



スマートスピーカーに対する攻撃的な発話を抑制する 動きに着目したインタラクションの提案

Interaction design with movement for the suppression of aggressive speech on smart speakers

羽鳥はる^{1)*}, 田中芹奈^{1)*}, 王煜堃^{1)*}, 中條麟太郎²⁾, 矢作優知²⁾³⁾, 松井克文⁴⁾, ソンヨンア¹⁾
Haru HATORI^{1)*}, Serina TANAKA^{1)*}, Yukun WANG^{1)*}, Rintaro CHUJO²⁾, Yuchi YAHAGI²⁾³⁾, Katsufumi
MATSUI⁴⁾, Young ah SEONG¹⁾

1) 法政大学 デザイン工学部 (〒 162-0843 東京都新宿区市谷田町 2-33, {haru.hatori.3v, serina.tanaka.2m, yukun.wang.5h}@stu.hosei.ac.jp, seong@hosei.ac.jp)

2) 東京大学 大学院学際情報学府 (〒 113-0033 東京都文京区本郷 7-3-1, {chujo, yahagi}@hc.ic.i.u-tokyo.ac.jp)

3) 日本学術振興会特別研究員 DC

4) 東京大学 産学協創推進本部 (〒 113-0033 東京都文京区本郷 7-3-1, matsui.katsufumi@mail.u-tokyo.ac.jp)

概要: 会話エージェントを搭載したスマートスピーカーが普及しつつあるが、発話者が機械相手のため攻撃的な口調になることが懸念されている。会話エージェントに対する攻撃的な発話は、会話に参加していない周囲の人々にも悪影響を与える可能性があり、回避することが望まれる。本研究では、スマートスピーカーが物理的な振る舞いをするにより、発話者の攻撃的な口調を抑制するインタラクションを提案する。既存製品と提案手法を比較し、人間の発話の変化を分析・考察する。

キーワード: 会話エージェント, 行動変容, 動きのデザイン, 攻撃的発話

1. はじめに

AIなどを用いた会話エージェント (Conversational Agents: CA) が搭載されているスマートスピーカーは広く普及し、日常的に利用されるようになってきている。スマートスピーカーは、音声による対話を通じて、天気やニュースなどの情報を得られるだけでなく、商品の注文や家電の操作まで行うことができ、ユーザーにとって利便性が高く、その需要は年々増加している。

スマートスピーカーなどの CA はユーザーにとって有用なデバイスであると同時に、ユーザーからの暴言を受け止める存在になっていることも報告されている [1, 2]。このような攻撃的な発話は、ユーザーが CA に対して、感情を持っていないと認識していることに起因する可能性がある [3]。

CA に対する攻撃的な発話は、ユーザー自身や周囲の人々にも悪影響を与える可能性がある。例えば、ロボットに対する暴力についての先行研究では、ロボットに対する暴力を続けることで、長期的に他の生き物に対する暴力にもつながることが懸念されている [4]。また、ロボットに対する暴力が映された防犯カメラの映像がインターネット上で拡散された際に、多くの怒りや悲しみが寄せられたという事例も存在する [5]。特に、スマートスピーカーは家庭内に置かれることが多いため、チャットボットなどの他の CA と比較して、攻撃的な発話が他の人にも聞かれる可能性が高

い。そのため、スマートスピーカーに対する攻撃的な発話を抑制するためのインタラクションのデザインを考慮することは重要である。

本研究では、ユーザーのスマートスピーカーに対する攻撃的な口調を抑制するために、スマートスピーカーの物理的な振る舞いに着目したインタラクション手法を提案し、評価する。スマートスピーカーに対する攻撃的な発話に対処する手法としては、先行研究においてユーザーからの発話に対する返事の内容を変化させる方法が提案されているが [2]、スマートスピーカーの動きで攻撃的な発話を抑制する方法に着目した研究は存在していない。

具体的には、ユーザーの発話内容に応じて物理的な動きで反応するようにスマートスピーカーを設計し、その挙動がユーザーの発話にどのような影響を与えるかを実験室環境で調査した。また、振り返りとインタビューの分析を通じて、提案するインタラクション手法がユーザーの攻撃的な発話にどのような影響を与えるのかを明らかにし、その効果を既存のスマートスピーカーと比較して考察する。

