

アバターと音声変換技術が講話能力の向上に与える影響(1) : 話し手自身による主観評価による検討

The Impact of Avatar- and Voice Transformation Technology on Speaking Ability (1): Subjective Evaluation by Speakers Themselves

鷹阪 龍太[†], 安久 絵里子[†], 目黒 文乃[†], 葛岡 英明[‡], 原田 悦子[†]

Takawaki Ryuta, Eriko Ankyu, Ayano Meguro, Kuzuoka Hideaki, and Etsuko T. Harada

[†]筑波大学, [‡]東京大学

[†]University of Tsukuba, [‡]The University of Tokyo

{takawaki@tsukaiyasusa.jp, etharada@human.tsukuba.ac.jp}

概要

本研究では、腹話術形式の講話能力をとりあげ、アバターならびに音声変換技術を組み込んだ仮想的 2 者対話形式のシステムを用いることによる講話支援の可能性があるか否かを、システムを用いた「練習」段階時の講話ならびに前後の講話パフォーマンスを対象とした話し手自身の主観評価の結果から検討した。その結果、アバターの操作スキルについては短時間の練習で習得可能であることが確認でき、アバターの活用による支援の可能性が示唆された。その一方で、話し手自身の主観的評価においては、アバターおよび音声変換技術を利用した練習を経ても講話達成の向上は確認されなかった。

キーワード：遠隔対話システム、アバター、音声変換、プロテウス効果、腹話術形式の講話

1. 問題と目的

同一の人物が複数の声色を切り替えて直接話法形式で対話する話法の一つに腹話術がある。疑似的な対話形式であること、語り分けの面白さなどの特徴から腹話術を用いることは講話の有効な形態の一つであると考えられている一方で、腹話術人形の操作は複雑であり、人形の体や口を講話内容にそって適切に作動させるためには専門的な操作スキルが求められる点、対話形式で進められる講話の中で、話し手の交代にあわせて、キャラクターに適した声色を演じ分ける必要がある点、またこれらのスキルを講話の中で並行して用いる必要がある点が、腹話術形式での講話において大きな障壁となっている。また、これらの「腹話術のためのスキル」の習得には時間的、認知的資源を割く必要があること、こうした達成上の認知的負荷により、講話内容を十分に理解する、適切な間で発話を行う、聴衆へ聞こえやすいよう明瞭に発話するといった本来の講話に必要な能力の習得にも影響があるものと考えられる。

そこで、こうした「腹話術実施のためのスキル」利用とその習得という問題に対し、腹話術人形の操作については、操作者の身体運動を自動的に認識し操作対象のキャラクターの運動に変換するシステム（アバター）

を用いること、また話し手の声色の演じ分けについてはリアルタイムでの音声変換技術を用いることで、スキルの習得における時間的・認知的負荷を軽減し、本来の講話能力の習得を支援できる可能性が考えられる。

また、アバター利用時に、用いられる外見の変化によってユーザーの行動や態度が変わるというプロテウス効果(Yee & Bailenson, 2007; Yee et al., 2009)を踏まえると、仮想の対話相手の見た目を幼い子どもやキャラクターなどの想定に合わせて設定することで、対話相手の発話ターンでより上手く振る舞うことができ、話し手本人と対話相手との演じ分けが上達する可能性が考えられるが、こうしたプロテウス効果が聴覚的特徴においても生じうる可能性については検討が必要である。また対象に対して抱く「かわいい」という感情は、対象と共存したいという接近動機の増大をもたらすことが指摘されており(井原ら, 2011)、アバターに「かわいい音声」を付与することで、話し手にとっての仮想の対話相手に対する接近動機が高まり、練習時間の増加が生じ、結果として演じ分け能力の向上につながる可能性も考えられる。

そこで本研究では、アバターのみ、あるいはアバターと音声変換技術を組み込んだ仮想的 2 者対話システムを用いた腹話術的講話の練習段階を設けて、アバターおよび音声変換の有無を操作した 3 つの支援条件下での講話練習を実施することにより、これらの技術の利用によって講話能力が向上するか、システムを利用しない状況での講話能力も向上するかを検討するため、実験操作となる練習段階のほか、その前後にパペットを用いた講話を実施した。本報告では、練習段階ならびにその前後の講話実施についての話し手自身の主観評価・行動に基づく検討結果を報告する。

2. 方法

参加者 4 年制大学学生 30 名(男性 15 名, 女性 15 名,

平均年齢 20.83 歳, SD=1.61).

