

# 人型ロボットとの日常会話音声を用いた 認知症簡易スクリーニングのための基礎的検討

吉井謙太<sup>1</sup> 木村大毅<sup>2</sup> 小杉晋央<sup>2</sup> 新川香<sup>2</sup> 高瀬 俊郎<sup>2</sup> 小林 正朋<sup>2</sup>  
山田康智<sup>2</sup> 根本 みゆき<sup>3</sup> 渡辺 亮平<sup>3</sup> 塚田 恵鯉子<sup>3</sup> 太田 深秀<sup>3</sup>  
東 晋二<sup>4</sup> 根本 清貴<sup>3</sup> 新井 哲明<sup>3</sup> 西村雅史<sup>1</sup>

**概要:** 高齢者の認知機能低下を認知機能検査時の音声に含まれる特徴から見出すとする研究がこれまで数多く行われてきた。一方、介護現場ではロボットを用いた見守りや介護作業の省力化が期待されている。ロボットとの日常会話から認知機能低下の兆しを見出すことができれば、特別な検査を行うことなく認知症の早期発見に繋げられる可能性がある。本研究では、健常者群と軽度認知障害患者群に対して心理士が認知機能検査を実施した際の音声を収録した。また、同じ被験者群に対し事前設定したシナリオに基づき Wizard of Oz 法で対話制御を行った人型ロボットとの日常会話音声も収録した。認知機能検査の音声を分析した結果、無音時間長やフィラー長といった発話区間に関する特徴量において健常者群と軽度認知障害患者群間で有意差を得た。人型ロボットとの日常会話においては、認知機能検査と傾向や有意差のある特徴量が異なるものの、有意に異なる特徴量が得られた。また、人型ロボットとの日常会話収録後のアンケート結果から、人型ロボットとの日常会話を受け入れられることが示唆された。以上より、人型ロボットとの日常会話音声を用いた認知症簡易スクリーニングの可能性が示された。

## 1. はじめに

超高齢社会の日本において、認知症は深刻な社会問題となっている。認知症は初期段階で治療を開始することにより回復もしくは症状の進行抑制が可能と知られている。このため、認知症は早期発見が重要であり、特に認知症の前段階である軽度認知障害(Mild Cognitive Impairment)の発見が大変重要[1]であると言われている。認知症早期発見の試みとして、会話音声から認知機能低下の兆しを見出すとする研究が数多く報告されている[2][3]。これらの多くは、医師や心理士が実施した認知機能検査中の会話音声を対象としたものである。一方、介護現場ではロボットを用いた見守りや介護作業の省力化が期待されている[4]。ロボットと高齢者の日常会話から認知機能低下の兆しを見出すことができれば、特別な検査を行うことなく認知症の早期発見に繋げられる可能性がある。さらに、高齢者に日常会話を楽しんでもらいながら日常的に認知症簡易スクリーニングを実施できる可能性がある。

本研究では、実験参加者(健常者、軽度認知障害患者以下「MCI患者」)に対して専門の心理士が実施した認知機能検査時の会話音声と、日常会話を模擬したシナリオに基づいた人型ロボットとの会話音声を収録し、健常者とMCI患者の発話特徴の差異を分析した。さらに、従来の認知機能検査音声と人型ロボットとの日常会話音声で得られた発話特徴を比較し、人型ロボットとの日常会話を用いた認知症簡易スクリーニングの可能性を検討した。

## 2. 会話データ収録

### 2.1 認知機能検査

国際的に広く利用されている MMSE(Mini Mental State Examination)と呼ばれる認知機能検査を専門の心理士が実施した際の会話データを収録した。MMSE は見当識、記憶力を口頭で問う質問や、図形模写といった筆記を含む 11 問で構成された検査である。今回は、先行研究[2]において用いられていた時間の見当識と、場所の見当識を分析対象とした。時間、場所の見当識ともに口頭で質問し 5 点満点で評価する。時間の見当識は、検査実施時の年月日、曜日と季節を問う、場所の見当識は検査実施場所の都道府県名、市、地域の名前、建物の名前(または種類)と階数を問う質問である。

### 2.2 人型ロボットとの日常会話

ソフトバンクロボティクス社の人型ロボット「Pepper」を用い、人型ロボットとの日常会話音声の収録を行った。なお、収録方法は文献[5]と同様であるが、認知機能検査を受けた実験参加者と対応するように分析対象の実験参加者を変更した。ロボットとの会話データ収録の様子を図1に示す。実験は、Wizard of Ozの形式で実施した。オペレーターは実験参加者から見えない位置に座り、事前に用意した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。

