

音声と外見の適合性がリモート接客での質問数と会話時間に与える影響

渡邊 恵佑^{1*}, 水野 翔太¹, 山本 吉伸¹

¹ 福知山公立大学大学院地域情報学研究科

* Corresponding Author: Tel: 090-1144-6468, E-mail: 32456015@fukuchiyama.ac.jp

Abstract

This study examines the impact of voice and appearance consistency in digital avatars on customer interactions in remote service. An experiment analyzed how matching or mismatching voice and appearance influences the number of spontaneous questions and conversation duration, while psychological effects were assessed through a subjective evaluation questionnaire. The results showed that the consistent condition (male avatar + male voice) led to a slight increase in questions and conversation time. However, no statistically significant difference was observed, indicating that consistency did not have a decisive impact on conversation activeness or continuity. Even when participants perceived incongruity, it had little effect on ease of questioning, psychological stress, or conversation smoothness. This study focused on short-term interactions, suggesting that voice-appearance consistency may not significantly affect customer interactions in such cases. However, further research is needed to explore long-term effects and other potential factors in designing remote customer service systems.

Keywords

remote customer service system, voice and appearance consistency, customer behavior analysis

1 はじめに

近年、リモート接客システムの導入が進んでいる。小売業や飲食業では、人手不足の解消や業務の効率化の観点から、対面接客に代わる手段としてデジタルアバターを用いた接客が試みられている。

そして、音声と外見の適合性や印象が人間の認識や感情に与える影響については多くの研究が存在する一方で、接客場面における影響を直接検討した研究は限られている。そこで本研究では、デジタルアバターの外見（性別）と音声（全て男性の声）の適合性が、リモート接客における顧客の行動や心理にどのような影響を与えるのかを調査する。具体的には、アバターの外見と音声の一致・不一致が、自発的な質問数や会話時間といった行動指標および顧客の主観評価にどのように影響を及ぼすのかを分析する。

本研究の成果は、音声や外見の適合性が接客体験に与える影響を明らかにし、快適なリモート接客システムの設計指針を提供することで、リモート接客の実用性向上に寄与することが期待される。

本論文では、まず第2章で先行研究を概観し、音声と外見の適合性に関する既存の知見を整理する。続いて、第3章では本研究の実験デザインと手法について述べる。第4章では実験結果を報告し、最後に第5章で考察、第6章では、今後の研究の展望について記述する。

2 関連研究

デジタルアバターやバーチャルキャラクターの音声や外見が、ユーザーの知覚や行動に及ぼす影響について多くの研究が行われている。特に、音声と外見の適合性、合成音声の評価、バーチャルインタラクションにおける信頼、視覚と聴覚の統合、AIアバターの影響など、さまざまな観点から研究が進められている。

2.1 音声と外見の適合性が知覚や行動に与える影響

音声と外見の組み合わせがキャラクターの印象に与える影響については、数多くの研究で調査されている。(Takahashi and Maeshiro 2022) は、人物やキャラクターの外見が変わると、それに適合する声質（基本周波数、声道長）も変化することを明らかにし、適切なキャスティングや声質設計の必要性を示している。(Choi et al. 2023) では、バーチャルキャラクターの外見と声が mismatch する場合の影響を調査し、人間の声を持つキャラクターは、外見がロボットであってもより高い好感度と安全性の評価を受けることを明らかにした。

また、(Ferstl et al. 2021) は、会話型ソーシャルエージェントにおいて、発話とジェスチャーのリアリズムが高いほど好意的に評価されるが、外見のリアリズムは必ずしも高い方が望ましいわけではないことを示唆した。これらの研究は、音声と外見の適合性が知覚や行動に与える影響の複雑さを示している。

2.2 合成音声の評価と影響

バーチャルキャラクターの音声に関しては、人間の声と合成音声の違いも重要な研究テーマとなっている。(Cabral et al. 2017)は、合成音声よりも人間の声の方が表現力があり、理解しやすく、好ましいと評価される一方、キャラクターの信頼性や魅力には声の種類が決定的な影響を与えないことを示した。また、(Iwamoto and Nishimoto 2023)では、商品販促を目的とする自己推薦ロボットの感情表現において、視覚情報・聴覚情報・言語情報の3要素を比較し、特に声色のデザインが感情認識に大きな影響を与えることを示した。これらの研究は、合成音声や声のデザインがユーザーの印象形成にどのように関与するかを明らかにしている。

