

大規模言語モデルを用いた性格特性に基づく 対話ロボットの共発話ジェスチャー生成

池見侑亮¹ Baihaqi Muhammad Yeza^{2,3} Kruengkrai Canasai³ 中村泰³ 吉野幸一郎^{1,2,3}

¹ 東京科学大学 ² 奈良先端科学技術大学院大学

³ 理化学研究所ガーディアンロボットプロジェクト

ikemi.y.d23a@m.isct.ac.jp muhammad_yeza.baihaqi.lx2@naist.ac.jp

{canasai.kruengkrai,yutaka.nakamura}@riken.jp koichiro@c.titech.ac.jp

概要

人間とロボットのコミュニケーションにおいて、ロボットが一貫した性格を持つことは重要であり、その表現にはジェスチャーが大きな役割を果たす。しかし、速度などの物理パラメータを調整する既存のジェスチャー生成手法では、多様な性格表現に限界があった。そこで本研究では、大規模言語モデルを用いた共発話ジェスチャー生成手法を拡張し、指定された性格特性に基づいて共発話ジェスチャーを決定する手法を提案する。提案手法の有効性を検証するため、対話ロボット CommU を用いて提案手法による外向的 (Ext), 内向的 (Int), 及び先行研究による共発話ジェスチャー (baseline) を生成し、動画を用いた主観評価実験を行った。実験の結果、Ext 条件は Int 条件と比較して有意に高い外向性スコアを示し、提案手法によって知覚される外向性の強弱を制御可能であることが示された。一方で、Int 条件では baseline 条件と比較して動作の自然さやエンゲージメントの評価値が低下しており、性格特性の反映によるジェスチャーの質の低下といった課題も明らかとなった。

1 はじめに

人間とロボットのコミュニケーションにおいて、ロボットの性格は重要な役割を果たす。近年、ロボットは身近な存在となりつつあり、ロボットが社会の一員として人間とコミュニケーションをとるうえで彼らの一貫した性格は不可欠である [1]。多くの研究において、ロボットの性格に応じて人間とのコミュニケーションの質が変化することが示されている [1]。

ロボットに性格を持たせようとした場合、その性

格をどのように表現するかが課題となる。ロボットの性格表現においては、言語情報だけでなく、非言語情報であるジェスチャーが重要な役割を果たす。人間やバーチャルエージェントにおいて、動きの頻度や空間の使い方によって知覚される性格が異なることが示されている [2, 3]。ロボットにおいても同様に、動作の速度や大きさが知覚される性格に影響を与えることが示されている [4]。

これまで様々なジェスチャー生成手法が提案されているが、性格表現を伴う生成には依然として課題が残る。多くの既存研究では、ルールベースやデータ駆動の手法を用いたジェスチャー生成に加えて、速度などの物理パラメータを調整することで性格表現を実現している [5, 6, 7]。しかし、これは性格に応じてパラメータ調整のためのモデルを構築する必要があり、多様な性格を表現するには限界がある [7]。

そこで本研究では、大規模言語モデル (LLM) を用いた共発話ジェスチャー生成手法 [8] を拡張し、性格特性を反映する共発話ジェスチャー生成手法を提案する。LLM はロボットの社会的行動を含む、高度な行動計画能力を有しており [9]、ロボットの発話に合ったジェスチャー計画及び、それを実現する関節角度生成にも利用できる [8]。本研究では [8] で提案された手法を改良し、LLM を用いた動的かつ多様な性格の共発話ジェスチャー生成を実現する。

2 関連研究

2.1 ジェスチャー生成

ジェスチャー生成手法はこれまで、主にルールベース手法とデータ駆動手法が用いられてきた。ルールベース手法は言語情報から意味的に適切な動作を選択可能だが、ルール設計コストが高く、動

作の多様性に欠ける課題がある。一方、データ駆動手法は自然な動作生成が可能だが、特定の文脈や意味内容に基づいた制御が難しく、異なるロボットや状況への適応には再学習が必要となる [2]。

