

擬人化対象における外見と音声の整合性が印象形成に与える影響

Effects of Audiovisual Consistency on Impression Formation in Anthropomorphic Daily Objects

石川 菜月¹ 早瀬 光浩^{1*}
Nazuki Ishikawa¹ Mitsuhiro Hayase¹

¹ 梶山女学園大学

¹ Sugiyama Jogakuen University

Abstract: 本研究では、生成 AI により擬人化した対象を用い、外見と音声の整合性（マッチ度）が印象形成に与える影響を体系的に分析した。外見特徴と音声特性の組み合わせによる複数条件の刺激を作成し、印象評価実験を実施した。その結果、整合条件ではマッチ度や信頼感が高く、不整合条件では違和感が顕著に増大する傾向が確認された。さらに、音声そのものではなく、外見との整合性（マッチ度）が信頼感や好感度を規定する主要因であることが示された。本研究は、対話エージェントにおける視聴覚統合に基づく印象設計の基礎構造を示す。特に、視聴覚整合性が印象形成を媒介する基本構造を実証した。

1 はじめに

AI が急速に進化した現代では、AI を用いて生成された多様なコンテンツが広く普及している。特に SNS や動画配信プラットフォームにおいて、AI 技術によって生成された画像や動画、あるいは 3D アバタを用いて活動するバーチャル YouTuber (Vtuber) を目にする機会が増加している [1]。しかし、これらのコンテンツにおいて、キャラクターの外見と音声の間に違和感を覚える事例が見受けられる。このような現象から、キャラクターの外見ごとに「適合すると知覚される音声」が存在し、その知覚は個人差を伴う可能性があると考えられる [2]。

これまで、擬人化や AI キャラクターに関する多くの研究は人型やキャラクター性が高い対象に焦点を当ててきた [3, 4] が、擬人化表現を最小限に抑えた日常物体において、外見と音声の整合性（マッチ度）がユーザの印象形成に与える影響については十分に検討されていない。そこで本研究では、ボール、椅子、ぬいぐるみといった日常物体を対象とし、目や口といった最小限の擬人化表現および音声を付与したキャラクターを作成し、それらの要素がマッチ度（整合性）や信頼感、好感度といった印象評価にどのような影響を及ぼすかを体系的に調査・分析することを目的とする。

2 関連研究

擬人化とユーザの印象形成については、音声のピッチ（音高）や速度が親しみやすさなどの評価に影響を与えることが示されている一方で、キャラクターの外見の影響に比べて音声特性の影響は相対的に小さいという課題も指摘されている [5]。また、視聴覚情報の不一致が知覚に影響を与える現象としては「マクガーク効果」が知られており、視覚と聴覚の食い違いが誤った知覚を引き起こすことが理論的に示されている [6]。

本研究は、これら視聴覚統合に関する知見を背景に、日常物体の擬人化における「外見と音声の一致・不一致」が印象形成に与える影響を検討するものである。さらに、近年の生成 AI 技術の発展により、高品質な画像や音声の生成が容易となっている背景を踏まえ、本研究では Google Gemini による画像・動画生成と、VOICEVOX による音声合成を組み合わせることで、外見と音声の特性を精密に制御した実験刺激を作成し、その有用性と課題についても考察を行う。

3 擬人化動画の生成

3.1 擬人化対象の選定と分類

対象として、形状や使用目的が異なる「ボール」「椅子」「ぬいぐるみ」を選定した。これらにピクサー風の 3D 質感を持たせ、被験者が現実世界の物体に近いキャラクターとして認識できるよう配慮した。

*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



(a) ボール (b) いす (c) ぬいぐるみ

図 1: 実験で使用する生成した擬人キャラクター

生成した擬人化キャラクターの例を図 1 に示す。

3.2 画像・動画生成と音声の合成

Google Gemini を用い、統一されたプロンプトで「大きな目と小さな口」を持つ画像を生成した。音声は VOICEVOX の女性声・男性声各 1 種をベースに、音高を「低 (-0.5)」「ニュートラル」「高 (+0.5)」の 3 水準に調整した。最終的に、3 物体 × 2 性別 × 3 音高 の計 18 種類の動画を作成した。生成条件およびパラメータは全刺激で統一し、再現性を確保した。

