

特別賞

自然発話を活用した地域在住高齢者の認知機能
推定手法に関する研究

東京大学大学院 新領域創成科学研究科 人間環境学専攻
博士後期課程3年

五十嵐 俊治

1. 緒 言

1.1 研究の背景

日本の高齢化率は29.1%に達し、世界で最も高い水準にある。高齢社会白書によれば、総人口は減少する一方で、65歳以上の人口は3627万人と過去最多となり、2065年には全人口の38.4%を占めると予測されている[1]。平均寿命の延伸が進む中で、健康寿命とのギャップ拡大が課題となっており、QOLの低下が懸念されている。この状況を受けて、厚生労働省は「健康寿命延伸プラン」を策定し、2040年までに健康寿命を3年以上延ばすことを目指している[2]。同プランでは、①生活習慣の改善、②疾病の予防、③介護・認知症予防の3分野を中心に、地域間格差の解消を進めている。

ただし、こうした支援には継続的な介入が不可欠であり、人的リソースの不足が大きな障壁となっている。2025年には約245万人の介護人材が必要とされる見込みだが、現状の190万人から55万人の増員が求められている[3]。介護人材の慢性的な不足は制度創設以来の課題であり、労働人口の減少も相まって根本的な解決は困難とされている[4]。加えて、かつて主流だった多世代同居から、近年では独居・高齢者夫婦世帯への転換が進んでいる[5]。そのため、家族や同居者によるサポートが難しいケースも増加し、自立生活の継続には技術を活用した支援体制の整備が急務である。

「認知症疾患診療ガイドライン」[6]によれば、認知症とは後天的な脳障害によって知的機能が低下し、社会生活に支障をきたす状態である。記憶や見当識、言語、判断力などの機能低下は日常生活に深刻な影響を及ぼす[7,8]。認知症は加齢とともに発症率が上がり、2050年には全世界で1億5200万人に達する見込みである[9,10]。日本でも2040年には認知症高齢者が584万人、MCI(軽度認知障害)の高齢者が612万人に達するとされている[11]。

現在、認知症の根治的治療法は存在せず、早期発見と介入が鍵となる。PET(陽電子放射断層撮影法)[12]やMRI(磁気共鳴画像法)といった高度検査は有効であるが、高価かつ侵襲的であるため、MMSE(Mini-Mental State Examination)[13]やHDS-R(Hasegawa Dementia Scale)[14]といった簡易検査が一般的に用いられている。しかし、これらの検査は定型的で繰り返しの使用に適さず、検査されること自体が精神的負担をもたらす。実際、認知症検査中に不安や怒り、拒否反応を示す高齢者も少なくない[15]。こうした背景を受け、観察型の臨床認知症評価(CDR)[16]やN型高齢者精神状態評価尺度(NMスケール)[17]も使用されているが、主観的な判断が含まれるため、評価者のスキルに依存するリスクがある。実際に介護者が入所者の認知状態を把握しきれていない事例も報告されている[18]。

そこで注目されているのが、自然な発話に基づく非侵襲的な認知機能スクリーニングである[19](表1)。これは医療機関での拘束を避けながら、定期的にモニタリングを行う方法として有効であり、特にロボットなどの技術を活用した生活支援との親和性が高い。

第5期科学技術基本計画で提唱された「Society 5.0」(図1)では、ロボットによる見守りや健康促進、病気の早期発見、データ連携による最適治療の提供といった未来像が示されており[20]、音声技術を活用した認知機能推定はその実現に貢献すると期待されている。近年では、ロボットの他にもスマートスピーカーなど音声を使用するデバイスは各世帯に普及しつつあり、利用者から得られる発話内容を用いる認知機能推定手法も同様に受容されやすいと推測される。そこで、本研究では居宅環境において自然発話より得られる音声情報から認知機能推定手法の開発に焦点を当てることとする。

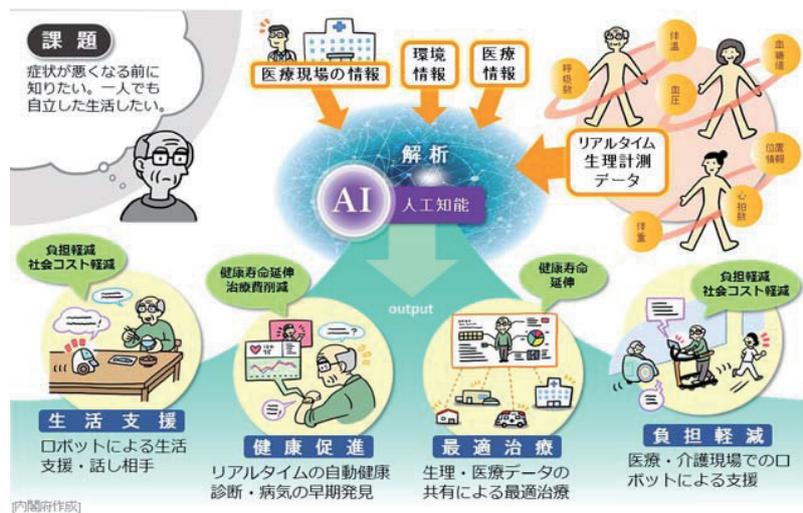


図1 内閣府 Society 5.0の医療・介護分野における新たな価値の事例[20]

