社会的ロボティクスにおける謝罪と信頼修復の要因分析

Investigating Apologies and Trust Repair for Social Robotics

尹 寧得 [1*]　山田 誠二 [2,1]
Nungduk Yun[1]　Seiji Yamada[2,1]

[1] 総合研究大学院大学
[1]The Graduate University for Advanced Studies(SOKENDAI)
[2] 国立情報学研究所
[2]National Institute of Informatics

Abstract: 近年、飲食店での配膳や接客など、サービスロボットが身近な存在となっている。しかし、機械に不具合が起きたあとは、それにより信頼性や信用が損なわれる可能性がある。人間と人間の場合、過ちを犯すと謝罪し許されることが一般的だ。人間とロボットの場合には人間ー人間と同じく許されるのか検証を行った。本研究では、非擬人化型の多関節マニピュレータを用い、サービスロボットが誤作動を起こし、人間だけ謝罪した際の反応と、人間とロボットの謝罪の違いによる信頼修復への影響を検証した。

1　はじめに

近年、飲食店でのサービスロボットによる料理の配膳が一般的に見るようになった。しかし、機械やロボットは必ずしも安定しておらず、エラーが発生すると信頼を失うことがあります。例えば、ロボットアームで物を運ぶ際に落とすなどのエラーが発生すると信頼を失う可能性がある。ロボットは効率性と一貫性を提供しますが、完璧ではありません。エラーが発生すると、ユーザーの信頼を失う可能性があり、ロボットサービスのより広範な普及を妨げる可能性があります。機械のエラーに対する人間による謝罪が、ロボットサービスへの信頼を維持できるかどうかは疑問です。人間と人間のインタラクションでは間違いなどがあれば謝れば治ることもある。しかし、人間とロボットの場合は同じくできるのか？特に非人型ロボットの場合、同様のメカニズムが人間とロボットの間において効果的かどうかは不明確です。

人間と人間のインタラクションでは間違いがあれば謝罪により修復できることがありますが、人間とロボットの場合も同様にできるのでしょうか。擬人化が高いロボットの謝罪については研究が行われていますが、ロボットアームのような非擬人化ロボットが human behavior で誤りを起こした場合、信頼を回復できるのかは不明です。現状では、ロボットがミスをした際に人間が謝罪する場合が多く見られます。

*連絡先：総合研究大学院大学
東京都千代田区
E-mail:ndyun@nii.ac.jp

そこで、本研究では、ロボットアームを使用し、信頼修復の効果を 2 つの実験で検証する。Study 1 ではロボット自身による言語的（verbal）、非言語的（non-verbal）、その両方の同時実行による謝罪の効果を検証する。これはアジアの文化における人間同士のインタラクションでは、お辞儀と謝罪を同時に行うことが一般的であることに基づいている。Study 2 では、人間とロボットの誤りと人間だけの誤りの比較を行う。これは現状でロボットのミスに対して人間が謝罪することが多いことから、その効果を検証する。

意図的にエラー（物の落下など）を発生させ、その後の信頼修復戦略を実施することで、実世界のシナリオをシミュレーションします。このアプローチにより、人間とロボットの信頼関係の修復力と、様々な修復戦略の有効性を検証します。この研究は、非擬人化ロボットの信頼修復によるサービスロボットの社会実装における実践的な知見を提供することを目指しています。

2　関連研究

Human-robot interaction において、謝罪はエラーの悪影響を軽減し信頼を維持する上で重要な役割を果たす。研究によると、ロボットの謝罪は信頼の侵害に対処し、失敗後のロボットに対する見方を改善する上で効果的であることが示されている [17, 14]。ロボットの謝罪における言語的内容、原因帰属、誠実性の認知などの様々な側面が研究されている [18]。謝罪に伴う負担やコストが重要であり、より大きな負担を伴う謝罪の

方が誠実で効果的だと認識されることが明らかになっている [13]。Okada らは複数のロボットによる謝罪のシナリオを探究し、特定の状況では 2 台のロボットからの謝罪が 1 台からの謝罪よりも効果的である可能性を見出した [13]。彼らの研究では、参加者は 2 台のロボットからの謝罪を 1 台からの謝罪よりも、許し、否定的な口コミ、信頼、利用意図の面で有意に好ましく評価した。他の研究では、言語的な謝罪に加えてミスの後の清掃など、謝罪するロボットの異なる役割も検討されている [13]。失敗の種類や修復戦略などの要因も謝罪の効果に影響を与えることが示されている [14]。信頼違反後の信頼再構築におけるロボットの謝罪、説明、信頼開示の効果が研究されている [3, 7]。ロボットの謝罪は信頼修復に役立つことが示されているが、その効果は謝罪のデザイン、失敗の文脈、ユーザーの認識に関連する様々な要因に依存する。