2.3 バーチャルインタラクションにおける信頼とモチベーション

バーチャル空間におけるアバターの役割や影響についても研究が進んでいる。(Guo et al. 2023)は、ホスピタリティ業界におけるアバターの活用を分析し、アバターへの認知的信頼が記憶に残るブランド体験を形成し、感情的信頼がブランド愛や忠誠心を高めることを示した。また、(Visser et al. 2024)では、AIアバターの外見とAIであることの開示がユーザーのモチベーションに与える影響を調査し、AIの開示がプレイの強度を増し、男性的な外見のAIアバターはプレイの強度を低下させる傾向があることを示した。これらの研究は、バーチャルインタラクションにおけるアバターのデザインが、ユーザーの行動や信頼形成に影響を与える可能性を示している。

2.4 視覚と聴覚の感情統合の影響

音声と外見の適合性は、視覚と聴覚の感情統合とも密接に関連している。(Izen et al. 2019)は、怒った表情と声を受けた後の知覚変化を調査し、感情の一致・不一致が知覚のバイアスに影響を与えることを示した。(Kokinous et al. 2015)では、動的な顔と声の統合において、感情の一致が早期の神経応答に影響を与えることを検証し、(Takagi et al. 2015)では、六つの基本感情における多感覚知覚が注意の指示によって変化することを示した。これらの研究は、視覚と聴覚の統合がどのように行われ、感情知覚や行動に影響を与えるかを理解する上で重要な知見を提供している。

2.5 本研究の位置づけ

これらの研究を踏まえ、本研究では、リモート接客におけるデジタルアバターの音声と外見の適合性が、顧客の行動や心理的受容に与える影響を調査する。特に、音声と外見の一致・不一致が、自発的な質問数や会話時間といった行動指標および主観評価にどのように影響するのかを分析し、リモート接客システムの設

計において考慮すべき要素を明らかにすることを目的とする。

3 実験概要

3.1 実験目的

本研究では、デジタルアバターを用いたリモート接客において、アバターの外見(性別)と音声(すべて男性の声)の適合性が、顧客の行動指標(自発的な質問数、会話時間)および主観評価(音声と外見の適合性、外見と音声を与える心理的印象)にどのような影響を与えるのかを調査することを目的とする。音声と外見の適合性が、会話のしづらさや不快感を引き起こし、結果として質問数や会話時間が減少する可能性がある一方で、不適合が逆に関心を引き、質問数や会話時間を増加させる可能性も考えられる。

したがって、本研究では数値データだけではなく、被験者の主観的評価も分析し、アバターの音声と外見の適合性が対話に与える影響を定量的、定性的に明らかにする。

3.2 実験デザイン

本研究では、音声条件をすべて「男性の声」に固定し、アバターの外見を「男性」または「女性」と設定することで、音声と外見の適合性が顧客の行動や主観評価に及ぼす影響を測定する。具体的には、音声と外見が一致する条件(男性アバター+男性の声)と、不一致の条件(女性アバター+男性の声)の2つの条件を比較する。アバターの外見は性別が明確に認識できるようにし、その他の特徴はシンプルにデザインした(図1, 図2参照)。

実験は、リモート接客を再現するためにZoomを使用し、Zoomのアバター機能を利用したアバターを通じて、テイクアウト専門のカフェに立ち寄り商品を選ぶという状況を再現するように設計された。被験者はカフェの顧客として、Zoom上に映るアバターと対話しながら、持ち帰り用のカフェ商品を選択するシナリオに従って行動した。実験環境の様子を図1に示す。実験室内には、Zoom上にアバターを表示し、その手前にカフェのメニューとマイクを配置した。被験者はこの環境内でリモート店員との対話を行いながら、商品を選ぶプロセスを体験した。

被験者は同じゼミに所属する学生9名とし、各条件に均等に割り当てた。接客者(店員)は、被験者と面識のない他ゼミの学生が担当し、実験の公平性を確保した。

また、本研究では、被験者が接客の様子をより注意深く観察できるよう、特定の目的を与えた。具体的には、被験者には「カフェチェーン本部のマネージャーであり、各店舗の接客を採点する担当者である」と説明し、実験の目的が単なる接客体験ではなく、店員の対応を評価するものであると意識させた。ただし、被験者には「店員には本部の採点担当者であることを悟られないよう、普段の顧客として自然に接客を受ける

ように」と指示した。このようにすることで、被験者が店員の言動や対応の違和感をより意識しやすくなり、音声と外見の適合性に対する影響を正確に捉えることができると考えた。

実験中の会話はカメラで録音・録画し、後に質問数や会話時間をカウントした。また、接客体験終了後に被験者には主観評価アンケートに回答してもらい、接客体験に対する印象や違和感についてのデータを収集した。



