

遠隔対話者の身体動作の提示による 音声コミュニケーションの円滑化

尾上 聡¹ 山本 健太¹ 田中 一晶^{1,2,a)} 中西 英之¹

受付日 2012年6月27日, 採録日 2013年1月11日

概要: 遠隔地の対話者を視覚的に提示する方法として, ビデオ, アバタ, 写真の3つの方法があげられる。しかし, これらの方法が音声コミュニケーションに与える効果を, 観察データから明らかにしたという報告はない。我々はそのような観察データを得るために, 1ターンの発話に含まれる途切れ回数と発話速度に着目した。そして, 我々は音声コミュニケーションに対する視覚情報の有効性を示すため, 提示する視覚情報を身体動作と外見の要因に分け, 4つの条件を設定した。その4つの条件は, 身体動作と外見を提示するビデオ条件, 身体動作のみを提示するアバタ条件, 外見のみを提示する写真条件, 視覚情報を提示しない音声のみの条件である。我々はこれらの条件下で対話実験を行い, 発話中に含まれる途切れ回数と発話速度を計測した。その結果, 外見の提示の有無にかかわらず, 身体動作の提示により途切れ回数が減少し発話速度が上昇していることが分かった。この結果は, 身体動作の提示により音声コミュニケーションが円滑化されることを示しており, さらにアバタがビデオを代替する可能性が示唆された。

キーワード: アバタ, ビデオ会議, 身体動作, 音声コミュニケーション

Smoothing Audio Communication via Presentation of a Remote Partner's Body Motion

SATOSHI ONOUE¹ KENTA YAMAMOTO¹ KAZUAKI TANAKA^{1,2,a)} HIDEYUKI NAKANISHI¹

Received: June 27, 2012, Accepted: January 11, 2013

Abstract: As methods to visually present a remote dialogue partner, there are video, avatar and photograph. However, there is no report that observational data showed an effect to audio communication by these methods. In order to obtain such data, we focused on the numbers of pauses included speech in audio communication and speech rate. We separated these methods into the factor of the partner's body motion and appearance, and thereby prepared four conditions: voice, photo, avatar and video. The video and the avatar conditions presented the body motion. The video and the photograph conditions presented the appearance. The audio condition didn't present either of them. We conducted an experiment in which the subjects spoke to the remote conversation partner under these conditions, and analyzed the number of pauses and the speech rate in the recorded speeches. As a result of the analyses, the presentation of the body motion reduced the number of pauses and increased the speech rate, regardless of the presentation of the appearance. These results indicate that the presentation of the body motion smoothens audio communication, and also imply the possibility that avatar replaces video.

Keywords: avatar, videoconferencing, body motion, audio communication

¹ 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, Graduate School
of Engineering, Osaka University, Suita, Osaka 565-0871,
Japan

² 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

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

1. はじめに

遠隔会議において, 対話者を視覚的に提示する方法として, ビデオ, アバタ, 写真の3つの方法が存在する。

ビデオを用いる方法は実際の対話者の映像を配信することによって, 対話者の表情, 容姿, 身振り手振りなどを提

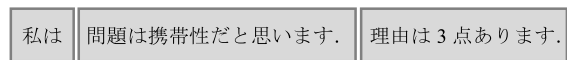
示することができる。長年、ビデオを遠隔会議で使用する必要性が議論されてきた [16]。観察データを用いた評価からその必要性を実証するために多くの研究が行われてきたが、いまだ明らかにされていない [1], [5], [14], [17], [20]。

次に、アバタを用いた方法は仮想空間上のアバタが対話者の代わりに動作することによって対話者の身体的な動作を提示することができる。近年のトラッキング技術の向上にともない、ユーザの顔の動きや身振り手振りなどをリアルタイムに反映させることが可能なアバタが、容易かつ安価に製作できるようになった。対話者の身体的な動作の情報を伝えようとすると、通常ビデオを使用する必要があるが、その場合身体的な動作の情報以外に対話者の外見の情報も伝えてしまう。しかし、リアルタイムに対話者の身体的な動作を反映するアバタを用いれば、外見の情報を伝えることなく、身体的な動作の情報のみを伝えることができる。このようにリアルタイムに対話者の身体的な動作を反映するアバタには利点があり、様々な研究がなされているが、音声コミュニケーションへの有効性については明らかになっていない [2], [3], [8], [11]。

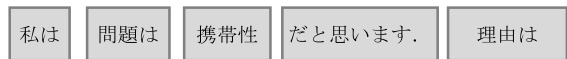
そして、写真を用いた方法は顔写真を用いることで対話者の外見を提示することができる。インスタントメッセージのユーザの多くは、自らの顔写真をユーザアイコンとして登録しており、インスタントメッセージのボイスチャット機能は頻繁に使用されている。このように顔写真を見ながら音声コミュニケーションを行うことはすでに多くの人々の間で行われているが、その有効性についてはアバタと同様に明らかになっていない [4], [21]。

