

みんなで聞くよ！〈NAMIDA⁰ Home〉による 共感的なコミュニケーションの構築

We are listening to you! Building empathic communication
using rapport-type agents “NAMIDA⁰ Home”

近藤祐太¹ 伏木ももこ¹ 大島直樹² 長谷川孔明¹ 岡田美智男¹

Yuta Kondo¹, Momoko Fushiki¹, Naoki Ohshima², 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

²豊橋技術科学大学 エレクトロニクス先端融合研究所

²Electronics-Inspired Interdisciplinary Research Institute,
Toyohashi University of Technology

Abstract: 相手に何気なく言葉を投げかけたとき、その相手の志向がこちらに向かい、また、こちらの会話に関心を示すような振る舞いを表出してくれたら、話を聞いてくれているという安心感や、私のことを理解してくれているという嬉しさのようなものを感じることができるだろう。このような話し手の発話を支える聞き手としての参加者が複数人居たら、どのような雰囲気形成されるのだろうか。本研究では、3つのエージェントにより共感的な応答を生成する多人数会話型のホームエージェント〈NAMIDA⁰ Home〉の構築を進めている。本発表では、〈NAMIDA⁰ Home〉のコンセプトおよびインタラクションデザインを説明し、実際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

何気ない会話における子どもたちの無邪気なやり取りに着目してみる。一緒になって会話を作り上げようとする子どもたちの協力的な姿勢を見ることができる。例えば、おばあちゃんと孫たちとの雑談の場面に例を考えてみよう。

おばあちゃん:「あのね、今日スーパーへ買い物に行ってきたの」

孫たち:「へー行ってきたんだー」「誰と、行ってきたのー?」「何を、買ってきたのー?」

無邪気に言葉を返す孫たちの積極的な姿勢にぎわいを感じつつも、おばあちゃんの伝えたいという主題に対して、孫たちも一緒に共感しながら、その出来事や経験をみんなで共有している。また、まだ伝えきれていない内容の一部を孫たちが聞き返すことで、おばあちゃんの会話の構成を補っている。積



図1. 〈NAMIDA⁰ Home〉

極的に話を聞いてくれるこのような孫たちの存在に嬉しさや疎通感のようなものを感じることができるだろう。話し手(おばあちゃん)と聞き手(孫たち)との双方の関わりによって、日常の何気ない雑談というものは構成されていることが分かる。

スマートスピーカーなどのホームエージェントが一般の家庭にも広く浸透している。エージェントと

の対話的なやり取りでは、ユーザからの特定の作業に対する要求と、その命令を受諾するエージェントからの応答に分けられる。例えば、ユーザからの「〇〇して!」という発話に対し、「分かりました。〇〇します」というような返答をエージェントは返す。

このようなスマートスピーカーとの関わりを前提としたとき、特に重要ではないが今日あった出来事を話しかけてみることはできるだろうか。そのように何気なく言葉を発するときにも、スマートスピーカーやホームエージェントなどから、孫たちとの会話のように、一緒に会話を構築しようと試みる振る舞いが現れてもおもしろい。

おばあちゃんの会話をみんなで聞く孫たちのように、ユーザが何気なく発した言葉にエージェントたちの志向が向き、今日あった出来事や経験に対する関心や共感を示してくれる。まだ伝えていない出来事の詳細を聞き返してくれる。そのような反応を生成するエージェントが複数いたら、共有感や安心感、嬉しさのようなものをスマートスピーカーからも感じることができるのではないだろうか。

このような発想のもと、本研究では、複数の聞き手として共感的な応答を生成するホームエージェント (NAMIDA⁰ Home) の構築を進めている(図 1)。本稿では、(NAMIDA⁰ Home) のコンセプトやインタラクションデザイン、ハードウェア構成、ソフトウェア構成などについて述べる。