シナリオは質問応答パートと雑談パートで構成されて



図1 人型ロボットとの  
会話音声収録の様子



図2 咽喉マイクと集音マイク

1 静岡大学大学院 総合科学技術研究所  
Shizuoka University  
2 IBM 東京基礎研究所  
IBM Research Tokyo  
3 筑波大学  
University of Tsukuba  
4 東京医科大学茨城医療センター  
Tokyo Medical University Ibaraki Medical Center

おり、質問応答パートではロボットが実験参加者に対して質問を投げかけ、それに対して実験参加者が答える1問1答形式とした。質問応答パートには、昨晚の食事内容や作り方など、記憶や論理的思考力を確かめる質問を織り交ぜた。一方、雑談パートでは、より日常会話を模擬するため実験参加者の応答に適した返事や自己開示を行った。なお、実験参加者がロボットに対して質問をしてきた場合は、「その質問にはお答えできません」といった返事や次の話題に移り、はぐらかすこととした。

### 2.3 収録機材

認知機能検査の際には心理士、実験参加者両者に、人型ロボットとの日常会話では実験参加者にネックバンド型の咽喉マイクと集音マイクを首に同時装着(図2)してもらい会話データの収録を行った (Linear PCM 形式/44.1k/ステレオ)。集音マイクは対象話者の口元から遠く離れた位置に居る人の発話や環境音も収録してしまう。このため発話重畳や大きな外部環境音発生した場合に正確な対象話者の発話区間推定が困難となる。そこで、本研究では上記の問題点を考慮し、外部環境音に頑健な咽喉マイクを併せて使用した[6]。咽喉マイクを用いることで、対象話者の発話タイミングおよび発話区間などの正確な検出が可能となり、より正確な分析が可能となる。

### 2.4 実験参加者

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健常と診断した45名、軽度認知障害(MCI)と診断した45名、合計90名を対象に認知機能検査音声と人型ロボットとの日常会話音声の分析を行った。本研究では早期発見への貢献を目標としているため、認知症患者ではなく認知症の前段階であるMCI患者に絞り分析を行った。分析対象とした実験参加者の詳細を表1に示す。

## 3. 分析特徴量

健常者とMCI患者の発話特徴の差異を分析するため、実験参加者の会話データから特徴量を抽出し、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2に示す。

まず咽喉マイクで収録された会話データを用いて発話区間とフィルターを人手でラベル付した。このラベルを用いて発話時間長からフィルター割合までの6種類の発話区間に関する特徴量を抽出した。咽喉マイクを用いることによって、発話重畳が発生した際に集音マイクでは正確な抽出が難しい反応時間といった特徴量の正しい抽出が可能となった。

次に、集音マイクで収録された会話データから2種類の音響特徴量抽出を行った。咽喉マイクは環境雑音に対して頑健であるが皮膚伝道音であるため通常の音響マイクとは特性が大きく異なる。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Praatというツールキットを用いて推定することとした。

表2 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]        | 心理士または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                 |
| フィルター長 [秒]      | 発話時間長における「えー」など、間を埋める発話時間の合算値        |
| フィルター割合 [割合]    | 発話時間長におけるフィルター長が占める割合                |
| Jitter (local)  | 声の高さの揺れ                              |
| Shimmer (local) | 声の大きさの揺れ                             |

## 4. 発話特徴分析

### 4.1 統計的有意差検定

表2に示した特徴量において、健常者とMCI患者間の統計的有意差検定を実施した。検定にはWelchのt検定を用いた。認知機能検査音声から特徴量を抽出し分析した結果を表3に、人型ロボットとの日常会話音声から特徴量を抽出し分析した結果を表4に示す。本研究では、p値が0.05以下の場合に有意差ありとし、同時に識別力を評価するために効果量[7]を用いる。今回は、t検定における代表的な効果量であるCohen's dを用いた。Cohen's dは以下の式で求められる。

$$d = \frac{|\text{実験群の平均} - \text{統制群の平均}|}{\sqrt{\frac{\text{実験群の標準偏差}^2 + \text{統制群の標準偏差}^2}{2}}}$$

Cohen's dにより、p値で有意差ありとなった2群間の平均値がどの程度離れているのか評価可能となる。効果量の目安としては、文献[7]に示されていた0.20(Small)、0.50(Medium)、0.80(Large)を用いた。