実験材料：交通安全に関する講話台本 警察官である講話者とロボタン（架空キャラクター：ロボタンと命名）の対話形式で構成される高齢者向け交通安全講話3本(原田ほか, 2022)を使用, 実施順は全員同一とした.

パペット 全ての参加者は, 講話内で「ロボタンとの会話」を演じる際にアバターもしくはパペットを使用するよう求めた. パペットは2種類が用意され(Figure 1), 実験開始前に参加者自身が一つを選択した.



Figure 1 実験で使用したパペット

アバター 「練習時間」に利用するアバターとして, 自分の手の動きで操作するアバター(Figure 2: 操作には ultraleap 社製の Stereo IR 170 Evaluation Kit を利用)をCGで合成した. 可能な操作は「体を左右に向ける」「うなづく」であり, 自動的なランダムな動きとして「ふわふわ上下に動く」を実行した.



Figure 2 実験で使用したアバターとアバターシステム利用時の様子

音声変換 実験参加者の音声を Roland 社製の VT-4 を用いてリアルタイムで変換した. 変換はピッチとフォルマントを対象に行い, 参加者自身が「かわいい」と感じる値に調整するよう求められ, 決定された.

評価項目 パフォーマンスに関する主観評価 講話のパフォーマンスに対する主観的評価を測定する6項目(Table 1)をVASで測定した(0-100).

パペット/アバターに対する生物感知覚 パペットやアバターに対して感じる生物らしさに関する検討を行うため, 長谷川・原田・栢野・大澤 (2015) の生物らしさ知覚を問う質問5項目, および小柳・鳴海・Jean・安藤・大村 (2020) を基に作成した2項目について, パペ

ットとアバターそれぞれに対して, まったくあてはまらない(1点)から「非常にあてはまる(6点)」の6件法で回答を7項目尋ねた (Table 2).

その他の測定項目 講話パフォーマンスへ影響すると考えられる参加者の特性として「人前で話すことに対する意識」, 腹話術形式の講話についての主観的印象, VT-4の操作/アバター操作の困難度を尋ねた.

Table 1 パフォーマンスの評価項目とVASのラベル

質問項目	左端(0)	右端(100)
①いまの話の出来は	不満だ	満足だ
②いまのあなたは	不安だ	安心している
③いまのあなたは	緊張している	リラックスしている
④台本は	話しにくい	話しやすい
⑤ロボタンは	かわいくない	かわいい
⑥ロボタンを動かすのは	難しい	簡単だ

Table 2 生物感知覚の測定項目

- | | |
|---|--------------------------|
| 1 | 実際に同じ部屋の中であなたのそばにいる感じがした |
| 2 | 話すのは楽しかった |
| 3 | 感情があるようだ |
| 4 | 生きているようだ |
| 5 | 目的をもって行動できそうだ |
| 6 | 計画を立てることができそうだ |
| 7 | 考えることができそうだ |

注: 全ての質問項目は主語「{パペット | アバター} (ロボタン) は～」の後に続けて使用された

練習時間 実験時の動画データを用いて, 実験者が「練習を始めてください」といった時点から, 参加者が実験者に練習の終了を意図する声掛けをした時点までを練習時間として計測した.

実験計画 練習条件(参加者間: パペットのみ条件/アバター+地声条件/アバター+VT-4条件)×試行段階(参加者内: pre/練習/post)の2要因混合計画. 一般にコミュニケーション課題では性差が見られること, 「かわいい」に関する評価についても性差が大きいとされることから, 各練習条件に男女5名ずつをランダムに割り当てた.