2 研究背景

2.1 ラポールトーク

Google Home や Amazon Echo などといったスマートスピーカーが近年に登場し、人とエージェントとの対話の機会が多くなっている。これらのスマートスピーカーはウェイクワードなどをきっかけにシステムは命令を受け付け、ユーザからの「〇〇して」などの音声による指示によって、家電の操作やメールの送信、電話への発信/応答などの機能の実現が行える。しかし、このような命令と受諾という人とロボットの関係は、どこかよそよそしく、相手の発話に耳を傾け、相手に共感を示す姿勢からは程遠い。ユーザからの命令が終了すると、エージェントの応答も終了してしまう。このようなやりとりは、タスク志向の会話に多いとされている。

社会学者の Tannen はこのような会話形態を「レポートトーク(report-talk)」と呼んでいる^[1]。また、非タスク志向の会話である「ラポールトーク(rapport-talk)」がある。ラポールトークは対話する 2 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

示したり、自分の主観を伝えるといった雑談に多いとされる。合理的というよりは、調和的・共感的な姿勢がより強く現れる会話形態といえる^[1]。

本研究では、人とホームエージェントとの対話において、ラポールトークの形成を目指すことにより、話し手であるユーザに対して共感的なインタラクションを生成する。ここで、1つのエージェントによる対峙的な応答よりも、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ている。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。

2.2 随伴的な聞き手の応答

人と人との会話においては、話し手の発話に対して、聞き手が何らかの応答を示すことで、その発話の意味が確定したり、また、「相手の発話をちゃんと理解していますよ」ということを社会的に表示 (social display) する役割がある。

例えば、「おはよう」という相手からの語りかけに対し、その発話が無視されてしまうならば、その「おはよう」の意味は宙に浮いてしまう。自分の存在価値も少し小さくなったようにも感じられるだろう。一方、相手が「おはよう」と返してくれるならば「自分の存在に気付いてくれているな」という気持ちを感じると同時に、相手との社会的なつながり (social bonding) のようなものも感じることができる。このように、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、その相互行為における意味や価値の生成を支えるという役割だけに限らず、聞き手の随伴的な応答は相手への理解の表出にも関わる。話し手に対して、発話内容を聞き手が理解できていることを示す方法にはいくつか存在する。Clark ら^[2]は言語的な表示の手法として4つを提示している。例えば、「うん」「そう」という相槌(acknowledgement)を示すもの、発話内容の一部を模倣するもの(demonstration)、発話内容の全部をそのまま模倣するもの(display)がある。これらの振る舞いを表出することで、相手への理解の表出を行っているという。

また、Goodwin^[3]は聞き手の身体による理解の表示手法として、話し手に視線を合わせる、話し手の目線の先に身体を移動させる、話し手にジェスチャーを行うといった行動も、話し手の会話を支える重要なリソースになると挙げている。

本研究では、このような身体的および言語的なレベルでの様々のグラウンディング(grounding)の手法を参考にし、エージェントの応答をデザインする。



図 2. $\langle \text{NAMIDA}^0 \text{ Home} \rangle$ とのインタラクション例

2.3 聞き返しの応答

その日に経験した出来事を相手に共有しようとする会話において、まだ相手から伝えられていない部分を補うようにエージェントたちが聞き返す。例えば、図 2 のインタラクション例に示すように、「買い物行ってきたよ」というユーザからの発話に対して、エージェントは「誰と行ってきたの?」という発話を生成する。相手と経験を共有するうえで、まだ相手に伝え切れていない、不足している情報の部分をエージェントたちが補うように聞き返す。このような聞き返しの振る舞いにより、ユーザからの語りかけに対して、「ここまでの話は理解したよ」「その先の話の聞きたいな」といったエージェントたちの興味、感心を示すことに繋がると考えている。また、ユーザの「もっと話したい」といった会話に対する参加姿勢(engagement)を高めることに繋がると考える。

本研究では、このような聞き返しの応答のデザインについても検討し、話し手と一緒に共感的な姿勢での会話の構築を目指すような関係論的なインタラクションを狙いとしている。

