているときの、実験結果を表5.10および図5.22に示す。視線の方向がR1/L1の方向の場合、すべての実験参加者が視線を一致すると感じる事ができたが、R2/L2では、一部の実験参加者は視線があうと感じることが無く、R3/L3の条件では、すべての実験参加者が視線を合わないと感じた。

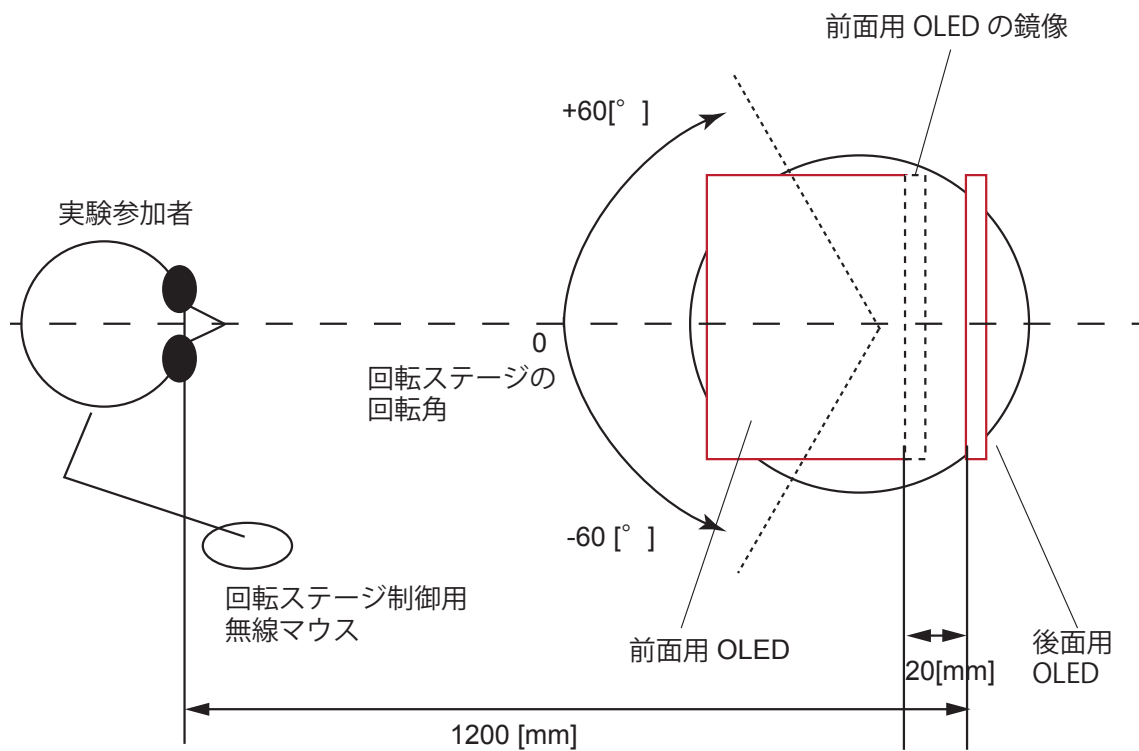


図 5.17 実験時のシステムの構成

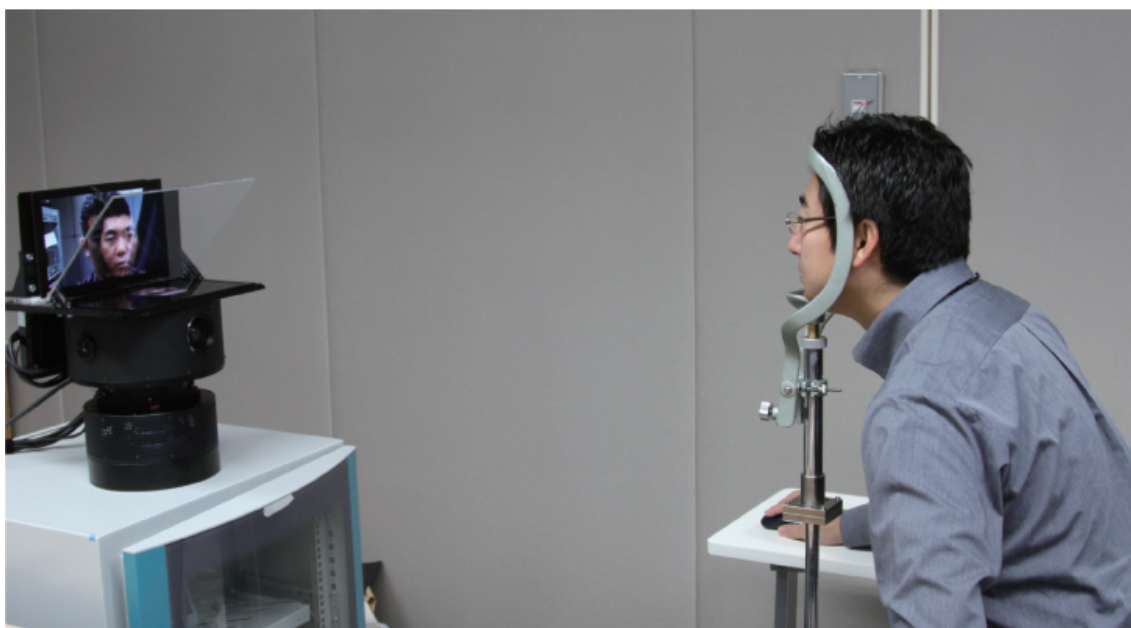


図 5.18 実験者が映像を観察している様子

表 5.7 正面を向いている評価映像による実験結果 (model1)

実験参加者	右端 [°]	最もよく視線が合う角度 [°]	左端 [°]	範囲 [°]
A	-2.93	-0.94	4.42	7.35
B	-2.99	-1.27	3.80	6.79
C	-8.68	3.34	8.17	16.85
D	-16.33	0.58	7.53	23.86
E	-4.37	0.52	3.68	8.05
F	-15.30	-0.12	14.89	30.19
G	-2.70	1.67	7.94	10.64
H	-17.31	-5.58	14.09	31.40
I	-9.95	1.21	11.21	21.16
J	-14.72	-2.30	8.05	22.77
平均値	-9.53	-0.29	8.38	17.90
標準偏差	3.98	6.03	2.45	19.08

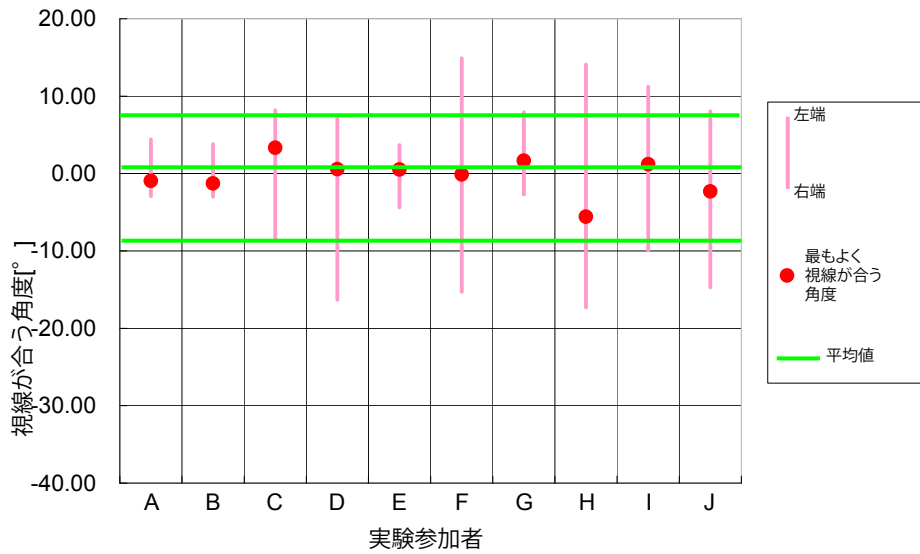


図 5.19 実験結果 (Model1)

表 5.8 正面を向いている評価映像による実験結果 (model2)

実験参加者	右端 [°]	最もよく視線が合う角度 [°]	左端 [°]	範囲 [° .]
A	-0.36	0.68	1.04	1.40
B	-8.13	0.52	3.22	11.34
C	-2.93	1.15	7.36	10.29
D	-28.64	2.88	16.33	44.97
E	-5.35	-1.78	5.35	10.70
F	-19.09	-0.06	12.25	31.34
G	-13.05	1.09	4.03	17.08
H	-18.92	-1.50	5.98	24.90
I	-9.66	-1.38	5.69	15.35
J	-9.14	0.70	3.74	12.88
平均値	-11.42	0.13	6.50	18.02
標準偏差	4.56	8.59	1.44	19.87

