

対話エージェントの相槌における態度表出に関する有効性の検証

A Study on the Effectiveness of Attitude Expressions in the Backchannel of Conversational Agent

片山 晋^{1*} 馬場 惇² 中西 惇也³ 吉川 雄一郎³

Shin Katayama¹ Jun Baba² Junya Nakanishi³ Yuichiro Yoshikawa³

¹ 名古屋大学 工学研究科

¹ Graduate School of Engineering, Nagoya University

² 株式会社サイバーエージェント

² CyberAgent, Inc.

³ 大阪大学 基礎工学研究科

³ Graduate School of Engineering Science, Osaka University

Abstract: 近年、ホテルや商業施設で接客を行う対話エージェントが活用されている。しかし、これらのエージェントは利用者に対して常に同調的な態度を示していることが多いため、長時間使用すると飽きられてしまう問題が存在する。そこで本研究では、エージェント自身が所有する態度に着目し、対話中の相槌において同意・不同意の態度表出を行う対話エージェントを提案する。被験者実験を行い、相槌における態度表出が利用者のエンゲージメントに与える影響を調査する。

1 はじめに

近年、対話可能なロボットが日常生活に普及しており、ホテルや商業施設においても接客業務を行う対話エージェントが活用されている。接客対話エージェントは一般的に、店舗案内や商品推薦などのタスクの遂行を求められるが、ホテルの部屋などの環境においては利用者の満足度を向上させるための雑談やおもてなしといった行動も求められる。しかし、雑談を行う接客対話エージェントは利用者に対して常に同調的な態度を示していることが多く、長時間使用すると発言の内容や対応がパターン化してしまうため、対話相手として認められず継続的な利用がされないことが課題となっている。満足度の高い接客体験をより多く提供するためには、利用者に継続的に利用したいと思わせる対話エージェントの実現が求められる。

そこで本研究では、エージェント自身が持つ態度の表出に着目し、利用者との対話において態度表出を行う対話エージェントを提案する。従来のエージェントのように常に同調するのではなく、エージェント自身が所有する同意・不同意などの態度を表出することで次に起きるエージェントの態度や発言に興味を持たせ、継続的な使用へ繋がると考えられる。また、相槌時に態度表出を行うことにより、意思を主張せずともエージェ

ントが所有する態度の伝達が可能になるため、円滑な対話の実現や利用者の興味などのエンゲージメントが上昇すると考えられる。

本研究における仮説は以下の2点である。

- 仮説 1: 常に同調的な態度を取るのではなく、同意・不同意などの態度を表出することでエージェントの意思表示を表現し、利用可能性を高めるのではないかと？
- 仮説 2: 態度の表出を行うことでエージェントに対する興味や注意などのエンゲージメントが上昇するのではないかと？

これらの仮説に基づき、相槌時に態度表出を行う質問応答対話エージェントのプロトタイプを実装する。また、被験者による対話実験を行うことによって、相槌における態度表出の有効性を検証する。

2 関連研究

人間とエージェントの円滑な対話コミュニケーションの実現において、相槌を用いた手法が提案されている。中でも、相槌を打つタイミングを予測する手法が多く提案されており、機械学習を用いて相槌のための発話終了タイミングを予測する研究 [1] や、深層学習を

*連絡先: 名古屋大学
〒464-8601 愛知県名古屋市千種区不老町
E-mail: shinsan@ucl.nuee.nagoya-u.ac.jp



図 1: Sota を用いた対話状況

用いて音響的な入力特徴に基づいて相槌のタイミングを予測する研究 [2] が行われている。

また、対話に応じて相槌のパターンを変更する研究も行われている。Kawahara ら [3] は言語情報と韻律情報を考慮して適切な相槌を選択する手法を提案し、自然な相槌を選択することで理解や共感の印象を与えることを示している。石田ら [4] は傾聴対話システムにおいて、ユーザ発話の音声認識や焦点解析を行い共感のための適切な応答を選択する手法を提案している。

これらの傾聴対話による手法では、利用者を怒らせないことや利用者の話のペースに合わせるなどの同調的な態度を取ることが求められている。しかし、意図なく常に同調の反応を行うと対等な対話相手として認められず、継続的に使用する際に飽きられてしまう課題が存在する。そのため、本研究では対話エージェント自身が所有する態度を表出することで、継続的な対話の実現を目指す。

