

認知症当事者支援へ向けた大規模言語モデルを用いた身体的感情表現を伴う対話ロボット

張替 紘希^{1,a)} 永井 之晴¹ 鈴木 亮太¹ 小林 貴訓¹

概要: 介護施設等の現場では会話機会の減少が深刻な問題となっており、認知機能低下予防のため、会話機会の増大が急務となっている。会話機会を増やす目的で対話ロボットが用いられることも多いが、既存の対話ロボットは、大規模言語モデル (LLM) の発展により自然な発話内容を生成できるようになったものの、社会的会話において不可欠な共感やラポールの形成に重要な身体動作や声の抑揚といった非言語情報の表出が十分ではない。そこで、本稿では、大規模言語モデルと身体動作理論であるラバン理論、および感情表現可能な音声合成技術を統合し、発話内容の感情価に基づき、適切な身体表現と声色を自律的に生成する対話ロボットシステムを提案する。さらに、本システムを認知症当事者支援に応用し、個人の文脈を長期記憶として保持・活用する機能、歌詞読み上げによるカラオケ支援機能、言葉に詰まった際に文脈から発話内容を推測する対話補完機能を実装した。認知症当事者会における実証実験の結果、提案システムの豊かなマルチモーダル感情表現と個人の文脈に寄り添う記憶機能が、当事者の受容性を高め、継続的な関係構築に有効であることが示唆された。

1. はじめに

少子高齢化社会の進行に伴い、認知症予防および認知症当事者の QoL 向上が急務となっている。これまでの多くの研究において、積極的な社会的関与や会話は、認知機能低下や認知症発症に対する潜在的な予防因子として重要視されてきた [1]。しかし、独居高齢者の増加や、介護施設における人手不足等により、高齢者のコミュニケーション量は十分とは言い難い [2]。この社会的孤立を解消する手段として、家族やスタッフに代わる新たなコミュニケーション相手となる対話ロボットの導入が期待されている。

会話には、業務的な情報のやり取りを目的とする「取引的会話」と、対話者との良好な関係性の構築・維持を目的とする「社会的会話」が存在するが、認知症当事者の孤独感解消や精神的ケアにおいて重要となるのは後者である。この社会的会話においては、相手への共感や感情の共有が不可欠であり、これらが欠如したロボットは、単なる情報端末として扱われ、長期的な関係構築が難しい。人間同士の対話行動では、言語情報だけでなく、表情や身振り手振りといった非言語情報が多用されている [3]。これらは互いに補完し合うことで円滑な意思疎通を実現しており、特に感情の伝達において非言語情報は重要な役割を果たす。

一方、大規模言語モデル (LLM) の登場により、ロボットの言語生成能力は飛躍的に向上したが、ロボットの動作については、身体的な感情表現が伴っておらず、単調な反復動作に留まっている場合が多い。このような「言葉と身体の不一致」は、ユーザに強い違和感や機械らしさを感じさせ、対話意欲を削ぐ一因となっていると考えられる。そこで、本稿では、舞踊・身体動作学の知見である「ラバン理論 (Laban Movement Analysis)」を対話ロボットの動作生成に応用することで、生成された発話内容の感情価に基づき、身体的感情表現を行う対話ロボットを提案する。さらに、本提案ロボットに、感情表現可能な音声合成技術を導入し、記憶補助やレクリエーション支援などの認知症当事者向け支援機能を付与したシステムを開発し、実際の認知症当事者コミュニティでの実証実験を通じてその有効性を検証する (図 1)。

