音声対話エージェントにおける CG らしさと人間らしさを統合したふるまい表現

Integrated Expressions of CG-specific and Human-like Behaviors in an Embodied Conversational Agent

稲葉　響[1*]　上乃　聖[1]　李　晃伸[1]
Hibiki Inaba[1]　Sei Ueno[1]　Akinobu Lee[1]

[1] 名古屋工業大学
[1] Nagoya Institute of Technology

Abstract: 平面ディスプレイ上の音声対話エージェントは存在感が乏しく，対話開始や継続に抵抗が生じやすい．本研究はカートゥーン調の CG らしいふるまいと人間の動作を模倣した人間らしいふるまいを，ユーザの積極性に応じて切り替える統合手法を提案する．18 名の主観評価実験により，CG らしいふるまいが話しかけを促し，かつユーザの様子に基づく切替そのものが緊張や抵抗感を軽減する効果があることが示された．

1　はじめに

音声対話エージェントの中でも身体性を持つ外見を備えたものは Embodied Conversational Agent（ECA）と呼ばれる [1]．近年，大規模言語モデル（LLM）やマルチモーダル LLM の発展により機械との高度な対話が実現できるようになってきており，実社会における ECA の活用が期待されている．

平面ディスプレイ上に表示される CG ベースの ECA は，実空間に存在しないため存在感に乏しく [2]，自発的に話しかけることや対話を継続することに抵抗を感じやすいという課題がある．この課題に対してエージェントの身体的なふるまいの役割に着目した研究があり，ふるまいがユーザの印象や対話エンゲージメントに影響することが示されている [3, 4]．

人型エージェントのふるまいは人間の動作を模倣するよう作られるのが一般的であるが，アニメ調あるいはカートゥーン調のエージェントにおいては，アニメあるいは CG 特有の誇張されたふるまいが可能である．先行研究では，実際の対話インタラクションにおいてそれぞれのふるまいスタイルが対話対象としての認知に与える影響を調査し，CG らしいふるまいが生命感や存在感を感じさせ興味や関心を集めやすい一方で，人間らしいふるまいのほうが知性があり対話に集中できるとしている [5]．

本研究ではこの先行研究を発展させ，ユーザの状況に応じて人間らしいふるまいと CG らしいふるまいを切り替える統合的なふるまい表現手法を提案する．対話への積極性に応じてふるまいの人間らしさと CG らしさを調整することで，話しかけ前から対話中まで一貫して話しかけやすく話しやすい音声対話エージェントを実現する．以下，ふるまいの定義と提案システムについて述べ，評価実験を行った結果を報告する．

*連絡先：名古屋工業大学
〒 466-8555 愛知県名古屋市昭和区御器所町
E-mail: h.inaba.892@stn.nitech.ac.jp

2　エージェントのふるまい表現

2.1　ふるまいの表現手法

音声対話エージェントのふるまいは人間が対話中に行うような動作や表情を取り入れることが多いが，その表現手法については，人間が実際に行う動作を模倣して再現することが主流である．モーションキャプチャやソフトウェア上でのトレースにより人間の自然な動きを取り入れ，実際の人間の動きに基づいた自然なふるまいを実現している．本研究ではこのような表現手法を「人間らしいふるまい」とする．

一方，エージェントにおいては，キャラクター性の高いエージェントの身体表現として人間らしいふるまいを模倣するよりも CG キャラクターとしての特性を活かした表現のほうが好ましいという考え方から，アニメーションや漫画におけるカートゥーン調表現を取り入れたふるまいが採用されることがある．カートゥーン調表現は，アニメーション作品における動作の強調の典型的な技法に則ったものであり，代表例として，キャラクタの感情を記号などで視覚的に表現する漫符や，身