図1 実験環境の様子



図2 女性アバターデザイン

3.3 測定項目

本実験では、アバターの音声と外見の適合性が対話体験に与える影響を測定するため、被験者にアンケートを実施した。まず、「店員の声と外見の組み合わせに違和感を感じましたか?」という質問を行い、「感じる」と回答した被験者と「感じない」と回答した被験者で、それぞれ異なる設問を提示した。

違和感を感じた被験者には、対話のしやすさや質問のしやすさ、心理的ストレスの有無について尋ね、音声と外見の不適合が対話に及ぼす影響を測定した。一方、違和感を感じなかった被験者には、安心感や信頼感、不快感の有無について質問し、音声と外見の適合性がポジティブな影響を与えたかを評価した。

これらの質問の回答は、5段階評価（「全く感じなかった」～「非常に感じた」）で尋ね、違和感や心理

的影響の度合いを測定した。また、数値データでは把握しにくい被験者の主観的な印象や具体的な気づきを収集するため、自由記述の設問も設けた。自由記述では、店員の声と外見の組み合わせに対する印象や、それが対話全体にどのような影響を与えたかについて記述してもらい、対話体験全体の感想や気づいた点を自由に回答できる形式とした。

3.4 実験手順

本実験では、被験者に対してリモート接客システムを体験してもらい、アバターの音声と外見の適合性が顧客の行動や主観評価に及ぼす影響を調査した。実験開始前に、被験者には実験の目的と流れについて説明を行い、「新しい接客システムを体験する」という趣旨を伝えた。被験者は、テイクアウト専門のカフェの顧客として、アバター店員の接客を受ける役割を担い、商品選択や購入判断を行った。

まず、ディスプレイ上に表示されたアバターが店員として、被験者に対して商品案内を行う。その後、被験者は店員との対話を通じて商品を選択し、購入の判断を行う。実際の商品は提供されないが、被験者には自由に購入の可否を決定できるようにし、普段の接客体験に近い状況を再現した。購入判断が完了した時点で実験は終了し、被験者は直後に主観評価アンケートの回答を行った。

実験では、統一された接客シナリオを用いたが、被験者の反応に応じて店員は柔軟に対応できる形式とした。店員は、被験者に挨拶をした後ディスプレイの前に設置されたメニュー表があることを伝え、そこから自由に商品を選ぶよう促した。被験者が質問した場合には、店員は適切に回答し、注文を受け付けた後に接客を終了した。

被験者には、実験中の会話が録音・録画されること、普段通りの自然な会話を心がけること、実験終了後に他の被験者に内容を伝えないことを事前に説明した。

4 結果

4.1 質問数,会話時間

本実験では、被験者ごとの質問数および会話時間を測定し、それらの関係性を分析した。図3はアバターの外見による質問数の違いを示し、図4は会話時間を示している。

図3の結果から、質問数には個人差があり、男性アバター (P1~P5) と女性アバター (P6~P9) で多少の違いが見られた。質問数の平均を比較すると、男性アバターが平均1.8回、女性アバターが平均1.0回の質問があり、男性アバターの方がやや多い傾向が見られた。しかし、データの分布が正規性を満たさない可能性を考慮し、統計的な検定として独立サンプルt検定 ($t = 0.97, p = 0.363$) およびMann-Whitney U検定 ($U = 14.0, p = 0.373$) を実施した。その結果、統計的に有意な差は認められなかった。

同様に、図4の結果では、男性アバターの平均会話時間は約1.7分、女性アバターは約1.0分であり、男性アバターの方がやや多い傾向が見られたが、独立サンプルt検定 ($t = 1.40, p = 0.213$) およびMann-Whitney U検定 ($U = 13.0, p = 0.556$) の結果、会話時間の差も統計的に有意ではなかった。そのため、男性アバターの方がやや会話時間が長い傾向は見られるものの、統計的に明確な差があるとは言えない。

これらの結果から、男性アバターの方が質問数および会話時間においてやや多い傾向はあるものの、統計的に有意な差は認められず、アバターの外見と音声の適合性が会話の長さに大きな影響を与えたとは結論づけられないことが示された。