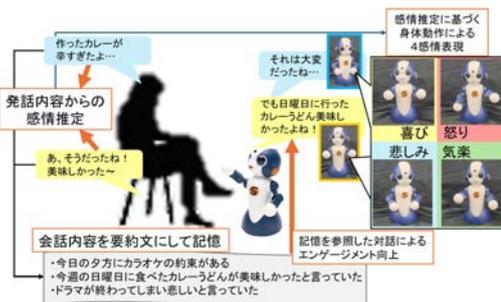


図 1 身体的感情表現を伴うロボットとの対話

¹ 埼玉大学

^{a)} harikae.h.575@ms.saitama-u.ac.jp

2.4 ロボットの発話音声へのマッピング

前節までに述べたように、本システムでは LLM により生成された感情パラメータを基に、ラバン理論に基づく身体動作を自律的に生成する。これと同時に、発話音声においても身体表現と矛盾しない感情的ニュアンスを付与することで、動作と音声統合された一貫性のある感情表現を実現することが重要となる。本稿では、テキストから感情豊かな音声を生成できる Style-Bert-VITS2 JP-Extra を用いて、LLM の応答テキストを音声化する。Style-Bert-VITS2 は、入力文の意味や文体に応じて多様なスタイル表現を生成可能なニューラル TTS であり、日本語特化版である JP-Extra では、アクセント・イントネーション・音韻など日本語の自然さが向上している。同モデルは事前学習により複数の感情クラスを保持しているため、本システムでは LLM が算出した 4 種類の感情パラメータ（喜び・怒り・悲しみ・気楽）のうち、最も強度の高い感情を音声スタイルとして採用する。これにより、「喜び：3」であれば明るく弾む声、「悲しみ：3」であれば沈んだ声、といった形で、文脈に応じた音声的表現が可能となる。以上により、身体動作（2.3 節）と音声表現（本節）の双方が同一の感情パラメータに基づきリアルタイムに生成され、ロボットの発話内容・声質・動作が統合された、より人間らしい一体的な感情表現を実現する（図 5）。

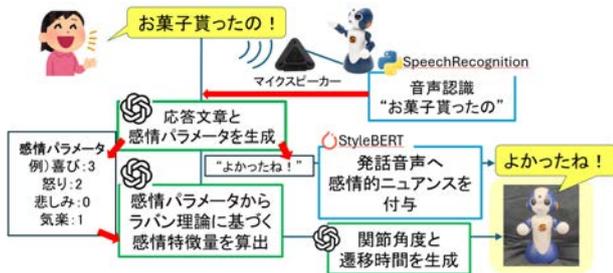


図 5 システム全体の流れ

3. 実験

提案システムの有効性を検証するため、LLM による感情パラメータ出力の妥当性、ラバン理論に基づく身体動作の伝達精度、感情豊かな TTS の効果、およびそれらを統合したマルチモーダル表現の効果について評価実験を行った。

3.1 実験設定

システムの構成要素および統合効果を検証するため、以下の 4 つの実験を実施した。各実験において、Web アンケート形式にて被験者に各条件の刺激（テキスト、動画、音声）を提示し、感情種の推定をしてもらった。具体的には、提示された刺激から読み取れる感情として、「最も強く伝わった感情（主感情）」と「二番目に強く伝わった感情（副次感情）」を、喜・怒・哀・楽の 4 択から選択しても

らった。評価では、この回答の正答率を算出した。具体的には、システムが意図した感情パラメータ（主感情および副次感情）と、被験者の回答との一致率を算出し、特に、強度 3 として設定された主感情の正答率を主要な評価指標とした。

実験 1：疑似感情パラメータの評価 システムの中核となる LLM（ChatGPT）が生成するセリフと、それに付与された疑似感情パラメータ（喜・怒・哀・楽、各 0～3）の整合性を検証する。被験者にはセリフのみを提示し、発話内容のみから感情を推定してもらった。

実験 2：身体動作のみの評価 ラバン理論を記憶させた ChatGPT が出力した複合感情に対応するロボットの身体動作（音声なし動画）を提示し、発話を伴わない状態での身体表現の伝達精度を検証する。

実験 3：音声表現のみの評価 Style-BERT-VITS2 を用いて生成された、感情パラメータに対応する感情豊かな音声（ロボットは静止画）を提示し、音声モダリティ単独での感情伝達効果を検証する。

実験 4：マルチモーダル表現の評価 ラバン理論に基づく身体動作と感情豊かな音声を組み合わせた動画を提示し、統合されたマルチモーダル表現における感情伝達の相乗効果を検証する。