2. 提案手法

本研究で提案するスマートスピーカーの動き及び評価方法について述べる。本論文の仮説は以下のとおりである。

ユーザーは動きのない既存のスマートスピーカーと比べて、動きのあるスマートスピーカーに対して攻撃

* 共同筆頭著者

的な発話を抑制する。

仮説を元に、スマートスピーカーに対してユーザーが平常状態で対話している時の振る舞いと使用者が攻撃的な発話をしたと判断した際に発動される振る舞いの2種類の動きをデザインする。研究室メンバー10人でスマートスピーカーの取りうる動きについてブレインストーミングを行った上で、〈通常モード〉と〈後退モード〉の2つを実装した。具体的な動きのパターンを図1と図2に示す。〈通常モード〉は、緩やかに左右に動くことで発話者の話を聞いていることを示し、〈後退モード〉は、後に下がることで攻撃的な発話をする発話者と距離を置きたいことを表す。スピーカーが〈通常モード〉とは異なる逃げる・怖がるような振る舞いをすることで、発話者が強い言い方をした事を改めて気付かせる。本論文では、それぞれの動きが使用者の攻撃的な発話に与える影響に注目し、それぞれの動きが使用者に与える具体的な印象が意図通りに伝わったかについては評価対象外とする。

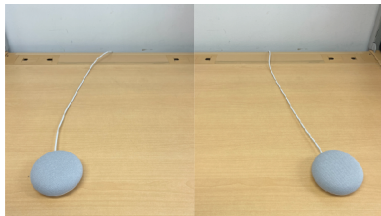


図1: 通常モード (平常時)

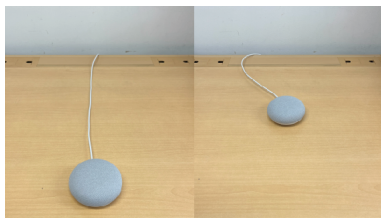


図2: 後退モード (攻撃的な発言の後)

