社団法人　電子情報通信学会
THE INSTITUTE OF ELECTRONICS,
INFORMATION AND COMMUNICATION ENGINEERS

信学技報
TECHNICAL REPORT OF IEICE.

# 仮想エージェントの心の理論の程度・ソーシャルスキルトレーニングがユーザの信頼性に与える影響

田中　宏季† 　佐賀　健志† 　岩内　厚大† 　中村　　哲†

† 奈良先端科学技術大学院大学 情報科学領域
E-mail: †hiroki-tan@is.naist.jp

**あらまし**　これまでに我々は, 仮想エージェントを用いたソーシャルスキルトレーニングシステムを構築し, その訓練効果を調査してきた。我々は新たに, 断る, 依頼する, 気持ちを伝える, 耳を傾ける課題の訓練を構築した。本研究では, 22 から 35 歳の 29 名の実験参加者により, 仮想エージェントによる見た目, 心の理論の程度を調整した動画の視聴, ソーシャルスキルトレーニングを行い, その各段階のユーザのシステムへの受容性, 信頼性などを質問項目により調査した。結果として, 仮想エージェントの心の理論の程度によりユーザの信頼性が有意に変化することを確認した。また, ソーシャルスキルトレーニング前後で, システムへの信頼性が有意に向上することを確認した。

**キーワード**　仮想エージェント, ソーシャルスキルトレーニング, 心の理論, 受容性, 信頼性

# Trustworthiness of Virtual Agents in terms of Theory of Mind Levels and Social Skills Training

Hiroki TANAKA†, Takeshi SAGA†, Kota IWAUCHI†, and Satoshi NAKAMURA†

† Division of Information Science, Nara Institute of Science and Technology
E-mail: †hiroki-tan@is.naist.jp

**Key words**　Virtual agent, social skills training, theory of mind, acceptability, trustworthiness

## 1. ま え が き

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, 医療機関などのリワークプログラムやデイケアプログラムの中で人間のトレーナ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 を模倣したシステムを開発し, 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]。システムは人間の行動のモデリング, リアルタイムの行動評価およびフィードバックを含んでいる [3]。また, システムへの受容性および信頼性が高いほど治療同盟（治療者と参加者の信頼関係）が構築され, 訓練効果の向上も見込めることから, 仮想エージェントの見た目について, ユーザの受容性, 信頼性への影響を調査した [4]。

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の機能, SST の内容が信頼性に影響を及ぼすかの調査は十分になされていない。本研究では仮想エージェントの信頼性に影響を及ぼす要因として, 心の理論に焦点を当てる [5], [6]。先行研究では, ロボットが心の理論をどの程度有しているかをユーザが知覚することで, ユーザのロボットに対する信頼性が変化することが報告されている。具体的には, ロボットの誤信念課題に対する対話シナリオの動画をユーザが視聴した際, 心の理論の程度に応じてその後のロボットの推薦をユーザが信じるかが変化することを確認している [7]。

以上の点を考慮し, 本研究で我々は新たな課題を含んだ SST システムを構築し, 仮想エージェントによる見た目, 心の理論の程度を調整した動画の視聴, 断るもしくは依頼するのシナリオにおける SST を行い, その各段階のユーザの信頼性を質問項目により調査した。結果として, 仮想エージェントの心の理論の程度によりユーザの受容性が有意に変化することを確認した。また, SST 後では, 信頼性が有意に向上することを確認した。本稿ではその詳細をまとめる。

## 2. ソーシャルスキルトレーニングシステム

我々は Greta プラットフォーム [8] を用いて SST システムを構築した [3]。システムは音声認識, 応答選択, 音声合成を有し, 表情やジェスチャーを生成することができる。また, 仮想エー