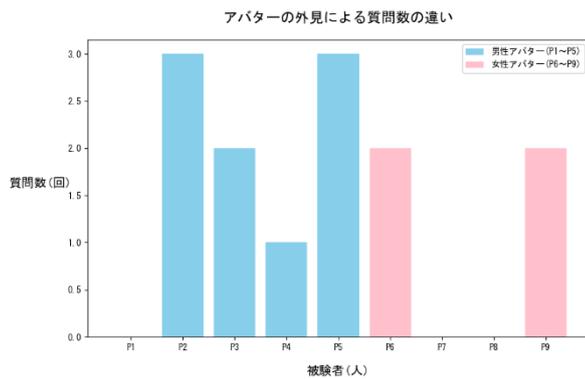


図3 質問数

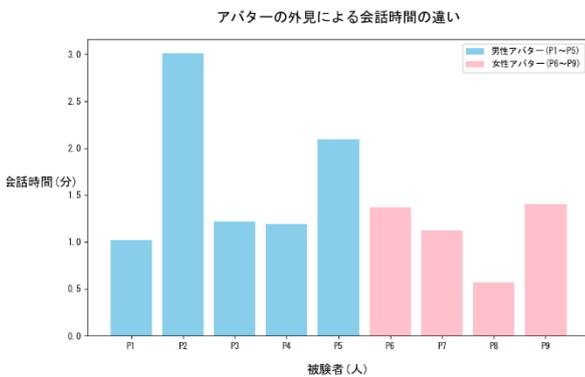


図4 会話時間

4.2 アンケート結果

本実験では、アバターの音声と外見の適合性が対話体験に与える影響を測定するため、被験者にアンケートを実施した。その結果、音声と外見が一致している条件（男性アバター）では、すべての被験者が違和感を感じなかったと回答し、音声と外見が不一致の条件（女性アバター）では、すべての被験者が違和感を感じたと回答した。

そして、違和感を感じた被験者においては、その違和感が対話行動に及ぼす影響は限定的であった。たとえば、「質問へのためらい」に関しては、違和感を感じた被験者・感じなかった被験者のどちらも低い評価となり、音声と外見の適合性が質問のしやすさに大き

な影響を与えたとは言えなかった。また、「対話のしづらさ」や「心理的ストレス」に関する評価も、違和感を感じた被験者であっても特に高くなり、音声と外見の違和感が対話の快適さに直接的な悪影響を及ぼしたとは言えない結果となった。

さらに、自由記述においても「最初は違和感を感じたが、対話には影響しなかった」「一度聞いてからは特に気にすることはなかった」といった意見が見られた。このことから、音声と外見の適合性が最初の印象に影響を与える可能性はあるものの、対話のしやすさや心理的ストレスに持続的な影響を与えるわけではないことが示唆される。

以上の結果から、音声と外見の適合性が対話の第一印象に影響を与える可能性は高いが、その影響が持続的なものであるか、また対話の継続性や心理的負担にどの程度関与するのかについては、さらなる検討が必要である。

5 考察

本実験の結果、アバターの音声と外見の適合性は接客場面における質問数および会話時間に有意な影響を与えなかったことが示された。

行動指標である質問数および会話時間については、「男性アバター+男の声」（一致条件）の方がやや多い傾向が見られたものの、その差は統計的に有意ではなく、音声と外見の適合性が接客中の対話の活発さや継続性に決定的な影響を与えるとは言えない。

また、アンケート結果からも、音声と外見の適合性が接客時の対話のしやすさに大きな影響を与えたとは言えなかった。音声と外見のミスマッチによる違和感を感じた被験者はいたものの、「質問へのためらい」「対話のしづらさ」「心理的ストレス」といった評価には影響がなかった。このことから、音声と外見の不一致による違和感が生じたとしても、それが接客のスムーズさや顧客の行動に直接的な悪影響を与えるわけではない可能性が示唆される。

さらに、自由記述の中には「最初は違和感を感じたが、対話には影響しなかった」「一度聞いてからは特に気にすることはなかった」といった意見も見られた。これにより、音声と外見のミスマッチによる違和感は、接客の初期段階では意識されるものの、実際の対話が進行するにつれてその影響が軽減される可能性がある。特に、接客の主な目的が「商品選択や購入」である場合、アバターの外見や音声の違和感が多少あったとしても、それが対話の流れを妨げる要因にはなりにくいと考えられる。

以上の結果を踏まえると、音声と外見の適合性が顧客の行動や態度に及ぼす影響は、場面や目的によって変化する可能性がある。たとえば、アバターが顧客と長時間継続的に対話する場面では、音声と外見の適合性がより重要な要素となることが考えられる。しかし、今回のように比較的短時間の接客が中心となる場面では、その影響は限定的であり、適合性の違和感が

対話の質を著しく低下させる要因にはならない可能性が高い。

本実験では、サンプル数が少なく、被験者の属性も限定的であったため、これらの解釈は仮説的なものである。今後の研究ではより多様な条件下での検証が必要である。加えて、異なる業種や接客スタイルにおけるアバターの音声・外見の影響を詳細に分析することで、より実用的なデザインガイドラインの策定につながる可能性がある。

