

音声対話エージェントを利用した認知症患者の状態把握支援の試み

Assessing the Status of Elderly People with Dementia Using Spoken Dialogue Agent

齋藤直子*¹
Naoko Saito

岡田将吾*¹
Shogo Okada

新田克己*¹
Katsumi Nitta

林佑樹*²
Yuki Hayashi

中野有紀子*²
Yukiko Nakano

*¹ 東京工業大学大学院総合理工学研究科
Tokyo Institute of Technology

*² 成蹊大学理工学部
Seikei University

The number of elderly people with dementia is rapidly increasing in our country. They often worry their family or their care worker in the conversation because they have difficulty in their memory, speech, etc. To support them, we have developed spoken dialogue agent that can analyze the cognitive status of the patients based on their utterances observed in conversations with the agent. In this study, we introduce the agent's question scenario that have a tree structure. For increasing utterance of patients, we prepare the scenario consists of multiple topics and the agent select the question's topic that be adapted to user's preference or background. In addition, we analyze nonverbal information by observing utterances and videos of patients, and report the result of the analyzing.

1. はじめに

社会の高齢化に伴い、認知症患者の数が増加傾向にある。患者の世話をする家族や介護士は、患者との日常的な対話から身体的および精神的健康状態を把握する。しかし記憶障害や見当識障害等により対話能力の落ちた患者と対話を行うことは容易ではない。同じ内容のことを何度も話し聞かされたり、投げかけた質問トピックとは異なるトピックの応答をされたりすることで、対話の相手には負担となる。そこで我々は、家族や介護士に代わり患者の話を傾聴する音声対話エージェントの開発に焦点を当て、開発を進めてきた。対話エージェントを導入することによる、アプリケーションの一つとして本研究では患者とエージェントの対話内容をレポート化し、作成したレポートを家族や介護士が見たときに患者の身体的および精神的状態を把握することの支援を目指す。

患者に関する詳細な状態把握を行うことのできるレポートを作成するには、患者の発話を促すと共に、より多くの情報を含む発話を得る対話戦略が必要である。我々が開発した対話エージェント[比企野 2011](図1)は、質問シナリオのエージェント質問項目を固定していた。そのため、質問シナリオ内にある質問トピックに関する患者の興味と記憶量に依存して、患者の発話量が増減していた。また、予め決められた質問内容に対する発話しか得られないため、患者の発話から得られる情報が単調であった。本研究では、複数のトピックから構成されるツリー型の質問シナリオを導入することにより、対話トピックをユーザ個人に適応させ、多数のトピックに関する患者の発話を誘発する戦略を取る。ここでエージェントはユーザの発話状態・聞く姿勢などに応じて、適宜質問内容のトピックを切り替えたり、場合によっては質問せず、相手の発言を傾聴する必要がある。

一般的なユーザとの対話において、ユーザの発話状態の推定により、システムが発話するか否かを決定したりトピックを切り替えたりするためのモデルは[神谷 2010]、[西村 2006]などで提案されている。しかし認知症患者がユーザの場合、どのようにユ



図1: 対話エージェントの外観

ーザ状態の推定を行い、エージェント発話の判断をしたり、話題を切り替えたりすればよいかは自明ではない。そこで本論文では[比企野 2011]で取得された、患者とエージェントとの対話実験データから、患者の発話状態や聞く姿勢に関する非言語情報を分析し、エージェント発話タイミングの判断や、話題の切り替えのモデルを構築するための手がかりを考察する。

2. 先行研究

認知症患者に限らず、高齢者と対話を行うシステムは様々な研究がなされている。[山本 2010]では、患者との自然な対話を目指し、予め用意しておいたシステムとユーザ間の対話履歴の連鎖構造を利用している。対話履歴に基づき、システムは最も遷移確率が高い応答を選択する構造となっている。認知症患者は症状の進行程度により、利用時点ごとに同じ質問内容でも異なる内容の応答をする可能性がある。そのため新たな発話が対話履歴に存在せず、遷移確率が求められないことが頻発すると考えられる。[Yasuda 2013]では、認知症患者に特化した対話システムの構築を行った。ここでは、患者のプロフィールを予めシステムに入力しておき、患者に適応したトピックでの対話を実現した。患者の性別や趣味を考慮したトピックで対話を行うこ