手続き 実験の説明完了後, 課題準備として講話台本を紙媒体で一度目を通した後, 同じ台本をiPadの講話台本表示画面で呈示し, 参加者が台本の内容を把握できた段階で, 机横のカメラに向かって講話を行うよう求めた (pre パフォーマンス). 課題に当たっては, 言い間違えや言いよどみがあっても, そのままパフォーマンス

ンスを続けるよう教示した。課題終了後、参加者は自分のパフォーマンスについて自己評価を行った。

次に、練習課題を実施した。練習課題では、すべての参加者がヘッドセットをつけ、コンピュータディスプレイを見て行うよう、設定を変更し、話者正面に設置した web カメラにより参加者が話す様子およびパペットの動きを録画した。練習条件のうち、パペットのみ条件（以下「なし条件」）では、pre パフォーマンスと同じパペットを使用して練習を行うよう求め、話者自身の声ならびに Web カメラで撮影した画像がそのまま提示された。アバター+地声条件（以下「地声条件」）では、練習課題ではアバターを用いて講話の練習を行うよう求め、アバター操作方法の教示後、基本動作の実施を確認後、参加者の様子とアバターが動く CG を合成した画像をモニター画面にリアルタイムに投影した。ヘッドセットからは、参加者の発する音声がそのまま聞こえるよう設定した。アバター+VT-4 条件（「VT-4 条件」）では、地声条件と同様にモニターに画像を提示すると同時に、音声は話者がボタンを押している間、話者音声が VT-4 で変換されたものが提示されるよう設定された。二者（警察官/ロボタン）の短い会話文を呈示し、音声を切り替えて発話する操作を獲得したことを確認のち、練習を実施した。

練習時間は最大 20 分とし、参加者の申告によって 20 分を待たずに練習を中断することが可能であった。練習終了後、参加者に練習段階でのパフォーマンスについての自己評価を求めた。

その後、pre パフォーマンスと同様に、パペットを用いた講話撮影(post パフォーマンス)を行った。その後、事後質問紙への回答とインタビューを行った。実験の所要時間は 1 時間程度であった。

3. 結果

実験時の観察から講話の方法などに大きく性差が見られたため、練習条件（なし条件 / 地声条 / VT-4 条件）×試行段階（pre / 練習 / post）×性別（男性 / 女性）の三要因混合計画分散分析を実施した。明確な有意差あるいは一定以上の効果量があった結果のみ、詳細に報告する。

練習時間 アバターの使用および音声変換の有無が接近動機へ与える影響を検討するため、練習時間を常用対数変換したものを従属変数として、性別（男性 / 女性）×練習条件（なし条件 / 地声条 / VT-4 条件）の二

要因分散分析を実施した。その結果、性別、練習条件の主効果および両者の交互作用はいずれも有意ではなく、アバターの使用や音声変換の実施は、講話の練習時間に影響を与えていなかった（Table 3）。

Table 3 性別、条件別の平均練習時間（秒）

性別	練習条件	平均値	標準偏差
女性	なし	841.200	327.314
	地声	864.400	336.151
	VT-4	635.800	299.290
男性	なし	871.600	198.008
	地声	738.600	222.496
	VT-4	985.400	292.105

次に、パフォーマンスの主観評価への影響を検討するため、性別（男性 / 女性）×練習条件（なし条件 / 地声条 / VT-4 条件）×試行（pre / 練習 / post）を独立変数、それぞれの評価項目得点を従属変数とした分散分析を実施した（Table 4）。その結果、「講話の出来」にすべての主効果並びに交互作用に有意な効果は見られなかった。

「安心感」については試行段階の主効果（ $F(2, 48) = 3.81, p \eta^2 = .14, p < .05$ ）が見られ、多重比較の結果、pre 試行と比較して post 試行における安心感得点が有意に上昇していた（ $t(24) = 2.63, d = 0.46, p < .05$ ）。性別、練習条件の主効果および全ての交互作用は有意ではなかった。

「リラックスの度あい」についても試行段階の主効果（ $F(2, 48) = 25.00, p \eta^2 = .51, p < .001$ ）が見られ、多重比較の結果、pre 試行と練習試行の間（ $t(24) = 4.66, d = 0.75, p < .001$ ）、練習試行と post 試行の間（ $t(24) = 2.44, d = 0.42, p < .05$ ）で「リラックスの度合い」がそれぞれ有意に上昇していた。性別、練習条件の主効果、および全ての交互作用は有意ではなかった。

