

認知症患者のための語りかけエージェント —患者の支援と状態把握を目指して—

中野 有紀子*¹, 酒井 洋一*², 野中 裕子*²

A Conversational Agent for People with Dementia: Supporting and Assessing the Patients

Yukiko NAKANO*², Yoichi SAKAI*², Yuko NONAKA*²

ABSTRACT : With the goal of supporting people with Dementia and reducing the caregivers' physical and mental burden, this paper proposes a conversational agent that can serve as a conversation partner for people with Dementia. First, we developed an autonomous conversational agent that can generate backchannel feedback during the user's speech, such as head nod and verbal acknowledgement based on the acoustic information in user's speech: pitch and intensity. Then, we propose an assessing system for people with Dementia which recognizes the users' speech and head movements and records their verbal and nonverbal behaviors during the conversation with the agent. The behavioral records are expected to be useful for the caregivers to know the mental status of the patients.

Keywords : conversational agent, Dementia, listener agent, feedback, speech

(Received September 21, 2012)

1. はじめに

高齢者人口の急速な増加に伴い、高齢者用施設、病院、家庭等様々な場面において、高齢者への支援が必要とされている。しかし現状では介護者が不足しており、高齢化する日本社会において重大な問題となっている。特に、認知症患者は高齢であることに加え、記憶障害のために、薬の飲み忘れ、同じことを何度も尋ねる、時として精神的に不安定な状態になるといった状況が起り、家族を中心とした介護者の負担が特に大きい。

一方、情報技術やロボット技術による高齢者の支援が試みられており、高齢者の自立支援、見守り、高齢者間のコミュニティづくり等を目的とした、社会的ロボットの研究・開発がすすめられている[1, 2]。これらを実現するためには、ロボット技術に加え、センシング技術、通信技術等の情報技術を統合したシステムを開発することが必要となる。また、認知症患者のためには、特に薬の飲み忘れなどに関する注意喚起、精神的安定を保つため

の見守り、さらには認知症患者の状況を評価し、介護者に通知する等の機能を実現することが重要であると考えられる。

そこで本研究では、患者に精神的安定を与えるとともに、認知症患者の介護者の負担を軽減させることができる情報通信技術として、認知症患者の話し相手となる擬人化エージェント(アニメーションキャラクタ)を提案・実装する。我々はこの擬人化エージェントを「語りかけエージェント」と呼ぶ。擬人化エージェントを用いることにより、人間同士の会話と類似した反応をユーザから引き出すことができることが分かっており[3]、この特性を利用することにより、将来的には、患者が日常と違う様子であった場合に、これをエージェントとの会話を通して自動的にセンシングすることができる可能性もある。また、通信システムと統合すると、これを介護者に通知することもできるであろう。本研究では認知症患者とその家族の方々にご協力をいただきながら、データ収集、分析、評価実験を重ね、そこから得られた知見を統合した認知症患者支援システムを提案する。

*¹: 理工学部情報科学科准教授 (y.nakano@st.seikei.ac.jp)

