

音声と外見が AI 受付アシスタントに与える印象

Effects of Vocal and Visual Appearance on Impressions of an AI Receptionist

垣見向夏花¹ 早瀬光浩^{1*}
Himari Kakimi¹ Mitsuhiro Hayase¹

¹ 梶山女学園大学

¹ Sugiyama Jogakuen University

Abstract: 本研究は、音声特性と外見の組み合わせが AI 受付アシスタントの印象に与える影響を調査する。本物の人のような見た目の「リアルキャラクタ」と、カートゥーンのような「アニメキャラクタ」を用い、音声の発話速度や音高を変化させた動画で実験を実施する。結果、外見の影響が声質を上回ることが示唆され、リアルキャラクタが高評価を得た。今後は男性被験者や異なる業種での評価を予定している。

1 はじめに

AI 技術の進展に伴い、企業や施設での受付業務を自動化する AI 受付アシスタントが注目されている。音声合成技術やキャラクタ生成技術の進化により、人間に近い対話体験の提供が可能となってきたが、音声特性や外見が与える印象が利用者の満足度や信頼性にどのように影響を与えるかについての具体的な研究は限られている。

本研究の目的は、音声特性（発話速度・音高）と外見（リアルキャラクタ・アニメキャラクタ）の組み合わせが AI 受付アシスタントの印象評価に与える影響を明らかにすることである。

2 関連研究

音声特性が印象評価に与える影響は、ロボットや AI を対象とした研究において注目されている。岩川らは、発話速度と声域が性格印象に与える影響を調査し、速い発話速度が勤勉性を高め、低い声域が協調性を弱める傾向を報告している [1]。

キャラクタの外見が利用者の印象に与える影響は多くの研究で示唆されている。しかし、音声特性と外見の相互作用についての研究は十分に行われていない。



(a) リアルキャラクタ



(b) アニメキャラクタ

図 1: ImageFX で作成した AI 受付アシスタントの画像

3 実験デザイン

本研究では、音声と外見の組み合わせによる印象評価を調査する。

音声合成は VOICEVOX を用いて、発話速度 (0.9, 1.0, 1.3) と音高 (-0.07, 0.00, 0.07) の異なる 9 種類の音声を生成する。この設定は自然な会話範囲内での違いを評価可能とするためである。

画像生成は ImageFX[2] を使用し、リアルキャラクタとアニメキャラクタを作成する。それぞれのプロンプトはリアルキャラクタが「Japanese female, receptionist20s, seatedfacingfront, dakhair」、アニメキャラクタが「Character Girl, Front」である。これらの外見は、受け手に異なる印象を与えることを目的としている。図 1 は作成した AI 受付アシスタントの画像を示している。図 1(a) がリアルキャラクタ、図 1(b) がアニメキャラクタである。

動画生成は SadTalker[3] を用いて、音声と画像を組み合わせた 4 秒程度の動画を 18 種類作成する。SadTalker

*連絡先: 梶山女学園大学
〒464-8662 名古屋市千種区星が丘本町 17 番 3 号
E-mail: mhayase@sugiyama-u.ac.jp

- Q1. 明るくて、親しみやすいと思う
- Q2. 賢くて、不快感があると思う
- Q3. 控えめでおとなしいと思う
- Q4. 冷静で気分が安定していると思う
- Q5. 抑揚がなく機械的な声だと思う
- Q6. 冷たくて、感情がこもっていないと思う
- Q7. 早口で聞き取りにくいと思う
- Q8. はっきり、聞き取りやすいと思う
- Q9. 上品で大人っぽいと思う
- Q10. 会社受付にいても違和感がないと思う

図 2: 印象評価のアンケート項目

とは、音声と静止画像を入力として、3D モーション係数（表情、頭部の動き）を活用し、自然な表情や頭部の動きを生成するシステムである。この手法により、リアルで自然な話す顔動画を作成できる。