「台本の話しやすさ」については試行段階の主効果（ $F(2, 48) = 7.79, p \eta^2 = .25, p < .01$ ）が見られ、多重比較の結果、pre 試行および post 試行と比較して練習試行における台本の話しやすさ得点が有意に低かった（pre-練習： $t(24) = 3.60, d = 0.93, p < .01$, post-練習： $t(24) = 2.43, d = 0.48, p < .05$ ）。性別、練習条件の主効果および全ての交互作用は有意ではなかった。

「ロボタンのかわいさ」については試行段階の主効果（ $F(2, 24) = 9.56, p \eta^2 = .29, p < .01$ ）が見られ、多

重比較の結果、練習試行が他の2試行と比較して有意にロボタンのかわいさ得点が低かった (pre-練習: $t(24) = 2.89, d = 0.59, p < .05$, post-練習: $t(24) = 3.61, d = 0.75, p < .01$). 性別、練習条件の主効果は有意ではなかった。条件×試行段階の交互作用は有意ではなかったものの一定の効果量が示され ($F(4, 48) = 2.47, p \eta^2 = .17, n.s.$), 練習試行における練習条件の単純主効果は有意であったことから ($F(2, 72) = 5.81, p \eta^2 = .33, p < .01$), 多重比較を行った結果、なし条件と地声条件の間 ($t(72) = 2.88, d = 2.31, p < .05$) および、なし条件とVT-4条件の間 ($t(72) = 3.01, d = 2.42, p < .05$) に有意差が見られた (Figure 3)。このため、段階の主効果についても、台本自体の差がもたらした相違ではなく、練習段階で新たに導入されたCGアバターに対する「かわいさ」評価がパペットよりも低下していたことを示す結果と思われる。

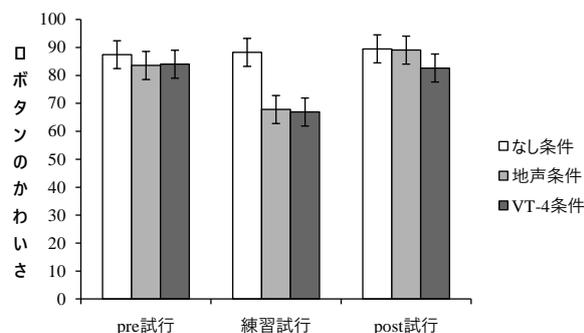


Figure 3 試行段階別のかわいさ評定平均

「ロボタンの操作の簡単さ」については、試行段階の主効果 ($F(2, 48) = 38.79, p \eta^2 = .62, p < .001$) が見られ、多重比較の結果、pre試行とpost試行の間 ($t(24) = 4.66, d = 0.75, p < .001$), 練習試行とpost試行の間 ($t(24) = 2.44, d = 0.42, p < .05$) で「操作の簡単さ」がそれぞれ有意に上昇していた。また、練習条件×試行段階の交互作用が有意であり ($F(4, 48) = 2.91, p \eta^2 = .20, p < .05$), 地声条件とVT-4条件における試行段階の単純主効果は、post試行が他の2試行と比較して有意に操作の簡単さ得点が高いというパターンであったが、なし条件では、pre試行とpost試行の間にのみ有意差が見られるというパターンであった。

性別、練習条件の主効果および性別×練習条件×試行段階の交互作用は有意ではなかった。

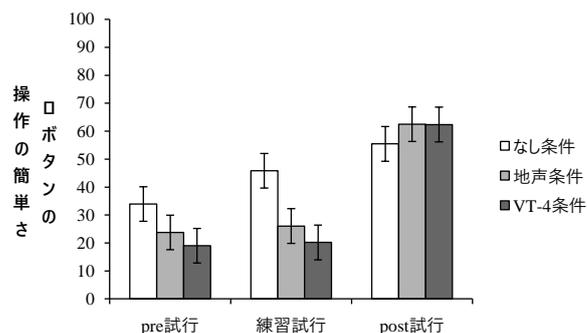


Figure 4 ロボタンの操作の簡単さ得点

VT-4による音声変換がアバターの印象に与えた影響を検討するため、アバターを利用した2条件のみを対象として、アバターに対する生物知覚を従属変数とした練習条件 (地声条件 / VT-4条件) ×性別 (男性 / 女性) 二要因分散分析を実施した。その結果、性別の主効果および練習条件の主効果は有意ではなかったが、両者の交互作用が有意であった ($F(1, 16) = 7.56, p \eta^2 = .32, p < .05$)。単純主効果検定の結果、女性における練習条件の単純主効果は有意ではなかったが ($F(1, 16) = 3.09, p \eta^2 = .28, n.s.$), 男性における練習条件の単純主効果は有意であり ($F(1, 16) = 4.53, p \eta^2 = .36, p < .05$), VT-4条件の男性は地声条件の男性と比較してアバターの生物らしさを低く知覚する傾向がみられた (Figure 5)。