とで、患者の発話量を増やすことを目的としている。また、エージェントの質問に対する患者の応答までの時間は、患者毎または質問毎に個人差が大きい。これを考慮し、エージェントが患者の発話を待機する時間を、対話開始前に調整できる機能を搭載している。これらの機能をもとに対話実験を実施し、システムの有効性を確認している。しかし患者のプロフィールや認知症の進行程度が変化するとシステムの調整を行う必要があるため、本研究では自動的にトピックを適応することでより有益なシステムを目指すことを考えた。

一方、対話エージェントと高齢者の対話分析を行ったものとして、高齢者が対話システムと話すトピックに関する研究が存在する[Vardoulakis 2012]。これは、人間が遠隔操作を行っている対話エージェントと独居高齢者に日常的に対話を行ってもらい、その対話内容を分析した研究である。Wizard of Oz 法を用いた対話実験の結果、高齢者が話した頻度の高いトピックは「天気」「家族」「Storytelling」であった。本研究では、トピック選択に際しこの3トピックを参考に質問シナリオを構成した。

3. 認知症患者のための対話エージェントシステム

我々は先行研究に基づき、認知症患者と非タスク指向型対話を行うエージェント(図1)を構築した[比企野 2011]。対話を円滑に進めるためには、エージェントのフィードバックが重要である。よって本エージェントでは患者発話の音声情報から適切なタイミングでの相槌や頷きを行い、ユーザの発話終了を待ってから次の質問に進むという機能を搭載した。過去に対話実験を実施した結果、適切なタイミングでエージェントのフィードバックを行うことの有効性を確認した。本エージェントを用いて、現在までに認知症患者 32 名、健常者 9 名のデータを収集した。さらに、患者の発話の言語情報からトピックと感情を推定し、対話内容をレポート化するシステムを構築した[齋藤 2013]。患者の発話内容を分析することで、患者の身体的および精神的な健康状態を把握する材料となるようなレポートを生成するシステムを実現した。

3.1 システム構成

既存の対話エージェントシステム[比企野 2011]では患者発話の非言語情報のみを扱っており、患者発話の言語情報についてシステムは認識していなかった。よって本研究では、まずエージェントに音声認識の機能を追加し、さらに先行研究[齋藤 2013]の対話レポート化システムを統合した。これにより、エージェントと患者の対話を行い、さらにレポート化する一連の流れを自動化した。

システムの構成を図2に示す。既存のエージェントでは、ユーザの発話音声のピッチと無音区間を検出し、エージェントの発話タイミングを図っていた。本研究では、さらに発話音声から得られる言い淀みや、録画した画像から得られる視線や頷き等の情報からエージェントの発話タイミングを図るモデルを検討している。ユーザ発話から一定時間が経過すると発話終了とみなし、検出した発話に音声認識を行う。ここでは Google 音声認識を使用している。その間にはエージェントは頷きや相槌を生成する。音声認識結果により質問シナリオの分岐を行い、次のエージェント発話内容を決定する。決定した発話内容を、ADOBE FLASH CS4 PROFESSIONAL で作成したアニメーションと共に合成音声として出力し、ユーザと対話を行う。音声合成には、SAP15 対応のスピーチエンジンである xpNavo2 を使用した。対

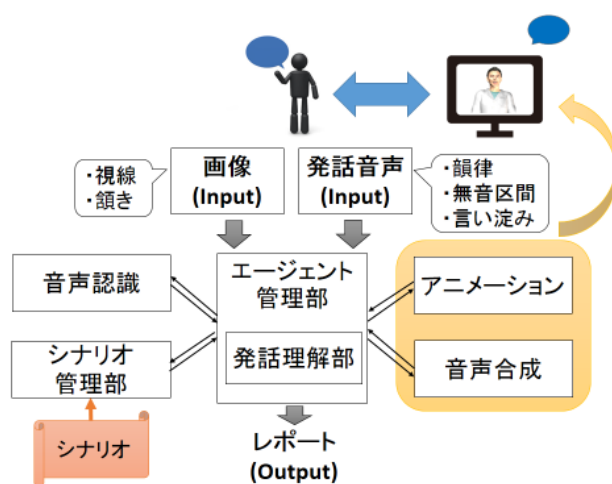


図2: システム構成図

話が終了したら音声認識結果や検出した非言語情報をもとに対話レポートを作成し、出力する。

対話レポートでは、先行研究[齋藤 2013]のシステムを利用し、音声認識結果から患者発話のトピックと感情の推定を行う。トピック推定により、患者の発話がエージェントの質問トピックと一致しているかを判定できる。これにより、各トピックに関する記憶の明瞭さや、患者の対話能力を推測する材料となる。また、感情推定を行うことで、患者の不安や興奮といった認知症の症状を検出する材料となる可能性が考えられる。これらの推定結果に、5 章で述べる非言語情報に関する知見を加えると、より家族や介護士にとって有益なレポートとなるのではないかと推察できる。

