

# 自動ソーシャルスキルトレーニングに向けた 仮想エージェントの設計

## Design of Embodied Conversational Agents for Social Skills Training

田中 宏季\* 中村 哲

Hiroki Tanaka Satoshi Nakamura

奈良先端科学技術大学院大学

Nara Institute of Science and Technology

**Abstract:** Social skills training by human instructors is a well-established method with which users can obtain appropriate social interaction skills and strengthen their social self-efficacy. Our previous works automated social skills training by developing a virtual agent that teaches social skills through interaction. This study attempts to investigate the effect of virtual agent design towards social skills training. We believe that our findings are important in designing a personalized automated social skills trainer.

### 1 はじめに

ソーシャルスキルトレーニング (SST) は自閉スペクトラム症をはじめとした幅広く社会的コミュニケーションを苦手としている人々に適用されている方法であり、医療機関や各種の就労支援施設、作業所、学校、職場などで人間のトレーナーにより実施されている [1]。我々はこれまで、仮想エージェントを用いて SST の自動化を行う研究を進めており、人間の行なう SST を模倣した自動 SST を開発し、自閉スペクトラム症の小児や成人での訓練効果を確認している [2, 3, 4]。システムは人間の行動のビデオモデリング、リアルタイムの行動認識およびフィードバックを含んでいる。実験により、システムが従来の本を読むトレーニングや、上手な動画を視聴するトレーニングと比較して、ソーシャルスキルの訓練に有効であることを報告した。また人間と話す場合と比較し、仮想エージェントと話す方がより話しやすく緊張が少ないことを確認している。自動 SST は、小児から成人の男女、自閉スペクトラム症、統合失調症までの種々の集団を対象としている。しかしながら、仮想エージェントの設計に関して、どのような仮想エージェントがより好まれるのか、より受け入れられるかの調査はなされていない。自動 SST が、本人に適切して受け入れられるためには、相性などを考慮した詳細な調査が必要だと考える。本研究では、人間およびロボットではなく、仮想エージェントの中での設計の比較を試みる。仮想エージェントは、人間や

ロボットと比べ色々な設計を試しやすいという側面もあるため、種々のキャラクターを用意した。

仮想エージェントにおける設計に関して、これまでの先行研究では、外見、動きの観点から種々の仮想エージェントの比較 [5, 6, 7]、リアリズムの比較 [8]、対話場面の深刻度に対する、体や目のパーツの割合および大きさを比較した際の適切さ [9]、高齢者向けの声の種類 [10]、性差、人種がユーザの自己効力感へ及ぼす影響 [11] などについて調査がなされてきた。我々は、これらの知見および評価尺度を活用し、より好ましく受け入れやすい自動 SST の構築に向けた仮想エージェントの設計について調査することに取り組む。

本研究では、ソーシャルスキルの訓練をするための種々の仮想エージェントの設計を用意し、好ましさ、受け入れやすさ、リアルさ、親しみやすさなどについて評価を行い、またどのような要素が好ましさに関連しているかについて調査した。仮想エージェントの設計と、ユーザの性別、年齢、自閉症傾向による好ましさの違いについても調査した。本研究の目的について以下に整理する。

1. どの仮想エージェントの設計が、好ましく、受け入れられるかを調査
2. 好ましさに関連している要素を調査
3. 性別、年齢、自閉症傾向による好ましさの関係性を調査

\*連絡先：奈良先端科学技術大学院大学  
〒 631-0192 奈良県生駒市高山町 8916-5  
E-mail: hiroki-tan@is.naist.jp

