

身体動作の提示による遠隔対話の円滑化

尾上 聡[†] 山本 健太[†]
田中 一晶^{†,‡} 中西 英之[†]

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つの方法が挙げられる。しかし、これらの方法による音声対話への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、発話中の途切れの頻度と割合に注目した。そして、我々は視覚的に提示する情報について身体動作と外見の要因に分け、音声対話への効果を確認するため4つの条件を設定した。設定した4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、何も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、発話中の途切れの回数と割合を調べた。その結果、外見の提示の有無に関らず、身体動作の提示によって途切れの回数と割合が減少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対話が円滑化していること示しており、アバタがビデオを代替する可能性も示唆している。

Presentation of a Partner's Body Motion Smoothens Remote Conversation

SATOSHI ONOUE,[†] KENTA YAMAMOTO,[†] KAZUAKI TANAKA,^{†,‡}
and HIDEYUKI NAKANISHI[†]

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 spoken dialogue by these methods. In order to obtain such data, we focused on the frequency and proportion of pauses included speech in spoken dialogue. We separated information of the partner's visually presentation into the factor of the partner's body motion and the partner's appearance, set up four conditions to evaluate the effect. These conditions consisted of the video condition and the avatar condition to present the body motion, the video condition and the photograph condition to present the appearance and the audio condition to present nothing. We conducted an experiment that the subjects spoken about topics under these conditions, and analyzed the number and proportion of pauses included the subject's speech. The analyses revealed that the presentation of the body motion reduced the number and proportion, regardless of presentation of the appearance. This results show the presentation of the body motion to smoothen spoken dialogue, also indicate the possibility that avatar replaces video.

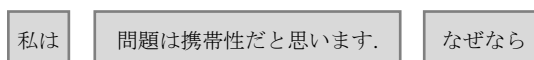
1. はじめに

遠隔地で音声対話を行う際、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。

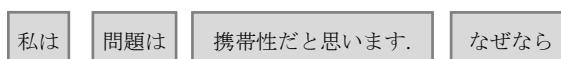
長年、遠隔会議でビデオを用いることの必要性が議論されてきた¹³⁾。その必要性を観察データから明らかにするために、多くの研究が行われてきたが、未だ明らかにされていない^{1),5),12),14),16)}。

近年のトラッキング技術の向上に伴い、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することが可能なアバタが、容易かつ安価に製作できるよ

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の例

うになった。アバタはユーザの外見の情報を伝えず、身体的な動きの情報のみを伝えることができる。しかし、アバタを用いることで、音声対話が改善されるかどうかは分かっていない^{2),3),7),8)}。

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いる。イ

[†] 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

[‡] 独立行政法人科学技術振興機構, CREST
Japan Science and Technology Agency, CREST

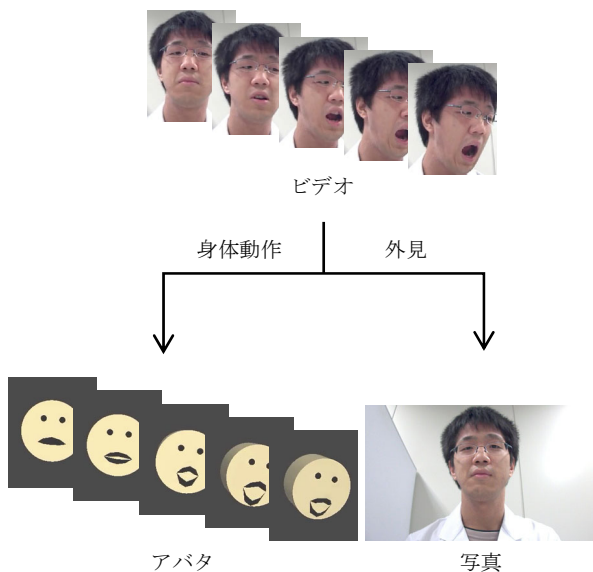


図2 ビデオに含まれる身体動作と外見の情報

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が、音声対話で対話者の顔写真を提示することの効果は明らかになっていない^{4),17)}。