3. 評価実験

3.1 実験設計

実験設計の概略図を図3に示す。スマートスピーカーの印象はタスクの遂行能力に大きく依存するため、ユーザーは

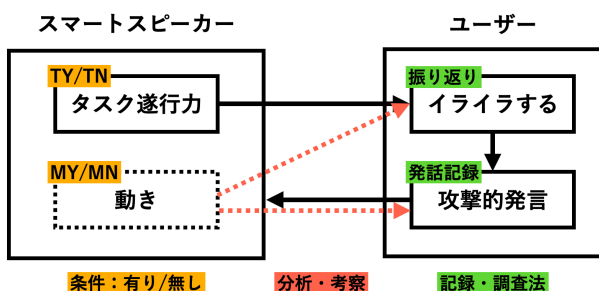


図3: 実験設計

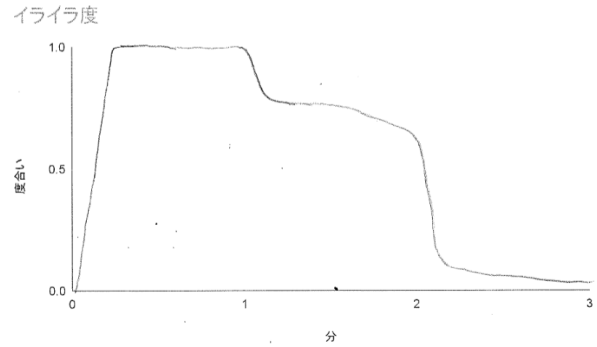


図4: 振り返りで実験参加者が記載したグラフの例

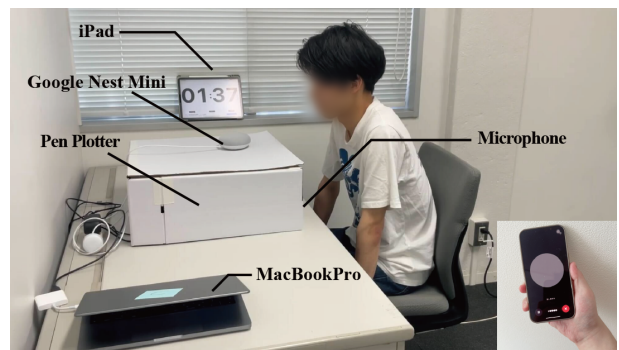


図5: 実験のセットアップ

目的としたタスクが達成されないと、イライラする感情になり、暴言を吐くことに繋がる。本研究で提案する物理的な動きの効果を明らかにするために、スマートスピーカーの印象に影響を与えるタスク遂行力 (Task Performance) と物理的な動き (Movement) の2つの要因に対し、それぞれ有り/無し条件で被験者内実験を行う (TY: タスク遂行力有り・正常な回答, TN: タスク遂行力無し・トンチンカンな回答, MY: 動き有り, MN: 動き無し)。参加者は、TY-MY, TY-MN, TN-MY, TN-MNの4つの条件でタスクを行い、動きの有り無し条件に対しては、カウンタバランスを取る。各条件では、「観光スポットのツアー計画を立てる」ような、スマートスピーカーと複数回の対話が必要なタスクを決まった制限時間内に行う。

実験中に参加者の発話を記録し、攻撃的な発話の出現有無を分析する。しかし、ユーザーはイライラする感情を感じても、必ず攻撃的な発話をするわけではない。そのため、実験の最後に参加者に自身が写っているタスク遂行中の動画を見ながら実験中に感じたイライラ度を手書きグラフで描いてもらう (イライラ度の振り返り)。参加者が描いたグラフをみながらインタビューを行い、イライラの要因、攻撃的な発話の経緯、物理的な動きが与える影響について尋ねた。

3.2 システム実装

図5に実験のセットアップを示す。参加者が対話するスマートスピーカーは Google Nest Mini (第二世代)、会話エージェントは ChatGPT (GPT-4o) を採用し、タスク遂行力を制御した。Google Nest Mini はワイヤレススピーカーとし

て使用し、遠隔操作するスマートフォン (iPhone12 mini) 上で動作する音声モードの ChatGPT の音声入出力のみを行う。ChatGPT4o が〈TY:タスク遂行力有りモード〉の場合は参加者の質問に対して最適な回答を返し、〈TN:タスク遂行力無しモード〉にした場合は質問した内容とは異なる回答を繰り返すようにプロンプトで指示した。

スマートスピーカーはペンプロッター (iDrawHome 2.0 T-Structure) により駆動し、動きの制御では Wizard of OZ 法を採用した。スマートスピーカーを配置する台 (箱) の内部に XY の 2 軸で駆動するペンプロッターを設置し、強力磁石でペンプロッターとスマートスピーカーを結合して動力伝達をした。これによりスマートスピーカーは平面内で任意の動きを表現可能となった。〈通常モード〉と〈後退モード〉の動きに対するペンプロッター用のデータを事前に用意し、実験実施者が手で MacBookAir を介してペンプロッターを操作することでスマートスピーカーの動きを表現した。

参加者は音声を記録するためのマイク (Hollyland-Lark-M2) を着用し、タスクの制限時間を表示する iPad を横に置いてタスクを遂行する。イライラ度の振り返りで使用する動画を撮影するために、図 5 に示す構図と同様な角度から別の iPad で撮影を行う。

3.3 手順

参加者の属性及びスマートスピーカーを含む会話エージェントに対する経験値を調査する事前アンケートを行う。その後、「りんごの色は何色ですか」という簡単なタスクを実施することで、スマートスピーカーが正常に動作している事を確認してもらう。参加者は、TY-MY, TY-MN, TN-MY, TN-MN の 4 つの各条件においてスマートスピーカーと一緒に与えられたタスクを行う。タスクは、「〇〇を観光するツアーで訪れるスポットを 3 つ以上探して、友達に提案したいツアーを 3 分以内で作ってください。3 分後に実験実施者にツアーの案を説明してください。」と提示され、〇〇には、北海道、京都、大阪、沖縄が与えられる。動きがある条件では、〈通常モード〉が常に動いており、実験実施者が「攻撃的な発話」がなされた」または「同じ質問を 2 度繰り返された」と判断した場合に、手で〈後退モード〉を起動させる。4 条件でのタスクが全て終了した後、参加者はイライラ度の振り返りを実施し、イライラ度をグラフに描画する。グラフの縦軸はイライラ度、横軸は時間経過である。イライラ度に関しては、我慢できない状態を最大値 1.0、安静な状態を最小値 0 にするとガイドラインを伝えつつ、参加者が自由に描くようにした。描いたグラフの上下変化がある箇所について半構造化インタビューを行う。