被験者は女性 15 名（21.3 ± 0.7 歳）である。評価には Ten Item Personality Inventory (TIPI-J) を参考に、「親しみやすさ」「堅さ」「落ち着き」など 10 項目であり、7 段階評価で回答を求める。図 2 に、具体的な質問項目を示す。また、動画はリアルキャラクター、アニメキャラクターごとにランダムに提示する。

4 実験結果と考察

4.1 実験結果

音声特性と外見の組み合わせが印象評価に与える影響を分析した。図 3 および図 4 は、各質問項目の平均分布を示している。実験結果から、リアルキャラクターとアニメキャラクターに対する印象評価には明確な違いが認められた。

まず、リアルキャラクター（図 3）は「親しみやすさ」（図 3(a)）や「上品で大人っぽい」（図 3(i)）といった項目で高評価を得ており、特にフォーマルな場面での信頼性や堅実性が強調されている。一方で、「抑揚がなく機械的な声」（図 3(e)）や「冷たくて感情がこもっていない」（図 3(f)）の評価は低く、これらの要素に対する改善が課題として挙げられる。

一方、アニメキャラクター（図 4）は「親しみやすさ」（図 4(a)）や「はっきり聞き取りやすい」（図 4(h)）の項目で高い評価を得ており、特にカジュアルな親近感や

柔らかさが利用者に好まれる傾向が見られた。しかし、「冷静で気分が安定している」（図 4(d)）や「会社受付にいても違和感がない」（図 4(j)）の評価はリアルキャラクターと比べて低く、業務適応性や信頼感においてリアルキャラクターに比べて低い評価が得られた。

これらの結果から、外見の違いが印象評価に与える影響は大きく、特に業務環境での利用においてはリアルキャラクターの方が適していると考えられる。また、音声特性の影響は外見に比べて相対的に小さく、外見が印象形成の重要な要因であることが確認された。

4.2 統計的仮説検定

次に、フリードマン検定で有意差が認められたため、中屋の変法を用いて事後検定を行い、音声特性と外見の組み合わせによる具体的な違いを調査した。図 5 と図 6 に結果を示す。

リアルキャラクターに対する印象評価について、「明るくて、親しみやすいと思う」（図 5(a)）では、発話速度が遅く音高が高い組み合わせ（速さ 0.9, 音高+0.07）が最も高い評価を得た。一方で、速い発話速度（速さ 1.3）の場合には評価が低下する傾向が確認された。「冷静で気分が安定していると思う」（図 5(d)）では、発話速度が遅い組み合わせ（速さ 0.9）が特に高評価を受け、速い発話速度（速さ 1.3）では安定感が低いと評価された。「上品で大人っぽいと思う」（図 5(i)）では、発話速度が遅く音高が低い場合（速さ 0.9, 音高 0.00）が最も高い評価を得たのに対し、速い発話速度や高音の組み合わせでは評価が低い傾向が見られた。また、「会社受付にいても違和感がないと思う」（図 5(j)）でも、発話速度が遅い（速さ 0.9）組み合わせが最適と評価された一方で、速い発話速度では適合性が低いと判断された。

一方、「抑揚がなく機械的な声だと思う」（図 5(e)）や「冷たくて感情がこもっていないと思う」（図 5(f)）といった項目では、音声特性の影響が比較的小さく、リアルキャラクターの外見による信頼感が音声特性を上回る結果となった。特に、速さ 1.3, 音高-0.07 の組み合わせでは若干の評価低下が見られたが、全体的には外見の影響が顕著であった。

これらの結果から、リアルキャラクターは「親しみやすさ」や「上品さ」を強調することで高評価を得やすい一方で、発話速度や音高の設定が印象に大きく影響することが確認された。特に、「冷静さ」や「業務適応性」においては、発話速度が遅く安定した声の組み合わせが最適であることが示唆された。

