

## 高齢者の会話における反応性の定量化と可視化

2A1-NFC-6-10

— 会話エージェントを用いた高齢者の状況把握を目指して —  
 Visualization and Quantification of the Conversational Responsiveness of the Elderly  
 -Towards Understanding the Cognitive Status of the Elderly using Conversational Agents-

野中 裕子\*1  
Yuko Nonaka

酒井 洋一\*1  
Yoichi Sakai

\*1 成蹊大学大学院理工学研究科理工学専攻  
Seikei University

安田 清\*2  
Kiyoshi Yasuda

中野 有紀子\*3  
Yukiko Nakano

\*2 千葉労災病院/京都工芸繊維大学 \*3 成蹊大学理工学部情報科学科  
Chiba Rosai Hospital Seikei University

With the goal of measuring and assessing the cognitive status of elderly people, this paper proposes a method of measuring the responsiveness of elderly people in conversations with a conversational agent using speech and head motion data. The system is also capable of speech recognition and language understanding functionalities, which are potentially useful for evaluating the cognitive status of elderly people. We also propose two ways of visualization of the responsiveness; graph representation, facial icons. Iconic presentation is expected to be useful for the families to intuitively understand the results of the analysis.

## 1. はじめに

少子高齢化により高齢者支援の必要性が高まっている。しかし、高齢者を支える介護者の数は高齢者数に比べ不足しているのが現状である。特に認知症患者は記憶障害があるため、介護者の負担はさらに大きい。このような状況において、認知的障害を持つ高齢者の支援を目的とした情報技術の研究・開発がすすめられており、その中で、患者の状態把握や認知状態の評価は重要な課題の一つとなっている[Pollack 2005]。

我々は、認知症患者のための会話エージェントを開発し、すでにその有効性について評価実験を行っている [酒井 2012]。その結果、認知症患者はエージェントに対して抵抗なくコミュニケーションし、会話エージェントとの会話が成立することが確認された。このことから、認知症患者等の高齢者の支援に会話エージェントの使用することは有効であると考えられる。

そこで本研究では、高齢者の状況把握を行うための会話エージェントを開発することを目的とする。具体的には、質問することにより高齢者に語りかけを行う会話エージェントとの対話において、ユーザである高齢者の発話・傾き・頭部の動作から、エージェントとのコミュニケーションに対する反応性やコミュニケーション活性化度を定量的に評価する手法を提案する。さらに、介護者が高齢者の様子を知るために、定量化した結果を可視化した方法を2種類提案する。

## 2. 語りかけエージェント

まず、実装した語りかけエージェントについて述べる。エージェントは、何種類かの特定の質問に答えるとともに、エージェントから質問をすることによりユーザである高齢者に語りかけ、高齢者の発言に対し、傾きや「はい」、「ええ」といった相槌を返す。ま

連絡先: 野中裕子, 成蹊大学大学院理工学研究科,  
dm126214@cc.seikei.ac.jp



図 1. 語りかけエージェントアニメーションデザイン

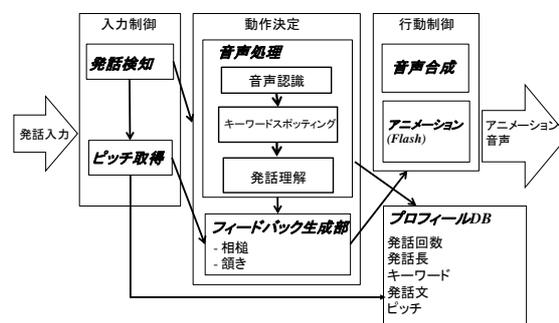


図 2. システム構成

た、「私は今日も元気です」「絵を描くのが趣味です」といった返答も適宜行う。エージェントの外観を図1に示す。また、病院スタッフが整形外科の入院患者に対して日常的に行う質問を事前に調査し、それに基づきエージェントの行う質問を選定した。

