

# ロボット会議：物理的実体を介した身体動作の提示による ソーシャルテレプレゼンスの強化

田中 一品<sup>1,2</sup> 中西 英之<sup>1,a)</sup> 石黒 浩<sup>3</sup>

受付日 2015年4月6日, 採録日 2015年10月2日

**概要：**ロボット会議とは、ロボットを介して遠隔地にいる相手と対話する新たな遠隔会議の方法である。映像ではなく、物理的な実体を持つロボットで身体動作を提示することで、従来の遠隔会議よりも対面に近い感覚で会話できることが期待されている。本研究の目的は、遠隔会議における実体の効果を検証し、実体を持たない従来の遠隔会議に対するロボット会議の優位性の有無を明らかにすることである。我々は、様々なメディアを介して対話する被験者実験を実施し、従来の遠隔会議とロボット会議を比較した。その結果、実体で身体動作を伝達すると対話相手と対面で会話している感覚が強化されることが分かった。さらに、ロボット会議は対面の会話で感じるような緊張感を生み出す可能性も示唆された。

**キーワード：**ロボット, アバター, テレプレゼンス, 緊張感, ビデオ会議, 音声対話, 対面会話

## Robot Conferencing: Presenting Physically Embodied Body Motions Enhances Social Telepresence

KAZUAKI TANAKA<sup>1,2</sup> HIDEYUKI NAKANISHI<sup>1,a)</sup> HIROSHI ISHIGURO<sup>3</sup>

Received: April 6, 2015, Accepted: October 2, 2015

**Abstract:** Robot conferencing is a new communication method which people talk with a remote conversation partner through a robot. This method might produce the feeling like a face-to-face meeting more than previous methods by presenting the bodily motions of the partner with a physical embodiment. The purpose of this study is to clarify the effects of physical embodiment on distant communication to investigate the superiority of robot conferencing to previous methods. We conducted an experiment in which subjects talked with a partner through various media to compare the robot conferencing with existing media, e.g., video, avatar and voice chats, that has no physical embodiment. As a result, we observed that the physical embodiment enhanced social telepresence that is the sense of resembling face-to-face interaction. However, it was also implied that the physical embodiment produces the nervousness as in the case of a face-to-face meeting.

**Keywords:** robot, avatar, telepresence, nervousness, videoconferencing, audio communication, face-to-face

### 1. はじめに

電話のような音声のみの対話では、相手のソーシャルテレプレゼンス（遠隔地にいる相手と対面している感覚の度合 [6]）が希薄になるという問題がある。相手の映像を提示するビデオ会議はソーシャルテレプレゼンスを強化する最

も一般的な方法であることが知られている [5], [10]。近年では、ビデオ会議は面接や相談サービスなど初対面の相手との会話にも使用される一般的な遠隔対話メディアになりつつある。しかし、ビデオ会議で感じられる存在感は、対面と比較するとまだ十分とはいえない。

遠隔会議をより対面での会話に近づけるため、ロボット会議が研究され始めた。ロボット会議とは、映像ではなく物理的実体（以下、実体と呼ぶ）を持つロボットで対話相手の身体動作を伝えるものである。対話相手の身体の代替となるロボットがユーザと同じ場所に物理的に存在するという実体の効果によって、より対面に近い感覚で対話できることが期待される。いくつかの研究でビデオ会議に対するロボット会議の優位性が示されており [15], [20], 人に酷

<sup>1</sup> 大阪大学大学院工学研究科知能・機能創成工学専攻  
Department of Adaptive Machine Systems, Osaka University, Suita, Osaka 565-0871, Japan

<sup>2</sup> 独立行政法人科学技術振興機構, CREST  
CREST, Japan Science and Technology Agency, Chiyoda, Tokyo 102-0076, Japan

<sup>3</sup> 大阪大学大学院基礎工学研究科システム創成専攻  
Department of Systems Innovation, Osaka University, Toyonaka, Osaka 560-8531, Japan

a) nakanishi@ams.eng.osaka-u.ac.jp

似した外見を持つロボットを介したロボット会議はビデオ会議やボイスチャットよりも高いソーシャルテレプレゼンスを生み出すことが報告されている [20]. しかし, 対話相手に合わせてリアルな外見を持つロボットを1つ1つ製作することはコストの面から現実的ではない. そこで, 人と分かる顔を持つが特定の人物には酷似していない万人向けの遠隔対話用ロボットが開発されている [19]. しかしながら, そのようなリアルな外見を持たないロボットを用いる場合でも, ビデオ会議に対する優位性があるかどうかは明らかになっていない. 本研究では, ロボット会議のどのような要因がソーシャルテレプレゼンスを強化するのかを明らかにするため, 実体の有無と伝達する情報の種類に要因を分けてロボット会議と従来の遠隔会議を比較した.

ロボット会議と同様に対話相手の身体動作を伝達するメディアとして, 近年, avatar Kinect に代表されるようなアバタチャットが容易に利用できるようになった. アバタチャットは, 対話相手の外見を直接伝えずに身体動作のみを伝える点においてロボット会議と類似しているが, 実体を持たない点において異なっている. アバタチャットの効果を調査した研究が多く行われており [3], [7], [11], [23], アバタチャットやビデオ会議では, 音声のみのメディアと比較して, ソーシャルテレプレゼンスが強化されることや [7], [11], 途切れの少ない滑らかな発話になること [23] が報告されている. これらの効果はロボット会議においても得られる可能性があるが, ロボットを用いることは映像を用いるよりもコストがかかるため, 実体があることの有効性が示されなければロボット会議の価値は損なわれる. 我々の先行研究では, ロボットの持つ実体が遠隔会議にどのような影響を与えるのかソーシャルテレプレゼンスの側面から調査を行った [24]. 本研究では, ソーシャルテレプレゼンスに加え, 発話の円滑さの側面からも調査し, これらの関係性について分析した.

## 2. 関連研究

ロボット会議に関する研究は多く行われており, 様々な遠隔対話用ロボットが提案されている. ロボットらしいデザインの頭部や [12], [15], 人間らしい頭部を持つもの [19], [20], また, ユーザの頭部の映像を表示するものがある [14], [22]. 先行研究では, これらのロボットを通して顔の動きを伝達する効果について主に研究されてきた.