アニメキャラクターに対する印象評価の結果について、「明るくて、親しみやすいと思う」（図 6(a)）では、発話速度が遅く音高が高い組み合わせ（速さ 0.9, 音高+0.07）が最も高い評価を得た一方、速い発話速度（速さ 1.3）

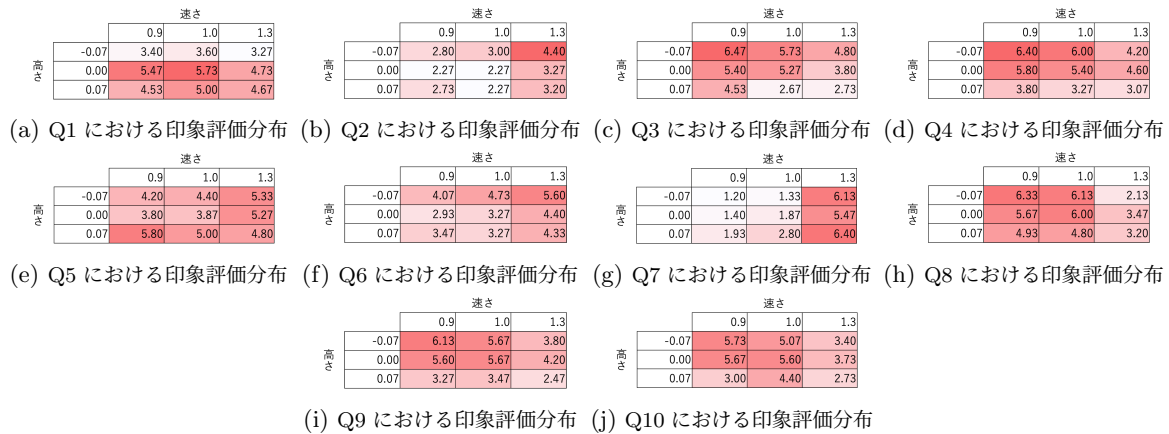


図 3: リアルキャラクターに対する質問項目ごとの平均分布

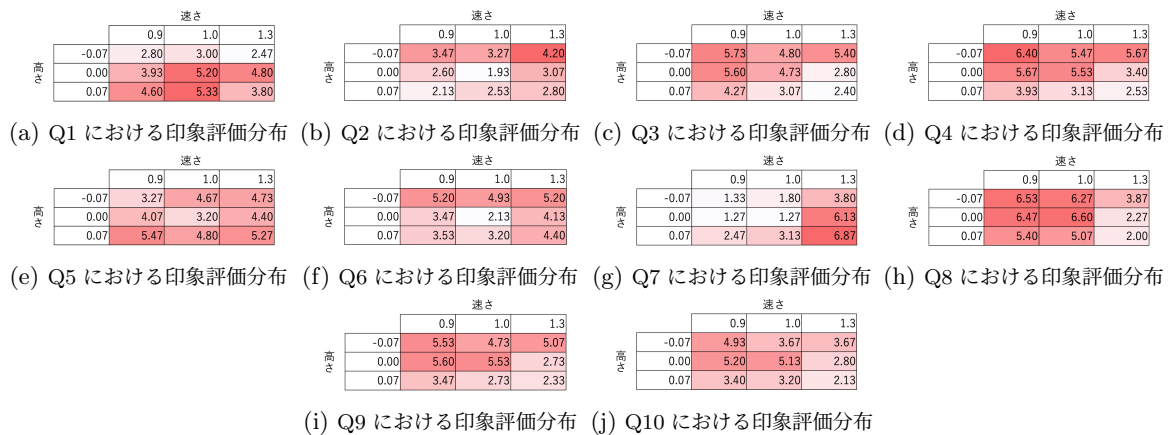


図 4: アニメキャラクターに対する質問項目ごとの平均分布

の場合には評価が低下する傾向が確認された。また、「冷静で気分が安定していると思う」(図 6(d)) では、発話速度が遅い組み合わせ(速さ 0.9)が高評価を受けたが、速い発話速度(速さ 1.3)では評価が低い傾向が見られた。「上品で大人っぽいと思う」(図 6(i))においても、発話速度が遅く音高が低い場合(速さ 0.9, 音高 0.00)が最も高評価を得たが、速い発話速度や高音の組み合わせでは評価が低くなる傾向が確認された。

