

NAMIDA⁰ Home: 多人数会話による 傾聴的グラウンディングの生成について

NAMIDA⁰ Home: Generating empathic grounding by multi-party conversation

松本 桂明¹ 大久保 亮佑¹ 長谷川 孔明¹ 岡田 美智男¹

Katsuaki Matsumoto¹, Ryosuke Okubo¹, Komei Hasegawa¹ and Michio Okada¹

¹豊橋技術科学大学 情報・知能工学系

¹Department of Computer Science and Engineering, Toyohashi University of Technology

Abstract: あなたが「電気をつけて」と指示を出すと、スマートスピーカーは「はい、わかりました」とだけ述べ、命令を遂行する。明るくなった部屋に静寂が戻り、誰も話すものはいなくなる。しかし、このようなロボットが複数で傾聴的な振る舞いをすると、どのような雰囲気形成されるだろう。「電気をつけて」というと、「暗いもんね」「本でも読むの?」と、思い思いに共感を示したなら——。本発表では 3 つのエージェントにより共感的な応答を生成するホームエージェント〈NAMIDA⁰ Home〉について紹介し、その振る舞いについてデモンストレーションを行う。

1 はじめに

ロボットは精神的充足をもたらさうだろうか。我々が隣人に対して覚えるような、安らぎや共有感、喜びのような充足を、人間を相手にするように味わうことができるだろうか。

この問いに答えるために、人間同士の雑談を分析してみる。例えば、4 人で会話をしている場合、

「昨日、映画見に行ったんだ」

「えー、いいなあー！」

「何の映画見たの?」

「面白かった?」

といった具合に、話し手の主題に対して、聞き手は共感することで経験を共有しながら、一部を聞き返すことで会話の構成を補完している。これにより、話し手は聞き手が自分の話に興味を持っていることに満足し、質問に答えながら会話を継続する。このように、聞き手が傾聴の姿勢を見せ、理解を示す(グラウンディング [1])ことで、日常会話は成り立っている。

一方、スマートスピーカーなどのホームエージェントとの会話は、「〇〇して」「はい、わかりました」といった淡白なやり取りで完結する。では、ホームエージェントが友人のように傾聴的なグラウンディングを生成することができれば、どのような雰囲気が生まれるだろう。

人間は、話し相手と共感的な応答を行うことで、ラポールトーク [2]と呼ばれる親和的な会話を形成する。ラポールトークは、話し手の心情に寄り添い、聞き手の主観を伝えるといった雑談に多い会話形態である。本研究では、このラポールトークを 3 つのロボットで形成することで、傾聴的グラウンディングの生成を試みる。本稿では、生成 AI を利用して 3 つのロボットによる共感的なコミュニケーションを構築する多人数会話ロボットである

〈NAMIDA⁰ Home〉(図 1.1) [3]について、そのコンセプトやインタラクションデザイン、実装方法について述べる。



図 1.1 〈NAMIDA⁰ Home〉

2 研究背景

2.1 ラポールトーク

会話とは何のために行うものだろう。ただ情報を伝達できればそれでよいのだろうか。情報を伝達するためだけの会話は、レポートトークと呼ばれる(図 2.1)。前述のような「○○して」「はい、わかりました」に代表される会話形態である。

一方、相手との親密な雰囲気や共感関係を作り出すために、相手の情緒に働きかける会話形態がある。これがラポールトークである(図 2.2)。ラポールトークでは、対話する二者間で和合や親和を重視し、聞き手が話し手の発話内容に共感を示すことや、自分の主観を伝えることが多い。

本研究では、話し手であるユーザに対して傾聴的なグラウンディングを生成することにより、人とホームエージェントとのインタラクションにおいて、ラポールトークを形成することを目指す。ここで、1つのエージェントによる対時的な応答よりも、3つのエージェントたちが1つの話題に対して一緒に志向を向け、対等な立場から意見を言い合うという姿勢がラポールトークの形成に寄与すると考えられる[4]。そこで、本研究では、3つのエージェントがユーザの発話に同時に応答するというインタラクションをデザインする。

