

遠隔会議におけるテレプレゼンスロボットのための  
非言語的な意思伝達に関する研究

長谷川 孔明

システム情報工学研究科  
筑波大学

2015年 3月

# 目次

第1章 序論	1
1.1 研究背景	1
1.2 関連研究	4
1.2.1 遠隔コミュニケーションを支援するロボット	4
1.2.2 コミュニケーションにおける身ぶり	4
1.2.3 身ぶりを表出するテレプレゼンスロボット	5
1.3 研究目的	7
第2章 発話衝突減少のための身ぶりの表出	8
2.1 意識的身ぶりと無意識的身ぶり	8
2.2 無意識的身ぶり取得方法	8
2.3 テレプレゼンスロボットの必要条件	9
2.4 実装	10
2.4.1 システム構成	10
2.4.2 テレプレゼンスロボット	10
2.4.3 モーションキャプチャ操作	13
2.4.4 受動型コントローラ操作	15
2.4.5 音声・動作の遅延について	16
2.5 実証実験	17
2.5.1 実験目的	17
2.5.2 比較条件	17
2.5.3 実験時の会話タスク	17
2.5.4 実験環境	19
2.5.5 実験の流れ	22
2.5.6 評価方法	24
2.6 実験結果と考察	25
2.6.1 実験結果	25
2.6.2 考察	29
2.6.3 結論	31
第3章 発話機会獲得のための身ぶりの表出	32
3.1 発話交替と参与役割	32
3.2 発話機会獲得のための身ぶり	32
3.3 身ぶりの誇張方法	34
3.4 実装	36
3.4.1 システム構成	36
3.4.2 モーションキャプチャデバイス	36
3.4.3 頭部方向の誇張	36
3.4.4 うなずき動作の誇張	37
3.4.5 相槌の誇張	37
3.4.6 音声・動作の遅延について	37
3.5 実証実験	38
3.5.1 実験目的	38

3.5.2	比較条件	38
3.5.3	実験時の会話タスク	38
3.5.4	実験環境	40
3.5.5	会話後の主観評価アンケート	45
3.5.6	実験の流れ	45
3.5.7	評価方法	47
3.6	実験結果と考察	48
3.6.1	実験結果	48
3.6.2	考察	54
第4章	結論	57
	謝辞	58
	参考文献	59
	著者公開論文リスト	62

## 図表目次

表 1-1 会議の機能と適した参加人数の関係 .....	2
図 2-1 無意識的身ぶり表出テレプレゼンスロボットのコンセプト .....	9
図 2-2 システム構成 .....	10
図 2-3 テレプレゼンスロボット外観 .....	11
表 2-1 KRS-2552HV 仕様 .....	11
図 2-4 テレプレゼンスロボットの機構 .....	12
図 2-5 Kinect 外観 .....	13
図 2-6 Kinect の骨格認識で取得可能な関節点 .....	14
図 2-7 faceAPI の動作時の様子 .....	15
図 2-8 受動型コントローラ外観 .....	16
表 2-2 砂漠生き残り問題の道具リスト .....	18
図 2-9 実験環境 .....	20
図 2-10 会議室側の様子 .....	21
図 2-11 遠隔地側の様子 .....	21
図 2-12 実験の流れ .....	23
表 2-3 セッション中の総ターン数の平均 .....	26
図 2-13 ロボットにより表出された意識的身ぶりの回数 .....	26
図 2-14 ロボットにより表出された無意識的身ぶりの回数 .....	27
図 2-15 遠隔参加者のターン取得率 .....	27
図 2-16 発話衝突回数 .....	28
図 3-1 参与役割の状態遷移図 .....	33
表 3-1 頭部方向の誇張 .....	35
表 3-2 うなずき動作の誇張 .....	35
表 3-3 相槌の誇張 .....	35
図 3-2 システム構成 .....	36
表 3-4 砂漠生き残り問題の道具リスト .....	39
図 3-3 実験環境(誇張あり条件, 誇張なし条件) .....	41
図 3-4 会議室の様子(誇張あり条件, 誇張なし条件) .....	42
図 3-5 遠隔地の様子 .....	42
図 3-6 会議室の様子(ビデオチャット条件) .....	43
図 3-7 実験環境(ビデオチャット条件) .....	44
表 3-5 会議室側の参加者に遠隔参加者の印象を評価してもらうための質問項目 .....	45
表 3-6 遠隔参加者に会議室側の参加者の印象を評価してもらうための質問項目 .....	45
図 3-8 実験の流れ .....	46
図 3-9 遠隔参加者のターン取得率 .....	48
図 3-10 総ターン数の平均 .....	49
図 3-11 遠隔参加者のターン取得数 .....	49
図 3-12 「遠隔参加者があなたや話をしている人を見ていると感じた」の質問項目の結果 .....	51
図 3-13 「遠隔参加者が話に対する反応を返してくれていると感じた」の質問項目の結果 .....	51

果 .....	51
図 3-14 「遠隔参加者が話したそうにしている様子に気付いた」の質問項目の結果...	52
図 3-15 「遠隔参加者の動きに気づきやすかった」の質問項目の結果 .....	52
図 3-16 「他の参加者が話したそうにしている様子に気付いた」の質問項目の結果...	53

# 第1章 序論

## 1.1 研究背景

会議について、高橋は「会議とは、三人以上の参加者が共通の場で対面し、情報の伝達や加工を行って問題の解決をはかること」と定義している[37]。さらに、共通の場で対面してコミュニケーションを行うことを「対面コミュニケーション」と称し、「対面コミュニケーション」のメリットとして以下の5つをあげている。

1. 大勢に同時に情報伝達できる
2. 相互の情報交換が可能である
3. 伝達情報に対する反応がわかる
4. 参加者同士の一体感が高まる
5. 参加者の知識向上に役立つ

会議はさまざまな問題を解決するために一般的に行われているものであり、会議の際には「対面コミュニケーション」を行えるということが重要となっている。

また、同じく高橋は会議を機能別に4種類、参加人数により3種類に分類している[37]。まず、機能別の分類として、伝達会議、調整会議、創造会議、決定会議の4種類がある。伝達会議とは、情報の伝達とその確認を目指す会議のことを指す。調整会議とは、組織間や大きな組織の中での部門間で様々な調整を行う必要があるため、目標に対して部門ごとの行動が適切であるか、また部門ごとに重複行動など無駄がないかなどをチェックするものである。創造会議は、問題解決会議とも呼ばれるもので、さまざまな課題に対して、問題点を分析したり、その解決策を考えたりするための会議である問題解決のために行うため、この創造会議が最も大切な会議の形式と言える。最後に、決定会議とは、企業などの組織の行動を決定するために意思決定をするための会議である。日本では、会議の前にすべての根まわしがすんでいて、決定会議といってもただ議決をとるだけの形式的なものになっている場合が多い。しかしながら、本来は討論の場であり、各自が自分の知恵を闘わせて討議しあう場である。また、これらの機能別の会議は、一度の会議に必ずしも1種類であるわけではなく、一度の会議に複数の種類の会議の機能が含まれる場合もある。

また、会議の参加人数の分類については、小集団会議、中集団会議、大集団会議の3種類がある。小集団会議とは、人数としては3人以上10人以下の会議である。小集団活動やしっかりした討議が行えるため、創造会議や決定会議に適した規模の会議である。中集団会議は、10人から30人程度の会議である。部長会や職場会といった各種委員会の会議などを行う規模の会議であり、伝達会議や調整会議に適した人数である。最後に、大集団会議とは、30人以上の大人数の会議である。議長団と議事運営が事前にしっかり決まっている様な会議であり、組合大会、社員総会、株主総会などの伝達会議を行うのが中心の会議の規模である。

会議の機能別の分類とそれらに適した参加人数の分類の関係を表 1-1 に示す。このように、会議の機能別、参加人数別の分類を見ると、伝達会議や調整会議は多くの人に情報の伝達や情報の交換を行うことが主である。そして、議長や司会などの役職が決まっていることが一般的である。一方で、創造会議や決定会議は情報伝達や情報交換だけでなく、議論を行う会議であり、司会などの役職を定めずに各自が積極的に討議に参加するという特徴がある。そのため、参加者同士でのやりとりも多くなるので前述した「対面コミュニケーション」の

メリットの恩恵がより大きい会議の種類であるといえる。

また、このような「対面コミュニケーション」のメリットは、言葉のやりとりだけでなく人同士が直接会ってコミュニケーションを行う際に言葉以外の情報もやりとりしているために現れるものである。人同士が行うコミュニケーションは、バーバルコミュニケーションとノンバーバルコミュニケーションに分類される[42]。バーバルコミュニケーションは言語を用いたものである。一方、ノンバーバルコミュニケーションは周辺言語や視線、表情、対人距離、身ぶりといった言語以外の手段を用いるコミュニケーションである。さらにバーバルコミュニケーションよりもノンバーバルコミュニケーションによってより多くの情報が伝達されると言われている[18]。そのため遠隔会話でも対面時のようにノンバーバルコミュニケーションを実現することが重要となる。

表 1-1 会議の機能と適した参加人数の関係

	会議の機能	小集団会議 (3～10人)	中集団会議 (10～30人)	大集団会議 (30人以上)
伝達会議	情報伝達とその確認		○	○
調整会議	組織間の情報調整		○	
創造会議	問題点の分析、解決策の立案	○		
決定会議	組織行動の意思決定	○		

一方で、通信技術の発達により、遠隔地にいる人と会話を行う手段としてビデオチャットが普及している。さらに、現在では1対1だけではなく複数人での同時利用も可能となっており、会議を遠隔地間で行う遠隔会議も可能となり広く利用されている。しかしながら、ビデオチャットや遠隔会議といった映像と音声のみを用いた遠隔会話の場面では発話交替がうまくいかずに発話衝突という問題が生じる。発話衝突とは二人以上の人が同時に発話している状態である。遠隔会話では音声遅延が300msを超えると特に発話衝突の発生が顕著になると報告されている[45]。また、玉木らによると遠隔会議の場面では対面会議と比較して発話衝突が30倍近く起こると報告している[38]。発話衝突が起こると発話をあきらめて中断する傾向が高く、沈黙が発生し消極的な会話になりかねない。また、有意義な意見が発話衝突により妨げられる可能性もある。さらに発話衝突が頻発すれば会話の中断により会議時間が無駄に長くなるという問題にも繋がる。

発話衝突の原因として発話予備動作が認知されにくくなるという点があげられる[38]。対面での会話では、参加者の発話予備動作を読み取ることにより誰がいつ発話し始めるのかを判断している。これにより発話衝突を回避し、円滑な発話交替を実現している。しかしながら、遠隔会議の場面では発話予備動作が認知されにくいことにより発話衝突が起こりやすいと示唆されている。

一方で、実世界のコミュニケーションにおいて映像ではなく実体を持つロボットを用いることの優位性を示唆する研究がある。Kiddらは実在のロボットと映像上のバーチャルエージェントを比較する実験を行い、実世界でのコミュニケーションでは実在のロボットのほうが適していることを報告している[15]。また、Shinozawaらは実在するロボットと映像上のエージェントからユーザに提案を行う状況において、実在するロボットのほうが適切であることを示している[30]。

さらに、ロボットを介して非言語的な情報を伝達し遠隔コミュニケーションを支援する研究が多数行われている[23][27][32]。このように、あたかもその人の分身であるかのように遠隔地に存在感を表出するロボットのことをテレプレゼンスロボットまたはレイグジスタンスロボットと呼ぶ[34][35]。これらのことから、映像では認知されにくくなる発話予備動

作を 3 次元的な実体を持つロボットを用いて伝達することにより，発話衝突を回避することが出来ると考えられる。

また，発話予備動作の具体的例として「腕組みをほどく」「身を前に乗り出す」「話者の方へ向きを変える」などがある[42]。これらは人が無意識的に行っている動作であると考えられ，発話予備動作を伝達するには，人が会話中に相手に見せようと意識的に行う動作だけではなく，このような無意識的な動作こそテレプレゼンスロボットにより伝達する必要があると考えられる。



## 1.2 関連研究

### 1.2.1 遠隔コミュニケーションを支援するロボット

既存の遠隔会話を支援するロボットの研究は、ロボットを用いてノンバーバルコミュニケーションを実現することを目的としている。

既存のテレプレゼンスロボットとして、Paulosらはロボットに搭載したディスプレイに操作者の顔を表示させる P<sub>RoP</sub> と呼ばれるロボットを開発し、テレプレゼンスについて報告している[24]。さらに、Anybots社のQBもP<sub>RoP</sub>と同様に遠隔操作で移動が可能なロボットにディスプレイを搭載したテレプレゼンスロボットである[2]。そして、InTouchTechnologies社のRP-7も同様な構成のテレプレゼンスロボットである[8]。これらのロボットはノンバーバル情報である表情と対人距離の表出が可能である。しかしながら、発話予備動作に関わる頭部、体幹、腕部の動きが表現できないと考えられる。

また、鈴木らは遠隔会議支援ロボットシステムを開発している[33]。これは卓上サイズの移動ロボットにパン・チルトカメラを搭載した構成となっている。カメラのパン・チルト動作と移動を組み合わせることで注意喚起能力を高められることを報告している。このロボットはノンバーバル情報である視線方向と対人距離の表出が可能である。そのため、発話予備動作に関連する頭部動作は表現できるが、体幹、腕部の動きが表現できないと考えられる。また、遠隔操作者の顔を表出するディスプレイ等も搭載していないため、会話時に重要となる表情を伝達することが不可能である。

大塚らはMM-Spaceと呼ばれる、間・空間を隔てた会話の状況を実世界に再構成するためシステムを開発している[22]。これは、透過型スクリーンに遠隔参加者の等身大の人物像が投影される構成となっており、さらにそのスクリーンはアクチュエータが接続されている。このアクチュエータにより、遠隔参加者の顔のパン・チルトの動きに合わせて人物を投影したスクリーンをパン・チルトすることが可能となっている。スクリーン自体がパン・チルトすることにより遠隔参加者の視線方向をより分かりやすくすることを目指したシステムである。しかしながら、スクリーンに投影されるは胸より上部分のみとなっているため、腕部の動きが表現できないと考えられる。また、スクリーンを大きくしたとしても、表出される腕の動きについてはスクリーン中の映像上の動きに限られる。

これら既存のテレプレゼンスロボットに関する研究や、製品化されたテレプレゼンスロボットは、ロボットの身体性のもつ存在感や対人距離、視線方向に着目したものが多い。しかし、人同士が対面してコミュニケーションを行う場面において伝達される非言語的な情報はこれだけに限らない。例えば、人は対面で会話する際、身振りを用いて相手とコミュニケーションを行うのが普通である。しかしながら、ここにあげたテレプレゼンスロボットは腕の機構を持たないため身振りを表出することは出来ない。

### 1.2.2 コミュニケーションにおける身ぶり

一口に身振りといっても人がコミュニケーションの際に用いる身振りの種類は様々である。

エックマンによると、身振りは表象、例示的動作、情動表出、調整子、身体操作に分類される[44]。表象とはVサインやサムアップのようなメッセージを意図的に伝達するために使われる身振りである。例示的動作は発話内容を協調したり補足したりする動作であり、言葉と一緒に聞かなければその意味は明確には分からない。情動表出は情緒的な状態や反応を示す身振りのことである。調整子は会話の進行を円滑にするための動作である。身体操作は頭をかいたり唇をなめたりといった、身体のある部分を使ってほかのものに働きかける動作のことである。

また、喜多は身振りを機能によって、伝達内容の表現、コミュニケーションのメタ調節、情動的きずなづくりに分類している[26]。伝達内容の表現には、VサインやOKサインといった形と意味が社会的な取り決めによって恣意的に決められている「エンブレム」、形と意味が固定しておらず動きと指示対象の形の類似性に基づいている「描画的ジェスチャー」、指差しのような時空間的隣接性によって指示対象を示す「直示的ジェスチャー」がある。コミュニケーションのメタ調節には、会話の流れを調節する機能が含まれる。これは、話す順番（ターン）を相手に渡すつもりがあるかないかを示す、ターンの保持、ターンをとりたいという意思表示を身振りによって行うものである。情動的きずなづくりは、例えば話し手と聞き手が同時にうなづくといったようにコミュニケーションの参加者が同時に同じ身体の動きをすることがある。これにより会話の参加者間に情動的なきずながつくられるというものである。

これらのことからコミュニケーションにおける身振りには、伝達内容の表現、話者交替の意思表示、親近感の形成といった機能があることが分かる。

また、Cassellらは、人の対面コミュニケーションの場面において、話し手が意識的に身振りを表出している場面でなくとも、聞き手は常に身振りから情報を得ていることを示唆している[3]。よって、意識的に表出された身振りだけでなく無意識的に表出された身振りも聞き手に影響を与えており、コミュニケーション中に何らかの役割を果たしていると考えられる。

### 1.2.3 身ぶりを表出するテレプレゼンスロボット

井上らは遠隔地間のコミュニケーション用ロボットシステムとしてテレコミュニケーターを開発している[7]。テレコミュニケーターの一つであるウェアラブルミニチュアヒューマノイドロボットは、人の肩に乗る大きさの小型ヒューマノイドロボットである。腕部各4自由度、頭部3自由度を有しており腕部と頭部の動きを表現可能である。しかしながら、この自由度の頭部と腕部をもつモデルにおいては、ディスプレイを搭載したものが開発されておらず、表情を伝達することは不可能である。

石黒らは、外見やしぐさが人間と極めて近いアンドロイドロボットである ReplieeQ2[9]や、ある特定の人間の分身ともいえる Geminoid[16]を用い、遠隔地での対話における存在感の伝達効果を研究している。これらのアンドロイドロボットは、口の動きや頷き、呼吸時の胸部の動きなど細かい動作により人間らしい表現を行っている。しかしながら、石黒らはアンドロイドの外見やしぐさによる人間らしさは評価しているものの、会話中の動作が発話交替にもたらす影響については特に言及していない。さらに、腕を上げたり腰を捻ったりといった大きな動作は行うことができない点や、仮想的な人物あるいは特定の人物に外見を似せているために操作者とアンドロイドの外見の違いから違和感を感じてしまうと考えられる点から、遠隔会議において発話予備動作を表出するデバイスとしては不向きであると考えられる。

また、同じく石黒らは、科学館という不特定多数の人が来場する環境で Robovie-M という高さ 29cm の小型ヒューマノイドロボットを用いた会話実験を行っている[21][29]。Robovie-M は全身で 22 自由度を有し、身振りを行いながら会話が可能である。多人数会話である点は遠隔会議と同一であるが、科学館の来場者が対話相手であるため頻繁に入れ替わりが起こる可能性があり、遠隔会議の環境とは違いがある。また、ロボットとの対人距離や印象についての評価はしているものの、発話交替や発話衝突については言及されていない。

吉崎らはモーションキャプチャを利用して直感的な操縦を行うことができるテレプレゼンスロボットを提案している[46]。モーションキャプチャのデバイスとして Kinect を利用しており、得られたデータから V-Sido と呼ばれる姿勢制御システムを用いることにより転倒しない補正をかけながらロボットに動作を行わせることが可能となっている。実装されたテレプ

レゼンスロボットは頭部にディスプレイと脚部各 6 自由度, 腕部各 3 自由度, 胴部 1 自由度, 頭部 1 自由度を有している。これにより, 表情, 体幹, 腕部の動きを表出可能である。しかしながら, 頭部は 1 自由度しか有していないため, うなずきや首をかしげるといった発話予備動作を伝達することは不可能である。

Tojo らは, 多人数会話に参加するロボットを開発している[40]。この会話ロボットは頭部の方向や体幹方向, 身ぶりといった非言語情報を出して会話に参加することが可能である。しかしながら, これは自律型のロボットであり遠隔会話に用いるシステムではないものとなっている。

既存の身振りを表出するテレプレゼンスロボットの研究として, 葛岡らの GestureMan-3.5 があげられる[41]。GestureMan-3.5 は遠隔操作型のロボットで前後移動と回転が可能であり, 頭部には操作者の志向を表現するための頭と 3 眼カメラを搭載している。また, GestureMan-3.5 の右側面には 5 自由度のロボットアームを搭載しており, 先端には指差しをしている状態の手の模型が取り付けられている。これにより, 対象物を指し示すことが可能である。しかし, GestureMan-3.5 で扱う主な身振りは指さしであり, これは伝達内容の表現を行う身振りの 1 つである「直示的ジェスチャー」にあたる。そのため, 扱っている身振りは話者交替の意思表示の身振りとは異なるものである。

また, 館らが開発した TELESAR II は, 人と同じサイズの人型マスタ・スレーブロボットである[36]。操作者は外骨格型のマスタコクピットを用いて人型スレーブロボットを操作することが可能であり, あたかも遠隔の環境にいるような感覚で精密な作業を行うテレグジスタンスを目指している。ロボットのスレーブアームは, 方直交 3 自由度, 肘 1 自由度, 手首直交 3 自由度, 手 8 自由度 5 本指という自由度を持ち繊細な作業を行うことが出来る。TELESAR II は遠隔環境での精密な動作の実現を目指しているが, その反面, 操作のためのシステムも大規模となり, 少人数の参加する会議や会話の場面での運用は現実的ではない。

Adalgeirsson らは MeBot と呼ばれる卓上サイズの移動ロボットを開発している[1]。MeBot は顔の映像を表示するパン・チルトと前後移動が可能なディスプレイを頭部として取り付けている。また, 各 3 自由度の腕の機構を有している。このロボットはノンバーバル情報である表情, 視線方向, 対人距離, 身ぶりの表出が可能である。そのため, 発話予備動作に関する頭部, 体幹, 腕部の全動作が表現できる。MeBot は発話予備動作を表出するモダリティとしては十分だと考えられる。しかしながら, MeBot の身ぶりの操作方法は MeBot の腕と同じ機構のコントローラで操作するという方法である。この方法では発話予備動作を操作に反映できないと考えられる。我々がこのように考える理由については次章で詳しく説明する。また, MeBot を評価する実験では, 1 対 1 の会話タスクにおいてロボットが動く場合と動かない場合についての印象の違いを主に評価している。そのため発話衝突や発話交替, 発話予備動作については一切評価していない。そのため MeBot の操作方法によって表出される身ぶりが発話衝突の減少に有効であるかどうかは明らかになっていない。

以上に述べたように, 表情を表出するディスプレイに加え, うなずきや身を乗り出すといった発話予備動作を表出するための頭部, 腕部, 体幹に十分な自由度を有すと共に無意識的身ぶりを伝達できる操作方法のテレプレゼンスロボットは見受けられない。また, 発話衝突の減少や発話交替に関してテレプレゼンスロボットにより表出される身ぶりの有用性は未だ確認されていない。