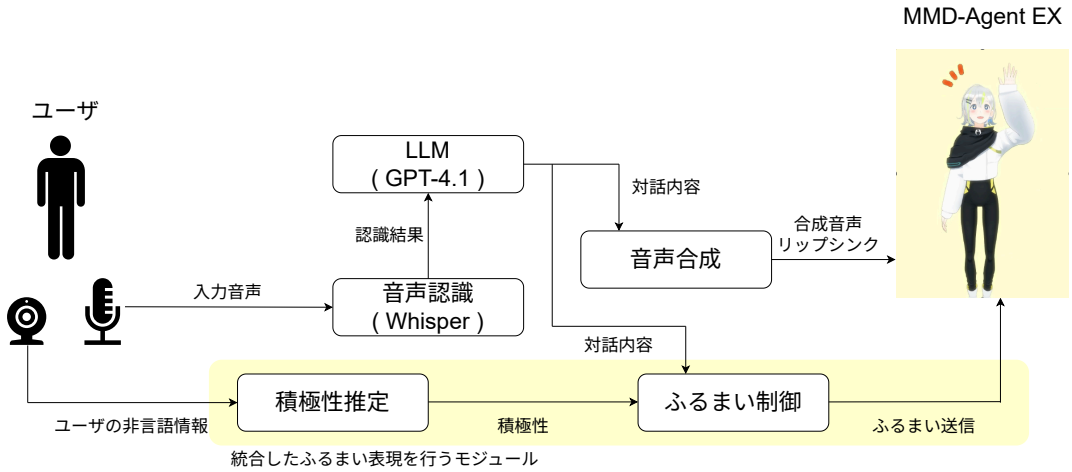

図 1: システムの構成図

体動作の躍動感を強調する誇張ジェスチャ[6] などが挙げられる．

先行研究 [5] では，音声対話エージェントにおける CG らしいふるまいの効果と影響を調査し，CG らしいふるまいはユーザが認知するエージェントの生命感や存在感を高める一方で，じっくり話し合う相手としての認知は弱く，人間らしいふるまいのほうがエージェントに対する知性を感じやすいという傾向が示されている．

2.2 統合したふるまい表現

ユーザのシステムに対する積極性を定義し，その積極性に基づいて人間らしいふるまいと CG らしいふるまいを切り替えるふるまい表現手法を提案する．ここでの積極性とは，ユーザが対話エージェントにどのくらい関与しているかの度合いを表す指標とする．

積極性は対話前から対話中にかけて変化するものとする．対話前では，ユーザがエージェントから離れた位置にいる，あるいはエージェントに意識を向けていない状態を積極性が低いとし，ユーザがエージェントに接近している状態を積極性が高いとする．対話中では，ユーザがエージェントと向き合い発話を継続している状態を積極性が高いとし，視線を外している，あるいは発話が途切れているなどの状態を積極性が低いとする．

この積極性に基づいて人間らしいふるまいと CG らしいふるまいを以下のように調整あるいは切り替えて表出する．（1）積極性が低い状態では CG らしいふるまいによってエージェントの存在感を強調し，ユーザの興味や関心を喚起しやすくする．（2）積極性が高い状態では，人間らしいふるまいによってエージェントにより知性を感じさせる．（3）対話の途中でユーザがよそ見をするなどして積極性が低下した場合，再び CG らしいふるまいを用いることでユーザの注意を引き戻し対話への復帰を促す．これにより，話しかけ前から対話中まで一貫して話しかけやすいふるまいを行う．

3 システム構成

3.1 実験システム

システムの構成図を図 1 に示す．通常の音声対話システムの構成に加えて，ユーザの状態や行動から積極性を推定するモジュールと，推定された積極性に従って対話内容にあったふるまいを制御するモジュールを新たに加える．

積極性の推定について，先行研究では対話中のエンゲージメントの推定を相槌，うなずき，視線，笑いなどの非言語的ふるまいやエージェントとの物理距離から推定する試みがあり [7, 8]，本研究においてもカメラや音声情報から積極性を推定するものとする．ただし本研究においてはふるまい統合の効果を検証することを実験の目的とするため，この部分は WoZ (Wizard of Oz）法，すなわち実験者がライブカメラ映像をもとに積極性を逐次推定する手法で動かす．

ふるまい制御モジュールは，推定された積極性に基づいてふるまいの人間らしさと CG らしさを統合して表出する．ふるまいの統合方法については，いずれかを選択して再生するバイナリ手法から，度合いに応じて中間的な動作を生成する（あらかじめ定義された両モーションを線形補間するなど）が考えられるが，本研究では単純なバイナリ手法を採用し，一定の積極性からどちらを再生するかを実験者が判断するものとする．

図 2: ふるまい例（対話前）

CG エージェントの表示および制御は MMDAgent-EX[9] を用いる．モデルには “ジェネ（Gene）”[10] を使用する．音声対話部には，Whisper[11] による音声認識，GPT-4.1 による応答生成，TTS による音声合成を用いた．