先行研究では、観察データから音声コミュニケーションを評価する指標として、同時発話の回数および同時発話の継続時間などが用いられている。これらの指標は、複数の対話者間における話者交替の円滑さを評価するものである。しかし、2者間での音声コミュニケーションにおいて対話者を視覚的に提示することによる有効性を、これらの話者交替を評価する指標で示すことに成功した研究はない [20]。本研究では、2者間での音声コミュニケーションにおける対話者の視覚的提示（写真、アバタ、ビデオ）の有効性を示すため、新たな評価指標として1ターンの発話に含まれる途切れに着目した。本評価指標では途切れの頻度が少なく発話速度が速い発話を滑らかな発話、途切れの頻度が多く発話速度が遅い発話をぎこちない発話と定義する。図1にそれらの発話の例を示す。ブロックはひと続きの発話を表しており、それらの間隔は途切れを表している。途切れが頻出し流暢でない発話では、聞き手は話し手に対して不安を感じるという報告がある [9]。したがって、そのよう発話は良好とはいえない。我々は滑らかな発話であれば音声コミュニケーションが円滑であると考え、途切れの頻度の減少と発話速度の上昇は音声コミュニケーションの改善であると定義した。本研究では1ターンの発話に含まれてい

滑らかな発話



ぎこちない発話



時間

図1 滑らかな発話とぎこちない発話

Fig. 1 Example of smooth speech and awkward speech.

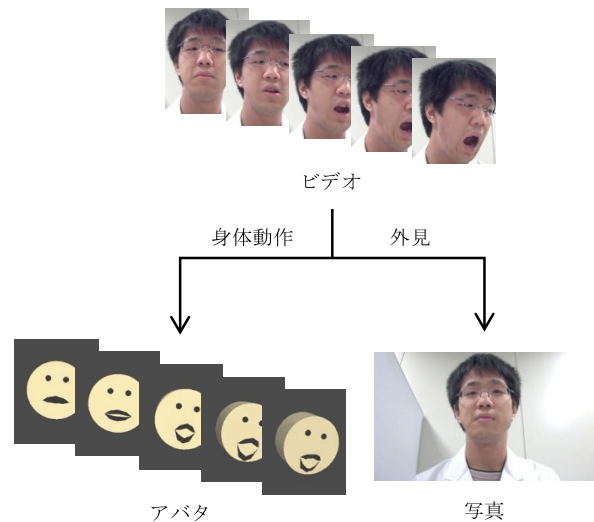


図2 ビデオに含まれる身体動作と外見の情報

Fig. 2 Motion and appearance in video.

る途切れの頻度と発話速度について分析する。

次に、本研究では遠隔メディアに含まれる情報の要因についても着目した。遠隔対話メディアを比較した研究のほとんどは、ビデオを1つの要因として扱っていた [5], [14], [20]。ビデオを用いた対話では、聞き手の身体的な動作や表情などの視覚的な手がかりによって、音声のみの対話より自発的な発話が継続することが報告されている [18]。さらに、聞き手の顔が見えることによって言い争いになりにくくなるという報告もある [23]。これらの報告から、ビデオは聞き手の身体的な動きと外見の2つの情報を伝達しており、それぞれの情報は会話に影響を与えていると考えられる。そこで、我々は遠隔対話メディアに含まれる情報を聞き手の顔きといった身体動作の要因と聞き手の外見の要因に分けた。

これらの着目点をふまえ、本研究では遠隔対話メディアを聞き手の身体動作の要因と聞き手の外見の要因によって分類し、1ターンの発話に含まれる途切れの頻度と発話速度の分析によって評価した。図2に示すように、ビデオは聞き手の身体動作と外見の情報を話し手に伝達していると仮定する。次に、アバタは聞き手の身体動作の情報のみを伝達し、写真は聞き手の外見の情報のみを伝達していると仮定する。これらの仮定により、身体動作による効果と外見による効果に分けて分析することが可能となる。

2. 関連研究

対話でビデオ会議を用いる有効性を調査するため、多くの実験が行われてきた。そのいくつかの実験では、視覚情報の共有によってコミュニケーションが円滑になることが示されている [5], [14], [20]。しかし、会話構造 [1], [5], [20] やタスクパフォーマンス [1], [14], [17] などの観察データによる評価では、2者間対話でのビデオの利用が音声コミュニケーションを改善するかは明らかにされていない。

近年、リアルタイムにユーザの動きを反映するアバタを使用して、身体動作を含む非言語情報の重要性を示す研究がなされている [2], [3], [8], [11]。アンケートによる主観評価において、アバタを用いた対話は気を許して話しやすい [2]、またはアバタを用いた協調作業は相手の存在感と信頼感が向上する [3] という報告がある。しかし、このようなアバタに音声コミュニケーションを改善する効果があるかどうかについて観察データによる評価では明らかになっていない。

ビデオとアバタに加え、写真もまた対話者の提示方法の1つとして研究されている [4], [21]。アンケートによる主観の評価において、音声のみの対話よりも顔写真を用いた対話は話者を判断しやすいという報告がある [4]。しかし、いまだに写真が音声コミュニケーションに及ぼす影響については観察データから明らかにされていない。