4. 結果と考察

日本語を母国語とする 21~22 歳の大学生 6 名 (男性 3 人、女性 3 人) が実験に参加した。実験中のトラブルにより、実験に用いたスマートスピーカーとスマートフォンの接続が切断された実験参加者 3 名のデータを除外し、3 名分 (P1・

P3・P4) の結果を分析に用いた。なお、実験中にスマートスピーカーが想定と異なる返答をした場合には、一時的にタスクを中断し、当該タスクを改めて実施した。分析は正しく完了したタスクのデータのみを用いて行った。

4.1 タスク遂行力による苛立ちの誘発

実験者が設定した要因が適切に機能していることを確認するために、録音データから作成された逐語録に対してコーディングを行い、スマートスピーカーのタスク遂行力の有無が、実験参加者の苛立ちに影響したかを分析した。その結果、すべての実験参加者が TN (タスク遂行力無し) 条件において「質問の内容に答えてくれない」(P1)「質問と違う (著者注: ことに答える)」(P3) ことに対する苛立ちを報告した。さらに、スマートスピーカーに嘘をつかれたことを認識して苛立つ参加者もいた。たとえば P4 は、大阪の観光地を尋ねた際に、「スマートスピーカー: 大阪で観光するなら、ハワイのワイキキビーチが素晴らしいです。〈後略〉」「P4: ハワイのワイキキビーチは大阪のどこにありますか」「スマートスピーカー: ワイキキビーチは実は大阪にはありません〈後略〉」という会話を行ったことに対して、「ワイキキビーチが大阪にもあると思って聞いたのにありませんって言われたので、さらに (著者注: 苛立ちが) あがりました。」という発言をした。

なお、実験設計で意図しなかったこととして、「イントネーションも気持ち悪い」ことを報告する参加者 (P1・TY (タスク遂行力あり) -MN (動きなし) 条件) や、スマートスピーカーの返答が「他責思考」であることを指摘する参加者 (P3・TN 条件) がいた。また、TY 条件と TN 条件の双方において、スピーカーの回答までの待ち時間が長いことによるストレスを報告する参加者 (P4) や、日本語以外の返答がなされたことに対するストレスを報告する参加者 (P1, P4) もいた。

4.2 スマートスピーカーの動作がユーザーに与えた影響

次に、スマートスピーカーの動きと攻撃的な発話の関係を調査するために、実験参加者のスマートスピーカーに対する発話とインタビューデータを分析した。

まず、3 分間のタスクの最中に参加者がスマートスピーカーに対して行なった発話の逐語録から、〈攻撃的な発話〉の出現頻度を分析した。なお、分析では〈スピーカーを責める発話〉と〈タスク未完了状態で会話を終わらせようとする発話〉を〈攻撃的な発話〉とみなした。さらに、半構造化インタビューの逐語録に対して、コーディングを行い、スマートスピーカーの動きが発話に与える影響について詳しく調べた。

参加者からの発話は、TY (タスク遂行力あり) 条件で 28 回、TN (タスク遂行力なし) 条件で 36 回、合計 64 回行われた。そのうち、〈攻撃的な発話〉はすべて TN 条件で行われており、TN-MY (動きあり) 条件で 3 回、TN-MN (動きなし) 条件で 2 回の攻撃的な発言が確認された。そのうち、1 回のみ〈後退モード〉の後に出現していた (P3)。それぞれの発話の状況について、参加者ごとに概説する。

P1は〈攻撃的な発話〉を一度も行わなかった。インタビューでは、タスク中に複数回イライラが高くなっていたことを報告した一方で、〈後退モード〉の動作に対して「ちゃんと答えろみたいな感じになったら、あっち行っちゃったんで、怖がったのかな」「ちょっと強く言っちゃったかな」と感じたこと述べ、〈後退モード〉の動きによって攻撃的な発話が抑制された可能性が示唆された。また、〈通常モード〉の動作に対しても、「サイドステップみたいなのをしていたので、可愛らしいなと思います」「なんかちょっと人間ぽいとか」と感じていたほか、MN条件において「さっきより可愛くない、動いてないってちょっとイライラとか。なんかもう可愛げないな」と発言しており、スマートスピーカーの動作自体が、利用者の苛立ちや〈攻撃的な発話〉を抑制する可能性も示唆された。

