

人型ロボットとの会話における軽度認知症者の発話特徴分析
Speech Characteristics Analysis of Mild Cognitive Impairment in Conversation
with a Humanoid Robot

吉井 謙太[†] 阿部 元樹[†] 木村 大毅[‡] 小杉 晋央[‡] 新川 香[‡] 高瀬 俊郎[‡]
Kenta Yoshii Motoki Abe Daiki Kimura Akihiro Kosugi Kaoru Shinkawa Toshiro Takase
小林 正朋[‡] 山田 康智[‡] 根本 みゆき[§] 渡辺 亮平[§] 塚田 恵鯉子[§]
Masatomo Kobayashi Yasunori Yamada Miyuki Nemoto Ryohei Watanabe Eriko Tsukada
太田 深秀[§] 東 晋二^{||} 根本 清貴[§] 新井 哲明[§] 西村 雅史[†]
Miho Ota Shinji Higashi Kiyotaka Nemoto Tetsuaki Arai Masafumi Nishimura

1. はじめに

超高齢社会の日本において、認知症は深刻な問題となっている。認知症の早期発見を目的として、高齢者の認知能力を発話から見出すとする研究がこれまで数多く行われてきた[1][2]。一方、介護現場ではロボットを用いた見守りが検討されている[3]。ロボットと高齢者の日常会話から認知能力低下の兆しを見つけることができれば、認知症の早期発見につながれる可能性がある。

本研究では事前に設定したシナリオに基づいて、人型ロボットと実験参加者（健常者、軽度認知障害患者 以下「MCI 患者」）との対話を収録し、健常者と MCI 患者との発話特徴の差異について分析を行なったので報告する。

2. 対話データ収録

2.1 収録方法

ソフトバンクロボティクス社の人型ロボット「Pepper」を用いて、日常会話を模擬した対話データの収録を行った。ロボットとの対話データ収録の様子を図 1 に示す。ロボットの発話に対して実験参加者が応答する形で音声収録を実施した。なお、収録時にはネックバンド型の咽喉マイクと集音マイクを首に装着(図 2)してもらっている[4]。

実験は Wizard of Oz の形式で実施した。オペレータは実験参加者から見えない位置に座り、事前に用意した対話シナリオを基本としつつ、実験参加者の応答に応じて話題の遷移や発話内容の選択を適宜行った。

2.2 対話シナリオ

対話シナリオは質問応答と雑談で構成されており、質問応答ではロボットが実験参加者に対して質問を投げかけ(例「昨日は何を食べましたか?」)、それに対して実験参加者が答える一問一答形式とした。一方、雑談では日常会話に近い対話を模擬するため、実験参加者の応答に適した返事(例「よく眠ると一日幸せな気分になりますよね」、「よく眠れないと一日辛いですよね」)や自己開示(例「私はリンゴが大好きなんです」)を行った。実験参加者がロボットに対して質問をしてきた場合(例「リンゴを食べれるの?」)は、「その質問にはお答えできません」という返事や次の話題に移るなどして、はぐらかすこととした。

[†]静岡大学大学院 総合科学技術研究科 Shizuoka University

[‡]日本アイ・ビー・エム株式会社 IBM Japan, Ltd.

[§]筑波大学 University of Tsukuba

^{||}東京医科大学茨城医療センター Tokyo Medical University
Ibaraki Medical Center

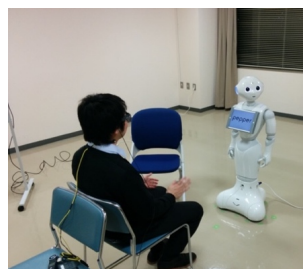


図 1 収録の様子

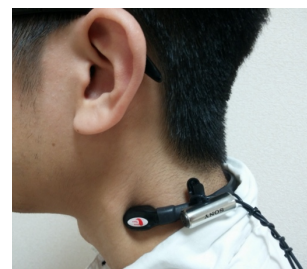


図 2 収録用マイクワホン

3. 対話データと特徴量

実験参加者のうち、認知機能検査や血液検査など様々な検査の結果から医師が軽度認知障害 (MCI) と診断した 45 名(平均 70.2 歳)と健常と診断した 45 名(平均 74.3 歳)、合計 90 名の対話データを対象に発話特徴分析を行った。