本研究がこれらの関連研究と異なる点は、音声コミュニケーションの評価手法として、1ターンの発話に含まれる途切れの頻度と発話速度といった会話の細かい構造に着目した点である (図 1)。また、遠隔対話メディアに含まれる情報に着目し、ビデオを身体動作の要因と外見の要因に分けたことも関連研究と異なるもう1つの点である (図 2)。我々は、この2つの要因が途切れの頻度と発話速度に及ぼす影響を調査した。

3. 実験

3.1 仮説

本研究の仮説は以下のとおりである。

仮説 1: 音声のみの対話と比較して、聞き手の身体動作と外見を伝達するビデオを提示した場合、話し手の発話の途切れの頻度が減少し、発話速度が上昇する。

仮説 2: 音声のみの対話と比較して、聞き手の身体動作のみを伝達するアバタを提示した場合でも、話し手の発話の途切れの頻度が減少し、発話速度が上昇する。

ビデオを用いた対話では、音声のみの対話と比較して自発的な発話が維持できることが示されている [18]。したがって、ビデオを用いた対話では音声のみの対話よりも滑らかな発話になると予想する。アバタを用いた対話では、音声のみの対話よりも発話量が多いことが示されている [6]。したがって、ビデオだけでなく、聞き手の身体動作

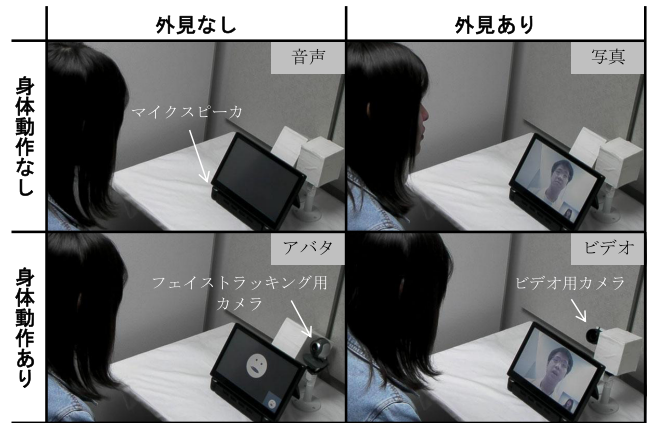


図 3 実験条件

Fig. 3 Conditions of the experiment.

のみを伝達するアバタを提示することでも発話が滑らかになると予想する。

3.2 実験条件

この仮説を実証するために、身体動作および外見をそれぞれ要因として設定し、身体動作2水準(有・無)および外見2水準(有・無)の2要因2水準の実験計画で行った。我々は図 3 に示すように2×2の4つの実験条件を設定した。被験者が話し手役、実験者が聞き手役となり、被験者と実験者は全4条件ともビデオ会議システムが置かれた机の前に座った状態で対話実験を行った。このビデオ会議システムは、マイクスピーカ、ディスプレイ、ビデオ用カメラとフェイストラッキング用カメラで構成されている。使用したディスプレイは、通常のビデオ会議システムで使用されるディスプレイよりも小型の10インチのワイドスクリーンの液晶ディスプレイを使用した。これは聞き手の表示サイズが小さくても、視覚情報の提示が音声コミュニケーションを改善するか確認するためである。各条件について以下に示す。

音声条件 (身体動作なし, 外見なし): この条件は一般的なボイスチャットと同じである。被験者はマイクスピーカのみを介して聞き手に話しかける。被験者に実験者と視覚情報を共有していないことを直感的に認識させるため、ディスプレイには何も表示せず、2台のカメラは白い箱で覆う。本研究では音声コミュニケーションに対する視覚情報の有効性を示そうとしているため、身体動作なし、外見なしの条件を互いに視覚情報がない状態を統制条件とし音声条件と設定した。

写真条件 (身体動作なし, 外見あり): この条件はインスタントメッセージのボイスチャットと同じである。写真条件と音声条件の異なる点は、ディスプレイに実験者の顔写真を表示するという点である。また、被験者が実験者のディスプレイにどのように自身の写真が表示されているか確認するため、被験者側のディスプレイの右下に自身の顔

写真を小さく表示する。被験者と実験者の顔写真は実験開始前に撮影したものをを用いる。音声条件と同様に、2台のカメラを白い箱で覆い、対話は音声のみでのやりとりであることを被験者に伝える。

アバタ条件（身体動作あり、外見なし）：この条件はユーザの顔の特徴を持たないアバタを用いて対話する。図2で示すように、本研究で用いたアバタはユーザの頭部と唇の動きを再現するが、ユーザの外見は伝達しない。アバタ条件では、ビデオ用カメラを白い箱で覆い、フェイストラッキング用カメラを使用する。アバタの3次元モデルは、円筒状の頭部、唇と眼球で構成されている。以下、アバタのデザインの詳細について述べる。頭部の形状は球状より円筒状の頭部が顔の向きの変化を理解しやすいと考え、円筒状の頭部を用いた。写真条件とビデオ条件でディスプレイに表示した実験者の顔のサイズとアバタの表示サイズが同じになるように調整した。目の形状は球形をしており、円筒状の頭部の平面側に固定されている。次に、アバタの動作の生成方法について述べる。まず、フェイストラッキング用ソフトウェア（faceAPI）によって、ユーザの頭部と唇の動作情報を30fpsで取得する。そして、取得された頭部の位置・姿勢（3次元座標、pitch/roll/yawの角度）と唇の形状（14頂点の3次元座標）からアバタの動作を生成する。各フェイストラッキング用カメラから撮影された映像を用いてフェイストラッキングを行い、各ディスプレイにアバタが表示されるまでの遅延時間は約300msである。被験者が実験者側のディスプレイに自身のアバタがどのように表示されているか、またアバタがどれだけ正確に動きを再現できているか確認できるように、被験者側のディスプレイの右下に自身のアバタを小さく表示する。