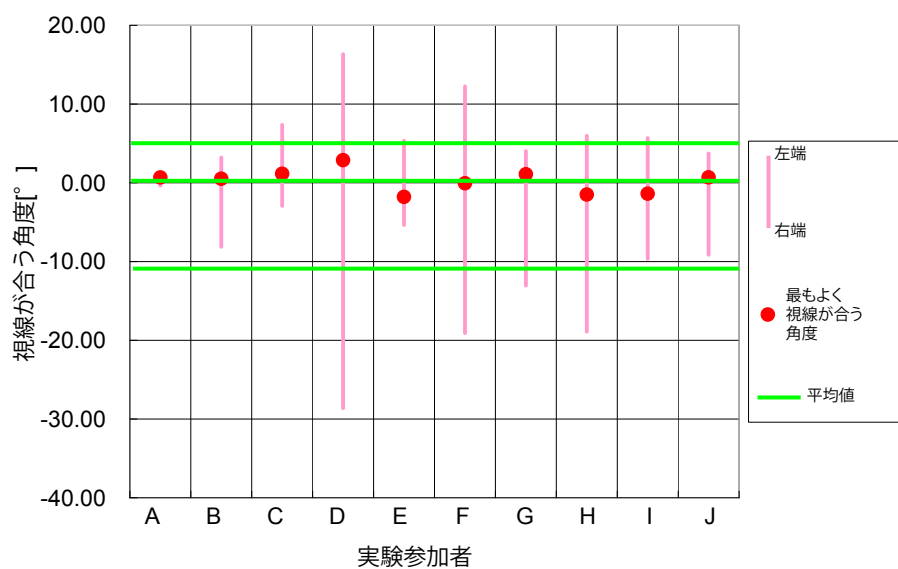


図 5.20 実験結果 (Model2)

表 5.9 正面を向いている評価映像による実験結果 (model3)

実験参加者	右端 [°]	最もよく視線が合う角度 [°]	左端 [°]	範囲 [°]
A	-0.59	1.79	2.05	2.63
B	-2.67	0.07	7.31	9.98
C	-8.57	0.81	13.34	21.91
D	-38.53	6.94	19.09	57.62
E	-5.06	0.58	7.94	13.00
F	-12.02	2.99	12.54	24.55
G	1.84	5.81	9.14	7.30
H	-15.64	-3.62	13.92	29.56
I	-14.43	-0.12	12.77	27.20
J	-5.35	3.05	12.06	17.41
平均値	-10.10	1.83	11.01	21.11
標準偏差	4.63	11.55	3.06	23.17

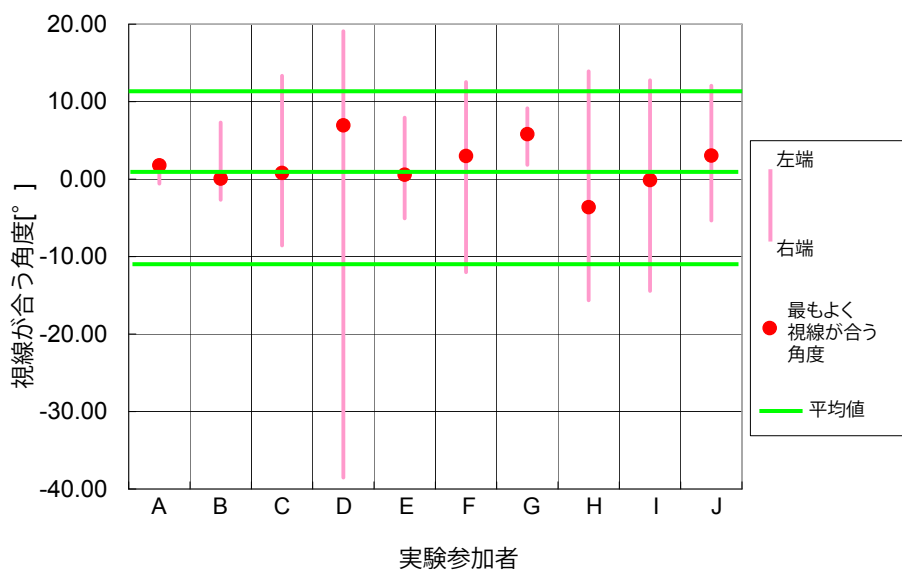


図 5.21 実験結果 (Model3)

表 5.10 評価が用映像の視線が左右に変化したときに最も視線が合うと感じる角度を測定した結果

Model 1					
	R2	R1	Center	L1	L2
平均値 [°]	-15.59	-9.36	-0.29	11.64	16.50
標準偏差 [°]	7.82	6.80	2.45	8.55	10.93
視線が合うと感じた 実験参加者の数	8	10	10	10	8
Model 2					
	R2	R1	Center	L1	L2
平均値 [°]	-14.34	-9.88	0.13	8.83	17.28
標準偏差 [°]	11.80	5.36	1.45	4.84	11.99
視線が合うと感じた 実験参加者の数	6	10	10	10	9
Model 3					
	R2	R1	Center	L1	L2
平均値 [°]	-18.90	-10.02	1.69	15.66	-14.52
標準偏差 [°]	15.038	8.44	3.21	10.98	10.65
視線が合うと感じた 実験参加者の数	8	10	10	9	4

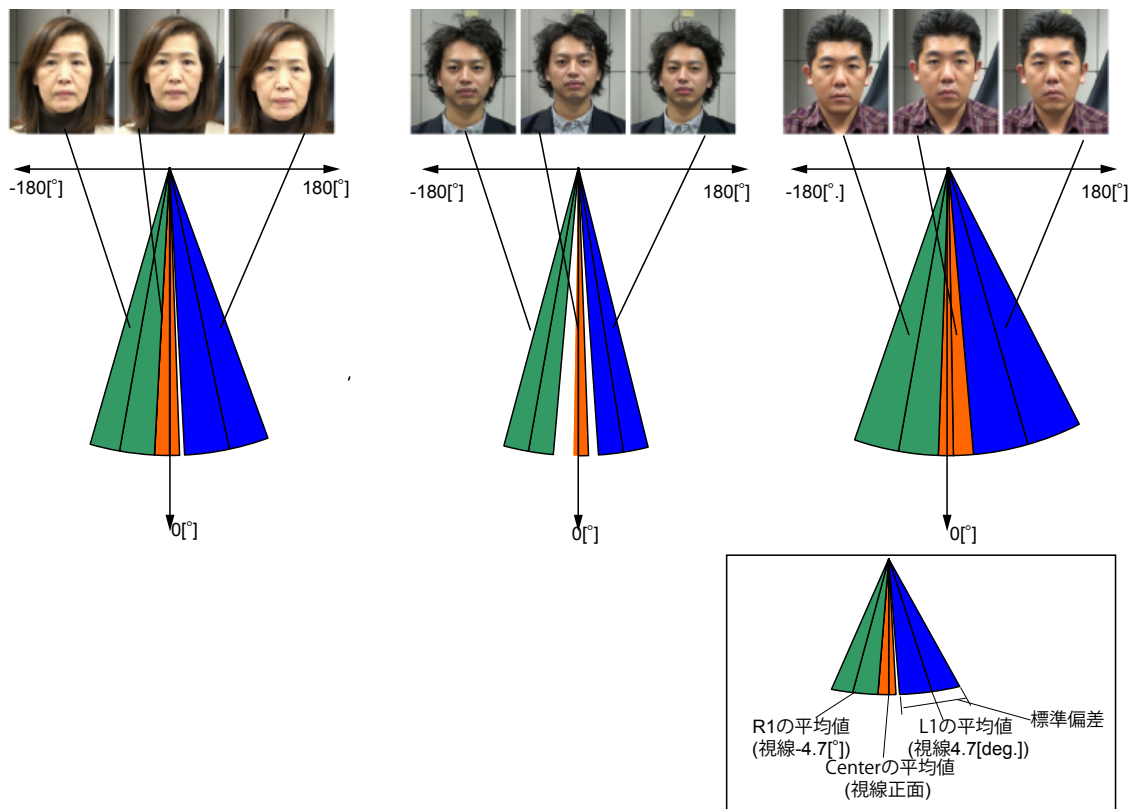


図 5.22 プロトタイプシステムによる実験結果

5.7 考察

5.7.1 2枚の映像のずれによって表現される視線や顔の方向の情報

今回の方式で、人の顔を表示する際、前面の映像と後面の映像は、距離センサの値を使って生成される。前面の映像には、目や鼻、口など部分が含まれる。後面の映像には、顔の輪郭や頬、耳などが含まれる。2枚の画面を積層したDFD映像を見ている人の視点が変わると、前面の映像と後面の映像の想定的な位置に変化があり、運動視差を生じるため、映像が立体的に見える。この時、表示されている目や鼻、口などの位置と、顔の輪郭、耳などの位置も変化するため、顔の向きや視線の方向が変わったように感じられると考えられる。

この結果により、2つの目的を同時に達成できた。つまり、「指向性のある映像表示：遠隔から参加している参加者と、特定の方向にいる参加者のみ目があると感じられること。」と「方向の制御：遠隔から参加している参加者の視線、顔の向きは制御可能であること。」