4 印象評価実験

4.1 実験の目的と仮説

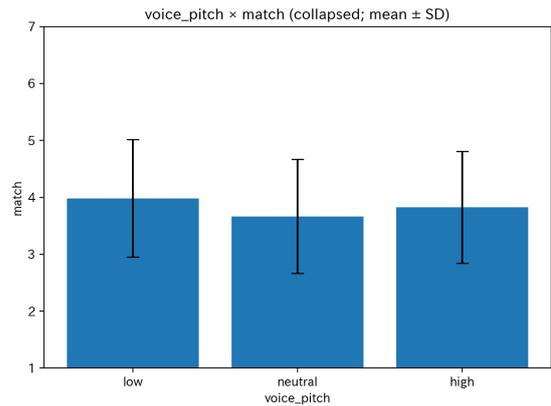
本実験の目的は、日常物体を擬人化した AI キャラクターにおいて、音声特性（音高）および外見と音声のマッチ度が、視聴者の印象評価に与える影響を明らかにすることである。特に、視聴者が「外見と音声調和している」と知覚する条件において、エージェントへの信頼感や好感度が高まると予測した。本研究では、以下の 2 つの仮説を設定する。

仮説 1: 外見と音声のマッチ度が高い条件ほど、信頼感も高く評価される傾向が示される。

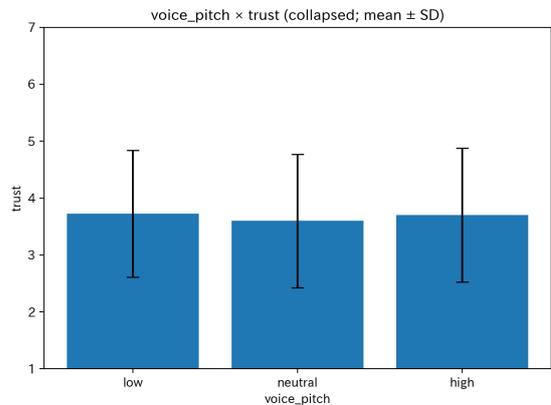
仮説 2: 音声の高さは、外見と音声のマッチ度に有意な影響を与える。

4.2 実験環境と手順

実験は Yahoo!クラウドソーシングを用い、ランダムに抽出された 45 名（有効回答）を対象に実施した。被験者は、作成した 18 種類の動画（3 物体 × 2 性別 × 3 音高）をランダムな順序で視聴し、「マッチ度」「信頼感」「好きさ」「親しみやすさ」「自然さ」などの項目について、7 段階リッカート尺度（1：全くそう思わない～7：非常にそう思う）で評価を行った。なお、「マッチ度」は、外見と音声の調和感に関する単一の主観評価



(a) 音高別マッチ度



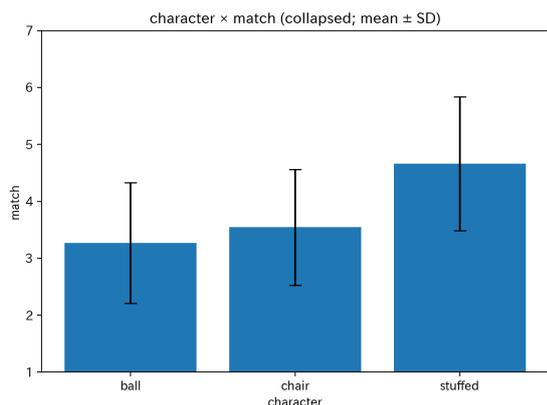
(b) 音高別信頼感

図 2: 音声の高さ（低・中・高）がマッチ度 (a) および信頼感 (b) に与える影響（平均 ± SD）

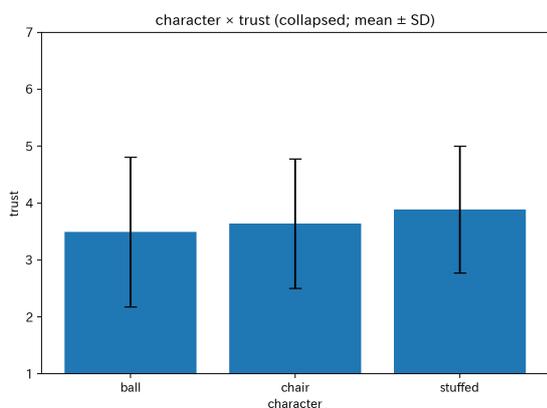
項目として測定した。刺激提示順は被験者ごとにランダム化し、順序効果の影響を最小化した。分析の信頼性を担保するため、回答時間が極端に短いデータや未回答項目を含むデータはあらかじめ除外している。