### 1.3 研究目的

遠隔コミュニケーションでは身振りが認知されにくくなるため、発話の予備動作などの話者交替の意思表示の情報が十分に伝達されなくなる。そのため話者交替が適切に行われず、発話衝突が起こるといった問題が発生する。そこで、テレプレゼンスロボットを介して身振りを表出することにより、身振りが認知されやすくなり話者交替が円滑に行えるようになることを考える。

一方、人同士のコミュニケーションでは意識的なものだけでなく無意識的に表出された身振りも情報を伝えているといわれている。そのため、テレプレゼンスロボットにおいても意識的な身振りのみを表出するものでは身振り情報を十分に伝達できていないと考えられる。よって、無意識的な身振りも含めて表出することが可能な操作方法が必要であると考えられる。

本研究では無意識的身ぶりを伝達して表出可能なテレプレゼンスロボットを提案する。このテレプレゼンスロボットは発話予備動作を表出する方法として身ぶりの操作方法に着目して設計を行う。また、無意識的身ぶりの伝達が発話衝突回避に有効であることを確かめるための実験を行う。

更に、本研究では遠隔多人数会話において発話機会を獲得するためのテレプレゼンスロボットを提案する。

提案システムでは、遠隔参加者の身ぶりを検出し、他の参加者が話し手となっている際にテレプレゼンスロボットが身ぶりを誇張表現することによって、遠隔参加者が聞き手の役割を得やすくなるよう支援を行う。身ぶりの誇張を行うことにより、遠隔参加者が聞き手になる機会と聞き手になってから次話者として発話を行う機会を増加させることができるものと考えられる。

## 第2章 発話衝突減少のための身ぶりの表出

### 2.1 意識的身ぶりと無意識的身ぶり

本研究で扱う「意識的身ぶり」と「無意識的身ぶり」についての定義を行う。まず「意識的身ぶり」は、相手に見せるという目的で表出した身ぶりとして定義する。例えば、「あそこの」と言いながら指さしをするといったものや「ボールがこう飛んで来て」と言いながら手を握って拳をつくることによりボールを表現し、ボールの飛ぶ起動を手の動きで表すといったものである。次に「無意識的身ぶり」は、相手に見せるという目的は無く表出した身ぶりとして定義する。例えば、頭をかく、口元に手をやる、身を乗り出すと言ったものである。前項で紹介した MeBot の身ぶりの操作方法は、ロボットと同じ機構のコントローラを動かすことによって行う。そのため、操作者が意識的に表出しようとした身ぶりのみが操作として現れ、ロボットの身ぶりとして表出されると考えられる。そして、この操作方法では人の癖などの無意識的な身ぶりは表出されないと考えられる。しかし、玉木らが挙げている発話予備動作の例には、「身を前に乗り出す」「手を口および顔周辺へ持っていき、もしくはそこから下ろす動作」といった人が無意識的に行っている動作が多く見受けられる。さらに、Cassell らは、人は対面コミュニケーションの場面において、話し手が意識的に身ぶりを表出している場面でなくとも、聞き手は常に身ぶりから情報を得ていることを示唆している[3]。よって、人が無意識的に行っている身ぶりも相手に影響を与えておりコミュニケーションにおいて何かしらの役割を果たしていると考えられる。これらのことから、無意識的身ぶりには発話予備動作となるものが含まれていると推察できる。そして、無意識的身ぶりを表出することにより発話衝突の回避を行うことが期待できる。そこで本研究では、テレプレゼンスロボットに無意識的身ぶりを表出させることにより、発話衝突を回避するシステムを実現する。

### 2.2 無意識的身ぶり取得方法

無意識的身ぶりを表出するためには、まずその身ぶり自体を取得する方法が必要となる。身ぶりを意識的か無意識的かに分別し、無意識的身ぶりのみを抽出するような方法は困難である。しかしながら、無意識的身ぶりも何かしらの体の動作である。そのため操作者のすべての動作を取得すれば、意識的か無意識的かの分別は出来ないが、無意識的身ぶりを含んだ動きをロボットの操作に反映することが可能である。

無意識的身振りを取得するための操作方法の必要条件は下記である。

1. 操作者の身体動作を取得できること
2. 操作者を拘束しないこと

まず、1.について、無意識的身振りは意識的ではないにしろ操作者の身体動作として表れるため、操作者の身体動作すべてを取得すれば、意識的身振りと無意識的身振りの両方が取得できると考えられる。2.については、操作者の体に何らかの装置を取り付けるなどして拘束を行うと、普段の対面でのコミュニケーションでは表出される無意識的身振りが抑制される可能性があるためである。

このような操作方法を実現させるための手段として、モーションキャプチャの技術を利用することがあげられる。モーションキャプチャの手法は様々なものがあるが、身体に装着する必要のある外骨格型のものではなく、赤外線カメラを用いる光学的な手法を用いれば操作者の拘束を最小限にして身振りを取得することが可能である。

提案システムのコンセプトを図 2-1 に示す。既存の遠隔会話支援ロボットでは、遠隔参加者が受動型コントローラを操作することによりロボットの身ぶりを表出していた。しかし、この手法では遠隔参加者が意識的に伝えようとした身ぶりしか操作に反映されないと考えられる。そのため、発話予備動作のような無意識的に行っている場合が多い動作についてはロボットを介して十分に伝達することが出来ない。そして、発話予備動作が十分に伝わらないために発話衝突が発生しやすくなると考えられる。一方、本研究で提案する手法は遠隔参加者の動きをモーションキャプチャで取得し、それをロボットの身ぶりの表出に反映させるというものである。遠隔参加者の無意識的な身ぶりも含めたすべて動きを操作に反映させることができると考えられるため、無意識的に表出される発話予備動作を、ロボットを介して他の参加者に伝えることができると考えられる。そのため、伝達した発話予備動作から発話を予期することができ発話衝突が減少することが期待される。

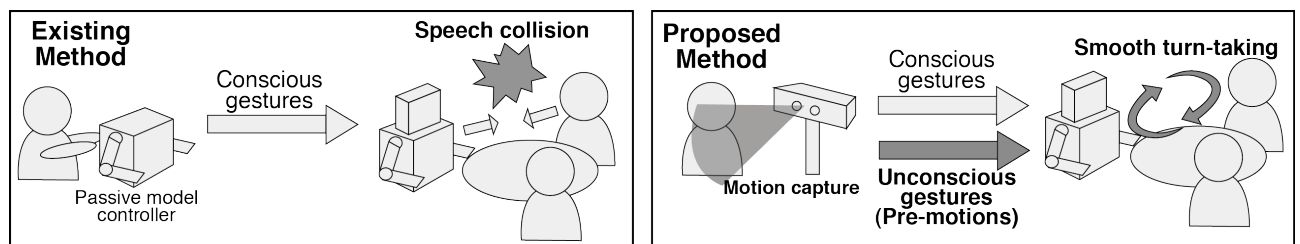


図 2-1 無意識的身ぶり表出テレプレゼンスロボットのコンセプト

## 2.3 テレプレゼンスロボットの必要条件

提案するテレプレゼンスロボットの必要条件は下記の通りである。

1. 人型であること
2. 小型で卓上に置ける大きさであること
3. 腕の機構を持つこと
4. 身の乗り出しと左右の捻りが行えること
5. 頭部が 3 自由度で動作を行えること
6. 顔の映像を表示できること
7. 音声を伝達できること

まず、1.について、テレプレゼンスロボットは人として認識されやすい形状であるほど存在感が増すため人型である必要がある。2.について、コミュニケーションの場の1つである会議を例にとると、参加人数により会議の役割が変化するといわれている[37]。その中でも 3~10 人ほどの人数でブレインストーミングや問題解決策の議論などを行う創造会議は、話者交替も頻繁に起こり参加者の表情や身振りが重要となる。そのため、少人数で行われる会議はロボットを用いる必要性が高いと考えられ、テレプレゼンスロボットを用いる会議の規模としては 3~10 人程度が参加する少人数のものを想定する。そして、少人数の会議であればロボットの小ささによる存在感の低下が起こりにくく、卓上に置けるほどの大きさである方が取り回しがしやすいと考えられる。そのため、卓上における大きさである必要がある。3.4.5.は全て身振りを表出するために必要である。また、5.の頭部動作に関しては、視線方向を表出するためにも必要となる。6.7.は既存の Web 会議でも用いられている要素であるためテレプレゼンスロボットの場合にも必要である。

## 2.4 実装

### 2.4.1 システム構成

実装したテレプレゼンスロボットシステムの構成を図 2-2 に示す。遠隔側では、遠隔参加者の頭部、体幹、腕部の動作をモーションキャプチャで取得し、遠隔参加者の表情と音声は Web カメラとマイクにより取得する。遠隔参加者側で取得した頭部、体幹、腕部の動作、表情、音声の情報は制御 PC によりインターネットを介し隣在側の制御 PC へと送られる。この通信の実装は SkypeAPI を用いて行った。隣在側では受け取った情報をもとにテレプレゼンスロボットにより遠隔参加者の身ぶりや表情を隣在参加者へと提示する。また、音声は制御 PC のスピーカを用いて提示する。隣在側から遠隔側への情報伝達は既存の Web 会議と同じ構成である。隣在側では、テレプレゼンスロボットの背後に設置した Web カメラとマイクからそれぞれ隣在参加者の映像と音声を取得する。映像と音声は Skype を用いて行い、制御 PC によりインターネットを介して遠隔側の制御 PC へと送られる。遠隔側では、映像はディスプレイを、音声はスピーカを用いて操作者へと提示される。

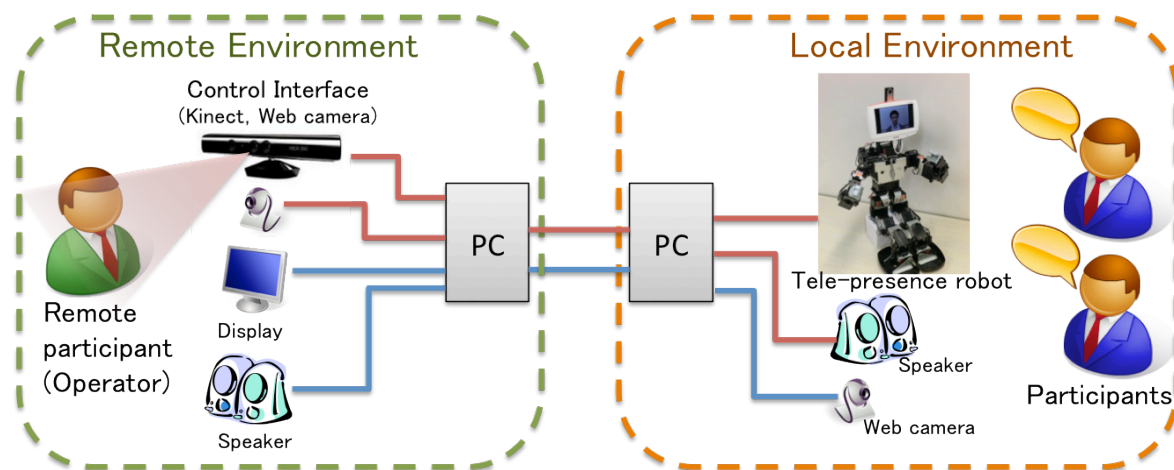


図 2-2 システム構成

### 2.4.2 テレプレゼンスロボット

実装したテレプレゼンスロボットの外観を図 2-3 に示す。ロボットのプラットフォームとして近藤科学株式会社製ヒューノイドロボット KHR-3HV を用い、本研究でのテレプレゼンスロボットの仕様を満たすように小型ディスプレイとサーボモータの追加を行った。このテレプレゼンスロボットに用いられているサーボモータは全て、近藤科学株式会社の KRS-2552HV である。KRS-2552HV の仕様をに表 2-1 示す。また、サーボモータの追加を行った後のテレプレゼンスロボットの機構を図 2-4 に示す。身ぶりの表現度を高めるため頭部 2 自由度、腕部各 1 自由度のサーボモータの追加を行い、頭部 3 自由度、腕部各 4 自由度、腰部の左右の捻り 1 自由度、身の乗り出し仰げ反り 1 自由度を用いた身ぶりの表現が可能である。これにより遠隔参加者の頭部と体幹、腕部動作を隣在参加者へと表出する。本研究では着座状態で会議を行うことを想定し、脚部の動きは表現しないものとした。そのため、図 2-4 のうち  $S_{12}$ ,  $S_{13}$ ,  $S_{16}$ ,  $S_{17}$ ,  $S_{18}$ ,  $S_{19}$  の各サーボモータは本システムでは稼働させない。ただし、遠隔参加者の代理のロボットとして人型である必要があるため、足であることが認識できるように稼働させない脚部のサーボモータも取り付けただままにした。

遠隔参加者の表情を表出については、テレプレゼンスロボットの頭部に取り付けた小型ディスプレイを用いる。このディスプレイに遠隔参加者の顔の映像を映し出すことにより、表情の表出を行う。本研究では、テレプレゼンスロボットの顔として違和感ができるだけ少な

くなるよう大きさを考慮し、株式会社磁気研究所の 4.3 インチ小型 USB ディスプレイ Mimo um-430 を用いた。

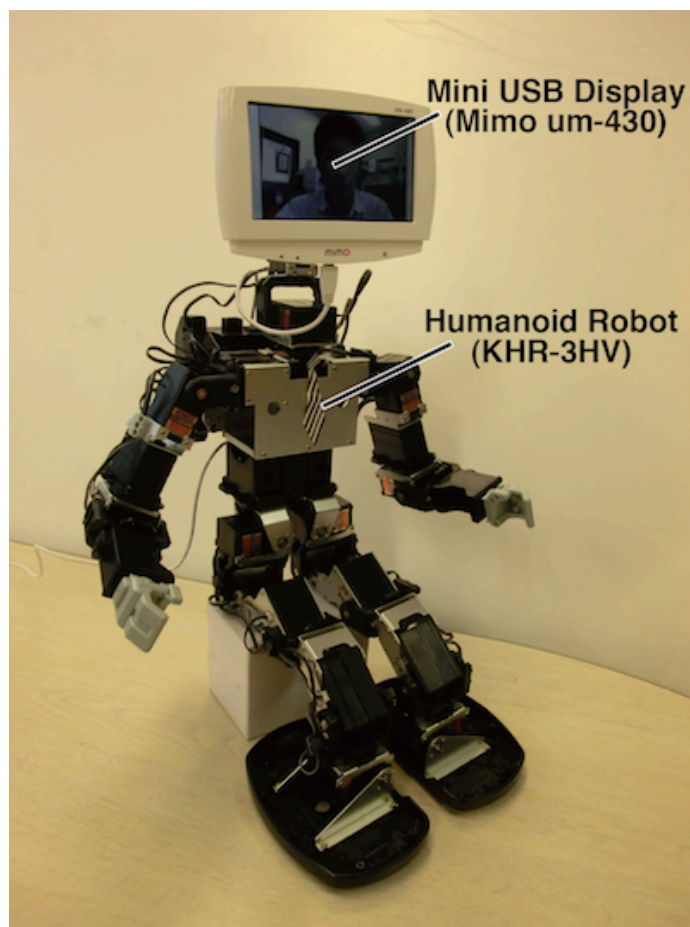


図 2-3 テレプレゼンスロボット外観

表 2-1 KRS-2552HV 仕様

寸法	L41[mm]×W21[mm]×H30.5[mm] (突起部除く)
定格電圧	9~12[V]
重量	41.5[g]
最大動作角度	270[° ] (±135[° ])
最大連続消費電流	訳 1.3[A] (11,1[V]時)
最大トルク	14.0[kg・m]
最大スピード	0.14[sec/60° ]
ギヤ比	31240:1



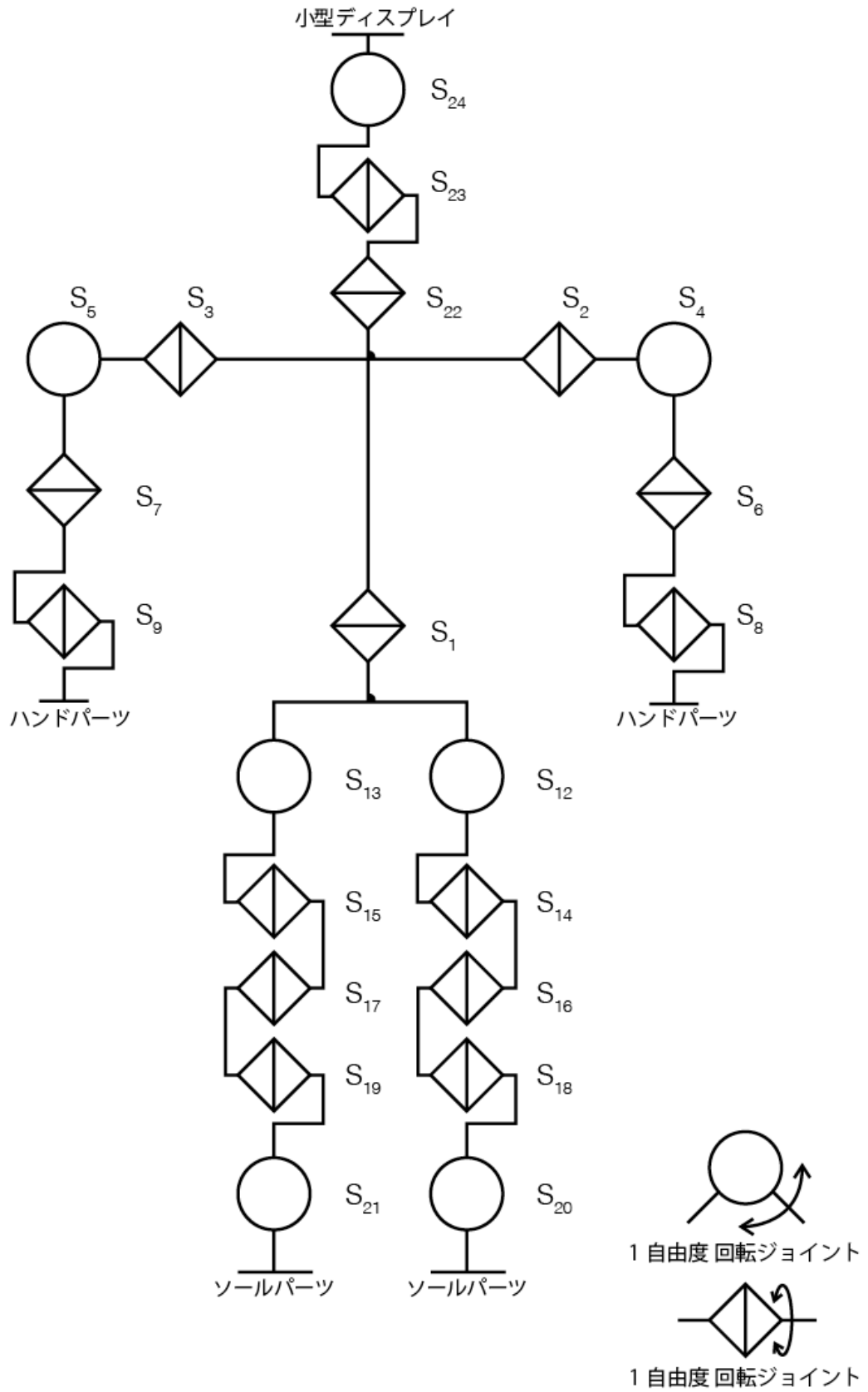


図 2-4 テレプレゼンスロボットの機構

### 2.4.3 モーションキャプチャ操作

体幹と腕部の動作を取得するためのモーションキャプチャデバイスとして Microsoft 社製の Kinect を用いた[19]. Kinect の外観を図 2-1 に示す. Kinect は同社のソフトウェア開発キット KinectforWindowsSDK を用いることにより深度センサーからの情報をもとに人物のボーンを認識し, 体の動きを各関節や手先位置などの 3 次元情報として取得することが可能である(図 2-6). Kinect はモーションキャプチャデバイスとしては比較的安価である. また, 一般的なモーションキャプチャシステムは動作を取りたい人の体に専用のマーカを取り付ける必要がある. それに対し Kinect は, 体にマーカをつけることなく動作を取得でき, 手軽に運用できるという特徴がある. そのため今回はモーションキャプチャデバイスとして Kinect を用いた. Kinect を用いたモーションキャプチャにより, 体幹, 首, 右肩, 右肘, 右手首, 左肩, 左肘, 左手首といった体の各部位の 3 次元座標を取得することが可能である. 本システムでは, 取得した各部位の 3 次元座標から体幹の方向ベクトルや上腕のベクトルなどを算出し, それらのベクトルの角度をロボットの各関節のサーボモータの目的角として動作のマッピングを行った.

頭部動作の取得には Web カメラと SeeingMachines 社の faceAPI を用いた[28]. faceAPI の動作の様子を図 2-7 に示す. faceAPI は, Web カメラから入力された画像を元にリアルタイムに顔の 3 次元位置(x, y, z 座標)と 3 次元姿勢(パン, チルト, ロール)を取得可能である. faceAPI は一般的に普及しているスペックの Web カメラの利用を前提としたソフトウェアである. そのため, 本システムでは Web カメラとしてロアス社の MCM-14BK を用いた. この Web カメラは 130 万画素で解像度は  $1280 \times 1024$  pixel である. 本システムでは, faceAPI で取得された 3 次元姿勢(パン, チルト, ロール)の各角度をテレプレゼンスロボットの頭部のパン, チルト, ロールにあたるサーボモータの関節角に 1:1 に対応させ, 頭部動作を行うようにした. faceAPI の頭部姿勢の計測範囲は, パン $-30^{\circ} \sim 30^{\circ}$ , チルト $-20^{\circ} \sim 45^{\circ}$ , ロール $-90^{\circ} \sim 90^{\circ}$ となっており, テレプレゼンスロボットの頭部サーボモータの動作可能角度は, パン $-70^{\circ} \sim 70^{\circ}$ , チルト $-15^{\circ} \sim 30^{\circ}$ , ロール $-65^{\circ} \sim 65^{\circ}$ である. ロボットの頭部動作角度は他パーツとの干渉を考慮して制限しており, 計測可能範囲よりも狭くなっているが, うなずく動作や見ている方向を表現するには十分な動作範囲である.