これらの課題を解決するために LLM の利用が注目されている。[9] は、LLM が対話の社会的文脈を理解し、ロボットの行動計画を適切に決定できることを示した。さらに、この能力を具体的な身体動作の生成へと拡張する試みも行われており、[8] は対話ロボット CommU の関節角度系列を発話内容から直接生成している。[10] はマスコットロボットの感情表現動作を生成することに成功している。これらの研究は、LLM が言語や感情に関する文脈を解釈し、それをもとに動作を生成する能力を有することを実証している。

2.2 性格特性を反映したジェスチャー生成

ロボットやエージェントの性格表現に関する研究は数多く行われている。特に、動作の速度、頻度、空間的広がりといった非言語特徴と、観察者が知覚する性格との関係性については詳細な調査がなされており [11, 3, 4, 12]、これらの知見はジェスチャーによる性格表現の基礎となっている。

これらの知見を応用し、[5, 6] らはルールベースにより選択されたジェスチャーに性格による調整を加えることで意図した性格を知覚させることに成功している。しかし、ルールベース手法に依存した生成のため、ジェスチャーの多様性に欠ける課題がある。また、[7] はデータ駆動手法により生成されたジェスチャーに対して調整を行う手法を提案したが、外向性や低い神経症傾向など、一部の性格特性の表現においては十分な効果が得られないことが報告されている。

3 提案手法

本研究では、[8] で提案された、LLM 駆動のジェスチャー生成手法である LLM generating joint angles (LLM-GJA) を拡張し、性格特性を反映可能な生成手法を提案する。[8] では発話テキストの長さに応じてジェスチャー数を決定していたが、性格特性は動作の頻度やタイミングに強く影響する。そこで本手法では、入力として目的の性格 q を受け付け、ジェスチャーの動作内容だけでなく、頻度も LLM に動的に決定させる。

提案手法の処理の流れを図 1 に示す。まず発話

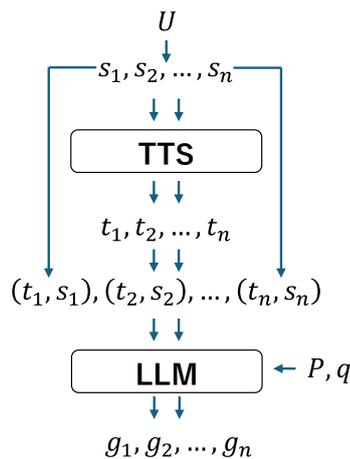


図 1 提案手法

内容 U を並列化のために文単位 s_1, s_2, \dots, s_n に分割する。

$$U = (s_1, s_2, \dots, s_n). \quad (1)$$

次に、各 s_i に text-to-speech (TTS) 変換を行い、発話時間 t_i を得る。タスクを表すプロンプト P 、目標の性格 q と (t_i, s_i) を入力として文 s_i のジェスチャー g_i が生成され、発話全体のジェスチャー (g_1, g_2, \dots, g_n) を得る。ここで、生成される g_i は以下のような停止時間 t_{stop} 、動作時間 t_{move} 、目標関節角度 a の組の系列である。

$$g_i = ((t_{\text{stop},1}^{(i)}, t_{\text{move},1}^{(i)}, a_1^{(i)}), \dots, (t_{\text{stop},m}^{(i)}, t_{\text{move},m}^{(i)}, a_m^{(i)})) \quad (2)$$

ジェスチャーは t_{move} だけ静止後、 t_{move} ほどの時間をかけて目標関節角度 a へと動作することの繰り返しにより実行される。

4 実験

4.1 動作環境

本提案手法の有効性を検証するため、対話ロボット CommU¹⁾を用いたジェスチャー生成実験を行った。CommU の関節は頭部、眼球、まぶた、胴体、腕部、口に合計 14 の自由度を有する。本実験では、発話に合わせて自動開閉する口を除く 13 自由度を LLM による生成対象とした。LLM には gpt-oss-120b[13] を採用し、TTS には Google 翻訳のテキスト読み上げ機能へのインターフェースを提供する Python ライブラリである gTTS²⁾を利用した。