- 個人に関する基本的質問:「名前は何ですか?」「何歳ですか?」等
- 体調について:「足の具合はどうですか?」「リハビリはやっていますか」等
- 趣味について:「テレビを見ますか?」「本は読みますか?」等
- 過去の思い出について:「運動会はありましたか?」「人前でお遊戯をしましたか?」等
- 自宅に関して:「家の畑で何を育てていますか?」「庭はありますか?」等

## 2.1 システム構成

本システムは入力制御, 動作決定, 行動制御の3つのモジュールで構成されている。図2にシステム構成を示す。以下, 各モジュールについて説明する。

入力制御モジュールでは, 発話検知とピッチ取得を行う。発話検知では, マイク入力の音量よりユーザの発話の有無を検知する。判定のための閾値は個人ごとに設定をする。また, 発話が検知されるとピッチ取得を行う。

動作決定モジュールでは, 入力制御モジュールで取得した音声情報をもとに, エージェントの内部状態の更新と行動の決定を行う。動作決定モジュールはフィードバック生成部と音声処理部から構成されている。フィードバック生成部ではユーザのピッチ情報から話し方や相槌といったエージェントのフィードバックの生成を行い, 音声処理部では検知された発話音声から音声認識と言語理解を行う。

音声認識のために, 認知症患者の発話に特徴に現れる単語を辞書登録し, さらに認知症患者と病院スタッフとの会話の概要の書きおこしを言語モデルの学習に用いた。

言語理解には, ユーザからの特定の質問に答えられるように, 想定される質問-応答ペアを用意した。音声認識結果と想定質問との類似度を計算し, 最も類似度の高い想定質問を言語理解結果とした。想定質問の作成では, 整形外科病棟に勤める看護師に行ったアンケートを参考にした。アンケートは以下のような項目からなる。

- 患者からの訴えの内容
- 訴えの具体的な言葉や言い方
- 効果的な返答方法

アンケートの回答は約50個あったが, 類似した回答がいくつか見られたため, 以下の6つのカテゴリに分類し, 想定質問を作成した。

- トイレ関係
- 帰宅関係
- 食事関係
- 体調関係
- 家族関係
- その他

最後に, 行動制御モジュールは, 音声合成により作成されたエージェントの発話音声ファイルを再生すると同時にアニメーションを実行する[比企野 2011]。

## 3. 収集データと基礎的検討

コミュニケーションに対する反応性を定量化するため, 男性2名, 女性8名の計10名の認知症患者による語りかけエージェントとの会話を収録したビデオデータを使用した。会話の様子を図3に示す。



図3. 語りかけエージェントとの会話の様子

年齢平均は77.6歳, MMSEの平均は23.5である。MMSEとは認知機能テストにより与えられる評価指数である。最高が30であり, 21以下は認知障害がある可能性が高いことを示す。

会話エージェントとの会話の様子を収録したビデオを観察すると, エージェントの質問に対してリズムカルに反応している認知症患者と反応の少ない患者が見られた。リズムカルな反応をしている認知症患者は, エージェントの質問に対して素早く反応をしたり, 発話に対して話しなどの頭部動作が多くみられた。一方で, リズムカルな反応の少ない患者は, エージェントの質問に対して短い返答しか返さず, あまり動きがなくぼんやりしているような印象を受け, 積極的に対話をしようとしていない様子が見られた。そこで, このような点を定量化するために具体的には以下の分析指標を用いた。

- 話し
- 頭部移動量
- 発話間のポーズ
- 声の高さ(ピッチ)
- 発話量

本研究では, 音声データと頭部動作データからこれらの指標を算出し, コミュニケーションの反応性指標として用いる。

### 3.1 音声データ

会話エージェントとの会話中の発話音声から, ポーズ情報とピッチ情報を取得した。ポーズ時間が短い場合は, 会話エージェントの発話に対し, 素早く反応していることを表しており, 反対にポーズ時間が長ければ長いほどエージェントの発話にあまり反応していないと考えられる。また, 相手の発話に対する自分の意見などを話すときは, 独り言などを発するときなどに比べ, 声の高さは高くなると考えられる。そこで, 音声データとして発話間のポーズと声の高さを使用することにした。