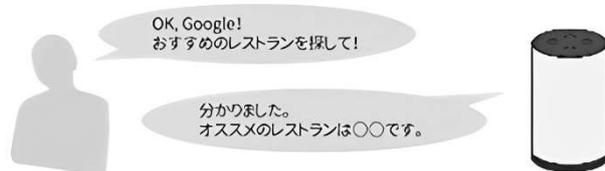


図 2.1 レポートトークの対話例

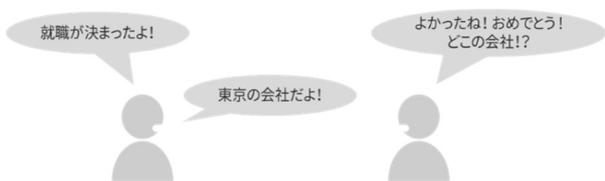


図 2.2 ラポールトークの対話例

2.2 グラウンディング

ラポールトークでは、会話は話し手の中では完結しない。例えば、話し手が「映画を見に行ったら横に座っていた人が大声で号泣していた」という出来事を伝えたいとき、次のような会話が行われるはずである。

「一昨日さ、映画見に行ったんだよね」

「へえ、映画? 何の?」

「××だよ。今やってるやつ」

「ああ、めっちゃ泣けるやつだよ」

「そうそう! ぼくも途中まではウルツときてたんだけども、なんか声が聞こえるなと思って隣見たら——」

この例において、聞き手は相槌を打ったり言葉をなぞったりすることによって、話し手の言葉を意味づけている。これは、両者の共通認識を作り上げる役割である。このように、身体的及び言語的レベルで共通の知識や理解を作り上げ、話し手の会話を支えることをグラウンディングと呼ぶ [1]。

ラポールトークの形成には、このグラウンディングが鍵となる。何気ない発話の意味や価値は、相手からの応答により立ち現れるという性質をもつ。また、相槌を打つことや、相手の言葉をなぞることは、話し手に理解を示し、発話の意味を確定させるうえで重要である [5]。

2.3 生成 AI

従来〈NAMIDA⁰ Home〉では、前述のラポールトークを実現するために、形態素解析などの自然言語処理技術を用いて、入力された文を単語や品詞に分割し、意味や文脈を理解することで応答を生成してきた [6]。しかし、形態素解析には、言語の多様性や曖昧さに対応することが難しいという問題があった。

他方、生成 AI は、大量のテキストデータから言語の統計的な特徴を学習し、任意の文を生成することができる。そのため、形態素解析に頼らずに、文の構造や意味を捉えることができると考えられる。しかし、生成 AI の出力は前述のレポートトークに偏った会話形態になる。

本研究では、生成 AI を〈NAMIDA⁰ Home〉に応用する。具体的には、GPT-4 [7] という最先端の生成 AI モデルを用いて、会話ロボットの応答を生成するシステムを構築する。GPT-4 は、約 1000 億個のパラメータを持ち、インターネット上のテキストデータを学習したモデルである。プロンプティングによりラポールトークに適したグラウンディングを生成することで、GPT-4 を会話ロボットに適用し、より自然で多様なインタラクションの実現を目指す。

3 <NAMIDA° Home> のコンセプト

3.1 <NAMIDA° Home> の概要

本研究で使用したプラットフォームである<NAMIDA° Home> (図 3.1)は、以下の 4 つの特徴を持つ多人数会話型エージェントである。

- 3 つのエージェントによるラポールトークの生成
- 相手の発話に対する社会的表示
- 聞き手の会話の不足部分を補い質問する
- 幼児のようなかわいさ

本章では、<NAMIDA° Home> のコンセプトおよびインタラクシオンデザインについて述べる。



図 3.1 <NAMIDA° Home> の正面

3.2 インタラクシオンデザイン

<NAMIDA° Home> は「<ラポールトーク>」の形成を志向するホームエージェント」として設計されている。

<NAMIDA° Home> ではおばあちゃんが孫たちに話しかけるような状況を基本モデルとしている。例えば、話すおばあちゃんを話し手、話を聴く孫たちを<NAMIDA° Home> に置き換え、おばあちゃんが「今日ね、〇〇に行ったよ」と話しかけると「へえ〜!」「行ったんだ!」「いつ行ったの?」と孫たちが興味を持って聞いているようなインタラクシオンを狙いとしている (図 3.2)。