4.3 実験結果：音声特性の影響

音声の高さ（ピッチ）がマッチ度に与える影響を検討するため分散分析を行った結果、音高の主効果が有意であった ($F(2, 86) = 7.33, p = 0.0011$)。事後検定の結果、「低い声」と「ニュートラル」、「ニュートラル」と「高い声」の間にそれぞれ有意差が認められたが ($p < .05$)、図 2(a) に示す通り「低い声」と「高い声」の間には有意差は認められなかった。一方で、図 2(b) 信頼感に対する音高の影響については、主効果は有意ではなかった ($F(2, 86) = 2.63, p = 0.077$)。このことから、音声の高さ単独では信頼感を直接決定づける要因にはなりにくいことが示唆された。

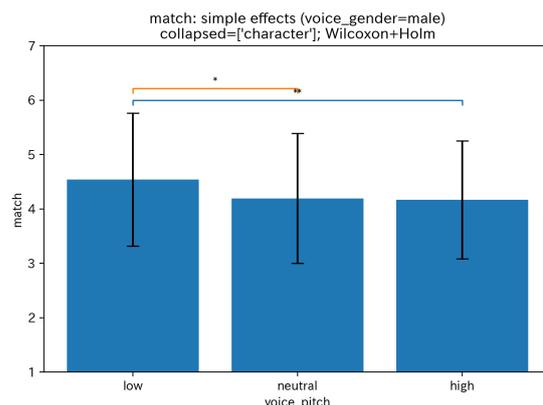


(a) 物体別マッチ度

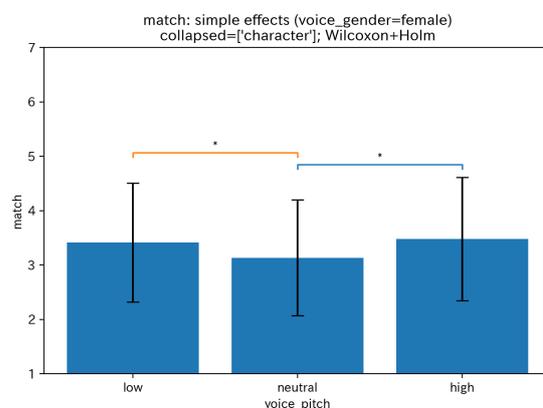


(b) 物体別信頼感

図 3: 擬人化対象 (ボール・椅子・ぬいぐるみ) がマッチ度 (a) および信頼感 (b) に与える影響 (平均±SD)



(a) 男性声の単純主効果



(b) 女性声の単純主効果

図 4: 音高×性別の交互作用における単純主効果 ((a): 男性声, (b): 女性声)

4.4 実験結果：外見と性別の影響

キャラクターの外見(物体)の違いが与える影響を分析したところ、図3の通り、マッチ度 ($F(2, 86) = 60.58, p < .00001$) および信頼感 ($F(2, 86) = 8.21, p = 0.00055$) の両項目において極めて強い主効果が確認された。物体別では、特に「ぬいぐるみ」が他の物体(ボール, 椅子)よりもマッチ度・信頼感ともに高く評価される傾向にあった。

また、音声の性別についても検討したところ、マッチ度 ($F(1, 43) = 56.64, p < .00001$) および信頼感 ($F(1, 43) = 26.03, p < .00001$) において有意差が認められ、全体として男性声よりも女性声の方が高く評価される傾向が示された。

さらに、音高と性別の交互作用を分析したところ、男性声(図4(a))と女性声(図4(b))では、音高の変化がマッチ度に与える影響のパターンが異なることが示された ($F(2, 86) = 4.48, p = 0.014$)。

4.5 相関分析による整合性の検証

外見と音声のマッチ度が他の印象項目に与える影響を明らかにするため相関分析を行った結果、図5に示す通り、マッチ度と信頼感の間には極めて強い正の相関が認められた ($r = 0.82, p < .001$)。同様に、マッチ度と好感の間にも強い正の相関が確認された ($r = 0.87, p < .001$)。以上の結果から、音声の高さがマッチ度に影響を与えるという仮説2は支持された。また、信頼感そのものへの音高の直接効果は認められなかったものの、マッチ度を介した間接的関係が確認されたことから、仮説1は部分的に支持された。