3 $\langle \text{NAMIDA}^0 \text{ Home} \rangle$ のコンセプト

$\langle \text{NAMIDA}^0 \text{ Home} \rangle$ は「3 つのエージェントによるラポルトークの生成」「相手の発話をちゃんと理解していますよという社会的表示」「会話の不足部分を補う関係論的なコミュニケーション」の 3 つを特徴としている。これらのコンセプトを実現するため、次のようにインタラクションをデザインした。

3.1 理解や共感の表出

$\langle \text{NAMIDA}^0 \text{ Home} \rangle$ 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、本研究ではラポルトーク生成手法の 1 つである Ratifying^[1]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を行うことで、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」というユーザの発話



図 3. $\langle \text{NAMIDA}^0 \text{ Home} \rangle$ の正面

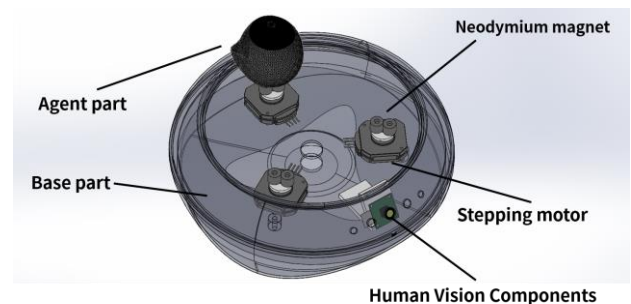


図 4. $\langle \text{NAMIDA}^0 \text{ Home} \rangle$ のハードウェア構成

に対して、「〇〇したんだ!」となぞるような返答を $\langle \text{NAMIDA}^0 \text{ Home} \rangle$ のエージェントの 1 つが行う。

また、 $\langle \text{NAMIDA}^0 \text{ Home} \rangle$ に搭載された顔認識カメラ(human vision components)によってユーザの顔を追随し、ユーザの発言中には 3 つのエージェントがユーザの顔を注視する。発言終了時には頷くといった動作を行う。これによりエージェントがユーザの発話に耳を傾けていること、関心を向け、理解しようとしていることを示す。

3.2 不足格情報に基づく聞き返しの応答

$\langle \text{NAMIDA}^0 \text{ Home} \rangle$ 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。このような不足格の判別を行うため、新たに格辞書を作成した。

格辞書には、日本語初級程度^[4]の基本的な動詞および形容詞などの約 450 語のデータベースを構築した。本研究では児童の雑談をモチーフにしているため、このような言語レベルのデータベースとしている。また、格辞書に合わせて、必須格辞書を構築している。例えば、「行く」という動詞に対しては「人が(行く)」というガ格や、「場所へ(行く)」というへ格などといった格を登録している。これらのデー

データベースに基づき、ユーザの発言から不足している格を特定する。

3.3 幼児のような可愛らしさと発話のデザイン

スマートスピーカーからの応答に対するユーザの要求水準を適切に下げることが、共感的なインタラクションを実現するうえで欠かせないデザイン要素であると考えられる。奈良県の5歳児の音声の元に合成された音声(wizard voice)を用いることで、エージェントから感じ取る印象をより幼児的なものにした。また、ユーザ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頷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ている。

4 〈NAMIDA⁰ Home〉の構築

4.1 概要

3章で述べたコンセプトを実現するため、〈NAMIDA⁰ Home〉を構築した。伏木ら^[5]によって構築されたリビングエージェント〈NAMIDA⁰〉をベースとし、対話機構と顔認識などホームエージェントとしての機能を追加することで〈NAMIDA⁰ Home〉を構築した。

4.2 ハードウェア構成

〈NAMIDA⁰ Home〉のハードウェア構成を図4に示す。〈NAMIDA⁰ Home〉のハードウェア構成は上部のエージェント部と下部のベース部に大きく分かれる。エージェント部では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、ベース部ではエージェントを回転させる3台のステッピングモータ、人の顔を追従するための顔検出カメラ HVC(Human Vision Components)、音声認識を行うための外付けマイクが配置されている。ここで、各エージェントはステッピング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おり、磁石が回転するとエージェントも回転する。