1.2 研究の目的

これまでに述べてきた日本の高齢者を取り巻く課題と認知機能推定についての先行研究における課題改善のため、認知機能低下の早期検知と介入判断を支援するシステムを提案することを目指し、本研究の目的を「居宅環境において自然発話より得られる音声情報を用いた認知機能推定手法の開発を行い、検査者としてAIエージェントを使用することによる高齢者への心理負担を評価すること」とする。

1.3 研究の目的

本研究の目的に対して、本研究における研究目標を以下のように定める。

- ① 認知機能を推定するための自然な発話内容とその適用範囲の明確化
- ② 自然な発話内容から抽出された特徴量による認知機能推定モデルの開発とその評価
- ③ 検査者としてAIエージェントを用いた場合の高齢者の心理的負担の評価

表1 検査手法の分類と Pros cons の比較

分類	検査手法	pros	cons
医療機器による検査	PET	以前から存在するエビデンスに基づいたロバストな手法。画像によるエビデンスを得ることができる	身体への侵襲性が高い
	MRI		30 - 60分程度の身体拘束を必要とする
簡易スクリーニング検査	HDS-R	短時間(10~15分)で実施できる。検査者の負担が少ない。	能力検査の形態による認知症のレッテル貼りやステイグマなどの精神的負荷
	MMSE		
生活行動観察	CDR	観察による検査のため、高齢者本人への指示は伴わない。	検査と比較すると客観性に欠け、評価者の知識や経験によって結果が異なる場合がある。
	NM Scale		
自然会話テスト	CANDy	認知症患者の会話の特徴を用いた検査のため高齢者本人への負担が少ない。	判定は医師や心理士が行うため、検査者には専門的な医学知識が必要となる。また、会話内容の指導がほとんどなく、話題が個人的なものになり、再現性に課題が残る。

2. 対人ベースでの自然発話による認知機能推定手法

2.1 手法

初対面での会話の例として、精神科や老年病科では治療における包括的な支援のため、患者に関連情報の聞き取りを行うインタビュー面接が実施されている。生活史のヒアリング項目は、実務で使用可能な内容が求められるため、まず著者が実際に病院での心理師のインタビュー面接に立ち会い、質問項目の調査を行った(調査期間：2021年9月～10月)。対象となる患者によって、質問項目は異なるため、その中から重要と思われる項目や重複項目を削除し、原案を作成した。その後、病院に勤務する公認心理師5名に項目を確認してもらい、質問順序の変更及び加筆修正を行った。

質問項目の内容は全体で5つのカテゴリ、30個の質問から構成されている(表2)。カテゴリとしては、①病院に来るまでの流れ、②生活史、③普段の生活、④興味・関心、⑤この後の予定に分けられ、それぞれのカテゴリに関連した質問がその下の階層に含まれる[21]。

表2 開発した質問項目

<p>① <u>・ここに来るまでの流れ</u></p> <p>Q1. ご自宅はどちらになりますか？</p> <p>Q2. 今日はここまで来るのに、どれくらい時間がかかりましたか。</p> <p>Q3. お家を出られた後、どのように来られましたか？</p> <p>Q4. そうなると、本日は何時くらいに家を出ましたか。</p>
<p>② <u>・生活史</u></p> <p>Q5. ご自身のお生まれはどちらになりますか？</p> <p>Q6. ご兄弟はいますか(いる場合、何人いらっしゃいましたか？)</p> <p>Q7. どちらの小学校に通われていましたか？</p> <p>Q8. 小学校卒業後はどうなされましたか？(どちらの中学校に通われていましたか？)</p> <p>Q9. 中学校卒業後はどうなされましたか？(どちらの高校に通われていましたか？)</p> <p>Q10. お仕事では、どのようなことをなさっていましたか？(思い出に残っている話がありますか？)</p> <p>Q11. ご結婚はされていらっしゃいますか？(それはいつですか？)</p> <p>Q12. お子さんはいらっしゃいますか？(お子さんは(それぞれ)どちらにお住まいですか？)</p>
<p>③ <u>・普段の生活</u></p> <p>Q13. 普段はどのようにお過ごしですか？(おおよその1週間のスケジュールを教えてください。)</p> <p>Q14. 朝は何時に起きて、何時に寝ていますか？</p> <p>Q15. 外出はどのくらいの頻度でされますか。(どこに行かれることが多いですか？)</p> <p>Q16. お風呂は毎日入っていますか。(湯船にはつかりますか？)</p> <p>Q17. 食事の支度はどうされていますか？(3食召し上がっていますか。/ 昨夜は何を食べましたか？)</p> <p>Q18. 家の掃除はどうされていますか？(どのくらいの頻度でされますか？)</p> <p>Q19. 洗濯はどうされていますか？(どのくらいの頻度でされますか？)</p>