と」の2つを実現していることを確認した。二画面積層表示によって表示される映像は、顔や視線の方向を直感的に感じることができる。

5.7.2 表示されている映像の視線表現の性能

プロトタイプシステムを用いた実験1の結果、表示されている顔の特徴により範囲はわずかな差はあるが、すべての実験参加者がある一定の方向でのみ目が合っていると感じることを確認した。目があっていると感じる範囲は、実験参加者によって個人差があるものの、平均で17.9~21.1°であった。仮に実験環境と同じ1.2mの距離から観察した場合には、36cm~46cmの範囲で目が合っていると感じることに相当する。会議室の中で特定の人間に向けて、視線や顔を向けるためには十分な指向性が得られていると考えられる。

実験2の結果より、プロトタイプ端末に表示されている映像中の顔で、視線を向ける角度が小さい条件(L1/R1)については、すべての実験参加者が視線を合うと感ずることができたが、視線を向ける角度が大きい条件(L2/R2)については、一部の実験参加者が視線を合わせることができず、それ以上になるとすべての実験参加者が視線を合わせることができなかった。このことから、遠隔参加者の視線ができるだけ正面方向をむいて撮影されるように、遠隔側の端末の表示と、撮影部分の設計が必要となる。

5.7.3 指向性の範囲の個人差について

ほとんどの実験参加者が、前面の映像と後面の映像をすこしシフトさせたとしても、立体的な1つの顔の映像として感じることができた。しかし、表5.7~5.9によれば、実験参加者Dのみ指向性が広いことを示している。実験後に実験参加者Dへインタビューしたところ、前面の映像を主に見ていたと回答した。そのため、前面の映像と、後面の映像はそれぞれ別々の映像として感じていた。この時、実験参加者は、ディスプレイの前面にある一枚の二次元映像を注視しているため、モナリザ効果によって、広い範囲で視線が一致していると感じたと考えられる。多くの利用者が1つの立体的な顔として映像を感じられるように表示方法は改良する必要がある。1つの方法としては、複数の方向で立体的な映像を表示できる技術[93]の応用が考えられる。

5.8 結言

本章では、視線や顔の向きを伝達するテレビ会議システムの実現のためには、「指向性のある映像表示」と「第三者的な視点における指向性の認識」と、「方向の制御」の3つの要

素が必要であることを示し、二画面積層表示を用いることで、それらを表現できる端末構成を提案した。二画面積層表示が、特定の方向にのみ視線が合うと感じられることを明らかにし、その範囲が約 17° ~ 21° に制限されることを示した。二次元ディスプレイでは、視線の方向や顔の向きを特定の方向に制限することはできないため、それと比較すると、狭い範囲に向きが制限されていることを示している。また、その視線方向は、機械的なディスプレイの回転だけでなく、映像の表示位置を左右方向へずらすという手法によっても、制御することが可能であり、繊細な視線の動きや顔の向きを表現する手法として有効である。また、実験に基づき、プロトタイプシステムを構築し、リアルタイムにコミュニケーション可能なシステムが実現可能であることを示した。今後は、必要な機能の追加を行った後、このシステムを利用し、実際の会議の中で二画面積層表示による視線表現の有用性について明らかにしていく予定である。

第 6 章

考察

6.1 緒言

本章では、本研究の成果を活用した遠隔協同作業環境について考察する。6.2 節では、4, 5 章にて得られた知見をまとめ、遠隔協同作業環境での活用方法とその効果について考察する。6.3 節で、さらに、活用領域を拡大するための、作業者の状態表現方法について考察する。6.4 節で、本研究成果の知識科学への貢献について考察する。

6.2 本研究成果の遠隔協同作業における活用

本研究では、円滑なコミュニケーションを支援する遠隔作業環境の構築を目指し、「非同期作業」と、「同期作業」を支援する 2 つの手法の研究を行った。4 章では、新たな人の状態推定手法を考案し、Daniel ら [21] や Laput ら [20] の手法のように、すべてのセンサーデータに教師情報を付与する手法と比較すると、人手で行う作業の負担を 60%~80% 軽減した。正解率は 90% 以上であり従来手法の正解率と同程度であることを示した。村上ら [66] の手法は予め蓄積されている業務知識を利用しているが、本研究はそうした蓄積された業務知識が無い遠隔協同作業においても、人手作業の負担軽減し、推定器を構築することができる。4 章の実験の結果、いずれの作業環境でも、図 3.4 の、在室中の非同期作業、同期作業に該当する状態を推定し、不在状態も含めて分類して推定できることを確認した。また、作業環境 1 は、コミュニケーション可否を区別して定義できた。本研究は人手の作業を軽減し推定器を構築できる、新しい人の状態手法について、その基本的な性能を実証した。

本推定手法を、2.3 節の形態 3 へ適用することにより、遠隔にいる作業者の状態を、オフィスに設置されているテレプレゼンスロボットの位置や表示装置の動作へマッピングすることができる。遠隔にいる作業者の作業状態は、テレプレゼンスロボットを介してオフィスにいる協同作業員へ表現することができる。これにより、オフィスの協同作業員は、テレプレゼンスロボットの動作から、遠隔にいる作業者の状態を認識し、適切なタイミングでコミュニケーションをとり、同期作業へ移行する環境を構築できる。作業推定結果をオフィス内のテレプレゼンスロボットの位置や動作へマッピングし、安全に移動する手法については、引き続き研究していきたい。

また、5 章の視線表現手法は、ディスプレイを見ている人は視線が合うと感じる範囲が、 17° ~ 21° 範囲に制限できることを確認した。それぞれの手法について、基本的な性能を実証した。装置は小型が可能であり、Misawa ら [24] の手法と同様に、テレプレゼンスロボットへ搭載可能な大きさで構成できる。作業員個人専用のスクリーンを製作することなく、ロ

ロボットを複数の作業者が共用して利用する環境を実現できる手法を確立できた。この表示手法をテレプレゼンスロボットへ搭載することにより、遠隔にいる作業者の視線情報を意図した通りに表現し、オフィスの会議室内で行われるフォーマルな会議への参加や、休憩などを行う共有スペースのインフォーマルなコミュニケーションへの参加も可能となる。本研究のプロトタイプシステムでは、映像・音声を遠隔に通信するシステムは構築できていないため、2枚の映像と音声を同期して伝送するシステムの開発が必要である。映像伝送については、3D テレビ放送にはステレオ映像と音声を伝送する方式が実現されており、こうした方式を参考に、通信システムを設計することができる。また、2枚の映像を撮影し生成する装置は、市場にあるスマートフォンやタブレット端末を活用することで、より小型化することが可能である。RGB カメラと3次元距離カメラは、スマートフォン [94, 95, 96] や、タブレット端末 [97] へ搭載されるようになってきた。これらの端末は、映像と合わせて同期して距離情報が取得可能であり、端末内で映像を生成し、伝送できる。これらの機器を使い、通信システムを構築することで、遠隔にいる作業者が容易に自らの映像を伝送し、オフィスにおける協同作業を円滑に進めることができると考えている。

本研究の成果を活用することで、遠隔にいる作業者の存在や状態は、自動的にオフィスに設置されたテレプレゼンスロボットの動作として表現され、オフィスにいる作業者が、遠隔にいる作業者の状態を認識することを支援できる可能性がある。遠隔にいる作業者とオフィスにいる作業者は互いの状態を認識しながら、シームレスに同期作業へ移行する環境の実現につながりうる。これは、岡田 [31] の協同の一般階層モデルにおいて非同期作業をしている時の共在と人に関する気づきのレベルが向上に貢献すると考える。また、視線表現手法がテレプレゼンスロボットへ搭載されることにより、遠隔にいる作業者が見ている人や、モノに関する情報は、オフィスの中でテレプレゼンスロボットを見ている人へ自然と伝達される。さらに、人とテレプレゼンスロボットの関係や、ロボットとモノの関係が遠隔にいる作業者の意図したとおりに伝わることにより、その空間で行われている作業の状況も自然とオフィスにいる作業者に伝達されるようになると思う。その結果、協同の一般階層モデルにおいて、同期作業を行っている時の「人」、「場/空間」、「モノ」に関する気づきのレベルの向上に貢献すると思われる。

共在と気づきのレベルの向上することにより、円滑なコミュニケーションが支援され、「意識の共有」が進むことで、効率的に遠隔協同作業を行うことが期待される。今後は、本技術を実際の遠隔協同作業環境へ適用し、その効果について実証していきたい。