図 2-5 Kinect 外観

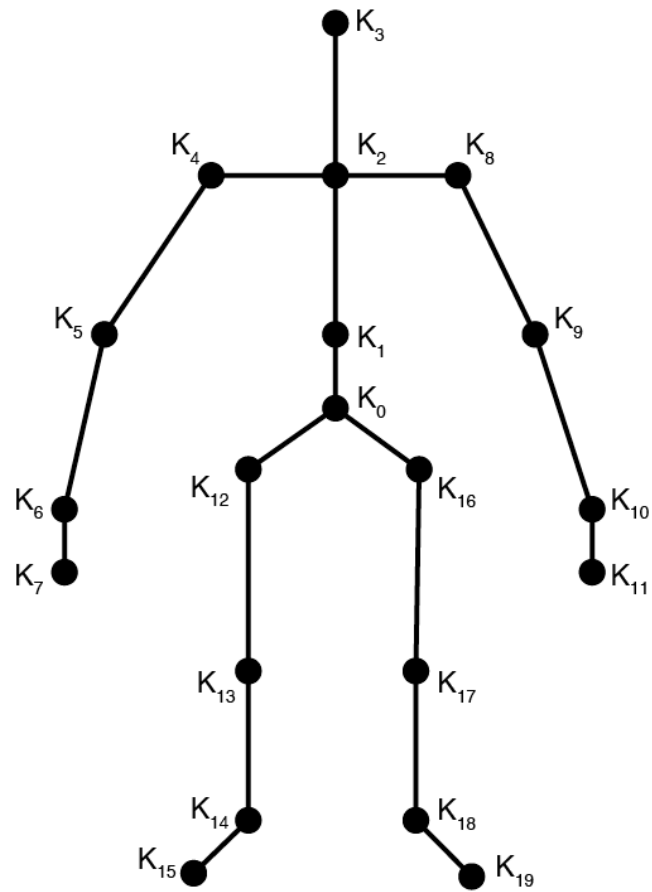


図 2-6 Kinect の骨格認識で取得可能な関節点

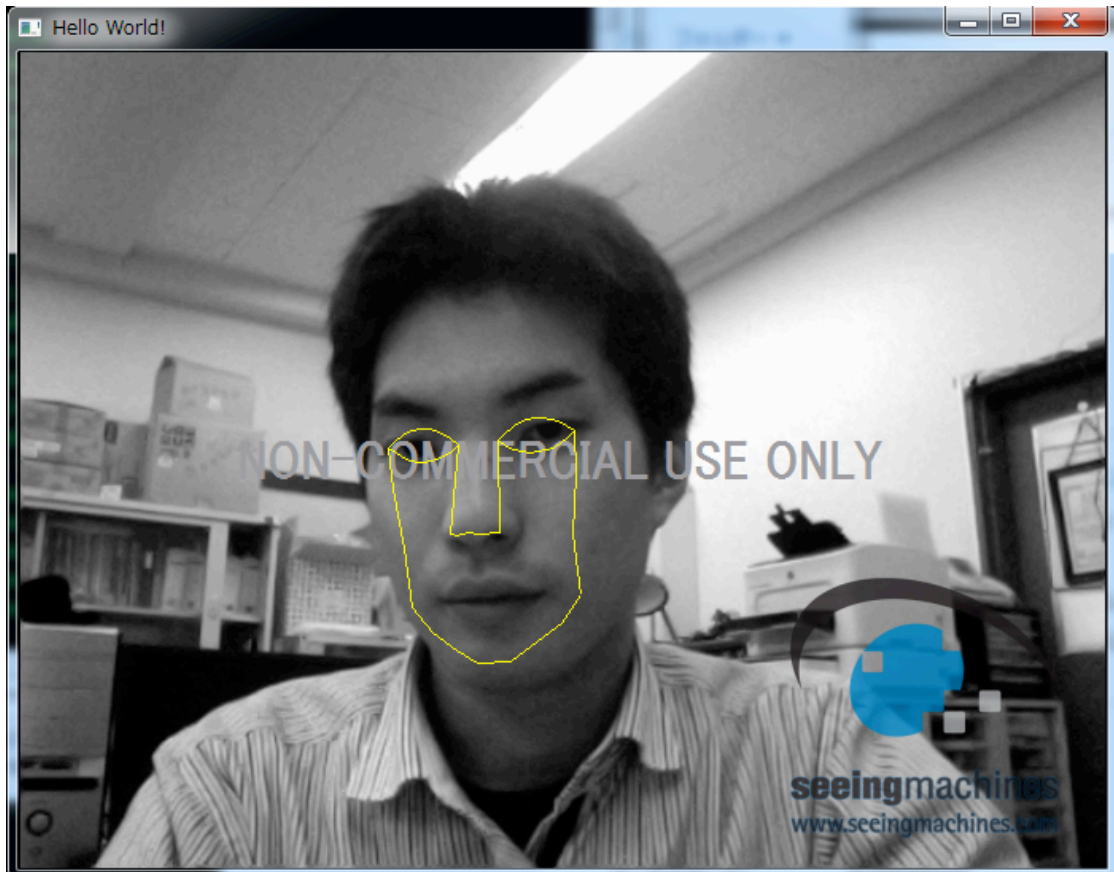


図 2-7 faceAPI の動作時の様子

#### 2.4.4 受動型コントローラ操作

提案するシステムの比較対象として、コントローラ操作により身ぶりを出表するシステムの実装を行った。システム全体の構成と用いるテレプレゼンスロボットはモーションキャプチャ操作を行うシステムと同一のものであるが、体幹と腕部の操作方法のみをコントローラ操作に変更する。コントローラ操作を用いたシステムは、関連研究の章で取り上げた MeBot の操作方法を参考にして実装を行った。具体的には、テレプレゼンスロボットの腕の動きについては、ロボットと同じ機構を持つコントローラによりマスタ・スレーブ方式で操作し、頭部の動きについては faceAPI を用いた操作を行う。

コントローラとしては、テレプレゼンスロボットと同様にサーボモータを追加し頭部のサーボモータのみを取り除いたもう一台の KHR-3HV を用いた(図 2-8)。遠隔地側では、藤科学独自のデータ通信規格である ICS 通信を用いて、受動型コントローラの各サーボモータの現在の角度データを制御 PC へと取り込む。その後、モーションキャプチャ操作の実装と同様に、SkypeAPI を用いてサーボモータの角度データを送信する。会議室側では、受信したコントローラのサーボモータ角度データをそのままテレプレゼンスロボットの各サーボモータの目標角度として、制御マイコンへと命令を送信する。このようにして、コントローラを用いてマスタ・スレーブ方式でテレプレゼンスロボットを介した身ぶり出表を行う。

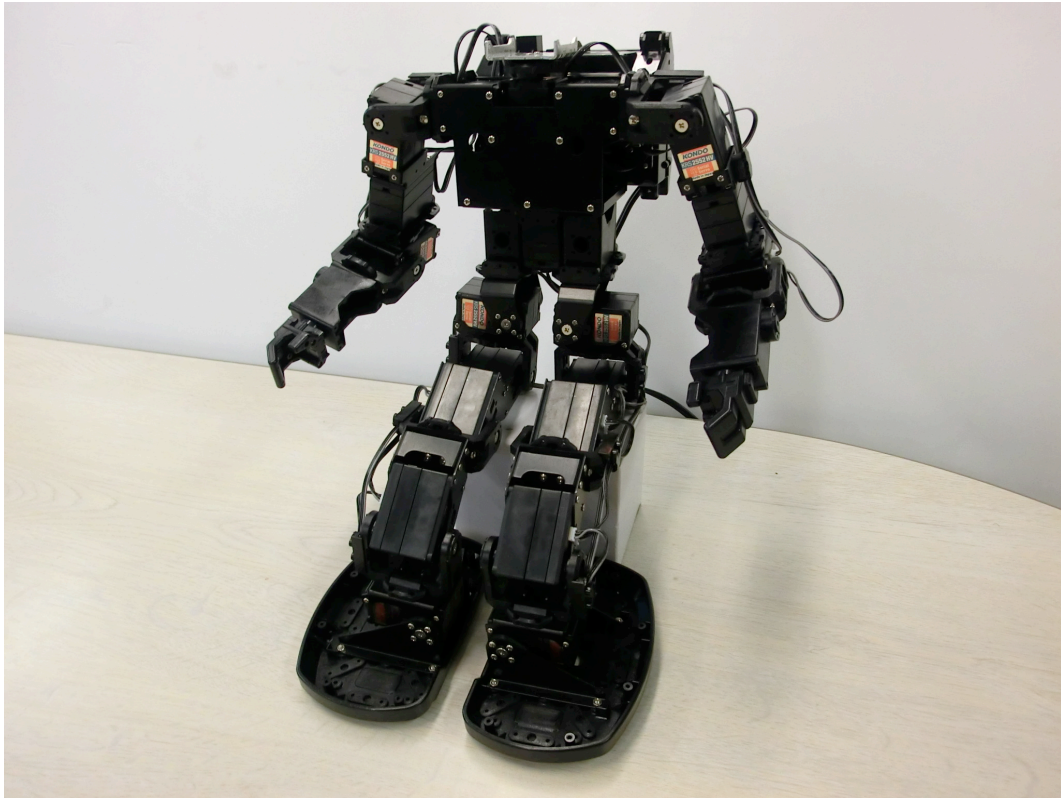


図 2-8 受動型コントローラ外観

#### 2.4.5 音声・動作の遅延について

音声の遅延が 300ms を超えると特に発話衝突の発生が顕著になると報告されている[45]. 我々が実装したシステムにおける音声の遅延は 100ms 程度であり, 音声遅延による発話衝突への影響はほぼないと考えられる. また, テレプレゼンスロボットの動作の遅延について, うなづくような小さな動きについては, 約 100ms 程度の遅延があった. 一方で, 腕を上げるような大きな動きでは, モーションキャプチャ操作にて約 600ms, 受動型コントローラ操作にて約 500ms の動作遅延がみられた. 玉木らは, 発話予備動作が行われた後にその動作がその後の発話確率に影響を与える持続時間があると報告している[38]. それによると, 手の動きと頭の動きが 1s, 頷きが 3s の持続時間を有しているとしており, 我々が実装したシステムのロボットの動作の遅延はこれらの持続時間よりも短いため, 遅延により発話予備動作としての効果が無くなることは無いと考えられる.

## 2.5 実証実験

### 2.5.1 実験目的

この実験の目的は、テレプレゼンスロボットの表出する無意識的身ぶりが会話中の発話衝突の減少に有効であるかを明らかにすることである。

### 2.5.2 比較条件

本実験では、提案手法である「モーションキャプチャ操作条件」と「受動型コントローラ操作条件」「対面会話条件」の3種類を比較条件とする。受動型コントローラ操作条件では遠隔参加者は受動型コントローラを用いてテレプレゼンスロボットを操作し、ロボットにより表出される身ぶりは意識的身ぶりのみになると考えられる。一方、モーションキャプチャ操作条件では意識的身ぶりに加えて無意識的身ぶりも表出されると考えられる。また、対面会話条件はテレプレゼンスロボットを利用せずすべての参加者が直接会って会話を行うものであり、ロボットを介さない自然な会話との比較を行うために用意した条件である。これらの比較により無意識的身ぶりが発話衝突に与える影響を検証する。

### 2.5.3 実験時の会話タスク

本実験では、会議の機能別の種類として創造会議と決定会議を想定し、会話タスクとして砂漠生き残り問題を用いた。また、会議の参加人数については、創造会議や決定会議に適している3人から10人の小集団会議を想定し、よりシンプルな実験条件にするために最少人数である3人での会議とした。

砂漠生き残り問題とは、砂漠で遭難しているという状況を想定し、リストアップされた道具について生き残るために必要なものの優先順位を議論により決定するというタスクである。実際に会話タスクに用いる道具のリストを表2-2に示す。この会話タスクは1回あたり4分間で行う。この会話1回を1セッションと呼ぶ。また、遠隔参加者となる被験者と操作条件をそれぞれ変更し、会話タスクに用いる道具の組み合わせも別のものに変更して各セッションを行う。1グループにつき3人の被験者がいるため、被験者と比較条件の組合せは、被験者A、B、Cのうち1人が遠隔参加者となり提案手法で会話に参加する場合は3通り、被験者A、B、Cのうち1人が既存手法で会話に参加する場合は3通り、被験者A、B、C全員が対面会話を行う場合は1通りとなる。よって、1グループあたりのセッション数は合計7セッションとなる。

表 2-2 砂漠生き残り問題の道具リスト

	道具
セッション1	<ul style="list-style-type: none"> <li>○ 1人につき1リットルの水</li> <li>○ チョコレート</li> <li>○ 食塩の錠剤</li> </ul>
セッション2	<ul style="list-style-type: none"> <li>○ 防水シート</li> <li>○ 毛布</li> <li>○ 軽装のコート</li> </ul>
セッション3	<ul style="list-style-type: none"> <li>○ 拳銃</li> <li>○ ナイフ</li> <li>○ フライパン</li> </ul>
セッション4	<ul style="list-style-type: none"> <li>○ マッチ</li> <li>○ 懐中電灯</li> <li>○ 赤と白模様のパラシュート</li> </ul>
セッション5	<ul style="list-style-type: none"> <li>○ 「食用に適する砂漠の動植物」という本</li> <li>○ その地域の航空地図</li> <li>○ 携帯電話</li> </ul>
セッション6	<ul style="list-style-type: none"> <li>○ ガーゼと包帯</li> <li>○ 人数分のサングラス</li> <li>○ ウォッカ</li> </ul>
セッション7	<ul style="list-style-type: none"> <li>○ コンパス(方位磁石)</li> <li>○ 消毒薬</li> <li>○ 化粧用の鏡</li> </ul>

#### 2.5.4 実験環境

実験環境の構成を図 2-9 に、会議室側と遠隔地側の様子をそれぞれ図 2-10 と図 2-11 に示す。発話衝突は複数人が参加する遠隔会話で発生しやすい。そのため本実験では、3 人参加の多人数会話を想定し、そのうちの 1 人が遠隔参加者として遠隔地からロボットを操作して会話に参加する。湯浅らによると多人数映像会話の場面において、視線一致環境では視線不一致環境に比べて身振りの表出回数が半分になるという報告がある[12]。そのため、視線の不一致により身振りの表出が減少することを避けるため、会議室側にはそれぞれの参加者を捉えるカメラを設置した。そして、遠隔地側では遠隔参加者から見た対話者の角度が対面状態と一致するように配置を調整した 2 台のディスプレイそれぞれにカメラの映像を映した。これにより遠隔参加者が左の参加者に視線を向けている場合と右の参加者に視線を向けている場合の顔の向きが明確に変化する。これにより、テレプレゼンスロボットの頭部の左右動作が視線方向を表しやすくなる。また、会議室側に設置した 2 つのカメラはそれぞれ参加者をとらえると共に、テレプレゼンスロボットの後ろ姿の一部も入るように調整した。これにより、遠隔参加者はディスプレイの端に映るテレプレゼンスロボットを見ることができ、自分の意図したようにロボットが動いているか確認できるようになっている。

また、システムの動作テストを行った際に、ネットワークの不安定によりロボットの動作にラグが発生することが分かっていた。ラグが会話に与える影響を出来る限り少なくし、操作条件の違いが会話に与える影響を捉えやすくする実験設定が望ましい。そのため、本実験ではネットワークを介しての通信は行わず、会議室側の参加者と同一の部屋にて遠隔参加者がロボットの操作を行うものとした。その際に、会議室側の参加者と遠隔参加者の間に衝立を設け、互いの姿が直接見えないように配慮した。同一の部屋にて実験を行うため、音声チャットは用いずに衝立を挟んでそのままの声で会話を行うこととした。

また、会話中に手元に何かしらのものがあると、それをいじるような動作が発生してしまう可能性がある。そのような会話とは関係なくものをいじるような動作が表れやすくなると、会話中の身ぶりが減少したり、身ぶりの分類が困難になったりすると考え、会話中の被験者には道具リストの書かれた紙などは持たせずに会話を行わせることとした。



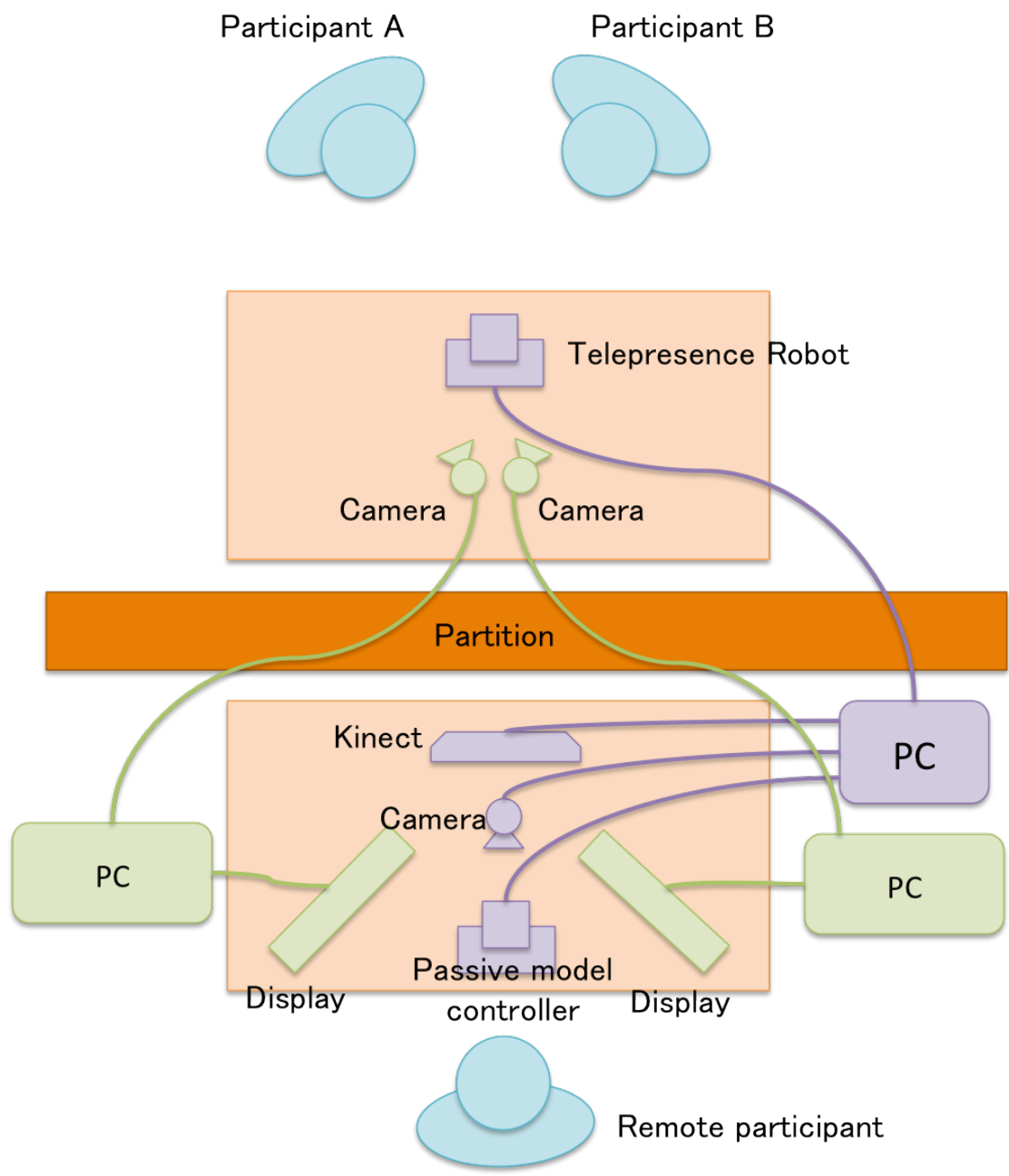


图 2-9 实验环境

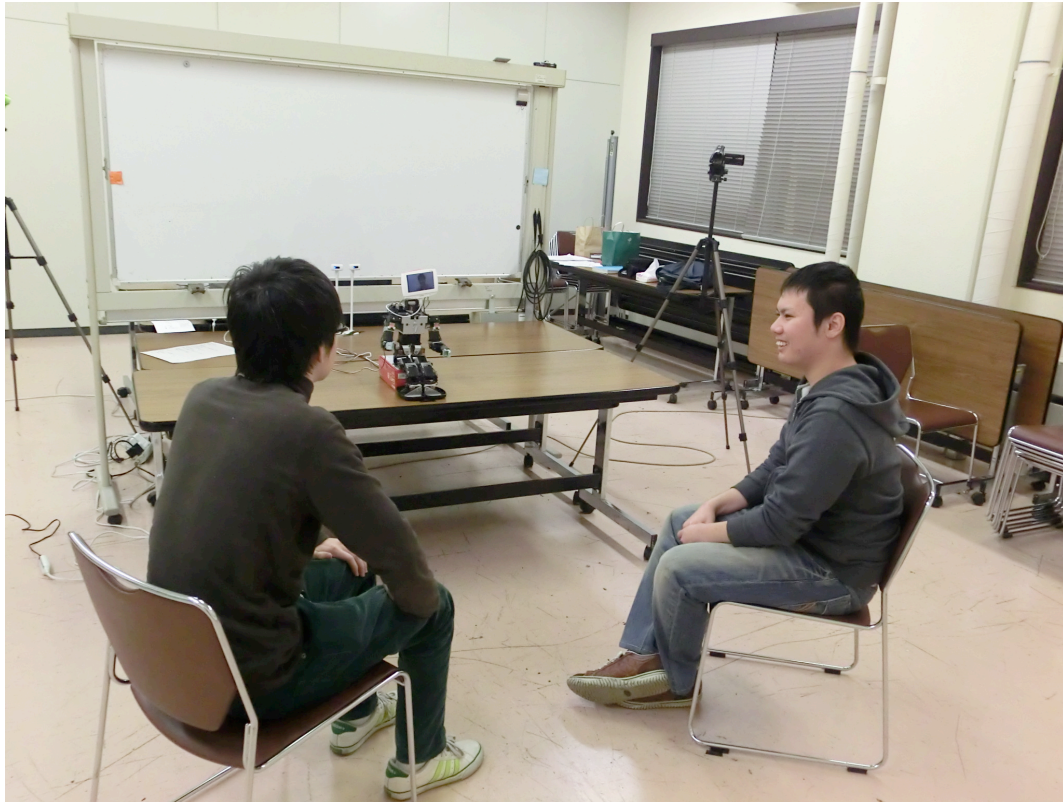


図 2-10 会議室側の様子

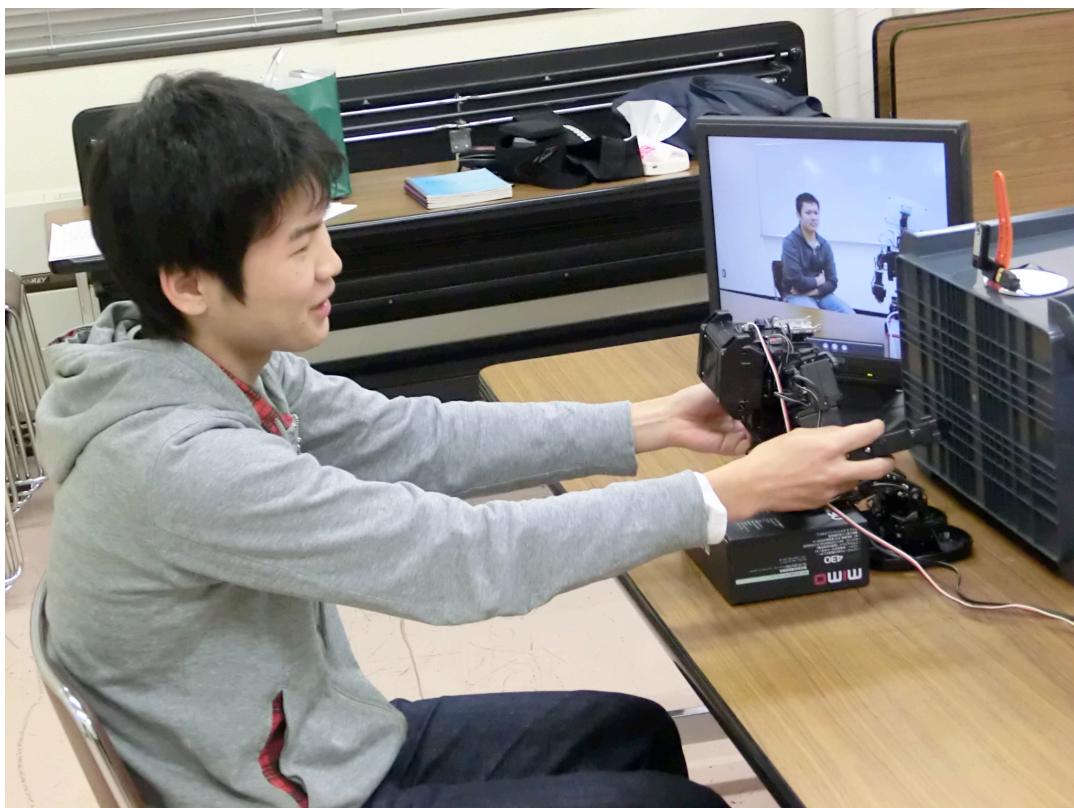


図 2-11 遠隔地側の様子

### 2.5.5 実験の流れ

実験の流れを図 2-12 に示す。

まず実験全体に関する説明を行った後、ロボットの操作方法について説明する。その後、3人の被験者にそれぞれに実際にモーションキャプチャ操作とコントローラ操作の両方でロボットを操作してもらい、1分間ほど操作を練習させる。次に、砂漠生き残り問題についての説明を行い、会話を行う前の各被験者の意見を収集するために、優先順位の事前調査として道具リストに各セッションでの3つずつの道具に優先順位をつけてもらう。