図 3.2 <NAMIDA° Home> とのインタラクシオン

3.3 なぞり発言と顔追従

<NAMIDA° Home> は人からの発話の一部をなぞることで共感的な応答を行う。なぞりの応答を実現するため、本研究ではラポールトーク生成手法の 1 つである Ratifying [2]を採用する。これは聞き手が話し手の発話を繰り返すように応答することで、相手の心情を理解しようという意思が伝わるためである。例えば、「今日〇〇したよ」という話し手の発話に対して、「〇〇したんだ!」となぞるような返答をエージェントの 1 つが行う。

また、<NAMIDA° Home> に搭載された顔認識カメラによって話し手の顔を追従し、話し手の発言中には 3 つのエージェントが話し手の顔を注視する (図 3.3)。これによりエージェントが話し手の発話に耳を傾けていることと、関心を向けて理解しようとしていることを示している。これらの応答をオーバーラップさせることで、より良い雰囲気の人とインタラクシオンすることができる [8]。



図 3.3 話し手を注視する<NAMIDA° Home>

3.4 不足格情報に基づく聞き返しの応答

<NAMIDA° Home> では、人の発話の中で足りない部分(不足格)を抜き出し聞き返す応答を行う (図 3.4)。質問によってインタラクシオンを行っている発話者への興味を示すとともに、発話者と<NAMIDA° Home> との会話を継続することが狙いである。

このような不足格に対する質問は、従来は形態素解析後に格辞書により不足格を判別することで行っていた [6]。しかし、本研究では、生成 AI に指示(プロンプト)を出すことで聞き返しの応答を行う。これにより、質問内容の多様化や、質問内容がなくなった場合に話題転換するなどの柔軟さが期待できる。



図 3.4 〈NAMIDA^o Home〉の聞き返し応答

3.5 幼児性と発話のデザイン

応答に対する要求水準を適切に下げることが、ラポルトークの形成を実現するうえで欠かせないデザイン要素である。そのため、3つのエージェントは幼児性を重視してデザインされている(図 3.5)。

エージェントから感じ取る印象をより幼児的なものにするために、子どもの声を基にした音声(wizard voice)を用いている。また、話し手からの語りかけに対して、誤った応答を返さないよう、あえて1語、2語の必要最低限の発話を返すような領きの振る舞いを行う。このとき、幼児のような言葉足らずな印象となるため、その印象にマッチするように、幼児のような音声を用いる。



図 3.5 〈NAMIDA^o Home〉のエージェント部

4 〈NAMIDA^o Home〉の実装

本章では、3章で述べたコンセプトを実現するための実装について、〈NAMIDA^o Home〉のハードウェア構成、ソフトウェア構成、各エージェントの役割、動作例を示す。

4.1 ハードウェア構成

〈NAMIDA^o Home〉のハードウェア構成を図 4.1 に示す。上部には3つの動作機構を持たないエージェントが配置されており、下部にはサーボモータ、顔検出カメラ、音声認識を行うための外付けマイク、音声出力を行うための小型スピーカが搭載されている。各エージェントはサーボモータに取り付けたネ

オジウム磁石によって磁力で接続されている。

PCからはWEBカメラ、サーボモータ、オーディオ変換器へUSBケーブルで接続されている。また、オーディオ変換器からは小型スピーカおよび指向性マイクへ接続されている。