*²: 理工学研究科理工学専攻情報科学コース修士学生

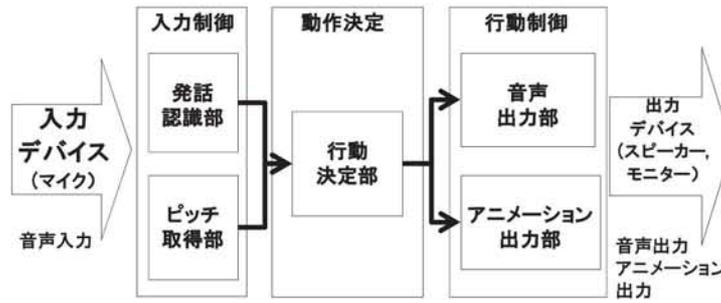


図1 語りかけエージェントのシステム構成



図2 エージェントアニメーション

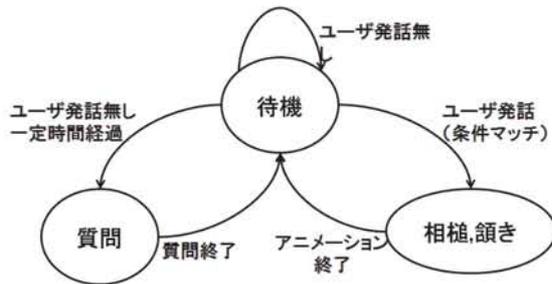


図3 エージェント行動の状態遷移モデル

2. 語りかけエージェントシステム

本研究では、最も基本的な社会的シグナルとして、相手（患者）の発話中に相槌（「ええ」「はい」）をうち、無言での頷きを表出する機能に加え、相手の発話終了を待つて応答を返すために、発話終了認識機能を実装した。システムの構成を図1に、エージェントアニメーションの例を図2に示す。

入力制御モジュール：発話認識部ではマイクから入力された音声の音量から、発話入力の有無を判定する。声の大きさは人によって異なるので、判定の閾値は適宜調整することとした。次に、ピッチ取得部では、発話入力であると判定された音声に対して、ピッチを算出する。これらの値は動作決定モジュールに出力される。尚、発話

表1 フィードバック生成ルール

条件	エージェントの行動
現在のピッチ n が $n-1$ から $n-16$ までの平均ピッチより 5% 低下	頷き
発話終了後 2 秒以内の無音があり再度発話が発出される	相槌(ええ)
発話終了後 2 秒以上の無音 or 発話終了直前のピッチ下り幅が 2% 以内 or 発話終了直前のピッチが発話平均ピッチより 10% 以上上昇	相槌(はい)

入力の有無の判定のためのマイク入力の音量取得には ActionScript を利用し、1 秒以上の無音区間が検出されると、そこで発話が終了したとみなした。また、ピッチ取得部は Pitch&RhythmMonitor を使用した。

動作決定モジュール：行動決定部では、入力制御モジュールから得られた音声情報を用いて、エージェントの状態の更新と行動の決定を行う。図3にエージェント行動の状態遷移図を示す。エージェントは質問を実行すると待機状態に入り、発話認識部においてユーザの発話が発出されると、ピッチ取得部から得られた韻律の情報をフィードバック生成ルール（表1）に適用し、相槌と頷きを決定する。これらのルールは人対エージェントのコミュニケーションにおけるエージェントの適切なフィードバック行動を提案した[4]と人間同士の会話を分析した[5]を参考にして作成した。頷きは最新の 16 個のピッチ情報から決定されるため、ユーザの発話途中にも実行されるが、相槌は発話終了ごとに決定される。相槌や頷きが実施されると、また待機状態に戻る。エージェントの質問後、一定以上の時間が経過してもユーザからの応答がない場合には、次の質問に移行する。

行動制御モジュール：行動制御モジュールは、キャラクターのアニメーションの実行と音声ファイルの再生を行うが、これにはWoZ実験のために実装した機構をそのまま利用した。

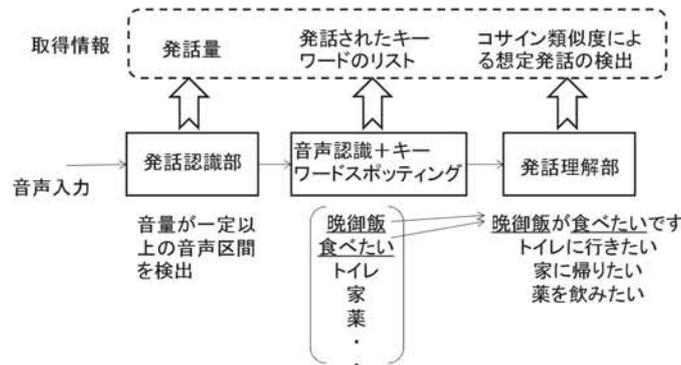


図4 音声入力からの発話量、言語情報の取得

3. 語りかけエージェントの改良

患者の認知状態の測定・評価の機能を実現するためには、ユーザがどのようなことを話し、活発にコミュニケーションを行っていたか否か、あるいは通常とコミュニケーションの仕方が異なっていたかを検出する必要がある。また、統合的な支援システムとしては、それを介護者に通知し、患者の状態を知る手助けとなる必要となる。そこで本節では、語りかけエージェントを発展させることにより、ユーザである認知症患者の状態を評価する指標となるデータを取得し、介護者が患者の様子を知る手助けとなる統合システムを提案する。