表 1　SST のシナリオ例。ここでは, 断るスキルを例にしている。ここでは, トレーナがレイ, ユーザが田中という呼び名。

| 発話者 | 応答 |
|---|---|
| トレーナ | こんにちは, 私はレイと言います。あなたの名前はなんですか？ |
| ユーザ | こんにちは, 私は田中です。 |
| トレーナ | 最近は, 元気でしたか？ |
| ユーザ | はい, 元気でした。 |
| トレーナ | そうでしたか。今日は一緒に断るスキルを学びます。人から何かをしてほしいと頼まれても, いつもそれに応えられるとは限りません。とても忙しいかもしれませんし, できそうにないと感じたり, 頼まれたことをやりたくなかったりすることもあります（途中省略）。わかりましたか？ |
| ユーザ | はい, わかりました。 |
| トレーナ | では, 実際にロールプレイを始めてみましょう。準備はいいですか？ |
| ユーザ | はい, 大丈夫です。 |
| トレーナ | すみません, つかぬことをお伺いしますが, 車はお持ちですか？ |
| ユーザ | あ, 持っていないです。 |
| トレーナ | そうなんですね。ここから最寄りの駅まで遠いので, どなたかお知り合いの方にお願いして, 私を駅まで送っていただけないでしょうか？ |
| ユーザ | ごめんなさい, 私ここら辺には知り合いがいないので, 送ってくことができないです。申し訳ないです。 |
| トレーナ | そこをなんとか, お願いできないでしょうか? |
| ユーザ | うーん, ごめんなさい, 難しいです。すみません。 |
| トレーナ | わかりました。急にお願いして申し訳なかったです。なんとかしますので, ありがとうございました。 |
| ユーザ | すみません, ありがとうございます。 |
| トレーナ | はい, 以上でロールプレイは終了です。スコアを計算しますので, 少々お待ちください。 |
| トレーナ | （フィードバック画面提示）では, フィードバックです。（フィードバック読み上げ） |

ジェントの見た目の質感や性別などを変更することが可能である。

### 2.1　課題設定

我々は精神科医師と議論し, SST の基本訓練モデル [1] を参考に課題を作成した。エージェントは初めに簡単な挨拶をしたあと, 実験参加者に対して訓練課題の重要性を説明することで動機づけを行う。シナリオ例を表 1 に示す。我々は, 断る, 耳を傾ける, 依頼する, 気持ちを伝えるの 4 課題に関してのシナリオを作成した。これはベラック方式（ステップバイステップ方式）における基本 4 課題である。

### 2.2　ロールプレイ, 評価, フィードバック

ロールプレイは 4 課題に関してそれぞれ 7 種類のバリエーションを用意した。これは, これまでに収録した精神科医師による SST データ [9] を参考にして作成した。ロールプレイ中は音声認識によりユーザの発話を認識し, 予め用意したキーワード（はい/わかりました/持っています, など）が含まれているかに応じてシステムの応答を行った。また, 音声認識結果にキーワードが直接含まれていない場合は, 事前学習済み BERT-base モデルを用いて, 文レベルで上記キーワードとのコサイン距離を計算し, 最も距離の近いキーワードと組み合わせた。対話破綻が少なくなる様, 予備検証により応答選択モジュールを改善した。

我々はロールプレイ中の動画からスコア評価器を構築 [9] し, ユーザの行動指標から精神科医師が 5 段階で評価した 7 項目: アイコンタクト, 体の向き, 表情, 声の変化, 明瞭さ, 流暢さ, 社会的妥当性をマルチモーダル特徴（Praat, Openface などを使用）およびランダムフォレストにより評価した。評価結果に応じてレーダーチャートおよび肯定のコメント, 修正のコメントを動画と共に画面に提示, およびコメントを仮想エージェントが読み上げた。レーダーチャートでは評価値を示している。また SST では肯定のコメント（正の強化の意図）を初めに提示することが重要とされている [1]。図 1 にシステムの様子を示す。

## 3.　実験方法

今回の実験では, 構築した SST システムを使用し, 以下の 3 点の調査項目を明らかにすることを目指す。

（1）　仮想エージェントの心の理論の程度がトレーナとしての受容性, 信頼性へ及ぼす影響

（2）　仮想エージェントの心の理論の程度がトレーニング後のトレーナとしての受容性, 信頼性へ及ぼす影響

（3）　SST によりトレーナとしての受容性, 信頼へ及ぼす影響

### 3.1　実験参加者

本研究は奈良先端科学技術大学院大学の倫理承認を得て行われた。本研究では, 22 から 35 歳の 29 名のデータを収集した。実験は奈良先端科学技術大学院大学にて対面で行われた。全実験参加者に書面で説明を行い, 参加の同意を得た。実験参加者から, 対人応答性尺度-2 [10], Kikuchi's scale of social skills-18 [11], 新版 STAI 状態-特性不安 [12] の質問紙を取得した。本報告では質問紙分析は対象としない。また, 本実験ではその他, 眼球運動計測, SST システムのフィードバック評価に関してもデータ取得しているが, 対象としない。本研究により取得した動画データなどは筆頭著者に連絡の上, 提供が可能である。

