

親しみや安心感を感じさせる「話しかけ」ができる AI エージェントの開発研究

The research and development of AI agents that can talk to people in a way that gives them a sense of familiarity and relief

専修大学ネットワーク情報学部 小杉尚子, 石井健太郎

新潟医療福祉大学・医療技術学部・診療放射線学科 児玉直樹

School of Network and Information, Senshu University Naoko KOSUGI, Kentaro ISHII

Department of Radiological Technology, Faculty of Medical Technology, Niigata University of Health and Welfare, Naoki KODAMA

Keywords: AI agent, talking, elderly people with dementia

Abstract

This paper describes research into the development of an artificial intelligence (AI)-based "talking" agent for elderly people with dementia. The number of elderly people with dementia is increasing worldwide, and the burden on caregivers has become a social problem because of the extreme difficulty of caring for elderly people with dementia. Behavioral and psychological symptoms of dementia such as "wandering" and "apathy" are increasing the burden on caregivers, and "apathy" has attracted attention in recent years as a factor increasing the psychological burden on caregivers. One possible way to recover "apathy" is to promote spontaneous behavior, such as talking, among the elderly people with dementia. However, Japan is facing a serious shortage of caregivers due to the low birthrate and aging population, and it is difficult for caregivers to have enough time to talk with elderly people with dementia. Therefore, we have developed an AI agent that "talks to" elderly people with dementia, aiming to create an environment in which elderly people with dementia can speak anytime, anywhere, and as much as they want. In this research, we first create a dialogue scenario that facilitates the talk with elderly people with dementia. We have also developed a prototype system of an AI agent that can interact with elderly people with dementia in a scenario-based manner. Our prototype uses Google Cloud Text-to-Speech as the speech synthesis technology and Open AI's Whisper as the speech recognition technology. We confirmed that the AI agent has the potential to facilitate the speech of elderly people with dementia by interviewing caregivers who are engaged in the care of elderly people with dementia.

1. はじめに

近年の ICT 技術の進化発展には目覚ましいものがあり、ChatGPT 等の AI 技術や、ドローン等に使用されるような高性能デバイスの小型化・低価格化は、我々の日常生活を急速に変化させている。カメラやマイクなどの一般的なデバイスも高性能化・小型化・低価格化が進んでおり、人の表情認識、動作認識、音声認識といった認識技術や自然言語処理技術と AI を組み合わせた「対話技術」等も急速に進化発展しており、使用する状況や方法がある程度限定すれば、あたかも人と会話しているような気持ちを人に与

える AI エージェントの開発が可能となりつつある[1]。

認知症高齢者数は世界的に増加しており、認知機能の障害によって介護は非常に困難で、介護者の負担軽減が喫緊かつ重大な社会課題になっている。介護負担を増大させているのは「徘徊」、「行動障害」、「無為・無関心」等の「行動・心理症状」である[2]が、これまで「無為・無関心」は介護負担とはあまり関係ないと考えられてきた。しかし認知症高齢者の「無為・無関心」は介護者の介護モチベーションを低下させ介護者の心をじわじわと蝕むため、「徘徊」や「行動障害」などと同様に、介護負担増大の要因として

関心が高まっている。したがって、認知症高齢者の「無為・無関心」を改善させる必要があるが、「無為・無関心」といった、いわゆる精神の陰性的な症状の改善には時間と人手がかかる。にもかかわらず、日本は少子高齢化が急速に進んでおり、時間や人手をかけて介護するどころか、そもそも介護人材の確保すら困難な状況である。

そこで本研究では、認知症高齢者の「無為・無関心」の改善策として発話等の自発行動の促進に着目し、特に「発話」を促すことに特化した AI エージェントを開発する。認知症高齢者がこの AI エージェントと会話を重ねることで、意欲的に発話するようになり、それが「無為・無関心」の改善に繋がることが期待される。そして認知症高齢者の自発的な発話の増加によって、介護者とその発話内容を知ることによって、認知症高齢者の「人となり」に関する理解を深め、介護モチベーションの向上にも繋がることが期待される。

