

## 語りかけエージェントとの会話における韻律情報の分析

野中 裕子<sup>1</sup> 酒井 洋一<sup>1</sup> 安田 清<sup>2</sup> 林 佑樹<sup>3</sup> 中野 有紀子<sup>3</sup>成蹊大学大学院 理工学研究科 理工学専攻<sup>1</sup>千葉労災病院/京都工芸繊維大学<sup>2</sup>成蹊大学 理工学部 情報科学科<sup>3</sup>

## 1. はじめに

現在、少子高齢化により高齢者、特に記憶障害などが生じる認知症患者への支援の必要性が高まっている。そのため、このような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。

そこで本研究では、認知症高齢者の日々の健康状態・認知状態の把握を目指し、語りかけエージェントとの対話における高齢者のピッチ・ポーズ・発話長・抑揚といった音声情報の分析を行った。

## 2. 語りかけエージェント

本研究で開発した語りかけエージェントは、一問一答式で質問を行うことでユーザである高齢者に語りかけ、高齢者の発話の音声的特徴に応じて、頷きや相槌、さらに「私は今日も元気です」といったコメントを返す。ただし、音声認識や言語理解を行っているわけではない。

## 3. 対話における反応性の観察

高齢者の状態把握にむけて、まず高齢者と語りかけエージェントとの対話の様子を観察した。図1に実験の様子を示す。女性8名、男性2名の認知症高齢者(年齢平均:77.6, MMSEの平均:23.5)のビデオデータから得られた被験者の発話を、2名のアノテータによって、反応の良い発話(HR)とそうでない発話(LR)の2種類に分類してもらった。そして、発話間のポーズ・ピッチ・発話長・頷き時間に着目し、2名が一致してHRと判定した発話群と、LRと判定した発話群間で各情報の平均値に差があるかを統計的に検定した。

(a) 発話間のポーズ: エージェントの発話終了から被験者の発話開始までの時間

(b) 声の高さ(ピッチ): 被験者の発話区間のピッチの平均

(c) 発話長: エージェントが質問をしてから(話題が切り替わってから)、次の話題に移行するまでの被験者の発話総時間



図1. エージェントとの対話様子

(d) 頷き時間: エージェント発話中に被験者が頷いている時間

平均値およびt検定の結果を表1に示す。表1より、HRとLRの間で有意差が見られた。HRはポーズが短く、ピッチが高く、発話長が長く、頷きを長く行うことが分かった。

表1. HRおよびLRの平均, 有意差

情報	HRの平均	LRの平均	有意差
ポーズ(秒)	0.98	1.57	$t(95) = -2.82$ $p < 0.05$
ピッチ(Hz)	178.6	137.1	$t(105) = 2.99$ $p < 0.01$
発話長(秒)	4.77	1.60	$t(105) = 5.83$ $p < 0.01$
頷き時間(秒)	0.61	0.34	$t(105) = 1.86$ $p < 0.1$

## 4. クラスタリングによる発話の分類

さらに、健常者である女性9名の高齢者(平均年齢:77.7)についても同様の対話データ収集実験を行った。この実験における被験者を本稿では健常者と呼ぶ。

収集した健常者の音声データと、3章で収集した認知症高齢者の音声データを用いて分析を行った。ここでは、前の実験で有意差が見られた、(a)発話間のポーズ、(b)声の高さ、(c)発話長に加え、抑揚に着目して音声情報を分析した。

(e) 抑揚: 被験者の発話区間におけるピッチの最高値と最低値の差

データマイニングツール Weka で実装されている k-means 法を用いて、音声情報に基づき2つのクラスタに分類し、各クラスタにおける各データの平均を算出した。表2にその結果を示す。

Analysis of prosodic information in the interaction with the agent speaks  
Yuko Nonaka<sup>1</sup>, Yoichi Sakai<sup>1</sup>, Kiyoshi Yasuda<sup>2</sup>, Yuki Hayashi<sup>3</sup> Yukiko Nakano<sup>3</sup>