6.3 本研究成果の活用領域の拡大

本研究で提案した手法により、複数の作業者がいるオフィスに対し遠隔にいる作業者がテレプレゼンスロボットでオフィス環境へ参加する形態(2.3節の形態3)の遠隔協同作業環境を構築できる。しかし、今後はすべての参加者がそれぞれ異なる環境で働く形態も想定される。テレプレゼンスロボットを1つのオフィスへ設置しおこなう遠隔協同作業環境だけではなく、それぞれが異なる実環境で作業する遠隔協同作業への支援も必要となる。従来のコミュニケーションシステムでは、(2.3節の形態1)のように、映像を使って互いの状態を共有しながら、コミュニケーションを行う形態や、(2.3節の形態2)のような、仮想的な空間を共有しながら行う形態が考えられる。形態1では、在宅勤務などで自宅を使う場合のプライバシーへの配慮が必要であり、常時に作業状態を共有し続けることが困難である。形態2は、仮想空間に作業者の状態をそれぞれ反映させ、それぞれの作業者は、作業中に仮想空間の状態を確認する必要がある。作業をしながら、遠隔にいる作業者の状態を自然と認識することは困難になる。そこで、遠隔の作業者の状態を、それぞれの作業環境にいるアバタを使って表現する形態(図6.1)を考える。アバタは、HMDを使い実空間に遠隔の作業者のアバタの映像を重ね合わせて表現する。協同作業同士が、互いに相手の空間を訪れているような状況を再現し、表示されているアバタに声をかけた時点で、インフォーマルにコミュニケーションを開始し、同期作業へ移行する。

筆者は、実際の生活を続けながら離れて生活している家族がいた場合に、その状態を互い

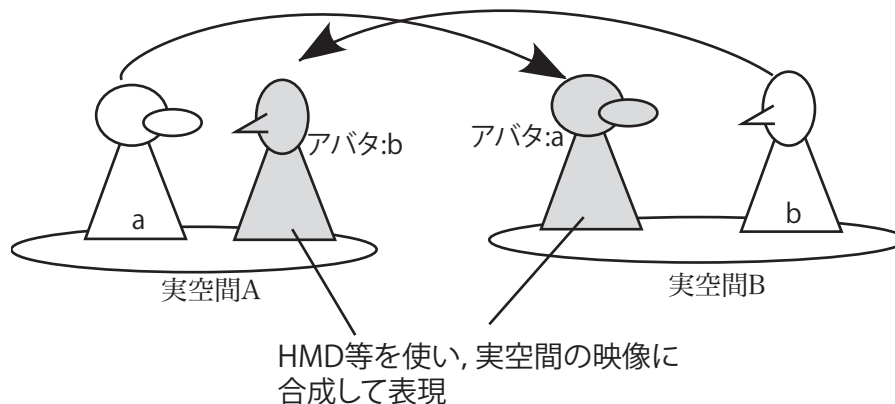


図 6.1 それぞれの空間に応じた情報を表現する形態

の生活空間で表現する手法として、生活融合通信 [98, 29] を提案している。図 6.2 は、生活融合通信のコンセプトのイメージである。図 6.3 は、遠隔にいる人の動きを異なる配置の空間で表現するプロトタイプシステムによる実験の様子である。左側にいる人の動作と関連しているオブジェクトを推定し、右側の空間では、その動作を表現するのに適したオブジェクトを決定し、アバタの姿勢をオブジェクトの配置に応じて変更して表示している。

この手法を、遠隔協同作業環境へ適用させることで、それぞれの作業環境において、遠隔にいる協同作業者の存在や状態を認識しながら作業を行い、シームレスに同期作業へ移行することも可能と考えている。

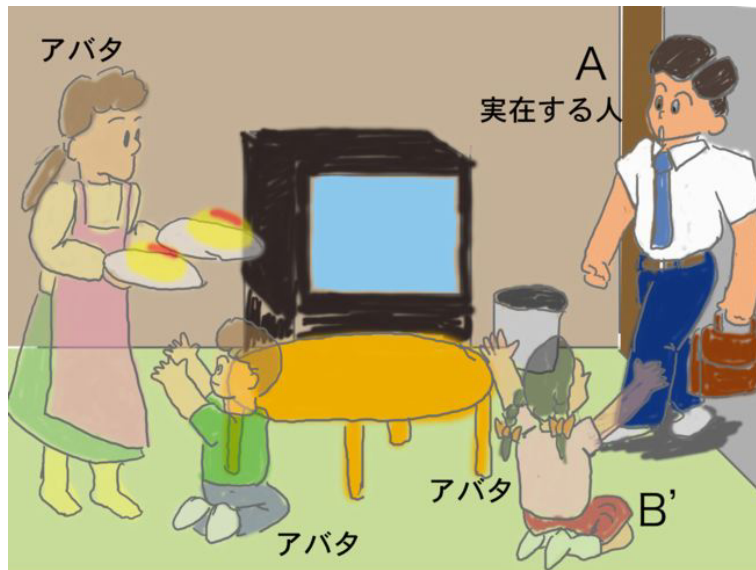
今後は、作業者の状態に応じて、遠隔にいる協同作業者の作業空間で表示するアバタの映像の場所や姿勢を決定するアルゴリズムの研究が必要となる。また、異なる場所で人の顔の表情 [99] やアバタの姿勢を変化させる手法 [100] が提案されており、現実空間の中で、VR空間と実際の空間を融合して HMD で表示させながら安全に歩行できる環境を目指した研究 [101] もある。こうした研究も参考にし、映像として協同作業者を表現する手法について検討する。

さらに、映像を用いずに遠隔にいる協同作業者の状態を表現する手法も必要である。同じ空間で、協同作業を行う場合は、人同士の会話の声や作業している時の音などからも遠隔の作業者の状態を認知している。Winsneski ら [102] は、部屋の中で情報を表現する手法を提案し、宮島ら [103] は遠隔にいる家族の様子を自然と感じられる通信システムを提案している。在宅勤務を自宅で行うことを想定すると、多くの人を部屋の中で表現することは困難である。音や光の変化などによって自然と遠隔にいる作業者の状態を認知できる表現手法を検討していきたい。

6.4 知識科学への貢献

企業や組織がおこなうサービス開発、技術開発は、複数の人が連携し協力して新たな価値を創造する作業であり、松下ら [30] が定義する協同：「複数の人が協力することにより、一人ではなし得ない、あるいはなすことが非常に困難な新しい価値を創造すること」と同じである。アレンら [1] は、技術情報のコミュニケーションのあり方が新製品のイノベーションプロセスに決定的な影響を及ぼしているとしている。近年は、複数の企業や、組織が連携したサービス開発や技術開発の機会が増え、テレワークの導入企業が増加しており、遠隔で協同作業をする機会は増加する傾向にある。企業や組織によるサービス開発や技術開発が、遠隔にいる人同士の協同作業として推進される機会も増加していくと考えられる。

