

人間への類似度を段階付けされた顔画像と音声の印象評価

The evaluate of impression for face graphics and voice which have the levels of likeness to human

田中 かおり*¹

TANAKA kaori

小島 一晃*²

KOJIMA Kazuaki

松居 辰則*²

MATSUI Tatsunori

*¹早稲田大学 大学院人間科学研究科

Graduate School of Human Sciences, Waseda University

*²早稲田大学 人間科学学術院

Faculty of Human Sciences, Waseda University

In this study, we discuss design issues for user-friendly computers that have human-like expression and communicate with human. HCI studies have pointed out that familiarity with tools can improve their usability. Therefore, we conducted an experiment to confirm a hypothesis that degree of human-likeness of computer agents can influence people's impression. We focused two components, face representation and speech voice. We asked subjects to evaluate impression computer agents designed by controlling the face and voice. The result indicated that representation of low degree of human-likeness increased friendly impression for the agents.

1. はじめに

1.1 擬人化エージェント研究と問題点

近年、人間とのコミュニケーションを重視した機械の研究が注目されている。その一端として、人と機械の自然でスムーズなインタラクションの検討と実現を目的とする擬人化エージェントの研究がある。擬人化エージェントとは、人の外見や知能を模倣したロボットやシステムを指す。一般的なエージェントの役割はユーザの操作負担を軽減することだが、擬人化エージェントでは更に表情や発話といった人間の持つ機能を備えることで、従来のマウスやキーボード、エラーメッセージのような人間にとって自然ではない情報伝達形式のギャップを解消することが期待されている。

その実現可能性は、次のような人間の性質によって支持される。人は自分とインタラクションを持つ機械が社会的な振る舞いをすると、無意識のうちに相手に擬似的な人格を見出すという性質 [1] と、人は全く動かない無機物より、発話したり視線を持つ無機物に親近感を覚える性質 [2] の二つである。つまり、人は機械に擬似的な人格を見出しやすい傾向があるため、機械とのインタラクションでも人間を相手にしたような自然なインタラクションを実現できる可能性は高い。しかし、それが好意的なものであるとは無条件に肯定はできず、機械の表現や振る舞いに人間らしさが感じられるかに依存するといえる。

このような観点からユーザフレンドリーな擬人化エージェントを追求すると、その外見や動作は現実の人間に近いものになると思われがちである。近年の技術の進歩によって、黙って座っていれば人間と区別のつかないロボットなども現れ始めている。しかし実際にそれらが動いている場面に遭遇すると、人間はロボットに対して強い違和感や嫌悪感を覚える。これは「不気味の谷」現象 (森 1970) と呼ばれ、人間への類似度と親しみやすさがある程度高めた機械に起こる印象の急激な落ち込みを指す。この現象は実証がなされていないが、人間を模したロボットや CG の分野で多く観察されている。

このことから、機械の人間への類似性と親和性はセットで考える必要があることがうかがえる。既存の顔画像の研究では、イラスト調のもの (中沢正幸 2004)、写実調のもの (パズラテ・レジェブ 2006)、3DCG (羽田拓郎 2003) などの様々な種類の画像

が使用されており、全体を体系化して述べるのは難しい。一方発話は音声のピッチレンジ [4] や特徴量 [3] などの明確な物理量と印象が結び付けられ、顔画像の印象評価に比べ解析が進んでいる。

以上のように画像と音声それぞれでも、人間への類似性に着目した既存研究は少ない。また擬人化エージェントには画像と音声の両方が備わっていることを考えると、最終的には複合的な印象評価を行う必要があるといえる。

1.2 本研究の目的

本研究は、顔画像と発話の 2 つの刺激それぞれから、人と擬人化エージェントが情報交流を行うとき、エージェントの持つ表現の違いが印象にどのような影響を与えるのかを明らかにすることを目的とした一連の流れの中の一部である。既に顔画像については実験を行っており [5]、ここではリアルな表現よりデフォルメした表現の顔を持つエージェントの方が親和性が高いと言う結果が出ている。発話実験も同様に、作成した音声刺激を人間の機械的かの観点で程度付けし、それが人の行動と印象形成に与える影響を調べる。

2. 音声の印象評価実験実験

2.1 材料

音声の印象評価実験をするにあたり、最初に刺激抽出のための予備実験を行った。まず肉声と合成音を STRAIGHT を用いて段階的に混ぜ合わせた。モーフィングは 5%刻みとし、数値が大きくなるにつれて合成音声の割合が高くなるよう設定した。そうして作成された 0%から 100%までの 21 個の音声刺激に対して、予備実験として同異性の判定と機械性 - 人間性の評価を行った。被験者は 20 ~ 32 歳までの大学生 8 名で、男女比は男 : 女 = 5 : 3 だった。

評価は作成時の混合割合 0%と 5%のように隣り合った割合同士 20 組の組み合わせで行った。その結果、同異性において 0%-5%、15%-20%、70%-75% で有意差が見られた (図 1)。ここから次の印象評価実験に使用する刺激として、0% (肉声) と 15%、70% の合成音声を抽出した。また同時に測定した機械性 - 人間性の評価では、0%-5%、40%-45%、95%-100% で有意差が抽出され (図 2)、同異性とは異なる結果となった。両方に共通する特徴としては、x 軸と y 軸が単純な右肩下がりの比例関係ではなく、一度 30%付近を底にして 100%に向けやや上昇している点が挙げられる。