④ ・興味・関心

- Q20. 最近、テレビやネットで気になっているニュースについて教えてください。
- Q21. 最近あった悲しい出来事について教えてください。(それはいつ起こったことですか?)
- Q22. 最近あった不安な出来事について教えてください。
- Q23. 最近、腹が立った出来事について教えてください。
- Q24. 嫌な気分になった最近の出来事について教えてください。
- Q25. 最近あった出来事で驚いたことを教えてください。
- Q26. 最近あった楽しい出来事について教えてください。(それはいつ頃の出来事ですか?)
- Q27. あなたが尊敬している人について教えてください。
- Q28. 最近夢中になっていることは何ですか?

⑤ ・この後の予定・時間の流れ

- Q29. 今日この後のご予定はどうなっていますか? (どのようにご自宅までお帰りになりますか。)
- Q30. 前回の受診日はいつでしたか?

2.2 評価

2.2.1 評価環境

開発した質問項目によって、認知症患者からのデータを取得し認知機能の推定ができるか評価するため、病院における評価実験を実施した。参加者の募集は、2022年8月から9月にかけて65歳以上の高齢者、医師により認知症の診断を受けている患者という条件で絞り込みを行い、研究概要の説明を行ったうえで同意を得ることのできた方に参加してもらった。

実験に参加したのは72歳から91歳の29名(男性7名、女性22名)であった。同意書については、本人のみが来院している場合は本人から、親族が付き添いで来ている場合には親族の方からも同意を得た。本研究は、東京大学医学部附属病院の倫理審査委員会の承認を得て実施された。

場所は東大医学部附属病院の検査室にて実施した。検査室内の机に、参加者と質問者が対面で座り、質問を実施した。レコーダーと小型カメラ(Gopro hero10)によって音声とタイムラプス画像が記録されている。図2は病院の心理検査室の様子を示す。a)はアクリルボードで仕切られたテーブル。b)奥側に質問者が座り、手前側に参加者を座らせた。



図2 検査室における検査環境 (a) (b) [21]

検査内容は、a) MMSE、b) GDS、c) インテーク面接における生活史のヒアリングに分けられる。a) MMSE は認知機能の推定のために行い、b) GDS は、高齢者用の抑うつ評価のために実施した。対象とした全ての参加者から発話データを取得することができ、MMSE の結果、MMSE が20点以下の中等度の認知症が12名、MMSE が21点以上23点以下の軽度の認知症が8名、MMSE が24点以上・27点以下の軽度認知障害 (MCI) が9名であった。また、GDS の結果、27名はGDS のカットオフ値である7点を下回ったが、2名がカットオフ値よりも高い値を記録した。

2.2.2 評価手法

また、認知機能の分類を行う教師付き学習として、中程度認知症群、軽度認知症群、MCI 群の三値分類と、それぞれの二値分類を行うモデルを構築した。自然言語処理研究のベースラインとされている日本語学習済みのBERT (Bidirectional Encoder Representations from Transformers) [22]を用い、日本語の辞書としてはNEologd、文章のトークナイゼーションにMeCabを用いた[23]。ファインチューニングにおけるパラメータとしては、バッチサイズ: 1、学習率 (Adam): $2e-5$ として設定し、過学習防止のためにEarly stoppingの実装を行った。

中程度認知症群の12名からは、334件の回答が得られた。合計36734語であり、1件あたりの平均は109.98語であった。また、軽度認知症群の8名からは、234件の回答が得られた。合計25647語であり、1件あたりの平均は109.60語であった。MCI群の9名からは、256件の回答が得られた。合計28423語であり、1件あたりの平均は107.26語であった。実験参加者の属性ごとにデータ数の偏りがでる不均衡データに対応する手段としてEasy data augmentationによるオーグメンテーションを実施した。

2.3 結果

2.3.1 多値分類および二値分類の結果

中等度認知症群、軽度認知症群、MCI群の三群を対象とした多値分類モデルの検証を行った結果、学習データにおける正解率は0.919、検証データでは0.530、テストデータでは0.405となった。中等度認知症群と軽度認知症群を区別する二値分類モデルでは、学習データでの正解率は0.821、検証データで0.482、テストデータで0.488であった。