1) <https://www.vstone.co.jp/products/commu/index.html>

2) <https://github.com/pndurette/gTTS>

4.2 テストケース

テストケースは [8] で用いられた, disagreeing, emphasizing, greeting, questioning, suggesting の発話からそれぞれ1つずつ選んだ計5つの発話内容をもとにジェスチャー生成を行った。詳細は付録 A.1 に記載されている。

4.3 ユーザー実験

提案手法が意図した性格特性を反映した動作を生成できるか、および生成された動作が対話ロボットとしての自然さを損なわないかを検証するため、アンケートを用いた主観評価実験を行った。比較対象として、[8]を参考にしたベースライン (baseline) と、提案手法による外向的 (Ext), および内向的 (Int) の計3条件を設定した。これら3条件に対し、前述した5つの発話内容をもとに合計15本の共発話ジェスチャーの動画を撮影した。今回は純粋にジェスチャーを評価するため、発話内容は同一とする。

被験者には著者らの所属する研究室のメンバー計2名を対象とした。実験は Google フォームを用いた。フォームの各セクション上部に動画が配置され、被験者は動画を視聴した後、その直下に続く質問項目に回答した。評価指標には、[8]で用いられたジェスチャーの質に関する自然さ (NA), 時間的一貫性 (TC), 意味的一貫性 (SC), エンゲージメント (EN) に関する12項目に加えて、ロボットから知覚される性格特性を測定するための TIPI-J [14] の10項目を7段階リッカート尺度で収集した。

4.4 分析手法

収集したデータに対し、反復測定分散分析 (repeated measures ANOVA) を実施し、条件間 (baseline, Int, Ext) に統計的な有意差が存在するかを確認した。さらに、主効果が認められた項目については Bonferroni 補正を用いた事後検定を行い、具体的にどの条件間で評価値に差が生じているかを検証した。

5 結果

5.1 ジェスチャーの質に関する分析

提案手法が、ジェスチャーの自然さや一貫性といったジェスチャーの質に及ぼす影響を検証した。各評価指標について反復測定分散分析および

Bonferroni 補正を用いた事後検定 (Post-hoc test) を行った結果を表 2 に示す。データの詳細な統計量及び分布については表 1, 図 2 を、生成されたジェスチャーの様子については付録 A.2 を参照されたし。

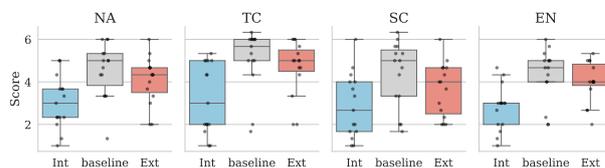


図 2 各手法における評価指標ごとの分布

ANOVA により全指標において有意差が確認されたが、事後検定では TC と SC においては、有意水準 ($p < .05$) には達しなかった。NA と EN においては、baseline 条件と Int 条件間に有意差が確認され、表 1 から Int 条件は baseline 条件に比べてジェスチャーの質が低下していることが確認できる。対して、Ext 条件と baseline 条件間には差が見られなかった。

5.2 性格に関する分析

条件ごとに知覚された性格は図 3 の通りである。特に、知覚された外向性スコアに対する反復測定分散分析の結果、手法間に有意な主効果が確認された ($F(2, 4) = 56.46, p < .05$)。外向性スコアの詳細については表 1 を参照されたし。