3 実装

本プロトタイプでは、質問応答対話エージェントを実装する。利用者の行きたい場所に関する対話シナリオを想定し、エージェントが投げかけた質問に対して利用者が応答、その応答に対してエージェントが態度を表出しながら相槌や発話を行う。本研究における対話エージェントを実装するにあたり、Vstone 社のコミュニケーションロボット Sota¹(図 1)を使用する。Sota はプログラムを通して指定したテキストの発話やジェスチャー動作を行うことが可能である。利用者の発話内容は MacBook Pro に内蔵されたマイクを用いて収集し、Google Cloud Speech API によるリアルタイム音声認識を用いて発話音声を変換を行う。エージェントの態度表出は、発話内容の制御と音声合成による韻律やピッチの変化、首の傾き・傾げによるジェス

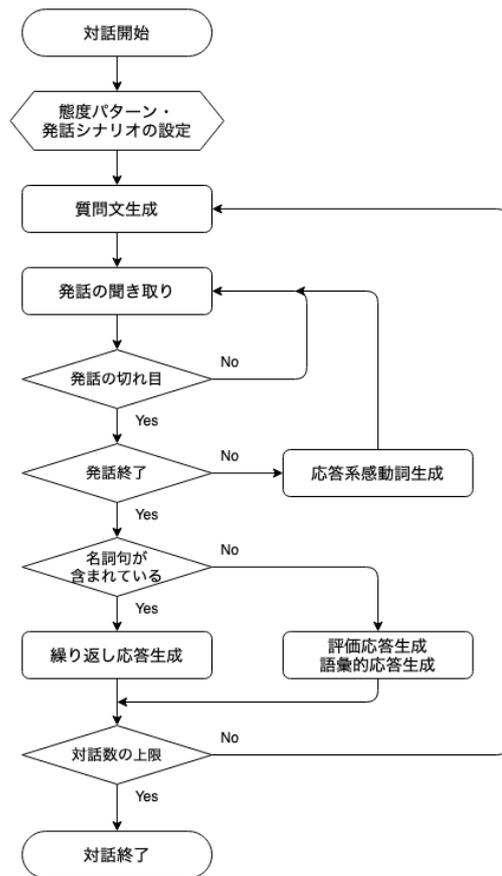


図 2: 対話時における対応手順のフローチャート

チャー動作によって実現する。対話におけるエージェントの全体的な対応手順を図 2 に記述する。質問応答対話の流れは以下の通りである。

1. エージェントが質問文生成を行う。
2. 利用者の応答を聞き取り、対話途中で相槌生成を行う。
3. 利用者の発話が終了した時点で、それに対する相槌生成を行う。

これらの手順を制限時間または質問数の上限に達するまで繰り返し行う。生成された質問文や相槌は、プログラムを介して Sota に随時発話させる。

3.1 質問文生成

本プロトタイプは、利用者の行きたい場所についての雑談をするシナリオを想定し、対話を行う。行きたい場所については、行きたい国、食べに行きたい料理、行きたいイベントの 3 つを設定する。質問文は以下の順番で生成される。

¹<https://www.vstone.co.jp/products/sota/>

表 1: 相槌の例

	中立	同意	不同意
応答系 感動詞	うん はい うんうん	うん! はい! うんうん!	うん... うーん んん?
語彙的応答 評価応答	なるほど そうなんです そういうことなんです	いいね! わかります! いいですね!	そっかあ そうなのかあ いやあ〜

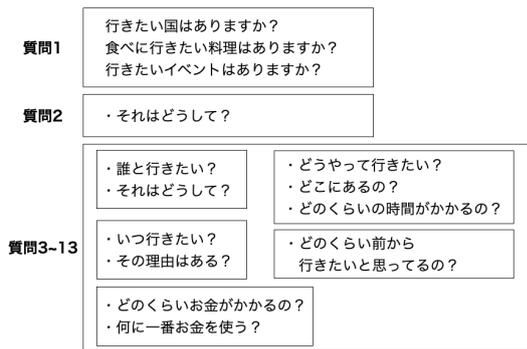


図 3: 生成される質問文の一覧

- 質問 1: 行きたい場所についての質問を行う。
- 質問 2: どうして行きたいか理由を質問する。
- 質問 3-13: 誰と行きたい, いつ行きたいなどについてランダムな順序で質問する。