3.2 対話タスク

対話タスクは，旅行代理店に設置された案内エージェントによる旅行相談タスクとする．ユーザが国内旅行の行き先に悩んでいるという設定の下，エージェントがユーザの希望を聞きながら複数の観光地を提案し，各地の魅力を伝える．LLM の対話用プロンプトは旅行案内対話を想定し，行き先の希望確認，同行者の確認，重視点の確認，行き先の提案，観光地の詳細説明の順で対話を進行するよう設計した．

このタスク内容に沿って，内部状態や感情を表現するリアクション（全身モーション）を用意した．ユーザがエージェントと対話を開始する前の状態のためのふるまい 1 種（呼びかけ），および対話中のふるまい 6 種（考え中・提案，回想，焦り，同情，高揚，喜び）の計 7 種類を設計し，これらに対して，CG らしいふるまいと人間らしいふるまい 2 種類のふるまいモーションを作成した．人間らしいふるまいは，案内業務の店員の動作を行う人間の動作をキャプチャしたものをベースに表情の変化などの調整を施した．CG らしいふるまいは，漫符および誇張ジェスチャ（予備動作，両端詰め，後追いの工夫，副次アクション）のカートゥーン調特有表現を手作業で作成した．

例として対話前のふるまいを図 2 に示す．CG らしいふるまいでは気づきの漫符やキャラクターらしい大袈裟な手振りを行うが，人間らしいふるまいでは実際の店員のように控えめなお辞儀のみを行っている．

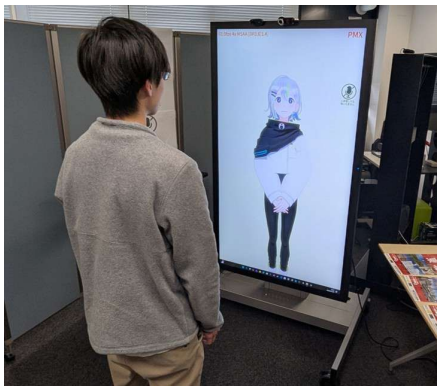

図 3: 実験中の様子

4 評価実験

4.1 実験設定

実験では，提案手法（ふるまいの切り替えを行う），CG らしいふるまいのみ，人間らしいふるまいのみの 3 種類のエージェントを比較する．実験参加者は 3 種類のエージェントのいずれか 1 つと対話を行う．

実験室内に縦型の大型ディスプレイを設置して等身大でエージェントを表示する．実験参加者に旅行対話を行うエージェントであることが自然に認知できるよう，周囲に旅行パンフレットなどを配備する．実験中の様子を図 3 に示す．手順は実験参加者が実験室に入室するところから開始し，実験参加者は実験室に設置されたエージェントに自発的に話しかける，あるいは旅行パンフレットを閲覧するなど，自由な行動を通じて旅行の行き先を決定する．

なお，本実験は初見のエージェントに対する評価を行うため，エージェントとの対話は強制せず，対話の開始および継続は実験参加者の自主判断に委ねる．また，実験には制限時間を設けず，実験参加者が旅行先を決定した時点で実験を終了とする．

提案手法では WoZ 法により，実験者が部屋の様子を別室で観察しながら適宜ふるまい再生を行う．参加者が入室した時点は対話前の CG らしいふるまいを再生し，対話中は基本的に人間らしいふるまいを行う．対話中に参加者がエージェントと異なる方向を注視し始めた場合には，CG らしいふるまいへと切り替える．

4.2 評価方法

主観評価アンケートを収集する．実験参加者は対話前および対話中での CG エージェントに対する評価（表 1）に加え，音声対話システムや対話エージェントに対する知識と利用経験，CG キャラクタに対する親和性を問うアンケートに回答する．評価は 7 段階のリッカー

表 1: CG エージェントに対する評価アンケート

	対話前での質問
(1)	エージェントがこちらに気づいているような動きをした
(2)	エージェントの様子から話しかけやすそうな印象を受けた
(3)	エージェントの動作によって「話しかけてほしい」と呼び掛けているように感じた
(4)	エージェントに声をかけるのに緊張感を感じた
	対話中での質問
(5)	エージェントがこちらの要望を汲み取ろうとする姿勢が感じられた
(6)	エージェントがこちらの発言を理解して反応しているように見えた
(7)	エージェントの動作が気になった
(8)	エージェントと話を続けるのに緊張感を感じた
(9)	エージェントと話を続けるのに抵抗感やストレスを感じた