表1 実験参加者詳細

|       | 性別 [人] |    |    | 年齢 [歳] |      |       | 時間の見当識得点 [点] |      |     | 場所の見当識得点 [点] |      |     |
|-------|--------|----|----|--------|------|-------|--------------|------|-----|--------------|------|-----|
|       | 男性     | 女性 | 合計 | 平均     | 標準偏差 | 範囲    | 平均           | 標準偏差 | 範囲  | 平均           | 標準偏差 | 範囲  |
| 健常者   | 20     | 25 | 45 | 71.0   | 4.6  | 61-80 | 4.6          | 0.6  | 3-5 | 4.7          | 0.5  | 3-5 |
| MCI患者 | 26     | 19 | 45 | 74.6   | 4.9  | 64-87 | 4.5          | 0.7  | 2-5 | 4.6          | 0.5  | 4-5 |

## 4.2 分析結果

認知機能検査音声进行分析した結果、調べた特徴量 8 個中 5 個で有意差を得た。フィルラー割合を除いた発話区間に関する特徴量で有意差が得られたが、音響特徴量においては有意差が得られなかった。効果量では、無音時間と無音時間割合、フィルラー長において目安である 0.2(Small)を上回っており健常者、MCI 患者間に差があると示された。

人型ロボットとの日常会話音声においては、8 個中 6 個の特徴量で有意差が得られ、認知機能検査音声より多くの特徴量で有意差を得た。発話区間に関する特徴量では、無音時間割合とフィルラー割合以外で有意差を得た。音響特徴量では Jitter(local), Shimmer(local)共に有意差が得られた。効果量は、反応時間、Jitter(local)で目安である 0.2(Small)を上回っており、健常者、MCI 患者間で差があると示された。

## 4.3 考察

### 4.3.1 認知機能検査音声における発話特徴

本研究では認知機能検査 MMSE の時間と場所の見当識を分析対象とした。これらは認知症、特にアルツハイマー病の初期に失われると知られている[8]。時間と場所の見当識の得点(表 1)において Welch の t 検定を実施したところ、健常者と MCI 患者間で有意差は得られず、得点のみで認知機能低下の兆しを見出すのは困難である。しかし、認知機能検査音声から抽出した特徴量では、無音時間長と無音時間割合、フィルラー長において健常者、MCI 患者間で有意差が得られ、効果量の目安を上回った。得点に現れない認知機能低下の兆しを認知機能検査音声から見出すことが可能と示された。無音時間長やフィルラー長において有意差を得られたのは、認知機能低下により MCI 患者は時間や場所を思考するのに時間を要し、無音時間や間を繋ぐためのフィルラーが増加したのであろう。

認知機能検査音声の結果において興味深いのは、p 値では有意差ありと示されたが効果量で目安を下回った発話時間長と反応時間である。健常者と MCI 患者の、発話時間長平均値を比較すると健常者は MCI 患者に比べ 0.89 秒短い。見当識がしっかりしている健常者は端的な回答をしている一方、認知機能が低下している MCI 患者はフィルラーや無音

時間が長くなり、発話時間長も長くなったのであろう。しかし、発話時間長は実験参加者が発話を開始してから終了するまでの時間であり、無音時間やフィルラーも含まれる。このため、無音時間長やフィルラー長の影響を受け有意差ありとなった可能性がある。今後、無音時間長やフィルラー長を除いて分析を行う必要がある。反応時間は実験参加者が最初の発話までの時間である。時間や場所を聞く見当識では最初の発話が「今年は…年です」や「ここは…です」と定型化されやすい点と、MCI 患者はフィルラーが増加しており、「えー」などの発話を最初に行っている可能性を考慮すると、反応時間は健常者、MCI 患者間の差が小さくなりやすい特徴量であると考えられる。このため、効果量の目安を下回ったからといって認知機能低下の兆しを見出すのに有効な特徴量ではないと結論づけるのは早計である。認知機能検査の得点との相関分析を行い、認知機能低下とどのような関係があるか分析を進める予定である。