その次に、会話タスクを行う。まず実験者が今回の会話で話し合ってもらう3つの道具を口頭で読み上げる。その後、提示した3つの道具について実際の会話を4分間行わせる。実験全体の説明の際に、4分間のうちに3人で優先順位の意見をまとめること、4分経つ前に意見がまとまったとしても、道具の他の使い方がないかなど4分経つまで会話を続けることを被験者に指示した。また、会話開始から3分でベルを1鈴、4分で2鈴鳴らすことにより会話時間の経過を被験者に伝えることとした。なお、被験者が時間の経過を意識しすぎないように時計やタイマーなどは被験者からは見えない位置に置いた。会話が終了したら、その会話で決定した道具の優先順位を実験者に口頭で伝えてもらう。これらの流れを1回の会話タスクとし、遠隔参加者訳になる被験者や比較条件を変えて合計7回の会話タスクを行わせる。比較条件の順序については順序効果の影響がキャンセルされるよう考慮した。また、比較条件と話し合う道具の組み合わせが異なるようにした。

実験の最後のアンケートでは、ロボットや操作方法に対する印象や気付いたことについて自由記述形式で回答させる。

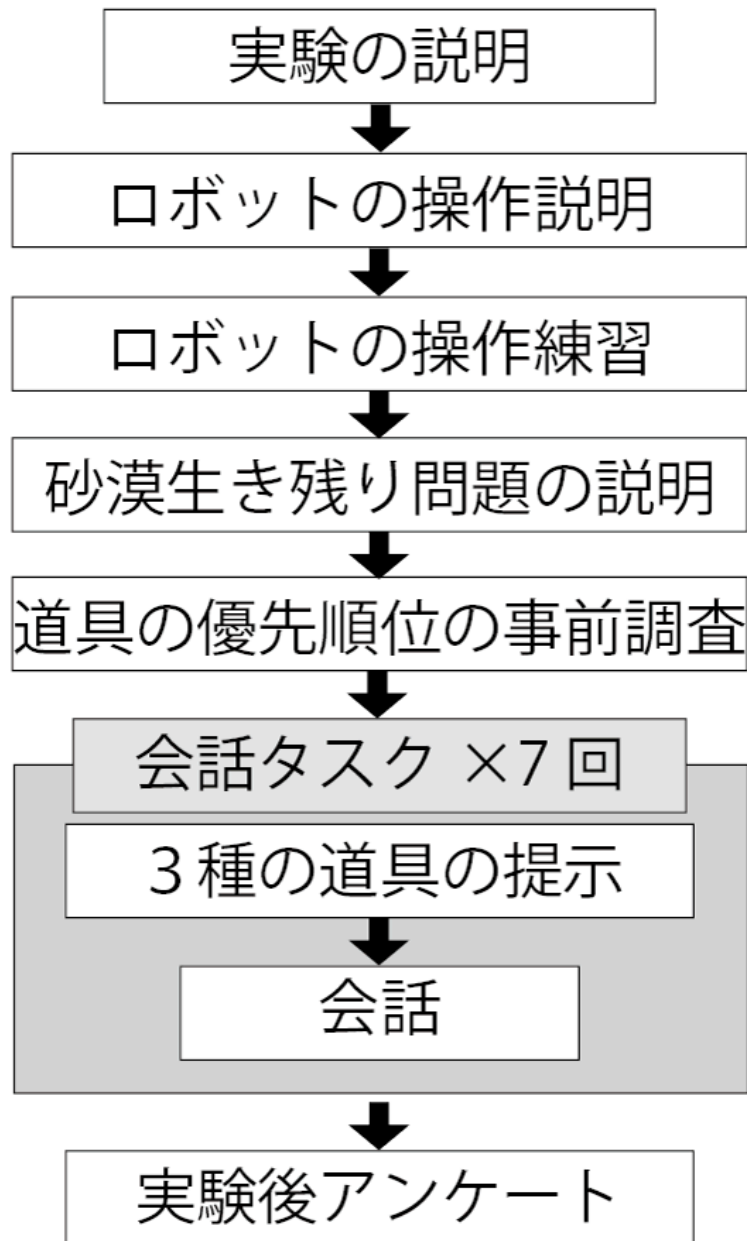


図 2-12 実験の流れ

### 2.5.6 評価方法

実験中の会話の様子を観察することによりモーションキャプチャ操作条件と受動型コントローラ操作条件での会話の評価を行う。評価の指標としては、会話中の発話衝突の回数と参加者の発話ターン取得率を用いる。発話ターン取得率は、会話中における全発話ターンのうち、ある参加者が獲得した発話ターンの割合とする。

受動型コントローラ操作条件では意識的身ぶりしか表出されないが、モーションキャプチャ操作条件では意識的身ぶりに加えて無意識的身ぶりも表出され、その身ぶりが発話予備動作の役割を有していると考えられる。そこで、以下の仮説を立て検証を行う。

仮説 1:モーションキャプチャ操作条件は受動型コントローラ操作条件に比べてロボットを介した参加者の発話ターン取得率が高い

仮説 2:モーションキャプチャ操作条件は受動型コントローラ操作条件に比べて発話衝突の回数が少ない

これらの仮説が検証されれば、モーションキャプチャを用いた操作方法により無意識的な身ぶりを表出でき、その身ぶりが発話衝突の回避と円滑な発話交替に有効であることが明らかになると考える。

## 2.6 実験結果と考察

### 2.6.1 実験結果

7組(合計21人)の被験者に対して実験を行い、ビデオカメラを用いて実験時の会話の様子を録画した。被験者は20歳代の男性17人、女性4人であった。実験後にビデオカメラの映像を見返し、テレプレゼンスロボットにより表出された身ぶりとは各参加者の発話回数、会話中の発話衝突回数のカウントを行った。

各操作条件におけるテレプレゼンスロボットにより表出された身ぶりのカウントについて、頭部の操作手法は提案手法と既存手法において同一のものをを用いたため、今回は頭部の動作による身ぶりは評価から除外し、体幹と腕による身ぶりを観察して分類と回数のカウントを行った。1グループあたり提案手法と既存手法で3セッションずつを行い、7グループ分のデータを取得したため、各条件で合計21セッションのデータから身ぶりを観察した。いずれかの条件において全セッションで合計5回以上観察された身ぶりについて、セッションあたりの平均回数を図2-13と図2-14に示す。グラフ中のエラーバーは計測値の標準偏差を表す。

図2-13に示した「手を前に出す」「片手を上げる」「手で方向を示す」といった身ぶりは他の参加者の注意を引くために行われていたため、相手に見せる目的で表出した身ぶりと考えられ「意識的身ぶり」とした。図2-14に示した「話者へ身を乗り出す」「話者へ体を向ける」「手を顔にやる」といった身ぶりは相手に見せる目的は無く表出した身ぶりと考えられ「無意識的身ぶり」とした。各身ぶりについてMann-WhitneyのU検定を行ったところ、「手で方向を示す」身ぶりとは「手を顔にやる」身ぶりについてモーションキャプチャ操作条件と受動型コントローラ操作条件の間に有意水準5%の有意差がみられた。

各条件における参加者の発話ターンの取得率を図2-15に示す。提案手法と既存手法についてはそれぞれ合計21セッション、対面会話条件は1グループあたり1セッションを行ったため合計7セッションのデータから発話ターン数のカウントを行った。また、条件別の平均総ターン数は表2-3のようになり、検定の結果、各条件の間に有意な差は見うけられなかった。ターン取得率の3条件間において反復測定1元配置分散分析を行った結果、有意水準1%の有意差がみられた。3条件間の組合せのどこに差があるのかをより詳細に明らかにするために、多重比較としてBonferroniの調整を用いた対応のあるt検定を行った。その結果、モーションキャプチャ操作条件と対面会話条件の間で有意水準1%の有意差がみられ、受動型コントローラ操作条件と対面会話条件の間で有意水準5%の有意差がみられた。このことから、直接会って会話を行う対面会話条件と比べて、遠隔地からロボットを介して会話に参加する場合に遠隔参加者のターン取得率が有意に少なくなることが明らかとなった。

各条件における1セッション中の発話衝突回数の平均を図2-16に示す。発話衝突回数の3条件間においてKruskal-Wallis検定を行った結果、有意水準5%で有意差がみられた。そこで3条件間のどの組合せに差があるのかより詳細に明らかにするために、多重比較としてBonferroniの調整を用いたMann-WhitneyのU検定を行った。その結果、モーションキャプチャ操作条件と受動型コントローラ操作条件の間に有意水準5%の有意差がみられた。このことから、受動型コントローラ操作条件に比べてモーションキャプチャ操作条件での発話衝突回数が有意に少なくなることが明らかとなった。

表 2-3 セッション中の総ターン数の平均

	Average of total turns per session	Standard Deviation
Proposed method	40.3	10
Existing method	46.5	9.48
Face-to-face	43.6	9.12

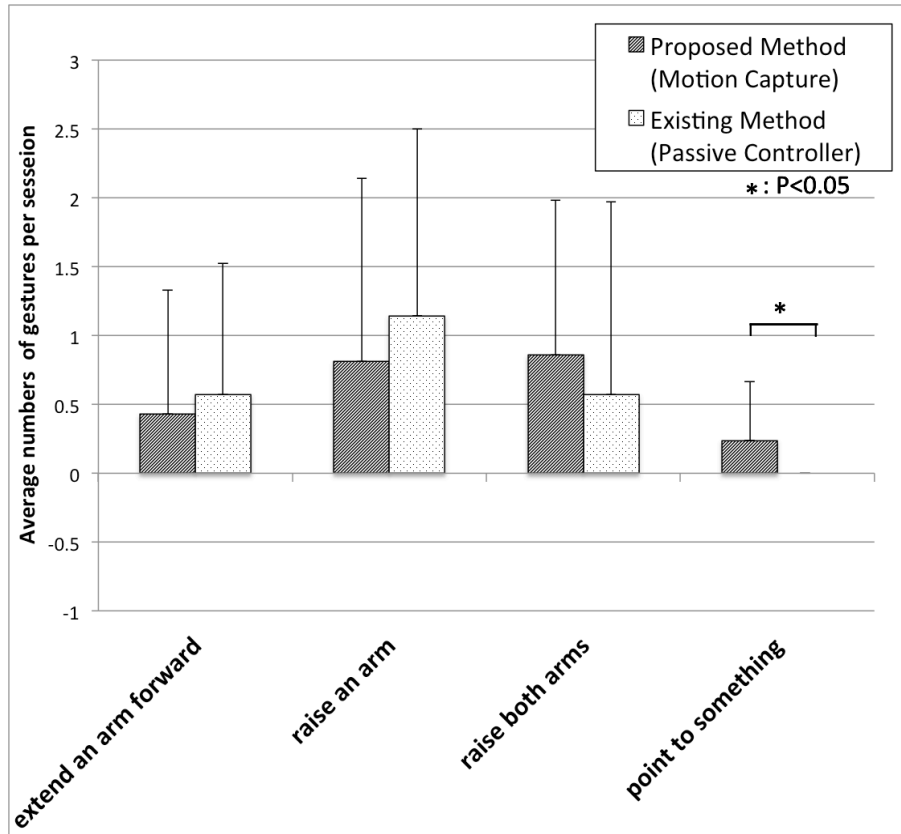


図 2-13 ロボットにより表出された意識的身ぶりの回数

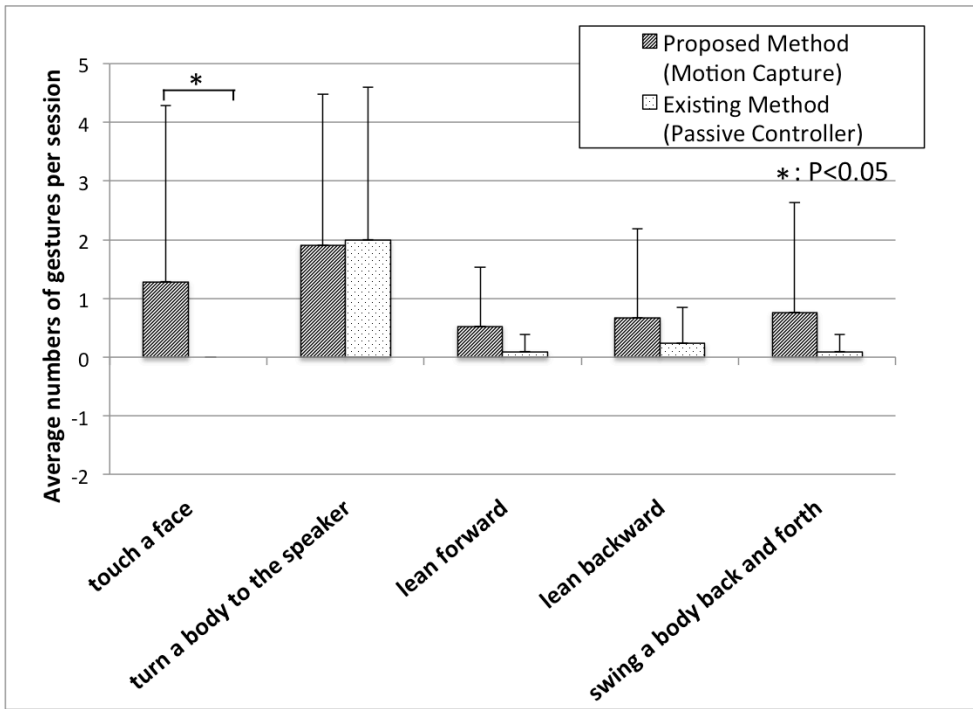


図 2-14 ロボットにより表出された無意識的身ぶりの回数

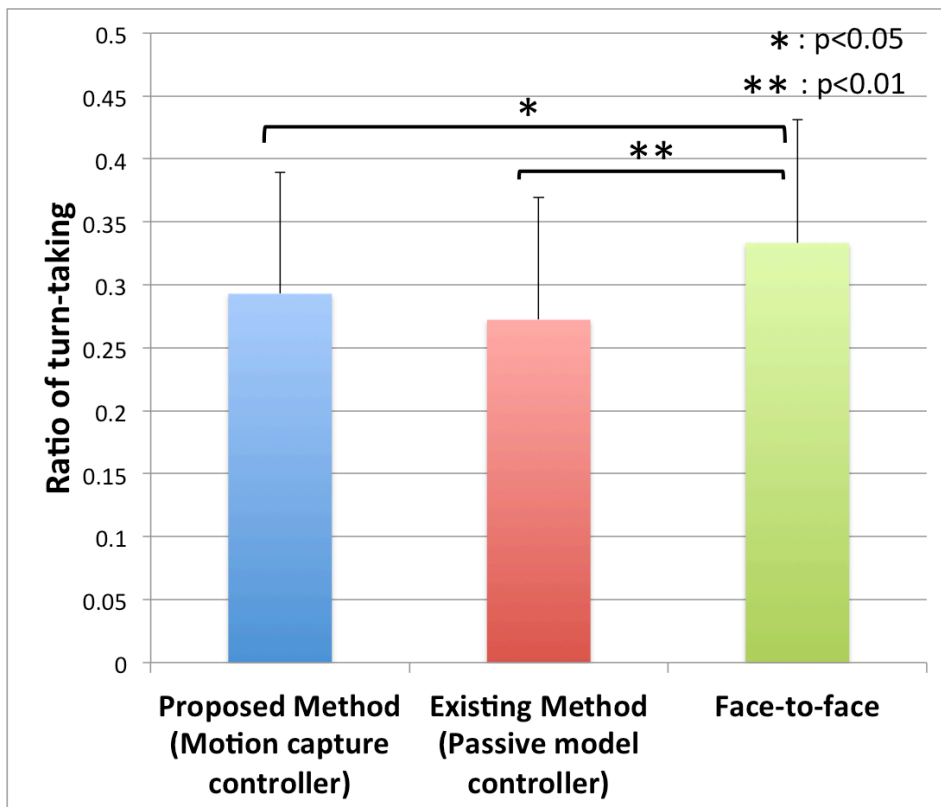


図 2-15 遠隔参加者のターン取得率



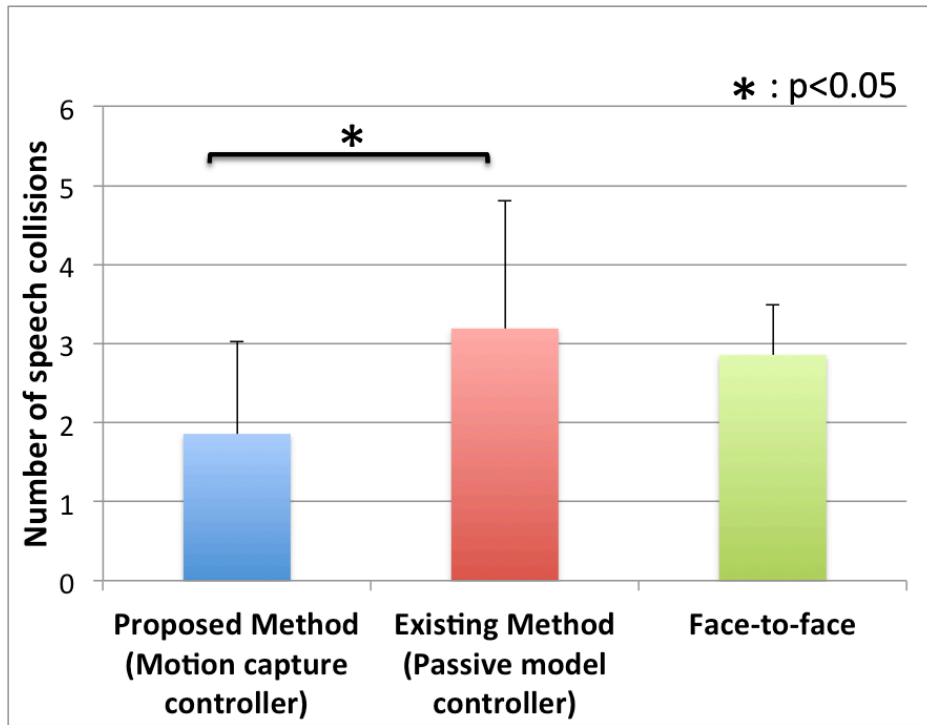


図 2-16 発話衝突回数

## 2.6.2 考察

テレプレゼンスロボットが表出する無意識的な身ぶりについては、図 2-14 の結果から「話者へ体を向ける」身ぶりについては両条件ではほぼ同数であった。これは、受動型コントローラ操作条件の際に半数以上の被験者が、現在注目している参加者の方にロボットの体を向けるという操作をこまめに行っていたため操作条件の違いにより明確な差がでなかったと考えられる。「話者へ身を乗り出す」「体をのけぞる」「体を前後に揺らす」といった身ぶりについては、有意差は確認されなかったもののモーションキャプチャ操作条件のほうが表出回数が多い傾向であった。「手を顔にやる」身ぶりについては受動型コントローラ操作条件では表出されず、モーションキャプチャ操作条件でのみ表出されていたことがわかる。以上のことから、全体としては既存手法と比較して提案手法であるモーションキャプチャ操作条件のほうがより「無意識的身ぶり」を表出していたといえる。

また、図 2-13 の結果からモーションキャプチャ操作条件と受動型コントローラ操作条件の双方で、手を上げて相手の注意を引くといった意識的な身ぶりが表出できていた。「手で方向を示す」身ぶりについてはモーションキャプチャ操作条件でのみ見うけられた。これは、モーションキャプチャ操作のほうがより直感的な操作であったためであると考えられる。

仮説 1 の「モーションキャプチャ操作条件は受動型コントローラ操作条件に比べてロボットを介した参加者の発話ターン取得率が高い」については、図 2-15 の結果からは仮説の立証はされなかった。実験結果から、モーションキャプチャ操作条件と受動型コントローラ操作条件のターン取得率はどちらも対面会話条件のターン取得率より減少する傾向にあることがわかった。そのため、ターンの取得率は操作条件だけではなく他の要因が考えられる。要因の 1 つとして、遠隔参加者の得る隣在側の情報が映像と音声のみの既存の遠隔会話と同じものであることがあげられる。そのため遠隔参加者が隣在参加者の発話予備動作を見逃している可能性が高い。このことが要因となり操作者がターン取得の機会を逃していたと考えられる。今後はターンの取得率に関しても対面条件と同等なものを目指す必要がある。