ロボット会議におけるソーシャルテレプレゼンスは, 音声のみの対話やビデオ会議よりも高いことが示されている [20]. しかし, この先行研究で用いられたロボットは, 人に近いリアルな外見を持っており, 実体で身体動作を提示したことで, リアルな外見を持っていること, どちらの要因がソーシャルテレプレゼンスを強化したのかは明らかにされていない. また, このロボットは人の全身が再現されていたのに対し, ビデオ映像は頭部のみを表示していた.

頭部だけの映像はソーシャルテレプレゼンスを低下させることが分かっており [18], 等身大の映像を使用したビデオ会議に対してもロボット会議の優位性があるかどうかは不明である. 本研究では, これらの疑問を解明するため, 特定の人物の外見を持たない遠隔対話用ロボット [19] を用い, 等身大の遠隔対話メディアと比較した.

ビデオ会議に関する先行研究では, 対話相手の映像を表示したディスプレイを物理的に移動させる方法で対話相手の前後移動を提示すると, ソーシャルテレプレゼンスが強化されることが報告されている [17]. また, 人とインタラクションを行う自律ロボットの研究分野では, ロボットの持つ実体の効果が検証されており, 実体のあるロボットの動きは映像で提示されたエージェントの動きよりも高い存在感を生むことが分かっている [2], [13]. 遠隔対話用ロボットにおいても, 実体をともなった身体動作の提示がソーシャルテレプレゼンスを強化する可能性がある.

ビデオ会議に対するロボット会議の優位性は観察データからも示されており, 対話相手の視線をロボットで提示すると, 映像で提示するよりも相手の状態が把握しやすくなること報告されている [15]. ロボットの頭部によって視線が明確化されることは既知であるため [15], [22], 我々は音声対話への影響に着目した. 遠隔対話メディアの違いによる音声対話への影響を調べた研究では, 主にビデオ会議における会話構造(話者交代や同時発話など) [1], [4], [21] が観察されてきた. これは, 多人数会話においてビデオ映像が話者を特定するうえで有効に働くためであり, 二者間の対話ではメディアの違いによる影響は示されていない. 一方, 話者の発話に含まれる途切れを観察した研究では, 二者間の対話において映像による身体動作の伝達が途切れを減少させ, 発話を円滑化することが示されている [23]. 本研究では, 実体をともなった身体動作の伝達がさらに発話を円滑化するかを調べるため, 途切れの頻度および, 途切れの時間的割合を観察した. これらの具体的な測定方法については, 3.4.2 項で詳しく説明する.

ビデオ会議に対するロボット会議の優位性を示した前述の先行研究 [15], [20] では, ロボット会議やビデオ会議を単一の要因として扱っている. これに対し, 実体と, 伝達情報の要因に分けて遠隔対話メディアを比較することが本研究の大きな特徴である (図 1).

## 3. 実験

### 3.1 仮説

本研究では, 従来の遠隔会議に対するロボット会議の優位性を検証する実験を行った. ロボット会議の特徴は, 対話相手の身体動作を伝達すること, その身体動作を実体で提示することである. 我々はこれらの特徴がソーシャルテレプレゼンスを強化すると考え, 以下の仮説を立てた.

**仮説 1:** 実体はソーシャルテレプレゼンスを強化する.

**仮説 2:** 身体動作の伝達はソーシャルテレプレゼンスを強

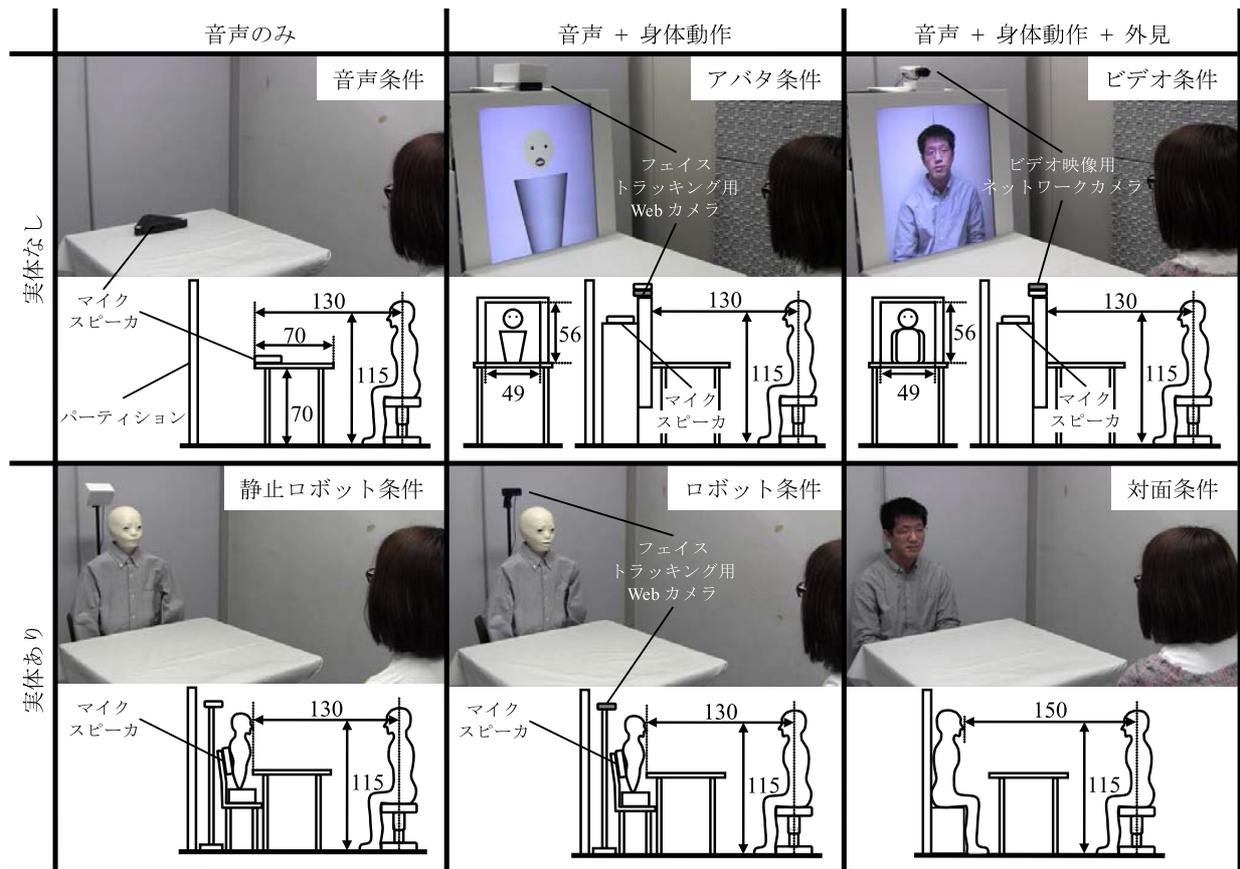


図 1 実験条件と実験環境 (単位: cm)

Fig. 1 Experimental conditions and setups (length unit: centimeters).