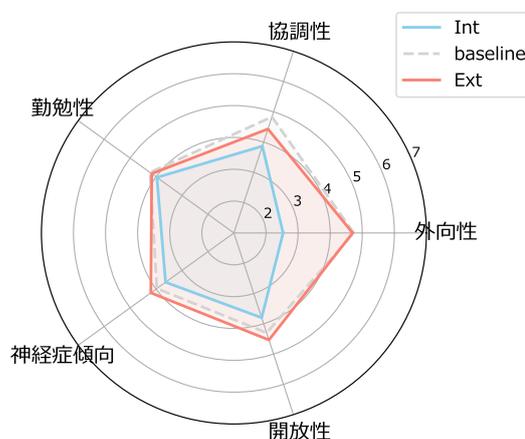


図 3 生成手法ごとの知覚される性格の平均スコア

続いて Bonferroni 法による多重比較を行った結果、Ext 条件は Int 条件と比較して有意に高い外向性スコアを示した ($p < .05$)。このことから、外向性の強弱の制御が可能であることが統計的に示された。baseline と Ext の間には有意差は認められず、Int は baseline に対して有意傾向 ($p = .054$) で低い外向性スコアを示した。

表1 各シナリオおよび全体における外向性, NA, TC, SC, EN に関する結果. 括弧の外は平均値, 内は標準偏差. 太字はシナリオ全体での手法間での最大スコア.

Method	Scenario	外向性	NA	TC	SC	EN
baseline	disagreeing	2.67 (0.764)	3.67 (0.577)	4.22 (2.27)	3.11 (2.22)	2.78 (1.07)
	greeting	5.17 (1.04)	5.00 (0.333)	5.78 (0.385)	4.89 (1.39)	4.44 (0.509)
	emphasizing	5.83 (0.577)	5.78 (0.385)	6.11 (0.192)	5.67 (0.667)	5.56 (0.509)
	questioning	4.67 (0.289)	3.00 (1.53)	3.78 (1.58)	3.56 (1.68)	3.89 (1.64)
	suggesting	5.33 (0.764)	5.33 (0.577)	5.56 (0.509)	4.67 (0.333)	5.00 (0.333)
	total	4.73 (1.29)	4.56 (1.28)	5.09 (1.43)	4.38 (1.54)	4.33 (1.27)
Ext	disagreeing	4.33 (1.76)	3.89 (1.64)	4.11 (1.84)	3.89 (1.64)	3.78 (1.68)
	greeting	5.00 (0.866)	4.67 (1.20)	5.22 (0.694)	4.00 (2.00)	4.33 (0.577)
	emphasizing	4.50 (0.866)	4.00 (0.88)	5.00 (0.667)	3.56 (1.17)	4.22 (0.694)
	questioning	4.50 (1.00)	4.33 (1.00)	4.78 (1.35)	3.78 (1.02)	3.89 (1.17)
	suggesting	5.17 (0.577)	3.56 (1.39)	4.22 (2.04)	3.11 (1.17)	3.78 (1.02)
	total	4.70 (0.978)	4.09 (1.13)	4.67 (1.29)	3.67 (1.27)	4.00 (0.959)
Int	disagreeing	3.20 (0.85)	1.56 (0.694)	2.00 (1.00)	1.89 (0.839)	1.44 (0.509)
	greeting	3.10 (0.90)	3.22 (0.77)	3.67 (1.76)	2.22 (1.26)	2.89 (0.192)
	emphasizing	3.30 (0.80)	2.78 (0.694)	3.00 (1.73)	4.11 (2.01)	2.67 (0.577)
	questioning	3.20 (0.85)	4.78 (0.385)	5.11 (0.192)	4.00 (0.667)	4.33 (0.333)
	suggesting	3.20 (0.85)	2.78 (0.77)	2.44 (1.71)	2.22 (1.58)	2.56 (0.77)
	total	2.53 (0.767)	3.02 (1.22)	3.24 (1.64)	2.89 (1.52)	2.78 (1.05)

表2 ジェスチャーの質に関する評価指標に対する反復測定分散分析および事後検定の結果

Metric	ANOVA		Post-hoc (p_{adj})		
	$F(2, 4)$	p	vs. Int.	vs. Ext.	Int. vs. Ext.
NA	86.3	.0005*	.0131**	.235	.0346**
TC	12.8	.0182*	.107	.844	.287
SC	7.44	.0449*	.247	.103	.720
EN	49.7	.0015*	.0224**	.306	.0945

(† : $p < .10$, * : $p < .05$, ** : $p < .05$, ns : not significant)

Post-hoc 列の “vs.” は baseline との比較を示す.