仮説 2 の「モーションキャプチャ操作条件は受動型コントローラ操作条件に比べて発話衝突の回数が少ない」については、図 2-16 の結果からモーションキャプチャ操作条件の発話衝突回数が有意に少ないことが確認された。そのため仮説は立証されたといえる。このことから、モーションキャプチャ操作により無意識的な動作をロボットを介して伝達することは発話衝突の回避に有効であるといえる。

本研究は、玉木らの研究等において指摘されている「対面会話に比べて遠隔会話の方が発話衝突が起りやすい」という知見[38]に基づき提案を行っている。しかしながら、図 2-16 では、「対面会話と提案手法の間」と「対面会話と既存手法の間」に有意差が見うけられず、対面会話に比べて遠隔会話の方が、発話衝突が起りやすいという結果にはならなかった。これについて 2 つの原因が考えられる。1 つ目の原因は、図 2-15 から分かるように提案手法や既存手法を用いた遠隔会話では、対面会話と比べて遠隔参加者のターン取得率が有意に減少していることである。このことから遠隔参加者は対面会話と比べて会話に参加する機会自体が減少していたことが分かる。そのため母数である発話回数自体が遠隔会話と対面会話とで異なると考えられ、発話衝突回数を単純に比較することはできない。2 つ目の原因は、玉木らの実験と今回の実験での実験環境の違いである。玉木らの実験では遠隔会話の際には参加者全員が Web 会議システムを用いていた。そのため、他の参加者全員の発話予備動作に気付きにくくなり発話衝突が起りやすくなっていた。一方、今回の実験では二人の参加者が対面会話をしているところに一人の遠隔参加者がロボットを介して会話に参加するというものであった。そのため、既存手法のロボットを介している遠隔参加者の発話予備動作は伝達されづらいが、もう一人の参加者については対面会話と同一であるため発話予備動作は伝わりやすいままである。したがって、玉木らの実験環境よりも発話衝突は起りづらくなるも

のと考えられる。これら2つの原因から、今回の実験結果として対面会話と遠隔会話における発話衝突回数に有意な差が表れなかったものと考えられる。

各操作時の会話の印象や3人対面時とロボットを介した参加者がいる時の印象などを実験後に回答してもらった記述形式のアンケートでは、いくつかの内容が多くの特験者に共通して表れていた。モーションキャプチャ操作条件の際の印象については、「自分が特に意識した動きではないものもロボットに反映されたことがリアルだと思った」「自分の何げない動きにも反応してしまう」という記述がみられた。実験の際に被験者には、この実験が無意識的身ぶりを伝達することの有無を比較する実験だということは伝えていなかった。それにも関わらず、このような無意識的身ぶりに関する意見が得られたのは、ロボットの後ろ姿が遠隔参加者の見るディスプレイに入るようにカメラをセットしており、リアルタイムに操作しているロボットの動作が確認できたためだと考えられる。ロボットの動作を操作者に見えるようにしたのは、自分の代理となるロボットが他者からどのように見えているのか、思ったように動いているのかを操作者が確認できるようにするためであった。アンケートでこのような意見が得られたことは、提案手法が無意識的身ぶりを表出していることを表すと同時に、ロボットの表出する無意識的身ぶりを操作者が見ることで無意識的身ぶりが意識にあがってしまうという問題が発生していることも表している。今後は本システムで操作者にフィードバックする情報に関しても慎重に考慮する必要がある。

受動型コントローラ操作条件での印象として、「人と話しながらの操作は難しかった」「話をしながらロボットを動かすことに違和感があった。気付いたらロボットから手が離れていた」「操作し忘れることがある」といった記述がみられた。これは実験のビデオでも確認されており、話が盛り上がるとコントローラをつかんではいても全く動かさなかったり、完全にコントローラから手を放して会話をしたりという様子がみられた。また、相手から見えないにも関わらずコントローラを放した状態で手ぶりをするという場面もみられた。これは、コントローラ操作でロボットを動かすということを被験者が初めて体験したため操作の熟練度が足りなかったとも考えられるが、人は相手に何かを伝えるためだけでなく、自身が発話する際に単語検索や情報の組織化を行う過程に身振りが付随するという仮説がある[16][20]。このことから、会話をしながら身ぶりを操作するということが通常の会話中の行動からかなり逸脱したものであり、同時にこなすことは難しいため意識的な身ぶりの表現にも支障が出ると考えられる。このように、無意識的身ぶりの減少だけが要因ではなく、このコントローラ操作と会話を同時に行うことが認知的負荷を高くし、会話や発話交替、相手の発話予備動作への注意が逸れやすくなったことも要因となり、発話衝突が起こりやすくなった可能性も考えられる。もし、コントローラ操作で認知的負荷が高くなることが主な要因であった場合、提案手法であるモーションキャプチャ操作条件と通常のビデオ会議はどちらもコントローラ操作の必要がないため、発話衝突の回数に差が表れないという可能性もある。しかしながら、モーションキャプチャ操作条件と通常のビデオ会議では身ぶりを表現する手段が3次元的なロボットと2次元的な映像という大きな違いがあり、身ぶりの注目のされやすさや存在感、3次元的な方向を表す正確さでは、ロボットのほうが影響力があると考えられる。そのため、通常のテレビ会議と比較し提案手法のほうが発話衝突を抑えることが出来ると推測される。この点については今後、通常のビデオ会議についても比較条件として追加し効果の検証を行う必要がある。

3人対面時とロボットを介した参加者がいる時の印象の違いについては、「対面している時の方が説得感があった。ロボットだと、テレビのように画面の向こう側にいる感覚だった」「ロボット越しの相手に対しては話をふりにくかった」といった記述がみられた。また、自身がロボットを介して参加している場合の印象については、「3人で会話しているというより、2対1で会話している感じがした」「やや会話の流れから疎外されて、会話におじゃましている感じがした」という記述がみられた。このことから、普通に参加している人はロボット参

与者に話しかけることに多少なりとも抵抗を感じると同時に、ロボットを介して参与している人は会話に入りづらい印象を受けていることがわかる。これは、図 2-15 にて示したターン取得率の結果とも一致しており、会話への入りやすさを向上させることが今後必要であると考えられる。

また、実験のビデオの様子から、3 人対面時の場合と比べロボットを介して参与している場合に身ぶりが減少している傾向があった。直接会って会話を行う場合と比較して遠隔地からディスプレイ越しに会話を行った場合に表出する身ぶりが減少することは、玉木らの研究や片山らの研究においても確認されている[12][38]。このことから、遠隔参加者の身ぶりをそのまま伝達するだけでは、対面時と同等の身ぶりは表出できないと考えられる。そのため、遠隔参加者の身ぶりをロボットで表出する際に誇張したり、身ぶりの回数を増加させたりする変換を行い、対面時と同等の効果を得られるようにする必要がある。

### 2.6.3 結論

本章では遠隔会話における発話衝突の問題を減少させるために、無意識的身ぶりを伝達して表出可能なテレプレゼンスロボットを提案し、実験により発話衝突が減少することを確認した。音声と映像のみの遠隔会話の場面では発話予備動作が認知されにくくなり、発話衝突という問題が発生する。そこで、3 次元的な実体を持つロボットを介して身ぶりを表出することにより身ぶりが認知されやすくなると考えた。また、身ぶりの中でも無意識的に行っている身ぶりが発話予備動作を有するため話者交替において重要であると考え、提案したのが無意識的身ぶりを表出可能なテレプレゼンスロボットである。そして意識的身ぶりのみを表出する受動型コントローラ操作条件と、提案手法である無意識的身ぶりも表出するモーションキャプチャ操作条件を比較する実験を行い、提案手法が発話衝突を減少させることを確認した。

一方で、遠隔参加者のターンの取得率については、対面時と比較してテレプレゼンスロボットを介して遠隔地から参与した場合に減少するという課題が残った。これは遠隔会議と対面での会議の間での大きな違いであり、遠隔参加者のターン取得率が減少する課題を解決することで、遠隔での会話を対面時の理想的な状態に近づけることができると考える。

## 第3章 発話機会獲得のための身ぶりの表出

### 3.1 発話交替と参与役割

ゴフマンは会話中の参与構造というものを提案している。参与構造の中では参与者の参与役割というものを定義している[5]。参与役割には、例えば、現在話をしている話し手やその話し手から注意を向けられてながら話を聞いている聞き手、話し手に注意を向けられてはいないが会話の輪に入って話を聞いている某参与者などがある。会話の参与者はこの参与役割を、視線を手がかりにして理解していると言われている[43]。そして、多人数会話ではこの参与役割が状況により入れ替わりながら会話が進んでいくこととなる。また、次に誰がどの参与役割になるかは、発話内容や非言語情報（視線方向、姿勢、身ぶり等）によって決まるといわれている。さらに、某参与者にくらべて、聞き手となっている参与者が次の話者になりやすいともいわれている。そのため発話機会を得る前段階として、他の参与者が話し手となっている時にその聞き手の参与役割になることが重要であるといえる[4]。

### 3.2 発話機会獲得のための身ぶり

ゴフマンの提唱した参与役割をもとに我々が作成した参与役割の状態遷移図を図 3-1 に示す。某参与者と比べて聞き手の役割になっている参与者が一般的に次話者になりやすいといわれている。そのため、他の参与者が話し手である際にその話し手の注目を得て聞き手になっておくことが、次話者になるために重要である。話し手の注目を得るためには、非言語的な手段で話し手に対して、自分は話に興味を持って聞いているということを示す必要がある。その際に用いられる具体的な非言語情報として以下のものがあげられる。

- 頭部方向
- うなずき動作
- 相槌

Goodwinによると、話し手は視線を使って誰に話しかけているのかを示し、受け手はその視線を話し手に向けることで、話し手の話を聞いているということを伝えている[6]。そして、このような話し手と受け手の間の相互の注目が会話の成立に不可欠であるといっている。また、Kendonは、視線には他人の行動をモニタリングする機能、自らの態度や意図を表出する機能、会話の流れを調整する機能を有していると主張している[14]。参与者が誰を注視しているか計測するには視線方向を計測する必要がある。しかしながら、視線方向を計測する装置は未だ高価であり、遠隔会議システムに導入するという点では非現実的である。その一方で、より容易に計測が可能な頭部方向を視線の代用として用いるという試みが行われている[10][13][25][31]。頭部方向は自身の興味の方角を示すため、それを話し手に向けることは話に興味を持っているということを示す上で重要になってくる。うなずき動作については、相手の話に合わせて行うことにより、話を聞いているということを示すことが可能である。ここであげた相槌とは、「うんうん」「あー」「そうそう」のような声による相槌のことである。声による相槌も、うなずきと同様に相手の話に合わせて行うことにより、話を聞いているということを示す効果があると考えられる。話を聞いている際にこれらの動作を表出することにより、話し手の注意を引き、聞き手になりやすくなると思われる。

しかしながら、対面で会話を行う場合と比較して、ビデオチャットのようにディスプレイに映った相手に対して会話を行う際には身ぶりの頻度が減少するといわれている[38]。また、

そのような場合には表出される身ぶり自体も小さい動作になると考えられる。そのため、頻度が減少し動きが小さくなった動作をそのまま伝達しても対面会話時と同等の効果は得られない。そこで、受け手になっていることを表す非言語情報を誇張して表出する必要がある。

そこで、遠隔多人数会話において発話機会を獲得するためのテレプレゼンスロボットを提案する。

提案システムでは、遠隔参加者の身ぶりを検出し、他の参加者が話し手となっている際にテレプレゼンスロボットが身ぶりを誇張表現することによって、遠隔参加者が聞き手の役割を得やすくなるよう支援を行う。身ぶりの誇張を行うことにより、遠隔参加者が聞き手になる機会と聞き手になってから次話者として発話を行う機会を増加させることができるものと考えられる。

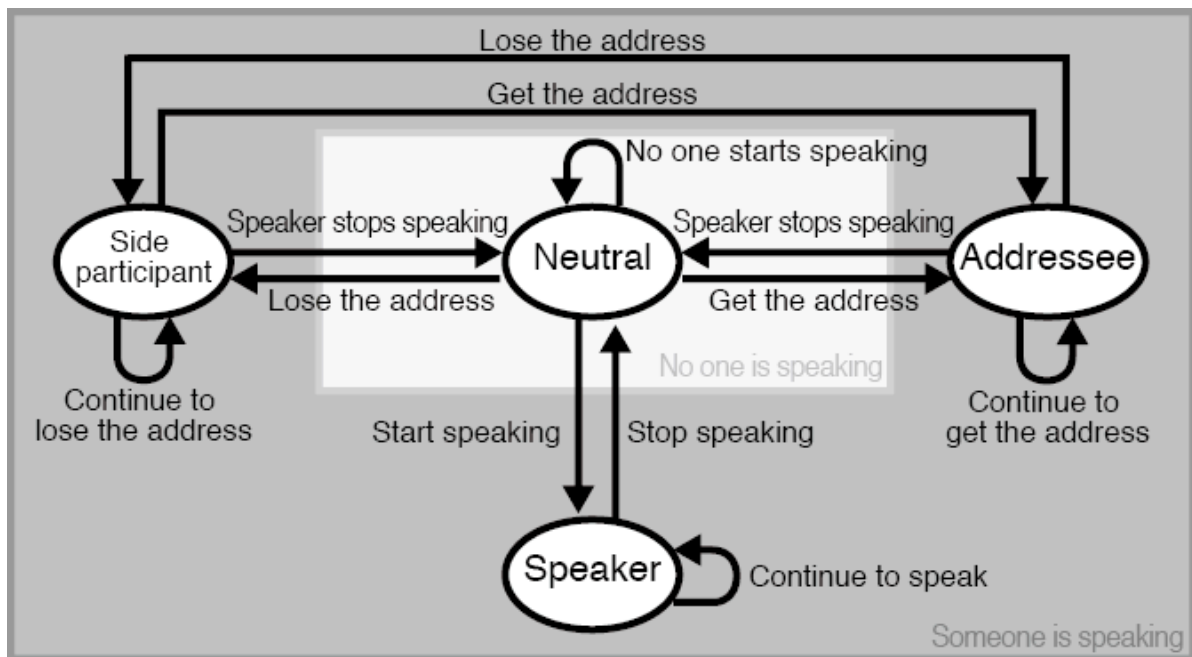


図 3-1 参与役割の状態遷移図

### 3.3 身ぶりの誇張方法

頭部方向の誇張方法を表 3-1 に示す。複数人での会話において遠隔参加者の頭部のパン方向の動作は、現在の話者のほうを見るために頻出しやすい動作である。しかしながら、対面での会話と比べて注目している方向に体ごと向くということは遠隔参加者にとってあまりない行動である。体幹方向というのは、頭部方向よりも変化する頻度が少なく一定であるため、主要な興味の対象を示すと考えられる。そのため、遠隔参加者の頭部のパン方向の変化が少ない時に、その方向を主要な興味の対象を示す方向と考え、それをテレプレゼンスロボットの体幹方向で表出することで、頭部方向の誇張を行う。

うなずき動作の誇張方法を表 3-2 に示す。対面時と比べ遠隔参加者のうなずき動作は小さくなると考えられる。そのため遠隔参加者の頭部のチルト方向の動作を大きくしてロボットに表出させる誇張を行う。テレプレゼンスロボットの頭部のチルト動作角度に限界がある場合には、体幹のチルト動作も併用することで大きくうなずいているように見せる誇張を行う。

相槌の誇張方法を表 3-3 に示す。遠隔参加者はうなずきの頻度も少なくなると考えられる。しかしながら、声による相槌については遠隔参加者であっても減少しにくい情報である。そこで、遠隔参加者が声による相槌を行った際に、テレプレゼンスロボットはうなずき動作を表出することにより、音声情報だけであった相槌にうなずきという視覚的な情報を付与して相槌を誇張する。

表 3-1 頭部方向の誇張

	遠隔参加者	ロボット
頭部のパン方向	○	○
体幹のパン方向	×	○

表 3-2 うなずき動作の誇張

	遠隔参加者	ロボット
頭部のチルト方向	○	○
体幹のチルト方向	×	○

表 3-3 相槌の誇張

	遠隔参加者	ロボット
相槌の音声	○	○
頭部のチルト方向	×	○



## 3.4 実装

### 3.4.1 システム構成

システムの構成を図 3-2 に示す。テレプレゼンスロボットは第 2 章のものと同じ構成であり、映像と音声のやりとりも同様に Skype を用いて行う。遠隔参加者側で取得したトラッキングデータの送信は、SkypeAPI の提供が終了したため、代わりにソケット通信を用いて制御用コンピュータ間の通信を行うように変更を施した。また、遠隔参加者への音声の提示はヘッドフォンまたはイヤフォンを用いて行うものとした。これは、遠隔地側で遠隔参加者の相槌の音声を取得する必要があるため、会議室側の音声をスピーカで提示すると、相槌の音声取得に支障が出るからである。

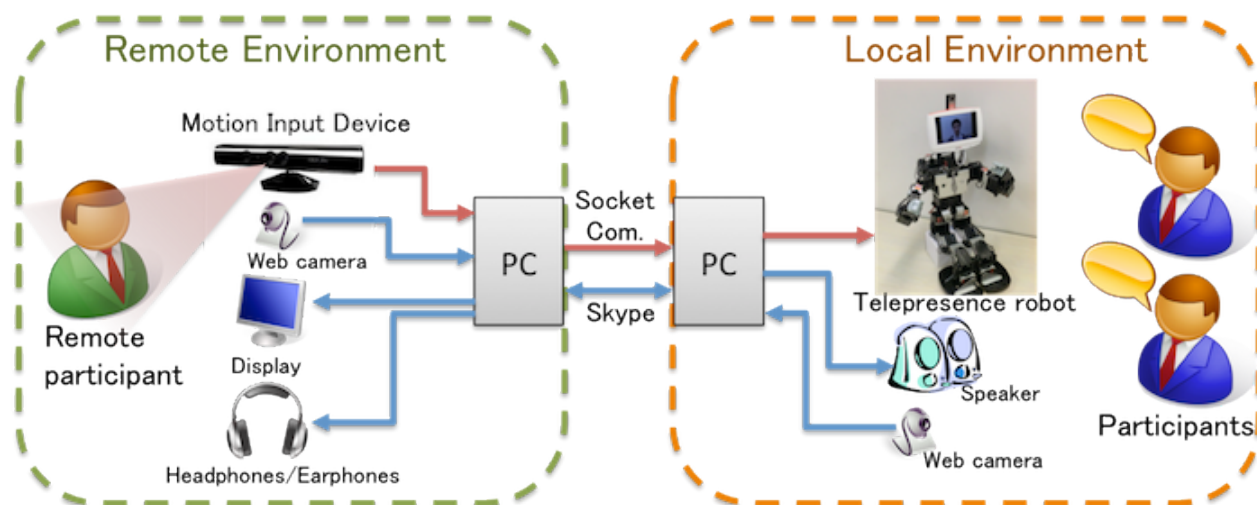


図 3-2 システム構成

### 3.4.2 モーションキャプチャデバイス

遠隔参加者の動作を取得するためのモーションキャプチャデバイスとしては第 2 章と同様に Microsoft 社製の Kinect を用いた。また、Kinect は同社のソフトウェア開発キット Kinect for Windows SDK を用いることにより、人物の骨格を認識の他に顔認識を行い顔の位置と方向を取得することが可能である。そのため、今回は faceAPI は用いずに Kinect によって顔の位置と方向を取得することとした。そして、Kinect はマイクロフォンを有しており、音圧を取得することが可能である。本システムでは、Kinect から得られる骨格認識データ、顔認識データ、音圧を入力情報として用いた。

### 3.4.3 頭部方向の誇張

頭部のパン方向については Kinect から得られる顔の方向のデータを利用した。遠隔参加者の頭部のパン方向の変化が少ない時に、その方向を主要な興味の対象を示す方向とし、テレプレゼンスロボットの体幹のパン方向の動作として表出させるために、頭部のパン方向の時間変化量の移動平均を算出した。その移動平均が閾値以下である時に、頭部のパン方向の変化が少ない状態だとし、その時の頭部のパン方向にテレプレゼンスロボットの体幹のパン方向を向かせるという実装を行った。

#### 3.4.4 うなずき動作の誇張

会話中において、頭部をチルト方向に動かす動作の大部分はうなずき動作を行うためであると考えられる。そのため、Kinect から得られる遠隔参加者の頭部のチルト方向の時間変化量が閾値以上である際に、テレプレゼンスロボットの頭部のチルト方向の目標角度を定数倍するという実装を行った。また、定数倍したあとの目標角度がロボットの頭部チルト方向の可動域を超える場合は、超えた分の角度を、体幹のチルト方向の目標角度に加算して体幹での誇張を行うものとした。

#### 3.4.5 相槌の誇張

Kinect はマイクフォンから取得される音圧データを元に簡易的に実装を行った。まず、ノイズを除去するために音圧データの時間的な移動平均を算出した。その移動平均の音圧に閾値を設け、その閾値以上ならば発話状態とした。相槌は短い発話だと仮定し、発話していない状態が現在時刻から一定時間前まで続いており、その前の時間に一定時間以内の発話状態があった場合には、相槌と判定するものとした。相槌と判定された際には、予め用意したうなずき動作をテレプレゼンスロボットの表出させるように実装を行った。

#### 3.4.6 音声・動作の遅延について

本システムにおける音声の遅延は 100ms 程度であった。また、テレプレゼンスロボットの動作の遅延について、うなずきのような小さな動きについては、約 100ms 程度の遅延があった。一方で、腕を上げるような大きな動きでは約 600ms の動作遅延がみられた。

うなずきの誇張を行った際には、遠隔参加者がうなずき始めてからテレプレゼンスロボットがうなずき始めるまでは、誇張なしの場合と変わらず約 100ms の遅延であった。誇張によりうなずきのための頭部チルト方向の動作角度が大きくなるため、遠隔参加者がうなずき終わってからテレプレゼンスロボットもうなずき終わるまでには約 300ms の遅延があった。誇張なしの場合には約 100ms であった。

また、相槌の誇張でのうなずき動作について、発声終了からうなずき始めるまでの間に約 500ms、発声終了からうなずきのチルト角度が最大になるまでに約 1000ms の遅延が見られた。

## 3.5 実証実験

### 3.5.1 実験目的

この実験の目的は、テレプレゼンスロボットの表出する誇張した身ぶりが会話中のターン取得に与える影響を明らかにすることである。

### 3.5.2 比較条件

本実験では、提案手法である「誇張あり条件」と既存手法である「誇張なし条件」、そして既存のビデオチャットを想定した「ディスプレイ条件」の3種類を比較条件とする。「誇張あり条件」は遠隔参加者がテレプレゼンスロボットを介して会話に参加し、その際にテレプレゼンスロボットが遠隔参加者の身ぶりを誇張して表出する実験条件である。「誇張なし条件」は、テレプレゼンスロボットを介して会話に参加するが身ぶりの誇張は行わず、そのままの身ぶりをテレプレゼンスロボットにより表出する。「ディスプレイ条件」では、既存のビデオチャットを想定しテレプレゼンスロボットの代わりに会議室側にディスプレイを配置し、遠隔参加者の姿をそのディスプレイ上に映した状態で会話を行うという実験条件である。現在一般的に利用されている遠隔会話手段であるビデオチャットと提案システムを比較するために「ディスプレイ条件」を用意した。