連絡先: 田中かおり, 早稲田大学大学院人間科学研究科, 埼玉県所沢市三ヶ島 2-579-15, mono_season101@asagi.waseda.jp

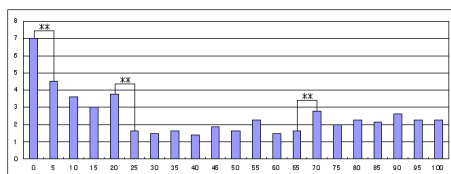


図 1: 同異性測定と検定結果

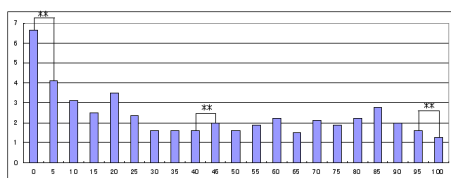


図 2: 機械性 - 人間性測定と検定結果

合成音声割合	追従率
0%	0.64
15%	0.75
70%	0.76
100%	0.90

表 1: 合成音声割合と追従率

2.2 手続き

次に印象評価実験を行った。刺激は予備実験で抽出した 0%(肉声),15%,70%に 100%合成音声を加えた計 4 つの音声を使用した。被験者は 19~23 歳の大学生 4 名だった。被験者にはパソコン画面に描かれた分かれ道 (図 3) の左右いずれかに進む際に、刺激音声から「右が正解です」「左が正解です」といったアドバイスが与えられた。この音声は案内ソフトから流れる機械の声だが、不具合があり時々間違った道を示すことがあるため、案内を参考にするのは自由であると教示した。間違った道を選ぶと行き止まりの画像 (図 4) が表示され、正解の道を選ぶと次の分かれ道が表示された。音声刺激は人毎に順番を変えて与えた。この操作を正解の道が 20 回選択されるまで続けさせ、最後に案内の音声に対する印象評価を行った。

2.3 結果と考察

その結果、道の選択について 0%(肉声) 音声の追従率 (音声と同じ道を選ぶ確立) が最も低く、100%合成音声の追従率が最も高くなった (表 2.3)。印象評価については今回は被験者数が少なかったためプロフィールを分析したが、特徴は見られなかった。

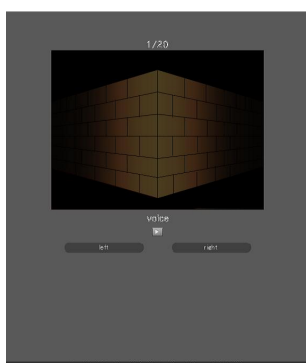


図 3: 分かれ道

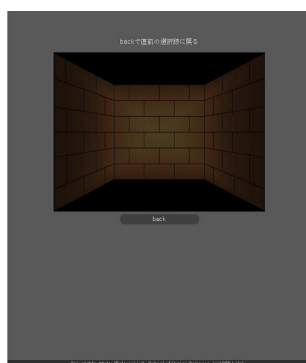


図 4: 行き止まり

プロフィール分析では特徴が抽出されなかったが、追従率を比較すると明確に差が現れた。特に加工 0% 音声 (肉声) と、100% 音声 (純粋合成音声) 間では、被験者の選択に顕著な違いがあった。これは、使用した形容詞群に表されない要因があることを示しているため、今後検討が必要である。また、被験者は肉声で

アドバイスされるより純粋合成音声でアドバイスされると指示と同じ道を選ぶ傾向があった。今回は情報提供者が機械であるとの教示のため、人間的な声がかえって違和感を生じさせたとも考えられる。顔画像の結果と合わせると、生身の人間ではないことが明らかな擬人化エージェントと音声を組み合わせた場合、人間にとって一番親しみやすいエージェントを実現するのは生身の顔の画像と肉声を持ったエージェントでは無いことが示唆される。

3. まとめ

本研究では擬人化エージェントを構成する要素として大きなファクターを占める顔画像と発話音声を複合的に評価することを最終目標とし、特に今回は発話音声について多く採り上げた。情報を提供する際、顔画像で写真のような写実的な顔よりデフォルメを施した顔に人は好印象を抱いた一方で、発話音声では印象評価に特徴的な差は見られなかったものの、提供された情報に対する被験者の反応に影響を与えることが分かった。特に今回は純粋な合成音声と肉声に比べて被験者の追従的な反応を引き出す傾向があることが示された。

今後は、以上のようにして個別に得られた顔画像と発話音声の印象評価を、いかに複合的にまとめていくかを課題とする。

参考文献

- [1] B.Reeves,C.Nass : The Media Equation: How People Treat Computers, Television, and New Media like Real People and Places, Cambridge University Press(1998)(邦訳:人はなぜコンピューターを人間として扱うか「メディアの等式」の心理学, 細馬宏通 訳, 翔泳社 (2001))
- [2] 米澤朋子, 山添丈丈, 内海章, 安部伸治: 視線コミュニケーションのためのぬいぐるみの視線行動の設計と分析, 電子情報通信学会論文誌. D, 情報・システム J92-D(1), pp81-92(2009)
- [3] 鈴木紀子, 竹内勇剛, 片桐恭弘: 音声の韻律が人間の行動と印象形成におよぼす影響, ヒューマンインタフェースシンポジウム 2003 論文集, pp.183-186(2002).
- [4] 河原達也, 小川一, 堂本修司: 声質に関する印象評価および心理的類似性と音響的類似性について, 電子情報通信学会技術研究報告. SP, 音声 95(140), pp7-13(1995)
- [5] 田中かおり, 小島一見, 松居辰則: デフォルメを段階付けされた擬人化エージェントに対する印象評価, 日本認知科学会第 26 回大会 発表論文集, pp260-262(2009) .