

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 The Impact of Avatar and voice Transformation Technology on Speaking Ability (2): Performance evaluation by listeners

安久 絵里子[†], 鷹阪 龍太[†], 目黒 文乃[†], 葛岡 英明[‡], 原田 悦子[†]
Eriko Ankyu, Takawaki Ryuta, Meguro Ayano, Kuzuoka Hideaki, and Etsuko T. Harada

[†]筑波大学, [‡]東京大学

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

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

概要

本研究では、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使用して講話練習を行うことで講話者の講話能力に変化が見られるか否かについて、第三者評価の結果より検討した。その結果、アバターと音声変換の同時利用により、「即時的な」講話能力を評価する練習段階と、「般化的な」講話能力を評価する post 段階での講話能力のいずれにおいても向上が見られた。この結果は講話者自身による主観評価との間にズレがあることを示しており、さらなる技術支援の可能性が示された。

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. 背景と目的

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。腹話術は聴衆の注意を惹き付ける特徴から公的機関による各種の講話で用いられることが多いが、話者の演じ分けと人形の操作を同時に行う話術として、認知的負荷の高い専門技術となっている。これに対し、近年の情報技術を用いて、アバターや音声変換技術を導入することにより、音声の切り替えや仮想の対話相手の操作を「容易にする」支援の可能性が考えられる。

話者の身体運動を自動的に変換してアバターの操作を行うことで、腹話術人形にくらべてより直感的に操作が可能となり、微細な動作や呼吸・瞬目などの基本動作などを自動的に組み込むことによる自然な生物感をもたらす可能性も考えられる。また、用いられるアバターの外見によってユーザーの行動や態度が変わるといふプロテウス効果(Yee & Bailenson, 2007; Yee et al., 2009)を踏まえると、仮想の対話相手の見た目や幼い子どもやキャラクターなど場面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、話者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。

また、リアルタイムでの音声変換技術を用いることで、話者自身が声色を変える話術をもたなくても「違う声」での発話が可能となり、話者はより講話内容や聞き手とのコミュニケーションに注力することが可能となると考えられる。さらに対象に対して抱く「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向上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、話者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、プロテウス効果による演じ分け能力の向上につながる可能性が考えられる。

そこで本研究では、アバターと音声変換技術を組み込んだ仮想的 2 者対話システムを想定し、アバターおよび音声変換技術の利用の有無を操作した条件下で講話練習を行った講話実験(鷹阪ほか, 2022)について、その講話動画の課題パフォーマンスに対して第三者評価を行った。仮想的 2 者対話システムの利用によって、システム利用時に即時的に得られる効果と、システムを利用しなくてもスキル/能力が習得されたという般化的な効果の両者を評価するため、パペットを用いて講話を行った pre 段階、システムを利用した練習段階、再度パペットを用いた post 段階のそれぞれに対して評価を行った。

2. 方法

実験計画 講話実験で実施された、講話者の性別(男性/女性) × 練習条件(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(pre/練習/post) の 3 要因混合計画に、評価者の性別(男性/女性)を加えた 4 要因混合計画であった。

参加者(評価者) 講話実験には参加していない 4 年

制大学の学生 37 名(男性 17, 女性 20, 平均年齢 20.03 歳, $SD=1.46$)が実験に参加した. 講話動画内の人物との面識がない参加者による評価を実施するため, 講話実験参加者を募集した大学とは異なる大学で参加募集を行った.

講話動画 鷹阪ら (2022) で収集された, 大学生 30 名を講話者とした, 試行段階 (pre/練習/post) 毎の講話動画を使用した. pre 段階と post 段階ではすべての講話者がパペットを用いて交通安全講話動画を撮影した. 練習段階では, 3 つの練習条件で講話練習を行っており, パペットのみ条件 (以下「なし条件」) では pre 段階と同様にパペットを使用しての練習を行い, アバター+地声条件 (以下「地声条件」) ではパペットの代わりにロボット型アバターを使用して, 講話者本人の地声での練習を求めた. アバター+VT-4 条件 (以下「VT-4 条件」) では, アバターの利用時に, VT-4 を用いて地声をリアルタイムで「かわいい音声」として変換をした. 練習段階では, 自由に練習を行った後に, 練習段階での講話動画を撮影した.

実験刺激 講話者 30 名が 3 試行段階毎に撮影した動画 90 本のうち, 技術上の理由で 9 本 (9 名による) を除

外し, 残りの計 81 本 (なし条件 26 本, 地声条件 28 本, VT-4 条件 27 本) を評価対象として使用した. 練習条件群ごとに講話者 4 名あるいは 3 名分の講話動画を 1 セットとして, 講話者 10 名ずつになるようセットの組合せを 6 パターン作成し, 評価実験参加者をランダムに割り当てた (表 1, 表 2). 各パターンには男女それぞれ 2 名以上が割り当てられ, 評価者 1 名あたりにつき, 講話者 10 名分の動画 (26-28 本) を評価した. 各評価者への動画の呈示順はパターン内でランダムであった.