4.3 ソフトウェア構成

〈NAMIDA⁰ Home〉のソフトウェア構成はメインフォームを中心とし、その下にエージェントをシリアル通信で操作するエージェントドライバ、音声対話処理を行う音声ドライバ、HVCカメラドライバが配置されている。また、音声ドライバは係り受け解析(日本語構文解析システム KNP)、音声認識(Intel RealsenseTM)、音声合成(Wizard Voice)から構成されている。

〈NAMIDA⁰ Home〉の応答処理の流れとして、音声認識のディクテーションを利用して、ユーザの発話を文字列として入力し、構文解析システムによって解析を行う。また、係り受け構造を特定することにより、ユーザの発言の一部(例: 述部)をなぞるような応答文、また、不足格を聞き返すような応答文を生成する。そして、応答文を音声合成システムに入力し、生成された wave ファイルをスピーカーから再生する。デモンストレーションでは、これらの動作を実演する予定である。

5 まとめと今後の展望

本研究では、話し手に対して共感的な応答を行うラポルトークを特徴とする多人数会話型のエージェント〈NAMIDA⁰ Home〉の提案および構築を行った。〈NAMIDA⁰ Home〉は幼児の雑談をモチーフとする、3つのエージェントによるラポルトークを特徴とした共感的なコミュニケーションの生成を目指している。今後は次のような実験や社会応用について考えている。

話し手に対して共感的な応答を行うロボットおよびエージェントの研究は近年になって盛んに行われてきている。神田ら^[6]は人の動きに追従し、アイコンタクトや指差しなどの協調動作を行う「聞いているという動作を行うロボット」の開発を行い、音声的要因よりも身体的要因が共感度や信頼度に対して影響を持ちやすいと述べている。本研究でも〈NAMIDA⁰ Home〉がユーザの信頼度や共感度にどのような印象を与えるのか、インタラクション実験を実施することでより詳細な調査を行う予定である。

また、下岡ら^[7]は大規模コーパスを用いて問い返しや共感応答・相槌応答を行う対話システムを開発し、特に高齢者に対してこれらの応答が効果的であると述べている。このことから、一人暮らしのお年寄りの良き話し相手として、〈NAMIDA⁰ Home〉を使用するなど、具体的な社会応用についても検討したい。

謝辞

本研究は、科研費補助金（基盤研究（B）18H03322）によって行われた。ここに記して感謝の意を表す。

参考文献

- [1] Deborah Tannen: You Just Don't Understand: Women and Men Conversation, New York Quill, 2001.
- [2] Herbert H. Clark, Edward F. Schaefer: Contributing to Discourse, 1989.
- [3] Charles Goodwin: Embodied Hearers and Speakers Constructing Talk and Action in Interaction, March 2009.
- [4] スリーエーネットワーク(著編): 『みんなの日本語初級 I 第2版』, スリーエーネットワーク, 2012.
- [5] 伏木ももこ, 田村真太郎, Nihan Karatas, 島崎景子, 因幡千尋, 那和一成, 岡田美智男: NAMIDA⁰: ミニマルなソーシャルエージェントとそのインタラクショナルデザイン, ヒューマンインタフェースシンポジウム 2017, P-35, 2017.
- [6] 神田崇行, 鎌島正幸, 今井倫太, 小野哲雄, 坂本大介, 石黒浩, 安西祐一郎: 人間型対話ロボットのための協調的身体動作の利用, 日本ロボット学会誌 Vol.23 No.7, 2005.
- [7] 下岡和也, 徳久良子, 吉村貴克, 星野博之, 渡部生聖: 音声対話ロボットのための傾聴システムの開発, 自然言語処理 Vol.24 No.1, 2017.