本論文の構成は以下である。第 2 章では認知症について簡単に説明する。第 3 章で、対話技術における発話促進に関連する従来技術などについて説明する。第 4 章ではこれらを受けて、我々が開発する AI エージェントの開発研究の全体像について述べ、第 5 章では AI エージェントが対象者に話しかけるためのシナリオ作成と話しかけ方について述べる。第 6 章では AI エージェントを試作開発するための、音声合成および音声認識技術に関する調査結果と、本研究開発で採用する技術の選定について報告する。第 7 章では AI エージェントの試作開発とそれに対する評価結果を報告し、第 8 章でまとめと今後の展開について述べる。

2. 認知症

認知症とは、正常に発達した認知機能が、後天的な何らかの障害により、生理的老化の範囲を超えて広汎かつ慢性的に低下した状態のこと[3]で、記憶、言語、視空間認知などの認知機能の障害と、BPSD (Behavioral and Psychological Symptoms of Dementia) と呼ばれる徘徊や無気力などの「行動・心理症状」からなる[2]。認知症には、現状では有効な治療法がないため、認知症高齢者への対応は介護が中心となるが、認知症高齢者の介護は非常に困難であるため、介護者の負担を軽減することが重要である。

認知症高齢者の介護を困難にするのは徘徊、不安、抑うつなどの BPSD であるが、これは適切なケアを行うことで発症を抑制したり、症状を軽減したりすることができるため、非薬物療法で対応することが推奨されている。認知症疾患診療ガイドライン 2017 [4] には、非薬物療法として運動療法、回想法、音楽療法などが挙げられている。

3. 話し出しやすい・話しやすいと感じられる「話しかけ」

話し出しやすい・話しやすいと感じられる「話しかけ」とは、どのようなものだろうか？

生成 AI の出現・発展によって、AI エージェントとの会話の質は目覚ましい進化を遂げている。しかし AI エージェントは、あくまでもこちらからの問いかけに答えるのであって、こちらが話し始めることを促すような仕組みは今のところ無い。

「会話に飢えている」と言われる高齢者・認知症高齢者だが、そもそも話す相手がいないことが多いだけでなく、同じ話を繰り返してしまったり、世代のギャップによって全く話が合わなかったりしたときに、相手が不快になっているような様子を見る経験を通して、おしゃべりを我慢してしまう高齢者も少なくない。そこで、人が話し出しやすく安心して話し続けられるような環境を生む対話技術について調査した。

対話技術研究について、現在は第 3 次ブームと言われているが、1950 年代に始まった第 1 次ブームに生まれた対話システムで有名なものの 1 つにマサチューセッツ工科大学で開発された ELIZA (イライザ) がある。「共感的理解」や「無条件の肯定的関心」などからなるロジャース心理学のカウンセリング手法[6, 7]を模倣する対話システムと言われており、特定のキーワードに対し、どのような反応をするかを記述したシナリオ (パターン) に基づいて対話を続ける。相手の発話の復唱や、「具体的には?」「もう少し詳しく教えて下さい」など、相手の話を深掘る質問を繰り返すだけなのだが、多くのユーザが ELIZA が自分のことをわかってくれたと解釈し、ELIZA との会話に没頭したと言われており[1]、「ELIZA 効果」とも呼ばれている。

4. AI エージェントの開発研究

本研究で開発する AI エージェントは、認知症高齢者と「雑談」をするのではなく、認知症高齢者の質問に答えるのでもなく、認知症高齢者が安心して話し続けられるような「話しかけ」ができる AI エージェントである。そこで第 3 章に記載した ELIZA 効果を参考に、認知症高齢者に対して質問と肯定的な反応を繰り返すことで、親しみや安心感を感じてもらえるような「話しかけ」をする AI エージェントを目指す。認知症高齢者は、自分に興味を持って様々なことを質問してくる AI エージェントに回答し、それに対して肯定的な反応が得られる体験を重ねることで、自発的かつ意欲的に安心して発話するようになることが期待される。