一方で、「堅くて、不快感があると思う」(図 6(h))や「冷たくて感情がこもっていないと思う」(図 6(f))といった項目では、音声特性が評価に影響を与えているものの、外見の影響が相対的に大きいことが示唆された。特に、速い発話速度(速さ 1.3)の場合には評価が低くなる傾向が見られたが、アニメキャラクターの外見がその影響を部分的に緩和していると考えられる。

また、「会社受付にいても違和感がないと思う」(図 6(j))においては、発話速度が遅く、安定した声の組み合わせ(速さ 0.9)が最も適切と評価されたが、速い発話速度では適合性が低下した。これらの結果から、アニメキャラクターは外見による親しみやすさや感情的な

魅力を活かしつつ、発話速度や音高の設定において慎重な調整が必要であることが示唆された。特に、業務適応性や信頼感を高めるためには、発話速度が遅く安定した組み合わせが効果的であると考えられる。

4.3 考察

本研究の結果から、音声特性と外見の組み合わせが AI 受付アシスタントの印象評価に及ぼす影響が明らかになった。特に、外見の影響が音声特性を上回ることが確認された。

音声特性に関しては、発話速度や音高が特定の印象評価に与える影響が認められた。例えば、発話速度が遅く、音高が高い設定(速さ 0.9, 音高+0.07)は「親しみやすさ」(Q1)の評価を高める一方、速い発話速度(速さ 1.3)は否定的な影響を与える傾向があった。また、「冷静で気分が安定している」(Q4)では発話速度が遅い組み合わせが好まれた。しかし、これらの影響は外見による影響に比べて相対的に小さく、音声特性の調整だけでは十分でない可能性が示唆された。