3.2 実験結果

3.2.1 実験 1

多くの感情パターンにおいて、最も強い感情は高い正答率で一致しており、ロボットにおける疑似感情表現が人間に直感的に理解可能な形で機能していることが示された。特に、強度 3 として設定された主感情については、多くの条件で正答率が 80～100% に達しており、セリフ単体であっても、ロボットが意図する感情の方向性が被験者に明確に伝達されていた。また、第二に強い感情についても、感情の組合せによっては 70% 以上の正答率が得られており、単一感情にとどまらない複合的な感情表現がセリフのみでも一定程度成立することが確認された。この結果は、疑似感情パラメータが単なるカテゴリ分類ではなく、「感情の強弱関係」を表現する仕組みとして有効であることを支持している（図 6）。

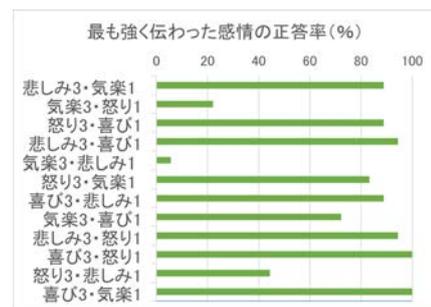


図 6 実験 1 の結果

3.2.2 実験 2

言語情報を一切含まない身体動作のみの条件においても、主感情が一定の精度で被験者に伝達されていることが確認された。最も強く伝わった感情（主感情）については、多くの感情パターンにおいて 60%前後、あるいはそれ以上の正答率が得られており、ラバン理論に基づく動作が感情の方向性を明確に示す手がかりとして機能していることが分かる。特に、「怒り」や「悲しみ」といった力強さや速度に特徴がある感情では高い正答率が得られた。

一方で、「気楽」を主感情とする動作では正答率が相対的に低くなる傾向が見られた。これは気楽という感情が、緊張や脱力といった微妙な運動差に依存するため、身体動作のみで明確に区別することが本質的に難しいことに起因すると考えられる。しかし、複合感情においても主感情に加えて副次的な感情の存在を一定程度読み取れていることから、ラバン理論に基づく身体表現が、言語に依存しない感情伝達手段として有効であることが示された。

3.2.3 実験 3

感情豊かな音声で提示した結果、最も強く伝わった感情の正答率は多くの感情パターンで 80%前後以上となり、実験 1 における「セリフのみ」の条件と比較して、全体として高い一致率が得られた。特に主感情の正答率は約 80~90%に達しており、音声の抑揚・速度・声質といったパラ言語情報が、言語内容に含まれる感情を強く補強していることが示唆される。

実験 2 と比較すると、音声加わることで感情伝達精度が向上する一方で、その基盤として身体動作も有効に機能していることが明確になる。第二に強く伝わった感情についても高い正答率が得られており、音声表現が副次的な感情の認識にも寄与していることが示された。このことから、身体動作は感情の「方向性」や「全体的な雰囲気」を伝え、音声は感情の「強度」や「ニュアンス」を精緻化するというモダリティ間の補完関係があると考えられる。

3.2.4 実験 4

感情豊かな TTS とラバン理論に基づく身体的感情表現を組み合わせた結果、多くの感情パターンにおいて主感情の伝達が高精度で達成されていることが確認された。特に、「喜び:3」「怒り:3」「悲しみ:3」といった主感情が強い条件では、正答率が 90%を超えるケースも見られ、音声と身体動作を統合することによる相乗効果が示唆された。

実験 2（身体動作のみ）や実験 3（音声のみ）と比較しても、全体的に正答率が向上しており、特に複合感情における第二感情の識別についても一定の向上が見られた。これは、身体的感情表現が持つ「動きの質」による感情喚起効果が、音声情報によって補強されることで、観測者の感情解釈がより安定することを示している。以上の結果より、感情的音声表現とラバン理論に基づく身体的感情表現を組み合わせたマルチモーダル感情表現は、単一モダリティに