#### (1) 発話間のポーズ

エージェントの発話終了後から患者の発話開始までの経過時間を調べた。最もポーズ時間が短い場合で0.48秒であり, 一方で最長のポーズ時間は2.01秒であった。また今回の10名の平均ポーズ時間1.32秒であった。

表1. ポーズの平均および標準偏差(秒)

|    | 最長   | 最短   | 平均   | 標準偏差 |
|----|------|------|------|------|
| 全体 | 0.48 | 2.01 | 1.32 | 0.49 |

#### (2) ピッチ

認知症患者による発話区間のピッチの平均を求めた。最も高いピッチは206.9Hzである。一方で, 最も低いピッチは142.6Hzであった。10名のピッチの平均値は181.4Hzであった。

表 2. ピッチの平均および標準偏差 (Hz)

|    | 最高    | 最低    | 平均    | 標準偏差 |
|----|-------|-------|-------|------|
| 男性 | 175.2 | 142.6 | 158.9 | 23.0 |
| 女性 | 206.9 | 163.9 | 187.0 | 17.2 |
| 全体 | 206.9 | 142.6 | 181.4 | 20.7 |

### 3.2 動作データ

認知症患者が会話エージェントと話している時の頭部の動作データの計測には顔認識ソフトウェア faceAPI を使用し、頭部の 3 次元位置座標と各軸における回転角度のデータを収集した。2 名の異なる認知症患者の同じ時間帯の頭部動作データの例を図 4 に示す。

二つのグラフを比べてみると、頭部が良く動いている人とあまり動いていない人がいることが分かる。このことから、会話エージェントと発話をしている際の頭部動作は人によって大きな違いがあるのではないかと考え、頭部動作の分析を行った。具体的には、傾きと傾き以外の頭部の動きの 2 種類についての検出を行った。

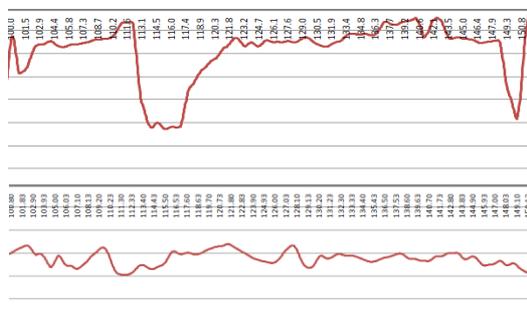


図 4. 頭部動作のグラフ

#### (1) 傾きの検出

傾きの検出には、頭部の上下位置座標(y 座標)の変化と左右方向への首振り(y 回転角度)のデータを使用した。まず、y 座標の変化を調べ、これをプロットしたグラフ中の 2 つの極値間の時間が 1 秒以内であり、かつそこでの頭部移動量が閾値以上であればこれを傾きとした。閾値は個人ごとに設定した。また、y 回転角度を使用し、正面を向いていない場合の頭部の上下動を傾きと認定しないようにした。エージェントの発話中、および終了後 3 秒以内に出現した傾きの回数をカウントした。ただし、現状のシステムでは、比較的大きい傾きは検出可能であるが、細かく連続的に行う傾きの検出はできていない。

#### (2) 頭部移動量

傾きの他にも、エージェントからの質問の直後に、返答を考えるために、視線をエージェントから逸らすなどの頭部の動きを行う人と特に変化のない人がいた。そこで、エージェント発話終了後 3 秒以内の頭部位置の移動量を計測した。これは、動きに変化がある人とそうでない人との頭部の動きを表す位置座標のデータを比べた際、エージェントの質問後 3 秒後までの位置座標データに変化がある人とそうでない人が見受けられた。よって、本研究ではエージェントの発話後 3 秒以内における頭部移動量の計測を行った。

計測は 30msec 毎に行い、以下の式を用いて過去 3 秒間の累積結果を頭部移動量とした。最も大きく動いている患者の頭部移動量は 0.39 であり、最も動きの少ない患者の頭部移動量は 0.007 であった。10 名の頭部移動量の平均は 0.23 であった。

faceAPI で得られたデータは 0 から 1 に正規化されているので、頭部移動量のデータに単位はない。

$$distance = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2 + (z_2 - z_1)^2}$$