### 4.3.2 人型ロボットとの日常会話音声における発話特徴

人型ロボットとの日常会話音声では、反応時間と Jitter(local)において健常者、MCI 患者間で有意差が得られ、効果量の目安を上回った。健常者に比べ MCI 患者の反応時間は平均的に長く、回答に時間を要している。認知機能検査に比べ日常会話は定型化されにくい。認知機能低下によって MCI 患者は回答を思考するのに時間を要したため、反応時間が長くなったのであろう。反応時間が日常会話において認知機能低下の兆しを見出すのに有効な特徴量の可能性がある。音響特徴量においては、Jitter(local)で効果量が目安となる値を上回り認知機能低下の兆しを見出すのに有効な特徴量であると示唆された。しかし、Jitter は年齢による喉の衰えや喉の疾患が影響している可能性がある。MCI 患者の平均年齢が健常者よりも高いことを考慮すると、認知機能検査の得点と相関分析を行い、認知機能低下に関する特徴であるか確認する必要がある。また、今回は質問応答パートと雑談パートを合わせ発話特徴の分析をした。今後は人型ロボットのどのような問かけが認知機能低下の兆しを見出すのに有効か検討するため、質問応答パートと雑談パートを分け分析する予定である。

表 3 認知機能検査音声の特徴量分析結果

| 特徴量名           | 健常者         | MCI 患者      | P value  | Cohen's d |
|----------------|-------------|-------------|----------|-----------|
|                | Mean (SD)   | Mean (SD)   |          |           |
| 発話時間長          | 2.25 (2.88) | 3.14 (6.03) | <0.01**  | 0.18      |
| 反応時間           | 0.53 (1.07) | 0.69 (1.02) | 0.02 *   | 0.16      |
| 無音時間長          | 0.42 (1.37) | 0.98 (3.26) | <0.01 ** | 0.22      |
| 無音時間割合         | 0.06 (0.14) | 0.11 (0.19) | <0.01 ** | 0.29      |
| フィルラー長         | 0.11 (0.32) | 0.19 (0.51) | <0.01 ** | 0.20      |
| フィルラー割合        | 0.04 (0.11) | 0.05 (0.12) | 0.18     | 0.09      |
| Jitter (local) | 0.04 (0.11) | 0.04 (0.01) | 0.44     | 0.05      |
| Shimmer(local) | 0.14 (0.04) | 0.14 (0.04) | 0.32     | 0.07      |

・ p<0.1 \* p<0.05 \*\* p<0.01

表 4 人型ロボットとの日常会話音声の特徴量分析結果

| 特徴量名           | 健常者         | MCI 患者      | P value  | Cohen's d |
|----------------|-------------|-------------|----------|-----------|
|                | Mean (SD)   | Mean (SD)   |          |           |
| 発話時間長          | 4.65 (7.42) | 3.80 (5.07) | <0.01 ** | 0.13      |
| 反応時間           | 0.88 (1.01) | 1.12 (1.33) | <0.01 ** | 0.21      |
| 無音時間長          | 1.30 (2.48) | 1.10 (2.11) | 0.048 *  | 0.09      |
| 無音時間割合         | 0.15 (0.21) | 0.14 (0.22) | 0.33     | 0.04      |
| フィルラー長         | 0.20 (0.60) | 0.54 (0.05) | 0.048 *  | 0.09      |
| フィルラー割合        | 0.03 (0.09) | 0.09 (0.16) | 0.16     | 0.06      |
| Jitter (local) | 0.03 (0.01) | 0.03 (0.00) | <0.01 ** | 0.21      |
| Shimmer(local) | 0.12 (0.03) | 0.13 (0.03) | <0.01 ** | 0.18      |

・ p<0.1 \* p<0.05 \*\* p<0.01

### 4.3.3 認知機能検査音声と人型ロボットとの日常会話音声の発話特徴比較

発話区間に関する特徴量は、認知機能検査音声に比べ、人型ロボットとの日常会話音声では無音時間割合で有意差を得られなかった。また、認知機能検査音声において発話時間長や無音時間長は健常に比べ MCI 患者が長いのに対し、人型ロボットとの日常会話音声では、MCI 患者の方が短いという異なる傾向を得た。人型ロボットとの日常会話では健常者に比べ MCI 患者は発話量が減少し、それに伴い無音時間も減ったのであろう。また、前述したように MCI 患者は反応時間が長くなるという傾向が得られている。発話量が減り反応が遅くなるというのは、文献[9]の“認知症者の多くは寡黙であり、反応も遅くなる”という知見と合致する。人との日常会話と同様の発話特徴を人型ロボットとの日常会話でも得られると示唆された。音響特徴量では、認知機能検査音声で有意差を得られなかった Jitter(local)と Shimmer(local)で有意差が得られ、人型ロボットとの日常会話特有の特徴量である可能性がある。