こうした背景の中で、知識科学としてイノベーションのための新たな知識共創環境を構築



単身赴任先



自宅

図 6.2 生活融合通信:ComAdapater による生活空間の実現イメージ

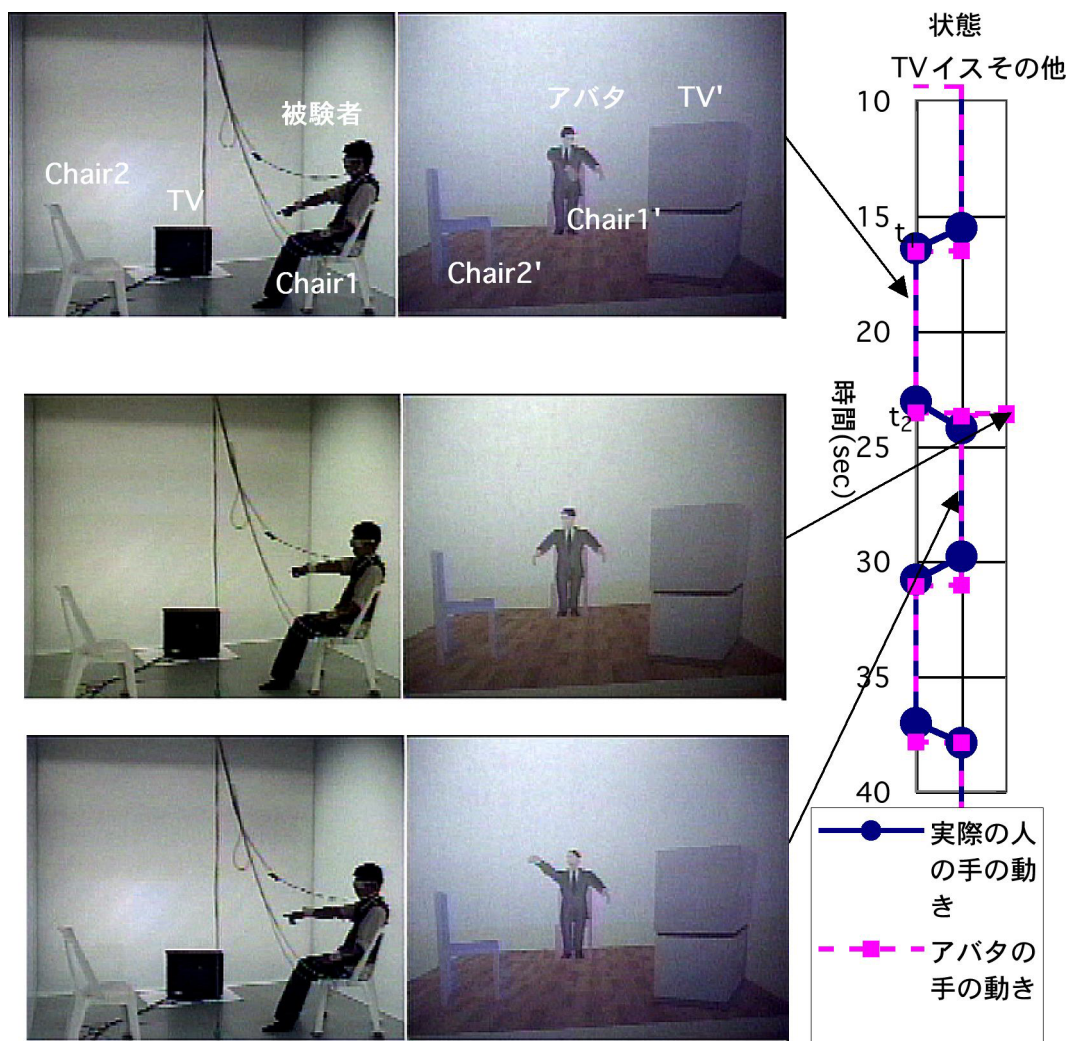


図 6.3 生活融合通信:ComAdapater のプロトタイプシステムの実験の様子

する技術の研究が求められる。本研究は、これまでに CSCW 研究や知識経営の研究において蓄積されてきた、アウェアネス支援や遠隔コミュニケーションに関する知見に着目し、新たに人の状態推定手法と、視線表現手法の 2 つを開発した。その手法の性能を検証するため、プロトタイプシステムを開発し、その基本的な性能を明らかにした。

6.5 結言

本章では、本研究の成果を活用し遠隔協同作業環境について考察した。6.2 節では、それぞれの研究成果を活用し、テレプレゼンスロボットを活用して非同期作業と同期作業をシームレスに遷移させる手法について考察した。6.3 節は、本研究成果の活用範囲を、それぞれの遠隔作業者が異なる環境にいる状況にも拡張し、円滑な遠隔協同作業環境を構築する手法について考察した。6.4 節は、本研究成果の知識科学への貢献について考察した。

第 7 章

結論

7.1 本論文のまとめ

企業や組織が革新的なサービスや技術を生み出すためには、様々な分野の高度な知識を持った人々の協同作業が必要となる。企業間、組織間が連携したサービス開発や技術開発の機会が増加している。国内ではワークライフバランスの向上や、多様な人材活用が求められており、テレワークの導入も増加する傾向にある。一方で、大規模な感染症の流行や災害が発生した際に、企業が業務を継続するための方法としても注目されており、遠隔で協同作業をする機会は増加している。

協同作業においては、協同作業間での気づき（アウェアネス）が重要とされている。人の存在や状態に対する気づきから、コミュニケーションが始まるきっかけが作られる。コミュニケーションをしている間であれば、会話をしている相手の視線によって、話題の対象などが自然と伝わる。遠隔協同作業では、人の存在や状態、周辺の場/空間/モノに関する気づきが不足するため問題となる。

本研究では円滑なコミュニケーションを支援する遠隔協同作業環境の構築を目指し、遠隔協同作業を「非同期作業」と「同期作業」に分類し、それぞれの作業を支援する技術について研究を行った。

「非同期作業」に対しては、協同作業同士が互いに状態を共有するためのプレゼンス共有機能を構築し、「非同期作業」から「同期作業」への移行を支援する環境の構築を目指した。作業者の状態推定には、大量の学習データの作成が必要となるが、本研究の手法により作業負担を60%~80%削減できることを実証した。

「同期作業」に対しては、発話と最も強い関連のある視線情報の表現に着目し、二画面積層表示による視線表現手法を新たに開発した。2枚の映像を積層した表示手法を用いることで、モナリザ効果を抑制し、約17°~21°の範囲でのみ視線が一致する映像を表示できる手法を確立した。

4章では、学習データの作成の負担を軽減するために、教師無しクラスタリング法や対話インタフェースを活用して人の状態を推定する推定器の構築手法を提案した。システムが自動的に行う3つのステップは、「データ収集」、「分類」、「選択」であり、この3つのステップに加え、作業者が手動で行う「ラベリング」のステップを加えて4つのステップで推定器を構築する。提案手法を評価するためにプロトタイプシステムを構築し、2つの作業環境に端末を設置し、2回の実験を行った。推定器の構築のために、作業者が閲覧する必要がある映像の時間は、収集した全映像と比較すると、約60%~80%少ない時間を閲覧するだけで、推定器を構築できることが確認された。システムによって同様の特徴の映像が選択されるこ

とで、作業状態の確認も容易になり、質的な面においても負担を軽減している。推定器の正解率は、90%以上となった。提案手法により、人の作業負担を軽減しながら、従来法と同等の性能を実現することを確認した。また、対話インタフェースを用いてラベリングする手法とすることで、システムの利用に関して習熟する負担の軽減にもつながった。また、作業者が対話をしながら、それぞれの作業環境毎に適した状態を定義することができ、コミュニケーションの可否をメッセージとして伝えることも可能となった。

5章では、視線や顔の向きを伝達するテレビ会議システムの実現のためには、「指向性のある映像表示」と「第三者的な視点における指向性の認識」と、「方向の制御」の3つの要素が必要であることを示し、二画面積層表示を用いることで、それらを表現できる映像表示手法を提案した。二画面積層表示が、特定の方向にのみ視線が合うと感じられることを明らかにし、その範囲が約 17° ～ 21° に制限されることを示した。二次元ディスプレイでは、視線の方向や顔の向きを特定の方向に制限することはできないため、比較すると狭い範囲に向きが制限されていることを示している。また、その視線方向は、機械的なディスプレイの回転だけでなく、映像の表示位置を左右方向へずらすという手法によっても、制御することが可能であり、繊細な視線の動きや顔の向きを表現する手法として有効である。

6章では、4章、5章の研究成果を活用し、人の状態推定結果をオフィスにあるテレプレゼンスロボットを用いて表現し、シームレスに同期作業へ移行する手法について考察した。さらに、筆者が提案した「生活融合通信」のコンセプトを発展させ、本研究成果の活用範囲を拡大し、協同作業者がそれぞれ異なる作業環境にいる形態でも、非同期作業から同期作業へシームレスに遷移する方法を考察した。そして、本研究が情報通信環境下での新たな知識創造環境の構築につながる成果であり、知識科学の基礎技術の確立に貢献することを述べた。

以上より、本論文は、新たな人の状態推定手法と、視線表現手法を確立し、その活用方法から「非同期作業」と「同期作業」をシームレスに遷移させる新たな遠隔協同作業環境を提示した。

7.2 今後の課題

人の状態推定機能については、テレプレゼンスロボットの動作として遠隔にいる作業者の状態表現を行い、継続的に遠隔で協同作業をする業務の中で活用しながら、状態共有機能の効果の検証を行う。作業者の間でコミュニケーション機会の増加や、コミュニケーションの中断抑止効果に着目して検証を行う。また、推定された作業状態をテレプレゼンスロボットの動作へマッピングしながら、オフィス内を安全に移動させる手法についても継続して研究を行う。

リアルタイムコミュニケーションにおける視線表現手法については、遠隔地との通信が可能なプロトタイプシステムを設計し、スマートフォンや、タブレット端末を用いて撮影装置の改良も加えて、遠隔協同作業環境におけるコミュニケーション支援の効果を検証する。

また、今後はすべての作業者がそれぞれ異なる実環境にしながら遠隔協同作業を行う機会も増加すると考えられるため、そうした環境においても、「非同期作業」と「同期作業」をシームレスに遷移させながら、円滑なコミュニケーションを支援する手法の研究について、引き続き行っていく。

謝辞

本研究を進めるにあたって、多くの方々に多大なご支援をいただきました。この場をお借りして感謝の意を示します。

指導教員である由井菌隆也准教授には、研究に関して様々のご指導、ご鞭撻を賜りました。社会人学生として研究に打ち込める環境をご提供いただいたことに深く感謝いたします。

博士後期課程の副テーマ研究において、明治大学の小林稔教授には、副テーマの実験や分析の手法に加え、幅広い観点で研究に関するアドバイス・ご指導いただきました。深謝いたします。

副テーマ指導教員の神田陽治教授および、副指導教員の永井由佳里教授には本研究テーマを進めるに当たり、幅広い観点で研究に関するアドバイス・ご指導を頂き、非常にお忙しい中ご尽力いただきました。深謝いたします。

外部審査委員の東京大学大学院 小川剛史准教授，内部審査委員 内平直志教授，金井秀明准教授，郷右近英臣准教授には貴重なお時間を割いて頂き，研究に関して様々で懇切なるご助言を賜りましたことを，心より感謝いたします。