化する。

遠隔対話メディアの違いが発話の円滑さに与える影響を調査した先行研究 [23] では、対話相手の身体動作を伝達するアバタチャットやビデオ会議は、音声のみを伝達するメディアと比較して、発話に含まれる途切れを減少させることが報告されている。さらに、アバタチャットやビデオ会議におけるソーシャルテレプレゼンスは、音声のみのメディアと比較して高いことも報告されている [7], [11]。我々は、ソーシャルテレプレゼンスの強化が途切れを減少させると仮定し、ロボット会議の特徴が発話を円滑にすると考えた。途切れに関する仮説は以下のとおりである。

仮説 3: 実体は途切れを減少させる。

仮説 4: 身体動作の伝達は途切れを減少させる。

### 3.2 実験条件

前節で述べた仮説を検証するため、実体の有無と伝達情報の違い (音声のみ, 音声+身体動作, 音声+身体動作+外見) の要因に分けた 2×3 の 6 つの実験条件を設定した (図 1)。遠隔会議の状況として被験者が不自然に感じることで不利になる条件を可能な限りなくすため、一般的に使用されている遠隔対話メディアをロボット会議の比較条件として設定した。実体のあるロボット会議と実体のないアバタチャットは、ともに対話相手の外見を伝えず身体

動作を伝えるメディアである。したがって、我々は、アバタチャットに実体を加えるとロボット会議になると仮定した。同様に、ビデオ会議に実体を加えると対面になると仮定した。また、ボイスチャットに実体を加えたメディアとして静止したロボットを設定した。これらのメディアが伝達する情報に関しては、ボイスチャットと静止ロボットは音声のみ、アバタチャットとロボット会議は音声と身体動作、ビデオ会議と対面は音声、身体動作、外見である。

予備実験において、カメラを用いない実験条件であってもどこから実験者に見られているのではないかと疑った被験者がいたため、対話環境が双方向であることを説明した。アバタ条件、ロボット条件、ビデオ条件では、実験環境が双方向であることを被験者に信じさせるために、各条件を実施する前に実験者から自分がどのように見えているか (アバタ、ロボット、ビデオ映像) を 7 インチのディスプレイで提示した。また、アバタやロボットが自動で動いているのではないかと疑った被験者もいたため、双方向の確認の際に、アバタ条件とロボット条件では、実験者と被験者自身の顔の動きがアバタやロボットで実際に再現されていることも確認させた。これらの確認に使用したディスプレイは対話に使用しないため実験の前に被験者から見えない位置に移動させた。音声の伝達には全条件で同じソフトウェア (Skype) を用いており、音声の遅延は約 100 ms

であった。ビデオ映像用ネットワークカメラやフェイストラッキング用 Web カメラは、使用しない条件ではカバーで隠した。各条件の詳細は下記のとおりである。

**対面条件 (実体あり, 音声 + 身体動作 + 外見) :** 被験者は実験者と直接向かい合って対話する。実験者は灰色のシャツを着用した。被験者から見て、実験者の顔の大きさがロボットの顔とほぼ同じ大きさに見えるように、被験者の座る位置から実験者までの距離を調整した (150 cm)。

**ビデオ条件 (実体なし, 音声 + 身体動作 + 外見) :** 被験者はディスプレイに表示された実験者の映像を見ながら対話する。実験者の映像は、実験者側のネットワークカメラ (解像度 1,280 × 720, フレームレート 30 fps) で撮影し、縦置きにした 40 インチのディスプレイに表示した。相手の映像がディスプレイ上に表示されていることを意識するとソーシャルプレゼンスが低下することが予備実験から分かったため、ディスプレイの枠を被験者の位置から見えないように背景のパーティションと同じ白色の板で隠した。これにより、ディスプレイの表示範囲は 56 × 49 cm になった。映像の遅延は約 200 ms であった。ネットワークカメラの画角は、実験者の頭部がロボットの頭部の横幅と同じ 13.5 cm になるように調整した (水平画角 87 度)。

**ロボット条件 (実体あり, 音声 + 身体動作) :** 被験者はロボットを見ながら実験者と対話する。被験者から見た実験者の頭部や上半身の大きさを実験条件間で統制するうえで、人の身体と乖離が小さく等身大に近いデザインのロボットを使用する必要がある。そのようなロボットとして Telenoid [19] を採用し、実験者の服装と同じ灰色のシャツを着せた。ロボットの首は 3 自由度 (pitch/roll/yaw), 口は開閉のみの 1 自由度である。実験者と被験者の顔の動きは、フェイストラッキングソフト (FaceAPI) を用いて取得した。頭部の向き (pitch/roll/yaw の角度), 口の形 (3 次元座標の 14 頂点) のトラッキングデータを 30 fps で取得し、ロボットで再現した。実験者の動作に対するロボットの動作の遅延は約 300 ms であった。

**アバタ条件 (実体なし, 音声 + 身体動作) :** 被験者はディスプレイに表示された 3DCG のアバタを見ながら実験者と対話する。アバタは、黒い目と唇が配置された肌色の円柱の頭部と、実験者が着用したシャツとほぼ同じ色である灰色の円錐の胴体で構成されている。頭部の直径は、ロボットの頭部の横幅と同じ 13.5 cm に設定した。予備実験では、ロボットに類似したモデルのアバタを使用したが、頷きの動作に気付きにくい被験者がいたため、円柱状の頭部を持つアバタを用いて顔の向きを明確化することで解決した。実験者と被験者の顔の動きは、ロボット条件と同様に FaceAPI を用いて取得し、アバタで再現した。実験者の動作に対するアバタの動作の遅延は約 300 ms であった。アバタはビデオ条件で使用したものと同一ディスプレイに表示した。