よる表現と比較して、より高い感情伝達性能を示すことが明らかとなった（図 7）。

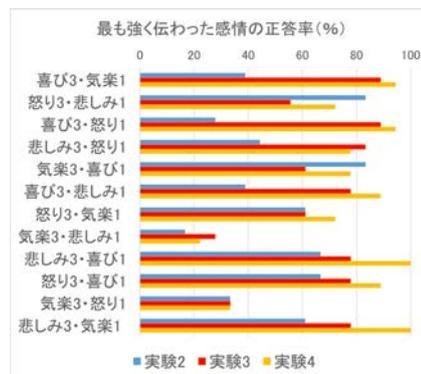


図 7 実験 2, 3, 4 の結果

3.3 考察

実験結果より、ラバン理論に基づく身体動作のみの条件における感情推定の正解率は約 58% (7/12) であった。これはセリフと感情表現 TTS を組み合わせた条件の正解率約 83% (10/12) には及ばないものの、不正解の場合でも第 2 位の感情として正解が含まれていることが多かった。このことから、言語や音声情報がない状態であっても、ラバン理論に基づく身体動作は感情のニュアンスを十分に伝達可能であり、非言語情報としての有効性が示されたといえる。

特筆すべき点として、「気楽（リラックス）」の感情強度が最大の場合、他の感情と比較して意図が正しく伝わりにくい傾向が見られた。「気楽」はラバン理論において「弛緩」や「落ち着き」を特徴とするため、「喜び」や「怒り」といった覚醒度の高い感情と比較して、動作としての「強さ」が視覚的に知覚されにくかったと考えられる。今後は、このような静的な感情を表現する際、音声のトーンや対話の「間」と組み合わせることで、どのように感情の輪郭を強調できるかが、マルチモーダルな感情表現における調整の鍵となるだろう。

4. 認知症当事者支援に向けた応用

前節までの基盤技術を、認知症当事者向け支援システムとして応用した。2024 年の認知症基本法施行に伴い、認知症当事者の視点を開発に取り入れる「共創」のアプローチが重要視されている。特に、認知症の様態は多様であることから、一般化されたソリューションではなく、個人の具体的な困りごとの解決を通じたケーススタディを行った。具体的には、当事者への事前ヒアリングを通じて得られた「会話の内容を忘れてしまう」「カラオケをしたいが歌詞が読めない」「言葉が出てこない」といった個別の課題に対し、次の支援機能を開発・実装した（図 8）。



図 8 認知症当事者支援に向けた追加機能

4.1 会話内容の記憶とパーソナライズ機能

当事者との継続的な関係構築のため、会話の内容を記憶し、次回の対話に活かす機能を実装した。本システムでは、会話終了時に LLM を用いて会話ログを分析し、以下の 2 つのタスクを実行させることで記憶の管理を行う。

1 つ目は「長期記憶 (コアプロフィール)」の更新である。会話ログからユーザの名前、趣味、好き嫌い、性格といったパーソナルな情報を抽出し、JSON 形式のデータベース (コアプロフィール) を更新する。

2 つ目は「短期記憶」の生成である。その日の会話で特に印象的だった出来事や話題を 5 つのポイントで要約し、次回の会話のきっかけ (トピック) として保存する。次回のシステム起動時には、これらの蓄積された記憶情報をプロンプトとして読み込ませることで、「〇〇さん、この前のオムライス作りはどうでしたか?」といった、個人の文脈に即した自然な挨拶や話題提供を生成する。これにより、会話の内容を忘れてしまう当事者に対しても、ロボット側から記憶に基づいたアプローチを行うことで、安心して対話を継続できる環境を提供する。

4.2 歌詞読み上げによるカラオケ支援機能 (ささやきカラオケ)