本研究のポイントは以下である。

1. 質問と肯定的な反応を繰り返す対話シナリオの作成
2. AI エージェントが話しかけるための音声合成技術

と、相手の回答を聞き取るための音声認識技術を用いたプロトタイプシステムの試作

3. 試作したプロトタイプシステムに対する評価

まずシナリオについてであるが、本研究ではトピック（例：春、和菓子、天気等）とキーワードの集合（例：桜、新緑、入学、団子、晴れ等）を多数用意し、それらを用いて「～と言えば？」「～の思い出は？」「～について聞かせて下さい」等の「問いかけ」を多用するシナリオ作成を想定する。そして高齢者の回答に対しては、「そうですね」「そうなんです」と肯定的な回答をし、その後にはほんの少し「自己開示」をすることで、AI エージェントに対して親しみを感じてもらえるようにする。詳細は第 5 章で述べる。

音声合成技術および音声認識技術については、現状で利用可能な複数の技術を抽出し、それらについて調査した上で、本研究開発に適切なものを選定する。そして選定した技術を用いてプロトタイプシステムを試作する。詳細は第 6 章、第 7 章で説明する。

5. シナリオ作成と話しかけ方について

本章では、「親しみや安心感を感じさせる「話しかけ」をして、質問と肯定的な回答をする AI エージェントの発話シナリオの作成について述べる。

このシナリオ作成では、非薬物療法の 1 つとして認知症高齢者に対して効果があるとされている音楽療法を、実際の認知症高齢者に行っている現役の音楽療法士の指導を受けながら作業を進めた。音楽療法とは「音楽の持つ生理的、心理的、社会的働きを用いて、心身の障害の回復、機能の維持改善、生活の質の向上、行動の変容などに向けて、音楽を意図的、計画的に使用すること」[8]とされており、歌唱（合唱）や楽器演奏などを通して、受療者のストレス発散や認知機能・運動機能の改善などを支援する。音楽療法は音楽療法士が行うのが一般的で、音楽療法士は歌唱や楽器演奏などの音楽活動の合間に対象者と対話し、それを通して対象者をより良く知って対象者に寄り添う。そこで本研究における対話シナリオの作成も音楽療法士の対話を参考にして作成することにした。音楽療法士には ELIZA 効果について説明し、それを尊重して頂きながら、実際の音楽療法セッションでも利用可能な話題や話の流れを指導して頂いた。指導内容の骨子は以下である。

1. まず、こちらから名乗る。
2. 最初の質問は「ご気分はいかがですか？」が良い。
3. 誰もが答えられる可能性が高く、安心して話せる話題としては、「天気」「好きな食べ物」が良い。
4. 最後に「ありがとうございました。」と感謝を述べる。

また、声は大きい方が良く、話すスピードはあまりゆっくり過ぎない方が良く、健全な成人の会話スピードよりも少しゆっくりであるくらいを目指す方が良く、等

のアドバイスを得た。さらに、「話しかけ」については、全体的にメリハリをつけることが非常に重要だとアドバイスを頂いた。具体的には、より伝えたい単語については、発話する直前に間を取ったり、より大きな声で発話したりするなどの配慮が望ましいとのことだった。

以上を踏まえて作成したシナリオの一部を示す。（我々が開発している AI エージェントは音楽療法士という想定で、名前は「さつき」である。）

- ```
=====
```
1. こんにちは。AI 音楽療法士のさつきです。5 月生まれなので名前は「さつき」といいます。今日、音楽をご一緒させて頂きますね。よろしくおねがいします。さて、今日のご気分はいかがですか？
  2. そうですね。私は朝ごはんをたくさん食べましたのでとても元気です。今日はそちらはどんなお天気ですか？
  3. そうなんです。私は晴れの日が好きです。青空の中にふわふわの白い雲を見付けると、嬉しくなり、綿あめが食べたくなります。あ、好きな食べ物は何ですか？
- ```
=====
```