## 2 方法

### 2.1 仮想エージェントの設計

本研究では、まず図1に示す通り仮想エージェントのイラスト一枚絵を用意した。仮想エージェントは、アニメーション専門のデザイン会社に依頼し設計した。全て正面を向き、感情表現はないものとした。(a)と(b)および(c)と(d)は年齢が10代後半で、性別を統一し、リアルさのみ変化するように設計された。(e)は小児への使用を想定し、無生物のものを作成した。(f)も小児への使用を想定し、人間以外の生物、ここでは犬として作成した。(g)に関しては、(a)および(b)と類似したリアルな3Dモデルの仮想エージェント ([www.renderhub.com/3d-models](http://www.renderhub.com/3d-models)) を選び、正面からのスクリーンショットを撮影した。(h)の仮想エージェントはマルチモーダル会話機能を有したGreta [12]においてデフォルトで用意されているエージェントで、Autodesk character generator (<https://charactergenerator.autodesk.com/>) で作成されたものである。フランス、英語圏の文化で主に使用することを想定された仮想エージェントである。(f)の仮想エージェントは、40代の日本人女性を想定して設計された。現在の自動SSTでは、(f)の仮想エージェントを使用している [4]。

上記画像に対して、「こんにちは、一緒にコミュニケーションの練習をしましょう」(5秒間の発話長)の文をGoogle Text-to-Speech (<https://cloud.google.com/text-to-speech?hl=ja>) で読み上げ、男女それぞれの声で埋め込んだ。(e)と(f)の仮想エージェントに関しては、小児向けを想定し、通常的女性音声合成よりもピッチを高く設定して作成した。

(h)および(i)のエージェントはGretaで動作する様に3Dモデルが作成されているため、Greta上で動作させその動画も収録した。動きにおいては、発話内容に同期して身振り(手を挙げる、自分の胸に手を当てる)を加えた。(h)と(i)で同一の行動を生成し、上記と同様の発話内容(8秒間の発話長)で日本語音声合成およびそのリップシンクを作成した(図2参照)。音声合成はCereproc (<https://www.cereproc.com/>) のYukiの声を使用した。

### 2.2 参加者

データ収集は、クラウドソーシングサービスのクラウドワークス (<https://crowdworks.jp/>) にて、オンラインで参加者を募集した。参加募集条件には、18歳以上の日本国籍を有する方という条件を設定した。各参加者の労力を分けるため、本研究では3回に分けてデータ収集を行った。それぞれ、データ1) n=305、データ2) n=305、データ3) n=302のデータを収集し

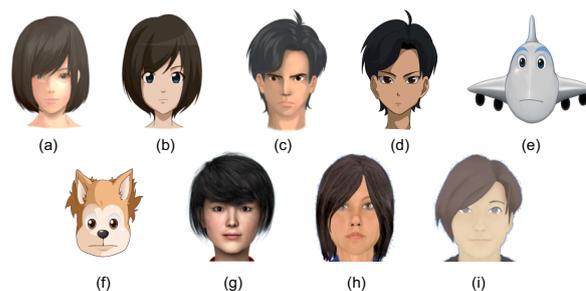


図1: 使用した仮想エージェントの画像

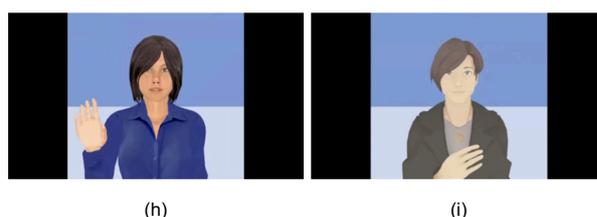


図2: 使用した仮想エージェントの動画

た。データ1は、画像に対して受け入れやすさ、好ましさ、親しみやすさ、特定の要素の好ましさ(目、顔、髪、声など)、自閉症傾向、失感情症について調査するために収集された。データ2は、画像に対してリアルさ、信頼できるか、不気味さについて調査するために収集された。データ3は、(h)、(i)の動画について調査するために収集された。

データ1に関して、年齢、男女比を図3に示している。データ2、3においても、年齢、性別に関してほぼ同じ分布となっていることを確認した。後に群を分けた分析を行い、そこで年齢に関して45歳を高低の基準とした(高年齢: n=84, 低年齢: n=221)。