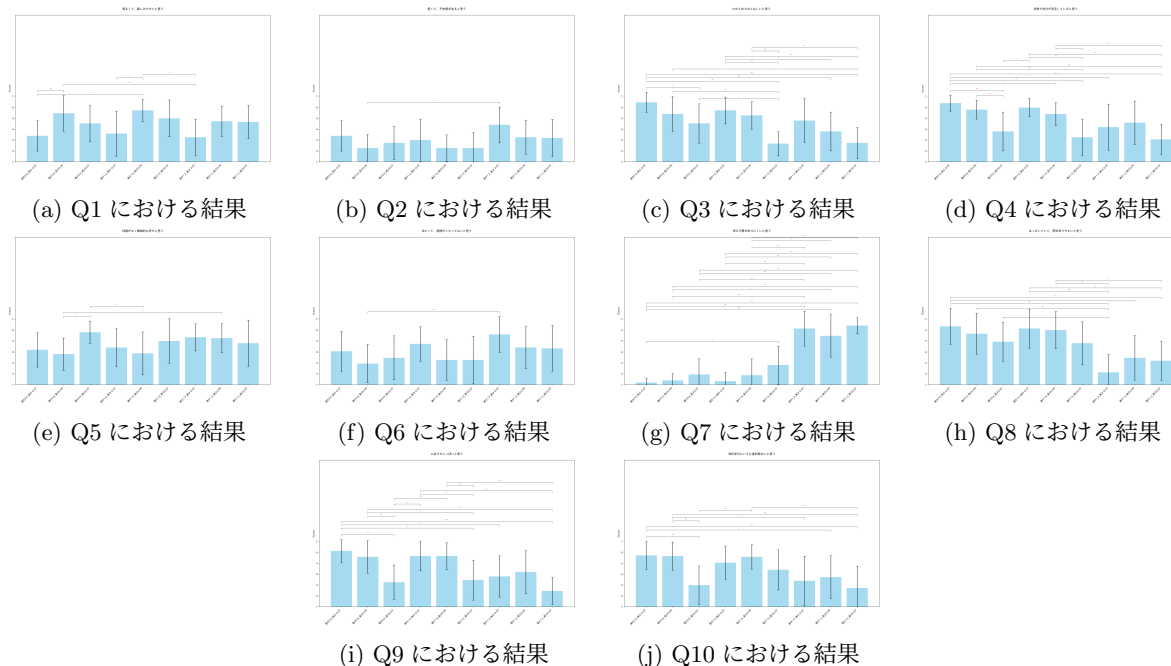


図 5: リアルキャラクタに対する発話速度と音高の組み合わせ間での有意差

外見に関しては、リアルキャラクタとアニメキャラクタの間に大きな印象の違いが認められた。リアルキャラクタは、「親しみやすさ」(Q1) や「上品で大人っぽい」(Q9) の評価が高く、フォーマルな場面での信頼性や堅実性を強調する特徴が確認された。また、「冷静で気分が安定している」(Q4) といった業務適応性を示す項目でも高い評価を得た。一方で、「抑揚がなく機械的な声だと思う」(Q5) や「冷たく感情がこもっていないと思う」(Q6) といった否定的な印象を与える要素も存在し、改善の余地があると考えられる。

アニメキャラクタは、「親しみやすさ」(Q1) や「はっきり聞き取りやすい」(Q8) など、カジュアルで親近感のある印象を与える傾向が強い結果が得られた。しかし、「冷静で気分が安定している」(Q4) や「会社受付にいても違和感がない」(Q10) といった業務適応性や信頼性を示す項目では、リアルキャラクタと比較して評価が低下した。このことは、アニメキャラクタが親しみやすさを活かしつつ、信頼感を高める工夫が必要であることを示している。

以上の結果から、AI 受付アシスタントのデザインにおいて、外見が印象形成において重要な役割を果たすことが確認された。特に業務環境では、リアルキャラクタが適している可能性が高い一方、アニメキャラクタは親近感を重視した環境で効果的であると考えられる。音声特性に関しては、発話速度や音高の微調整が印象改善に寄与する可能性があるが、外見とのバランスを考慮する必要がある。

5 おわりに

本研究では、AI 受付アシスタントにおける音声特性と外見の組み合わせが印象に与える影響を調査した。その結果、外見が印象形成において特に大きな役割を果たす一方、音声特性の調整も印象改善に一定の効果を持つことが明らかになった。本研究は、AI 受付アシスタントの設計指針を提供するだけでなく、AI キャラクタの社会受容性向上に貢献する知見を示した。

今後は、男性キャラクタや異なる業種での評価を実施し、さらに長期的なユーザー体験調査を行うことで、より一般化可能な知見の取得を目指す。これらの知見は、受付業務に限らず、接客、教育、介護など多様な分野での AI キャラクタ活用に応用可能であり、将来的には個人に最適化された AI キャラクタの設計へと発展することが期待される。

参考文献

- [1] 岩川壱, 加納政芳, ジメネスフェリックス, 早瀬光浩, 吉川大弘, "ロボットにおける発話速度と声域の違いが印象評価に与える影響", HSS2023, p5, 2023.
- [2] Google, ImageFX, <https://labs.google/fx/tools/image-fx>, accessed : 2025.01.15.
- [3] Zhang, Wenxuan and Cun, Xiaodong and Wang, Xuan and Zhang, Yong and Shen, Xi and Guo, Yu and Shan, Ying and Wang, Fei, "SadTalker: Learning Realistic 3D Motion Coefficients for Stylized Audio-Driven Single Image Talking Face Animation", arXiv preprint arXiv:2211.12194, 2022.



図 6: アニメキャラクターに対する発話速度と音高の組み合わせ間での有意差