音声対話を観察データから評価する指標として、同時発話の回数及び継続時間などがある。しかし、同時発話の回数及び継続時間では、対話者の表示方法による音声対話の改善は示されていない¹⁶⁾。そこで、本研究では評価指標として、1回のターンの発話に含まれる途切れの頻度と割合に着目した。途切れの頻度が少なく途切れの割合が低い発話を滑らかな発話、途切れの頻度が多く途切れの割合が高い発話をぎこちない発話と称し、図1にそれらの発話の例を示す。1つのブロックはひと続きの発言を表しており、ブロックとブロックの隙間が1回の途切れを表している。我々は途切れの頻度と割合の減少が音声会話の改善であると仮定し、1回のターンの発話に含まれている途切れの頻度と割合を分析する。

そして、一般的に過去の研究では、ビデオは一つの要因として扱われてきたが、ビデオを用いた対話には対話者の音声以外に対話者の身体的な動きと外見の2つの情報が含まれている。そこで、我々是对話者の顔ききといった身体動作の効果と対話者の外見の効果に分けて分析を行った。本研究では、図2に示すように、ビデオを用いた対話は身体動作と外見の情報を伝達していると仮定する。そして、アバタは身体動作の情報のみを伝達し、写真は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、身体動作の効果と外見の効果について分析することが可能となる。

2. 関連研究

ビデオ、アバタ、写真について議論された関連研究は非常に多く存在する。しかし、ビデオが2者間の音声対話に良い影響を与えたことを観察データで示した研究はほとんど存在しない。

音声会話に対するビデオ会議の優位性を調べるため、多くの実験が行われてきた。いくつかの実験では、音声のみの場合よりビデオを用いた場合のコミュニケーションの方が会話をする上で効果的であることが示されており^{5),12),16)}、また、多人数会話を容易に行えることを示している⁹⁾。しかし、2者間の音声対話では、会話構造^{1),5),16)}やタスクパフォーマンス^{1),12),14)}のような観察データでは、ビデオによる音声対話の改善を示すことができなかった。

最近、メディアを介したコミュニケーションの研究に関して、トラッキングシステムを用いて実時間で人間の動きを反映するアバタを使った研究がなされている^{2),3),7),8)}。しかし、アバタに音声対話を改善する効果があるかは分かっていない。

ビデオとアバタに加え、写真もまた対話者の提示方法の一つとして研究されている^{4),17)}。しかし、未だに、音声対話に関して、対話者の写真を提示することの影響は明らかになっていない。

本研究がこれらの関連研究と異なる点は、音声対話の評価手法として発話に含まれる途切れの頻度と割合といった会話の細かい構造に注目したことである(図1)。また、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、発話に含まれる途切れの頻度と割合にどのような影響を及ぼしているのかをそれぞれ調査した。