5章，6章の研究においては，日本電信電話株式会社 サイバースペース研究所，メディアインテリジェンス研究所の関係者の皆様の協力を頂き，研究コンセプトの確立，実験方法，実験環境，実験実施に協力いただいたことに深く感謝いたします。

また，研究者としての基礎をつくり，研究者の道へ導いていただいた元知識科学研究科研究科長 杉山公造教授に深く感謝いたします。

最後に私事で恐縮ですが、長い学生生活を支え続けてくれた妻・綾に、心より感謝の意を表わさせていただきます。ありがとうございました。

参考文献

- [1] トーマス J アレン・グンター W ヘン・糀谷 利雄 (訳)・富樫 経廣 (訳)・日揮株式会社 (監修)：「知的創造の現場: プロジェクトハウスが組織と人を変革する」, ダイヤモンド社 (2008)
- [2] JOIC・NEDO：「オープンイノベーション白書 第二版」(2019)
- [3] 総務省：「情報通信白書」(2019)
- [4] R. S. Fish, R. E. Kraut, B. L. Chalfonte: “The videowindow system in informal communication”, in Proceedings of the 1990 ACM Conference on Computer-Supported Cooperative Work, ser. CSCW ’ 90, pp.1–11, New York, NY, USA (1990)
- [5] S. Whittaker, J. Swanson, J. Kucan, C. Sidner: “Telenotes: Managing lightweight interactions in the desktop”, *ACM Trans. Comput.-Hum. Interact.*, Vol.4, No.2, pp.137–168 (1997-6)
- [6] G. Venolia, J. Tang, R. Cervantes, S. Bly, G. Robertson, B. Lee, K. Inkpen: “Embodied social proxy: Mediating interpersonal connection in hub-and-satellite teams”, in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ser. CHI ’ 10, pp.1049–1058, New York, NY, USA (2010)
- [7] M. K. Lee, L. Takayama: ““now, i have a body”: Uses and social norms for mobile remote presence in the workplace”, in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ser. CHI ’ 11, pp.33 – 42, New York, NY, USA (2011)
- [8] P. Dourish, V. Bellotti: “Awareness and coordination in shared workspaces”, in Proceedings of the 1992 ACM Conference on Computer-supported Cooperative Work, ser. CSCW ’92, pp.107–114 (1992)
- [9] D. D. Salvucci, N. A. Taatgen, J. P. Borst: “Toward a unified theory of the multi-tasking continuum: From concurrent performance to task switching, interruption,

- and resumption”, in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ser. CHI '09, pp.1819–1828, New York, NY, USA (2009)
- [10] G. Mark, V. M. Gonzalez, J. Harris: “No task left behind?: Examining the nature of fragmented work”, in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ser. CHI '05, pp.321–330, New York, NY, USA (2005)
- [11] 黒川 隆生:「ノンバーバルインターフェース」, オーム社 (1994)
- [12] A. Kendon: “Some functions of gaze-direction in social interaction”, *Acta Psychologica*, Vol.26, pp.22–63 (1967)
- [13] 「テレプレゼンスロボット Kubi」, <http://kubi-robot.com>, accessed on April 18, 2020.
- [14] “beam”, <https://suitabletech.com>, accessed on April 18, 2020.
- [15] “Double”, <https://www.doublerobotics.com>, accessed on April 18, 2020.
- [16] “temi”, <https://www.robotemi.com>, accessed on April 18, 2020.
- [17] 高橋 一成・桑田 喜隆:「プレゼンスサービスを活用したオフィスのコラボレーション支援」, 情報処理学会論文誌, Vol.48, No.1, pp.2–15 (2007-1)
- [18] M. Züger, C. Corley, A. N. Meyer, B. Li, T. Fritz, D. Shepherd, V. Augustine, P. Francis, N. Kraft, W. Snipes: “Reducing interruptions at work: A large-scale field study of flowlight”, in Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, ser. CHI '17, pp.61–72, New York, NY, USA (2017)
- [19] S. Hashimoto, T. Tanaka, K. Aoki, K. Fujita: “Improvement of interruptibility estimation during pc work by reflecting conversation status”, *IEICE Transactions on Information and Systems*, Vol.E97.D, No.12, pp.3171–3180 (2014)
- [20] G. Laput, Y. Zhang, C. Harrison: “Synthetic sensors: Towards general-purpose sensing”, in Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, ser. CHI '17, pp.3986–3999, New York, NY, USA (2017)
- [21] D. Avrahami, M. Patel, Y. Yamaura, S. Kratz: “Below the surface: Unobtrusive activity recognition for work surfaces using rf-radar sensing”, in 23rd International Conference on Intelligent User Interfaces, ser. IUI '18, pp.439–451, New York, NY, USA (2018)
- [22] I. Feldmann, W. Waizenegger, N. Atzpadin, O. Schreer: “Real-time depth estimation for immersive 3d videoconferencing”, in 2010 3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video, pp.1–4 (2010-6)

- [23] A. Jones, M. Lang, G. Fyffe, X. Yu, J. Busch, I. McDowall, M. Bolas, P. Debevec: “Achieving eye contact in a one-to-many 3d video teleconferencing system”, *ACM Trans. Graph.*, Vol.28, No.3, pp.64:1–64:8 (2009-7)
- [24] K. Misawa, Y. Ishiguro, J. Rekimoto: “Livemask: A telepresence surrogate system with a face-shaped screen for supporting nonverbal communication”, *Journal of Information Processing*, Vol.21, No.2, pp.295–303 (2013)
- [25] K. Iso, T. Yuizono, M. Kobayashi: “Worker state estimation method with reduced manual task for teleworking environment”, *International Journal of Informatics Society*, Vol.11, No.3, pp.177–185 (2020)
- [26] 磯 和之・由井 隆也: 「テレワーク環境に適応したプレゼンス共有のための作業状態推定システムの協調的な構築手法」, 日本バーチャルリアリティ学会論文誌, Vol.25, No.1, pp.21–30 (2020)
- [27] 磯 和之・伊達 宗和・高田 英明・安藤 康子・松浦 宣彦: 「視線の向きを表現可能な2画面積層表示を用いたテレビ会議システムの提案」, 情報処理学会論文誌, Vol.52, No.3, pp.1224–1233 (2011-3)
- [28] K. Iso, S. Ozawa, M. Date, H. Takada, Y. Andoh, N. Matsuura: “Video conference 3d display that fuses images to replicate gaze direction”, *Journal of Display Technology*, Vol.8, No.9, pp.511–520 (2012)
- [29] 磯 和之・八木 貴史・小林 稔・岩城 敏・石橋 聡: 「生活融合通信: 空間情報整合化機能“ComAdapter”」, 日本バーチャルリアリティ学会論文誌, Vol.9, No.2, pp.169–178 (2004)
- [30] 松下 温・岡田 謙一: 「コラボレーションとコミュニケーション」, 共立出版 (1995)
- [31] 岡田 謙一: 「社会の未来を拓くネットワーク情報共有空間: 1. 情報共有空間における協同」, 情報処理, Vol.48, No.2, pp.123–127 (2007-2)
- [32] 内平 直志: 「音声つぶやきによる気づきの収集と活用で看護・介護サービスの質を向上する」, サービスロジー, Vol.1, No.2, pp.14–17 (2014)
- [33] H. Kanai: “Introduction on a support system for awareness in an aware-environment enhanced group home: Awarerium”, in 2012 Seventh International Conference on Knowledge, Information and Creativity Support Systems, pp.165–168 (2012-11)
- [34] P. Dourish, S. Bly: “Portholes: Supporting awareness in a distributed work group”, in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ser. CHI ’ 92, pp.541–547, New York, NY, USA (1992)

- [35] C. Gutwin, S. Greenberg: “A descriptive framework of workspace awareness for real-time groupware”, *Comput. Supported Coop. Work*, Vol.11, No.3, pp.411–446 (2002-11)
- [36] 國藤 進・加藤 直孝・門脇 千恵・敷田 幹文:「知的グループウェアによるナレッジマネジメント, 4章 アウェアネス支援を取り入れたグループウェア」, 日科技連出版社, pp.63–98 (2001)
- [37] 「シスコ テレプレゼンスソリューション」, <https://www.cisco.com/c/ja.jp/products/collaboration-endpoints/telePresence-integration-solutions/index.html>, accessed on April 18, 2020.
- [38] 森川 治:「超鏡:魅力あるビデオ対話方式をめざして」, 情報処理学会論文誌, Vol.41, No.3, pp.815–822 (2000-3)
- [39] K. Hirata, Y. Harada, T. Takada, N. Yamashita, S. Aoyagi, Y. Shirai, K. Kaji, J. Yamato, K. Nakazawa: “Basic design of video communication system enabling users to move around in shared space”, *IEICE Transactions on Electronics*, Vol.92, No.11, pp.1387–1395 (2009)
- [40] S. A. Bly, S. R. Harrison, S. Irwin: “Media spaces: Bringing people together in a video, audio, and computing environment”, *Commun. ACM*, Vol.36, No.1, pp.28–46 (1993-1)
- [41] P. Luff, H. Kuzuoka, C. Heath, K. Yamazaki, J. Yamashita: *Creating Assemblies in Media Space: Recent Developments in Enhancing Access to Workspaces*, Springer London, pp.27–55 (2009)
- [42] 馬場口 登:「安全と安心のための画像処理技術: 5. プライバシーを考慮した映像サーベイランス」, 情報処理, Vol.48, No.1, pp.30–36 (2007-1)
- [43] M. Boyle, C. Edwards, S. Greenberg: “The effects of filtered video on awareness and privacy”, in *Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work*, ser. CSCW ’00, pp.1–10, New York, NY, USA (2000)
- [44] R. W. Root: “Design of a multi-media vehicle for social browsing”, in *Proceedings of the 1988 ACM Conference on Computer-Supported Cooperative Work*, ser. CSCW ’88, pp.25–38, New York, NY, USA (1988)
- [45] R. S. Fish, R. E. Kraut, R. W. Root, R. E. Rice: “Video as a technology for informal communication”, *Commun. ACM*, Vol.36, No.1, pp.48–61 (1993-1)
- [46] J. C. Tang, E. A. Isaacs, M. Rua: “Supporting distributed groups with a montage of lightweight interactions”, in *Proceedings of the 1994 ACM Conference on*