さらに、中等度認知症群とMCI群の分類では、学習データで0.956、検証データで0.633、

テストデータで0.767の正解率が得られた。また、軽度認知症群と MCI 群を分類したモデルでは、学習データでの正解率は0.965、検証データで0.660、テストデータで0.700であった。

2.3.2 オーグメンテーションを行った多値分類および二値分類の結果

データオーグメンテーションを適用した多値分類モデルの検証結果として、学習データの正解率は0.994、検証データは0.992、テストデータは0.991となり、大幅な精度向上が確認された。中等度認知症群と軽度認知症群の二値分類モデルでは、学習データでの正解率が0.984、検証データが0.971、テストデータが0.972であった。

また、中等度認知症群と MCI 群の分類においては、学習データが0.999、検証データが0.997、テストデータが0.996という高い正解率を示した。加えて、軽度認知症群と MCI 群の分類では、学習データで0.991、検証データで0.987、テストデータで0.985という結果が得られた(図3)。

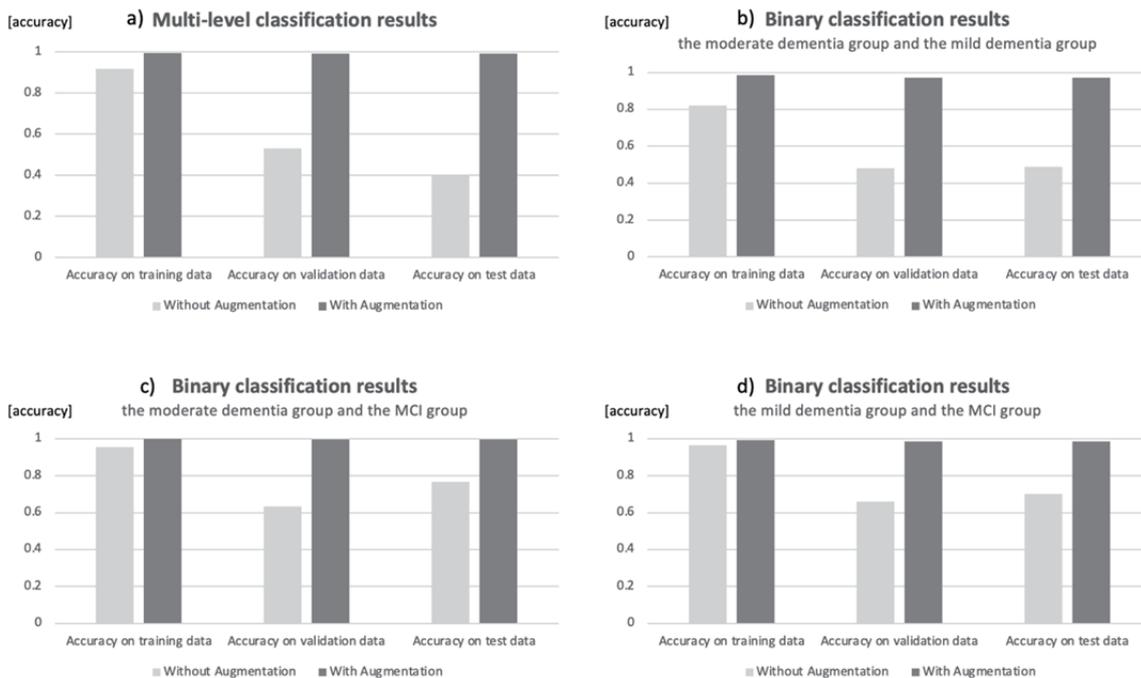


図3 多値分類および二値分類のオーグメンテーションの有無による比較[21]

2.4 分析

1回答あたりの平均無音秒数を3群で比較したところ、中等度認知症群(12名、1542サンプル)は2.776秒、軽度認知症群(8名、1390サンプル)は1.849秒、MCI群(9名、1895サンプル)は1.589秒という結果であった。この差に対して対応しないt検定を実施したところ、中等度と軽度認知症群、および中等度とMCI群との比較においては、いずれも $p < 0.01$ ($t = 4.056$, $t = 5.332$)と有意差が認められた。軽度認知症群とMCI群の比較についても、 $p < 0.05$ ($t = 2.032$)で有意差が見られた(図4)。これらの結果から、同一の半構造化インタビューにおいては、認知機能が低下するほど発話中の平均無音時間が長くなる傾向があることが示唆された。