このシナリオのポイントは、受療者に対する質問を投げた「問いかけ」を多用している点と、受療者の回答に対して、「そうですね」「そうなんです」と、最初に肯定的な反応を示したうえで、自身のことを少し話して「自己開示」を続けている点である。

6. 音声合成技術と音声認識技術

本章では、音声合成技術と音声認識技術について、現状で利用可能な複数の技術を抽出し、それらの技術調査結果をまとめ、最終的に本研究で試作する AI エージェントのプロトタイプシステムで使用する技術の選定を行う。

6.1. 音声合成技術

機械による音声合成技術は 1790 年頃から存在しており、1900 年代後半にコンピューターが実用化されるとともに、さまざまな技術が採用されてきた。2010 年代に入ってからディープラーニング（以下、DL）が登場し、多くの分野で応用され始めた。その結果、近年では DL を用いたより自然な音声合成技術が開発されている。代表的なものとして、Google Cloud Text-to-Speech, Amazon Polly, Microsoft Azure Text-to-Speech, CoeFont などがある。本論文ではこれらのうち、紙面の関係で Google Cloud Text-to-Speech と CoeFont について比較検討する。

まず、どちらも音声合成 Web API であるが、それぞれの大きな特徴として、前者（以下、GCTTS）は代表的なクラウドサービス提供者が提供していること、後者は日本語に特化していることが挙げられる。以下、それぞれについてさらに詳細にまとめる。

6.1.1 GCTTS

GCTTS は、Google Cloud 上で提供されている音声合成サービスで、2015年に発表された「WaveNet」と呼ばれる DL を用いた音声合成技術を利用している。

WaveNet は、音声合成において、数理モデルを利用せずに音声の波形を直接生成することを目的として開発された。「Alpha Go」を開発した DeepMind 社が Google 買収後に開発した技術である。従来の数理モデルによる音声合成技術では、音声の波形を生成するために短い音声を結合するという手法が用いられていた。しかし WaveNet では、CNN (Convolutional Neural Network の略。畳み込みニューラルネットワーク) を通して音声の波形を直接生成することで、音声の断片を結合する必要がなく、より自然な音声合成を実現している。2022年には WaveNet がより進化した「Neural2」というモデルが採用され、より自然な音声を合成できるようになった。2023年には Neural2 が日本語に対応し、8月31日の時点で、英語・フランス語・スペイン語・中国語・日本語など、様々な言語に対応している。SSML (Speech Synthesis Markup Language の略。音声合成のためのマークアップ言語) での音声合成リクエストにも対応している。

音声合成の品質は、従来の数理モデルを用いた Standard に加えて、WaveNet、Neural2 の3つのモデルから選択することができる。Standard モデルで作成した音声は機械的に作られたことが分かるレベルの品質であり、不自然なノイズなども発生することがある。一方、WaveNet モデルで作成した音声は Standard モデルと比較して自然な合成音声になっているが、Neural2 モデルと比較するとやや機械的な音声の印象を受ける。Neural2 モデルでは、注意して聴くとイントネーションが不自然な点がある程度で、かなり自然な合成音声になっている。mp3 形式で合成音声を出力する場合は、mp3 特有のノイズが発生するが、この問題は合成音声を wave 形式で出力することで解決できる。

利用料金はモデルに応じて異なる従量課金制で、100文字ごとに Standard では\$4、それ以外は\$16である。また月間あたりの無料枠は standard は 400 万文字、それ以外は 100 万文字である (2023/8/31 時点)。合成音声の速度や音高、間隔を空ける、強調するなどの処理は SSML によって制御することができる。

6.1.2 CoeFont

株式会社 CoeFont により運営されている日本語に特化した音声合成サービスで、2023年8月31日時点で利用できる音声モデルは日本語が大半を占め (5,000 種類以上)、英語向けのモデルが 4 つ、中国語向けのモデルが 2 つ存在した。声優やアナウンサーを含む著名人 (ひろゆき氏、フリーアナウンサーなど) の声を利用することができるだけでなく、サービスを利用することで特定の話者の声を学習したカスタム音声モデルを作成し、非公開で利用す