### 3.2　心の理論の程度

Theory of Mind（ToM）とは, 他者の感情や思考, 意図などの心的状態に対する推定, および他者は自己と異なる心的状態が存在していることを理解することを指す。先行研究において, サリーとアン課題に変更を加えたシナリオにより, エージェントの心の理論の程度において, ユーザの信頼性が低下することを確認している [7]。本研究ではそのシナリオを参考に, 3 者による対話を行い, 表 2 に示す様な, 仮想エージェントが誤信念課題を失敗する事例と成功する事例の 2 種類を用意し, 動画を収録した。仮想エージェントは, Wizard of Oz (WOZ) 法によってそのタイミング, 発話, 頷きが操作された（図 2）。これは SST 前に視聴してもらう動画として用意した。

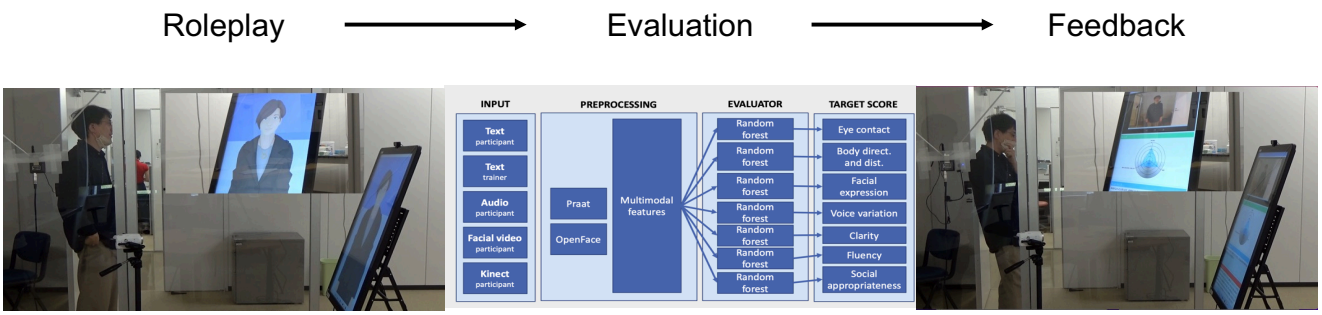


図 1　SST システムの様子。ロールプレイ, 評価（入力・処理・出力）, フィードバックから構成されている。写真上部に画面を拡大したシステムの切り抜きを載せている。また新型コロナウイルス対策としてアクリル板を設置している。

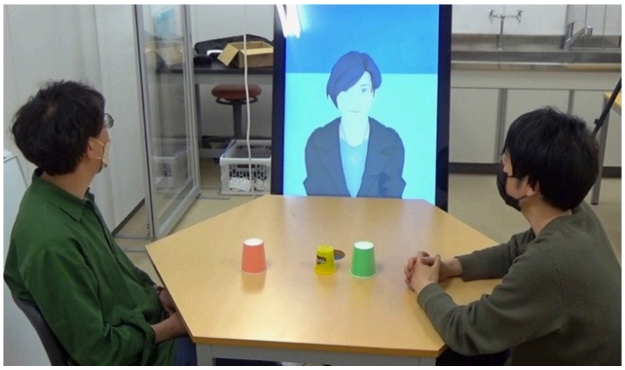

図 2　誤信念課題の動画のスナップショット。黄色のバケツ, 赤色のコップと緑色のコップを用意している。

### 3.3　質 問 項 目

先行研究 [4] に基づいて, 受容性, 信頼性, 親しみやすさ, 好ましさの主観評価を取得した。仮想エージェントに対し, 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。親しみやすさと好ましさに関しては本研究の対象から除外した。

### 3.4　実験の流れ

実験参加者を, 心の理論の事例および SST 課題に応じて人数が同比率になる様に, ランダムに群分けした。まず, 初めにユーザに SST システムと対面してもらい, 仮想エージェントが「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」と発話するのを視聴する。その後, 質問項目に回答する（これを Appearance 段階とする）。次に心の理論の動画を視聴する。これは群毎に異なる動画（High-level ToM もしくは Low-level ToM）を視聴する。その後, 質問項目に回答する（これを ToM Video 段階とする）。次に SST システムを使用し, 群毎に, 断る, あるいは依頼するの課題を実施する。最後に質問項目に回答する（これを SST 段階とする）。なお, SST の 2 課題間では質問項目に有意な差はなかった（t 検定, $p > 0.05$）。