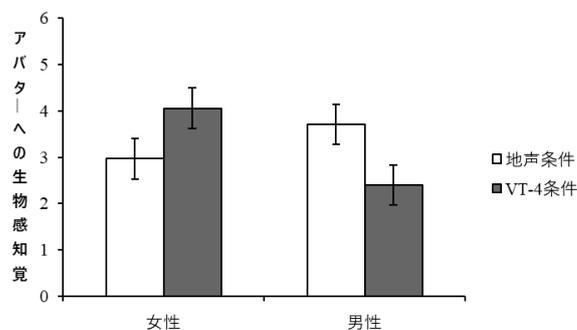


Figure 5 アバターの生物知覚への音声変換の影響

4. 考察

本研究では、アバターと音声変換技術を組み込んだ仮想システムを用いて講話練習を行うことで講話能力に与える影響について、講話者の主観評価をもとに検討を行った。

パフォーマンスの主観評価についての評定では、アバターの使用および音声変換の有無によって「講話の出来」には影響が見られなかった。また、練習試行をは

さんだ pre 試行と post 試行間でも有意な差がみられなかったことから、主観評価では練習によって講話のスキル獲得によるパフォーマンスの向上は見られなかったことが示唆された。一方で、「ロボタンの操作の簡単さ」については、pre 試行と post 試行間で有意に得点が向上しており、パペットの操作の仕方および、アバターの操作の仕方については、練習試行の間で主観的には十分に学習できたと評価されていることが示唆された。これらのことから、主観的な経験としては、練習試行の目的が講話の上達であるというよりは、アバターの操作の仕方を学習することが練習の目的であると認識されていた可能性がある。また、アバターの操作の習得については、post 試行における操作の簡単さ評定においてパペットと同様の評定値であることから、アバターの操作はパペットの操作と同程度の練習で習得可能であることが示され、アバターを用いた講話練習支援の可能性が示唆された。

アバターの使用および音声変換により、アバターに対する接近動機が高まるかどうかについては、講話の練習時間をもとに検討を行った。その結果、アバターの使用の有無および音声変換のいずれも練習時間に影響を与えてはならず、アバターの使用や音声変換によって接近動機が高まるとはいえない結果であった。また、かわいい音声の付与がアバターへの接近動機を高めると考えられたが、「ロボタンのかわいさ」評定において、アバターを使用した試行のみで「かわいさ得点」が有意に低く評定されており、パペットと比較してアバターが「かわいくない」と受け取られていることが示唆された。加えて、アバターを使用していた「地声条件」と「VT-4 条件」の間に有意差が見られなかったことから、ロボタンのかわいさ評定は音声変換の有無に影響を受けていないことが示唆された。これは原田ほか(2022)において、大学生が声の評価とアバター自体のかわいさの評価を独立させていることと一致している。

一方、アバターの生物らしさの知覚得点の結果では、男性において VT-4 による音声変換を加えるとアバターの生物らしさが有意に下がるという結果が得られた。これは、男性の方が女性と比較して「かわいい声」を調整する際に、地声からの調整幅が大きかったものと推測され、そうした調整幅の大きさが声のかわいさよりもまずは声に対する違和感をもたらし、その結果として、生物らしさの知覚が低減された可能性がある。こうした生物らしさの低下は、自分がどのような存在であるかといった自己の身体イメージを変容させるとされ

るプロテウス効果 ([6]小柳ほか,2020) にも負の影響を与える可能性が考えられる。したがって、音声変換によりアバターへの接近動機を高めるためには、より自然に感じられる音声への変換など、音声変換の手法について検討を行う必要がある。