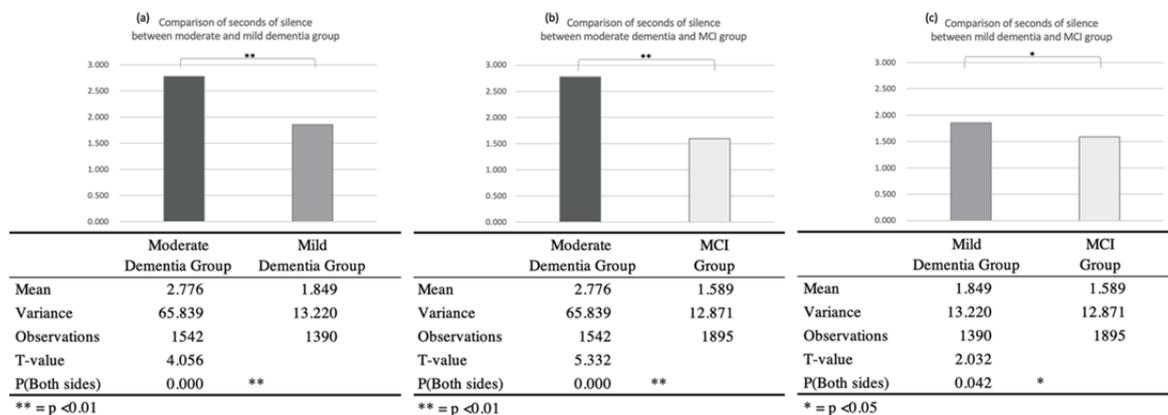


図4 (a) 中等度認知症群と軽度認知症群、(b) 中等度認知症群と MCI 群、(c) 軽度認知症群と MCI 群の沈黙秒数の比較 [24]

3. バーチャルエージェントを活用した事前発話による認知機能の推定

3.1 関連研究

日常会話を用いた認知機能のスクリーニング技術の開発は、被検者のみならず、介護者や医療提供者にとっても受診にかかる負担やコストを低減しうる手法として注目されている。会話型 AI エージェントは、医療リソースの逼迫やアクセス性の課題に対処するデジタル技術の一環として導入が進められており、視覚的には漫画調のキャラクターから高度な三次元アバターまで多様な形式が存在する [24, 25]。これらのエージェントは、自然言語処理や機械学習を活用し、モバイルやウェブ、音声インターフェースを通じてユーザーと対話する仕組みである。Walker らの研究では、イギリスにおいて事前録音された質問を用いたバーチャルエージェントとの対話を、軽度認知障害 (MCI) 群および健常群に対して実施したところ、MCI 群では発話時間が有意に短い傾向が見られた [27]。このことは、音声データの分析が認知機能障害のスクリーニングに資する可能性を示している。

さらに、音声インターフェースの要件に関する調査では、認知症患者、介護者、非認知症の高齢者それぞれに対して異なるニーズが明らかとなった。具体的には、認知症のある高齢者には励ましや導きのある応答が好まれる一方で、認知機能に問題のない高齢者に対しては、過剰な支援は不要であると認識される傾向が示されている [28]。

国内においては、NTT コミュニケーションズが2022年に「脳の健康チェックフリーダイヤル」の運用を開始した [29]。このサービスは、約20秒間の発話内容および音声特徴を AI が解析することで、認知機能の変化を把握可能とするものであり、ユーザーは非対面かつ匿名で認知機能の簡易チェックを行うことができる。また、中等度の認知症高齢者群を対象とした調査では、エージェントとの対話が成立し、好意的に受け入れられたほか、エージェントが研究参加者の興味関心に沿った話題を提示した際に発話量が増加するなど、アニメーション型エージェントの有用性が示唆された [30]。

ロボット技術に関しては、NEC が開発した Papero-i を用いた複数の研究が実施されている。五十嵐らの研究では、ロボットによる自己開示が高齢者の発話量および発話内容の多様性を向上させる効果が確認されている [31]。また、小林らは、Papero-i による自然会話から

認知症の兆候を検出し、異常が見られた場合には家族へ通知を行う早期発見システムを構築し、その有効性を報告している[32]。

これらの AI エージェントによるスクリーニングは、実施場所や専門人材の制約を受けにくく、人的偏見を伴わないことから、高齢者が抱える懸念や個人的な問題についても、人間相手よりも安心して語れる可能性がある[33]。こうした特徴を踏まえ、AI エージェントを用いた認知機能評価は、今後の社会実装に向けた重要なアプローチとして位置づけられる。

3.2 手法

3.2.1 研究参加者と取得データ

高齢者で、同地域のデイサービス利用者17名、シルバー人材センター所属者17名の計34名を対象とした(男性12名、女性22名)。当初は各19名を対象としていたが、各2名は調査期間中に死亡や転院などにより追跡が不可となった。なお、参加者は本調査の目的に同意を得られた者とし、認知機能が低下している高齢者については家族にも説明のうえ、同意を得た。

