

対話型 AI を搭載した小型ロボットとの会話失敗時の修復に みられる人の身体動作の特徴分析

Analysis of Human Body Movements Observed During the Repair of Conversation Failures in Small Robots Equipped with Conversational AI

田中 杏樹¹ 駒崎 俊剛¹

Tanaka Anjyu¹, Toshitake Komazaki¹

¹ 東京医療保健大学

¹Tokyo Health Care University

Abstract: This study analyzes the characteristics of human body movements during the repair of conversation failures in small robots equipped with conversational AI, focusing on speech content, gaze, facial orientation, and upper body posture.

1 はじめに

本稿では、タスクが明確に定まっていない対話機能を持ったロボットの会話中の失敗に対して、人どのように会話の修復に関与しているのか、また、会話の修復プロセスでどのような身体動作が見られるのか、これらを検討する。

2 関連研究

2.1 ロボットの失敗と受容度

医療や介護などの分野で対話機能を持つロボットを導入する際、より利用者に受け入れてもらうにはどのようなことを検討すればよいのだろうか。

まず、野村 (2009) は、人間にはロボットとの会話のような新規のコミュニケーション技術に対して過度に肯定的もしくは否定的な態度を持つ傾向があると指摘している[1]。

次に、ロボットと人の会話において人間の受容度を高めるには、グレイスによる「会話の4つの格率」[2]にそのような会話ができることが重要である。例えば、駅での案内ロボットと人との会話の場面を想定する。ロボットが切符の買い方の案内を開始し、その途中でユーザーが別の説明項目を指示しても、ロボットが現在読み上げている内容が終わらない限り、次の説明に移らなかったとする。これは、聞き手である人からすれば、必要以上の情報量を伝えようとしているにみえる。この場合は、会話の格率の「量」の格率から外れることになる。そして、ロボットが「失敗」しているとユーザーは捉え、その結

果、受容度が低くなるかもしれない。しかし、会話の格率から外れることもあるロボットの発話が、受容度を高める可能性も指摘されている。

谷郷ら(2016)は、商品紹介を担当するロボットがスライドの送り間違えや言葉の詰まりなどの不完全な振る舞いをした場合、完璧な振る舞いをするロボットよりも、「人間らしさ」や「親しみやすさ」の評価が高くなること述べている[3]。また、大谷 (2021) は、ロボットが砕けた話し方で商品説明する際、「親しみやすさ」については失敗頻度との間に有意な関連は認められなかったが、失敗がない場合と比較して、1 分間に 1 回失敗する場合の方がユーモアを感じさせる。また、ロボットに否定的な態度を持つ人はロボットの失敗行動によって親しみやすさの向上を図りやすいことを示唆した[4]。加えて、高橋 (2023) によれば、製品紹介ロボットが失敗することにより、製品紹介前に製品への興味が低い人たちは製品への興味が向上することが明らかになった[5]。これらの研究では、タスクが明確である商品説明という制度的会話時に利用されるロボットにおいては、会話中に意図的に失敗を組み込むことにより、ロボットの受容度を高め、説明している内容への興味を高める可能性を示唆している。

2.2 ロボットの発話に対する「修正」

前項では、タスクが明確である制度的会話時におけるロボットの発話の失敗が、人のロボットに対する受容度を高める可能性があることについて述べた。

これに対して、商品説明と比べて相対的にタスクが明確ではない雑談での失敗は、ロボットの受容度へ

どのような影響を与えるのだろうか。そこで、雑談中に発話を失敗するロボットを用いて、この失敗に対して、人がどのような発話や身体動作で応じているのか、ここに焦点をあてる。

本実験で用いたロボットは、次の機能を持っている。それは、(1)音声合成による発話機能、(2)発話に連動して形が変わる「口」を表示させる機能、(3)音声認識機能というものである[6]。一般的に人同士の会話であれば、音声情報以外にも、視線や身振り、他にも身体全体・局所の動き、お互いの立ち位置などを利用することができる。しかし、このような身体行為に伴う情報は、今回用いたロボットでは、先に示した機能の制約上利用することができない。一方、人はロボットの発話に対して、音声に加えて視線や身振りなどを利用できる。