ることもできる。カスタム音声モデルの作成は 1,000 円からと低額である。

音声合成の品質は利用する音声モデルの種類によってバラツキがある。無料で使用できる音声モデルは、GCTTS と比べて機械的な音声になっている印象が強い。

利用料金は月額制で、Free、Standard、Enterprise の3つのプランがある。合成した音声を商用利用できるのは月額 3,000 円の Standard プラン以上であり、Web API としての機能を利用するためには Enterprise プランの契約が必要となる。Enterprise プランは、最短契約期間は 2 か月で、月額 100,000 円の利用料に加えて、100 万文字ごとに 50,000 円が課金される。合成音声の速度の調整や間隔を空けるなどは、入力するテキストの句読点や鉤カッコなどの記号を利用することができる点が日本語に特化した特徴を表している。

6.1.3 音声合成技術の選定

GCTTS は、DL を用いた自然な音声合成技術を利用している点や、SSML での音声合成リクエストにも対応している点、12 ヶ月以降も無料枠が存在する点などの特徴があり、比較的低予算で利用できる利点がある。一方、CoeFont は特定の話者の声を学習することができる点が特徴的であり、今後様々な AI エージェントを開発する際には大きなメリットを感じるが、Web API として利用するためには月額 100,000 円と GCTTS に比べて高額である。これらのことを踏まえ、本研究の音声合成技術としては GCTTS を採用することとした。

6.2. 音声認識技術

音声をコンピューターが認識する場合、大きく分けると 2 つの工程に分けられる。

1. VAD 音声区間検知
2. Speech-to-Text 音声文字変換

6.2.1 VAD 音声区間検知

VAD (Voice Activity Detection) 音声区間検知とは、マイクから収録した音 (音声信号) の中から、発話者の発話開始と終了を検知する技術である。基本的な仕組みは、音声信号の大きさ (エネルギー) を計算し、一定の閾値を超える場合に音声が存在すると判断するが、音声以外の環境音や雑音を検知しないようにするためには、エネルギー計算だけでなく、周波数特性やスペクトル分析などの追加の処理を行うことが重要である。近年では DL の発達により、機械学習モデルを使用して、音声と非音声の区別をより正確に行うこともできる。機械学習モデルを用いた VAD は、一般的に従来のものよりも精度が高いため、AI エージェントツールにも機械学習モデルを用いた VAD を使用することが望ましい。

AI エージェントシステムをブラウザ上で動作する Web アプリケーションの形式で開発する場合、AI エージェントシステムに用いる VAD は JavaScript 上で動作する必

要がある。基本的に JavaScript がブラウザ上で動作する際は、Python のような機械学習に強い言語と比べて GPU 等の利用環境が充実していないため、計算に使用できるリソースが少ないことが多い。そのため、軽量で効率的な VAD を選択する必要がある。そこで考えられる方法は以下の 2 つである。

1. JavaScript のみで実装する
2. JavaScript+WebAssembly で実装する

1 の方法で実現できる場合、実装難易度は低いが、機械学習モデルを用いた VAD の実行速度が遅いというデメリットが考えられる。2 の方法の実装難易度は高いが、Python 向けの機械学習モデルを利用することができるなど、選択肢が広がり、実行速度が速いというメリットがある。実際に 1 の方法で実現できるかどうかを調べたところ、AI エージェントツールの音声認識機能に利用できる機械学習モデルを用いたライブラリを見つけることが出来なかった。一方、2 の方法では「vad-web」という機械学習モデルを用いた VAD を実行できるライブラリを見つけることが出来た。「vad-web」は WebAssembly を通じて、本来は JavaScript 上で動作しない ONNX (Open Neural Network Exchange) 向けの VAD モデルである「Silero VAD」を利用することが出来る。「Silero VAD」は音声と非音声を識別する機械学習モデルであり、音声信号の大きさで音声区間検知を行うことと比べて、より精度が高いという特徴がある。