手続き 動画の視聴およびパフォーマンス評価には Qualtrics を使用した. 各動画の長さは 1 分前後であり, 1 本視聴する毎に, 計 6 項目 (表 3) のパフォーマンスに関する評価項目について VAS での回答を求めた.

3. 結果

まず, パフォーマンスに関する評価項目について, 主成分分析を行うことにより評価項目の圧縮を試みたところ, 1 因子構造が妥当だと判断された (表 3). そこで全項目の平均値を算出し pre 試行と練習・post 試行のそれぞれの差分を従属変数として, 講話者の性別 (男性/女性) × 練習条件 (パペットのみ条件/アバター+地

表 1 講話動画の内訳

練習条件	セット名 動画本数 (除外本数), 講話者人数		
		A1	A2
なし条件	11 本(1), 4 名	8 本(1), 3 名	7 本(2), 3 名
	B1	B2	B3
地声条件	12 本(0), 4 名	8 本(1), 3 名	8 本(1), 3 名
	C1	C2	C3
VT-4 条件	11 本(1), 4 名	7 本(2), 3 名	9 本(0), 3 名

表 2 講話動画のパターン

	セット	男性	女性	計
1	A1, B2, C3	2	3	5
2	A1, B3, C2	4	2	6
3	A2, B1, C3	3	5	8
4	A2, B3, C1	3	3	6
5	A3, B1, C2	3	4	7
6	A3, B2, C1	2	3	5
	計	17	20	37

表 3 パフォーマンスに関する評価項目の主成分分析結果

項目	Factor1	共通性
Q4 2 人の登場人物の話分けは 下手だ—上手だ	.845	.714
Q2 今の講話は つまらなかった—面白かった	.808	.653
Q5 話しているときの視線や姿勢は 悪い—良い	.787	.619
Q3 講話を もう聞きたくない—もっと聞きたい	.774	.600
Q1 今の講話は わかりにくかった—わかりやすかった	.760	.577
Q6 腹話術の相手 (ロボタン) の動きは 悪い—良い	.746	.556
因子寄与	3.719	
α 係数	.876	

声条件/アバター+VT-4条件) × 評価者の性別 (男性/女性) × 試行段階 (練習/post) の4要因混合計画分散分析を実施した (図1)。その結果, 練習条件の主効果が有意であり ($F(2, 220) = 9.31, p < .001$, 偏 $\eta^2 = .08$), 多重比較の結果, 地声条件 (1.37) よりなし条件 (2.69) のほうがパフォーマンス評価が高くなったが, VT-4条件 (11.96) はいずれの条件とも有意な差は見られなかった。また, 試行段階の主効果が有意であり ($F(1, 220) = 6.80, p = 0.10$, 偏 $\eta^2 = .03$), 多重比較の結果, 練習段階 (7.20) は post 段階 (3.48) よりパフォーマンスの向上度合いが大きく評価された。講話者の性別と評価者の性別については有意な主効果が見られなかった (講話者の性別, $F(1, 220) = 0.77, p = .383$, 偏 $\eta^2 = .00$; 評価者の性別, $F(1, 220) = 0.93, p = .336$, 偏 $\eta^2 = .00$)。

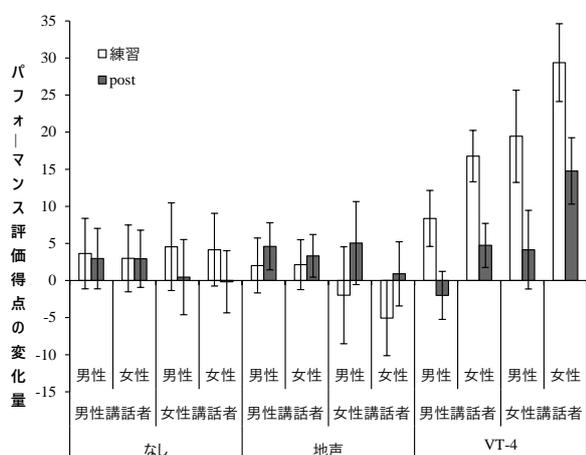


図1 パフォーマンス評価得点の変化量

また, 講話者性別と練習条件の交互作用 (図2) が有意であり ($F(2, 220) = 3.50, p = .032$, 偏 $\eta^2 = .03$), 講話者が男性のときは練習条件間で差はなかったが ($F(2, 220) = 1.19, p = .306$, 偏 $\eta^2 = .02$), 講話者が女性ときには, 地声条件 (-0.27) となし条件 (2.51) が VT-4 条件 (16.94) よりもパフォーマンス評価の変化量が小さかった (なし vs. VT-4, $t(220) = 3.36, d = 1.48, p = .003$; 地声 vs. VT-4, $t(220) = 3.80, d = 1.73, p = .001$)。さらに, 練習条件と試行段階の交互作用 (図3) が有意であり ($F(2, 220) = 12.83, p < .001$, 偏 $\eta^2 = .10$), なし条件と地声条件では試行段階による評価の差は見られなかったが (なし条件, $F(1, 220) = 0.81, p = .371$, 偏 $\eta^2 = .01$; 地声条件, $F(1, 220) = 2.95, p = .088$, 偏 $\eta^2 = .027$), VT-4 条件では post (5.41) より練習 (18.50) のほうがパフォーマンス評価の変化量が小さかった ($t(220) = 5.39, d = .64, p < .001$)。