本研究では、高齢者の認知機能と精神的負担に関する影響の差異を評価するために、複数の評価尺度と測定方法を用いた。認知機能の測定には杉下訳の MMSE-J、参加者の不安の度合いの測定には新版 STAI と VAS(Visual Analogue Scale) を使用した。うつ病が会話に与える影響を考慮し、老年期うつ病評価尺度(Geriatric Depression Scale:GDS-15)を用いて評価した。

3.2.2 AI エージェントの表示環境

本研究における AI エージェントは、対人対話を模倣するアバターを用いたシステムとして設計され、複数の技術モジュールを組み合わせて構築された。対人での自然発話との比較を行うため、発話をしているだろうと参加者が推測する人物と外見が著しく異なる印象を受けるアバターを使用してしまうと、検査結果に影響を与える可能性がある。そのため、エージェントの外観は筆者自身の容貌を模倣し、3D アバターとして VRoid Studio[34] を用いてモデリングし(図5)。ブラウザ上でアバターを表示するために、Pixiv が提供するオープン

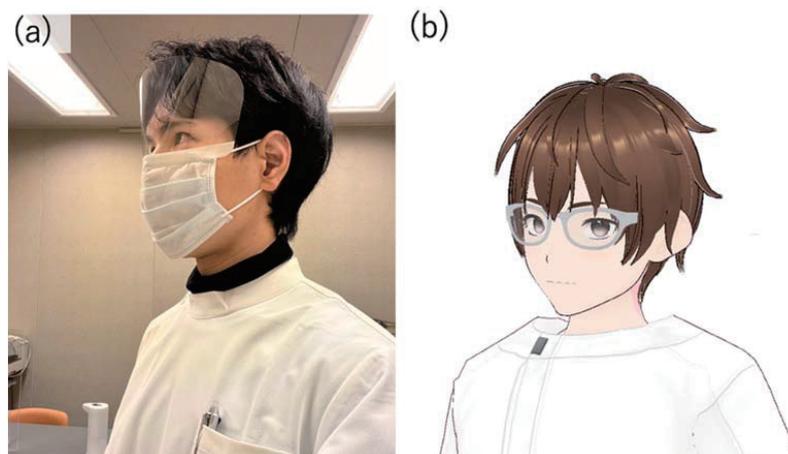


図5 (a) 対人会話を実施した筆者の外見および、(b) 筆者の外見を基にモデリングされた AI エージェントの外見[36]

ソースライブラリ @pixiv/three-VRM[35]を活用した。これは three.js ベースで VRM 形式の 3D アバターを表示可能とするものである。AI エージェントは MacBook Pro 上に表示され、背景には対話が行われる施設の実写画像を用い、実環境の一体感を演出した。

ユーザーの発話認識には、Web Speech API(Speech Recognition) [37]を採用した。エージェントとの発話が重なることで認識精度が低下することを避けるため、参加者にはピンマイクを装着してもらい、録音はボタン押下時のみに限定する設計とした。エージェントの音声出力には、Koeiro API[38]を利用し、自然な音声合成を実現した。また、音声に連動してアバターの口元が動くようリップシンク機能を組み込んでおり、視覚的なリアリティを高めている。

会話設計は、半構造化面接に基づくプロトコルにより構築された。すべての質問は事前に設定されており、各質問に対し、ユーザーの回答に応じた返答を AI エージェントが行った後、次の質問に進む形式を採用している。これにより、各トピックの会話の開始と終了が明確化され、対話の一貫性が保たれる。なお、回答への応答生成には ChatGPT API[39]を用いた。これにより、ユーザーの発話内容に即した柔軟な返答を実現しており、図6に全体概要を示す。