6.2.2 Speech-to-Text 音声文字変換

Speech-to-Text (以下、STT) とは、音声データをテキストデータに変換する技術である。音声認識は AI エージェントシステムの音声認識機能に不可欠であり、正確性と高速性が求められる。STT の実装にもいくつかのアプローチがあるが、VAD のように JavaScript+WebAssembly で実装する方法がなく、現時点では以下の 3 つの方法が考えられる。

1. STT のライブラリを利用する方法
 - (ア) 利用者の PC 上で STT を実行する場合
 - (イ) 外部サーバー上で STT を実行する場合
2. WebAPI 形式の STT サービスを利用する方法

STT のライブラリの選択肢には、Julius や Whisper などがあるが、STT の精度の高さから Whisper の利用が望ましい。Whisper は無料で使用でき、PC やサーバーを動かす以外の費用は発生しないというメリットがある。1 (ア) の場合は、AI エージェントシステムの利用者の PC 上に Python の実行環境を構築する必要があるが、1 (イ) や 2 の方法とは異なり、インターネットを経由した外部サーバーと通信する必要が無いため、STT の処理速度向上が見込めるというメリットもある。そして 3 つの方法の中では最も安価な方法である。しかし、Python の実行環境を利用者の PC 上に構築することは難しい場合があり、特定の PC 上での動作に限定されると、多くの利用者に AI エ

ージェントシステムを提供する上では大きなデメリットとなる。一方、今後 AI エージェントツールをブラウザ (つまり JavaScript) 上で動作させる計画であることを考慮すると、Python の実行環境を必要としない 1 (イ) や 2 の方法が、より多くの環境での導入に適している。

外部サーバーとして使用する予定のインフラである「Google Cloud Run」は、サーバーが一定時間利用されていない場合はサーバーがシャットダウンするように設定でき、利用時間のみ課金されるというメリットがある。しかし 1 (イ) の方法では、シャットダウンしたサーバーの起動時 (コールドスタート) に毎回 Whisper ライブラリをダウンロードする必要があり、Whisper の音声認識モデルのダウンロードには約 12 秒の時間がかかる。そのためコールドスタートが発生すると、STT の処理が約 12 秒遅延するという問題がある。この問題を解消するには、シャットダウンしない常時起動のサーバーを用意する必要があり、この場合は 1 日約\$2、30 日で約\$62 の費用 (Google Cloud Run vCPU2 メモリ 4GB を常時起動した場合) が発生する。しかし、AI エージェントツールの利用者が少ないときは、この費用が無駄になることが懸念される。

WebAPI 形式の STT サービスを利用する際 (2 の方法) は、セキュリティの観点から AI エージェントシステムが動作する PC から直接サービスにアクセスさせずに、外部サーバーを介して独自の認証システムを経由することが推奨される。そのため、2 の方法でも 1 (イ) と同じく「Google Cloud Run」を使用して外部サーバーを用意する必要がある。しかし、2 の方法では、1 (イ) の場合のように音声認識モデルをダウンロードする必要がないため、コールドスタート時の遅延は発生しない。そのためサーバーを常時起動する必要がなく、1 (イ) に比べてより安価に運用ができる。なお、利用可能な STT サービスとしては、Google Speech-to-Text や Whisper API などが挙げられる。その中で最も安価であり、音声認識性能も高い Whisper API を利用する事が望ましい。特に AI エージェントシステムの開発中や試験導入期間など、サービス利用者が限られているようなフェーズでは、費用効率と性能を考慮すると、STT サービスを利用しない時間には費用が発生せず、多くの利用者へ導入しやすい 2 の方法が最適だと考える。

6.2.3 音声認識技術の選定

6.2.1 および 6.2.2 より、AI エージェントシステムの VAD としては「vad-web」ライブラリを利用することとした。ただし、ブラウザ上で動作する機械学習モデルを用いた VAD は、2023 年 9 月の時点では「vad-web」の他に有効な選択肢がないため、他に同様のライブラリが公開された場合は再検討する必要があると考えられる。