上記 3 段階における群間およびトレーニング前後で統計的に比較を行った。質問回答の等分散性と正規性（コルモゴロフ・スミルノフ検定, $p > 0.05$）を確認した後, 心の理論の程度に関して ToM Video 段階での群間比較を, 対応のない t 検定により

表 2　仮想エージェントが誤信念課題を誤る事例。ここでは, トレーナがレイ, 対話者 A が田中, 対話者 B が岩内という呼び名。

| 発話者 | 応答 |
|---|---|
| トレーナ | はじめまして。田中さん, 岩内さん, こんにちは。 |
| 対話者 A | はじめまして, レイさん, こんにちは。 |
| 対話者 B | こんにちは。 |
| 対話者 B | 今からこの黄色いバケツを緑のコップの中に入れたいと思います。 |
| 対話者 A | はい, わかりました。 |
| トレーナ | なるほど, わかりました。 |
| 対話者 B | すみません, 用事を思い出したので一旦この部屋から出させていただきたます。 |
| 対話者 A | はい, わかりました。じゃあまた後で。 |
| トレーナ | はい, それでは後ほど。 |
| 対話者 B | （部屋から退出） |
| 対話者 A | それでは, 岩内さんがいなくなったので, このバケツを緑のコップから赤いコップに移動したいと思います。 |
| トレーナ | なるほど, わかりました。 |
| 対話者 A | ところで, レイさん, 岩内さんが帰ってきたときに, どちらのコップの方を探すと思いますか？ |
| トレーナ | えーと, そうですね。赤のコップの中にバケツがあるので, 岩内さんが探すのは赤のコップの中だと思います。 |
| 対話者 A | なるほど, わかりました。 |
| 対話者 B | （部屋に戻ってくる）すみません, お待たせしました。 |
| 対話者 B | （緑のコップの中を見る）あれ, バケツがなくなってる。 |

行った（調査項目 1）。心の理論の程度がトレーニング後の評価に影響を及ぼすかについて, Appearance 段階と SST 段階で, 分散分析による交互作用の有無を調査した（調査項目 2）。また, Appearance 段階と SST 段階で, 対応のある片側 t 検定を行った（調査項目 3）。なお有意水準は全て 5% としている。

## 4.　実 験 結 果

図 3 に受容性, 図 4 に信頼性の各段階における質問項目の棒グラフを示している。結果として, 心の理論の程度がトレー

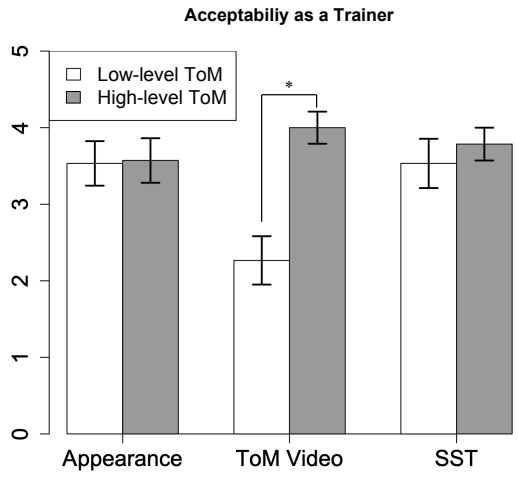


図 3　受容性に関しての質問項目の棒グラフ。エラーバーは標準誤差。

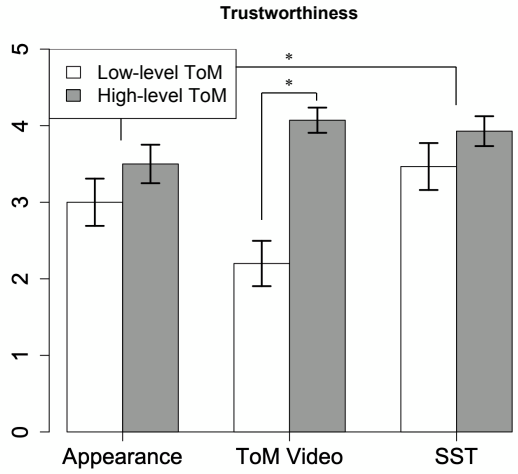


図 4　信頼性に関しての質問項目の棒グラフ。エラーバーは標準誤差。

ナとしての受容性, 信頼性へ有意に影響していることがわかる（$p < 0.05$）。心の理論の程度がトレーニング後の受容性, 信頼性へ及ぼす影響は分散分析により交互作用が認められなかった（$p > 0.05$）。また SST により, 信頼性が見た目よりも向上することを確認したが（$p < 0.05$）, 受容性では確認されなかった（$p > 0.05$）。これにより, 調査項目 1 および 3 の信頼性には有意差があることを確認したが, 調査項目 2 の心の理論の程度がトレーニング後の信頼性に及ぼす影響はないとされた。