2.1 HRI における信頼修復

人間とロボットのインタラクションに関する先行研究では、多くの研究者が信頼修復戦略を設計している。説明による方法は信頼の向上につながらないことが示されている [6]。Zhang らは、HRI における異なる種類の技術的失敗（論理的、意味的、構文的失敗）とそれらの人間の信頼への影響を検証し、さらに複数の信頼修復戦略（内部帰属の謝罪、外部帰属の謝罪、否定、修復なし）の有効性を比較している [18]。Chang らは、意図しないロボットの行動（拒絶と認識される行動）がサービスロボットへの人間の信頼に与える否定的な影響を研究し、拒絶と認識された後の信頼修復のための非言語的な改善戦略（好意的行動と謝罪）を検証している [4]。このことから、本研究では信頼修復戦略として、言語的、非言語的、両方（言語的および非言語的）による謝罪、そして人間による謝罪を選択した。Nayyar らは、ロボットの謝罪のタイミングが信頼修復に影響を与えることを示している [12]。Kraus らは、異なる謝罪戦略がロボットのエラー後の信頼修復にどのように影響し、それが人格化傾向などの個人差とどのように相互作用するかを研究している [9]。また、自動システムからの約束と謝罪は、複数の相互作用にわたって信頼修復に影響を与えることが示されている。この長期的な視点は、人間とロボットのインタラクションにおけるロボットのミスと修復戦略に関する実世界の文脈を理解する上で重要である [2][1]。Rogers らは、高リスクシナリオにおけるロボットの繰り返される欺瞞後の信頼修復戦略を研究しており、これは単なるエラーではなく意図的な違反後の信頼修復に関連している [14]。Kox らは、一人称シューティングタスクを用いた人間とエージェントのチームにおける、説明と後悔の表現という異なる信頼修復戦略の有効性を検証している [8]。

図 1: Mycobot280 M5Stack 版

3 実験方法

研究 1 と 2 では同じ質問紙による測定を用いたが、最後の動画で異なる信頼修復戦略を使用した。研究 1 では verbal と non-verbal による謝罪、研究 2 では人間による謝罪を検証した。G*Power によるサンプルサイズ計算（効果量 0.5、研究 1：n=159、研究 2：n=128）に基づき、動画視聴後のオンラインアンケート調査を実施した。参加者は Yahoo!クラウドソーシングから募集した。先行研究により、ライブとビデオベースの HRI 実験は概ね同等であることが示されている [16]。ただし、ビデオベースの HRI 実験は対面実験と比較して共感が低くなる場合がある [10, 15]。

3.1 ロボットプラットフォーム

本研究では、図 1 に示す Elephant Robotic 社の Mycobot280 M5Stack 版を使用し、研究 1 と 2 用に独自のエンドエフェクタを設計した。

3.2 測定方法

信頼度の測定には、Malle らによるマルチディメンショナル信頼尺度第 2 版（MDMT v2）のうち、パフォーマンス信頼（信頼性、有能性）と倫理的信頼（倫理性）を使用した [11]。回答は 7 段階リッカート尺度（1：全く同意しない〜7：強く同意する）で評価した。

3.3 実験手順

参加者は 3 つの動画を視聴した。1 つ目は図 2 に示すロボットによる 2 本のボトルのピック＆プレース作業、2 つ目は図 3 に示すボトルの意図的な落下、3 つ目は研究 1 では図 4 と 5 に示す 3 条件（verbal、non-verbal、

図 2: 研究 1 と 2：ボトルを pick and place する動作と初期信頼値の設定

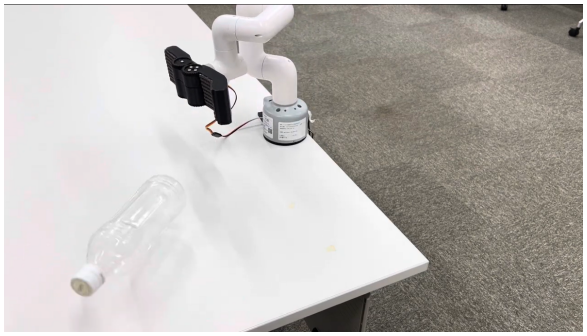

図 3: 研究 1 と 2： 2 番目のビデオについては、意図的なボトル落下.