- Computer Supported Cooperative Work, ser. CSCW ' 94, pp.23–34, New York, NY, USA (1994)
- [47] 本田 新九郎・富岡 展也・木村 尚亮・大澤 隆治・岡田 謙一・松下 温：「作業者の集中度に応じた在宅勤務環境の提供 – 仮想オフィスシステム valentine」, 情報処理学会論文誌, Vol.39, No.5, pp.1472–1483 (1998-5)
- [48] S. Sugawara, G. Suzuki, Y. Nagashima, M. Matsuura, H. Tanigawa, M. Moriuichi: “Interspace : Networked virtual world for visual communication”, *IEICE Transactions on Information and Systems, D*, Vol.77, No.12, pp.1344–1349 (1994)
- [49] 小林 稔・井原 雅行・島田 義弘・細田 真道・石橋 聡：「没入型仮想共有環境」, 電子情報通信学会誌, Vol.85, No.4, pp.259–264 (2002-4)
- [50] 廣瀬 通孝・小木 哲朗・玉川 憲・山田 俊郎：「没入型コミュニケーションのための高臨場感ビデオアバタ」, ヒューマンインタフェース学会論文誌, Vol.2, No.2, pp.55–62 (2000-5)
- [51] J. Leigh, A. E. Johnson: “Supporting transcontinental collaborative work in persistent virtual environments”, *IEEE Computer Graphics and Applications*, Vol.16, No.4, pp.47–51 (1996)
- [52] A. Voids, S. Voids, S. Greenberg, H. A. He: “Asymmetry in media spaces”, in Proceedings of the 2008 ACM Conference on Computer Supported Cooperative Work, ser. CSCW ' 08, p.313–322, New York, NY, USA (2008)
- [53] H. Kuzuoka, Y. Kodama, J. Xu, E. Myodo, E. Harada, H. Osawa: “Telepresence robot’s salutations to trigger informal conversation with teleworkers”, in Companion of the 2018 ACM Conference on Computer Supported Cooperative Work and Social Computing, ser. CSCW ' 18, p.233–236, New York, NY, USA (2018)
- [54] C. Gutwin, S. Greenberg: “Workspace awareness for groupware”, in Conference Companion on Human Factors in Computing Systems, ser. CHI ' 96, pp.208–209, New York, NY, USA (1996)
- [55] T. Erickson, D. N. Smith, W. A. Kellogg, M. Laff, J. T. Richards, E. Bradner: “Socially translucent systems: Social proxies, persistent conversation, and the design of “babble””, in Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ser. CHI ' 99, pp.72–79, New York, NY, USA (1999)
- [56] A. E. Milewski, T. M. Smith: “Providing presence cues to telephone users”, in Proceedings of the 2000 ACM Conference on Computer Supported Cooperative Work, ser. CSCW ' 00, pp.89–96, New York, NY, USA (2000)

- [57] N. Romero, G. McEwan, S. Greenberg: “A field study of community bar: (mis)-matches between theory and practice”, in Proceedings of the 2007 International ACM Conference on Supporting Group Work, ser. GROUP '07, pp.89–98, New York, NY, USA (2007)
- [58] “Skype”, <https://www.skype.com/>, accessed on April 18, 2020.
- [59] “Slack”, <https://slack.com/>, accessed on July 3, 2019.
- [60] A. Bulling, U. Blanke, B. Schiele: “A tutorial on human activity recognition using body-worn inertial sensors”, *ACM Comput. Surv.*, Vol.46, No.3, pp.33:1–33:33 (2014-1)
- [61] 大内 一成・土井 美和子:「携帯電話搭載センサによるリアルタイム生活行動認識システム」, 情報処理学会論文誌, Vol.53, No.7, pp.1675–1686 (2012-7)
- [62] K. Murao, T. Terada: “A combined-activity recognition method with accelerometers”, *Journal of Information Processing*, Vol.24, No.3, pp.512–521 (2016)
- [63] 江木 啓訓・尾澤 重知:「学習者センシングのための筆記行為の検知手法と評価」, 日本教育工学会論文誌, Vol.36, pp.181–184 (2012)
- [64] J. Fogarty, C. Au, S. E. Hudson: “Sensing from the basement: A feasibility study of unobtrusive and low-cost home activity recognition”, in Proceedings of the 19th Annual ACM Symposium on User Interface Software and Technology, ser. UIST '06, pp.91–100, New York, NY, USA (2006)
- [65] J. E. Froehlich, E. Larson, T. Campbell, C. Haggerty, J. Fogarty, S. N. Patel: “Hydrosense: Infrastructure-mediated single-point sensing of whole-home water activity”, in Proceedings of the 11th International Conference on Ubiquitous Computing, ser. UbiComp '09, pp.235–244, New York, NY, USA (2009)
- [66] 村上 知子・鳥居 健太郎・長 健太・内平 直志:「センサデータと業務知識からのトピックモデルを用いた看護業務行動の推定」, 人工知能学会論文誌, Vol.29, No.5, pp.427–435 (2014)
- [67] D. Riboni, T. Szytler, G. Civitarese, H. Stuckenschmidt: “Unsupervised recognition of interleaved activities of daily living through ontological and probabilistic reasoning”, in Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing, ser. UbiComp '16, pp.1–12, New York, NY, USA (2016)
- [68] H. Ishii, M. Kobayashi: “Clearboard: a seamless medium for shared drawing and conversation with eye contact”, in CHI '92: Proceedings of the SIGCHI conference

- on Human factors in computing systems, pp.525–532 (1992)
- [69] J. Li, M. Manavalan, S. D’ Angelo, D. Gergle: “Designing shared gaze awareness for remote collaboration”, in Proceedings of the 19th ACM Conference on Computer Supported Cooperative Work and Social Computing Companion, ser. CSCW ’16 Companion, pp.325–328, New York, NY, USA (2016)
- [70] S. D’ Angelo, D. Gergle: “An eye for design: Gaze visualizations for remote collaborative work”, in Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, ser. CHI ’18, pp.1–12, New York, NY, USA (2018)
- [71] R. Ishii, Y. I. Nakano, T. Nishida: “Gaze awareness in conversational agents: Estimating a user’s conversational engagement from eye gaze”, *ACM Trans. Interact. Intell. Syst.*, Vol.3, No.2, pp.11:1–11:25 (2013-8)
- [72] K. Okada, F. Maeda, Y. Ichikawaa, Y. Matsushita: “Multiparty videoconferencing at virtual social distance: Majic design”, in Proceedings of the 1994 ACM Conference on Computer Supported Cooperative Work, ser. CSCW ’94, pp.385–393, New York, NY, USA (1994)
- [73] A. J. Sellen: “Speech patterns in video-mediated conversations”, in CHI ’92: Proceedings of the SIGCHI conference on Human factors in computing systems, pp.49–59 (1992)
- [74] R. Vertegaal, I. Weevers, C. Sohn, C. Cheung: “Gaze-2: conveying eye contact in group video conferencing using eye-controlled camera direction”, in CHI ’03: Proceedings of the SIGCHI conference on Human factors in computing systems, pp.521–528 (2003)
- [75] S. Tachi, N. Kawakami, H. Nii, K. Watanabe, K. Minamizawa: “Telesarphone: Mutual telexistence master-slave communication system based on retroreflective projection technology”, *SICE Journal of Control, Measurement, and System Integration*, Vol.1, No.5, pp.335–344 (2008-9)
- [76] K. Ito, H. Kikuchi, H. Sakurai, I. Kobayashi, H. Yasunaga, H. Mori, K. Tokuyama, H. Ishikawa, K. Hayasaka, H. Yanagisawa: “360-degree autostereoscopic display”, in ACM SIGGRAPH 2010 Emerging Technologies, ser. SIGGRAPH ’10, p.1 (2010)
- [77] Y. Takaki, T. Dairiki: “72-directional display having VGA resolution for high-appearance image generation”, *Stereoscopic Displays and Virtual Reality Systems XIII*, Vol.6055, No.1, pp.307–314 (2006)
- [78] H. Kuzuoka, J. Kosaka, K. Yamazaki, Y. Suga, A. Yamazaki, P. Luff, C. Heath:

- “Mediating dual ecologies”, in CSCW '04: Proceedings of the 2004 ACM conference on Computer supported cooperative work, pp.477–486 (2004)
- [79] M. Otsuki, K. Maruyama, H. Kuzuoka, Y. SUZUKI: “Effects of enhanced gaze presentation on gaze leading in remote collaborative physical tasks”, in Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, pp.1–11, New York, NY, USA (2018)
- [80] N. P. Jouppi: “First steps towards mutually-immersive mobile telepresence”, in CSCW '02: Proceedings of the 2002 ACM conference on Computer supported cooperative work, pp.354–363 (2002)
- [81] 総務省：「平成30年通信利用動向調査報告書(企業編)」(2019)
- [82] S. Suyama, S. Ohtsuka, H. Takada, K. Uehira, S. Sakai: “Apparent 3-D image perceived from luminance-modulated two 2-D images displayed at different depths”, *Vision Research*, Vol.44, pp.785–793 (2004-4)
- [83] A. K. Jain, A. Topchy, M. H. C. Law, J. M. Buhmann: “Landscape of clustering algorithms”, in Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004., Vol.1, pp.260–263 (2004)
- [84] D. Xu, Y. Tian: “A comprehensive survey of clustering algorithms”, *Annals of Data Science*, Vol.2, pp.165–193 (2015)
- [85] “scikit-learn Machine Learning in Python”, <https://scikit-learn.org/>, accessed on May 10, 2019.
- [86] J. J. Lim, C. L. Zitnick, P. Dollár: “Sketch tokens: A learned mid-level representation for contour and object detection”, in 2013 IEEE Conference on Computer Vision and Pattern Recognition, pp.3158–3165 (2013-6)
- [87] M.-A. Storey, A. Zagalsky: “Disrupting developer productivity one bot at a time”, in Proceedings of the 2016 24th ACM SIGSOFT International Symposium on Foundations of Software Engineering, ser. FSE 2016, pp.928–931, NY, USA (2016)
- [88] K. Iso, M. Kobayashi, T. Yuizono: “A method for estimating worker states using a combination of ambient sensors for remote collaboration”, in Collaboration Technologies and Social Computing, pp.22–28 (2017)
- [89] K. Iso, M. Kobayashi, T. Yuizono: “A trial of ambient sensor method for worker states classifier”, in Proceedings of International Workshop on Informatics, pp.289–294 (2017)
- [90] 磯 和之・由井 蘭 隆也：「テレワーク環境におけるアンビエントセンサを用いた作業者

- の状態推定に関する実験」, 研究報告グループウェアとネットワークサービス (GN), No.19, pp.1 – 7 (2018-3)
- [91] “Hubot”, <https://hubot.github.com>, accessed on July 3, 2019.
- [92] 安藤 康子・伊達 宗和・高田 英明・磯 和之・松浦 宣彦:「距離画像カメラを用いたリアルタイム実写 DFD 表示」, 日本バーチャルリアリティ学会研究報告, サイバースペースと仮想都市研究会, Vol.15, No.CS-1, pp.23–26 (2010-3)
- [93] M. Date, Y. Andoh, H. Takada, Y. Otani, N. Matsuura: “Depth reproducibility of multiview depth-fused 3-D display”, in *Journal of the SID*, pp.470–475 (2010)
- [94] “SONY Xperia 1 II”, <https://www.sonymobile.co.jp/xperia/xperia1m2/>, accessed on May 1, 2020.
- [95] “Huawei P30 Pro”, <https://consumer.huawei.com/jp/phones/p30-pro/>, accessed on May 1, 2020.
- [96] “Samsung Galaxy Note 10+”, <https://www.galaxymobile.jp/galaxy-note10/camera/>, accessed on May 1, 2020.
- [97] “Apple iPad Pro”, <https://www.apple.com/jp/ipad-pro/>, accessed on May 1, 2020.
- [98] K. Iso, T. Yagi, M. Kobayashi, S. Iwaki, S. Ishibashi: “On natural living room communication with “ComAdapter”: Adapting to the differences in room structure”, in *CHI ’ 03 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA ’ 03, pp.716–717, New York, NY, USA (2003)
- [99] C.-F. Hsu, Y.-S. Wang, C.-L. Lei, K.-T. Chen: “Look at me! correcting eye gaze in live video communication”, *ACM Trans. Multimedia Comput. Commun. Appl.*, Vol.15, No.2, pp.38:1–38:21 (2019-6)
- [100] Y. Kim, H. Park, S. Bang, S. Lee: “Retargeting human-object interaction to virtual avatars”, *IEEE Transactions on Visualization and Computer Graphics*, Vol.22, No.11, pp.2405–2412 (2016-11)
- [101] 小沢 健悟・小川 剛史:「安全性を考慮した「歩き VR」システムにおける空間操作ツールの実装」, *VR 学研報*, Vol.24, No.CS-4, pp.11–12 (2020)
- [102] C. Wisneski, H. Ishii, A. Dahley, M. G. Gorbet, S. Brave, B. Ullmer, P. Yarin: “Ambient displays: Turning architectural space into an interface between people and digital information”, in *Proceedings of the First International Workshop on Cooperative Buildings, Integrating Information, Organization, and Architecture*, ser. CoBuild ’ 98, pp.22–32, Berlin, Heidelberg (1998)

- [103] 宮島 麻美・伊藤 良浩・伊東 昌子・渡邊 琢美：「つながり感通信：人間関係の維持・構築を目的としたコミュニケーション環境の設計と家族成員間における検証」, ヒューマンインタフェース学会論文誌, Vol.5, No.2, pp.171-180 (2003-5)

本論文に関する発表論文

学術論文

- [1] 磯 和之・由井 蘭 隆也：「テレワーク環境に適応したプレゼンス共有のための作業状態推定システムの協調的な構築手法」, 日本バーチャルリアリティ学会論文誌, Vol.25, No.1, pp.21–30 (2020)
- [2] K. Iso, T. Yuizono, M. Kobayashi: “Worker state estimation method with reduced manual task for teleworking environment”, *International Journal of Informatics Society*, Vol.11, No.3, pp.177–185 (2020)
- [3] K. Iso, S. Ozawa, M. Date, H. Takada, Y. Andoh, N. Matsuura: “Video conference 3d display that fuses images to replicate gaze direction”, *Journal of Display Technology*, Vol.8, No.9, pp.511–520 (2012)
- [4] 磯 和之・伊達 宗和・高田 英明・安藤 康子・松浦 宣彦：「視線の向きを表現可能な2画面積層表示を用いたテレビ会議システムの提案」, 情報処理学会論文誌, Vol.52, No.3, pp.1224–1233 (2011-3)
- [5] 磯 和之・八木 貴史・小林 稔・岩城 敏・石橋 聡：「生活融合通信：空間情報整合化機能 “ComAdapter”」, 日本バーチャルリアリティ学会論文誌, Vol.9, No.2, pp.169–178 (2004)

査読付き 国際会議

- [1] K. Iso, M. Kobayashi, T. Yuizono: “A trial of ambient sensor method for worker states classifier”, in Proceedings of International Workshop on Informatics, pp.289–294 (2017)
- [2] K. Iso, M. Kobayashi, T. Yuizono: “A method for estimating worker states using a combination of ambient sensors for remote collaboration”, in Collaboration Technologies and Social Computing, pp.22–28 (2017)
- [3] K. Iso, S. Ozawa, M. Date, H. Takada, Y. Andoh, and N. Matsuura: “Video Conference 3D Display That Fuses Images to Replicate Gaze Direction”, Proc. IEEE IAS2011, pp.1 – 6, Orlando, Florida, USA, October 9–13, (2011)
- [4] K. Iso, S. Ozawa, M. Date, H. Takada, Y. Andoh, and N. Matsuura “3D Video Conference system using fused images Display to Replicate Gaze Direction”, ECSCW 2011 extended abstracts, pp. 9–12, Aarhus, Denmark , September 24-28, (2011)
- [5] K. Iso, T. Yagi, M. Kobayashi, S. Iwaki, S. Ishibashi: “On natural living room communication with “ComAdapter”: Adapting to the differences in room structure”, in CHI ’ 03 Extended Abstracts on Human Factors in Computing Systems, ser. CHI EA ’ 03, pp.716–717, New York, NY, USA (2003)

国内 口頭発表

- [1] 磯 和之・由井 蘭 隆也:「テレワーク環境におけるアンビエントセンサを用いた作業者の状態推定に関する実験」, 研究報告グループウェアとネットワークサービス (GN), No.19, pp.1 – 7 (2018-3)