**静止ロボット条件 (実体あり, 音声のみ) :** 被験者は動か

ないロボットを見ながら実験者と対話する。予備実験においてロボットがトラブルで動いていないと誤解する被験者がいたため、動かないロボットと会話することをあらかじめ伝えてから実験を行った。

**音声条件 (実体なし, 音声のみ) :** 一般的なボイスチャットと同様である。予備実験では、他の条件と同様にマイクスピーカを被験者から見えない位置に設置したが、実験者の声がどこから聞こえてくるのか分からないことに不安を感じた被験者がいたため机の上に設置した。

### 3.3 実験タスク

発話に含まれる途切れを分析するうえで、できるだけ長い 1 ターンの発話データを得る必要があった。そこで、特定の電子機器の問題点とその改善案、および追加してほしい機能について述べるタスクを設定した。被験者の対話相手である実験者は、我々の研究室に所属する男性の大学院生が務めた。実験は、1 人の被験者が 6 つすべての条件を体験する被験者内計画で実施したため、異なる 6 つの電子機器 (電子ブックリーダー, 携帯ゲーム機, スマートフォン, ロボット掃除機, 携帯音楽プレイヤー, 3D テレビ) を話題として設定した。これらの話題は、被験者が問題点などの意見を述べやすいように、最近話題になった電子機器から選定した。話題は各条件で実験を開始する際に実験者から伝えた。被験者が話している間の実験者の行動は、被験者の発言を遮らないように、頷きながら相槌を打つのみとした。実験者が途中で発話を促すことはなかったが、ほとんどの被験者は 1 分以上話すことができた。

聞き手の主な役割は、話し手の発話に相槌を打つことであり、その際に提示される明確な身体動作として「はい」などの発話にともなう口唇の動きや首の縦ふり (頷き) をアバタやロボットで再現した。これらの身体動作に比べて瞬きや動揺などの微細な動きは目立たないことが予想されるため再現しなかった。また、話し手の方向に身体を向けることも聞き手の身体動作として考えられるが、相手が 1 人しかいない場合には行われなため身体の向きは正面に固定した。これらの理由から、聞き手の身体動作の代表として首と口の動きを提示することとした。

6 条件の順番と話題の順番はカウンターバランスをとるよう決定した。また、6 条件の前に被験者をタスクに慣れさせるため、対面条件で練習を行った。この練習において被験者は実験者の外見を事前に確認したため、ビデオ条件や対面条件を他の条件の前に行くかどうかで実験者の外見を知るタイミングが被験者ごとに異なることはなかった。練習での話題は IC カード型乗車券とした。

### 3.4 実験データの収集

実験には大学生 36 人 (男性 : 19 人, 女性 : 17 人) が被験者として参加した。



図 2 ソーシャルテレプレゼンスを評価するアンケート

Fig. 2 Questionnaire to evaluate the social telepresence.

### 3.4.1 アンケート

6つの条件で実験を行った後、被験者にアンケート(図2)を実施した。このアンケートでは、被験者が感じた対話相手(実験者)のソーシャルテレプレゼンスを測定するため、「同じ部屋の中で相手に話しかけている感覚があった」という質問について、9段階のリッカード尺度で6つの条件をそれぞれ評価してもらった。各条件に対応するアンケート項目の順番は、実験を行った順番に並べ替えており、各条件の実験装置の写真とともにアンケートに記した。この並べ替えと実験装置の写真は、アンケートに答える際に、実験の想起を助ける手がかりである。また、アンケートの後に、点数を付けた理由についてインタビューで尋ねた。

### 3.4.2 途切れの頻度と割合の測定

被験者の発話中に含まれる途切れの頻度と割合を測定するため、各実験条件での被験者の発話をすべて録音し、3人の実験者がアノテーションツール(ELAN)を用いて書き起こした。その際、途切れ(無音区間)を得るために、発話区間の開始時間と終了時間を、発話の波形を見ながら可能な限り正確に決定した。ただし、促音「っ」による無音区間は発話の円滑さとは無関係であるため無視した。また、50msよりも短い区間はホワイトノイズと被験者の発話を区別することが困難であったため、50ms以上の無音区間のみを途切れとした。ホワイトノイズはノイズフィルタによって除去することが可能であるが、非常に小さい音量の発言も除去されてしまうため、フィルタは使用しなかった。1人の実験者が書き起こしたデータは、途切れの判断ミスがないか他の2人の実験者が確認した。

図3に途切れの頻度と割合の算出方法を示す。まず、全体の発話から一定量の発話を抽出する必要がある。発話の冒頭は、被験者がその話題に興味を持っている場合、流暢に話す傾向があり、話題による影響が大きい。一方、発話の末尾は、他に話すことがないか考えながら話すため、途

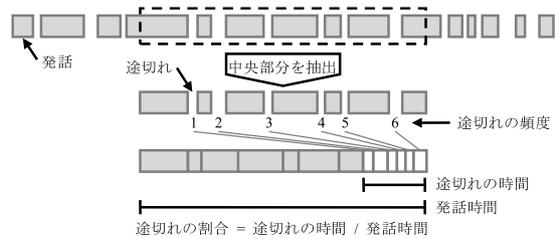


図 3 途切れの頻度および割合の算出方法

Fig. 3 Method to calculate the frequency of pauses and the percent pause time.