3.2 ツリー型質問シナリオの導入

患者の発話量を増やし、かつ多くの情報を含む発話を得るには、対話トピックを患者個人に適応させていく必要がある。認知症患者は特に、興味のないトピックや記憶が曖昧なトピックに関する発話量の少なさは顕著である。そのため本研究ではツリー型の質問シナリオを導入する。エージェントからの質問に対する患者の応答に音声認識を行い、その結果により分岐を辿り、患者個人に適応したトピックで対話を進める。しかし、患者は健常者と比較して音声認識結果の精度が低いことが経験的に分かっているため、患者の長い応答をシステムが正確に理解するのは困難である。そこで、エージェントの質問を「閉じた質問」と「開いた質問」の 2 種類の質問形式で構成する(表1)。本研究では、閉じた質問を「はい(肯定)」「いいえ(否定)」で答えられる質問、「開いた質問」を自由形式で答えられる質問と定義する。

表1: 「閉じた質問」と「開いた質問」の特徴

	閉じた質問	開いた質問
質問例	「朝ご飯は食べましたか？」	「お子さんは、どんな方ですか？」
患者の発話量	少ない	多い
音声認識の精度	良い	悪い

ツリーの末端のみ「開いた質問」、そのほかに「閉じた質問」を配置し、分岐は「閉じた質問」のみで行う。「閉じた質問」に対する応答では患者の発話長が短くなるのが予想される。これにより

音声認識の誤りによるシステムの発話理解の誤りを減らす狙いがある。患者の発話が「肯定」であるか「否定」であるかの判定には、言語情報を用いる。「肯定」を表す語群と「否定」を表す語群を作成し、患者発話に含まれる語と比較を行うことで判定する。こうして得られたトピックについて「開いた質問」を行い、患者からの自由な形式で応答を得る。ここでの「開いた質問」に対する応答は、ツリーの分岐結果によりトピックや前提条件が限定されている。そのため、音声認識結果の精度が低くなりがちな長い発話も、システムが発話内容を推測することが容易になると考えられる。さらに、ツリーの分岐を辿ることにより患者個人に適したトピックでエージェントが質問することができるようになる。これにより、患者の詳細な情報を含む発話を誘発する。

また、質問シナリオを作成するにあたり、高齢者がエージェントと話す頻度が高いトピックに関する研究[Vardoulakis 2012]と過去の対話実験データを参考にした。本研究では質問シナリオに含まれるエージェントのトピックを「天気」「Storytelling (患者の過去の話)」「家族」「健康」「食事」「趣味」の6個とする。各トピックについて複数のツリーを作成し、全てのツリーについて末端の質問まで到達したら、対話終了とみなす。

4. 対話中に表出する非言語情報の分析

4.1 エージェント・患者の対話データセット概要

本研究では、先行研究[比企野 2011]で得られた実験データを用いて、非言語情報の分析を行った。エージェントと対話を行ってもらった被験者は14名(男性6名, 女性8名)で、60~90歳(平均年齢77.8歳)の認知症患者である。認知症の進行程度に関する指標であるミニメンタルステート検査(MMSE)の平均得点は22.6点であり、軽度の認知症である患者が多い。対話では質問一応答が33ターン行われ、対話中における患者の録画と、患者発話の録音を行った。対話実験の様子を図3に示す。

4.2 分析対象の非言語情報の種類

本システムでは、エージェントの発話タイミングをはかるために患者の音声情報を利用している他は、言語情報のみを取り扱っている。しかし音声情報をはじめ視線や頭部動作といった非言語情報は、発話タイミングを図る以外にも様々な状態を認識でき、また出力レポートに反映させることができる。

認知症患者の症状は、言語情報と非言語情報の両側面に表れる。例として、不安や興奮といった精神症状は発話の言語的



図3: 対話実験の様子

内容と韻律情報に表れることがある。既存のエージェントでは言語情報のみ利用し感情推定を行っているが、非言語情報と合わせて推定を行うことで精度向上が望める。このように、認知症患者の状態把握を行う上で、非言語情報の認識は重要である。