3. 実験

3.1 仮説

本研究の仮説は以下の通りである。

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と割合が減少する。

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と割合が減少する。

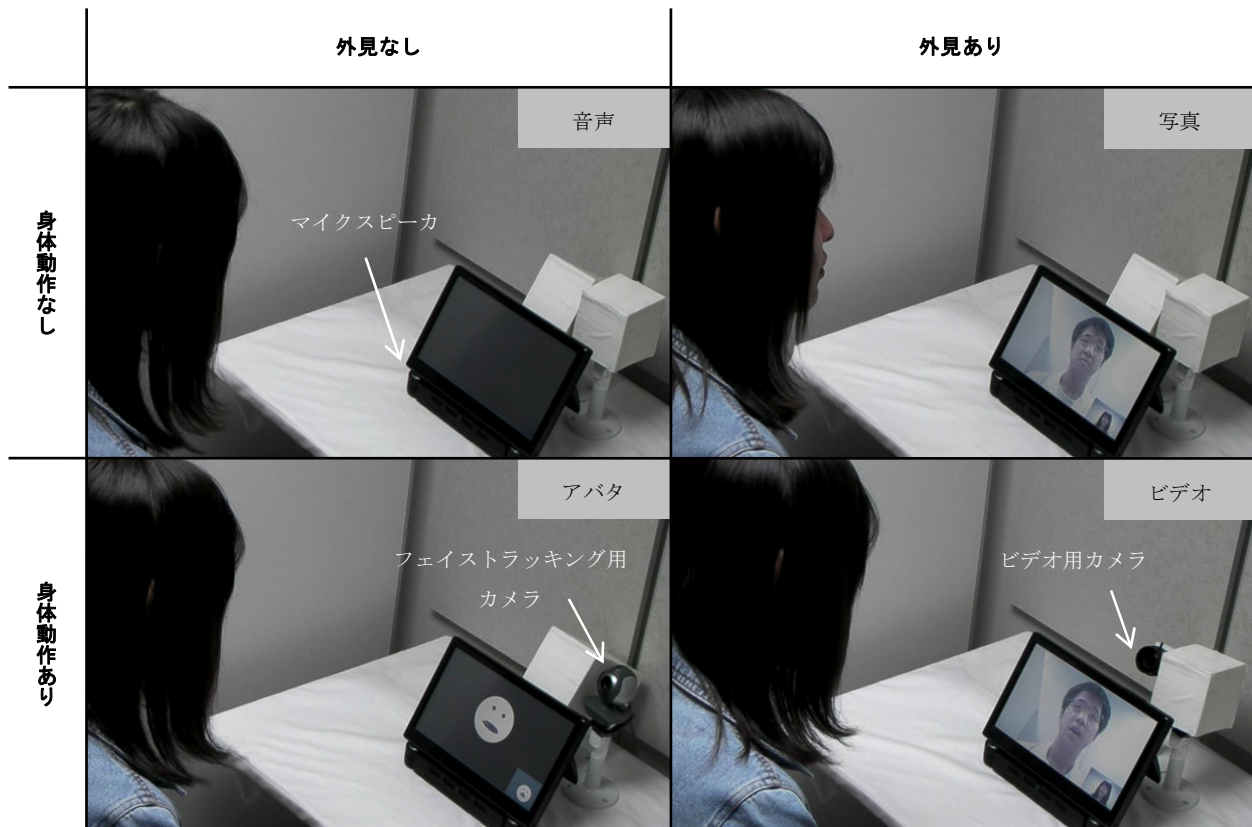


図3 実験条件

ビデオ会議システムにおいて、先行研究では、遠隔地の対話者の動きを強調することにより、対話者の存在感が向上することが示されている¹⁰⁾。聞き手の存在感によって、相手は話しやすくなると我々は考えており、身体動作は外見とは無関係に効果があると予想している。したがって、ビデオだけでなく、聞き手の動き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ことを期待している。

3.2 条件

仮説は身体動作と外見の2要因から成り、各要因は身体動作の有無と外見の有無の2水準を持つ。そして、この仮説を実証するため、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。話し手と聞き手ともに、4条件ともテレビ会議端末が置かれた机の前に座った状態で実験を行った。このテレビ会議端末は、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している。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。これは聞き手を視覚的に表示する領域が狭くても、音声対話が改善することを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。話し手はマイクスピーカのみを介して聞き手に話しかける。話し手に、聞き手と視覚的な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った。

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でのボイスチャットと同じである。この条件と音声条件の異なる点は、ディスプレイに聞き手の顔写真を表示するという点である。また、話し手が聞き手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、話し手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。話し手と聞き手の顔写真は実験の開始時に撮影したものをを用いる。また、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取りであることを話し手に伝える。