### 2.3 自閉症傾向および失感情症

データ1では、自閉症傾向を測定するため成人版対人応答性尺度-2 (SRS) [13]、失感情症を測定するためトロントアレキサイミア尺度-20 (TAS) [14]を収集した。両方とも総合得点を計算し、下位尺度は本論文では取り扱わないこととした。両者の関係性として、スピアマンの相関係数で0.67( $p < 0.05$ )となった。後の分析では、自閉症傾向を表すための尺度としてSRSを使用する。SRSの平均と標準偏差は、72.0 (SD: 29.3)であった。本研究では、81点 [15]のカットオフ値をSRSの高低の基準とした(高SRS: n=113, 低SRS: n=192)。

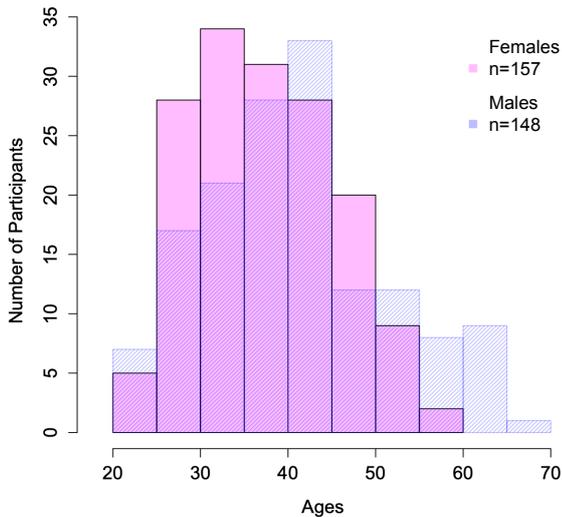


図 3: 参加者の年齢と性別

## 2.4 質問項目

本研究では、Esposito および Ring らの研究を参考に質問項目および尺度を作成した [5, 6, 10, 9]。質問項目は、トレーナとして受け入れることができる、話の聞き手として受け入れることができる、リアルだ、不気味だ、信頼できる、顔が好みだ、目が好みだ、声が好みだ、年齢が好みだ、好みだ、親しみやすい、で構成されている。各質問項目は、Google フォームを使用して回答された。データ 1 では、SRS および TAS の取得後に各質問項目への回答を行った。データ 3 では、動画が上半身全体を含んでいるので、上記に加えて、服装が好みだを追加した。参加者は、まず初めに全ての画像の一覧を見て (図 1)、全体の仮想エージェントの印象をつかんだ上で、個別の仮想エージェントを視聴しそれぞれの質問項目を 5 段階のリッカート尺度 (1: 全くそう思わない、5: とてもそう思う) にて評価した。

本研究での統計的解析には、R を使用した。質問項目の評定値に正規性を確認することができなかつた為 (コルモゴロフ・スミルノフ検定,  $p < 0.05$ )、仮想エージェント毎の違いに関して、クラスカル・ウォリス検定を用いた。質問項目間の関係性を見るため、スピアマンの相関係数を算出した。性別、年齢、SRS の群間毎の分析では、仮想エージェント・質問項目の全組み合わせにて効果量  $r$  を算出し、 $r$  の上位 3 個の組み合わせを報告する。画像と動画で、好ましさと親しみやすさに関して、対応なしのウィルコクソンの符号順位検定を行った。本研究では、統計的仮説検定の有意水準  $\alpha$  値を 0.05 に設定した。