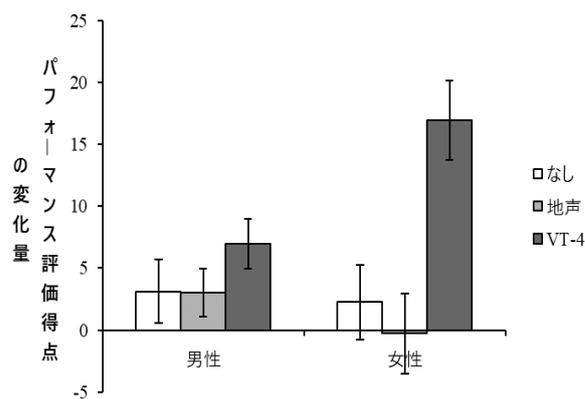


図2 パフォーマンス評価得点の変化量における講話者性別×練習条件の交互作用

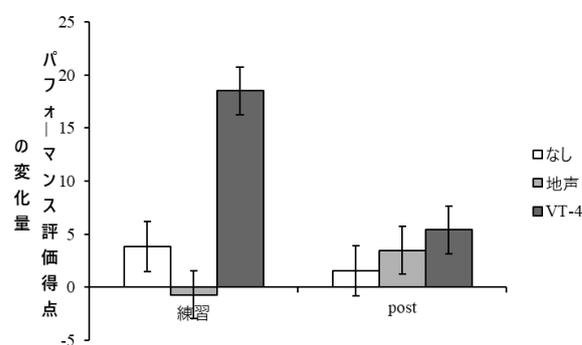


図3 パフォーマンス評価得点の変化量における練習条件×試行段階の交互作用

4. 考察

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, 腹話術形式での講話能力に与える影響について, 講話動画の第三者評価結果より検討した。その結果, システムを使わずパペットで練習したなし条件と, アバターを使用した音声変換を行わなかった地声条件では, 試行段階におけるパフォーマンス評価の変化量に差は見られなかった。一方で, アバターと音声変換を適用した VT-4 条件では pre 段階に比べ, 練習段階と post 段階のいずれにおいてもパフォーマンス評価が向上した。特に, アバターと音声変換技術を併用することによる即時的な効果を示す練習段階での講話評価の向上が強く見られたが, 一般化的な効果を示す post 段階, すなわち練習段階後のパペットを利用した講話でのパフォーマンス評価も向上している点は興味深い。地声条件では試行段階の効果は見られなかったことから, アバター利用だけではなく, 同時に音声変換を加えることが重要な要素であること

が示された。これより、アバターと音声変換技術を併用することが、講話そのものを支援するシステムとしてだけでなく、腹話術形式の講話において、練習のための学習ツールとしての利用可能性があることも示唆された。

一方で、講話練習を行った講話者の主観評価の結果（鷹阪ら, 2022）では、こうした練習段階での支援ツール利用の条件間差は見られなかった。また、講話者自身の判断による練習時間にも条件間の差は見られなかったことから、音声の付与による接近動機の向上は示されなかった。

さらに、第三者評価ではVT4条件でのパフォーマンス評価の向上が見られたが、主観評価では練習条件による差が見られなかったというズレが得られたことから、講話者にとっては自分自身の講話能力、あるいは講話のパフォーマンスを客観的に判断することが困難であることを示唆している。これは腹話術形式の2者対話という特殊な対話における特性に起因する現象であるか否かについては、さらに検討を行っていく必要がある。腹話術形式では話者の演技分けと人形の操作を同時に行う認知的負荷の高い状態で自身を客観的に評価する余裕がなかったことが一因である可能性が考えられるが、より一般的に人は自らの講話やスピーチのパフォーマンスを正しく評価することが難しい可能性も考えられる。いずれの場合においても、講話者が「自分の講話がより良いものになっている」ことを意識できるならば、講話を良い方向に向けてより一層練習していくなどの効果が見込まれること、また一般に学習は学習成果に関するメタ認知に依存することを考慮す

ると、講話のパフォーマンスを高めるための支援システムとして、「自分の行う講話について、客観的な評価ができるような環境をもたらす」システム設計の有効性が示唆されると考えられる。今後さらなる検討が必要であるといえよう。

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。アバターシステムの作成をしてくださった沖電気工業・鈴木雄介さん、畠中貴明さんに謝意を表したい。なお、本研究は筑波大学人間学群心理学類4年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] 鷹阪龍太, 安久絵里子, 目黒文乃, 葛岡英明, 原田悦子 (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 (1): 話し手自身による主観評価による検討, *日本認知科学会第39回大会*