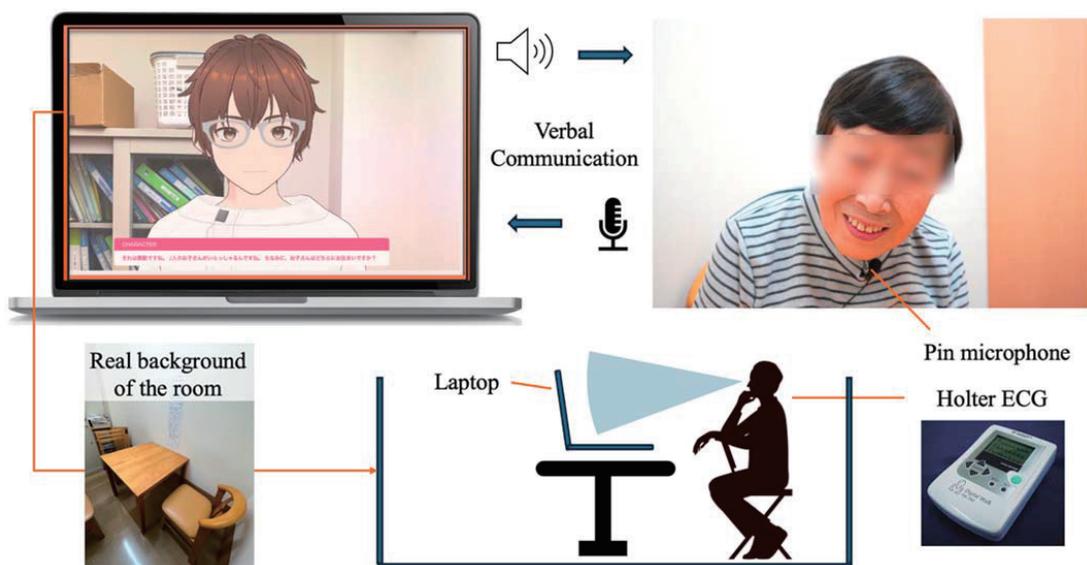


図6 AI エージェントによる認知機能推定のための日常会話システムのプロトコル[40]

3.3 評価

本研究では、参加者の心理的負担と認知機能への影響を評価するために実験的手法を採用した。実験の目的は、異なるインタラクティブ環境下における参加者の自記式の心理検査や心電図の変化を通して、心理的負担を可視化することである。

開発した質問項目によって、認知症患者からのデータ取得ができるか評価するため、MMSE と、人間との日常会話中における参加者の心理的負担に与える影響を比較する実験を実施した。研究は、地域在住の高齢者で認知症デイサービスセンターに通所する利用者 19名(男性3名、女性16名)を対象とした。本研究は、東京都立産業技術大学院大学の倫理審査の承認を得て実施した。

参加者それぞれに対し、MMSE(Mini-Mental State Examination)を用いた認知機能検査、人間との日常会話、およびAIを介した日常会話の3つの検査を実施した。まずSTAI(State-Trait Anxiety Inventory) [41,42]により、参加者の現在および通常時の不安尺度を測定した。さらに、VAS(Visual Analogue Scale) [43]を用いて、「現在、精神的な負担をどれくらい感じているか」についての自己評価を行った。

実験の各セッションの前には、参加者に5分間の開眼安静を指示し、心拍数の安定を確認した。その後、30分間の日常会話を行い、会話後には再びSTAIとVASを用いて、セッション中の精神的負担を評価した。次に、15分間のMMSEによる認知機能検査を実施し、同様にSTAIとVASで評価を行った。日常会話における質問の順序が参加者の回答成績に影響することを避けるため、どの参加者も統一して実施した。AIエージェントとの対話試験は、対人による対話試験から一ヶ月以上の間隔を空けて実施した。

また、検査者が人間およびAIエージェントである場合において、発話によって得られたテキストデータによって高齢者の健常群と認知機能低下群に文章的特徴量をもとに分類することができるか検証を実施した。10分割交差検証を行い、全てのデータで学習および検証がされるようにした。

3.4 結果

参加者の平均年齢は83.21歳であり、MMSEの平均スコアは14.05(標準偏差4.58)であった。これは、認知機能が低下している高齢者が研究対象として適切であることを示している。抑うつ症状を評価するGDSの平均スコアは3.21(標準偏差2.59)であり、GDSのカットオフ値を上回る参加者はおらず、抑うつ症状を示す参加者は確認されなかった。

検査者が人間である場合において、テキストデータによって高齢者の健常群と認知機能低下群の2値分類の結果、学習データでのAccuracyは0.947、テストデータでのAccuracyは0.882であった。検査者がAIエージェントである場合において、テキストデータによって高齢者の健常群と認知機能低下群の2値分類の結果、学習データでのAccuracyは1.000、テストデータでのAccuracyは0.971であった。

実験開始時のVAS(Visual Analogue Scale)の平均値は5.50(SD=10.56)であり対人間による日常会話後のVAS平均値は7.15(SD=12.32)、AIエージェントによるものでは実施後VAS平均値は9.76(SD=14.42)であった。また、MMSE実施後のVAS平均値は42.88(SD=29.13)であった。STAI(State-Trait Anxiety Inventory)による現在の状態不安の平均値は29.4(SD=5.45)であり、対人間による日常会話後のSTAI平均値は28.26(SD=5.98)、AIエージェントによるものでの実施後平均値は28.85(SD=5.71)であった。MMSE実施後の状態不安(STAI)の平均値は48.85(SD=13.85)であった(表3)。

表3 VASとSTAIによる平均値の差の比較[44]