音声認識技術としては、WebAPI 形式で提供される Whisper API を外部サーバーで使用方法を採用する

こととした。

7. AI エージェントの試作開発と評価

第3章から第6章の調査や検討を踏まえて、我々はAIエージェントのプロトタイプを試作した。AIエージェントの名前は「さつきさん」で、さつきさんは、第5章で作成したシナリオに基づいて、第6章で選定した Google Cloud Text-to-Speech による音声で対象者に語りかけ、Open AI 社の Whisper を用いてその回答を聞き取る。ただしこの聞き取りは、現状では、対象者の発話開始と発話終了のタイミングを検知しているだけで、聞き取った内容を形態素解析などを用いて素材化し、次の発話に利用しているわけではない。

試作した AI エージェント (さつきさん) をデイサービスや小規模多機能居宅介護施設で実際に認知症高齢者の介護に従事する介護スタッフに試用して頂いて、日ごろから認知症高齢者の行動や反応を間近で見ている介護のプロフェッショナルとして、認知症高齢者がさつきさんとの対話でどのような反応をすると考えられるか、さつきさんは認知症高齢者の発話を促進することができそうか、などについてヒアリングを行った。紙面の関係で、ヒアリング項目の一部を以下に示す。

- =====
1. 「さつきさん」との会話は楽しかったですか? 「さつきさん」との会話についてご意見をお聞かせください。
 2. 「さつきさん」とどんなことを話してみたいですか?
 3. 将来的にどのような利用方法が良いと思いますか? 「さつきさん」をどのように利用してみたいですか?
- =====

複数の介護スタッフから得た評価は①もっと言ったことに反応して欲しい、②名前を呼んで欲しい、の2点である。①については、音声認識によって聞き取った内容を形態素解析などを用いて単語に分解して復唱できるようにすること、音楽療法の現場で交わされる会話データを収集して反応パターンのバリエーションを増やすことで対応していきたい。②については、音楽療法の現場では非常に重要視されていることなので、できるだけ早く対象者の名前を聞き取って、お名前で呼びかけながら話しかけられるように改良したいと考えている。

「さつきさん」との会話について、各介護スタッフから得られた良い評価は、③自分の言ったことを広げてくれる感じが良いと思う、④身近な質問を選んでいるところは良い、⑤私はこういうのが好きですよ、などと言っていたのは親しみやすくて良かった、である。天気の話から雲を挙げ綿あめを連想させることや、天気や綿あめなどの身近な話題を選んでいること、そして対象者の回答を受けて、さつきさんが自分のことを少し話すという流れは好意的に捉えられたようである。

一方、各介護スタッフから得られた指摘事項は、⑥感情が足りない。事務的にしゃべっているように感じる、⑦共感するという感覚が足りない感じがする。「私もです」って言ってもらえると良い、⑧一方的に聞かれるだけではなくて、こちらからも話せ

ようになると良い。まだ心を許せない感じがしてしまう、である。これらの指摘に共通するキーワードは「共感」だと考える。今回の、介護スタッフに対するヒアリングでは、介護スタッフの業務時間内に実施したために時間的な制約があり、さつきさんと十分に対話を重ねることができなかった部分があるが、音楽療法士は「寄り添う」ということをとても重視するので、「寄り添ってかれている」という印象を与えられるような話しかけ方、声の出し方などについて調査を行って、改良していきたいと考えている。

将来的な利用方法としては、⑨施設など複数の利用者さんがいるところで会話できるようになると良い、⑩日々のデイサービスの中で、毎日の朝の会とかで活用できるとよい、⑪スタッフが記録を残したりしている間に AI が利用者さんの相手をしてくれるとよい、⑫独居の高齢者の方が受け入れやすいと思う、などの意見が得られた。集団と対話ができるようになるのは非常に難しいので、まずは独居の高齢者に使って頂けるように完成度を高めることに注力したい。そうすれば独居の高齢者だけでなく、介護施設などで、介護スタッフの方が記録を残したりしている間に、特定の利用者さんの相手ができるようになると考えている。