切れの多い発話になる場合が多い。そこで、全体の発話の中央部分を抽出することとした。この中央部分に含まれる途切れの数を途切れの頻度とする。そして、中央部分に含まれる途切れの合計時間を発話時間で割ったものを途切れの割合とする。発話の分量を量る単位は、発音の最小単位である音節を採用した。前述のとおり、ほとんどの被験者が1分以上話しており、1分に相当する音節数は約200であったため、抽出する発話の量は200音節とした。

## 4. 実験結果

図4にアンケートの結果、図5に途切れの頻度と途切れの割合の平均値をそれぞれ示す。図中の各点は、各実験条件の平均値であり、エラーバーは標準誤差である。

実験場面の録画において、すべての相槌と同時にアバタ、ロボットが頷いている(フェイストラッキングの結果が正しく反映されないことによる頷きの減少は起こっていない)ことを確認した。

### 4.1 ソーシャルテレプレゼンス

アンケートの結果について実体要因×伝達情報要因の2要因分散分析を行った。その結果、実体要因( $F(1, 35) = 36.955, p < .001$ )と伝達情報要因( $F(2, 70) = 279.603, p < .001$ )の主効果、および交互作用( $F(2, 70) = 14.794, p < .001$ )が有意であった。交互作用が有意であったため、Bonferroni補正法を用いて単純主効果の検定を行った。伝達情報が音声+身体動作の場合、および音声+身体動作+外見の場合、実体要因に有意差があり(それぞれ、 $F(1, 105) = 8.857, p < .01, F(1, 105) = 65.470, p < .001$ )、音声のみの場合は有意傾向であった( $F(1, 105) = 3.460, p = .086$ )。つまり、ソーシャルテレプレゼンスは、ロボット条件の方がアバタ条件よりも高いこと、対面条件の方がビデオ条件よりも高いことが分かった。これらの結果は仮説1(実体の効果)を支持する。ただし、伝達情報が音声のみの場合には、実体の効果は低いことも示された。

実体の有無にかかわらず伝達情報要因に有意差が認められた(それぞれ、 $F(2, 140) = 107.141, p < .001, F(2, 140) = 223.095, p < .001$ )。多重比較の結果、ソーシャルテレプレゼンスは、対面条件の方がロボット条件と静

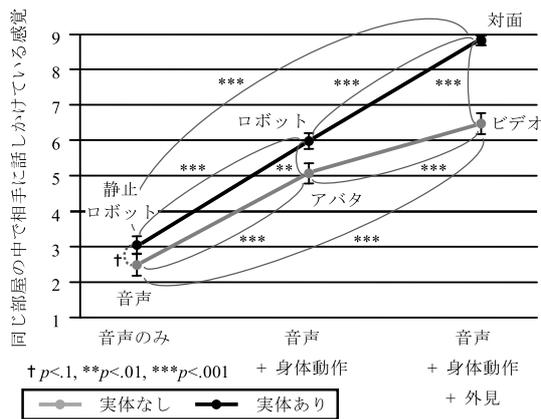


図 4 ソーシャルテレプレゼンスのアンケート結果

Fig. 4 Results of the questionnaire on social telepresence.

止ロボット条件よりも高いこと (ともに  $p < .001$ ), ロボット条件の方が静止ロボット条件よりも高いこと ( $p < .001$ ), ビデオ条件の方がアバタ条件と音声条件よりも高いこと (ともに  $p < .001$ ), アバタ条件の方が音声条件よりも高いこと ( $p < .001$ ) が分かった. これらの結果は仮説 2 (身体動作の効果) を支持する. また, 身体動作だけでなく外見の伝達の効果も示された.

#### 4.2 発話の円滑さ

3.4.2 項で述べたとおり, 途切れを分析するうえで発話の中央部分を抽出する必要があるが, 発話時間が 1 分に満たない場合, 分析可能な量の発話を抽出することができなかった. そこで, すべての条件で 1 分以上の発話が揃っている 26 人 (男性: 15 人, 女性: 11 人) のデータを対象として実体要因 × 伝達情報要因の 2 要因分散分析を行った.

まず, 途切れの頻度について分析した結果, 実体要因の主効果が有意であり ( $F(1, 25) = 8.004, p < .01$ ), 交互作用が有意傾向であった ( $F(2, 50) = 2.947, p = .062$ ). 交互作用は有意傾向であったが, Bonferroni 補正法を用いて単純主効果の検定を行ったところ, 伝達情報が音声 + 身体動作の場合, および音声 + 身体動作 + 外見の場合, 実体要因に有意差が認められた (それぞれ,  $F(1, 75) = 6.799, p < .05, F(1, 75) = 6.981, p < .05$ ). つまり, 途切れの頻度は, ロボット条件の方がアバタ条件よりも高いこと, 対面条件の方がビデオ条件よりも高いことが示された.

次に, 途切れの割合について分析した結果, 実体要因の主効果および交互作用が有意傾向であった (それぞれ,  $F(1, 25) = 3.174, p = .087, F(2, 50) = 2.947, p = .062$ ). 交互作用は有意傾向であったが, Bonferroni 補正法を用いて単純主効果の検定を行ったところ, 伝達情報が音声 + 身体動作の場合, および音声 + 身体動作 + 外見の場合, 実体要因に有意差が認められた (それぞれ,  $F(1, 75) = 4.647, p < .05, F(1, 75) = 4.369, p < .05$ ). つまり, 途切れの割合は途切れの頻度と同様に, ロボット条件の方がアバタ条件よりも高い

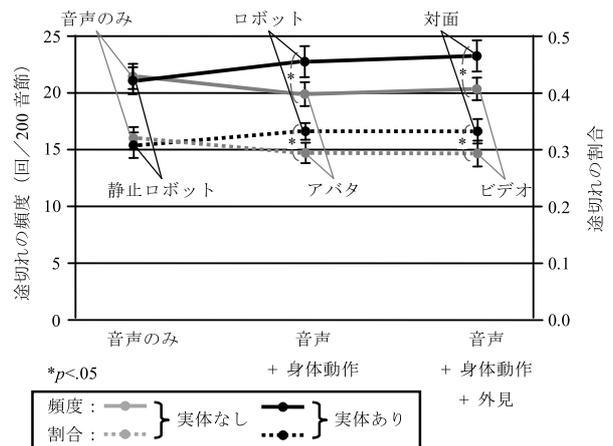


図 5 途切れの頻度と割合の平均値

Fig. 5 Frequency of pauses and percent pause time.

こと, 対面条件の方がビデオ条件よりも高いことが示された.

したがって, 途切れの分析結果は仮説 3 に反し, 実体があると途切れが増加する結果となった. ただし, 伝達情報が音声のみの場合には実体の効果は見られなかった. また, 伝達情報の効果は有意ではなく, 仮説 4 は示されなかった.

最後に, 途切れの分析対象とした 26 人のデータについて, 中央 200 音節中の実験者の相槌を数えた. 平均回数は音声: 7.2, 静止ロボット: 7.1, アバタ: 6.5, ロボット: 6.5, ビデオ: 6.3, 対面: 8.3 であった. 1 要因分散分析を行ったところ有意差が認められたため ( $F(5, 120) = 7.666, p < .001$ ), Bonferroni 補正法を用いて多重比較を行った結果, アバタ, ロボット, ビデオ条件は対面条件よりも有意に相槌の回数が少ないことが分かった (順に  $p < .01, p < .001, p < .01$ ). 相槌の回数が異なっていたことによる実験結果への影響については次章で考察する.