### 3.5.3 実験時の会話タスク

本実験では、会議の機能別の種類として創造会議と決定会議を想定し、会話タスクとして砂漠生き残り問題を用いた[11]。また、会議の参加人数については、創造会議や決定会議に適している3人から10人の小集団会議を想定し、よりシンプルな実験条件にするために最少人数である3人での会議とした。

砂漠生き残り問題とは、会話の参加者が砂漠で遭難したという状況を仮定してもらい、その際に提示する3種類の道具について砂漠で生き残るために必要度合いの優先順位を話し合って決定するという会話タスクである。今回の実験では表 3-4 に示す道具のリストを用いた。

この会話タスクは1回あたり5分間で行う。この会話1回を1セッションと呼ぶ。遠隔参加者となる被験者と実験条件をそれぞれ変更して実験を行う。1グループにつき被験者が3人いるため、「誇張あり条件」で3通り、「誇張なし条件」で3通り、「ディスプレイ条件」で3通りの合計9セッションを行う。

表 3-4 砂漠生き残り問題の道具リスト

セッション1	<input type="radio"/> 防水シート <input type="radio"/> 毛布 <input type="radio"/> 薄手のコート
セッション2	<input type="radio"/> 拳銃 <input type="radio"/> ナイフ <input type="radio"/> フライパン
セッション3	<input type="radio"/> ハサミ <input type="radio"/> ガムテープ <input type="radio"/> ハンマー
セッション4	<input type="radio"/> 「食用に適する砂漠の動植物」という本 <input type="radio"/> その地域の航空地図 <input type="radio"/> 携帯電話
セッション5	<input type="radio"/> ガーゼと包帯 <input type="radio"/> 人数分のサングラス <input type="radio"/> ウォッカ
セッション6	<input type="radio"/> 方位磁石 <input type="radio"/> 消毒薬 <input type="radio"/> 化粧用の鏡
セッション7	<input type="radio"/> 1人につき1リットルの水 <input type="radio"/> チョコレート <input type="radio"/> 食塩
セッション8	<input type="radio"/> ロープ <input type="radio"/> リュックサック <input type="radio"/> 雨傘
セッション9	<input type="radio"/> マッチ <input type="radio"/> 懐中電灯 <input type="radio"/> 赤と白模様のパラシュート

#### 3.5.4 実験環境

実験環境の構成を図 3-3 に示す。3 人参加の多人数会話を想定し、そのうちの 1 人が遠隔参加者として遠隔地を想定した別室からテレプレゼンスロボットを操作して会話に参加するという構成となっている。実際の実験時の会議室と遠隔地の様子をそれぞれ図 3-4 と図 3-5 に示す。「ディスプレイ条件」の際には、会議室に配置したテレプレゼンスロボットをディスプレイに置き換えて実験を行う(図 3-6, 図 3-7)。ディスプレイには遠隔参加者の上半身が映るようにし、テレプレゼンスロボットの大きさとディスプレイ上に映る人の大きさが近くなるようなサイズのディスプレイを用いて設置した。実際に用いたものは 24 インチの液晶ディスプレイ (DEL ST2410b) である。

また、音声の通話には一般に遠隔会議にも用いられている Skype を利用し、音質や音声の遅延なども実際の遠隔会議に近い状態になるようにした。今回の実験では、遠隔参加者の相槌を取得する必要があるため、他の参加者の話し声が外乱として強く作用する可能性があったため、遠隔参加者は会議室側の音声をヘッドフォンで聞くこととした。遠隔地側の音声は Web カメラのマイクロフォンで取得され、会議室側のテレプレゼンスロボットもしくはディスプレイの後ろの左右に配置したスピーカから出力される。

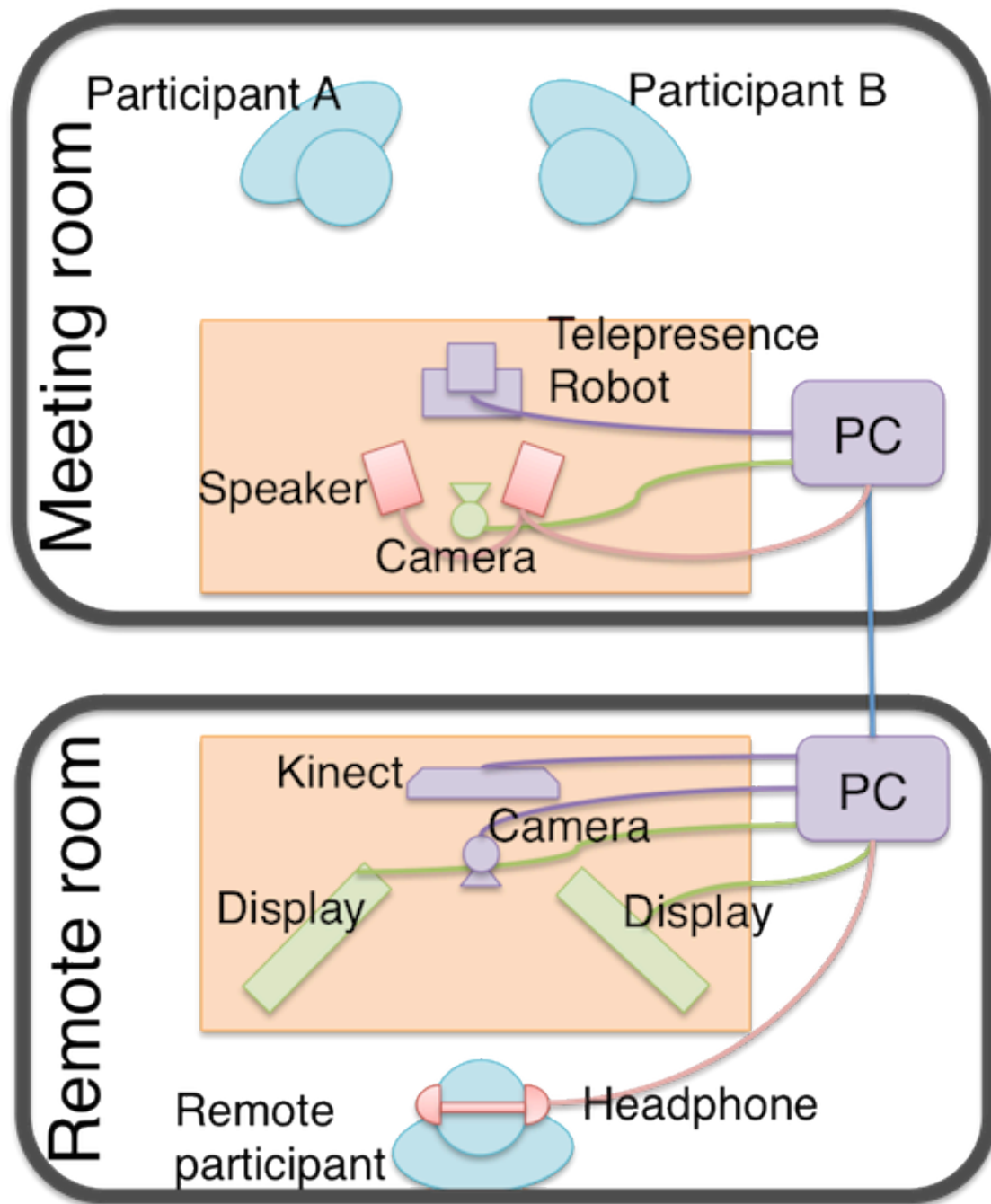


図 3-3 実験環境(誇張あり条件, 誇張なし条件)

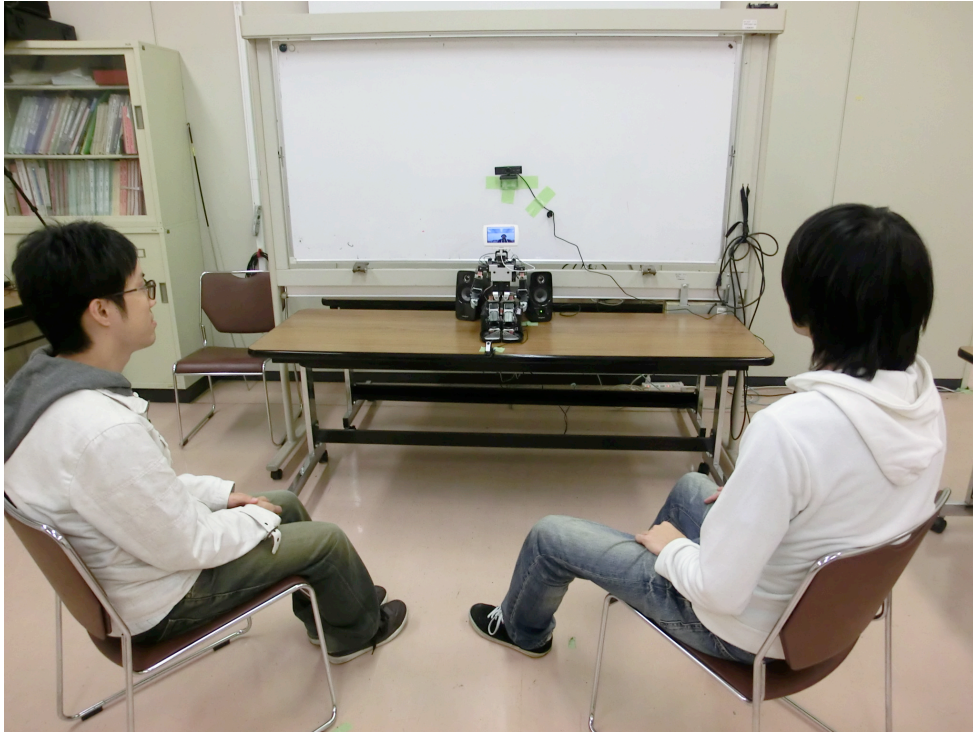


図 3-4 会議室の様子(誇張あり条件, 誇張なし条件)



図 3-5 遠隔地の様子



図 3-6 会議室の様子(ビデオチャット条件)



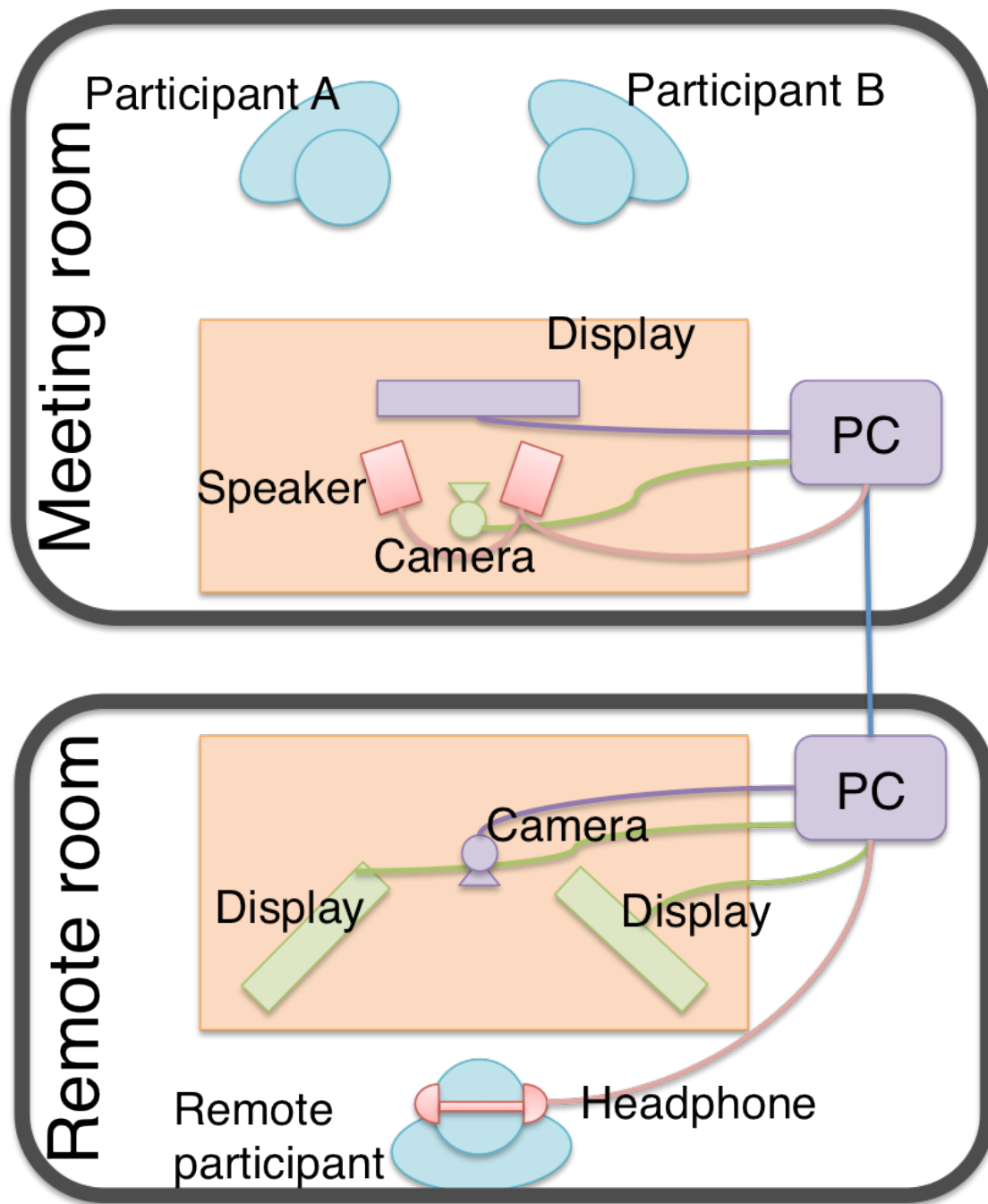


図 3-7 実験環境(ビデオチャット条件)

### 3.5.5 会話後の主観評価アンケート

各会話についての印象や、会話中に会議室側の参加者が遠隔参加者に対して、または、遠隔参加者が会議室側の参加者に対して、どのような印象を抱いたかの主観評価を取得するためにアンケートを行う。会議室側の参加者に遠隔参加者の印象を評価してもらうための質問項目は表 3-5 に示す 6 項目、遠隔参加者に会議室側の参加者の印象を評価してもらうための質問項目は表 3-6 に示す 6 項目を用意した。各質問項目について 7 段階リッカート尺度を用いた回答欄にて回答を行わせる。

表 3-5 会議室側の参加者に遠隔参加者の印象を評価してもらうための質問項目

質問番号	質問項目
M-1	遠隔参加者があなたや話している人を見ていると感じた
M-2	遠隔参加者が話を聞いていると感じた
M-3	遠隔参加者が話に対する反応を返してくれていると感じた
M-4	遠隔参加者が話したそうにしている様子に気付いた
M-5	自分が話をはじめてもいいタイミングがわかった
M-6	遠隔参加者の動きに気づきやすかった

表 3-6 遠隔参加者に会議室側の参加者の印象を評価してもらうための質問項目

質問番号	質問項目
R-1	他の参加者があなたや話している人を見ていると感じた
R-2	他の参加者が話を聞いていると感じた
R-3	他の参加者が話に対する反応を返してくれていると感じた
R-4	他の参加者が話したそうにしている様子に気付いた
R-5	自分が話をはじめてもいいタイミングがわかった
R-6	他の参加者の動きに気づきやすかった

### 3.5.6 実験の流れ

実験の流れを図 3-8 に示す。

まず実験全体に関する説明を行った後、ロボットの操作方法について説明する。その後、3人の被験者にそれぞれに実際にロボットを操作してもらい、1分間ほど操作を練習させる。次に、砂漠生き残り問題についての説明を行い、会話を行う前の各被験者の意見を収集するために、優先順位の事前調査として道具リストに各セッションでの3つずつの道具に優先順位をつけてもらう。

その次に、会話タスクを行う。まず実験者が今回の会話で話し合ってもらった3つの道具を読み上げる。その後、提示した3つの道具について実際の会話を5分間行わせる。実験全体の説明の際に、5分間のうちに3人で優先順位の意見をまとめること、5分経つ前に意見がまとまらなかったとしても、道具の他の使い方がないかなど5分経つまで会話を続けることを被験者に指示した。また、会話開始から4分でベルを1鈴、5分で2鈴鳴らすことにより会話時間の経過を被験者に伝えることとした。なお、被験者が時間の経過を意識しすぎないように

時計やタイマーなどは被験者からは見えない位置に置いた。会話が終了したら、その会話で決定した道具の優先順位を実験者に口頭で伝えてもらい、その後、被験者それぞれに会話後のアンケートを回答させる。これらの流れを1回の会話タスクとし、遠隔参与者訳になる被験者や比較条件を変えて合計9回の会話タスクを行わせる。比較条件の順序については順序効果の影響がキャンセルされるよう考慮した。また、比較条件と話し合う道具の組み合わせが異なるようにした。

実験の最後のアンケートでは、ロボットや操作方法に対する印象や気付いたことについて自由記述形式で回答させる。

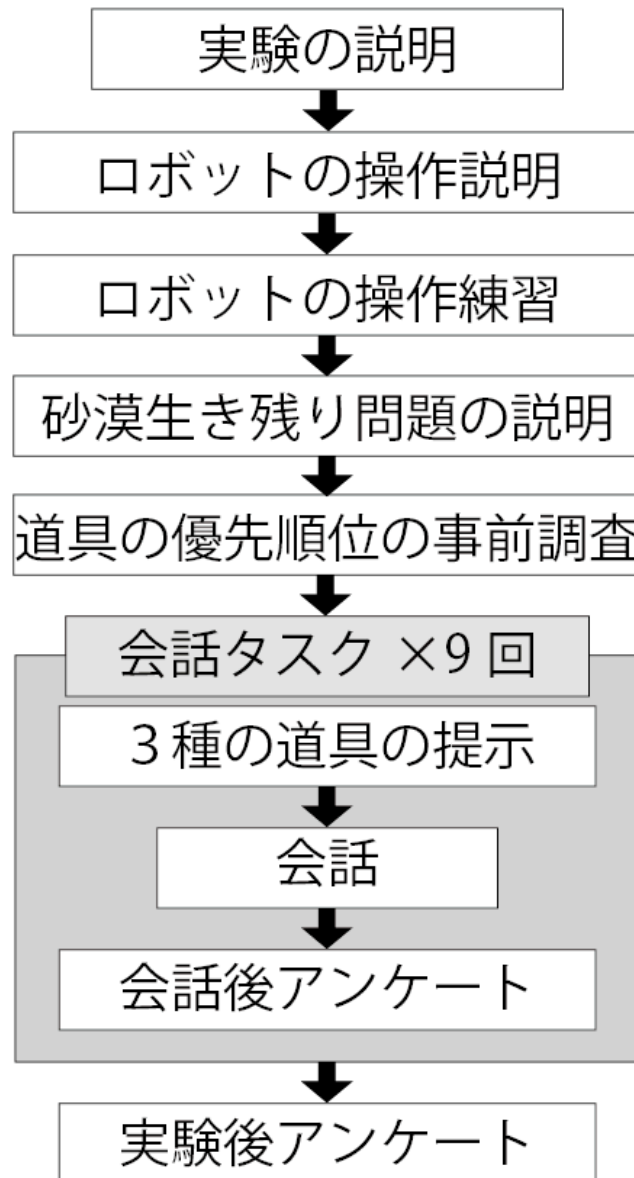


図 3-8 実験の流れ

### 3.5.7 評価方法

実験中の会話の様子を観察することにより 3 条件での会話の評価を行う。評価の指標としては、参加者の発話ターン取得率を用いる。

今回は以下の仮説を立てて検証を行う。

仮説 1: 誇張あり条件は誇張なし条件に比べて遠隔参加者のターン取得率が高い

仮説 2: テレプレゼンスロボットを用いた条件(誇張あり条件と誇張なし条件)はディスプレイ条件に比べて遠隔参加者のターン取得率が高い

仮説 1 については、誇張なし条件では遠隔参加者の減少したままの身ぶりをそのまま伝えるので、他の参加者の注目を十分引くことができず聞き手となったりターンを取得したりすることに不利であると考えられる。一方で、誇張あり条件では減少した身ぶりを誇張するため、誇張なし条件よりも他の参加者の注目を引くことができ、ターンの取得につながると考えたためこのような仮説を立てた。

仮説 2 については、ディスプレイ条件では遠隔参加者は平面上に映るため、存在感や向いている方向がわかりづらくなり会話中に重要な非言語情報の伝達が困難になると考えられる。一方でテレプレゼンスロボットを用いた条件では、ロボットの 3 次元的な存在感と向いている方向のわかりやすさ、身ぶりへの注目のしやすさなどから、遠隔参加者の表出した身ぶりが会話中に効果的に伝達され、その結果、遠隔参加者のターンの取得にも寄与すると考えられるためこのような仮説を立てた。

## 3.6 実験結果と考察

### 3.6.1 実験結果

3組（合計9人）の被験者に対して実験を行い、ビデオカメラを用いて実験時の会話の様子を録画した。被験者は20歳から30歳代の男性8人、女性1人であった。実験後にビデオカメラの映像を見返し、各参加者のターン取得回数のカウントを行った。

各条件における遠隔参加者のターンの取得率を図3-9に示す。グラフのエラーバーは標準偏差を示す。遠隔参加者のターンの取得率の3条件間においてFriedman検定を行った結果、有意差はみられなかった。

次に、各条件におけるセッションごとの総ターン数の平均を図3-10に示す。総ターン数の3条件間においてFriedman検定を行った結果、有意水準1%で有意差がみられた。そこで3条件間のどの組合せに差があるのかより詳細に明らかにするために、多重比較としてBonferroniの調整を用いたWilcoxonの符号順位検定を行った。その結果、誇張あり条件と誇張なし条件の間に有意水準5%の有意差がみられた。このことから、1セッションあたりの総ターン数は、誇張なし条件に比べて誇張あり条件において有意に多くなることが明らかとなった。

各条件における遠隔参加者のターン取得数を図3-11に示す。遠隔参加者のターン取得数の3条件間においてFriedman検定を行った結果、有意水準5%で有意差がみられた。そこで、多重比較としてBonferroniの調整を用いたWilcoxonの符号順位検定を行った。その結果、誇張あり条件と誇張なし条件の間に有意水準5%の有意差がみられた。また、誇張あり条件とディスプレイ条件の間に有意水準10%の有意傾向がみられた。このことから、誇張なし条件と比べて誇張あり条件のほうが遠隔参加者のターン取得数が多くなるといえる。また、ディスプレイ条件と比べた場合には誇張あり条件のほうが遠隔参加者のターン取得数が多くなる傾向がある。