ビデオ条件（身体動作あり、外見あり）：この条件は一般的なビデオチャットと同じである。ビデオ条件では、フェイストラッキング用のカメラを白い箱で覆い、ビデオ用のカメラを使用する。被験者側と実験者側のディスプレイには、対話する相手のビデオが表示され、ディスプレイの右下に自身のビデオが表示される。ビデオの解像度は640×375ピクセルあり、フレームレートは30fpsである。各ビデオ用カメラから撮影された映像が、各ディスプレイに表示されるまでの遅延時間は約150msである。

3.3 被験者

実験には、本大学の近辺に在住する学部生24名が参加した。3.4節で述べる実験タスクに慣れてもらうために、各被験者は実験開始前に初対面の実験者と対面で実験の練習を行い、その後全4条件で実験を行う。

3.4 実験タスク

途切れの頻度と発話速度の計測を安定させるため、被験者が一方的に長く話し続ける会話を収集する必要がある

た。なぜならば、被験者が話している途中で実験者へ問いかけた場合、実験者の返答が途切れの頻度や発話速度に影響を及ぼす可能性があるからである。したがって、被験者の実験者への問いかけを避ける必要があった。被験者が問いかけなしに一方的に話し続けるために、特定の電子機器の問題点とその改善点、その電子機器が普及するためには何をしたらよいのかということについて自由に意見を述べるタスクを設定した。全被験者は4つの実験条件と練習を含め5回の対話を行うため、話題は異なる5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、3Dテレビ）を設定した。話題の順序はランダムに設定した。実験者は実験開始時に話題を被験者に伝え、被験者はその話題についての意見を自由に述べた。被験者が意見を述べている間の実験者の行動は、被験者の発話への影響を最小限にするため相槌を打ちながら頷くのみとした。

我々は「〇〇分以上意見を述べてください」といった指示はしなかったため、被験者はいつでも対話を終了することができた。しかし、最近注目を浴びている電子機器を選んだため、ほとんどの被験者は全5つの電子機器の問題点と改善点にある程度知識があり、1分以上話し続けてもらうことができた。1分間の発話は途切れの頻度と発話速度を計測するうえでは十分な長さであった。

3.5 データ収集

収集した被験者の発話にはフィラーが含まれている。フィラーは、発話中の沈黙を避ける（途切れを埋める）ために発する言葉である。また、フィラーには、話し手の発話権を維持する役割があることも知られている[24]。本実験ではつねに発話権は被験者が保持しているため、そのような役割で使用されたフィラーはなかった。したがって、発話の分析の際には、フィラーを途切れとして扱った。我々は先行研究で使用されたフィラーのリスト[10]を参考にし、録音した音声からフィラーを識別した。分析を行うため、録音した発話の書き起こしを行い、図4に示すように発話に含まれるフィラーを取り除いた。促音「っ」は、発話のぎこちなさを示す途切れとは無関係であるため、途切れと判断しなかった。

途切れの頻度と発話速度を定量的に分析するためには、全体の発話から一定量の発話を抽出する必要がある。発話の開始部分では、被験者が、設定した話題の電子機器に直に触れるような体験があった場合、被験者はその体験を滞ることなく話す傾向があった。一方、発話の後半部分では、被験者は意見が思いつかなくなり、発話が途切れがちになる傾向があった。そこで、被験者の知識や経験に左右されずに発話を分析するため、発話の中央部分を抽出することとした。発話の分量を測る単位として、音節やモーラなどがある。モーラは、音節とは異なり、撥音「ん」、促音

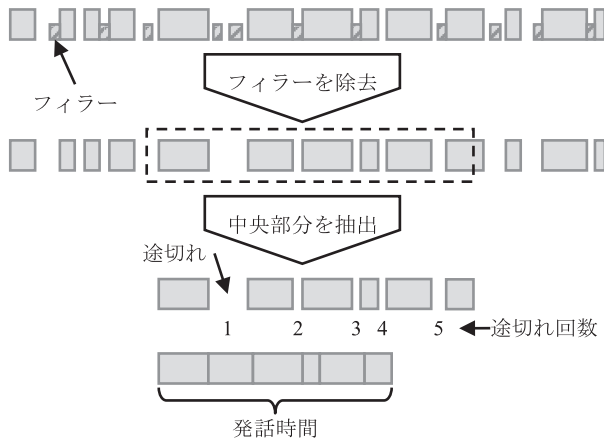


図 4 途切れ回数と発話時間を求める手順

Fig. 4 Procedure to count the number of pauses and measure the duration of speech.