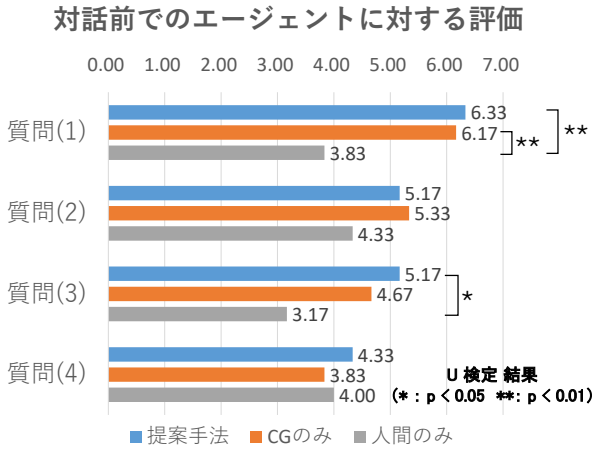

図 4: 対話前のエージェントに対する評価および検定結果

ト尺度を用いる．また，アンケート回答後に各設問に対する評価理由についてインタビューを行う．アンケート結果は平均スコアについてマン・ホイットニーの U 検定（両側検定，有意水準 5%）を行う．

4.3 実験結果

大学生・大学院生 18 名（男性 13 名，女性 5 名）の実験参加者の対話前評価結果を図 4，対話中の評価結果を図 5 に示す．

対話前でのエージェントに対する評価の分析

提案手法は，人間らしいふるまいのみを行うエージェントに対して，質問 (1) および (3) で有意に高い評価を示した．また，質問 (2) においても人間らしいふるまいのみを行うエージェントより高い評価となった．さ

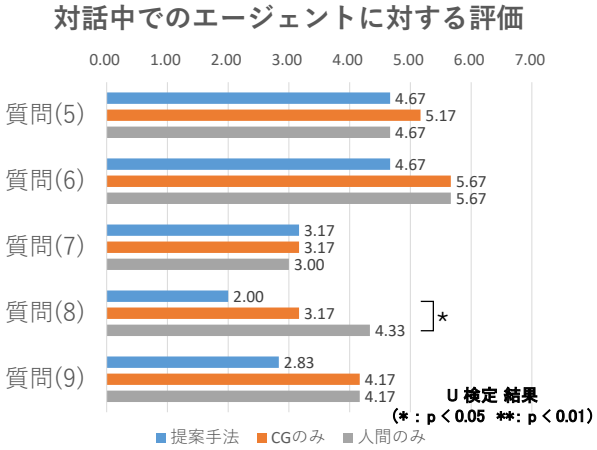

図 5: 対話中のエージェントに対する評価および検定結果

らに，CG らしいふるまいのみを行うエージェントも同様に，質問 (1) から (3) で人間らしいふるまいのみを行うエージェントより高い評価を得た．これらの結果から，CG らしいふるまい表現は対話前の段階において，エージェントの存在感を印象付け，「話しかけてほしい」という意図を伝える効果があることが示された．

一方，質問 (4) の「エージェントに声をかけるのに緊張感を感じた」については，エージェント間で有意な差は見られなかった．これは，ふるまいの影響よりも初対面のエージェントに自発的に話しかけるという負担が大きかったものと考えられる．

対話中でのエージェントに対する評価の分析

対話中の各手法の評価の平均スコアおよび検定結果を図 5 に示す．

質問 (8) では，提案手法が人間らしいふるまいのみを行うエージェントと比較して有意に低い値を示し，CG らしいふるまいのみを行うエージェントに対しても低い評価となった．質問 (9) においても同様の傾向が見られた．このことから，ユーザの対話に対する積極性に応じてふるまいを切り替える手法が，対話継続時の緊張感や抵抗感を低減する効果があるとみられる．

質問 (5) から (7) については，いずれのエージェント間でも有意な差が見られなかった．インタビューでは「エージェントが自分の発話に対して正しい応答をしていた」といった，ふるまい以外の面での評価理由が複数見られており，これらの質問項目はふるまいだけでなくエージェントの対話性能全体に対する回答になってしまったと考えられる．

表 2: 緊張感に関する評価スコア平均の推移

手法	(4) 対話前	(8) 対話中	増減
提案手法	4.33	2.00	-2.33
CG のみ	3.83	3.17	-0.66
人間のみ	4.00	4.33	+0.33