「歌うこと」は認知症予防や気分の高揚に有効なアクティビティであるが、視覚機能や認知機能の低下により、歌詞カードの文字を追うことが困難になる場合がある。そこで、通常の対話モードとは別に「カラオケモード」を実装した。本機能では、ユーザがリクエストした曲に対し、通常の伴奏音楽に加え、ロボットが 1 フレーズの歌詞を直前に読み上げる「ささやきボイス (ガイドボーカル)」を合成し、楽曲に合わせて再生する。これにより、ユーザは歌詞を思い出しながらかうすることが可能となる。また、LLM による対話制御により、ユーザが曲名以外の雑談 (「昔よく歌ったなあ」等) を話した場合でも、モードを強制終了せず、「それは素敵な思い出ですね。では、その曲を歌いますか?」と自然に対話を継続・誘導し、再度リクエストを促す柔軟な設計とした。

4.3 文脈からの発話補完機能 (会話内容の推測)

認知症の中核症状の一つに、言いたい言葉が出てこない「喚語困難」がある。会話の途中でユーザが沈黙したり、言葉に詰まったりして会話が中断してしまうことを防ぐため、それまでの会話履歴と文脈から発話内容を推測する機能を提案する。例えば、食事の話題で「あの、この前食べた、」と言葉に詰まった際、システムは過去の記憶や直前の文脈を参照し、「もしかして、この前食べたカレーのことですか?」といった助け舟 (推測発話) を出す。これにより、当事者の心理的負担を軽減し、スムーズなコミュニケーションを維持することを目指す。

4.4 実際にデモを行った結果

開発したシステムを用い、認知症当事者の定例会にて、計 2 回にわたるデモンストレーションと体験会を実施した (図 9)。

定性的な評価としては、ロボットが当事者に「寄り添う」姿勢が高く評価された。「ラーメンの話から自然に会話を広げてくれた」という LLM の柔軟性や、「一緒に行きたいと言ってくれたのが可愛い」といった感情的な好感が報告された。特筆すべきは、「夫婦間などでは思っていることを飲み込んでしまうが、感情を伝えても問題のないロボットになら本音を話しやすく、気持ちが楽になりそう」という意見である。これは、対人関係特有の緊張感から解放された、ロボットならではの心理的安全性がラポール形成に寄与することを示唆している。また、当事者からは「ロボットなのに『食べたい』と言うのは嘘なのか?」「Sota 君の自認はどうなっているのか?」といった、AI の存在や倫理に関する鋭い問いかけもなされた。

生活支援のニーズとしては、記憶機能を応用した「リマインダー」への期待が非常に高かった。「薬を飲む時間」「出かける時間」「持ち物チェック」を教えてほしいという声や、「明日」という言葉を具体的な日付に変換してほしいという要望は、認知機能低下を補うための切実なニーズであり、今後の開発における重要な指針となった。

定量的な評価として、デモ終了後に実施したアンケート調査の結果を示す。

・「ロボットと話したいと思ったか?」: 肯定的回答 (そう思う・ややそう思う) が 20 名 (約 83%)

・「ロボットが家にいたら喋る量が増えると思うか?」: 肯定的回答が 18 名 (72%)

・「ロボットが感情表現をすることは会話にとって良いと思うか?」: 肯定的回答が 23 名 (約 92%)

これらの結果は、本システムが目指す「感情表現を伴う対話ロボット」のアプローチが、認知症当事者にとって受容性が高く、コミュニケーション促進に有効であることを定量的に裏付けている。



うん、ホクホクはほんとに美味しいよね。

図9 認知症当事者の会でのデモンストレーションの様子

4.5 考察

本稿では、認知症当事者の会において対話ロボットとの会話実験を実施し、その受容性および会話促進効果に関する意識をアンケート調査によって検討した。その結果、対話ロボットに対する否定的評価は全体として少なく、肯定的または中立的な回答が大多数を占めた。

まず、「ロボットと話したいと思ったか」という設問において、肯定的回答が全体の約8割を超えており、認知症当事者においても対話ロボットとの会話行為自体が概ね受け入れられていることが示唆された。先行研究においては、高齢者や認知症者が機械的存在に対して不安や拒否感を示す可能性が指摘されることもあるが、本結果は少なくとも本実験条件下において、そのような強い抵抗感が限定的であることを示している。