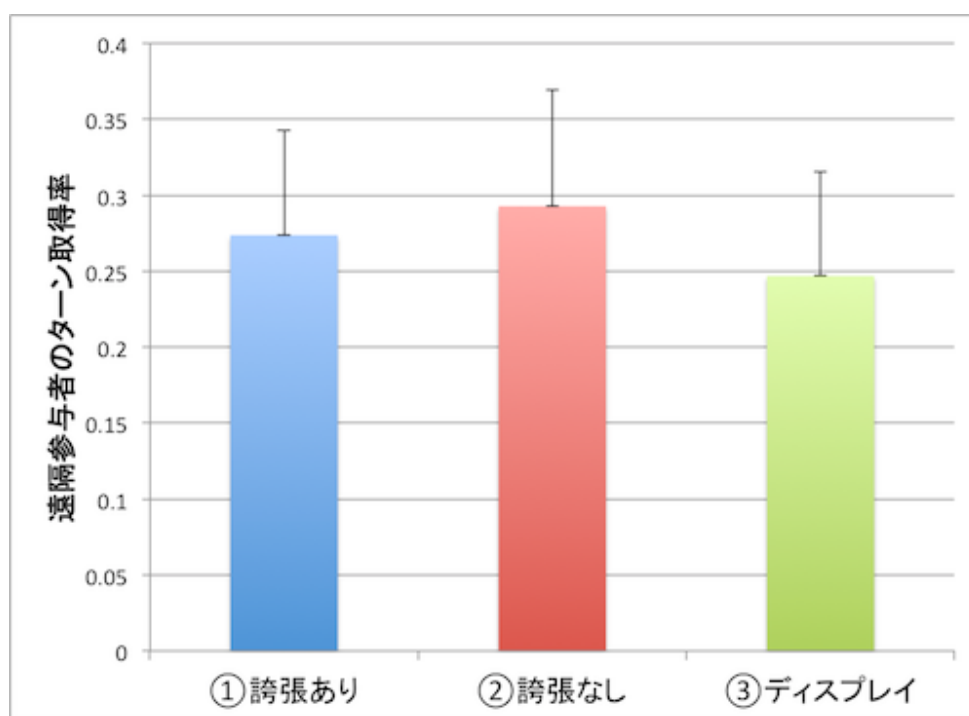


図 3-9 遠隔参加者のターン取得率

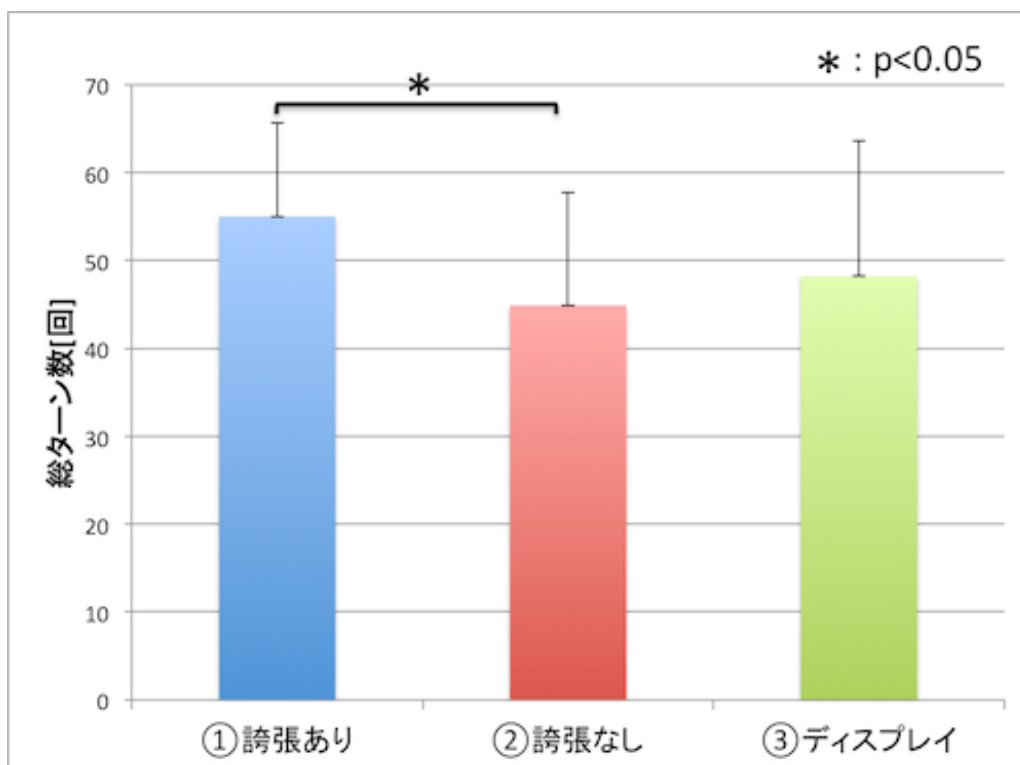


図 3-10 総ターン数の平均

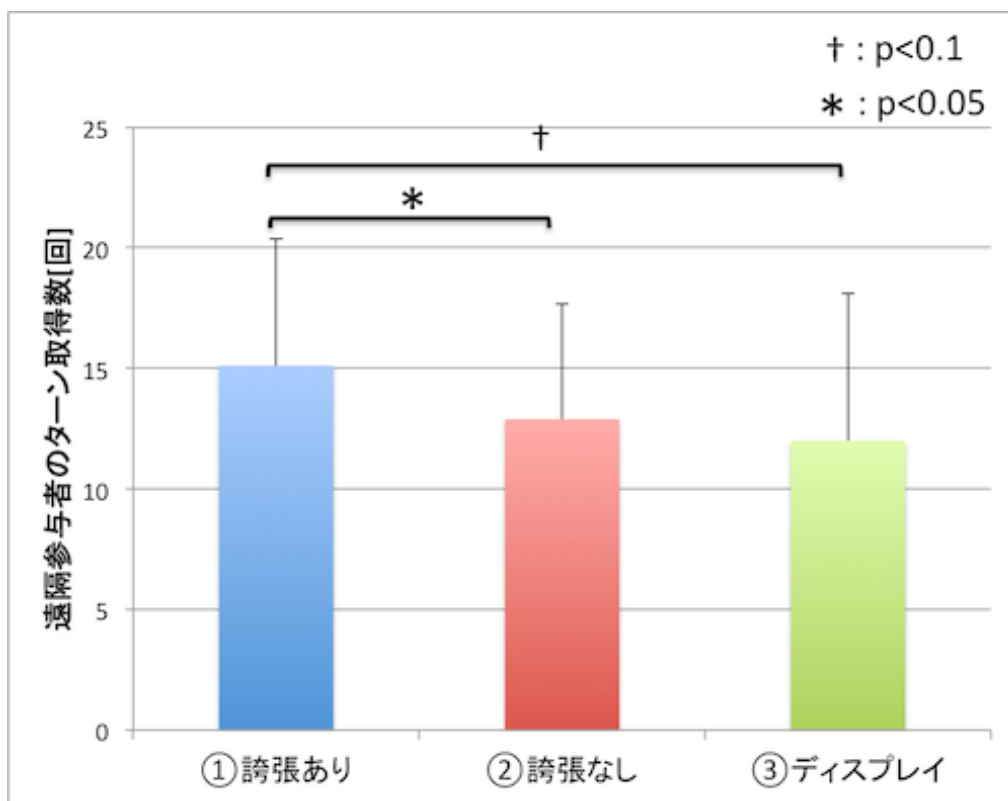


図 3-11 遠隔参加者のターン取得数

また、各条件における主観評価アンケートについて、条件間に有意差の見られた質問項目の結果を列挙する。

まず、会議室側の参加者に遠隔参加者の印象を評価してもらうためのアンケートについて、各質問項目での3条件間でFriedman検定を行った結果、「遠隔参加者があなたや話をしている人を見ていると感じた」「遠隔参加者が話に対する反応を返してくれていると感じた」「遠隔参加者が話したそうにしている様子に気付いた」「遠隔参加者の動きに気づきやすかった」の4つについて有意差がみられた。そこで、多重比較としてBonferroniの調整を用いたWilcoxonの符号順位検定を行った。

「遠隔参加者があなたや話をしている人を見ていると感じた」の質問項目については、誇張あり条件とビデオチャット条件の間に有意水準5%の有意差がみられた(図3-12)。「遠隔参加者が話に対する反応を返してくれていると感じた」の質問項目については、誇張あり条件とビデオチャット条件の間に有意水準10%の有意傾向がみられた(図3-13)。「遠隔参加者が話したそうにしている様子に気付いた」の質問項目については、誇張あり条件とビデオチャット条件の間に有意水準5%の有意差がみられた(図3-14)。「遠隔参加者の動きに気づきやすかった」の質問項目については、誇張あり条件とビデオチャット条件の間に有意水準5%の有意差がみられた。さらに、誇張なし条件とビデオチャット条件の間にも有意水準5%の有意差がみられた(図3-15)。

次に、遠隔参加者に会議室側の参加者の印象を評価してもらうためのアンケートについて、各質問項目での3条件間でFriedman検定を行った結果、「他の参加者が話したそうにしている様子に気付いた」という質問について有意差がみられた。多重比較としてBonferroniの調整を用いたWilcoxonの符号順位検定を行ったところ、誇張あり条件とビデオチャット条件の間に有意水準10%の有意傾向がみられた。さらに、誇張なし条件とビデオチャット条件の間にも有意水準10%の有意傾向がみられた(図3-16)。

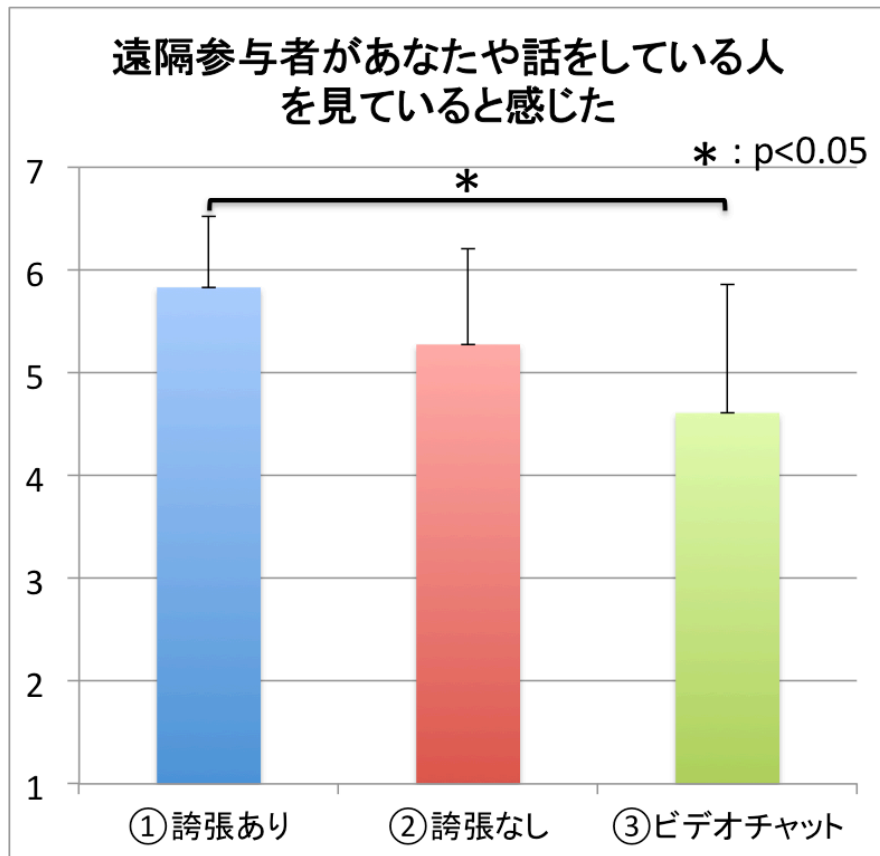


図 3-12 「遠隔参加者があなたや話をしている人を見ていると感じた」の質問項目の結果

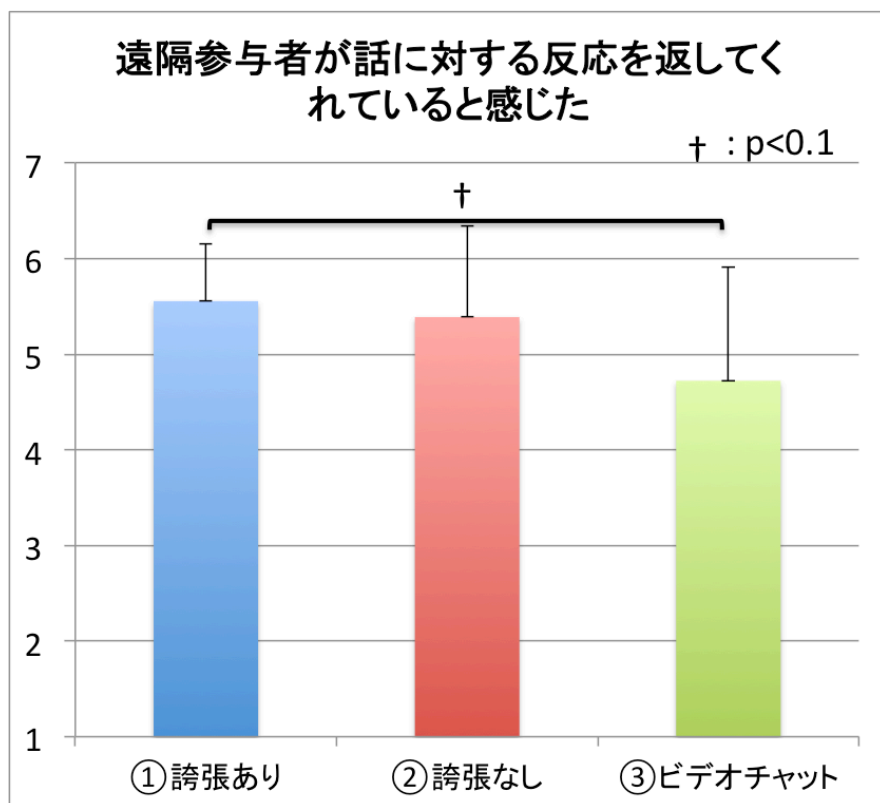


図 3-13 「遠隔参加者が話に対する反応を返してくれていると感じた」の質問項目の結果



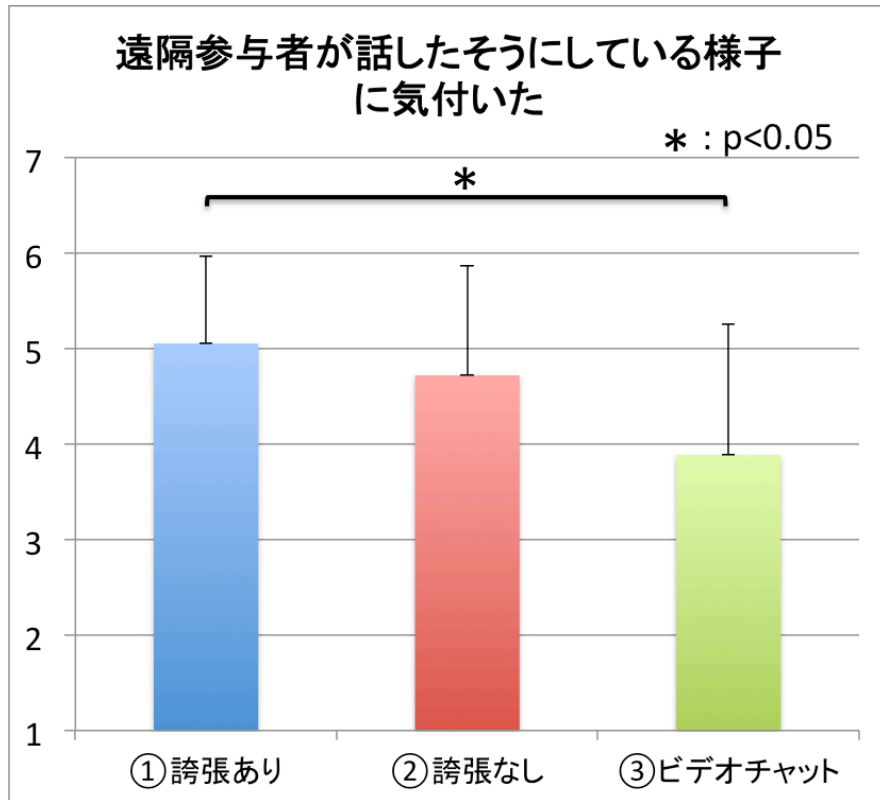


図 3-14 「遠隔参加者が話したそうにしている様子に気付いた」の質問項目の結果

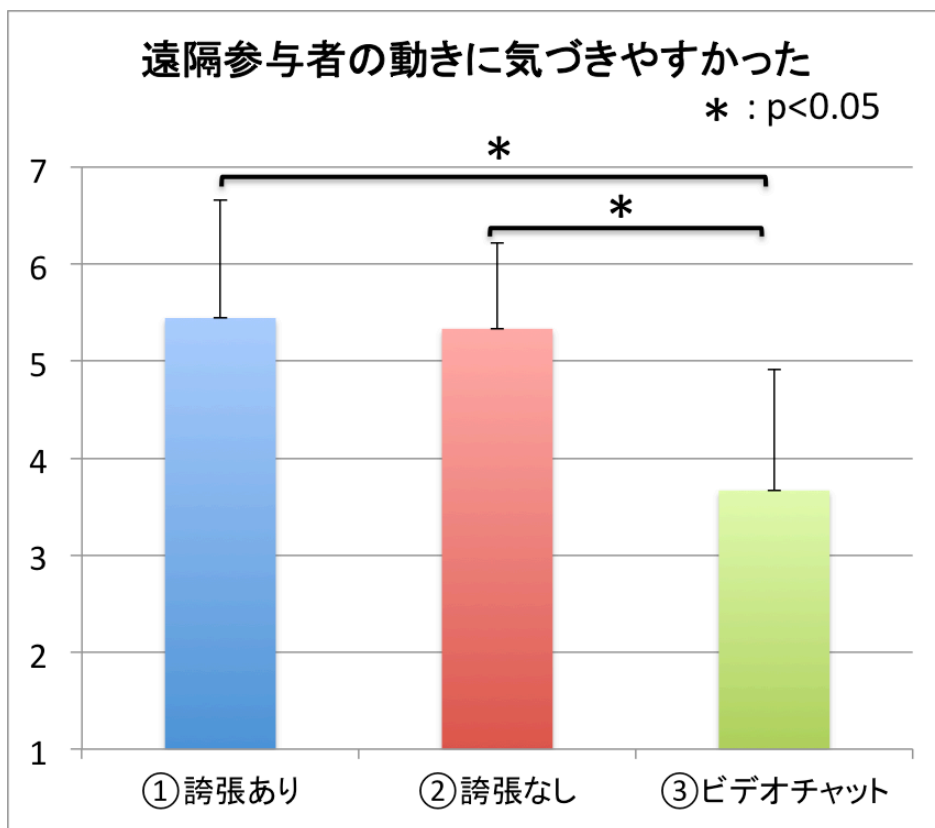


図 3-15 「遠隔参加者の動きに気づきやすかった」の質問項目の結果

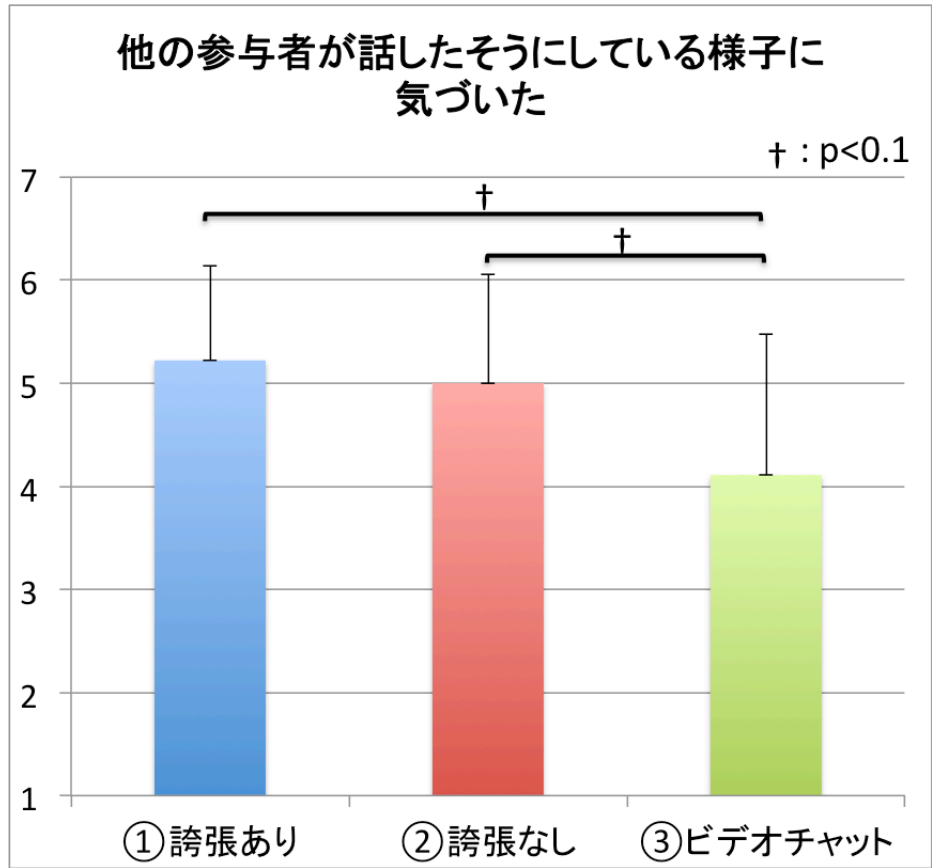


図 3-16 「他の参加者が話したそうにしている様子に気付いた」の質問項目の結果

### 3.6.2 考察

仮説 1 の「誇張あり条件は誇張なし条件に比べて遠隔参加者のターン取得率が高い」については、図 3-9 の結果から、誇張あり条件と誇張なし条件の間に有意な差はみられず、仮説の立証はされなかった。しかしながら、図 3-10 の結果から、誇張なし条件と比較して誇張あり条件のほうがセッションでの総ターン数が有意に多くなっていることがわかる。また、図 3-11 に示された遠隔参加者のターン取得数についても、誇張なし条件よりも誇張あり条件のほうが有意に多くなっている。このことから、誇張なし条件と比べて誇張あり条件ではセッション中のターンのやりとり自体が多く、より活発で盛り上がった会話であったと考えられる。