具体的には、図1に示す語りかけエージェントの入力制御モジュールを拡張することにより、ユーザの発話からの音声・言語情報、発話量、非言語的反応、コミュニケーション活発度の取得、また、動作決定モジュールの拡張により、ユーザ対エージェントの質問応答履歴の取得機能を実装した。また、これらの各機能について以下に実装方法を述べる。

(1) 発話量情報の取得：拡張した入力制御モジュールにおける音声処理の流れを図4に示す。入力制御モジュール中の発話認識部では、一定の音量以上の音声入力があると、発話と判定され、また音量が一定以下になると発話終了と判定する。この機能を利用し、患者の発話数と発話継続時間を計測し、これらの情報から患者の発話量を指標化することにより、普段と比較して発話量が極端に少ない場合にこれを検出することができる。

(2) 発話の言語情報の取得：発話認識部に音声認識機能を追加することにより、患者の発話中の言語情報を取得できるように拡張した。音声認識エンジンには、汎用大語彙連続音声認識エンジンjuliusを使用し、千葉労災病院の認知症患者のリハビリ中の訴えの内容の書き起こしを用いて、音声認識用の単語辞書・言語モデルを作成した。言語モデルの作成には、統計的言語モデルツールキ

ットPalmkitを使用した。図4に示されるように、まず、患者の発話音声を音声認識器により処理し、認識結果を保存する。これに加え、キーワードスポッティング機能を実装し、特定のキーワードが発声されたか否かを判定できるようにした。キーワードスポッティングは、各キーワードの音素列と音声認識結果内のN-gram音素列との編集距離を計算し、一定以下の編集距離、つまり音素列の類似度が一定以上であるキーワードが見つかった場合にはこれを出力する。高齢者であるために発音が不明瞭である場合もあるので、患者の状態を知る上で重要な単語を予め登録しておくことにより、文全体の認識には失敗しても、その中からキーワードを取り出すことができれば有用な情報となる。

さらに、想定発話をあらかじめ用意しておき、各想定発話内にふくまれる自立語の単語ベクトルとスポッティングされた自立語であるキーワードの単語ベクトルとのコサイン類似度を計算し、類似度が高い想定発話があった場合には、それを発話認識結果として出力する機能も実装した。

$$\text{Cos}(T, S) = \frac{U_T \cdot U_S}{|U_T| |U_S|}$$

ここで、Tは想定質問、Sは認識結果の文、 U_t , U_s は文T, Sに現れる自立語の頻度ベクトルを表す。この機能により、例えば、「ご飯が食べたい」、「足が痛い」のように頻繁に出現する発話がある場合は、これを想定発話として登録しておくこと、検出することができる。

(3) 非言語的反応の取得：音声認識を利用した言語情報の取得に加え、ディスプレイの上にWeb Cam等の小さなカメラを設置しておくことにより、患者の顔映像を収録することができる。システムは映像を収録するとともに、入力フレーム毎にヘッドトラッキングソフトウェアFaceAPIで処理することにより患者の頭部動作を計測し、蓄積することができる。患者の頭部の動作を分析してみると、図5に示されるように、相槌等を行いながらリズ