有意差のある特徴量や発話特徴の傾向は異なるが、人型ロボットとの日常会話音声において認知機能検査音声より多くの特徴量で有意差が得られた。また、識別力を効果量という指標で評価したところ、認知機能検査音声では3種類の特徴量で有意に異なる差を得た。人型ロボットとの日常会話音声では反応時間と Jitter(local)の2種類の特徴量で有意に異なる差が得られ、人型ロボットとの日常会話から認知機能低下の兆しを見出す可能性が示された。

## 5. アンケート結果

人型ロボットとの日常会話音声を取録後、ロボットとの会話が自然だったか、今後も話してみたいと思ったかについてアンケートを実施した。各項目0~5の5段階評価で、5に近いほど会話が自然と感じ、今後も話したいと思っていることを意味する。回答の集計が済んだ健常者43名、MCI患者42名の結果を図3に示す。

ロボットとの会話が自然であったかの評価は、MCI患者に比べ健常者は0.2ポイント低いという結果になった。健常者は認知機能が正常であるため、ロボットとの日常会話に違和感を感じた可能性がある。対して、MCI患者は認知機能の低下により違和感を感じにくい可能性がある。認知症患者に対しても人型ロボットとの日常会話音声を取録しているため、認知症患者のアンケート結果も集計し認知機能低下によりロボットとの会話の違和感に気づき辛くなるのか分析を行う予定である。

今後もロボットと話してみたいと思ったかの評価は健常者、MCI患者共に4ポイントを上回り、ロボットとの日常会話を今後も話したいと好意的に受け止めていると示された。これより、日常的にロボットとの日常会話を行なってくれる可能性が示された。

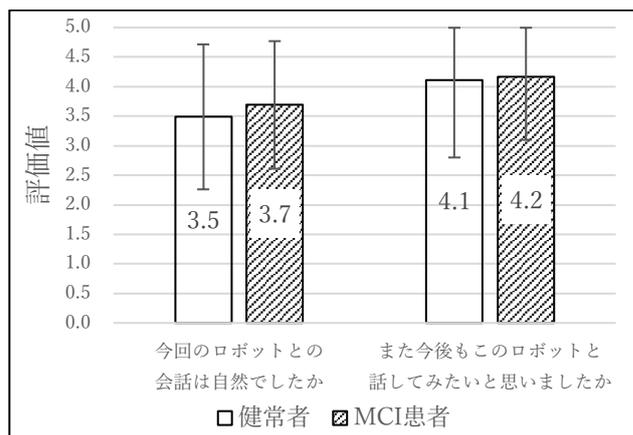


図3 アンケート結果

## 6. おわりに

本研究では、人型ロボットとの日常会話音声を用いた認知症簡易スクリーニングのための基礎的検討として従来手法の認知機能検査音声と比較を行った。発話特徴分析の結果、人型ロボットとの日常会話音声において認知機能検査音声より多くの特徴量で有意差が得られた。さらに、アンケート結果から人型ロボットとの日常会話が高齢者に受け入れられると示唆され、人型ロボットとの日常会話を用いた認知症の簡易スクリーニングの可能性を示した。今後は、有意差を得た特徴量と認知機能検査の得点との相関を分析し認知機能低下との関係を明らかにする予定である。

## 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] 羽生春夫, “1. 初期診断 (MCI を含む),” 日本内科学会雑誌, 100.8, pp.2109-2115, 2011.
- [2] Kato et al., “Detection of mild Alzheimer’s disease and mild cognitive impairment from elderly speech: Binary discrimination using logistic regression,” 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC). IEEE, pp. 5569-5572, 2015.
- [3] 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.
- [4] 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.
- [5] 吉井 他, “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,” 第19回情報科学技術フォーラム, E-001, pp.109-110, 2020.
- [6] 西村 雅史, 小林 悠一, 桐山 伸也, 峰野 博史, “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,” 音響学会講演論文集, 2-4-9, pp. 1309-1310, 2015.
- [7] 水本篤 他, “研究論文における効果量の報告のために。基礎的概念と注意点,” 英語教育研究 31, pp. 57-66, 2008.
- [8] 池田学, “認知症,” 高次脳機能研究 (旧 失語症研究), 29.2, pp. 222-228, 2009.
- [9] 池田学, “認知症者のコミュニケーション,” 高次脳機能研究 (旧 失語症研究), 35.3, pp. 292-296, 2015.