このように、身ぶりの誇張を行った際にセッション中のターン数が多くなった理由としては、身ぶりを誇張することにより、遠隔参加者の人が話を聞いている、会話にちゃんと参加しているということが会議室側の参加者にとってわかりやすくなったからであると考えられる。この傾向については主観評価アンケート結果から推察できる。

図 3-12 の「遠隔参加者があなたや話をしている人を見ていると感じた」の質問項目にて、有意差はみられなかったものの平均値としては誇張なし条件よりも誇張あり条件のほうが良い評価となっていた。このことから、身ぶりの誇張により遠隔参加者からより見られているという印象を会議室側の参加者に与えていると考えられる。この見られているという印象は頭部方向の誇張を行ったために印象が高くなったと示唆される。

また、図 3-13 の「遠隔参加者が話に対する反応を返してくれていると感じた」の質問項目にて、誇張なし条件よりも誇張あり条件の平均値が高くなっていた。これは、うなずきの大きさの誇張や、相槌の音声にあわせてうなずき動作を行わせる誇張を行ったために、より反応を返してくれているという印象を受けたと考えられる。

さらに、図 3-14 の「遠隔参加者が話したそうにしている様子に気付いた」の質問項目についても、誇張なし条件よりも誇張あり条件の平均値が高くなっていた。これは、前述のように「見られている」印象や「反応を返してくれている」という印象をより与えていたことから、遠隔参加者が参加役割でいう「聞き手」になりやすくなっており、話者から注目されている状態が多くなったために「話したそうにしている様子」に気づいてもらう機会が増えていたのではないかと考えられる。

このように、身ぶりを誇張することにより、遠隔参加者の人が話を聞いている、会話に反応しているということが会議室側の参加者の注意を引きと考えられる。そのため、会話にちゃんと参加しているということが会議室側の参加者にとってわかりやすくなったため、会議室側の参加者が遠隔参加者の様子を伺って待つようなことが少なくなり、セッション中の全体のターン数が多くなったと考えられる。また、全体のターン数が増加しても遠隔参加者のターン取得率が有意に減少しなかった理由としては、遠隔参加者が聞き手の参加役割になりやすくなったためと考えられる。そのため、全体のターン数が増加しても会議室側だけが盛り上がり過ぎてターンの交替を行うのではなく、遠隔参加者のターンの取りやすさは減少せずに会話に参加できていたといえる。

一方で、身ぶりの誇張がない場合は、身ぶりの大きさは小さくなり頻度も減少するため、得られる反応が少なく、ちゃんと話を聞いているのかわかりづらくなる。そのため、遠隔参加者の反応を待つことで会話の活発さが失われたのではないかと考えられる。一方で、誇張あり条件の場合は、遠隔参加者の反応がわかりやすいため、反応を待つような気を使うことなく、より対面会話に近い状態で会話ができたと考えられる。これらのことから、身ぶりの誇張を行うことにより、ターンの取得率を増加させることはできなかったものの、会話全体

の総ターン数を増加させることができたと言える。

仮説 2 の「テレプレゼンスロボットを用いた条件(誇張あり条件と誇張なし条件)はディスプレイ条件に比べて遠隔参加者のターン取得率が高い」については、図 3-9 の結果から、テレプレゼンスロボットを用いた場合とディスプレイを用いた場合で有意な差はみられず、仮説は立証されなかった。しかしながら、平均値としてみるとディスプレイ条件よりもテレプレゼンスロボットを用いた 2 つの条件のほうがターン取得率としては高くなっている。今回は被験者が 9 人と少なかったため統計的な有意差はみられなかったと考えられる。

主観評価アンケート結果をみると、図 3-15 に示す「遠隔参加者の動きに気づきやすかった」の質問項目については、誇張あり条件と誇張なし条件のテレプレゼンスロボットを用いた両方の条件では、ビデオチャット条件とくらべて有意に平均値が高くなっていた。この要因の一つとしては、2 次元的なディスプレイではなく 3 次元的な実体を持つロボットが動作することにより、顔を向けたり体幹を向けたりする動作の方向がわかりやすく、遠隔参加者が自分の方を見ようとする動きにより気づきやすくなったためと考えられる。また他に考えられる要因としてはテレプレゼンスロボットのサーボモータの動作音があげられる。ビデオチャット条件では遠隔参加者の映像をディスプレイ上に映すため、遠隔参加者が動いたとしても衣擦れ音程度の小さな音が出る可能性はあるが特に目立つような音は出ない。一方で、テレプレゼンスロボットを用いた条件の場合には、遠隔参加者の動きをロボットが表出するため、ロボットのサーボモータから動作音が発生する。また、遠隔参加者の動きがない場合であってもロボットの姿勢を維持するためにサーボモータが稼働し続けるために、静止中もサーボモータ動作音が発生し続けるが、動作時にはそれよりも大きい動作音が発生する。このようなロボットの動作に伴ったサーボモータの動作音によって、会議室側の参加者はロボットの動きに気づきやすくなり、それが遠隔参加者の動きに気づきやすいという印象を与える結果となったと考えられる。会話の様子を観察した結果、サーボモータの動作音に声が遮られて聞き返すような場面は見受けられなかったため、ロボットのサーボモータの動作音の大きさは、会話中の声が聞き取りづらくなるほどの影響を与えるものではなかった。そのため、サーボモータの動作音が会話を阻害するような悪影響を与えていた点は見受けられず、動作音によってロボットの動きに気づきやすくなるという点では、会話中の遠隔参加者の動きを際立たせて会議室側の参加者に伝えるメリットがあったと考えられる。

図 3-16 の遠隔参加者から会議室側の参加者の印象を評価した質問項目である「他の参加者が話したそうにしている様子に気付いた」について、誇張あり条件と誇張なし条件のテレプレゼンスロボットを用いた両方の条件では、ビデオチャット条件とくらべて平均値が高くなる傾向がみられた。このことは、上述のようにディスプレイではなくロボットが遠隔参加者の代わりに会議室で動くことにより会議室側の参加者の注意をより集めるため、会議室側の参加者は遠隔参加者の方を見ることが多くなったと考えられる。そして、遠隔参加者がその視線に気付いたために会議室側の参加者が話したそうにしている印象を受けたと考えられる。

また、実験後のアンケートでは、「ロボットを介している方が身ぶり手ぶりが多かったように思う」「ロボットを介している時は、顔の向きで話を聞いていると感じる」「(相手がロボットで参加している場合は) 前フリがあると発言するんだなというのが把握しやすい」「ディスプレイの場合は画面を見ないとどっちを向いているのか分からなかった」というような意見があった。このように、ロボットを用いることにより、その身ぶりや顔の方向がわかりやすくなっているということが伺える。

ロボットを操作した際の感想については、「ロボットが自分の意思にそぐわない動きをした

時に、そちらに注意が向いてしまった」「ロボットがどのようなふるまいをしているのかが詳しく分からず、不安を感じる」「操作のフィードバックがもうちょっと欲しかった」などの意見が得られた。遠隔参加者は、ディスプレイの端にテレプレゼンスロボットの後ろ姿の一部を見ることができるが、全体としてどのような振る舞いをしているのかは把握しづらい状態であったためこのような意見が多かったと考えられる。今後は遠隔参加者側へ提示する情報についても考慮していく必要がある。

## 第4章 結論

本研究では、遠隔会議において発話衝突と遠隔参加者の発話機会が減少するという問題を解決するための、テレプレゼンスロボットの身ぶり伝達方法を提案した。

発話衝突の問題については、無意識的身ぶりを伝達して表出可能なテレプレゼンスロボットを提案し、実験により発話衝突が減少することを確認した。音声と映像のみの遠隔会話の場面では発話予備動作が認知されにくくなり、発話衝突という問題が発生する。そこで、3次元的な実体を持つロボットを介して身ぶりを表出することにより身ぶりが認知されやすくなると考えた。また、身ぶりの中でも無意識的に行っている身ぶりが発話予備動作を有するため話者交替において重要であると考え、提案したのが無意識的身ぶりを表出可能なテレプレゼンスロボットである。そして意識的身ぶりのみを表出する受動型コントローラ操作条件と、提案手法である無意識的身ぶりも表出するモーションキャプチャ操作条件を比較する実験を行い、提案手法が発話衝突を減少させることを確認した。

遠隔参加者の発話機会が減少するという問題については、遠隔参加者の身ぶりが減少することが要因だと考え、遠隔参加者の身ぶりを誇張するテレプレゼンスロボットを提案した。提案手法である誇張あり条件と既存手法である誇張なし条件、既存のビデオチャットの状態を模したディスプレイ条件を比較する実験を行った。その結果、身ぶりの誇張により遠隔参加者のターンの取得率を増加させるには至らなかったものの、誇張なし条件とくらべて会話全体のターン数が増加する結果となった。

今後の課題としては、会議室側の遠隔参加者の存在をロボットで補強するだけでなく、遠隔参加者への情報提示を工夫することにより、遠隔参加者が疎外感を感じることなく会話に参加できるようにシステムを発展していくことが考えられる。発展の方向性としては、遠隔地側にもロボットを置き会議室側の参加者の身ぶりを伝達することや遠隔参加者に提示する映像上で人の特定の動きを強調するエフェクトを重畳表示させるなどが考えられる。それらを実現させるために、発話交替に関わりの強い特徴的な身ぶりを明らかにする必要があると考えられる。それ元、特徴的な身ぶりの表出に特化したロボットの設計や、それらの身ぶりの映像上での強調が可能になると考えられる。

## 謝辞

本論文執筆にあたり，多くの方の御指導，御助言，御協力をいただきました．ここに感謝の意を表します．

指導教員である本学システム情報系の中内靖教授には，多大な御指導・御助言を頂きました．学類の卒業研究の頃より6年間，研究の細部にわたり丁寧な御指導・御助言を賜りました．また，学会発表の機会を多く与えてくださり，見聞を広め貴重な経験を積むことができました．ここに深く御礼申し上げます．

本学システム情報系の葛岡英明教授，鈴木健嗣准教授，田中文英准教授，人間系 原田悦子教授には本研究を進めるにあたり，貴重な御意見・御指導を頂きました．ここに深く御礼申し上げます．

筑波大学知能機能システム専攻ヒューマン・ロボット・インタラクション研究室の諸氏には，研究打ち合わせやシステム開発，実験などを通じて多くの御協力ならびに有意義なご指摘・ご助言を頂きました．

以上の皆様に対して，ここに深く感謝いたします．

## 参考文献

- [1] Adalgeirsson, S.O. and Breazeal, C., MeBot: a robotic platform for socially embodied telepresence, Proceedings of Human-Robot Interaction 2010 (HRI2010), pp.15-22, 2010.
- [2] Anybots, QB (online), available from <<https://www.anybots.com>>, (参照日 2015年1月5日).
- [3] Cassell, J., McNeill, D. and McCullough, K.E., Speech-gesture mismatches: evidence for one underlying representation of linguistic and nonlinguistic information, Pragmatics and Cognition, Vol.7, No.1, pp.1-33, 1999.
- [4] Clark, H. H., Using Language, Cambridge University Press, 1996.
- [5] Goffman, E., Replies and Responses, Language in Society, No.5, pp.257-313, 1982.
- [6] Goodwin, C.: Conversational Organization: Interaction between Speakers and Hearers, Academic Press, 1981.
- [7] 井上順博, 妻木勇一, ウェアラブルミニチュアヒューマノイドロボットの開発, ロボティクス・メカトロニクス講演会'08 講演論文集 (2008), 2A1-B21.
- [8] InTouch Technologies, Inc., RP-7 (online), available from <<http://www.intouchhealth.com>>, (参照日 2015年1月5日).
- [9] 石黒浩, アンドロイドサイエンス人間を知るためのロボット研究, 毎日コミュニケーションズ, 2007.
- [10] 伊藤禎宣, 岩澤昭一郎, 土川仁, 康之, 間瀬健二, 片桐恭弘, 小暮潔, 萩田紀博: 装着型体験記録装置による対話インタラクションの判別機能実装と評価, ヒューマンインタフェース学会誌, Vol.7, No.1, pp.167-178, 2005.
- [11] J. C. Lafferty, and P. M. Eady: The desert survival problem, Experimental Learning Methods, Plymouth, MI, 1974.
- [12] 片山貴信, 武川直樹, 徳永弘子, 湯浅将英, 多人数映像会話における話し手の身振りとアクティビティの関係-視線一致と不一致環境により会話の質はどのように変わるか?-, 電子情報通信学会技術研究報告, Vol.108, No.487, pp.121-126, 2009.
- [13] Katzenmaier, M. Stiefelhagen, R. Achults, T.: Identifying the Addressee in Human-Human-Robot Interactions based on Head Pose and Speech, Proceedings of the 6th international conference on Multimodal interfaces, pp.13-15, 2004.
- [14] Kendon, A.: Some Functions of Gaze-Direction in Social Interaction, Acta Psychologica, Vol.26, pp.22-63, 1967.
- [15] Kidd, C., Breazeal, C.: Effect of a Robot on User Perceptions, Proceedings of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2004), Vol.4, pp.3559-3564, 2004.
- [16] Martha W. Alibali, Sotaro Kita, Amanda J. Young: Gesture and the process of speech production: We think, therefore we gesture, Language and Cognitive Processes, Vol.15, No.6, pp.593-613, 2000.
- [17] Matsui, D., Minato, T., MacDorman, K. F. and Ishiguro, H., Generating natural motion in an android by mapping human motion, Proceeding of IEEE/RSJ International Conference on Intelligent Robots and Systems, pp.1089-1096, 2005.
- [18] Mehrabian, A., Nonverbal communication, Aldine-Atherton, 1972.
- [19] Microsoft Corp.: Kinect (online), available from <<http://www.xbox.com/ja-JP/kinect>> (参照日 2015年1月5日).
- [20] Morrel-Samuels, Palmer Krauss, Robert M.: Word familiarity predicts temporal asynchrony of hand gestures and speech, Journal of Experimental Psychology: Learning Memory and Cognition, Vol.18, No.3, pp.615-622, 1992.
- [21] Nabe, S., Kanda, T., Hiraki, K., Ishiguro, H., Kogure, K. and Hagita, N., Analysis of human behavior to a communication robot in an open field, Proceedings of Human-Robot Interaction



- 2006 (HRI2006), pp.234-241, 2006.
- [22] 大塚和弘, 熊野史郎, 松田昌史, 大和淳司: MM-Space:頭部運動の物理的補強表現に基づく会話場再構成, 情報処理学会論文誌, Vol. 54, No.4, pp.1450-1461, 2013.
- [23] 小山慎哉, 上坂純一, 葛岡英明, 山崎敬一: ロボットを介した遠隔コミュニケーションに関する研究, 計測自動制御学会システムインテグレーション部門 講演会 SI2002, pp.93-94, 2002.
- [24] Paulos, E. and Canny, J., Social Tele-embodiment: Understanding presence, *Autonomous Robots*, Vol.11, No.1, pp.87-95, 2001.
- [25] Reidsma, D., Akker, R., Rienks, R., Poppe, R., Nijholt, A., Heylen, D., Zwiers, J. : Virtual Meeting Rooms: From Observation to Simulation, *AI & Society*, Vol.22, No.2, pp. 133-144,2007.
- [26] 齋藤洋典, 喜多壮太郎 編著: 認知科学の探究ジェスチャー・行為・意味, 共立出版, 2002.
- [27] 坂本大介, 神田崇行, 小野哲雄, 石黒浩, 萩田紀博: 遠隔存在感メディアとしてのアンドロイド・ロボットの可能性, 情報処理学会論文誌, Vol.48, No.12, pp.3729-3738, 2007.
- [28] Seeing Machines: faceAPI (online), available from <<http://www.seeingmachines.com/product/faceapi/>> (参照日 2015年1月5日).
- [29] Shimoi, M., Kanda, T., Ishiguro, H. and Hagita, N., Interactive humanoid robots for a science museum, *Proceedings of Human-Robot Interaction 2006 (HRI2006)*, pp.305-312, 2006.
- [30] Shinozawa, K., Naya, F., Yamato, J., Kogure, K.: Differences in Effect of Robot and Screen Agent Recommendations on Human Decision-Making, *International Journal of Human-Computer Studies*, Vol. 62, pp. 267-279, 2005.
- [31] Stiefelhagen, R., Yang, J., Waibel, A.: Modeling Focus of Attention for Meeting Index Based on Multiple Cues, *IEEE Transactions on Neural Networks*, Vol.13, No.4, pp.928-938, 2002.
- [32] 鈴木雄介, 福島寛之, 深澤伸一, 竹内晃一: 参加者の身体性を表現する遠隔会議支援ロボットシステムの評価, 情報処理学会研究報告.GN, Vol.2008, No.91, pp.113-118, 2008.
- [33] 鈴木雄介, 福島寛之, 深澤伸一, 竹内晃一, 遠隔会議支援ロボットシステムの注意喚起能力評価, 情報処理学会論文誌, Vol.51, No.1, pp.25-35, 2010.
- [34] 館暲, 阿部稔: テレイグジスタンスの研究 (第1報), 第21回 SICE 学術講演会, pp.167-168, 1982.
- [35] 館暲: 人工現実感, 日刊工業新聞社, 1992.
- [36] 多田隈理一郎, 梶本裕之, 川上直樹, 館暲: テレイグジスタンスの研究 (第36報), TELESAR II スレーブアームの開発, 計測自動制御学会第4回システムインテグレーション部門学術講演会 (SI2003), pp.592-593, 2003.
- [37] 高橋誠, 会議の進め方, 日経文庫, 1987.
- [38] 玉木秀和, 東野豪, 小林稔, 井原雅行, Web 会議における話者交替円滑化手法の検討, 画像電子学会 VMA 研究会, Vol.29, pp.9-18, 2011.
- [39] 玉木秀和, 中茂睦裕, 東野豪, 小林稔, 人のコミュニケーションリズムに着目した Web 会議円滑化手法, IEICE Technical Report MVE2009, pp.101-106, 2009.
- [40] Tojo, T., Matsusaka, Y., Ishii, T., Kobayashi, T., "A conversational robot utilizing facial and body expressions," in 2000 IEEE International Conference on Systems, Man, and Cybernetics, No.2, 2000.
- [41] 友野達也, 古澤洋将, 葛岡英明, 山下淳: コミュニケーションロボットと実世界ビデオアバタシステムにおける遠隔作業指示の比較に関する研究, 日本バーチャルリアリティ学会第11回大会論文集, pp.30-32, 2006.
- [42] Vargas, M.F., *Louder than words: an introduction to nonverbal communication*, Iowa State University Press, 1986.
- [43] Vertegaal, R., Slagter, R., Veer, G., Nijholt, A.: Eye Gaze Patterns in Conversations: There is More to Conventional Agents than Meets the Eyes, *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pp.301-308, 2001.

- [44] W.ラフリー・エンジェル著, 本名信行, 井出祥子, 谷林真理子編訳: 「ノンバーバル・コミュニケーション」, 大修館書店, 1981.
- [45] 鎧沢勇, 滝川啓, 大久保栄, 渡辺義郎, 衛星通信を利用した画像会議におけるエコー及び伝搬遅延の影響, 電子通信学会論文誌 B, Vol.64, No.11, pp.1281-1288, 1981.
- [46] 吉崎航, 加賀美聡, Kinect を用いた人型ロボットの全身制御, ロボティクス・メカトロニクス講演会'11 講演論文集, 2P2-L05, 2011.

# 著者公開論文リスト

## 査読のある学術雑誌

1. 長谷川孔明, 中内靖: テレプレゼンスロボットによる無意識的身ぶりの表出が発話交替に与える影響, 日本機械学会論文集, Vol.80, No.819, 2014.

## 査読のある国際会議論文

1. Komei Hasegawa, Yasushi Nakauchi: Preliminary Evaluation of a Telepresence Robot Conveying Pre-motions for Avoiding Speech Collisions, Proceedings of The 1st International Conference on Human-Agent Interaction (HAI2013), No. III-1-1, 2013.
2. Komei Hasegawa, Yasushi Nakauchi: Facilitation of Telepresence Robot Turn-Takings by Gesture Exaggeration, Proceedings of the 2014 IEEE/SICE International Symposium on System Integration, pp.650-654, 2014.

## 国際会議発表

1. Komei Hasegawa, Yasushi Nakauchi: Telepresence Robot Conveying Pre-motions for Avoiding Speech Collisions in Teleconference, Proceedings of The 22nd IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN2013), TuA1.2P.27, 2013.
2. Komei Hasegawa, Yasushi Nakauchi: Telepresence Robot that Exaggerates Non-verbal Cues for Taking Turns in Multi-party Teleconferences, Proceedings of The 2nd International Conference on Human-Agent Interaction (HAI2014), pp.293-296, 2014.

## 国内会議発表

1. 長谷川孔明, 中内靖: 遠隔会議ロボットのための非言語的な意思伝達に関する研究, 計測自動制御学会第 11 回システムインテグレーション部門講演会(SI2010)予稿集, No.1E1-6, pp.273-276, 2010.
2. 長谷川孔明, 中内靖: 遠隔会議におけるロボットを介した意思表示に関する研究, 日本機械学会ロボティクス・メカトロニクス講演会 2010(ROBOMECH2010)予稿集, No.2A2-B06, 2010.
3. 長谷川孔明, 中内靖: 非言語的な意思伝達を行う遠隔会議ロボットの提案, 第 28 回日本ロボット学会学術講演会(RSJ2010)予稿集, No.RSJ2010AC3C2-7, 2010.
4. 長谷川孔明, 中内靖: 無意識的身ぶりを表出可能なテレプレゼンスロボットの提案, 計測自動制御学会第 12 回システムインテグレーション部門講演会(SI2011)予稿集, No.1C3-1, pp.199-201, 2011.
5. 長谷川孔明, 中内靖: 無意識的身ぶりを表出するテレプレゼンスロボットの提案, HAI シンポジウム 2011 予稿集, No.III-1B-2-L, 2011.
6. 長谷川孔明, 中内靖: 遠隔対話ロボットによる無意識的身ぶりの表出と発話衝突回避について, 計測自動制御学会第 13 回システムインテグレーション部門講演会(SI2012)予稿集, No.1A3-3, 2012.
7. 長谷川孔明, 中内靖: テレプレゼンスロボットの身ぶりが発話交替に与える影響について, HAI シンポジウム 2012 予稿集, No.2B-2, 2012.

8. 長谷川孔明, 中内靖: 遠隔多人数会話における発話機会獲得のための身ぶり誇張テレプレゼンスロボットの提案, 第 32 回日本ロボット学会学術講演会(RSJ2014)予稿集, No.RSJ2014AC2E2-05, 2014.
9. 長谷川孔明, 中内靖: 遠隔参加者の発話機会増加のために身ぶりを誇張する遠隔会話ロボットの提案, 電子情報通信学会技術研究報告(クラウドネットワークロボット研究会), Vol. 114, No. 227, pp. 5-10, 2014.