5 考察

5.1 印象形成を媒介する「マッチ度」の役割

本研究の最も重要な知見は、音声の高さ(音高)が信頼感に直接影響を与えるのではなく、「外見と音声のマッチ度」を介して間接的に信頼感や好感度を規定す

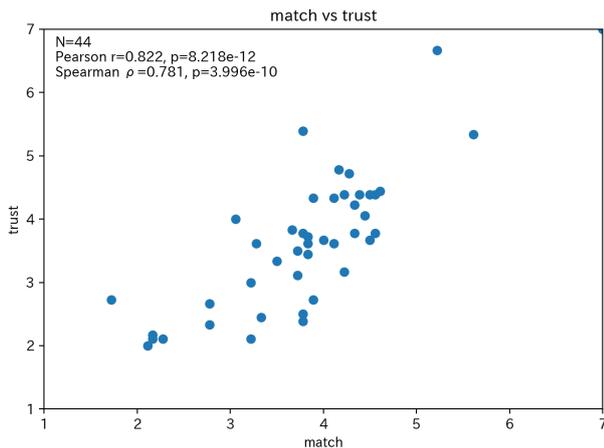


図 5: マッチ度と信頼感の相関関係 ($r = 0.82, p < .001$)

る点にある。相関分析の結果、マッチ度と信頼感の間には強い正の相関 ($r = 0.82$) が確認された。これは、ユーザが声そのものよりも「外見に適合した声」であるかという視聴覚整合性を優先して評価していることを示唆する。したがって、日常物体の擬人化においては、対象の形状や質感に適合した音声の選定が重要である。

5.2 擬人化対象による期待値の差異

対象物体(ボール, 椅子, むいぐるみ)の主効果は、マッチ度 ($F(2, 86) = 60.58$) および信頼感 ($F(2, 86) = 8.21$) の両項目で有意であった。特にむいぐるみは他の物体よりも高く評価される傾向が確認された。これは、むいぐるみが擬人的属性を想起させやすく、音声付与に対する心理的障壁が低いためと考えられる。本結果は、対象の生物らしさに応じて許容される音声特性が変化する可能性を示唆する。

5.3 音声の性別と音高の交互作用

全体として男性声よりも女性声の方がマッチ度・信頼感ともに高く評価される傾向 ($p < .00001$) が示された。さらに、音高と性別の間に有意な交互作用 ($F = 4.48, p = 0.014$) が認められ、女性声では音高の変化がマッチ度により敏感に反映された。これは、擬人化エージェントに対する期待や社会的バイアスが視聴覚統合に影響する可能性を示唆する。ただし、本結果は音色や文化的要因の影響も受ける可能性があり、今後の検証が必要である。

5.4 生成 AI を用いたエージェント設計への指針

本研究で得られた「音高がマッチ度を左右し、マッチ度が信頼感を媒介する」という構造は、生成 AI を用いたエージェント設計への指針となる。外見に対してランダムに音声を割り当てるのではなく、対象から想起される期待ピッチとの整合性を最適化することが重要である。今後は、より広いピッチ条件やイントネーションなど時間的要素の影響を検討する必要がある。

6 おわりに

本研究は、日常物体の擬人化エージェントにおける外見と音声のマッチ度が印象形成に与える影響を明らかにした。分析の結果、音声の高さ(ピッチ)はマッチ度に有意な影響を与え、マッチ度が高いほど信頼感や好感度も高まることが確認された。これにより、日常物体の擬人化においても視聴覚情報の整合性が極めて重要であることが示された。

今後の課題として、音高の範囲拡張や、発話の「間」やイントネーション等の時間的要素が与える影響の検証が挙げられる。本知見は、生成 AI を用いたエージェント設計における重要な指針であり、視聴覚の整合性が印象設計の基盤となる基本構造であることを示した。

参考文献

- [1] 佐藤慎, "バーチャル Youtuber におけるイメージと身体", 早稲田大学 修士論文, 2021.
- [2] 重野純, "マクガーク効果における個人差", 音声言語医学, Vol.39, No.3, pp.267-273, 1998.
- [3] 高橋日和, 真栄城哲也, "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", 計測自動制御学会 システム・情報部門 学術講演会 2021 (SICE-IS 2021), 2021.
- [4] 三村亮太郎, 佐藤弘喜, "音声案内キャラクターの外見および音声に対する印象評価 — ユーザー体験向上に向けたキャラクターの設計", 日本デザイン学会第 72 回研究発表大会概要集, 2025.
- [5] 垣見向夏花, 早瀬光浩, "音声と外観が AI 受付アシスタントに与える印象", HAI シンポジウム, 2025.
- [6] 竹内晶, "刺激に関する知識とバイアスがマクガーク効果に及ぼす影響 — 母語は錯覚にどう影響するのか —", 東京大学言語学論集, Vol.46, pp.271-320, 2024.