アバタ条件（身体動作あり、外見なし）：この条件はディスプレイに個人を特定しないデザインのアバタを表示する。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、そのユーザの外見は伝達しない。また、この条件は、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、唇と眼球で構成している。まず、アバタの頭部のデザインに関しては、球状の頭部より円筒状の頭部の方が顔の向きの変化を理解しやすいと考え、アバタの頭部には円筒状の頭部を用いた。次に、アバタの頭部の大きさに関しては、写真条件とビデオ条件でディスプレイに表示した聞き手の顔の大きさとほとんど同じである。そして、アバタに用いた色彩に関しては、使用する色彩を最小限に抑えるために、アバタの唇の色は赤ではなく、濃い灰色を用いた。また、眼球と背景も濃い灰色を用い、アバタの頭部は淡い黄色を用いた。また、アバタの頭部に用いた自由度は、並進3自由度と回転3自由度で合計6自由度である。そして、アバタの唇の形状は8点の特徴点の3次元位置から決定される。また、アバタの眼球は球体でアバタの顔に固定した。フェイストラッキング用ソフトウェア（faceAPI）は33ms毎にトラッキングデータを更新し、アバタはそれを反映して頭部と唇が動く。そして、話し手の頭部と唇の動きもまた取得し、聞き手側のディスプレイに表示される。また、話し手が、聞き手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、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、話し手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。また、この条件は、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。聞き手のビデオをディスプレイに表示し、ディスプレイの右下に小さく話し手のビデオを表示する。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。

以後、音声条件、写真条件、アバタ条件、ビデオ条件の4条件をメディア条件と呼ぶ。また、被験者が実験に慣れてもらうための条件として対面条件を設定した。

対面条件：この条件は一般的な対面環境である。話し手と聞き手は机を間に挟むように座り、対面での対話を行う。

今回の実験は、話し手役を被験者、聞き手役を実験者が行った。そして、この実験は各被験者が全5条件を行う被験者内計画で行った。全ての被験者は、下記に述べるタスクおよび実験者に慣れてもらうため、初回に対面条件での対話を行い、その後、メディア条件

での対話を行った。メディア条件の実験の順序はランダムに決めた。

3.3 タスク

途切れの頻度と割合を安定に計測するため、比較的長い発話を記録する必要がある。同時に、実験者の返答が被験者の発言に介入した場合、被験者の発言に影響を及ぼすため、実験者の介入を避ける必要もあった。そこで、我々は実験者の介入なしに、被験者が1分以上話し続けることができるようなタスクを用意した。被験者が話している間、実験者が行う身体動作は、相槌を打ちながら頷くのみとした。

実験者は、被験者に条件毎に電子機器の問題点と改善点と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を探ねた。全被験者は5条件での対話を行うため、我々は、電子ブックリーダー、携帯ゲーム機、スマートフォン、ロボット掃除機と3Dテレビの5つの電子機器の話題を採用した。また、5つの話題の順序はランダムに決めた。我々は事前に話題を被験者に伝えず、条件の開始時に話題を被験者に伝えた。

我々は被験者に一定時間話し続けるような指示はせず、被験者はいつでも対話を終了することができた。また、最近話題になっている電子機器を選んだため、ほとんどの被験者は電子機器の問題点と改善点についてある程度知っており、1分以上話し続けることができた。1分の発話はターンテイキングの分析には短すぎると思われるが、途切れの頻度と割合の変化を計測する上では十分な長さであった。

3.4 データ収集

本大学の近辺に在住する学部生9名が実験に参加し、全5条件で電子機器について話してもらった。そして、合計45個の発話を録音した。また、5条件の実験を実施した後、被験者にはアンケートに答えてもらった。