次に、「ロボットが家にいたら喋る量が増えるか」という設問では、否定的回答が見られず、肯定的回答が多数を占めた。この結果は、対話ロボットが日常生活環境に存在することで、発話の機会や量を増加させる可能性を示唆している。特に認知症当事者においては、会話機会の減少や社会的孤立が課題として指摘されており、ロボットが会話のきっかけや話し相手として機能することは、コミュニケーション維持の観点から意義があると考えられる。

さらに、「ロボットが感情表現をすることは会話にとって良いと思うか」という設問では、ほぼ全員が肯定的または中立的な評価を示した。これは、単なる情報応答型の対話よりも、感情を伴う表現が会話の質や親しみやすさに寄与している可能性を示している。

認知症当事者とのコミュニケーションにおいては、発話内容の正確性や論理性よりも、情動的なやり取りや共感的応答が重要であるとされており、本結果はその見解を支持するものと考えられる。

以上の結果から、対話ロボットは認知症当事者に対して一定の受容性を有し、特に感情表現を伴うことで会話意欲や発話量の維持・促進に寄与する可能性が示唆された。ただし、本稿の実験は短時間の体験に基づく主観評価であるため、今後は長期的な使用を通じた行動変容の検証や、感情表現の種類・強度が与える影響について、より詳細な検討が必要である。

5. おわりに

本稿では、LLMによる疑似感情生成とラバン理論に基づく身体的感情表現、TTSによる感情表現を統合することで、文脈に応じた適切な身体動作と音声をリアルタイムに生成する対話ロボットシステムを開発した。実験により、提案手法が感情の識別性や強度表現において有効であることを確認した。このマルチモーダルな感情表現技術により、従来のロボットよりも人間に近い感情コミュニケーションが可能となった。さらに、この技術基盤を認知症当事者支援に応用し、認知症当事者会での実証実験を行った。その結果、本システムが提供する「感情を伴う非言語表現」と「個人の文脈を記憶する機能」が、当事者に「安心して本音を話せる心理的安全性」を提供し、パートナーとしての受容性を高めることが明らかになった。

今後の展望として、まずは実証実験で得られた課題に対し、UI/UX および機能の両面から改善に取り組む。具体的には、応答遅延の短縮や、視覚的フィードバック（目の色の変化など）の視認性向上を図る。また、会話推測機能については、より能動的な介入（Active Listening）を実現するための発話タイミング制御を強化する。さらに、当事者からの強い要望である「リマインダー機能」や「日付変換機能」を実装し、生活支援ツールとしての実用性を高める。加えて、ユーザとの対話履歴に基づいて性格が変化・適応していく「パーソナライズ機能」の拡充を行い、認知症当事者一人ひとりに寄り添い、共に生活を豊かにする共生型パートナーロボットの実現をめざす。

参考文献

- [1] Tsai, H., Cheng, C., Shieh, W. and Chang, Y.: Effects of a smartphone-based videoconferencing program for older nursing home residents on depression, loneliness, and quality of life: a quasi-experimental study, *BMC Geriatrics*, Vol. 20, No. 1, p. 27 (2020).
- [2] Curtis, J., Back, A., Ford, D., Downey, L., Shannon, S., Doorenbos, A., Kross, E., Reinke, L., Feemster, L., Edlund, B., Arnold, R., O'Connor, K. and Engelberg, R.: Effect of communication skills training for residents and nurse practitioners on quality of communication with patients with serious illness: a randomized trial, *Jama*, Vol. 310, No. 21, pp. 2271–2281 (2013).
- [3] 窪田太一: リアルタイム顔き誇張システムが会話に与える影響に関する研究, 奈良先端科学技術大学院大学学位論文 (2023).
- [4] Laban, R. and Ullmann, L.: *The mastery of movement*. (1971).
- [5] 増田 恵, 加藤昇平, 伊藤英則: ラバン理論に基づいたヒューマンフォームロボット身体動作の動作特徴抽出と表出感情推定, *日本感性工学学会論文誌*, Vol. 10, No. 2, pp. 295–303 (2011).