本研究では、言語情報と非言語情報の両方を認識し活用することができる対話エージェントを目指し、過去の対話実験のデータについて非言語情報に関する分析を行った。

4.3 実験データ分析結果

本研究ではこれらのデータを用いて分析を行った。分析においては、エージェント発話と患者発話のタイミングに焦点を当て、次の2点に着目した。

- 患者発話とエージェント発話のオーバーラップ
- 患者が発話する際の言い淀み(無声休止)

分析の結果、患者の発話とエージェントの発話との間で、対話1セッションにつき平均7.4回のオーバーラップが存在した。そのうち約90%がエージェントの相槌と患者の発話がオーバーラップしており、残り約10%はエージェントの質問と患者の応答のオーバーラップであった。

発話中の言い淀みは、様々な種類に分類できる。例として有聲休止、無声休止、音節の引き伸ばし、言い直し等が挙げられる。本研究での分析においては、この中から無声休止を取り上げる。なぜなら、本研究で扱う対話エージェントは患者の無音区間を検出することで患者の発話終了とみなしているため、患者が音を発している限りは発話終了と判定しない。つまり、患者が有音休止や音節の引き伸ばしのために言い淀んでいる時間についてエージェントは患者の発話終了を待機することができる。よって、本エージェントが自然な対話の流れを保ち、かつ患者が発話をする妨げにはならないと考えられる。そこで無音休止とその他の言い淀みの回数について、対話実験を記録した動画を分析した。ここでは、首をかしげる動作やエージェントから視線を外し何かを考えているような動作をしたときに無声休止であると判定した。分析結果の例を表2に示す。

表2: 言い淀みに関する分析結果

患者 No.	言い淀み	
	無声休止	その他
1	27	7
2	9	6
3	18	4
4	10	7
5	21	1

5. トピック選択・エージェント発話タイミングモデル構築のための考察

5.1 非言語情報の分析結果からの考察

まず患者発話とエージェント発話のオーバーラップに関して、エージェントの相槌生成のタイミングに問題があると考えられる。オーバーラップが起ると、患者は発話を途中で止めてしまい、対話の流れが止まってしまう。そして全体の発話量の低下に結びついていると考えられる。現在本システムでは患者の発話後に一定の無音区間を検出すると相槌や次の質問を生成する仕組みとなっている。よって、この無音区間の後にエージェントが

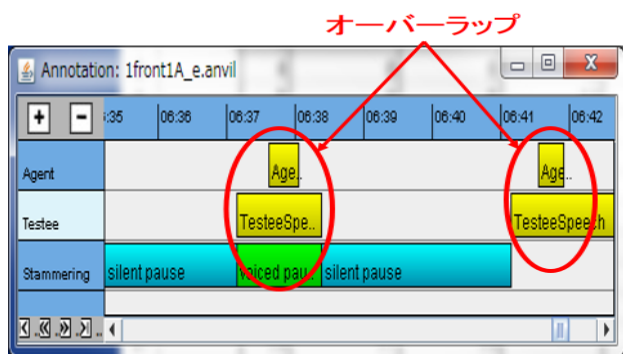


図4: 患者発話とエージェント発話のオーバーラップ例

頷きや質問を生成するタイミングと、患者が次の発話をしようとしたタイミングが一致するとオーバーラップが起きてしまう(図4)。しかし動画を観察すると、このオーバーラップ数は個人差が大きく、前述のタイミングが一致しない患者についてはオーバーラップが全く起こらないこともあった。つまり、患者の発話タイミングは個人差が大きいので、予め一定値に決められた無音区間を検出するのではなく、個人に最適した発話タイミングで質問や相槌を行うことがエージェントの機能として必要である。

次に、言い淀みの中でも無声休止を表す非言語情報については、「首を傾げる動作」と「エージェントから視線を外す動作」の2パターンが多いことがわかった。よって頭部動作や視線等の情報から無声休止を検出することで、患者が答えを考えている間にエージェントが次の質問を投げかけてしまう回数を減らすことができると考えられる。しかし認知症患者は、質問内容に対する記憶が曖昧であったり、考えている間に質問内容がわからなくなってしまうことがある。よって無声休止に関しては、休止が長時間になったときには次の質問を生成してしまう等の対応が必要であると考えられる。また今回は無声休止についてのみ分析を行ったが、有声休止や言い直し等も有益な情報である。これらの言い淀みは、言語障害や記憶障害に直結するデータとなる可能性が高いため、本システムにおけるレポート化の際に重要である。