途切れの頻度と割合を分析するため、一定量の発話を抽出する必要がある。発話の開始部分と終了部分には、被験者によって個人差があると考えられる。たとえば、実験の話題で用いる電子機器を被験者が購入しようとしていた場合、開始部分では滑らかになる傾向があった。また、被験者が無理に話し続けようとした場合、発話の終了部分で被験者はぎこちなくなる傾向があった。そのため、我々は発話の中央部分を抽出した。また、発話の分量を測る単位として、最小の発音の単位である音節数を用いた。図4に途切れの回数と時間を求める手順を示す。図中の各ブロックは発言を表し、各ブロックの隙間は途切れを表している。また、下記の式に基づいて途切れの割合を計算した。ここで、全体の時間とは、抽出した中央部分の途切れを含めた

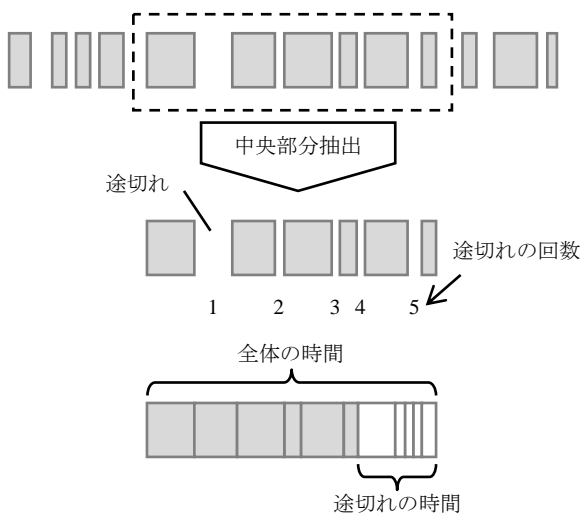


図4 途切れの回数と時間を求める手順

全体の時間である。

途切れの割合 = 途切れの時間 / 全体の時間

全ての被験者の発話は1分以上あり、1分間の発話は日本語で約200音節であった。したがって、我々は分析のために被験者の発話の中央200音節を抽出した。

分析者の個人差を除外するため、我々は途切れの時間幅の最大、最小の閾値を設定せず、発話中に無音区間があるならば、それを途切れであると判別した。しかし、ホワイトノイズにより、50 ms以下の無音区間を途切れと判別することが困難だったため、50 ms以下の無音区間を我々は途切れとしなかった。また、フィルタリングによって音声データのホワイトノイズを除去した場合、音量の小さい発言までが除去される可能性があったため、フィルタリングすることはしなかった。

録音した発話がどのように処理されたかは以下の通りである。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) を使用して、発言の開始時刻と終了時刻を入力し、その発言全てを書き起こした。録音した発話の書き起こしが終了後、書き起こしのデータはテキストファイルにエクスポートした。そして、そのテキストファイルのデータをスプレッドシートに貼り付け、発話の中央部分の途切れの回数と途切れの割合を求めるために表計算ソフトを使用した。

4. 実験結果

図5に、メディア条件の途切れの回数と割合の平均値ならびに標準誤差含むグラフを示す。メディア条件全体の中央200音節中の途切れの回数の平均値は約20回、途切れの割合の平均値は約0.294であった。図5から、身体動作を提示しない条件 (音声条件と写真