表 3: 提案手法における緊張感の評価スコア（提案手法）

参加者	対話前 (Q4)	対話中 (Q8)	スコア差
A	5	1	-4†
B	1	1	0
C	2	2	0
D	6	5	-1
E	6	2	-4†
F	6	1	-5†

†：スコア差が − 3 以下の参加者

提案手法についての分析

CG らしいふるまいと人間らしいふるまいを切り替える提案手法について分析する．

表 2 は緊張感に関する質問 (4)（対話前）と質問 (8)（対話中）の評価スコア平均の推移である．対話前は手法間で差がないが，対話中では提案手法が従来手法より低い値を示した．

この差の要因を分析する．表 3 に提案手法における質問 (4) と質問 (8) の個別スコアとその差を示す．対話前，対話中でスコアが 3 以上低下した参加者が 3 名見られた．この参加者のうち 2 名は対話経験が 2 回以下であった．このような評価の低下は従来手法では見られなかった．このことから，特に対話経験の少ないユーザにおいて，これから話そうというタイミングでふるまいを切り替えることが緊張感や不安感を軽減させる効果が生じたと考えられる．

5 むすび

本研究では，音声対話エージェントに対する話しかけにくさ，および対話の続けにくさといった課題を改善することを目的として，CG らしいふるまいと人間らしいふるまいをユーザの対話に対する積極性に応じて切り替える統合的なふるまい表現手法を提案した．

主観評価実験の結果，CG らしいふるまいが対話前の段階でエージェントの存在感を印象付け,「話しかけてほしい」という意図を伝える効果があること，また提案手法であるふるまいの切り替えが対話経験の少ないユーザにおいて緊張感や不安感を軽減する効果があることが示唆された．

今後は，ユーザの視線や表情，対話内容などのマルチモーダルな情報から積極性を自動推定するモジュールの設計や，より幅広い属性の参加者を対象とした実験を行う予定である．

参考文献

[1] Justine Cassell. Embodied conversational agents: airepresentation and intelligence in user interfaces. *AI magazine*, Vol. 22, No. 4, pp. 67–67, 2001.

[2] Aaron Powers, Sara Kiesler, Susan Fussell, and Cristen Torrey. Comparing a computer agent with a humanoid robot. In *Proceedings of the ACM/IEEE international conference on Human-robot interaction*, pp. 145–152, 2007.

[3] Angelo Cafaro, Hannes Högni Vilhjálmsson, and Timothy Bickmore. First impressions in human–agent virtual encounters. *ACM Transactions on Computer-Human Interaction (TOCHI)*, Vol. 23, No. 4, pp. 1–40, 2016.

[4] Kirsten Bergmann, Friederike Eyssel, and Stefan Kopp. A second chance to make a first impression? how appearance and nonverbal behavior affect perceived warmth and competence of virtual agents over time. In *International conference on intelligent virtual agents*, pp. 126–138. Springer, 2012.

[5] 川又 朱莉，上乃 聖，李 晃伸. 身体性を持つ CG 対話エージェントにおけるカートゥーン調表現の方法論および比較評価. HAI シンポジウム 2025, pp. G–10, 2025.

[6] Frank Thomas and Ollie Johnston. *The Illusion of Life: Disney Animation*. Hyperion, New York, 1981. Contains the ”12 Basic Principles of Animation”.

[7] 井上昂治, 高梨克也, 河原達也. 潜在キャラクタモデルによるリアルタイム対話エンゲージメント推定. 人工知能学会研究会資料 言語・音声理解と対話処理研究会 81 回 (2017/10), p. 22. 一般社団法人 人工知能学会, 2017.

[8] Jian Bi, Fang-chao Hu, Yu-jin Wang, Ming-nan Luo, and Miao He. A method based on interpretable machine learning for recognizing the intensity of human engagement intention. *Scientific Reports*, Vol. 13, No. 1, p. 2537, 2023.

[9] Akinobu Lee. MMDAgent-EX, December 2023.

[10] Lee Akinobu. CG Cybernetic Avatar model ”Gene”. https://github.com/mmdagent-ex/gene, December 2025.

[11] Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavey, and Ilya Sutskever. Robust speech recognition via large-scale weak supervision. In *International conference on machine learning*, pp. 28492–28518. PMLR, 2023.