発話特徴分析に利用した特徴量を表 1 に示す。まず咽喉マイクで収録された対話データから発話区間に関する 6 種類の特徴量を抽出した。なお、咽喉マイクを利用することにより、ロボットの発話や動作音などの雑音に影響されることなく、発話区間を正確に検出することができる。一方、咽喉マイクは装着位置による観測値のばらつきが大きく、音響特徴量の抽出には適していない。このため、音響特徴量は咽喉マイクで得られた発話区間を参照し、集音マイク側データの当該部分から推定することとした。今回用いた音響特徴量は Shimmer と Jitter の 2 種類である。

表 1 特徴量

特徴量名	概要
発話時間長 [秒]	1 ターン中の実験参加者の発話開始点から終了点までの時間長
反応時間 [秒]	人型ロボットの発話終了点から実験参加者の発話開始点までの時間
無音時間長 [秒]	発話時間長における無音時間(300ms 以上)の合算値
無音時間割合 [割合]	発話時間長における無音時間長が占める割合
フィラー長[秒]	1 ターン中の「えー」や「あー」など、間を埋める発話時間の合算値
フィラー割合 [割合]	発話時間長におけるフィラーが占める割合
Shimmer (local)	声の大きさの揺れ
Jitter (local)	声の高さの揺れ

4. 発話特徴分析

4.1 統計的有意差検定

表1に示した特徴量において、健常者とMCI患者間で統計的有意差検定を行なった。割合に対してはZ検定を用い、それ以外にはWelchのt検定を用いた。特徴量分析の結果を表2に示す。

4.2 結果・考察

発話区間に関する特徴量では、発話時間長と反応時間で極めて有意な差が得られた。

発話時間長では、健常者に比べMCI患者の発話時間が有意に短く、発話量が減っていると考えられる。ロボットからの問いかけの違いにより発話時間長に差が出るのかを検証するため、雑談・質問応答に分けて統計的有意差検定を行った。表3にその結果を示す。雑談・質問応答ともにMCI患者で有意に発話時間長が短いことが確認された。平均値を比較すると、雑談においてMCI患者は健常者に比べ平均0.36秒短く、質問応答では1.24秒短いという結果が得られた。対話データを聞いたところ、MCI患者はロボットの質問に対して「はい」など一言で終わるケースが度々確認された。思い出しや手順の説明など雑談に比べ複雑な返答を求められる質問応答では、認知機能の低下によって返答が困難になり、発話時間長が短くなったのではないかと考えられる。

反応時間では、MCI患者において遅くなる傾向が得られ、返答を思考する時間が健常者に比べ長くなっていると考えられる。発話時間長と同じく問いかけの違いにより反応時間に変化があるか分析したところ、雑談では有意な差が得られなかった。一方、質問応答では極めて有意な差が得られた(表3)。これは、質問応答においてMCI患者の反応時間が特に遅くなっていることを意味する。MCI患者はロボットからの質問に対し、返答を思考するのに時間がかかってしまい、反応時間が遅くなると考えられる。

以上の結果から、発話時間長や質問応答における反応時間が認知機能低下の兆しを見出すのに有効な特徴量となる可能性があると言える。

反応時間にはロボットの発話中に実験参加者が話し始め

表2 特徴量分析

特徴量名	健常者		MCI患者		p値
	平均	標準偏差	平均	標準偏差	
発話時間長	4.62	7.41	3.86	5.16	0.008 **
反応時間	0.90	1.02	1.13	1.30	<0.001 ***
無音区間長	1.30	2.47	1.15	2.16	0.138
無音時間割合	0.03	0.10	0.02	0.09	0.059
フィルター長	0.21	0.27	0.20	0.27	0.216
フィルター割合	0.20	0.61	0.16	0.56	0.189
Shimmer (local)	0.17	0.04	0.17	0.05	0.115
Jitter (local)	0.07	0.03	0.08	0.05	0.045 *