質問文は図 3 に記載されている合計 13 件で構成され, 全ての質問文が発話された時点で対話が終了する。

3.2 相槌生成

本システムにおける相槌の種類は, 吉田ら [5] によって提案された相槌の種類から, 以下の 4 種類を採用した。

- 応答系感動詞
「ああ」「うん」「ええ」「おお」など
利用者が対話をしている途中でも切れ目が生じた場合に生成を行う。話の切れ目は音声認識の中間結果や最終結果が 0.4 秒間返ってこなかった場合とする。
- 繰り返し応答
他の話者の発話の一部, もしくは全体を繰り返す反応 「〇〇ですか」
形態素解析を行い, 名詞句が存在する場合はその名詞句を用いて繰り返し応答を行う。

● 語彙的応答

「なるほど」「確かに」「そう(ですね)」など態度が中立の場合に「なるほど」などの語彙的応答を行う。

● 評価応答

「おもしろい(な)」「すごい」「こわ」など同意や不同意などの態度が設定された場合に, それらの態度に基づいた評価応答を行う。

3.3 態度表出

対話中におけるエージェントの態度として, 中立(ニュートラル), 同意, 不同意の 3 種類を設定する。応答系感動詞, 語彙的応答・評価応答の内容を事前設定したテキストからランダムに選択し, 相槌発話時や質問発話時の音声のトーンやピッチなどを態度に基づいて変更する。音声のトーンやピッチを変更することで同意・不同意の態度を強調する。表 1 に態度別における相槌の内容を示す。

4 被験者実験

エージェントが対話中に態度を表出することで利用可能性やエンゲージメントが向上する仮説について, 被験者実験による検証を行う。被験者は男女 18 名(男性 8 名, 女性 10 名)の大学生・大学院生であり, 一人につき約 5 分間の対話を 3 回, 合計約 15 分間の対話を行う。

4.1 実験手順

被験者には, プロトタイプとして実装した対話エージェントと (A) 常に中立, (B) 常に同意, (C) 前半に不同意・後半に同意, の 3 パターンの態度別に対話をしてもらう。対話トピックは (1) 行きたい国, (2) 行きたいイベント, (3) 食べに行きたい料理の三つであり, 対話パターンとトピックの順序はカウンターバランスを

表 2: 対話内容の例

発話者	発話内容 (中立シナリオ)
ロボット	何を食べに行きたいですか？
ユーザ	チャーハンを食べに行きたいです.
ロボット	チャーハンですか (繰り返し応答)
ロボット	それはどうして？
ユーザ	近所にある中華料理屋の
ロボット	うん (応答系感動詞)
ユーザ	チャーハンが美味しいんだよね
ロボット	なるほど (語彙的応答)
発話者	発話内容 (同意シナリオ)
ロボット	どのイベントに行きたいですか？
ユーザ	肉フェスです.
ロボット	肉フェスですか！ (繰り返し応答)
ロボット	それはどうして？
ユーザ	美味しいお肉が好きだから
ロボット	うんうん！ (応答系感動詞)
ユーザ	食べに行きたいんだよね
ロボット	いいですね！ (評価応答)
発話者	発話内容 (不同意シナリオ)
ロボット	どこの国に行きたいですか？
ユーザ	ベトナムです.
ロボット	ベトナムですかー (繰り返し応答)
ロボット	それはどうして？
ユーザ	前に修学旅行で一度行ったのが
ロボット	うーん (応答系感動詞)
ユーザ	楽しかったからもう一度行きたいな
ロボット	そっかあ (評価応答)

取って行う。(C) 前半に不同意・後半に同意を設定した理由として、本研究の目的である常に同じ態度を取るエージェントではなく、話の内容によって態度を変化させるエージェントの効果を確認するためである。実験手順は以下の通りである。

1. 最初の約2分半、1つ目の「行きたい〇〇」について対話を行う。
2. 途中でロボットが「それではもう一つについてもお聞きします。」と話したら、2つ目の行きたい〇〇) について約2分半、対話をしてもらう。
3. 対話後に対話の印象等に関するアンケート記入してもらう。