#### 5. 考察

実験結果から, 実体をともなう身体動作の伝達は対話相手のソーシャルテレプレゼンスを強化することが示唆された. 被験者へのインタビューでは, 目の前にロボットが物理的に存在するために, ロボット条件ではアバタ条件よりも実験者が目の前にいるように感じたという意見が得られた (36 人中 7 人). 一方, 静止ロボット条件も物理的な実体があるが, 音声条件との間に有意な差は見られなかった. インタビューでは, 実験者の反応を見ることができないため静止ロボット条件は音声条件と変わらないという意見が得られており (36 人中 3 人), アンケートでは 36 人中 8 人の被験者が音声条件と静止ロボット条件に同じ点数を付けていた. さらに, ロボットが動いたときに実験者がいるように感じたと述べた被験者もいた (36 人中 5 人). これらの被験者の意見は, 身体動作を伝達する場合に遠隔対話メディアが実体を持つことが効果的に働くという実験結果を支持するものである. これらの結果は, 実体を持たないアバタチャットに対するロボット会議の優位性を示している.

身体動作を伝達するうえで、実体を用いる場合と映像を用いる場合の違いの1つとして運動視差の有無があげられる。実体を用いる場合、運動視差から得られる奥行情報によって動作の視認性が向上することが考えられる。予備実験においてロボットに類似したアバタの頷き動作に気付きにくい被験者がいたのは(3.2節参照)、この奥行情報の欠如が影響していたのかもしれない。先行研究では、人の視線の移動に合わせて遠隔地のカメラを移動させる方法で再現された運動視差が、ビデオ会議におけるソーシャルテレプレゼンスを強化することが報告されている[16]。実体によるソーシャルテレプレゼンスの強化は、運動視差による動作の視認性の向上が寄与していた可能性がある。

遠隔対話メディアが伝達する情報に関しては、身体動作だけでなく外見もソーシャルテレプレゼンスを強化することが分かった。この結果は、対話相手の外見を伝えないアバタチャットとロボット会議の欠点を示すものである。しかし、図4を見ると、ロボット条件は、対話相手の外見を伝えていないにもかかわらず、ビデオ条件と同等であることが分かる。実際、アンケートでは半数近くの被験者がロボット条件にビデオ条件以上の点数を付けていた(36人中16人)。これは、実体によるソーシャルテレプレゼンスの強化が、外見を伝えないことによるソーシャルテレプレゼンスの低下を相殺したためだと考えられる。したがって、ビデオ会議に対するロボット会議の優位性を示した先行研究[20]の結果は、ロボットが人に近いリアルな外見を持っていたことが起因していたと考えられる。

発話の円滑化の効果はソーシャルテレプレゼンスの強化によるものであると予想していた。しかし、予想に反して、ソーシャルテレプレゼンスを強化した実体は途切れを増加させる結果となった。一方、実体のないアバタ条件とビデオ条件に関しては、統計的有意差は見られなかったが、途切れの頻度と割合(図5)が音声条件および静止ロボット条件よりも低いことがうかがえる。これは、先行研究[23]において報告された身体動作の伝達による発話の円滑化の効果が表れたものと推測される。統計的有意差が見られなかった原因は現在調査中であるが、本実験ではアバタやビデオを等身大で表示しているなど、先行研究とは実験環境にいくつかの差異があるためと思われる。

実体による途切れの増加は、対面会話で感じるような緊張感が起因していた可能性がある。インタビューでは、36人中16人の被験者が緊張感について言及していた。そのうち、緊張を感じた条件として15人が対面条件を、5人がロボット条件をあげていた。一方、ビデオ条件とアバタ条件をあげた被験者は2人のみであり、静止ロボット条件と音声条件をあげた被験者はいなかった。つまり、実体で身体動作を提示することが緊張感を与えた可能性がある。社会心理学の分野では、社交不安を感じると途切れの頻度[9]や時間的な割合[8]が増加することが知られており、対面

条件やロボット条件における途切れの増加はこの社交不安によるものと推測される。社交不安とは、対話中に感じる不快感のことであり、対面で緊張する感覚もその1つである。ロボット会議は実体の効果によって対面で感じるような緊張感を生み出す可能性がある。

実験では、実体と伝達情報に要因を分けて実験条件を比較したが、一部の条件間にはこれらの要因以外の差異も含まれていた。まず、アバタ条件とロボット条件の比較において実体の有無以外に見た目の差異が影響を与えていた可能性がある。アンケートにおいてロボット条件にアバタ条件よりも低いスコアをつけた被験者や、両者に大きな差をつけなかった被験者も存在した。それらの被験者のほとんどは、アバタよりもロボットの見た目の異様さに言及しており(36人中、アバタ:3人、ロボット:16人)、アバタを好む傾向があった。また、ロボットに異様さを感じた16人には、アバタのように抽象化した方が良いと述べた被験者が7人、よりリアルにした方が良いと述べた被験者が6人、約半数ずつ含まれており、ロボットの見た目は不気味の谷に位置するものであった可能性が推測される。したがって、ロボットの見た目をより抽象化した場合には、アバタに対する優位性がより顕著に表れた可能性がある。また、ロボットで対話相手の外見を再現する場合には不気味の谷を越えるほどリアルにする必要があり、リアルな外見を提示することでビデオ会議に対しても優位性が得られる可能性がある。途切れの増加に関しても、ロボットや実験者の見た目が起因していた可能性が懸念されるが、これらの条件と同じ見た目を持つ静止ロボット条件やビデオ条件において途切れが増加していなかったことから、途切れへの影響は見た目よりも実体の方が大きかったと思われる。

音声条件とアバタ条件の比較において身体動作の有無以外に対話相手の代理となる存在の有無も影響を与えていた可能性がある。しかし、対話相手の代理として設置した静止ロボットにはソーシャルテレプレゼンスを強化する効果や途切れを増加させる効果がなかったことから、対話相手の代理があることによる実験結果への影響は実体や身体動作の影響と比べて大きくなかったと考えられる。