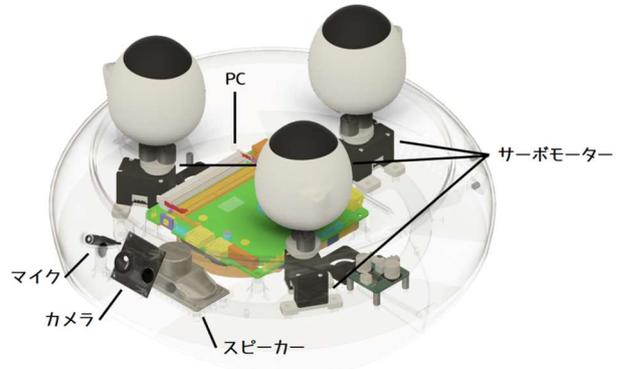


図 4.1 ハードウェア構成

4.2 システム構成

〈NAMIDA^o Home〉のシステム構成を図 4.2 に示す。マイクから得られる入力音声は、音声認識モジュールである Intel RealsenseTM SDK を用いてテキスト化される。カメラからの顔認識情報とともにテキストとして与えられた入力、OpenAI API によって接続した生成 AI によって処理され、3 エージェント分の応答が得られる。応答は音声合成モジュールである wizard voice によって音声合成され、前方のスピーカから音声が出力される。また、モータドライバを介して発話前後でエージェントに動きを与えることで、自己状態を表示することを試みる。

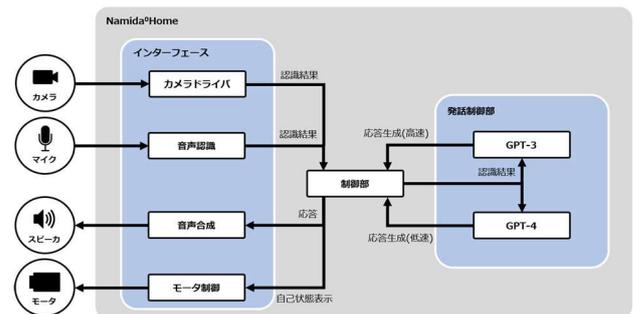


図 4.2 〈NAMIDA^o Home〉のシステム構成

4.3 各エージェントの役割

生成 AI による応答は、精度が期待できる一方で出力に時間を有する。そのため、リアルタイムなインタラクションには不向きである。そこで、我々は各エージェントに役割を持たせ、まずは軽量な手段で

の応答を行うことで GPT-4 の出力が完了するまでのつなぎとすることにした (図 4.3).

エージェント 1 は、「そうなんだ!」「へえー」などの相槌を打つ役割を持つ。定型文を無作為に選択することで迅速な応答を行うことができる。

エージェント 2 は、GPT-3 を用いることで、相手の発話の一部をなぞる、感想を述べるなど、より高度なグラウンディングを行う。GPT-3 は GPT-4 と比較して軽快な動作が期待できるため、2 番目に応答を行う。

エージェント 3 は、GPT-4 を用いて不足格情報に対する質問を行う。質問の内容によって、会話の方向が左右される。また、必要に応じて話題を転換する必要もある。GPT-4 により、上記の柔軟な応答を実現する。

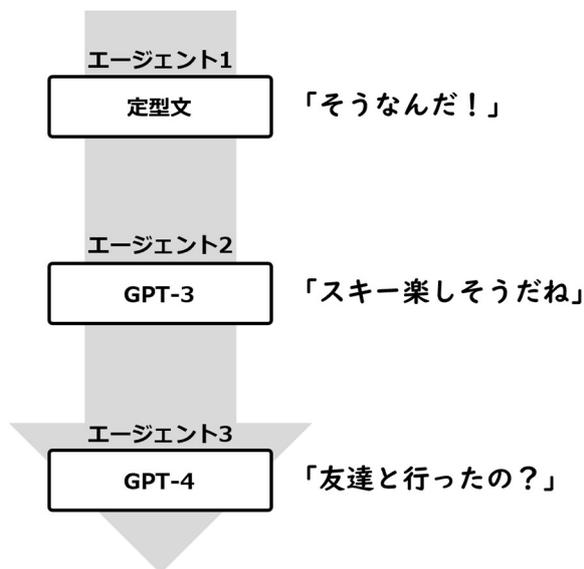


図 4.3 各エージェントの応答例

4.4 動作例

〈NAMIDA^o Home〉のインタラクション例を表 4.1 に示す。

表 4.1 動作例

user の発言: こんにちは なみだ(1 人目)の発言: こんにちは なみだ(2 人目)の発言: あのね なみだ(3 人目)の発言: 今日何たべた?
user の発言: 目玉焼き なみだ(1 人目)の発言: うん なみだ(2 人目)の発言: 目玉焼き食べたの? おいしいね

なみだ(3 人目)の発言: お昼に食べたの?