このような状況下で、Schegloff ら[7]が示した修復の4つのタイプうちどのような修復の仕方が見られるのか、そのプロセスにおいて人はどのような身体動作を示すのか、検討する。

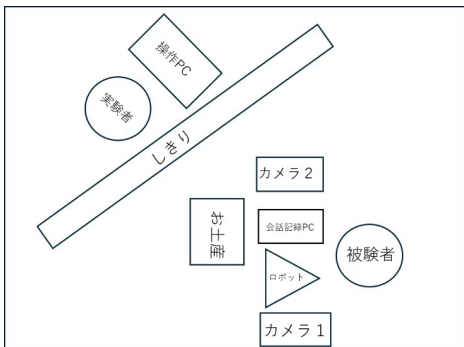
3 目的

本稿では、ロボットと人との相互行為において利用できる資源に制約のある状況下で、ロボットの発話を修復する際に見られる人の身体動作の特徴を検討する。

4 方法

4.1 実験機材

被験者と実施者の間にはしきりを置き、被験者の両端にはロボットとの会話時の様子を撮影するカメラを配置した。利用したロボットは、M 5Stack[6]とこれに連携するノート PC である。M 5Stack は、ディスプレイに「口」の動きを表示、マイクやスピーカー、そして対話型 AI との通信を担う。ノート PC は、音声解析「VOSK」、音声合成「voicevox_engine」を担う。これらの機材と LLM との連携は、アールティ社の設定手順に従い設定した。



4.2 実験手順

被験者は、20 歳代の大学生であり、研究内容を説明し、実験参加の同意を得た 5 名である。まず、被験者には、ロボットとの会話の前に、野村[8]と大谷ら[4]の評価尺度を用いた事前質問紙によりロボットの受容度を測定した。その後、被験者は 4.1 で示したロボットと 15 程度の会話をする。会話の途中でロボット側の発話に失敗としての言い間違いを生じさせる。会話終了後、事前質問紙と同じ項目の事後質問紙によりロボットに対する受容度を測定した。最後に、被験者はロボットとの会話風景を撮影した動画を見ながら、テンショングラフを記入した。このテンショングラフを見ながら、被験者に対して筆者はインタビューを行った。

なお、本研究の実施にあたり東京医療保健大学ヒトに関する研究倫理委員会の承認を得ている。（承認番号: S24-02B）

5 結果と考察

実験は、2024 年 12 月から 1 月にかけて 5 人が実験に参加した。それぞれの実験時間は、同意取得・インタビューも含めて各回 40 分前後であった。本稿では、このうち 2 回の会話を検討する。抜粋では会話に加えて身体動作も記述するために、トランスクリプト記号は坂井田ら[9]の記法を参考にした。

5.1 被験者 1（C さん）の身体動作

抜粋 1 での 01 行目の発話の前にロボットが、ロボットの名前を発話している。その後、01 行目でロボットから「あなたは」と質問されている。ただし、「あなたの名前は」のように明確に問われていないため C がすぐに答えられず 3 秒の無音区間がある。03 行目で、C の名前を答える。その後、ロボットから応答がないため 05 行目と 06 行目でロボットから視線を外す、向ける、を繰り返している。07 行目の最後に首を傾げる。08 行目でロボットが C の名前の 1 文字目「し」を発話し終わる直前にロボットへ視線を向ける。C の名前の 2 文字を間違えて発話している。ここで、ロボットが発話を失敗して、トラブル源となった。C はロボットの間違いに気が付いているが、ロボットが 3 文字目の発話を終了するまで、ロボットへ視線を向けている。ロボットが発話を終了すると視線を外すが、すぐに視線を向ける。13 行目でロボットはもう一度「しいた」と発話する。C はロボットを見ているが、訂正をせずに待っている。まだ修復されていないので、C は 16 行目でロボットの発話を訂正している。その後、ロボットは 18 行目で「しゅうた」と発話し、直後に C は訂正が適切で