頷き動作の遅延は3.2節で述べたとおりアバタ条件とロボット条件で最も大きく、音声に対して約200ms遅れていた。インタビューでは、アバタ条件とロボット条件の遅延に言及した被験者は36人中それぞれ2人のみであり目立つものではなかったと思われるが、遅延によってソーシャルテレプレゼンスが低下した可能性がある。また、途切れに関しては、遅延がない対面条件と遅延があるロボット条件でともに増加しており、ビデオ条件よりも遅延が大きいアバタ条件で増加していないことから、遅延による影響は実体の影響と比べて大きくなかったと思われる。

音声の遅延については実験者の相槌の頻度に影響を与えていた可能性がある。対面条件よりも相槌の回数が少な

かったアバタ、ロボット、ビデオ条件は(4.2節)、相手の振舞いが見えるため、被験者が話し続けようとした場合に、相槌のタイミングだと判断しても遅延による同時発話を避けようとして相槌を控えがちになっていた可能性がある。対面条件では遅延を気にする必要がないため、この差が相槌の回数の差となって表れたと考えられる。しかしながら、相槌が多かった対面条件において、少なかったロボット条件と同様に途切れが増加したことや、相槌が少なかったアバタ条件で途切れが増加しなかったことから、相槌の頻度の影響は実体の影響と比べて大きくなかったと思われる。一方、ソーシャルテレプレゼンスに関しては、アバタ条件とロボット条件において相槌は実験者が聞いていることを被験者に伝える唯一の手段であるため、相槌の減少はマイナスに働いた可能性が懸念される。相槌の回数とアンケートのスコアの相関係数を求めたところ、アバタ：-0.235、ロボット：-0.078、ビデオ：0.017であり、相槌の回数が少ないとスコアが低くなる傾向は見られなかった。したがって、相槌の減少による実験結果への影響は大きいものではなかったと考えられる。

実験では、1人の実験者が被験者の対話相手を務めたため、その性別や年齢は固定であった。実験は被験者内計画で行われたため、その影響がたとえあったとしても条件間で相殺されたと思われる。実験者と被験者はタスクの練習の際に初めて顔を合わせる初対面の関係であった。会話における緊張感は両者の社会的関係に依存すると考えられるため、実体によって生じたと推測される緊張感は、友達同士の対話では表れない可能性がある。そのような社会的関係の影響によって遠隔対話メディアの効果の表れ方が異なるかを明らかにすることは今後の課題である。

アバタやロボットで提示した身体動作は相槌にともなう首や口の動きのみであったがソーシャルテレプレゼンスを強化する効果が確認できた。瞬きや動揺など本研究で扱わなかった身体動作を追加することでその効果が増すことが期待される。しかしながら、ソーシャルテレプレゼンスを強化するうえで必要最小限の身体動作は明らかになっておらず、首か口いずれかの動作や、3者以上の対話であれば話し手に身体を向ける動作の提示のみでも十分であった可能性もある。また、本研究では、話し手のソーシャルテレプレゼンスを強化する要因についても明らかになっていない。話し手はつねに発話動作を提示できるため、身体動作の提示がより効果的である可能性が考えられるが、相槌のように相手の発話に対してインタラクティブに行われるものではないため、アバタやロボットが自動的に動作を再生していると感じられた場合には有効に働かない可能性も考えられる。これらを明らかにすることは今後の課題である。

## 6. まとめ

本研究では、実体で対話相手の身体動作を伝達するとい

うロボット会議の特徴が遠隔会議に与える影響を検証するため、実体の有無と伝達情報の違い(音声のみ、音声+身体動作、音声+身体動作+外見)の要因に分けて、ロボット会議と従来の遠隔対話メディアおよび対面会話を比較した。その結果、実体のあるロボットで身体動作を伝達するとソーシャルテレプレゼンスが強化されることが分かった。この結果から、映像で身体動作を伝達するアバタチャットに対するロボット会議の優位性が示された。しかし、ロボット会議が伝達しない対話相手の外見もソーシャルテレプレゼンスを強化することが分かった。結果として、外見の欠如によるマイナス効果を実体があることによるプラス効果で補うため、ロボット会議とビデオ会議のソーシャルテレプレゼンスは同等であった。先行研究では、ビデオ会議に対するロボット会議の様々な優位性が議論されてきた。しかし、ソーシャルテレプレゼンスに関しては、ロボットがリアルな外見を持たない場合、必ずしもロボット会議に優位性があるとはいえない。

さらに、実体で対話相手の身体動作を伝達すると、発話に含まれる途切れが増加することも明らかになった。この途切れの増加は、実体が与える緊張感によるものと推測される。したがって、ロボット会議は対面会話で感じるような緊張感を生み出す可能性があり、面接や講義などある程度の緊張感を必要とする用途で使用できるかもしれない。

謝辞 本研究は、JST CREST「人の存在を伝達する携帯型遠隔操作アンドロイドの研究開発」、基盤研究(B)「ソーシャルテレプレゼンスのためのロボットエンハンスドディスプレイ」、挑戦的萌芽研究「気配伝達型ソーシャルテレプレゼンスの研究」、KDDI財団「人間クラウドのためのロボティックアバター」、倉田記念日立科学技術財団「監視感を最小化しつつ存在感を最大化するミニマルロボティックメディア」からの支援を受けた。

## 参考文献

- [1] Anderson, A.H., Newlands, A., Mullin, J., Fleming, A., Doherty-Sneddon, G. and Van Der Velden, J.M.: Impact of Video-Mediated Communication on Simulated Service Encounters, *Interacting with Computers*, Vol.8, No.2, pp.193-206 (1996).
- [2] Bainbridge, W.A., Hart, J., Kim, E.S. and Scassellati, B.: The benefits of interactions with physically present robots over video-displayed agents, *International Journal of Social Robotics*, Vol.3, No.1, pp.41-52 (2011).
- [3] Bente, G., Ruggenberg, S., Kramer, N.C. and Eschenburg, F.: Avatar-Mediated Networking: Increasing Social Presence and Interpersonal Trust in Net-Based Collaborations, *Human Communication Research*, Vol.34, No.2, pp.287-318 (2008).
- [4] Daly-Jones, O., Monk, A.F. and Watts, L.: Some Advantages of Video Conferencing over High-quality Audio Conferencing: Fluency and Awareness of Attentional Focus, *International Journal of Human-Computer Studies*, Vol.49, No.1, pp.21-58 (1998).
- [5] de Greef, P. and Ijsselstein, W.: Social Presence in a

Home Tele-Application, *CyberPsychology and Behavior*, Vol.4, No.2, pp.307–315 (2001).