「っ」、長音「ー」を含んでいる。日本語では、発話の分量を測る単位としてモーラを用いることが一般的であるが、本実験では、書き起こしの際に発話に含まれる長音の有無を判断することが困難であったため、音節を用いた。図 4 に途切れ回数と発話時間を求める手順を示す。ブロックは発話を表し、斜線のブロックはフィルターを表し、それらの隙間は途切れを表している。途切れの頻度と発話速度を分析するために、中央部分の発話の途切れ回数を数え、その発話時間を求めた。

すべての被験者の発話は 1 分以上あり、1 分間の発話は約 200 音節であったため、発話の中央 200 音節を抽出した。したがって、分析には中央 200 音節内の途切れ回数と発話速度 (syllables/sec) を使用した。

録音した発話の処理工程は以下のとおりである。我々は、録音した発話を書き起こすためにマルチメディア注釈ツール (ELAN) を使用し、可能な限り正確に発話の開始時刻と終了時刻を入力し、その区間の発話すべてを書き起こした。録音した音声に入ったホワイトノイズにより、50ms 以下の無音区間を途切れと判断するには困難であった。そのため、50ms より長い無音区間を途切れとした。ホワイトノイズの除去方法としては録音した音声にフィルタリングを行うことが考えられるが、音量の小さい発話までが除去される可能性があった。それゆえ、録音した音声データにフィルタリングを施すことはしなかった。ELAN を使用した書き起こしのデータから、表計算ソフトを用いて音節数と途切れ回数と発話時間を求めた。

4. 実験結果

実験は 1 人の被験者が全 4 条件を行うため、その順序をラテン方格法で設定し、被験者間でカウンターバランスをとった。

身体動作の提示の有無または外見の提示の有無によって途切れ回数が減少するかどうか調べるために、身体動作

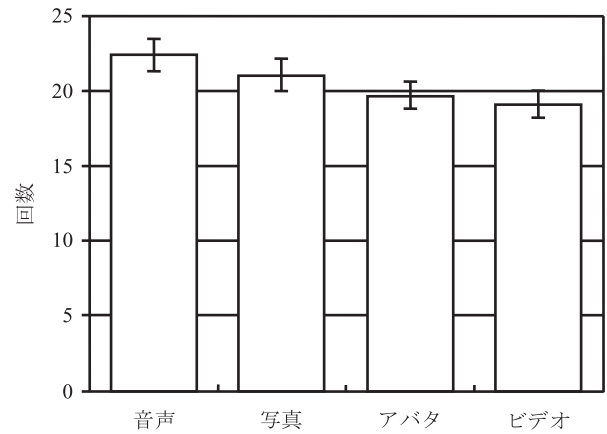


図 5 平均途切れ回数

Fig. 5 Average number of pauses.

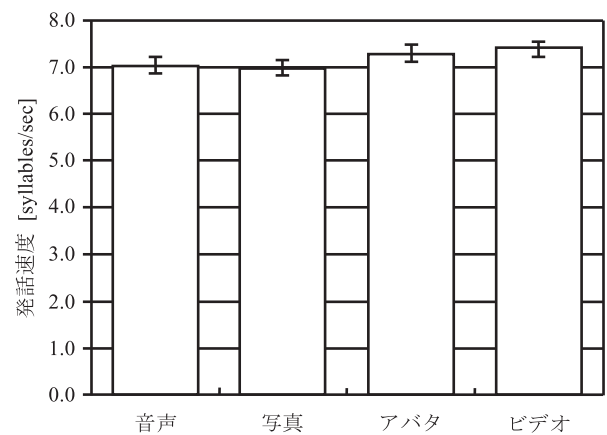


図 6 平均発話速度

Fig. 6 Average speech rate.

(有・無) と外見 (有・無) を独立変数、途切れ回数を従属変数とする対応ありの 2 要因分散分析を行った。各条件の平均値ならびに標準誤差を図 5 に示す。分析の結果、途切れ回数には身体動作要因の主効果のみ有意であり、身体動作を提示した条件は提示しない条件に比べ途切れ回数が減少していた ($F(1, 23) = 13.31, p < .01$)。外見要因の主効果および交互作用は有意ではなかった。

次に、身体動作の提示の有無または外見の提示の有無によって、発話速度が上昇するかどうか調べるために、身体動作 (有・無) と外見 (有・無) を独立変数、発話速度を従属変数とする対応ありの 2 要因分散分析を行った。各条件の平均値ならびに標準誤差を図 6 に示す。分析の結果、発話速度には身体動作要因の主効果のみ有意であり、身体動作を提示した条件は提示しない条件に比べ発話速度が上昇していた ($F(1, 23) = 19.01, p < .01$)。外見要因の主効果および交互作用は有意ではなかった。

途切れ回数と発話速度の分析の結果、身体動作の提示によって、途切れ回数が有意に減少し、発話速度が有意に上昇したことが分かった。このことは、実験者の身体動作の提示によって、被験者の発話が滑らかになったことを意味