6 結論

本研究では、リモート接客におけるアバターの音声と外見の適合性が、顧客の対話行動や心理的受容に与える影響を分析した。行動指標である質問数および会話時間の結果から、「男性アバター+男の声」（一致条件）の方がわずかに多い傾向が見られたものの、統計的に有意な差は認められなかった。そのため、音声と外見の一致・不一致が接客中の対話の活発さや継続性に決定的な影響を与えたとは言いきれない。

また、アンケートの結果からも、音声と外見の適合性が接客時の対話のしやすさに大きな影響を与えたとは言えず、違和感を感じた場合であっても「質問へのためらい」「対話のしづらさ」「心理的ストレス」といった評価にはほとんど影響が見られなかった。

音声と外見の違和感は、顧客の行動や心理的負担にはほとんど影響を及ぼさないことが示された。

7 今後の課題

本研究では、シミュレーション環境での接客実験を実施したが、実際の商品販売を伴う接客状況においても同様の結果が得られるかは検証されていない。そのため、今後の課題として、実際の販売環境においてアバターによるリモート接客を行い、顧客の購買行動や満足度への影響を調査することが求められる。また、この実験結果を基に、リモート接客システムの実用化に向けた設計を進め、より実際のサービス環境に適したアバターの音声・外見のデザインや、接客シナリオの構築を行うことが重要である。

8 謝辞

本研究の遂行にあたり、福知山公立大学大学院地域情報学研究科の山本教授には、研究の構想から論文執筆に至るまで貴重なご指導と温かいご助言を賜りました。研究の方向性を定める上での確かなアドバイスをいただき、多くの示唆を与えてくださいました。深く感謝申し上げます。

また、本研究の実験にご協力いただいた山本ゼミの皆様には、実験の参加およびアンケートへの回答を通じて、貴重なデータ収集にご協力いただきました。さらに、店員役を担当してくださった外部の生徒の皆様にも、多大なるご協力を賜りました。深く御礼申し上げます。皆様のご協力により、本研究の成果を得ることができました。心より御礼申し上げます。

最後に、研究活動を進める上で助言をくださった研究室の皆様にも感謝いたします。皆様の支えがなければ、本研究を無事に完了することはできませんでした。ここに深く感謝の意を表します。

9 参考文献

- Iwamoto, T., & Nishimoto, K. (2023). 商品販促を目的とする自己推薦ロボットの感情表現を構成する各要素情報が感情認識に与える影響. *日本感性工学会論文誌*, 22(2), 85-94.
- Takahashi, H., & Maeshiro, T. (2022). 人物・キャラクターの外見から受ける性格特性の印象と声の関係. 第 49 回知能システムシンポジウム, オンライン開催, 2022年3月14日-15日.
- Cabral, J. P., Cowan, B. R., Zibrek, K., & McDonnell, R. (2017). The influence of synthetic voice on the evaluation of a virtual character. *Proceedings of INTERSPEECH 2017*, 229-233.
- Choi, M., Koiliias, A., Volonte, M., Kao, D., & Mousas, C. (2023). Exploring the appearance and voice mismatch of virtual characters. *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, 555-560.
- Ferstl, Y., Thomas, S., Guiard, C., Ennis, C., & McDonnell, R. (2021). Human or robot? Investigating voice, appearance and gesture motion realism of conversational social agents. In *IVA '21: ACM International Conference on Intelligent Virtual Agents, Virtual Event, Japan, September 14-17, 2021*. Association for Computing Machinery.
- Guo, Y.-M., Ng, W.-L., Hao, F., Zhang, C., Liu, S.-X., & Aman, A. M. (2023). Trust in virtual interaction: The role of avatars in sustainable customer relationships. *Sustainability*, 15(18), 14026.
- Izen, S. C., Lapp, H. E., Harris, D. A., Hunter, R. G., & Ciaramitaro, V. M. (2019). Seeing a face in a crowd of emotional voices: Changes in perception and cortisol in response to emotional information across the senses. *Brain Sciences*, 9(8), 176.
- Kokinous, J., Kotz, S. A., Tavano, A., & Schröger, E. (2015). The role of emotion in dynamic audiovisual integration of faces and voices. *Social Cognition and Affective Neuroscience*, 10(5), 713-720.
- Takagi, S., Hiramatsu, S., Tabei, K., & Tanaka, A. (2015). Multisensory perception of the six basic emotions is modulated by attentional instruction and unattended modality. *Frontiers in Integrative Neuroscience*, 9(1).
- Visser, B., van der Putten, P., & Zohrehvand, A. (2024). The impacts of AI avatar appearance and disclosure on user motivation. In *Data Science and Artificial Intelligence* (pp. 142-155).