[6] Finn, K.E., Sellen, A.J. and Wilbur, S.B.: Video-Mediated Communication, *Lawrence Erlbaum Associates* (1997).

[7] Garau, M., Slater, M., Bee, S. and Sasse, M.A.: The Impact of Eye Gaze on Communication Using Humanoid Avatars, *Proc. CHI2001*, pp.309–316 (2001).

[8] Goberman, A.M., Hughes, S. and Haydock, T.: Acoustic characteristics of public speaking: Anxiety and practice effects, *Journal of Speech Communication*, Vol.53, No.6, pp.867–876 (2011).

[9] Harrigan, J.A., Suarez, I. and Hartman, J.S.: Effect of Speech Errors on Observers' Judgments of Anxious and Defensive Individuals, *Journal of Research in Personality*, Vol.28, No.4, pp.505–529 (1994).

[10] Isaacs, E.A. and Tang, J.C.: What Video Can and Can't Do for Collaboration: A Case Study, *Multimedia Systems*, Vol.2, No.2, pp.63–73 (1994).

[11] Kang, S., Watt, J.H. and Ala, S.K.: Communicators' Perceptions of Social Presence as a Function of Avatar Realism in Small Display Mobile Communication Devices, *Proc. HICSS2008* (2008).

[12] Kuzuoka, H., Yamazaki, K., Yamazaki, A., Kosaka, J., Suga, Y. and Heath, C.: Dual Ecologies of Robot as Communication Media: Thoughts on Coordinating Orientations and Projectability, *Proc. CHI2004*, pp.183–190 (2004).

[13] Lee, K.M., Jung, Y., Kim, J. and Kim, S.R.: Are physically embodied social agents better than disembodied social agents?: The effects of physical embodiment, tactile interaction, and people's loneliness in human-robot interaction, *International Journal of Human-Computer Studies*, Vol.64, No.10, pp.962–973 (2006).

[14] Misawa, K., Ishiguro, Y. and Rekimoto, J.: LiveMask: A telepresence surrogate system with a face-shaped screen for supporting nonverbal communication, *Proc. AVI2012*, pp.394–397 (2012).

[15] Morita, T., Mase, K., Hirano, Y. and Kajita, S.: Reciprocal Attentive Communication in Remote Meeting with a Humanoid Robot, *Proc. ICMI2007*, pp.228–235 (2007).

[16] Nakanishi, H., Murakami, Y. and Kato, K.: Movable Cameras Enhance Social Telepresence in Media Spaces, *Proc. CHI2009*, pp.433–442 (2009).

[17] Nakanishi, H., Kato, K. and Ishiguro, H.: Zoom Cameras and Movable Displays Enhance Social Telepresence, *Proc. CHI2011*, pp.63–72 (2011).

[18] Nguyen, D.T. and Canny, J.: More than Face-to-Face: Empathy Effects of Video Framing, *Proc. CHI2009*, pp.423–432 (2009).

[19] Ogawa, K., Nishio, S., Koda, K., Balistreri, G., Watanabe, T. and Ishiguro, H.: Exploring the Natural Reaction of Young and Aged Person with Telenoid in a Real World, *Journal of Advanced Computational Intelligence and Intelligent Informatics*, Vol.15, No.5, pp.592–597 (2011).

[20] Sakamoto, D., Kanda, T., Ono, T., Ishiguro, H. and Hagita, N.: Android as a Telecommunication Medium with a Human-like Presence, *Proc. HRI2007*, pp.193–200 (2007).

[21] Sellen, A.J.: Remote Conversations: The Effects of Mediating Talk with Technology, *Human-Computer Interaction*, Vol.10, No.4, pp.401–444 (1995).

[22] Sirkin, D. and Ju, W.: Consistency in physical and

on-screen action improves perceptions of telepresence robots, *Proc. HRI2012*, pp.57–64 (2012).

[23] Tanaka, K., Nakanishi, H. and Ishiguro, H.: Appearance, Motion, and Embodiment: Unpacking Avatars by Fine-grained Communication Analysis, *Concurrency and Computation: Practice and Experience*, Vol.27, No.11, pp.2706–2724 (2015).

[24] Tanaka, K., Nakanishi, H. and Ishiguro, H.: Physical Embodiment Can Produce Robot Operator's Pseudo Presence, *Frontiers in ICT*, Vol.2, No.8 (2015).



田中 一晶 (正会員)

2006年京都工芸繊維大学工芸学部電子情報工学科卒業。2008年同大学大学院工芸科学研究科情報工学専攻博士前期課程修了。2011年京都工芸繊維大学大学院工芸科学研究科情報工学専攻博士後期課程修了。博士(工学)。

同年大阪大学大学院工学研究科知能・機能創成工学専攻特任助教。2015年より関西学院大学理工学部特任講師。人とインタラクションを行うエージェントやロボットの設計に興味を持つ。



中西 英之 (正会員)

1996年京都大学工学部情報工学科卒業。1998年同大学大学院工学研究科情報工学専攻修士課程修了。同年日本学術振興会特別研究員。2001年京都大学大学院情報学研究科社会情報学専攻博士課程修了。博士(情報学)。

同年同専攻助手。2006年より大阪大学大学院工学研究科知能・機能創成工学専攻准教授。空間共有や存在感伝達のためのアバタやロボットに興味を持つ。2002年度情報処理学会坂井記念特別賞。2004年度テレコムシステム技術賞。2006年度科学技術分野の文部科学大臣表彰科学技術賞。



石黒 浩 (正会員)

1991年大阪大学大学院基礎工学研究科博士課程修了。工学博士。その後、京都大学情報学研究科助教授、大阪大学大学院工学研究科教授等を経て、2009年より大阪大学大学院基礎工学研究科教授。2013年大阪大学特別教

授。ATR 石黒浩特別研究所客員所長(ATRフェロー)。専門は、ロボット学、アンドロイドサイエンス、センサネットワーク等。2011年大阪文化賞受賞。2015年文部科学大臣表彰受賞。