1: Graduate School of Science and Technology, Seikei University

2: Chiba Rosai Hospital/Kyoto Institute of Technology

3: Dept. of Computer and Information Science, Seikei University

表 2. クラスタリング結果

クラスタ	発話長 (秒)	ピッチ (Hz)	ポーズ (秒)	抑揚 (Hz)
a	2.53	172.4	1.18	93.3
b	5.32	192.6	0.79	235.7

表 2 より, クラスタ b に分類されたデータは, クラスタ a に分類されたデータと比べて, 発話長が長く, ピッチが高く, ポーズが短く, 抑揚が大きいことが示された. 表 1, 2 より, 自動でクラスタリングした結果と, 人手によって反応性をアノテーションした結果とが, 同様の傾向を持つことが示された. つまり, 反応性の評価にクラスタリングが有効であることが示された.

### 5. クラスタリングに基づくユーザの分類

クラスタ a を C\_LR, クラスタ b を C\_HR とし, 被験者の各発話を分類した. さらに, 各被験者について, 全発話における各クラスタの割合を算出した. そして, 以下に示す基準に基づき, 被験者を反応性高・低の 2 群に分類した.

- ・ C\_LR の割合が 0.6 以上: 反応性低群に分類
- ・ C\_LR の割合が 0.6 より下: 反応性高群に分類

結果を表 3, 4 に示す. 表は, 左から被験者番号, 全体に対する C\_LR の割合, 全体に対する C\_HR の割合を表している.

表 3, 4 より認知症高齢者の多くは反応性低群に分類され, 健常者の多くは反応性高群に分類されていることが示された.

### 6. おわりに

本研究では, 健常者および認知症高齢者の語りかけエージェントとの対話における音声データの分析を行った. そして, 算出した 4 つの音声情報を用いてクラスタリングを行うことにより, 発話長が長く, ピッチが高く, ポーズが短く, 抑揚が大きいクラスタと, そうでないクラスタに分類されることが明らかとなった. そして前者のクラスタには健常者が多く, 後者のクラスタには認知症患者が多く分類されることが示された.

本稿ではエージェントの一発話における被験者の音声情報を分析したが, 今後は, 対話のトピックに注目し, 話題単位で反応性を算出する方法を検討する. これにより, 各被験者について一連の対話の中のどのようなトピックに興味があるのかを推定できる.

表 3. 認知症高齢者の結果

	C_LR の割合	C_HR の割合
1(反応性低)	0.87	0.13
2(反応性低)	0.71	0.29
3(反応性高)	0.5	0.5
4(反応性低)	0.7	0.3
5(反応性低)	0.78	0.22
6(反応性低)	0.95	0.05
7(反応性低)	0.7	0.3
8(反応性低)	0.85	0.15
9(反応性高)	0.3	0.7
10(反応性低)	0.67	0.33

表 4. 健常者の結果

	C_LR の割合	C_HR の割合
1 反応性高)	0.28	0.72
2 反応性高)	0.37	0.63
3(反応性高)	0.5	0.5
4(反応性高)	0.15	0.85
5(反応性低)	0.75	0.25
6(反応性高)	0.29	0.71
7(反応性高)	0.3	0.7
8(反応性高)	0.45	0.55
9(反応性高)	0.35	0.65

さらに, 非言語情報に加えて言語情報を高齢者の状態把握に取り入れることで, 反応性が良くなかった状況の発話内容を把握し, 得られた結果を医者や家族のためのレポートとして出力するシステムを検討していく予定である.

### 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*, Vol.26, No.2, pp.9-25 (2005)
2. 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」, 第 73 回情報処理学会講演論文集, 4ZA-7 (2011)
3. 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 回情報処理学会講演論文集, 3ZC-8 (2012)
4. Yuko N., Yoichi S., Kiyoshi Y., and Yukiko N.: "Towards Assessing the Communication Responsiveness of People with Dementia," In *Proc. of 12th International Conference on Intelligent Virtual Agents*, pp.496-498 (2012)