表 3. 頭部移動量の平均と標準偏差

|    | 最大   | 最小    | 平均   | 標準偏差 |
|----|------|-------|------|------|
| 全体 | 0.39 | 0.007 | 0.23 | 0.09 |

表 4. 各データにおける反応性の結果

| 発話                      | 頭部移動量 | ピッチ | ポーズ | 傾き |
|-------------------------|-------|-----|-----|----|
| あなたの、好きな、食べ物は、<br>なんですか | 5     | 2   | 5   | 1  |
| 趣味は、何か、ありますか            | 5     | 1   | 5   | 1  |
| 普段、本は、読みますか             | 5     | 5   | 1   | 1  |
| どんな本が、好きですか。            | 1     | 4   | 4   | 0  |
| テレビ番組は、好きですか。           | 4     | 1   | 5   | 0  |
| どんな、テレビ番組を、見ますか         | 5     | 5   | 2   | 0  |

## 4. 反応性の定量化

3 節で得られたデータを使い、会話における反応性の定量化を行う。

### 4.1 評価方法

評価方法として、データを値の小さいものから順に 20 パーセントタイル区切りで等分し、各群に 1 から 5 の評価値を与えた。この評価値が定量化された会話における反応性である。評価の値が高いほど良い反応であることを表すものとした。頭部移動量・ピッチ・ポーズそれぞれに対してこの方法を使用し、5 段階評価を付与する。また、傾きに関してはこの方法は行わず、傾きありの場合は 1、傾きなしの場合は 0 と分類した。表 4 はある患者の反応性を示す。「発話」は会話エージェントの質問内容を示す。

表 4 より、傾きのある前半 3 つの質問は頭部移動量も多く、さらに最初の 2 つはポーズも高く(ポーズが短く)、エージェントの質問に素早く反応している。よって、この 6 つの質問の中では最初の 2 つがより反応の良い状態であると考えられる。また、最後の質問は頭部移動量が高い数値であるが傾きは行っていないことから、このときの頭部の動きは傾き以外の動きであったことになる。

この表から、このユーザは傾きがなくなった後半から、エージェントとの会話に飽きてきたのではないかと考えられる。また、最初の 2 つの質問では、返答するまでの時間が短いですがピッチがあまり高くない。それに対して、3 つ目の質問では、返答までの時間が長いですがピッチが高い。このようなパラメータ間の相関関係についても今後分析してゆく必要がある。

## 5. 可視化

会話への反応性の可視化方法として、4 種類の指標(頭部移動量・傾き・ピッチ・ポーズ)をすべてグラフ化して表示する「詳細表示方式」と、4 種類のデータを統合し、エージェントの各発話(質問)に対する反応性を表す顔マークを表示する「簡易表示方式」の 2 種類の方法を提案する。

### 5.1 詳細表示方式

詳細表示方式では、各発話に与えられた 4 種類の指標(頭部移動量・傾き・ピッチ・ポーズ)をすべてグラフ化する。図 3 に一例を示す。横軸がエージェントの発話内容、縦軸が反応性を表す。