## 3 結果

まず仮想エージェントによる評定の違いについて報告する。クラスカル・ウォリス検定により、好ましさと親しみやすさで仮想エージェントによる違いについて有意差を確認している ( $p < 0.05$ )。リアルさに関して、元々の設計の狙い通りに分布していることがわかる ((b) < (a) < (g) の順番となっている)。参加者の中で最も好まれた仮想エージェントは (b) となり、平均 3.29 (SD: 1.0) となった (図 4, Likable 参照)。その他の質問項目においても (b) は高く評価されていることがわかる。男性の仮想エージェント (c) および (d)、人間以外の (e) と (f) の仮想エージェントでは、好ましさが (b) と比較して低くなっていることがわかる。(h) は、好ましさが低くなり、親しみやすさも低いことがわかる。(h) と (i) に関して、画像と動画を比較すると、動画にすることにより、好ましさは有意に変化しないことがわかった ( $p > 0.05$ )。しかしながら、(h) の仮想エージェントでは親しみやすさが有意に向上していることがわかった ( $p < 0.05$ ) (図 5 参照)。

続いて質問項目間の相関関係の結果を述べる。図 6 に相関行列を示す。顔と好ましさに高い相関が見られた ( $\rho=0.78, p < 0.05$ )。また、トレーナとして受け入れられるかと、聞き手として受け入れられるかについても高い相関が見られた ( $\rho=0.80, p < 0.05$ )。一方、声の好ましさと他の質問項目に関しては、有意な差は確認しているものの、比較的低い相関係数となった ( $p < 0.05$ )。

表 1 は、性差において効果量  $r$  の高い仮想エージェント・質問項目の組み合わせ上位 3 個を載せている。同様に、表 2 は年齢の高低、3 は SRS の高低に関して、効果量  $r$  の高い組み合わせ上位 3 個を載せている。性別に関しては、(g) の顔および全体の好ましさと、トレーナとして受け入れられるかに関して、男性が女性より高い値を示した。年齢に関しては、特に (i) の目および顔の好ましさとに関して、高い年齢群の方が低い年齢群よりも高い値を示した。SRS に関しては、(g) の顔および髪の色に関して、高い SRS 値群の方が低い SRS 値群よりも高い値を示した。図 7 にキャラクター (g) における SRS 高低間の差を示す ( $p < 0.05$ )。

表 1: 性差と関係のある仮想エージェントと質問項目 ( $all, p < 0.05$ )

| 仮想エージェント | 質問項目    | r    | 傾向            |
|----------|---------|------|---------------|
| (g)      | Face    | 0.29 | Male > Female |
| (g)      | Likable | 0.25 | Male > Female |
| (g)      | Trainer | 0.25 | Male > Female |