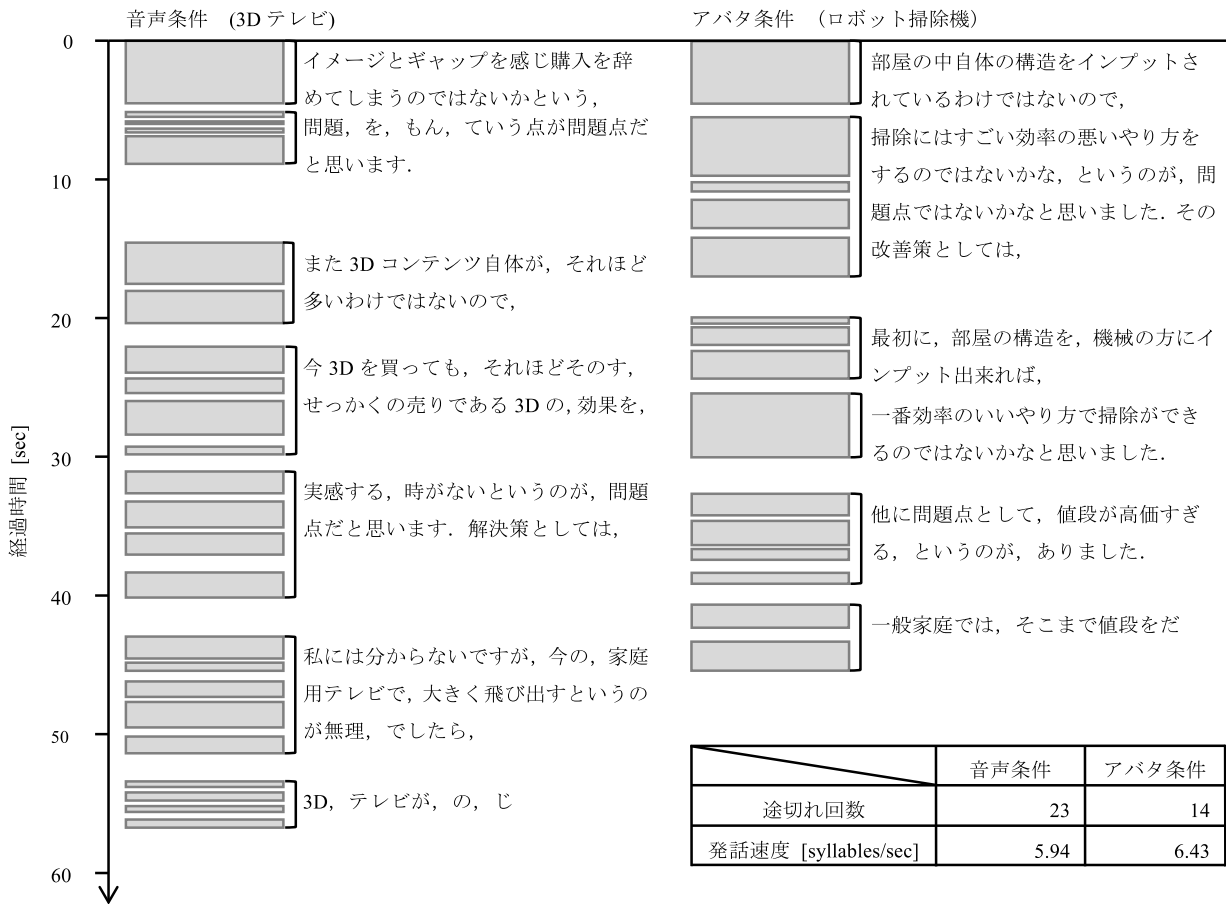


図 7 被験者の発話例
Fig. 7 Speech example.

している。以上の結果は、ビデオとアバタが音声コミュニケーションを円滑にするという我々の仮説を支持するものである。

図 7 は、ある被験者の中央 200 音節中の発話の例を示しており、各ブロックは被験者の発話を表し、各ブロックの間隔は途切れを表している。また、図 7 では、各ブロックに並列して、発話の書き起こし文を記し、文章中のカンマ「,」は途切れを表している。図 7 より、音声条件よりアバタ条件は途切れ回数が少なく発話速度が速いことが分かる。

5. 考察

本実験では、被験者に実験者のビデオを提示することで、被験者が実験者に向けた発話は滑らかになった。さらに、実験者の外見は再現せず、実験者の頭部と唇の動きを再現するアバタを提示することでも、ビデオを提示する場合と同様に、実験者に対する被験者の発話は滑らかになった。本実験で見られた滑らかな発話は、途切れ回数が少ないため 1 発話の音節数が多く、かつ発話速度が速い発話であった。