図 4: 研究 2 信頼修復戦略: verbal 条件（すみませんという）

図 5: 研究 1 信頼修復戦略： non-verbal 条件と both(verbal と non-verbal) 条件

両方）の信頼修復戦略、研究 2 では図 6 と 7 に示す 2 条件（人間の謝罪、人間とロボットの謝罪）を視聴した。各動画視聴後に質問紙による評価を行った。実験時間は 15〜30 分程度で、報酬として 100 円を支払った。

4 研究 1：謝罪による信頼修復戦略：言語的、非言語的、両方

テキストと音声による説明方法に有意差が見られないこと [6]、謝罪を含む言語的内容 [18]、および非言語的要素 [4] に基づき、言語的（verbal）、非言語的（non-verbal）、両方（verbal and non-verbal）の条件を設定した。これらの組み合わせが信頼修復に与える影響を検証する。本研究では以下の仮説を立てた：

- **H1** 両方の条件で信頼修復が最も高くなる

4.1 実験デザイン

1 要因 3 水準（信頼修復戦略：verbal、non-verbal、both）の参加者間計画による一元配置分散分析を実施した。verbal 条件では、ロボットは動かず、音声生成にはテキスト読み上げ Web サービス[1]を使用した。

[1]https://ondoku3.com/ja/

4.2 研究 1 の参加者

G*Power 計算 [5] に基づき、各条件 53 名、計 159 名のサンプルサイズを設定した。Yahoo!クラウドソーシングで 299 名（男性 237 名、女性 61 名）が参加し、各条件上位 53 名のデータを分析に使用した。年齢は 21-81 歳（M=49.25、SD=10.44）。実験は本機関の倫理委員会の承認を得た。

4.3 研究 1 の結果

分散分析の結果、戦略の主効果が有意であった（$F(2,156)=7.35$、$p < .001$、$\eta^2 p = 0.086$）。事後検定より、verbal 戦略は non-verbal 戦略（平均差=-0.7975、p=.004、d=-0.6310）および両方の戦略（平均差=-0.8314、p=.003、d=-0.6578）と比較して有意に低かった。non-verbal と両方の戦略間に有意差はなかった（平均差=-0.0340、p=.990）。平均信頼度は、両方の戦略（M=3.56）、non-verbal 戦略（M=3.53）、verbal 戦略（M=2.73）の順となった。

図 6: 研究 2 信頼修復戦略： 人間だけの誤り条件.

図 7: 研究 2 信頼修復戦略: 人間とロボットの誤り条件.

5 研究 2：信頼修復戦略 - 人間のみの謝罪 vs 人間とロボットの謝罪

人間とロボットが共に謝罪する効果を検証した先行研究は見当たらない。そこで以下の仮説を立てた：

- **H2** 人間とロボットが共に謝罪する場合に信頼修復が最も高くなる

5.1 研究 2 の参加者と方法

G*Power 計算により各条件 64 名、計 128 名のサンプルサイズを設定。199 名（男性 138 名、女性 61 名）が参加し、各条件上位 64 名を分析。年齢は 22-75 歳（M=50.18、SD=11.40）。

5.2 研究 2 の結果

独立サンプルの t 検定の結果、人間のみの謝罪群と人間とロボットの謝罪群の間に有意差が見られた（t(126)=-4.17、p¡.001、d=-0.736）。人間とロボットが共に謝罪する条件で信頼度が高く、H2 が支持された

6 考察

仮説 H1「verbal と non-verbal の両方の条件で信頼修復が最も高くなる」は部分的に支持された。verbal 条件と比較して、non-verbal および両方の条件で信頼が有意に高くなった。ただし、non-verbal と両方の条件間には有意差が見られなかった。仮説 H2「人間とロボットが共に謝罪する場合に信頼修復が最も高くなる」は支持された。人間のみの謝罪と比較して、人間とロボットの共同謝罪で信頼が有意に高くなった。これらの結果から、非人型ロボットにおいても非言語的要素を含む謝罪戦略が効果的であることが示された。また、サービスロボットのエラーに対しては、人間単独の謝罪よりも人間とロボットの共同謝罪が信頼修復に有効であることが明らかになった。エラー発生後、信頼は大きく低下したものの、適切な謝罪戦略により部分的な修復が可能であることが示された。この知見は、サービスロボットの設計やヒューマン・ロボット・インタラクションにおける信頼修復戦略の構築に貢献すると考えられる。

7 結論

本研究は、非人型サービスロボットの謝罪戦略設計とその信頼修復への影響について調査した。結果から、non-verbal な謝罪や人間とロボットの協調的な謝罪の有効性が示された。これらの知見は、エラーや誤解が生じた際でも人間との良好な関係を維持できるロボットシステムの開発を行う。

参考文献

[1] Yusuf Albayram, Theodore Jensen, Mohammad Maifi Hasan Khan, Md Abdullah Al Fahim, Ross Buck, and Emil Coman. Investigating the effects of (empty) promises on human-automation interaction and trust repair. In *Proceedings of the 8th International Conference on Human-Agent Interaction*, HAI '20, pp. 6–14, New York, NY, USA, November 2020. ACM.