8. まとめと今後の展開

本論文では、「親しみや安心感を感じさせる「話しかけ」ができる AI エージェントの開発研究」について述べた。少子化による介護者不足が懸念されるいま、特に対象者が増え、かつその介護が難しいと言われている認知症高齢者については、介護者の確保とその負担軽減が重要な課題となっている。そして、その課題解決のための技術利用に対する期待は高まっている。本研究では認知症の介護負担に関連する症状の1つである「無為・無関心」を改善するために、認知症高齢者の自発的な発話を促進する AI エージェントの開発を進めている。

現在開発中の AI エージェントは ELIZA 効果を参考に、認知症高齢者が AI エージェントに親しみや安心感を感じ、その結果として自発的な発話が促進されることを目指している。そこで本研究では「問いかけ」と「肯定的な回答」を繰り返す発話シナリオを作成し、それを Google Cloud Text-to-Speech による音声で対象者に語りかけ、Open AI 社の Whisper を用いてその回答を聞き取る、AI エージェントのプロトタイプシステムを開発した。

開発したプロトタイプシステムは、実際に認知症高齢者の介護に従事する介護スタッフに試用して頂いて、AI エージェントとの会話についてヒアリングを行った。その結果、「私はこういうのが好きですよ、などと言っていたのは親しみやすくて良かった」などの好意的な意見を複数頂いた。「共感が足りない」といった問題点の指摘も多数頂いたが、将来的な利用方法については、「スタッフが記録を残したりしている間に AI が利用者さんの相手をしてくれるとよい」をはじめ、様々なアイデアや改良のポイントを確認することができた。

今後は、名前を呼びながら会話できるように改良すると

共に、対象者の回答の中から単語を拾って、それを復唱できるように改良したい。そして最終的には拾った単語に関連する話題を提供できるように改良したいと考えている。また、対象者の返答に問題があると考えられる場合の対応シナリオの作成も課題である。認知症の場合は、「物盗られ妄想」と呼ばれる、本当は盗られていないのに、大事なものを盗られたと訴える症状が出ることもあり、このような場合は肯定的な回答をするのは避けるべきである。認知症専門医や認知症高齢者を対象とした音楽療法の臨床経験が豊富な音楽療法士の指導を仰ぎながら対応シナリオを作成したいと考えている。

謝辞

AI エージェントのプロトタイプシステムの開発に協力して下さった株式会社 dott の清水俊之介氏、AI エージェントのシナリオ作成に協力して下さった NAT 音楽療法事務所の認定音楽療法士の相川直子氏に感謝いたします。また、介護スタッフのヒアリング調査にご協力いただいた、「福祉ネットワーク・やえやま」の當山房子先生と、「あかゆら」の介護スタッフの皆さまに感謝いたします。本研究は新潟医療福祉大学学長裁量研究費、AMED の課題番号:JP21dk0207056、専修大学情報科学研究所共同研究助成の助成を受けました。

参考文献

- [1]東中竜一郎, 他. Python でつくる対話システム, オーム社, 2020
- [2]認知症ちえのわ net, <https://chienowa-net.com/>.
- [3]「臨床精神医学」編集委員会編集, 精神科臨床評価検査法マニュアル, 2004
- [4]日本神経学会監修, 認知症疾患臨床ガイドライン 2017
- [5]内閣府, 令和 4 年版高齢社会白書, http://www8.cao.go.jp/kourei/whitepaper/w-2022/zenbun/04pdf_index.html
- [6]カール・ロジャーズ (著), 畠瀬稔 (監修), 加藤久子/東口千津子 (翻訳), ロジャーズのカウンセリング(個人セラピー)の実際. コスモスライブラリー, 2007
- [7]斧原藍, カール・ロジャーズに学ぶ「聴く姿勢」. <https://www.ritsumeihuman.com/essay/essay-2405/>
- [8]一般社団法人日本音楽療法学会, <https://www.jmta.jp>.