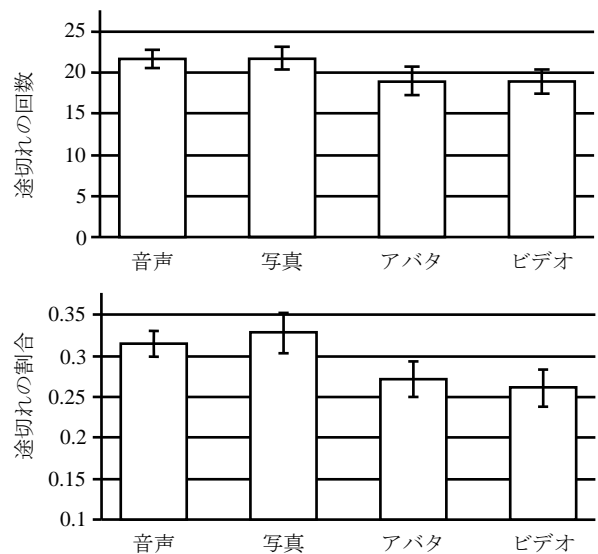


図5 途切れの回数と割合の平均値

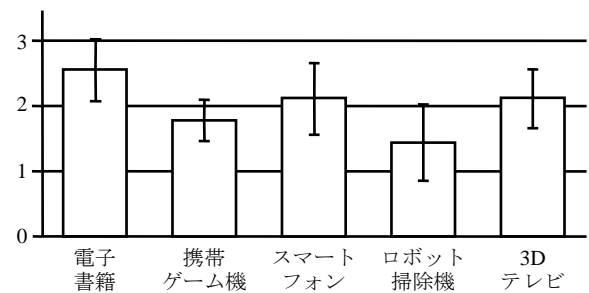


図6 アンケート結果

条件)と比較して、身体動作を提示した条件 (アバタ条件とビデオ条件)は途切れの回数と割合が減少していることが分かる。

そして、身体動作と外見の要因の効果を見るために、中央200音節中の途切れの回数と割合を分析した。ただし、対面条件は他の条件とは異なり、常に初回に行ったため、分析対象から外した。

我々は身体動作と外見の要因による効果を調べるため4つの条件を比較した。各要因は2水準あり、人の発話には個人差があるため、各被験者は全条件で実験を行った。したがって、この実験で得られた観察データは対応ありの2要因分散分析で比較した。分析の結果、途切れの回数には身体動作の主効果が見られた ($F(1,8)=23.585, p<.01$)。また、途切れの回数で外見の主効果は見られず、交互作用も見られなかった。一方、途切れの割合についても、身体動作の主効果が見られた ($F(1,8)=12.785, p<.01$)。そして、途切れの割合での外見の主効果は見られず、交互作用も見られなかった。このことは、実験者の身体動作を視覚的に提示することが被験者の発話を滑らかにすることを意味し、実験者の外見の欠如は発話を滑らかにすることに影響しないことを意味する。そして、これらの結果は