[2] Minja Axelsson, Micol Spitale, and Hatice Gunes. "oh, sorry, I think I interrupted you": Designing repair strategies for robotic longitudinal well-being coaching. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 13–22, New York, NY, USA, March 2024. ACM.

[3] Anthony L Baker, Elizabeth K Phillips, Daniel Ullman, and Joseph R Keebler. Toward an understanding of trust repair in human-robot interaction: Current research and future directions. *ACM Trans. Interact. Intell. Syst.*, Vol. 8, No. 4, pp. 1–30, December 2018.

[4] Xiaoyu Chang, Yanheng Li, Sijia Liu, Ling Ma, and Ray Lc. "sorry to keep you waiting": Recovering from negative consequences resulting from service robot unintended rejection. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, Vol. 10, pp. 96–105, New York, NY, USA, March 2024. ACM.

[5] Edgar Erdfelder, Franz FAul, Axel Buchner, and Albert Georg Lang. Statistical power analyses using G*Power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, Vol. 41, No. 4, pp. 1149–1160, 2009.

[6] Kasper Hald, Katharina Weitz, Elisabeth André, and Matthias Rehm. " an error occurred! " - trust repair with virtual robot using levels of mistake explanation. In *Proceedings of the 9th International Conference on Human-Agent Interaction*, HAI '21, pp. 218–226, New York, NY, USA, November 2021. ACM.

[7] Yngve Kelch, Annette Kluge, and Laura Kunold. Would you trust a robot that distrusts you? In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, New York, NY, USA, March 2024. ACM.

[8] E S Kox, J H Kerstholt, T F Hueting, and P W de Vries. Trust repair in human-agent teams: The effectiveness of explanations and expressing regret. In *Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems*, AAMAS '22, pp. 1944–1946, Richland, SC, May 2022. International Foundation for Autonomous Agents and Multiagent Systems.

[9] Johannes Maria Kraus, Julia Merger, Felix Gröner, and Jessica Pätz. 'sorry' says the robot: The tendency to anthropomorphize and technology affinity affect trust in repair strategies after error. In *Companion of the 2023 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '23, pp. 436–441, New York, NY, USA, March 2023. ACM.

[10] Sonya S Kwak, Yunkyung Kim, Eunho Kim, Christine Shin, and Kwangsu Cho. What makes people empathize with an emotional robot?: The impact of agency and physical embodiment on human empathy for a robot. In *2013 IEEE RO-MAN*, pp. 180–185, August 2013.

[11] Bertram F Malle and Daniel Ullman. A multidimensional conception and measure of human-robot trust. In *Trust in Human-Robot Interaction*, pp. 3–25. Elsevier, 2021.

[12] Mollik Nayyar and Alan R Wagner. When should a robot apologize? understanding how timing affects human-robot trust repair. In *Social Robotics*, Lecture notes in computer science, pp. 265–274. Springer International Publishing, Cham, 2018.

[13] Yuka Okada, Mitsuhiko Kimoto, Takamasa Iio, Katsunori Shimohara, and Masahiro Shiomi. Two is better than one: Apologies from two robots are preferred. *PLoS One*, Vol. 18, No. 2, p. e0281604, February 2023.

[14] Kantwon Rogers, Reiden John Allen Webber, Jinhee Chang, Geronimo Gorostiaga Zubizarreta, and Ayanna Howard. Lie, repent, repeat: Exploring apologies after repeated robot deception. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, pp. 602–610, New York, NY, USA, March 2024. ACM.

[15] Stela H Seo, Denise Geiskkovitch, Masayuki Nakane, Corey King, and James E Young. Poor thing! would you feel sorry for a simulated robot? a comparison of empathy toward a physical and a simulated robot. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction*, HRI '15, pp. 125–132, New York, NY, USA, March 2015. Association for Computing Machinery.

[16] Sarah N. Woods, Michael L. Walters, Kheng Lee Koay, and Kerstin Dautenhahn. Methodological issues in HRI: A comparison of live and video-based methods in robot to human approach direction trials. *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, pp. 51–58, 2006.

[17] Jin Xu and Ayanna Howard. Evaluating the impact of emotional apology on human-robot trust. In *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pp. 1655–1661. IEEE, August 2022.

[18] Xinyi Zhang, Sun Kyong Lee, Whani Kim, and Sowon Hahn. "sorry, it was my fault": Repairing trust in human-robot interactions. *Int. J. Hum. Comput. Stud.*, Vol. 175, No. 103031, p. 103031, July 2023.