表3 外向性スコアに対する事後検定の結果.

Comparison Pair	P_{adj}
Int vs. baseline	.0543
Int vs. Ext	.0465**
baseline vs. Ext	1.000

(** : $p < .05$)

6 考察

実験結果より Ext 条件と baseline 条件の間には外向性スコアに有意差が見られなかった. これは, baseline 条件が元来外向的なジェスチャーを生成する傾向にあることが示唆され, 生成時のプロンプトで用いられた one-shot の具体例などの影響について調査が必要だと考える.

Int 条件は結果から有意傾向に低い外向性を示すジェスチャーを生成したが, 同時にジェスチャーの質も有意に低下した. これは, LLM が内向的な性格をより控えめな動作で表現しようとする傾向による

ものと考えられる. また, 表1によると, questioning の発話内容については Int 条件が高い質を示しており, 発話内容と性格のミスマッチの影響についても今後の調査が必要だと考える.

図3によると, 外向性以外にも開放性や協調性のスコアに差が見られる. TIPI-J では各因子間の相関関係は中程度以下であることから [14], これは生成手法に起因するものと考えられ, [7] でも確認された意図しない性格への影響は今後の課題である.

7 おわりに

本研究では, LLM を活用して対話ロボットの性格特性を反映した共発話ジェスチャーを生成する手法を提案し, 対話ロボット CommU を用いた主観評価実験によりその有効性を検証した. 実験の結果, Ext 条件は内向的 Int 条件と比較して有意に高い外向性スコアを示し, 提案手法によって外向性の強弱を制御可能であることが実証された. 内向的条件では baseline 条件と比較してジェスチャーの質が低下する傾向が見られ, 性格特性反映とジェスチャーの質の維持とのトレードオフが課題となった. 今後は, 内向的な性格表現における品質の向上や, 意図しない他の性格特性への影響の制御, ジェスチャーに加えて発話内容も考慮した性格表現の実現が求められる.