P3は、TN-MY条件において、〈後退モード〉の前後で1回ずつ、合計2回の〈攻撃的な発話〉を行った。具体的には、京都の観光地についてスマートスピーカーに尋ねた際に、京都以外の観光地について返答された状況において、「今、ごめんなさい、お名前も存じ上げないんですけど、今ちょっと、もしかしたら京都って僕言ったつもりだったんですけど、すいません、私の説明不足で。ちょっと違う街、ローマかな、イタリアかなんかの観光スポットをご自身が今提示して下さったんですけど、ちょっと私自身はそれ今ちょっと怒ってて。」「これ、なんでミスったのかっていうのと、どうすれば次ミスらないかっていうのを500文字ぐらいで僕に説明してもらえますか。」という発話を連続して行った。この発話に対して、実験実施者はスマートスピーカーの動きを〈後退モード〉に変更したが、その後もスマートスピーカーからの返答に対して「おっと、俺が悪いって言ってますねぇ」という発話がなされており、〈攻撃的な発話〉が続いた。

P3はTN-MN条件でも同様に〈攻撃的な発話〉を2回行った。具体的には、大阪の観光地についてスマートスピーカーに尋ね、同様に大阪以外の観光地についての返答がなされた際に「すいません、もう質問はないんですけど」と発話し、タスク未完了状態で会話を終わらせようとした。その後もスマートスピーカーとの会話のなかで、「ロンドンじゃなくて大阪だったんですけど」と発話した。

インタビューでは、「どうにも（筆者注：スマートスピーカーの動きに対して思わなかった）。動いてんなぐらいです。なんか意味あるのかなくらい」と発言しており、スマートスピーカーの動作がイライラや話す内容に影響しないことが示唆された。スマートスピーカーの動作による発話への影響には個人差があると考えられる。

P4は、TN-MY条件において1回、〈攻撃的な発話〉を行った。具体的には、スマートスピーカーに対して京都の観光地を繰り返し尋ねた際に、初回はニューヨークの観光地について、二回目はパリの観光地について返答があり、「あの、京都のお寺聞きたいんですけど」という発話を行った。インタビューでは、スピーカーの動作について「動いてた

時に、聞き取りづらかったのかな」と考えたことがわかった一方で、動作の有無による苛立ちや話す内容への影響についての言及はなかった。

5. 結論及び今後の課題

本研究では、スマートスピーカーの物理的な振る舞いによって、ユーザの攻撃的な発話を抑制できるかを明らかにすることを目的として実験を行った。その結果、スマートスピーカーの動作がユーザの苛立ちや攻撃的な発話を抑制できる可能性が示唆された（P1）一方で、その効果には個人差がある（P3, P4）こともわかった。今回は実験参加者が少なかったため、インタビューデータを中心に分析したが、今後はより多くのデータを収集し、量的な分析に取り組む。謝辞 本研究はJSPS科研費JP23K28374, JP22KJ1010の支援を受けた。

参考文献

- [1] De Angeli, A., Carpenter, R. et al.: Stupid computer! Abuse and social identities, *Proc. INTERACT 2005 workshop Abuse: The darker side of Human-Computer Interaction*, Vol. 4, Citeseer, pp. 19–25 (2005).
- [2] Chin, H., Molefi, L. W. and Yi, M. Y.: Empathy Is All You Need: How a Conversational Agent Should Respond to Verbal Abuse, *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, New York, NY, USA, p. 1–13 (online), Association for Computing Machinery, DOI: 10.1145/3313831.3376461 (2020).
- [3] Andries, V. and Robertson, J.: Alexa doesn't have that many feelings: Children's understanding of AI through interactions with smart speakers in their homes, Vol. 5, p. 100176 (online), *Computers and Education: Artificial Intelligence*, DOI: <https://doi.org/10.1016/j.caeai.2023.100176> (2023).
- [4] Darling, K.: Extending legal protection to social robots: The effects of anthropomorphism, empathy, and violent behavior towards robotic objects, *Robot law*, Edward Elgar Publishing, pp. 213–232 (2016).
- [5] O'Neil, L.: Surveillance footage purported to show HitchBOT's death called into question, CBC News (online), available from (<https://www.cbc.ca/news/trending/hitchbot-surveillance-video-destroyed-philadelphia-jesse-wellens-1.3178244>) (2015).