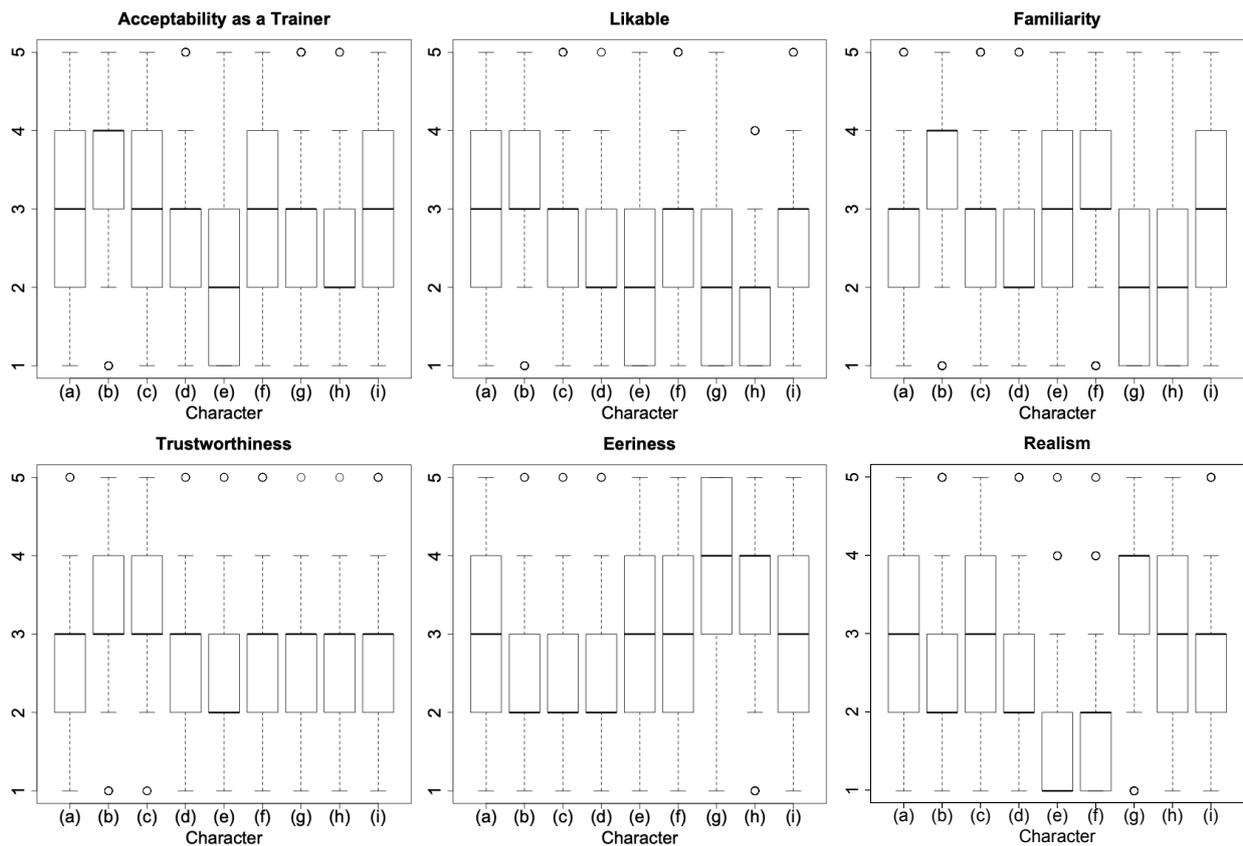


図 4: 画像に対しての評定

表 2: 年齢の高低と関係のある仮想エージェントと質問項目 ( $all, p < 0.05$ )

| 仮想エージェント | 質問項目     | r    | 傾向         |
|----------|----------|------|------------|
| (i)      | Eye      | 0.21 | High > Low |
| (i)      | Face     | 0.19 | High > Low |
| (a)      | Listener | 0.17 | High < Low |

表 3: SRS の高低と関係のある仮想エージェントと質問項目 ( $all, p < 0.05$ )

| 仮想エージェント | 質問項目 | r    | 傾向         |
|----------|------|------|------------|
| (g)      | Eye  | 0.19 | High > Low |
| (g)      | Hair | 0.18 | High > Low |
| (h)      | Face | 0.16 | High > Low |

## 4 考察

本研究では、自動 SST の設計にあたって、仮想エージェントの受け入れやすさ、好ましさ、およびその他の種々の尺度に関して調査を行った。以下のそれぞれの項目についての考察をまとめる。

### 4.1 仮想エージェントの設計

本研究において、仮想エージェントによる評価の違いを確認することができた。まず仮想エージェントの設計のリアルさに関しては (a),(b),(g) で統制できていると考えられる。全体として、(b) の仮想エージェントが好まれる結果となった。(b) は、元々10代女性のアニメらしい絵として設計されたものである。日本人がこの様なアニメ動画の視聴に慣れていることもあると考えられ、親しみやすさも高い値を示している可能性がある。一方、(g),(h) および無生物、動物などの仮想エージェントは、あまり好まれない結果となった。低年齢の自閉スペクトラム症では、電車などの仮想エージェントがより好まれる可能性もあるということから [16]、今後低年齢ではどうなるかの検討も必要である。動画に関してもほぼ同様の傾向が見られるが、親しみやすさでは、動きが入ったことにより、(h) の評価が画像に比較して改善したと考えられる。