あることをうなずきながら 20 行目で発語している。ここで修復が終わったと考えられる。

この抜粋の中で、ロボットにより自己修復・他者開始のタイプの修復が見られる。また、このプロセスで見られるロボットへ適切に情報が伝わっていないこと示す人の身体動作として、(1)視線を外す動作(08 行目)、(2)動作がない長い無音部分(09 行目、15 行目)が挙げられる。これらの動作をロボットが検出できれば、修復の手掛かりになると考えられる。同様に訂正が終わったことの動作として「うなずく」ことが、検出できれば「そう」という C の発語とあわせて、修復の終了の手掛かりになると考えられる。

抜粋 1

会話開始から 41 秒後

(.) は 0.2 秒以下の短い無音区間、() 内は無音時間の秒数、R はロボットの発話、C は被験者の発話

c は被験者の発話以外の身体動作を示す

「+」は被験者の身体動作の開始点

「->」は同一の身体動作の継続を示す

- 01 R: あなたは?
c: +ロボットを見つめる
- 02 (2.5)
- 03 C: じぶんは かわさきしゅうた
- 04 (3)
- 05 c: +視線を外す (1) +視線を向ける
- 06 c: (1.5) +視線を R から外す (0.5) +視線を R へ向ける
- 07 c: (1.6) +視線を R 外す (0.4) +首傾げる
- 08 R: し(.) い(.) た?
c: +視線を R へ向ける----->
- 09 (3)
- 10 c: +視線を R から外す----->
- 11 (0.5)
- 12 c: +視線を R へ向ける
- 13 R: し (1) い (1.5) た?
c: +視線を R へ向ける----->
- 15 (4)
- 16 C: しゅうた(.) しゅうた(.) だよ
- 17 (4)
- 18 R: しゅうた
- 19 (0.6)
- 20 C: そうしゅうた
c: +うなずく

5.2 被験者 2 (D さん) の身体動作

抜粋 2 でも抜粋 1 と同様に 01 行目の発話の前にロボットが、ロボットの名前を発話している。その後、01 行目でロボットから「あなたは」と質問され

ている。ただし、「あなたの名前は」のように明確に聞かれていない。そのため、03 行目で、D は「あなたは」とだけ発語し、何を答えれば良いのかが分からず、言い終わる直前に笑っている。ここでの D の発語が一番目のトラブル源となっていると考えられる。その後、D は 05 行目で 01 行目のロボットに質問に対して名前を答えるべきか、確認し、修復を開始している。これに対して、06 行目でロボットは、「うーん」としか答えず、D は、名前を答えれば良いのか判断できかねている。続いて、D は 10 行目で名前を答える。12 行目では指の動きを使って漢字を示している。

14 行目では、ロボットが D の名前を発話している。ここで名前を答えるということにはおいては、修復が終了したと考えられる。

しかし、ロボットの一語目の発語は名前の一部を間違っている。ここが二番目のトラブル源になっていると考えられる。16 行目で D がロボットの発語を訂正し、ロボットが 18 行目で正しい名前を発話する。

この抜粋の中では、03 行目から始まる人 (D) がトラブル源になったケースでは、人による自己修復・自己開始のタイプの修復が見られる。14 行目から始まるロボットがトラブル源になったケースでは、ロボットによる自己修復・他者開始のタイプの修復が見られる。

この一連のプロセスで、C と異なり D は終始ロボットへ視線を向けている。そのため 5.1 のように視線を外す動作の検出によるトラブルの発生の手掛かりを得ることができない。

抜粋 2

会話開始から 1 分 38 秒後

(.) は 0.2 秒以下の短い無音区間、() 内は無音時間の秒数、R はロボットの発話、D は被験者の発話

d は被験者の発話以外の身体動作を示す

「+」は被験者の身体動作の開始点

「->」は同一の身体動作の継続を示す

- 01 R: あなたは
- 02 (0.5)
- 03 D: あなたは
d: +視線を R へ向ける
d: +笑う
- 04 (0.5)
- 05 D: スタックちゃんが名前だとしたら 僕も名前を答えた方がよくて
d: +視線を R へ向ける----->
----->
d: +笑う