本研究のパフォーマンスの評価は、講話者の主観的な評価に基づいており、練習前後でのパフォーマンスの主観的な向上も見られなかった。この点については、参加者がアバターの操作の仕方やパペットを用いた講話形態の習得などに練習目標を設定していたために、たとえ実践された講話のパフォーマンスが向上していたとしても、その向上が主観的な評価には現れにくかった可能性が考えられる。そのため、アバターおよび音声変換技術による講話パフォーマンスの支援効果については、主観的な評価のみではなく、第三者評価によるパフォーマンスの検討が必要であると考えられた(安久ほか, 2022)。

5. 謝辞

本研究は沖電気工業株式会社との共同研究の成果の一部である。また、本研究は筑波大学人間学群心理学類の目黒文乃氏の卒業研究「“かわいい”感性に対する音声の影響—アバター対話システムを用いた検討—」の一部をもとに再分析をしたものである。

文献

- [1] Nick Yee & Jeremy Bailenson (2007) The Proteus Effect: The Effect of Transformed Self-Representation on Behavior, *Human Communication Research*, 33, pp. 271-290
- [2] Nick Yee, Jeremy Bailenson, Nicolas Ducheneaut (2009) The Proteus Effect Implications of Transformed Digital Self-Representation on Online and Offline Behavior, *Communication Research*, Vol. 36, No. 2, pp. 285-312
- [3] 井原なみは, 入野野宏 (2011) 幼さの程度による“かわいい”のカテゴリ分類, *広島大学大学院総合科学研究科紀要*, 人間科学研究, 6, 13-17.
- [4] 原田・安久・鷹阪・目黒・葛岡 (2022) アバター評価と“かわいい”音声：音声の影響と加齢変化の検討 日本認知科学会第 39 回大会.
- [5] 長谷川莉子・原田悦子・栢野航・大澤博隆 (2015). エージェント利用システムに対する生物らしさ知覚の検討—年齢群間比較— 日本認知心理学会第 13 回大会発表論文集, 5.
- [6] 小柳陽光・鳴海拓志・Jean-Luc.Lugrin・安藤英由樹・大村廉 (2020). ドラゴンアバターを用いたプロテウス効果の生起による高所に対する恐怖の抑制 日本バーチャルリアリティ学会論文誌, 25 (1), 2-11.
- [7] 安久・鷹阪・目黒・葛岡・原田(2022) アバターと音声変換技術が講話能力の向上に与える影響(2)：聞き手によるパフォーマンス評価 日本認知科学会第 39 回大会.

Table 4 条件別のパフォーマンスの主観評価得点

質問項目	試行条件 性別/練習条件	pre試行			練習試行			post試行		
		なし	地声	VT-4	なし	地声	VT-4	なし	地声	VT-4
講話の出来に満足	男性	53.29	50.66	63.82	57.76	56.84	48.82	59.08	65.13	56.45
	女性	44.87	46.45	61.32	63.55	60.66	67.50	55.79	65.26	70.00
いまのあなたは 安心している	男性	50.92	68.82	70.00	55.92	70.79	75.92	63.29	79.21	75.92
	女性	58.95	64.74	61.18	54.21	73.42	64.74	67.11	86.97	62.76
いまのあなたは リラックスしている	男性	40.66	57.11	56.58	49.21	79.21	72.63	56.71	79.87	81.32
	女性	48.68	56.05	51.32	71.71	70.79	61.84	72.37	89.34	67.11
台本は話しやすい	男性	67.24	82.89	73.42	55.79	72.63	34.08	58.16	82.76	64.61
	女性	84.74	77.89	77.76	61.18	69.21	73.29	76.71	79.08	64.61
ロボタンはかわいい	男性	81.18	90.53	85.39	85.53	78.03	67.24	85.66	90.13	82.24
	女性	93.68	76.58	82.63	90.92	57.63	66.58	93.29	88.03	83.03
ロボタンを動かすのは 簡単だ	男性	46.32	29.74	15.26	52.24	29.61	17.89	58.55	65.66	62.76
	女性	21.58	17.76	22.76	39.47	22.50	22.50	52.37	59.34	61.97

ⁱ 本研究ではVT-4での変換設定についての記録が取られておらず、調整値についてはデータに基づく検討が不可能であった。