外見があり身体動作がある条件はビデオ条件、外見があり身体条件がない条件は写真条件とした。これに対応させ

た場合、外見がなく身体動作がある条件はアバタ条件、外見がなく身体動作がない条件は静止したアバタを提示する条件となると考えられる。本研究では、音声コミュニケーションに対する視覚情報の有効性を示すため、外見がなく身体動作がない音声条件を用いた。これによって、身体動作がない場合の外見要因の比較が、静止したアバタを提示する条件と写真条件から音声条件と写真条件となり、視覚情報の差が大きくなっている。それにもかかわらず、実験者の外見の提示について、途切れ回数の有意な減少は見られず、また発話速度の有意な上昇も見られなかった。本実験では、話し手と聞き手は初対面であり、親密な関係ではなかった。お互いに面識がある場合や親密な関係にある場合には、外見を提示することにより発話中の途切れに影響を与えるかもしれない。したがって、この実験結果から音声コミュニケーションに外見が不要かどうかを判断することはできない。

本実験では、ビデオ、アバタ、写真を表示するために小型のディスプレイを使用した。そのため、実験者の表示サイズは通常のビデオ会議より小さいサイズでの対話であった。本実験での対話者の表示サイズより小さいサイズでも、身体動作の提示によって音声コミュニケーションを円滑にするか否か検証する必要がある。なぜなら、対話者の

表示サイズが小さい携帯型テレビ電話でも、身体動作の提示によって音声コミュニケーションを円滑にするかもしれない。また、対話者を等身大で表示したビデオとアバタにも音声コミュニケーションを円滑にする効果があるか調べる必要がある。先行研究では、対話者を等身大で表示したビデオが対話者の存在感を強化させることが示されている [15]。しかし、等身大で表示したビデオとアバタが音声コミュニケーションを円滑にする効果があるか否かは分かっていない。

本研究では、フェイストラッキング技術により、頭部の位置・姿勢と唇の形状からアバタの動作を生成した。しかし、アバタは対話者の身体動作を忠実に再現する必要はないかもしれない。人間が操作するアバタと自動的に動くエージェントでは、存在感に有意差はないという報告がある [22]。途切れ回数の減少または発話速度の上昇が潜在意識による現象であるならば、身体動作が実際の聞き手の動作ではないと話し手が認識している場合でも、身体動作の提示は効果的かもしれない。過去の研究では、発話音声から話し手や聞き手の身体動作を推定し、アバタを用いて身体動作を共有する技術が開発されている [13]。その技術によって推定された身体動作は、音声コミュニケーションを円滑にする可能性がある。音声コミュニケーションを円滑にするうえで、身体動作の忠実性がどの程度必要であるのか検証することも今後の課題である。

本研究で使用したアバタは、3次元のCGのアニメーションである。アバタに物理的な身体を持たせることによって、発話を滑らかにする効果が増すかもしれない。先行研究では、特定の人物に酷似したアンドロイドロボットを用いた遠隔会議は、ビデオ会議よりも強い存在感があることが示されている [19]。一方、特定の人物の外見を持たないヒューマノイドロボットを用いた会議では、ビデオ会議より相手の状態が把握しやすくなるということが示されている [25]。特定の人物の外見に酷似したロボットと比較して、人にしか見えないが年齢や性別が分からない見かけを持つロボットの方が安価で汎用的であると思われる。今後、我々はそのようなロボットを使用して実験を行う予定である。

本実験で被験者にアバタで提示した身体動作は、頭部と唇の動きによる頷きのみであり、使用したアバタには眉や手はなかった。このことは、表情や腕の動きを提示しなくても音声コミュニケーションを円滑にすることを意味する。しかし、眉や手の動きの提示は、音声コミュニケーションを円滑にする効果を増すかもしれない。また、物理的な空間の位置移動を共有する機能を音声コミュニケーションに追加したシステムが開発されており [7], [12]、位置移動の提示も音声コミュニケーションを円滑にするかもしれない。今後、様々な身体的な動作の提示についても音声コミュニケーションを円滑にする効果があるか検討する必要がある。

6. 結論

ビデオの提示は、2者間での音声コミュニケーションでは不必要であると考えられてきた。しかし、話し手に聞き手の身体動作と外見を提示するビデオは、音声コミュニケーションを円滑にするということが分かった。そして、聞き手の身体動作のみを提示するアバタも、ビデオと同様に音声コミュニケーションを円滑にする効果があることを示した。したがって、この結果は、音声コミュニケーションを円滑にするには、外見を提示しなくても身体動作の提示だけでも十分であることを意味しており、アバタがビデオを代替する可能性が示唆された。

謝辞 本研究は、若手研究 (A) 「テレロボティックメディアによる社会的テレプレゼンスの支援」、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 (S) 「遠隔操作アンドロイドによる存在感の研究」、グローバル COE プログラム 「認知脳理解に基づく未来工学創成」、異分野融合による方法的革新を目指した人文・社会科学研究推進事業 「手話コミュニティにおける遠隔コミュニケーション環境の提案」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] Bailenson, J.N., Yee, N., Merget, D. and Schroeder, R.: The Effect of Behavioral Realism and Form Realism of Real-Time Avatar Faces on Verbal Disclosure, *Nonverbal Disclosure, Emotion Recognition, and Copresence in Dyadic Interaction, Presence: Teleoperators & Virtual Environments*, Vol.15, No.4, pp.359-372 (2006).
- [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] Colburn, R.A., Cohen, M.F., Drucker, S.M., Tiernan, S.L. and Gupta, A.: Graphical Enhancements for Voice Only Conference Calls, Microsoft Research Technical Report, MSR-TR-2001-95 (2001).
- [5] 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).
- [6] Dodds, T.J., Mohler, B.J. and Bühlhoff, H.H.: Talk to the Virtual Hands: Self-Animated Avatars Improve Communication in Head-Mounted Display Virtual Environments, *PLoS ONE*, Vol.6, No.10, pp.1-12 (2011).
- [7] Flintham, M., Anastasi, R., Benford, S., Hemmings, T., Crabtree, A., Greenhalgh, C., Rodden, T., Tandavanitj, N., Adams, M. and Row-Farr, J.: Where On-Line Meets On-The-Streets: Experiences With Mobile Mixed Reality Games, *Proc. CHI 2003*, Vol.5, No.1, pp.569-576