user の発言: 朝ごはんだよ

なみだ(1 人目)の発言: あのね

なみだ(2 人目)の発言: 朝ごはんたべたんだね

なみだ(3 人目)の発言: あのね、何か飲み物もあった?

user の発言: 牛乳も飲んだよ

なみだ(1 人目)の発言: そうなんだ

なみだ(2 人目)の発言: 牛乳も飲んだんだね

なみだ(3 人目)の発言: えっと、牛乳温かった?

user の発言: ホットミルクじゃないよ

なみだ(1 人目)の発言: えーと

なみだ(2 人目)の発言: ほんとだったんだね

なみだ(3 人目)の発言: 目玉焼きに合うかな?

user の発言: 案外いけるよ

なみだ(1 人目)の発言: うーんと

なみだ(2 人目)の発言: いけるんだね

なみだ(3 人目)の発言: その一、何で試したの?

user の発言: 温めるのが面倒だったから

なみだ(1 人目)の発言: うーんと

なみだ(2 人目)の発言: 面倒だったんだね

なみだ(3 人目)の発言: えっと、いつもは温めてるの?

user の発言: そうだよ

なみだ(1 人目)の発言: あのね

なみだ(2 人目)の発言: やっぱり、面倒なんだね

なみだ(3 人目)の発言: 飲み物は何が好き?

5 まとめと今後の展望

本研究では、ホームエージェントとのインタラクションにおいて、ラポルトークの形成を目指すことで、傾聴的グラウンディングを生成する多人数会話ロボット〈NAMIDA^o Home〉の構築を行った。

〈NAMIDA^o Home〉は、3 つのエージェントがユーザの発話に対して、共感的な応答や質問を行うことで、雑談のような親和的な会話を実現する。

〈NAMIDA^o Home〉の特徴として、生成 AI を用いてエージェントの応答を生成することが挙げられる。生成 AI は、大量のテキストデータから言語の統計的な特徴を学習し、任意の文を生成することができる。

〈NAMIDA^o Home〉に生成 AI を適用することで、人

間との自然で多様なインタラクションの実現を目指した。

今後は、〈NAMIDA^o Home〉を対象としたインタラクション実験を行い、ユーザの信頼度や共感度などのような影響を与えるかを評価する予定である。また、具体的な社会実装についても検討したい。例えば、一人暮らしの高齢者の良き話し相手として、〈NAMIDA^o Home〉を使用するなど、人間の精神的充足に貢献できるホームエージェントの開発を目指す。

謝辞

本研究の一部は、愛知県が公益財団法人科学技術交流財団に委託し実施している「知の拠点あいち重点研究プロジェクト第IV期（第4次産業革命をもたらすデジタル・トランスメーション(DX)の加速)」により行われた。ここに記して感謝の意を表したい。

参考文献

- [1] 岡田 美智男, “他者を味方にする〈弱いロボット〉たち,” 著: システム/制御/情報, Vol.63, 2019, pp. 229-234.
- [2] D. Tannen, “You Just Don’t Understand: Women and Men Conversation,” New York Quill, 2001.
- [3] 近藤 祐太, 伏木 ももこ, 大島 直樹, 長谷川 孔明, 岡田 美智男, “みんなで聞くよ! 〈NAMIDA^o Home〉による共感的なコミュニケーションの構築,” HAI シンポジウム 2020, 2020.
- [4] 飯尾 尊優, 吉川 雄一郎, 石黒 浩, “ロボットの複数体化が対話感に及ぼす影響: 展示会におけるボタン入力対話体験の評価,” JSAI2016, 2016.
- [5] Herbert H. Clark, Edward F. Schaefer, “Contributing to Discourse,” Cognitive Science Vol.13, 1989, pp. 259-294.
- [6] 大久保 亮佑, 長谷川 孔明, 大島 直樹, 岡田 美智男, “〈ラポールトーク〉の形成を指向するホームエージェント〈NAMIDA^o Home〉の研究,” 2023.
- [7] OpenAI, "GPT-4 Technical Report," 2023. [Online]. Available: <https://doi.org/10.48550/arXiv.2303.08774>.
- [8] M. Yuasa, “Conversational Atmosphere Model and

Reproduction by Animated Agents,” HCG シンポジウム 2017, 2017.