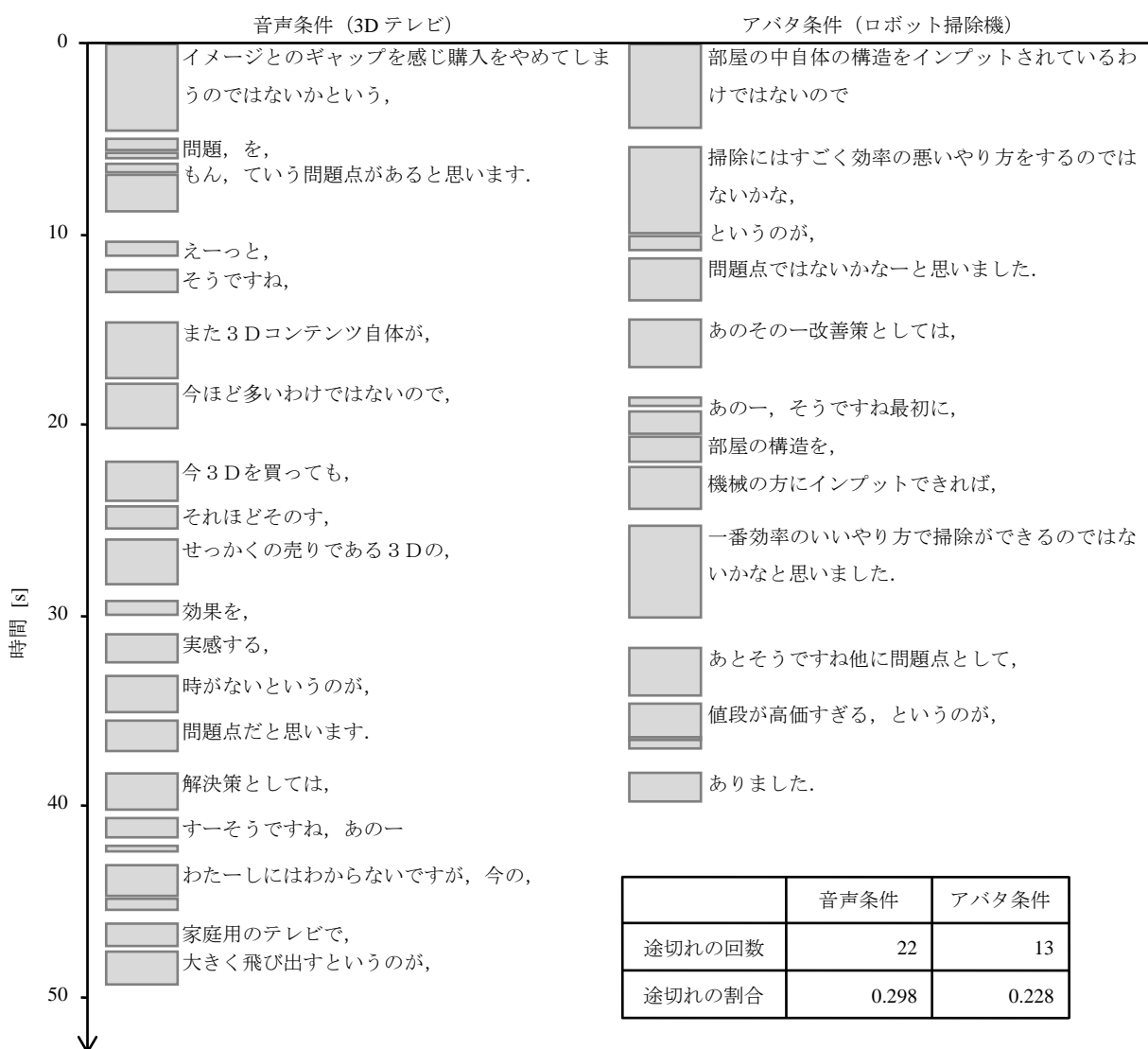


図7 ある被験者の中央 200 音節中の発話

ビデオとアバタが音声対話を滑らかにするという我々の仮説を支持するものである。

次に、話題の種類による話しやすさへの影響を調べるため、話題について最も話しやすいものから最も話しにくいものまで、計 5 段階で被験者に評価してもらった。評価してもらった順位をもとに、1 位から 5 位の話しやすさの順に 4 点から 0 点と得点化した。図 6 に、各話題の得点の平均値と標準偏差を含むグラフを示す。

全被験者は全条件を行った後に各話題について評価しているため、アンケート結果に対して対応ありの 1 要因分散分析を行った。その結果、話題の種類には有意差は見られず、よって、話題の種類は全体的な実験結果に影響しなかったと思われる。

図 7 では、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言を表し、各

ブロックの隙間は途切れを表している。同様の 200 音節の発話量で、アバタ条件は音声条件よりも途切れの回数と途切れの割合が減少していることがわかる。中央部分からの 200 音節の時間幅は約 1 分であった。図 7 より、対話中に、実験者の身体動作を見ることができない場合、被験者の発話は途切れ途切れになっていることが分かる。

以上の結果を次のようにまとめる。身体動作の提示は被験者の発話を滑らかにするのに十分であった。一方、外見の提示に発話を滑らかにする効果は見られなかった。そして、話題の種類による話しやすさへの影響は見られなかった。

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、その実験者に向けた発話を滑らかにした。さら

に、実験者の外見は再現せず、その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、ビデオと同様に被験者の発話を滑らかにした。