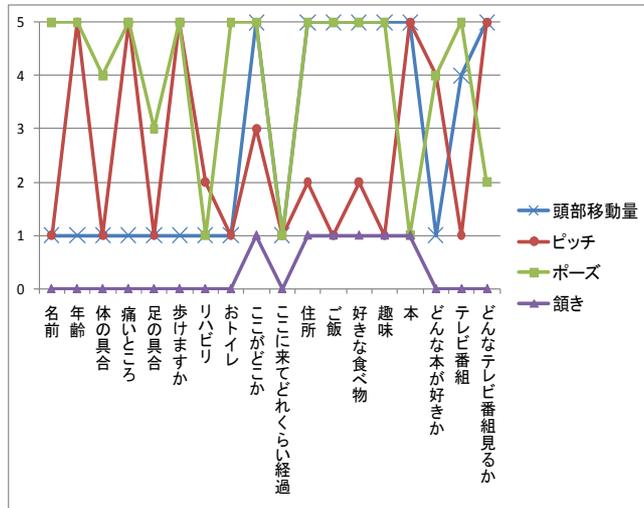


図 5. グラフを用いた詳細表示

このように、グラフで表すことで、他の発話との比較が可能になる。また頭部移動量と傾きに相関関係があることが分かる。どのデータの反応性が高いときに良い反応であるのか、4 種類のデータのうちいくつ以上において高い反応性が観察されると全体として良い反応と言えるのか等、より詳細に検討していく必要がある。また、「本は好きか?」という質問から「どんな本が好きか?」という質問になった時、頭部移動量が大きく低下しているが、ポーズが上昇している。このような相関関係についても、今後さらに詳しく分析する必要がある。

### 5.2 簡易表示方式

簡易表示方式では頭部移動量・ピッチ・ポーズ・傾きの 4 種類のデータを統合して評価し、1 つの顔マークに代表させる。そのために、各発話の反応性をこれら 4 種類のデータにより特徴づけられる 4 次元空間の点とみなす。これをデータマイニングツール Weka で実装されている k-means 法により 4 つにクラスタリングした。次に、クラスタごとで平均を求め、平均値が高いクラスタから順に高い評価をつけ、評価にあわせて図 6 のような顔マークを各クラスタに与えた。左が最も反応が良いことを表し、右にいくにつれて反応が良くない状態を表す。



図 6. 顔マーク

詳細表示で示したデータの一部を、顔マークにより表現した場合、図 7 のようになる。顔マークを使用することで、この患者がエージェントの発話に対して反応が良いかどうかを一目で判断することができる。この場合、「少し歩けますか?」という質問にはよく反応をしているが、それ以外の質問にはあまり反応していないことが一目で分かるようになる。簡易表示方式の場合、4

種類のデータがそれぞれどのような結果であったかは考える必要がないので、専門的な知識を持っていない患者の家族が患者の様子を知るには有効な可視化方法となると考えられる。

また、複数の質問に対する反応性の評価を統合し、トピックごとに反応性としてあらわす場合は、トピック毎に 1 から 4 の発話の評価値の平均値を求め、これに対してマークを新たに付与する。

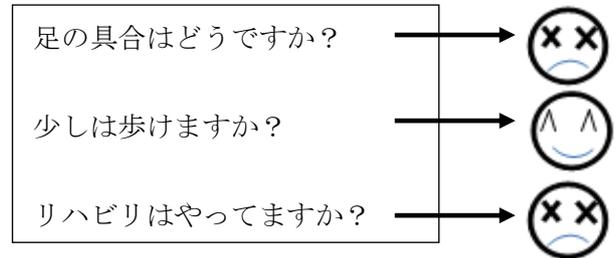


図 7. 顔マークによる簡易表示

## 6. おわりに

本研究では、高齢者の認知状態の把握および評価をするためのシステム実現を目指し、会話における反応性の定量化を行った。動作データから傾きと頭部移動量を、音声データから声の高さと発話間のポーズを算出し、それらのデータから会話における反応性を算出する方式を実装した。また、算出した反応性に対して詳細表示方式と簡易表示方式の 2 種類についての可視化を行った。これにより、エージェントとの対話における患者の反応、また質問内容による反応の違いを視覚的に表すことができる。

今後は反応性を継続的に調べ、時間経過や日による反応性の変化を分析してゆくことにより、本手法により認知機能の評価が可能であるかを検討してゆく。また、被験者数を増やしてより多くのデータを検証することにより、より精度の高いシステムを実装する必要がある。

### 参考文献

- [Pollack 2005] Martha E. Pollack: Intelligent Technology for an Aging Population: The Use of AI to Assist Elders with Cognitive Impairment, in *AI Magazine*, 2005.
- [酒井 2012] 酒井洋一, 野中裕子, 安田清, 中野有紀子: 認知症患者のコミュニケーション特性の評定における会話エージェントの利用: 第 74 回情報処理学会, 2012.
- [比企野 2011] 比企野純太, 中野有紀子: 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: 2011