

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

阿部元樹[†] 梅井良太[‡] 木村大毅[¶] 小杉晋央[¶] 新川香[¶] 高瀬俊郎[#]
 山田康智[¶] 根本みゆき[§] 渡辺亮平[§] 塚田恵鯉子[§] 太田深秀[§]
 東晋二[§] 根本清貴[§] 新井哲明[§] 西村雅史[†]

静岡大学[†] 無所属[‡] 日本アイ・ビー・エム (株) 東京基礎研究所[¶]
 日本アイ・ビー・エム (株) グローバル・ビジネス・サービス[#] 筑波大学[§]

1. はじめに

認知症者の簡易スクリーニングを目的として、会話音声から認知症者の特徴を見出そうとする研究が数多く報告されている[1][2]。これらの研究は心理検査時の音声データを対象としており、日常会話を対象としたものではない。今後介護現場での普及が期待される人型ロボットとの日常的な対話から認知症に関連した兆候を見出すことができれば、高齢者に負担をかけることなく早期の段階の認知症を発見できる可能性がある。

今回、認知症者を含む高齢者と人型ロボットとの日常会話を模擬した音声対話データを収集し、その特徴を明らかにするとともに、心理検査時の音声データにみられる特徴との比較を行なったので報告する。

2. 対話データ収集セットアップ

ここでは、高齢者と心理士が心理検査時に行う対話と、人型ロボットとの対話をそれぞれ収集した。なお、ロボットとの対話データの収録方法は文献[3]と同じであるが、今回被験者数を大幅に増やして実験を行なっている。

最初に、専門の心理士が MMSE(Mini Mental State Examination)[5]と呼ばれる認知機能検査を実施する。この際、対象となる高齢者と心理士にはそれぞれ咽喉マイクと集音マイクの装着を依頼し、検査中の対話データを話者ごとに2チャンネル収録した[3][4]。ここで咽喉マイクを利用するのは互いの発話タイミングを相手の発話や雑音に影響されることなく正確に検出するためである。

次に、人型ロボット「Pepper」を用いて、高齢者との日常会話を模擬した音声対話データを収集する。人型ロボットの対話制御は自動ではなく、事前に準備したスクリプトをオペレータが逐次的に選択する Wizard of Oz の形式とした。なお、対話は質問応答セクションと雑談セクションの2つから構成されている。質問応答セクションでは記憶能力、論理的思考能力に関する質問を行い(例「昨日は何を食べましたか?」「あなたが健康を保つためにしていることは何かありますか?」)、雑談セクション

では、人型ロボットが自己開示など非質問発話を行うこととした(例「私は健康のために毎日体操をしているんですよ」「私はりんごが大好きなんです」)。この際にも高齢者には心理士との対話の場合と同様のセットアップを用いた。

3. 収録データと分析対象特徴量

上記のセットアップに基づいて 61-87 歳の実験参加者 34 名(うち女性 22 名)から MMSE の時間・場所の見当識において合計 57 分、人型ロボットとの音声対話において合計 211 分、総計 268 分の音声データを収録した。MMSE の中から時間・場所の見当識を用いた理由はロボット質問応答がこれら見当識と同様に過去・現在の自身の記憶に関する質問を主としているためである。

今回、特に発話長や発話のタイミングに着目することとし、咽喉マイクで収録された音声データに対して、表 1 に示す 4 種類の特徴量抽出を行なった。これらの特徴量はいずれも先行研究[1][2]で有効性が示されたものである。発話区間検出については咽喉マイクを利用することで集音マイクよりも正確な分析が可能となる[4]。一方、咽喉マイクはスペクトル上の歪みや装着位置による観測値のばらつきが大きく、発話区間検出以外の特徴量分析には適していないため、今回の検討ではスペクトルに関する特徴量を分析対象に含めていない。

表 1: 分析対象とした音響特徴量

音響特徴量	概要
発話長 [秒]	1 ターン中の最初の発話の開始時点から最後の発話の終了時点までの時間長
反応時間 [秒]	心理士・人型ロボット側の発話終了時から実験参加者発話開始時までの長さ
無音区間長 [秒]	発話間のポーズなど、実験参加者の発話全体の中での無音区間の合算値
無音割合 [割合]	発話全体における無音区間長が占める割合

Identifying Dementia Characteristics using Spoken Dialogue with a Humanoid Robot

Motoki Abe[†], Ryota Togai[‡], Daiki Kimura[¶], Akihiro Kosugi[¶], Kaoru Shinkawa[¶], Toshiro Takase[¶], Yasunori Yamada[¶], Miyuki Nemoto[§], Ryohei Watanabe[§], Eriko Tsukada[§], Miho Ota[§], Shinji Higashi[§], Kiyotaka Nemoto[§], Tetsuaki Arai[§], Masafumi Nishimura[†]

[†] Shizuoka University

[‡] Freelance

[¶] IBM Research – Tokyo

[#] Global Business Service, IBM Japan

[§] University of Tsukuba

4. 分析結果

今回実施した MMSE においてスコアが 30 点満点中 23 点以下の場合、認知機能低下の可能性があると示唆されている[5]。本研究では、MMSE スコアと MMSE の見当識、質問応答、雑談における各音響特徴量との相関関係を分析した。結果を図 1、表 2 に示す。