本実験では、外見の要因を強調するため、身体動作も外見も無い条件として、アバタの静止画をディスプレイに表示するのではなく、何も表示しない音声条件を用いた。それにもかかわらず、本研究では実験者の外見を被験者に提示する写真条件での発話は音声条件より滑らかにはならなかった。実験者が被験者と面識があれば、この結果は変わるかもしれない。したがって、我々はこの実験結果から音声対話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、小型のディスプレイを使用した。そのため、実験者のビデオとアバタの表示サイズは小さかった。どの程度表示サイズが小さくても音声対話を円滑にするのか調べることも興味深いところである。携帯電話のテレビ電話さえも音声対話を円滑にする上で効果的かもしれない。また、等身大のビデオとアバタの利点を調べることも興味がある。対話者を等身大で表示することによる存在感の強化が、音声対話の円滑化に効果的なのか、話しやすく感じるのかはまだ分かっていない。

本研究では、フェイストラッキング技術によって実験者の身体動作からアバタの動きを生成した。しかし、アバタに関して、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。過去の研究では、発話の音声情報からアバタが話している動作や聞いている動作に変換する技術が開発されている¹¹⁾。その技術によって作り出された動作は、音声対話を改善する可能性がある。途切れの回数と割合の減少が潜在意識の現象であるならば、その動作が実際の相手の動きではないとユーザが分かっている場合でも、恣意的に作られた動作は効果的かもしれない。また、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¹⁸⁾。動作の忠実性がアバタを効果的にする上でどの程度必要であるのかを検証することも今後の課題である。

本研究で我々が使用したアバタは、2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¹⁵⁾。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¹⁹⁾。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。

本研究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った。眉や手の動き、位置の移動などを視覚的に提示することでも、音声対話を改善する可能性がある。我々のアバタには眉や手はない。このことは、音声対話に関して、表情や腕の動きを提示せずとも音声対話が滑らかになったことを意味する。しかし、眉や手の動きによって、発話を滑らかにする効果が増す可能性がある。また、物理的な空間における位置の移動を共有する機能を、音声対話に追加した先行研究がある^{6),9)}。位置の移動の共有が音声対話をどのように改善するかは興味深いところである。

6. 結論

ビデオの提示は、2者間での音声対話では不必要であると考えられてきた。しかし、話し手に聞き手をビデオで提示した場合、話し手の発話が滑らかになるということが分かった。そして、聞き手の外見を隠し、身体動作を提示するアバタでも、ビデオと同様に発話を滑らかにする効果があることを示した。したがって、この結果は、音声対話を円滑にするには身体動作の提示だけでも十分であることを示唆しており、アバタがビデオを代替する可能性があることも意味している。

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(研究領域:共生社会に向けた人間調和型情報技術の構築)」、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) Flintham, M., Anastasi, R., Benford, S., Hemmings T., Crabtree, A., Greenhalgh, C., Rodden, T., Tandavanitj, N., Adams, M. and Row-Farr, Ju.: Where On-Line Meets On-The-Streets: Experiences With Mobile Mixed Reality Games, *Proc. CHI 2003*, Vol.5, No.1, pp.569-576 (2003).
 - 7) 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).
 - 8) 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).
 - 9) 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).
 - 10) Nakanishi, H., Kato, K. and Ishiguro, H.: Zoom Cameras and Movable Displays Enhance Social Telepresence, *Proc. CHI 2011*, pp.63-72 (2011).
 - 11) Ogawa, H. and Watanabe, T.: InterRobot: Speech-Driven Embodied Interaction Robot, *Advanced Robotics*, Vol.15, No.3, pp.371-377 (2001).
 - 12) 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).
 - 13) Pye, R. and Williams, E.: Teleconferencing: Is Video Valuable or is Audio Adequate?, *Telecommunications Policy*, Vol.1, No.3, pp.230-241 (1977).
 - 14) 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).
 - 15) 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).
 - 16) Sellen, A.J.: Remote Conversations: The Effects of Mediating Talk with Technology, *Human-Computer Interaction*, Vol.10, No.4, pp.401-444 (1995).
 - 17) 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).
 - 18) 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).
 - 19) 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).