・ p < 0.1 * p < 0.05 ** p < 0.01 *** p < 0.001

る負の反応時間が含まれている。全ターンにおける負の反応時間割合にZ検定を行ったところ健常者とMCI患者間で有意な差は認められなかった(p > 0.1)。阿部ら[4]は認知機能検査であるMMSEのスコアが23点未満(認知症疑い)の実験参加者で負の反応時間を観測する傾向があったと報告しており、割合の比較だけでなく認知機能検査のスコア等との相関を分析する必要があると考えられる。

音響特徴量では、Jitter(local)において有意な差が得られた。Hallら[5]もShimmer(local, APQ3)やJitter(local)で有意な差が得られたと報告しており、音声の揺らぎの特徴量も認知機能低下を見出すのに有効な特徴量になり得ると考えられる。一方、原ら[6]は、ShimmerやJitterなど音声の揺らぎの値の増加は身体的QOLの低下と有意な関連性を示したと報告しており、認知機能の低下よりも身体的な衰えが影響している可能性がある。今後、年齢や認知機能検査スコアとの相関等の分析を進めていく必要があると考えている。

5. おわりに

本研究では人型ロボットとの対話による軽度認知障害スクリーニングの可能性に関する検証を行った。その結果、複数の特徴量で統計的に有意な差が得られ、ロボットとの対話から認知機能低下の兆しを見出す可能性を示すことができた。今後は人やタブレットを用いた対話と比較や認知機能検査スコアとの相関分析を行い、ロボット対話を用いた場合との比較検証を行う予定である。

謝辞

本研究はJSPS科研費JP19H01084の助成を受けたものである。

参考文献

- [1] 加藤 他, “高齢者音声韻律特徴を用いたHDS-Rスコアとの相関分析”, 人工知能学会論文誌, Vol.26, No.2, pp.347-352(2011).
- [2] Roark et al., “Spoken language derived measures for detecting mild cognitive impairment”, IEEE transactions on audio, speech, and language processing, Vol.19 No.7, pp.2081-2090(2011).
- [3] 二宮 恒樹, “コミュニケーションロボット「PALRO (パルロ)」の紹介とさがみロボット産業特区における取り組み”, 日本ロボット学会誌, Vol.33, No.8, pp.607-610(2015).
- [4] 阿部 他, “人型ロボットとの音声対話に見られる認知症者の発話特徴分析” 情報処理学会第81回全国大会, Vol.4, No.06(2019).
- [5] Hall et al., “Using Tablet-Based Assessment to Characterize Speech for Individuals with Dementia and Mild Cognitive Impairment: Preliminary Results.”, AMIA Summits on Translational Science Proceedings 2019, Vol. 2019 34-43(2019).
- [6] 原 他, “介護施設入所高齢者の健康関連QOLと音響学的分析手法を用いた音声機能との関連性.”, 日本老年医学会雑誌, Vol.52 No.4 pp. 391-398(2015).

表3 雑談・質問応答ごとの特徴量分析

特徴量名	健常者		MCI患者		p値
	平均	標準偏差	平均	標準偏差	
発話時間長 (雑談)	2.62	6.01	2.26	2.43	0.037 *
発話時間長 (質問応答)	6.52	9.57	5.28	6.37	0.013 *
反応時間 (雑談)	0.79	1.07	0.88	1.14	0.195
反応時間 (質問応答)	1.01	0.95	1.34	1.40	<0.001 ***

・ p < 0.1 * p < 0.05 ** p < 0.01 *** p < 0.001