参考文献

- [1] Kwan Min Lee, Wei Peng, Seung-A Jin, and Chang Yan. Can robots manifest personality?: An empirical test of personality recognition, social responses, and social presence in human–robot interaction. **Journal of communication**, Vol. 56, No. 4, pp. 754–772, 2006.
- [2] Simbarashe Nyatsanga, Taras Kucherenko, Chaitanya Ahuja, Gustav Eje Henter, and Michael Neff. A comprehensive review of data-driven co-speech gesture generation. In **Computer Graphics Forum**, Vol. 42, pp. 569–596. Wiley Online Library, 2023.
- [3] Funda Durupinar, Mubbasir Kapadia, Susan Deutsch, Michael Neff, and Norman I Badler. Perform: Perceptual approach for adding ocean personality to human motion using laban movement analysis. **ACM Transactions on Graphics (TOG)**, Vol. 36, No. 1, pp. 1–16, 2016.
- [4] Marieke van Otterdijk, Heqiu Song, Konstantinos Tsiakas, Ilka van Zeijl, and Emilia Barakova. Nonverbal cues expressing robot personality - a movement analysts perspective. In **2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)**, pp. 1181–1186, 2022.
- [5] Maryam Saberi, Steve DiPaola, and Ulysses Bernardet. Expressing personality through non-verbal behaviour in real-time interaction. **Frontiers in psychology**, Vol. 12, p. 660895, 2021.
- [6] Sinan Sonlu, Uğur Güdükbay, and Funda Durupinar. A conversational agent framework with multi-modal personality expression. **ACM Transactions on Graphics (TOG)**, Vol. 40, No. 1, pp. 1–16, 2021.
- [7] Sinan Sonlu, Halil Özgür Demir, and Uğur Güdükbay. Personality expression using co-speech gesture. **ACM Transactions on Applied Perception**, Vol. 22, No. 2, pp. 1–20, 2025.
- [8] Muhammad Yeza Baihaqi, Angel García Contreras, Seiya Kawano, and Koichiro Yoshino. Llm-driven approach for motion control in human-robot dialogue for elevating engagement. In **2025 34th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)**, pp. 544–550, 2025.
- [9] Karthik Mahadevan, Jonathan Chien, Noah Brown, Zhuo Xu, Carolina Parada, Fei Xia, Andy Zeng, Leila Takayama, and Dorsa Sadigh. Generative expressive robot behaviors using large language models. In **Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction**, pp. 482–491, 2024.
- [10] Yoongu Lim, Jaeuk Cho, Duk-Yeon Lee, Dongwoon Choi, and Dong-Wook Lee. Design and implementation of a companion robot with llm-based hierarchical emotion motion generation. **Applied Sciences**, Vol. 15, No. 23, p. 12759, 2025.
- [11] Ziya Erkoç, Serkan Demirci, Sinan Sonlu, and Uğur Güdükbay. Skeleton-based personality recognition using laban movement analysis. In **Understanding social behavior in dyadic and small group interactions**, pp. 74–87. PMLR, 2022.
- [12] Bart GW Craenen, Amol A Deshmukh, Mary Ellen Foster, and Alessandro Vinciarelli. Shaping gestures to shape personality: Big-five traits, godspeed scores and the similarity-attraction effect. In **AAMAS**, pp. 2221–2223, 2018.
- [13] OpenAI. gpt-oss-120b & gpt-oss-20b model card, 2025.
- [14] 小塩真司, 阿部晋吾. 日本語版 ten item personality inventory (tipi-j) 作成の試み. **パーソナリティ研究**, Vol. 21, No. 1, pp. 40–52, 2012.

A 付録

A.1 テストケース

表4 使用したテストケース

disagreeing	確かに、バルセロナFCの試合のチケットは高いかもしれませんが、でも、一生に一度の経験として、それだけの価値があります。忘れられない体験になるでしょう。ぜひ買うべきです。
emphasizing	そうですよ！ゴルナーグラート山頂からの眺めは、本当に息を呑むほど美しいです。晴れた日には、マッターホルンをはじめとするスイスの山々が一望できます。そして、山頂にはレストランもありますよ。
greeting	こんにちは！私はCommU、小さな雑談ロボットです。この実験に参加してくれてありがとうございます。あなたとお話できるのを楽しみにしています。
questioning	えええ、そうなんだ！アイスランドのレストランでは、熱々のラムスープや魚料理がとてもおいしいです。香川さんはスープやシチュー系の料理が好きですか？私は好きですよ。
suggesting	ブルーラグーンという広大な温泉がすぐ近くにあります。あそこで温かい湯に浸かりながらリラックスするのは、まさに至福のひとつです。香川さんもきっと気に入って、何度でも訪れたいと思いますよ。

A.2 生成されたジェスチャー

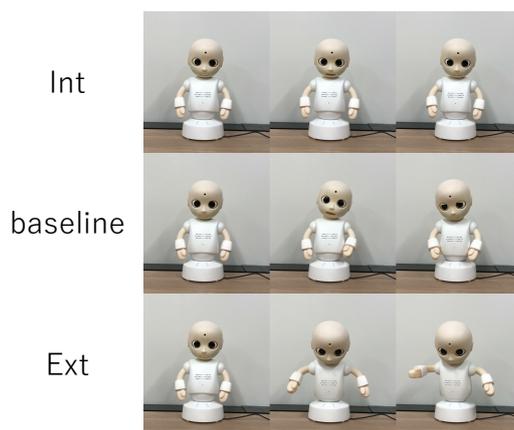


図4 生成されたジェスチャーの一例。初期ポーズから、「確かに、バルセロナFCの試合のチケットは高いかもしれませんが。」という発話に合わせたジェスチャーを行っている様子。