以上の過程を3回繰り返す被験者内実験を行い、対話後に口頭のインタビューを実施する。それぞれの態度表出のシナリオにおける実際の対話内容例を表2に示す。

4.2 評価指標

同意・不同意の態度を表出することの有効性を評価するために、対話の質や態度表出の印象を問うアンケートを実施する。Q1, Q2, Q4, Q7, Q11, Q13に関しては、傾聴対話システムにおける印象評価を行った井上ら [6] のロボットの対話の質に関するアンケート項目を参考する。それらに加え、態度表出に関する質問などを設置し、合計17つのアンケート項目を7段階のリッカート尺度によって定性的な評価を行う。また、対話中にロボットの後方に設置したカメラにより撮影された動画を用いて、目線の逸らした回数や目を合わせた秒数などのエンゲージメントを定量的な評価を行う。

5 実験結果

アンケートの内容及び評価結果を表3に、動画におけるエンゲージメント評価結果を表4に示す。

5.1 アンケート評価

常に同意を行う (B) と前半に不同意・後半に同意を行う提案手法 (C) の間に有意差があるかを確認するため、ノンパラメトリックな対応あり検定であるウィルコクソンの符号付順位検定を行った。(B) と (C) の標本を用いて検定を行った理由として、常に同調的な態度を取るのではなく、同意・不同意などの態度を表出することでエージェントの意思表示を表現し、利用可能性を高めるのではないかという仮説1を検証するためである。

「Q8. ロボットは自分の態度を表出していた」や「Q9. ロボットの反応から態度が読み取れた」といった指標において (B) の手法と比較して高い平均値となった。また、Q8($p = 0.04$) と Q9($p = 0.04$) の両方に有意差が認められたことから、提案手法が態度を表出し、伝達することが可能であったことが伺える。

「Q3. ロボットの反応に生命感を感じた」や「Q7. ロボットは話を理解していた」の指標においても、提案手法 (C) の平均が高い数値になり、Q3($p = 0.02$) と Q7($p = 0.03$) の両方にそれぞれ有意差が認められたため、不同意の態度を表出することで、全てにおいて同意するよりも話を理解していると思わせること、生命感を感じさせることが可能であったと考えられる。一方で、「Q11. ロボットとの対話に満足した」の指標では提案手法の評価平均値が低くなり有意差も認められなかったため、不同意になることで対話の満足度は少し下がったと考えられる。

有意差は認められなかったが、「Q12. ロボットに感情を揺さぶられた」、「Q15. ロボットの気持ちを考えて対

表 3: 7 段階のアンケートにおける質問内容と平均値

質問項目	A(中立)	B(同意)	C(不同意→同意)
Q1 ロボットが話した言葉は自然だった	4.277	3.888	4.166
Q2 ロボットはタイミングよく反応していた	4.555	3.777	3.111
Q3 ロボットの反応に生命感を感じた	3.666	3.055	3.666*
Q4 ロボットの反応の頻度は適切だった	4.555	4.222	4.055
Q5 ロボットは対話相手として満足だった	3.277	2.888	3.277
Q6 ロボットは話に興味を示してくれた	4.722	4.222	4.333
Q7 ロボットは話を理解していた	3.555	2.888	3.611*
Q8 ロボットは自分の態度を表出していた	3.444	3.444	4.166*
Q9 ロボットの反応から態度が読み取れた	3.666	3.388	4.222*
Q10 ロボットの相槌と発話の態度が一貫していた	5.055	4.166	3.777
Q11 ロボットとの対話に満足した	3.555	3.444	3.444
Q12 ロボットに感情を揺さぶられた	2.944	2.666	3.166*
Q13 ロボットとまた対話したいと思った	3.500	3.277	2.944
Q14 ロボットに意見を求めたいと思った	3.111	2.888	3.000
Q15 ロボットの気持ちを考えて対話を行なった	3.666	3.722	3.888
Q16 ロボットの気持ちを知りたいと思った	4.222	3.944	4.277
Q17 今後使用する時、このロボットとの対話は飽きないと思う	3.000	2.944	3.055

太字は最大平均値

* $p < 0.05$

表 4: 動画解析による 1 分あたりのエンゲージメント数の平均値

	A(中立)	B(同意)	C(不同意→同意)
目を逸らした回数(合計/分)	前半	5.47	5.50
	後半	6.28	5.85
目を合わせた秒数(合計/分)	前半	38.67	36.04
	後半	35.05	36.09