図 1、表 2 より、MMSE では 4 つの音響特徴量全てで負の相関が見られた。これは先行研究[1][2]で報告されている傾向と同様のものを示している。次に、質問応答では発話長、反応時間では相関は見られなかったものの、無音区間長と無音割合では負の相関がやや見られた。最後に、雑談では MMSE や質問応答と比較して無音区間長と無音割合の相関が弱いことが分かる。また、反応時間において正の相関が見られ、特に MMSE スコアが 23 点未満の実験参加者においては負の反応時間を観測する傾向にあった。これは人型ロボットの発話が終了する前に実験参加者が発話を開始する発話重畳が起きていることを示している。なお、この発話重畳はロボット発話間のショートポーズで多く発生していた。

以上より、MMSE と質問応答を比較した際、特に発話中のポーズに関する特徴については同様の傾向を示しており、人型ロボットによる質問対話においても先行研究の知見を活用できる可能性がある。また、雑談における無音区間長、無音割合及び反応時間の特徴は返答に関する思慮時間が少ない雑談特有の現象であると考えられる。

5. おわりに

本研究では人型ロボットと高齢者との音声対話データを収集し、心理検査時の対話音声データとの比較及び MMSE スコアと相関分析を行った。結果、ロボット対話の質問応答では無音区間長と無音割合で、雑談については、反応速度において MMSE スコアとの相関関係を確認出来た。また、従来の心理検査を対象とした分析と比較した時、ロボット対話における質問応答では同様の傾向が見られた一方、雑談では異なる傾向が確認された。

今後はスペクトルや基本周波数に関連した通常の音響分析についても集音マイク収録音を利用して実施する予定である。

参考文献

- [1] Beltrami et al., "Automatic Identification of Mild Cognitive Impairment through the Analysis of Italian Spontaneous Speech Productions", *In LREC*, pp. 2086-2093, 2016.
- [2] Khodabakhsh et al., "Evaluation of linguistic and prosodic features for detection of Alzheimer's disease in Turkish conversational speech", *EURASIP Journal on Audio, Speech, and Music Processing*, vol.2015, No.1, pp.1-15, 2015.
- [3] 阿部 他, "人型ロボットとの雑談対話における認知症者の特徴に関する予備的検討", 平成 30 年度電気・電子・情報関係学会東海支部連合大会講演論文集, L3-1, 2018.
- [4] 西村 他, "生体音と環境音の同時収録による高齢者の行動および身体状態認識に関する検討", 日本音響学会 2015 年春季研究発表会講演論文集, 2-4-9, pp.1309-1310, 2015.
- [5] 日本神経学会 認知症ガイドライン https://www.neurology-jp.org/guidelinem/deg1/sinkei_deg1_c_2012_03.pdf (取得日時: 2018/06/27)

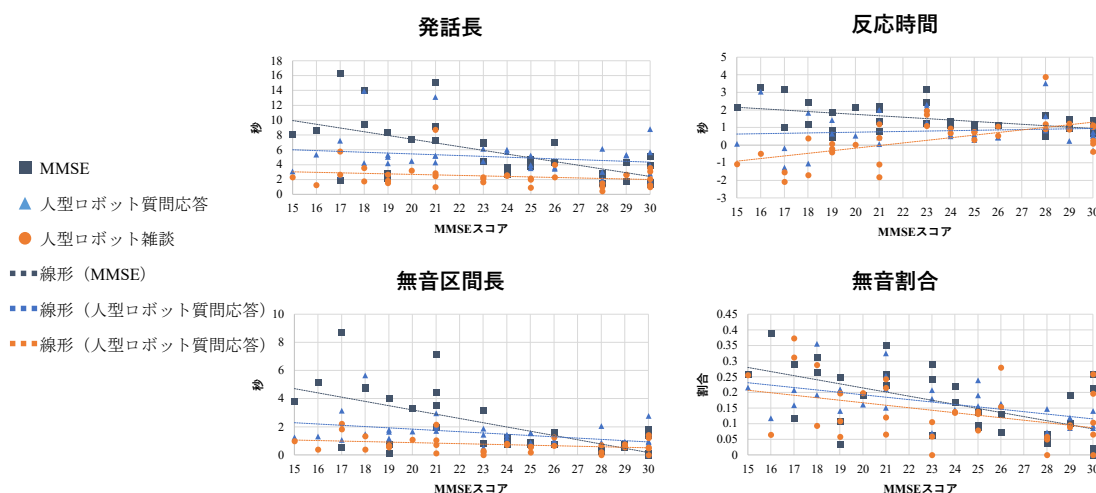


図 1 : MMSE スコアと各音響特徴量との散布図及び線形近似

表 2 : 各音響特徴量と MMSE スコアとの相関係数 (太字は係数が 0.4 以上/-0.4 以下のものを示す)

音響特徴量	MMSE	ロボット質問応答	ロボット雑談
発話長	-0.61	-0.20	-0.21
反応時間	-0.49	0.10	0.58
無音区間長	-0.65	-0.42	-0.30
無音割合	-0.58	-0.49	-0.39