- (2003).
- [8] Garau, M., Slater, M., Bee, S. and Sasse, M.A.: The Impact of Eye Gaze on Communication Using Humanoid Avatars, *Proc. CHI 2001*, pp.309-316 (2001).
 - [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] Ishihara, S. and Kinoshita, Y.: Filler Words as a Speaker Classification Feature, *Proc. SST 2010*, pp.34-37 (2010).
 - [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. HICSS 2008* (2008).
 - [12] Nakanishi, H., Koizumi, S., Ishida, T. and Ito, H.: Transcendent Communication: Location-Based Guidance for Large-Scale Public Spaces, *Proc. CHI 2004*, pp.655-662 (2004).
 - [13] Ogawa, H. and Watanabe, T.: InterRobot: Speech-Driven Embodied Interaction Robot, *Advanced Robotics*, Vol.15, No.3, pp.371-377 (2001).
 - [14] Olson, J.S., Olson, G.M. and Meader, D.K.: What Mix of Video and Audio is Useful for Small Groups Doing Remote Real-time Design Work?, *Proc. CHI 95*, pp.362-368 (1995).
 - [15] Prussog, A., Muhlbach, L. and Bocker, M.: Telepresence in Videocommunications, *Proc. Annual Meeting of Human Factors and Ergonomics Society*, pp.180-184 (1994).
 - [16] Pye, R. and Williams, E.: Teleconferencing: Is Video Valuable or is Audio Adequate?, *Telecommunications Policy*, Vol.1, No.3, pp.230-241 (1977).
 - [17] Radford, G.P., Morganstern, B.F., McMickle, C.W. and Lehr J.K.: The Impact of Four Conferencing Formats on The Efficiency and Quality of Small Group Decision Making in a Laboratory Experiment Setting, *Telematics and Informatics*, Vol.11, No.2, pp.97-109 (1994).
 - [18] Rutter, D.R., Stephenson, G.M. and Dewey, M.E.: Visual communication and the content and style of conversation, *Social Psychology*, Vol.20, pp.41-52 (1981).
 - [19] Sakamoto, D., Kanda, T., Ono, T., Ishiguro, H. and Hagita, N.: Android as a Telecommunication Medium with a Human-like Presence, *Proc. HRI 2007*, pp.193-200 (2007).
 - [20] Sellen, A.J.: Remote Conversations: The Effects of Mediating Talk with Technology, *Human-Computer Interaction*, Vol.10, No.4, pp.401-444 (1995).
 - [21] Tanis, M. and Postmes, T.: Two Faces of Anonymity: Paradoxical Effects of Cues to Identity in CMC, *Computers in Human Behavior*, Vol.23, No.2, pp.955-970 (2007).
 - [22] von der Putten, A.M., Kramer, N.C. and Gratch, J.: Who's there? Can a Virtual Agent Really Elicit Social Presence?, *Proc. PRESENCE 2009* (2009).
 - [23] Williams, E.: Experimental comparisons of face-to-face and mediated communication: A review, *Psychological Bulletin*, Vol.84, No.5, pp.963-976 (1977).
 - [24] 中島悦子: 自然談話に現れるフィラー—自然談話録音資料に基づいて, *アジア・日本研究センター紀要*, Vol.4, pp.1-23 (2008).
 - [25] 森田友幸, 間瀬健二, 平野 靖, 梶田将司, 岡留 剛: ヒューマノイドロボットを用いた遠隔コミュニケーションにおける注目伝達, *情報処理学会論文誌*, Vol.48, No.12, pp.3849-3858 (2007).



尾上 聡

2010年岡山県立大学情報工学部情報システム工学科卒業。2012年大阪大学大学院工学研究科知能・機能創成工学専攻博士前期課程修了。現在、富士通株式会社に所属。



山本 健太

2010年大阪大学工学部応用理工学科卒業。2012年同大学大学院工学研究科知能・機能創成工学専攻博士前期課程修了。現在、株式会社リンテックに所属。



田中 一晶 (正会員)

2006年京都工芸繊維大学工学部電子情報工学科卒業。2008年同大学大学院工学科学研究科情報工学専攻博士前期課程修了。2011年同大学院工学科学研究科情報工学専攻博士後期課程修了。博士(工学)。同年より、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。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。FIT2007論文賞。



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。2005年ジョージア工科大学客員研究員。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。遠隔会議を改善する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。