- 06 R: うーん
 07 (1)
 08 D: あ うん うん 了解
 d: +笑う----->
 d: +視線を R へ向ける---->
 09 (0.5)
 10 D: えっと おおにしたつやっています
 d: +笑う----->
 d: +視線を R へ向ける----->
 11 (1.2)
 12 D: 大きいっていう字に 東西南北の西 達成するの
 たつに
 d: +机に指で字を書くふりをする->+視線を R へ向
 ける---->
 D: よく漢文で出てくる也で 大西達也って言います
 d: +視線を R へ向ける----->+視線
 を R から外す
 13 (0.3)
 14 R: たくや (.) たつや
 d: +視線を R へ向ける
 15 (.)
 16 D: たくやじゃない (.) たつやですね
 d: +視線を R へ向ける----->
 d: +笑う
 17 (4)
 18 R: たつや
 19 (.)
 20 D: うん (.) そうそうそう
 d: +視線を R へ向ける->

6 総合考察と今後の課題

ロボットとの会話においても、5.1 のように修復のプロセスで訂正が終了したことを示す手掛かりとなる身体動作（「うなずく、視線を向けなおす」）が見られた。他方、5.2 のように「笑う」という動作が継続的に見られて、それが訂正を終了したものなのか判断しにくい身体動作がみられた。

また、今回用いたロボットには、人の表情や視線、それ以外の身体動作などを検出する仕組みがない。そのため文字情報でしか聞き手の理解を確かめるべきがない。

今後の課題として、ロボットの機能面では、以下の点が挙げられる。(1)ロボットに発話内容に応じた表情を表現する機能を加える。また、機体の向きを変えられるにする、(2)人の表情や身体動作を認識する機能をロボットに加える。また、認識した身体動作をテキスト情報として LLM へ渡し、会話の訂正の手掛かりにする。

次に、会話の場面では、商品説明場面以外の制度的な会話（例えば、問診場面や就職試験の面接など）場面での検討を行いたい。

謝辞

本研究を進めるにあって実験にご協力いただいた友人の皆さんに感謝申し上げます。また、本研究会でコメントをいただいた皆様にも感謝申し上げます。

参考文献

- [1] 野村竜也: ロボットの社会的受容に関する要因の探索: ユーザ研究の視点から, 情報処理学会関西支部支部大会講演論文集, (2009)
- [2] P.グレイス. 清塚邦彦, 飯田隆 (訳): 論理と会話, 勁草書房, (1998/2020)
- [3] 谷郷力丸, 高橋卓見, 廣田敦士, 早川博章, 岡夏樹, 西崎友規子: 失敗を演出したロボットはヒトと円滑な関係を築くか, 情報処理学会関西支部支部大会講演論文集, (2016)
- [4] 大谷空遊希, 麻生智史, 勝川慶一, 西崎友規子: ロボットの失敗頻度が親しみやすさに及ぼす影響, 情報処理学会関西支部支部大会講演論文集, (2021)
- [5] 高橋嘉一, 上村洋介, 古田真将, 尹鴻飛, 前野彩実, 西崎友規子: ロボットのミスは製品紹介に有効か: 製品に対する興味の個人差による検討, 情報処理学会関西支部支部大会講演論文集, (2023)
- [6] アールティ: ChatGPT と連携して喋ってもらった - スタックチャン アールティ ver. 組立キットブログ Part.5, (2024)
- [7] Schegloff, E. A., Jefferson, G. and Sacks, H.: The preference for self-correction in the organization of repair in conversation, *Language*, Vol. 53, No. 2, pp. 361-382.(1977).
- [8] 野村竜也: Human-Agent Interaction (HAI) における人の主観評価, *人工知能*, Vol. 31, No. 2, pp. 224-229, (2016)
- [9] 坂井田瑠衣, 榎本美香, 伝康晴, 坊農真弓: フィールドに依存した身体相互行為の組織化過程—歯科診療における「修正」のやりとり—, *人工知能学会研究会資料, SIG-SLUD-B503-04*, pp.17-22. (2016).