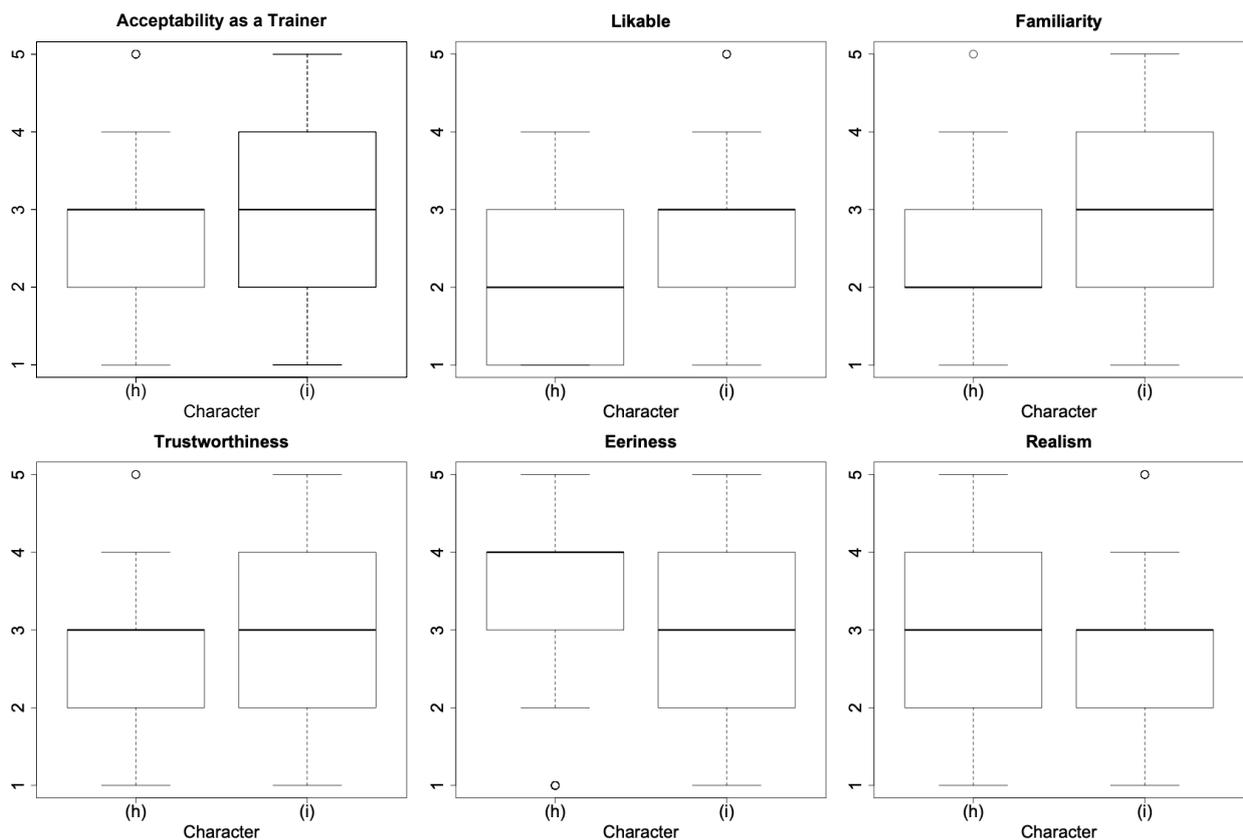


図 5: 動画に対する評価

## 4.2 好ましさと関連している質問項目

質問項目間で有意な相関が確認された ( $p < 0.05$ )。声は、仮想エージェント毎で同一のものを使用したこともあり、相関係数としては比較的低いものとなった。今回の場合は、トレーナとしての役割と聞き手としての役割で、継続したインタラクティブな対話をしていないことから差を確認することができなかったが、今後明示的に仮想エージェントの役割を与えた際の調査も必要となってくると考えられる。