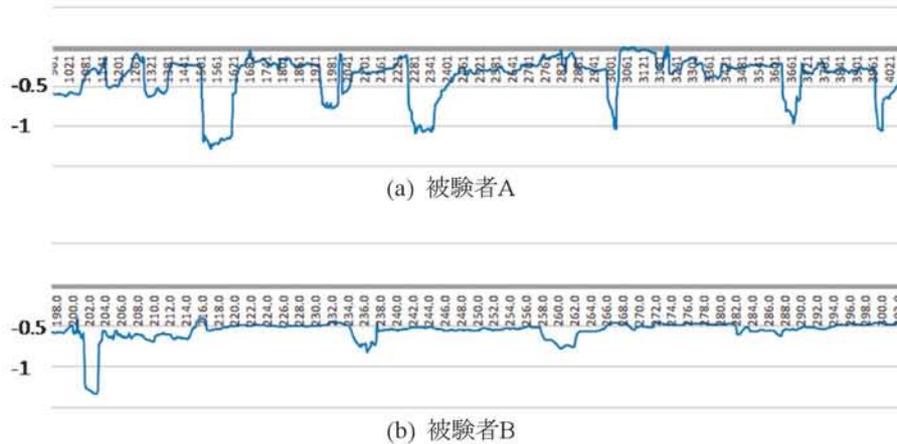


図5 2名の被験者における約100秒間の頭部の動き(X-rotation)

ミカルに頭部動作が行われている被験者(被験者A)と、ほとんど動きがみられない被験者(被験者B)がいた。

8名の被験者について、エージェントとの会話中の領き行動をアノテーションし、会話中の各領き行動の継続長の総計を集計した(これを領きの総量とする)。そして、領きの総量が10秒以下の被験者と10秒以上の被験者の2群において、会話中の被験者の発話割合を比較したところ、領き総量が10秒以上の群は10秒以下の群に比べて、被験者の発話時間の割合が大きくなる傾向がみられた($t(6)=-1.9999$, $p<0.1$)。この結果は、エージェントとの会話において領き行動が多くみられた被験者ほど発話が活発になることを示しており、頭部の動きや位置から領きを検出できれば、コミュニケーションの活発性を測ることができることを示唆している。また、頭部動作による会話参加態度推定方式を利用すれば[6]、頭部の動きから患者がエージェントとの会話に積極的であるか否かも推定できると考える。

4. おわりに

本稿では、認知症患者とその介護者を支援することを目的とし、アニメーションによる語りかけエージェントを提案・実装した。さらにこれを拡張し、認知症患者の音声・言語情報と頭部の動きをエージェントとの会話中に計測し、患者の状態を介護者がよりきめ細かく把握するためのシステムを実装した。

今後は、患者が興味を持つ話題選択するために、患者のプロフィールから興味を持ちそうな話題を推定したり、使用時の行動からシステムが学習する等の方法を検討する必要がある。また、本システムの使用において、患者はキーボードやマウスの操作を一切必要としないが、シ

ステムの開始と終了の操作を介護者や病院のスタッフに行っていただく必要がある。そのため、ブラウザ上でアニメーションを動作させる等の工夫をし、特殊な計算機環境を必要としないシステム設計を採用したが、今後は介護者のシステム操作をさらに簡便化するようシステムの改良を行ってゆく予定である。評価実験もさらに進め、継続的に本システムを利用していただき、より長い期間での評価実験を進めることにより、本システムの有効性の評価を行ってゆく予定である。

参考文献

- [1] Pollack, M., Intelligent technology for an aging population: The use of ai to assist elders with cognitive impairments, in AI Magazine. 2005. p. 9-24.
- [2] 米澤朋子他, 高齢者同士の TV 会話におけるユーザの参加状態に応じた介在型ロボット, in HAI シンポジウム 2010. 2010.
- [3] Reeves, B. and C. Nass, The Media Equation: How People Treat Computers, Televisions and New Media Like Real People and Places. 1996: CSLI Publications.
- [4] Gratch, J., et al., Virtual Rapport, in 6th International Conference on Intelligent Virtual Agents. 2006, Springer: Marina del Rey, CA.
- [5] Tsukahara, W. and N. Ward. Responding to Subtle, Fleeting Changes in the User's Internal State. in CHI 2001. 2001: ACM.
- [6] Ooko, R., R. Ishii, and I.N. Yukiko. Estimating a User's Conversational Engagement based on Head Pose Information. in IVA2011. 2011.