話を行った」,「Q16. ロボットの気持ちを知りたいと思った」,「Q17. 今後使用する時、このロボットとの対話は飽きないと思う」の指標においても、提案手法(C)が高い平均値を示したため、ロボットの気持ちを考えたりさせることで、飽きない対話を実現する手掛かりになったと考えられる。

5.2 エンゲージメント評価

対話実験中に撮影した動画を用いて、目を逸らした回数と目を合わせた時間について一人のアノテーターに注釈してもらった。被験者の合計対話秒数に誤差が発生したため、1分あたりの回数として計算を行った。被験者 18 人の動画を解析した結果、提案手法(C)の後半部分である同意の態度を表出する部分において、目を合わせた秒数の平均値が増加した。有意差は認められなかったが、不同意から同意に移り変わることで意図・態度を持っている印象を与え、より注意深く観察す

るようになったと考えられる。一方で、提案手法(C)によって目を逸らした回数が増える事象は確認できなかった。しかし、提案手法(C)の後半部分である同意では目線を逸らしている回数が増えているが目を合わせた秒数も増えているため、目線が外れても再度目線を合わせる行為が行われていたと考えられる。

5.3 考察

アンケート評価結果から、「Q13. ロボットとまた対話したいと思った」,「Q14. ロボットに意見を求めたいと思った」といった指標において提案手法に有意差が認められなかった。これについて被験者にインタビューを行ったところ、不同意されることで、そもそも対話の意欲がなくなってしまうという点や、意見を求めても不同意されることが予想できるといった点を指摘された。そのため、不同意の態度を表出する場合は会話内容などに応じて流動的に態度を変更し、一貫したルー

ルに基づいて態度の表出を行うことが重要であると考えられる。

また、本プロトタイプを用いた発話において、態度を表出するものが相槌内容や韻律だけのため、対話の内容が乏しく、中立・同意・不同意の態度を明示的に表出できなかった可能性があった。同じ質問を何度もしてくるロボットに対して飽きている被験者も多く、なぜその態度になっているのか理由を述べることであれば、明示的な態度の表出に繋がると考えられる。

6 おわりに

本研究では、エージェント自身が所有する態度に着目し、対話中の相槌において同意・不同意の態度表出を行う対話エージェントを提案した。被験者18名による対話実験を行い、相槌における態度表出が利用者のエンゲージメントに与える影響を調査した結果、相槌で不同意と同意の態度を表出するエージェントの対話が話の理解を促し、生命感を感じさせ、目を合わせる時間が増加した。今後は発話者の対話内容やこれまでの流れに応じてエージェント自身が流動的に態度の表出を変化させることでどのような影響を与えるのかを調査する。

参考文献

- [1] Divesh Lala, Pierrick Milhorat, Koji Inoue, Masanari Ishida, Katsuya Takanashi, and Tatsuya Kawahara. Attentive listening system with backchanneling, response generation and flexible turn-taking. In *Proceedings of the 18th Annual SIGdial Meeting on Discourse and Dialogue*, pp. 127–136, 2017.
- [2] Robin Ruede, Markus Müller, Sebastian Stüker, and Alex Waibel. Yeah, right, uh-huh: a deep learning backchannel predictor. In *Advanced Social Interaction with Agents*, pp. 247–258. Springer, 2019.
- [3] Tatsuya Kawahara, Takashi Yamaguchi, Koji Inoue, Katsuya Takanashi, and Nigel G Ward. Prediction and generation of backchannel form for attentive listening systems. 2016.
- [4] 石田真也, 井上昂治, 高梨克也, 河原達也ほか. 共感・発話促進のための多様な聞き手応答を生成する傾聴対話システム. 第80回全国大会講演論文集, Vol. 2018, No. 1, pp. 403–404, 2018.
- [5] 吉田奈央, 高梨克也, 伝康晴. 対話におけるあいづち表現の認定とその問題点について.
- [6] 井上昂治, 山本賢太, 中村静, 高梨克也, 河原達也ほか. 自律型アンドロイド erica による傾聴対話システムの評価. *SIG-SLUD*, Vol. 5, No. 02, pp. 19–24, 2019.