## 4.3 性別、年齢、自閉症傾向による関係性

(g)の女性仮想エージェントを男性の方がより好むと評価することがわかった。また、(b)の仮想エージェントでも男性が有意に高く評価することが確認されたことから、男性の参加者では女性の仮想エージェントをより好ましいと評価することがわかった。年齢に関して、(i)の仮想エージェントを高い年齢群の方で好むことがわかった。(i)は比較的年齢が高め(元々40代をターゲットに作られていることから、年齢が高い群では、自分と近い年代で落ち着き信頼できると評価されたと考

えられる。自閉症傾向が高い群は、(g)、(h)などの仮想エージェントに対して高い評価を示していた。この点については、今後、自閉スペクトラム症の顔の要素の認識特性など調べる必要があると考える。

## 5 まとめ

本研究では、自動SSTの設計にあたって、仮想エージェントの受け入れやすさ、好ましさなど種々の尺度に関して調査を行った。仮想エージェントの画像と動画を用意し、固定されたシナリオに対して、合計912名のクラウドワーカーが評価を試みた。結果として、女性のアニメ型の仮想エージェントが最も好まれることがわかった。また参加者の性別、年齢、自閉症傾向に関して有意に差のある仮想エージェントと質問項目の関係性も調査した。これらの知見は、個人に適応した仮想エージェントの設計で重要だと考える。

今後の予定としては、リアリズムのみについて統制した際[8]の影響を調査する必要がある。また、日本人以外での文化差などを考慮する必要がある。さらに、インタラクティブな対話システムに組み込んだ際に、口

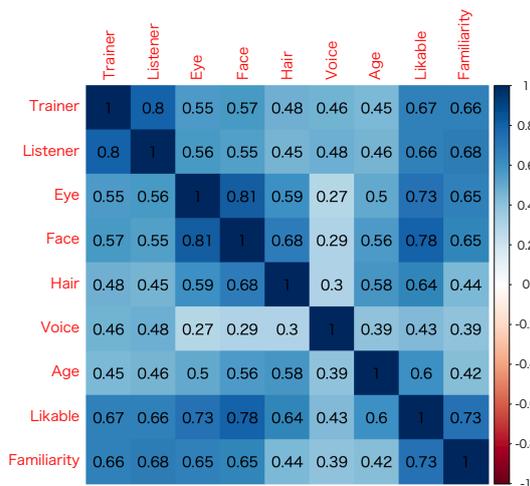


図 6: 質問項目間の相関行列

調、対話戦略が受け入れやすさに及ぼす影響についても調査する必要がある。

## 謝辞

本研究は、CREST 戦略的創造研究推進事業 (JP-MJCR19A5) および JSPS 科研費 (JP17H06101 および JP18K11437) の支援を受けて行われたものである。

## 参考文献

- [1] Alan S. Bellack, Kim T. Mueser, Susan Gingerich, and Julie Agresta: Social Skills Training for Schizophrenia, Second Edition: A Step-by-Step Guide. Guilford Publications (2013)
- [2] Hiroki Tanaka, Hideki Negoro, Hidemi Iwasaka, and Satoshi Nakamura: Embodied conversational agents for multimodal automated social skills training in people with autism spectrum disorders, Plos One, Vol.12, No.8, pp.1–15 (2017)
- [3] Hiroki Tanaka, Hidemi Iwasaka, Hideki Negoro, and Satoshi Nakamura: Analysis of Conversational Listening Skills toward Agent-based Social Skills Training, Journal on Multimodal User Interfaces, Vol. 14, 1, pp.73–82 (2020)
- [4] Hiroki Tanaka, Hidemi Iwasaka, Yasuhiro Matsuda, Kosuke Okazaki, and Satoshi Nakamura:

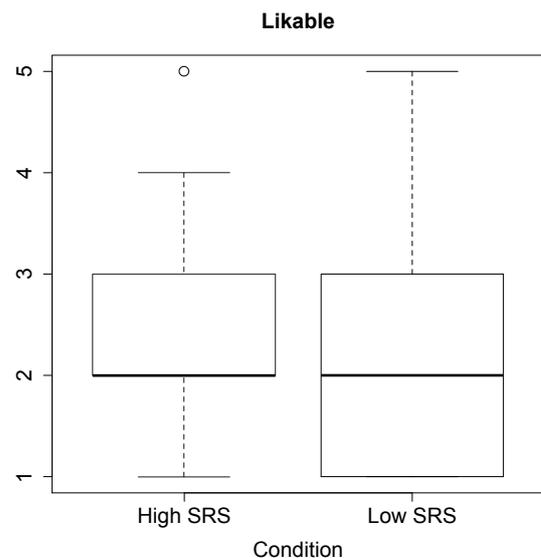


図 7: SRS の高低とキャラクター (g) の好ましさの関係

Analyzing Self-Efficacy and Summary Feedback in Automated Social Skills Training. [under review]

- [5] Anna Esposito, Terry Amorese, Marialucia Cuciniello, Antonietta M. Esposito, Alda Troncone, Maria Ines Torres, Stephan Schlogl and Gennaro Cordasco: Seniors' Acceptance of Virtual Humanoid Agents, Italian Forum of Ambient Assisted Living, pp.429–443 (2018)
- [6] Anna Esposito et al.: Elders prefer female robots with a high degree of human likeness, IEEE 23rd International Symposium on Consumer Technologies, pp.243–246 (2019)
- [7] Kazunori Terada, Liang Jing, and Seiji Yamada: Effects of Agent Appearance on Customer Buying Motivations on Online Shopping Sites, In Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems, pp.929–934 (2015)
- [8] Rachel McDonnell, Martin Breidt, and Heinrich H. Bulthoff: Render me real? investigating the effect of render style on the perception of animated virtual humans. ACM Trans. Graph. 31, 4, Article 91, pp.1–11 (2012)
- [9] Lazlo Ring, Dina Utami, and Timothy Bickmore: The Right Agent for the Job? The effects of

- agent visual appearance on task domain, International Conference on Intelligent Virtual Agents, pp.374–384 (2014)
- [10] Anna Esposito, Terry Amorese, Marialucia Cuciniello, Maria Teresa Riviello, Antonietta M. Esposito, Alda Troncone, and Gennaro Corrado: The Dependability of Voice on Elders' Acceptance of Humanoid Agents, *Interspeech*, pp.31–35 (2019)
- [11] Amy L. Baylor, and Kim Yanghee: Pedagogical Agent Design: The Impact of Agent Realism, Gender, Ethnicity, and Instructional Role, *Intelligent Tutoring Systems*, pp.592–603 (2004)
- [12] Isabella Poggi, Catherine Pelachaud, F. Rosis, Valeria Carofiglio, and Berardina Carolis: Greta. A Believable Embodied Conversational Agent, pp.3–25 (2005)
- [13] John N. Constantino, and Christian P. Gruber: Social Responsiveness Scale - Second Edition (SRS-2), WPS (2012)
- [14] R.Michael Bagby, James D.A. Parker, and Graeme J. Taylor: The twenty-item Toronto Alexithymia scale—I. Item selection and cross-validation of the factor structure, *Journal of Psychosomatic Research*, Vol. 38, 1, pp.23–32 (1994)
- [15] M. L. Bezemer, E. M. A. Blijd-Hoogewys, and M. Meek-Heekelaar: The Predictive Value of the AQ and the SRS-A in the Diagnosis of ASD in Adults in Clinical Practice, *Journal of Autism and Developmental Disorders* (2020)
- [16] Ofer Golan, Simon Baron-Cohen: Systemizing empathy: teaching adults with Asperger syndrome or high-functioning autism to recognize complex emotions using interactive multimedia. *Dev. Psychopathol.* 18(2), pp.591–617 (2006)