また対話全体について、患者の対話中の動作における非言語情報のパターンは一貫して変わらないことがわかった。例として、発話と頷きを同時に行う、発話するときにエージェントから視線を外す等があげられる。これらの特性を対話の始めにシステムが認識することで、その特性から外れた行動を患者が行ったときにその変化点を検出して対話戦略を変えるなどの対応をとることができる。これにより更に個人に最適した対話システムになると考えられる。

5.2 非言語情報の自動認識モジュールの導入方法

前項で得られた知見から、エージェントの発話タイミング決定や質問トピック選択のため非言語情報の自動認識を行うモジュールの導入を考える。前項の分析結果から、以下の2点を検出することが有用であることが示唆された。

① 発話中の言い淀み(無声休止)

② 対話中の非言語行動についての変化点

これらを検出するためには、音声・視線・姿勢・顔向き・頭部ジェスチャ等の様々な非言語情報が必要である。①の言い淀みについては、音声のパワー情報から検出する方法が提案されている[緒方 2009]。しかし認知症患者の発話における音声情報の変化は少ない傾向にあるので、視線や顔向き等と組み合わせで検出を行う等の対応を検討している。また、②に関しては過

去の実験データを利用し、患者の非言語行動に関する定常状態に関するモデルを構築し、その状態から逸脱する状態を時系列変化点検出の手法を用いて検出する。分析結果より患者によって、頭部ジェスチャ・姿勢・顔向きなどの変化量が異なることが示された。この結果は、非言語に関する定常状態が患者によって異なることを意味している。この問題に対処するため、予め非言語情報の定常状態を複数クラスに分類しておき、新規の患者と対話する際、対話の初期段階で、その患者の定常状態の分類を行うことで、定常状態を判別する方法を取る。その後は例として韻律から対話の盛り上がりを検出したり、視線の外れ方から興味度の推定等を行ったりすることを検討している。

6. まとめ

本研究では、認知症患者の状態把握支援に向けた音声対話エージェントシステムの開発を行った。さらに、過去の実験データの分析から、今後システムに取り入れるべき非言語情報について考察した。

今後は、開発した対話エージェントを用いて実際に患者との対話実験を行いその有用性を評価する。また5章で考察した非言語情報に関する機能について検討していく予定である。

参考文献

- [比企野 2011] 比企野純太, 中野有紀子, 安田清:「会話エージェントを利用した認知症患者のためのコミュニケーション支援」, 第 73 回情報処理学会全国大会論文集, 4.195-4.196, 2011.
- [神谷 2010] 神谷優貴, 大野誠寛, 松原茂樹, 柏岡秀紀:「同調的対話システムにおけるあいづち挿入タイミング」, 言語処理学会第 16 回年次大会発表論文集, 2010.
- [西村 2006] 西村良太, 北岡教英, 中川聖一:「対話における韻律変化・タイミングのモデル化と音声対話システムへの適用」, 人工知能学会言語・音声理解と対話処理研究会 48,37-42, 2010.
- [Yasuda 2013] K. Yasuda, J. Aoe, M. Fuketa: "Development of an Agent System for Conversing with Individuals with Dementia", The 27th Annual Conference of the Japanese Society for Artificial Intelligence, 2013.
- [山本 2010] 山本大介, 小林優佳, 土井美和子:「高齢者対話インタフェース-対話誘導による問診対話」, HAI シンポジウム, 2010.
- [Vardoulakis 2012] L. P. Vardoulakis, L. Ring, B. Barry, C. L. Sidner and T. Bickmore: "Designing relational agents as long term social companions for older adults", in Intelligent Virtual Agents, 2012, pp.289-302.
- [齋藤 2013] 齋藤直子, 林佑樹, 中野有紀子, 安田清:「認知症患者の状態把握支援のための発話内容認識手法」, 第 27 回人工知能学会全国大会, 2013.
- [緒方 2009] 緒方淳, 後藤真孝, 伊藤克亘:「有声・無声休止区間の自動検出を考慮したデコーディングによる自由発話音声認識の性能改善」, 電子情報通信学会論文誌 D 情報・システム 92.2, 2009.