## 5. ま と め

今回の実験では, 新たに構築した SST システムを使用し, 以下の 3 点の調査項目を明らかにすることを目指した。1) ToM の程度がトレーナとしての受容性, 信頼性へ及ぼす影響, 2) ToM の程度がトレーニング後のトレーナとしての受容性, 信頼性へ及ぼす影響, 3) SST により受容性, 信頼性が向上するかについて検証した。実験結果として, 調査項目 1 および 3 の信頼性は確認されたが, 調査項目 2 である心の理論の程度がトレーニング後の信頼性に及ぼす影響がないとされた。これは SST と心の理論のビデオが実験参加者にとっては独立したものと知覚された可能性がある。SST を行なっている際の仮想エージェントに心の理論の機能を組み込むなど [6], 心の理論の程度が SST に及ぼす影響を調べる直接的な実験設計の工夫が必要だと考える。今後は, SST システムの各モジュールの改良, また事前にゲームをするなど, その他の信頼性を高め関係づくりをしていくための SST 設計が必要だと考える。また, 今後は SST に心の理論の訓練を組み込む視点も重要だと考える [13]。

**謝辞**　本研究は CREST (グラント番号: JPMJCR19A5), の支援によって行われた。また, 本研究は JSPS 科研費 22K12151 の助成を受けたものです。

## 文　献

[1] A.S. Bellack, K.T. Mueser, S. Gingerich, and J. Agresta, Social skills training for schizophrenia: A step-by-step guide, Guilford Publications, 2013.

[2] H. Tanaka, H. Negoro, H. Iwasaka, and S. 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, 08 2017. https://doi.org/10.1371/journal.pone.0182151

[3] H. Tanaka, H. Iwasaka, Y. Matsuda, K. Okazaki, and S. Nakamura, “Analyzing self-efficacy and summary feedback in automated social skills training,” IEEE Open Journal of Engineering in Medicine and Biology, vol.2, pp.65–70, 2021.

[4] H. Tanaka, S. Nakamura, et al., “The acceptability of virtual characters as social skills trainers: Usability study,” JMIR human factors, vol.9, no.1, p.e35358, 2022.

[5] S. Baron-Cohen, “Theory of mind and autism: A review,” International review of research in mental retardation, vol.23, pp.169–184, 2000.

[6] N. Rabinowitz, F. Perbet, F. Song, C. Zhang, S.A. Eslami, and M. Botvinick, “Machine theory of mind,” International conference on machine learningPMLR, pp.4218–4227 2018.

[7] W. Mou, M. Ruocco, D. Zanatto, and A. Cangelosi, “When would you trust a robot? a study on trust and theory of mind in human-robot interactions,” 2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)IEEE, pp.956–962 2020.

[8] R. Niewiadomski, E. Bevacqua, M. Mancini, and C. Pelachaud, “Greta: an interactive expressive eca system,” Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems-Volume 2Citeseer, pp.1399–1400 2009.

[9] T. Saga, H. Tanaka, Y. Matuda, T. Morimoto, M. Uratani, K. Okazaki, Y. Fujimoto, and S. Nakamura, “Analysis of feedback contents and estimation of subjective scores in social skills training,” 2022 44th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), pp.••–••, 2022.

[10] J.N. Constantino and C.P. Gruber, Social responsiveness scale: SRS-2, Western psychological services Torrance, CA, 2012.

[11] K. Kikuchi, “Shakaiteki-sukiru-wo-hakaru-kiss-18 handbook,” The social skills are measured, The handbook of Kiss-18, Tokyo, Japan, Kawashima, pp.••–••, 2007.

[12] C.D. Spielberger, “State-trait anxiety inventory for adults,” ••, pp.••–••, 1983.

[13] S.A. Nijman, W. Veling, K. Greaves-Lord, M. Vos, C.E.R. Zandee, M.A. hetRot, C.N.W. Geraets, and G.H.M. Pijnenborg, “Dynamic interactive social cognition training in virtual reality (discovr) for people with a psychotic disorder: single-group feasibility and acceptability study,” JMIR mental health, vol.7, no.8, p.e17808, 2020.