	Age	GDS-15 Score	VAS				STAI				
			MMSE Score	Before Intervention	Human Conversation	AI Agent Conversation	MMSE	Before Intervention	Human Conversation	AI Agent Conversation	MMSE
Mean	78.71	2.48	21.09	5.50	7.15	9.76	42.88	29.41	28.26	28.85	48.85
SD	6.77	2.45	8.26	10.56	12.32	14.42	29.13	5.45	5.98	5.71	13.85

3.5 分析

4. 発話データに関する分析

認知機能による分類が可能になった要因分析のため、発話データに関する量的な分析を行った。健常高齢者群について、人間との対話においては、発話数の平均は123.96語であり、発話時間の平均は42.56秒、沈黙時間は4.16秒であった。また、AIとの対話における発話数の平均は21.86語であり、発話時間の平均は10.22秒、沈黙時間は3.76秒であった。(表4)。認知機能低下群について、人間との対話においては、発話数の平均は104.87語であり、発話時間の平均は24.24秒、沈黙時間は4.21秒であった。また、AIとの対話における発話数の平均は92.58語であり、発話時間の平均は23.42秒、沈黙時間は5.34秒であった(表5)。

表4 健常高齢者群における取得データの平均及びSD[45]

	Age	MMSE	Number of answer words per question		Speech time per question (sec)		Silence time per question (sec)	
			with human	with AI	with human	with AI	with human	with AI
Mean	74.73	28.67	123.96	21.86	42.56	10.22	4.16	3.76
SD	5.48	1.35	169.66	26.95	77.09	17.51	4.14	4.70

表5 認知機能低下群に所属する参加者における取得データの平均及びSD[45]

	Age	MMSE	Number of answer words per question		Speech time per question (sec)		Silence time per question (sec)	
			with human	with AI	with human	with AI	with human	with AI
Mean	82.59	13.47	104.87	92.58	24.24	23.42	4.21	5.34
SD	4.96	3.84	123.66	106.06	27.15	25.75	4.18	5.82

発話数について年齢および認知機能による影響を検討するため、健常高齢者群と認知機能低下群の発話語数について0.05%水準で不当分散の両側t検定を実施した。その結果、全体のデータにおいては有意差がみられなかった($t=4.97$)。また、人間との対話における健常高齢者群と認知機能低下群の発話語数を比較したところ、有意差はみられなかった($t=2.27$)。さらに、AIとの対話においてもシルバーと認知機能低下群の参加者の発話数に有意差はみられなかった($t=1.76$)。

健常高齢者群の参加者に関しては、AIとの対話と人間との対話の発話数を比較した結果、有意差はみられなかった($t=3.52$)。一方、認知機能低下群の参加者においては、AIとの対話と人間との対話の発話数に有意差が見られ、対人間の発話数の平均が92.58、対AIが104.88であった($t=0.01$)(図7)。

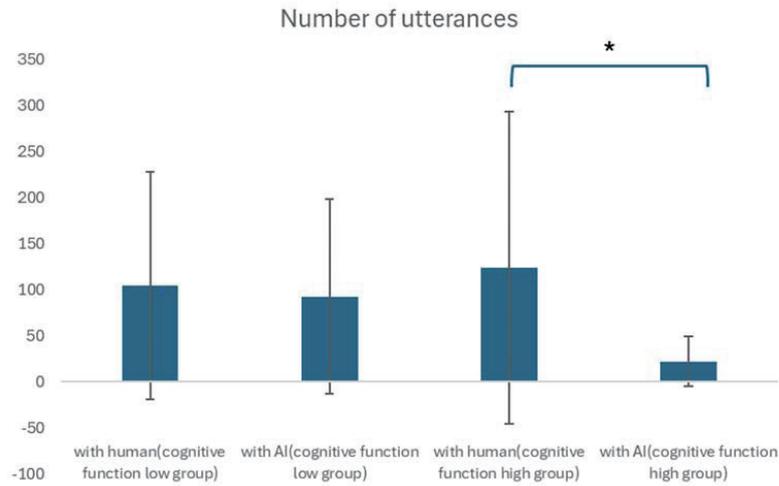


図7 発話数における高認知機能群と低認知機能群の比較[36]

5. 心理的負担の影響の比較分析

参加者の精神的負担を評価するために、Visual Analogue Scale(VAS)を用いた。34名の参加者について、面接前の状態と、対人会話実施後、AIエージェントとの会話実施後、およびMMSE実施後VASスコアに対し、Microsoft ExcelにてFriedman検定を実施した。p=1.5e-11であることから試験実施によりVASは変化した。Post-hocな2群の検定はBonferroniの調整による多重比較で行い、MMSE後のみ面接前の状態・対人会話実施後・AIエージェントとの会話実施後の各状況下に対し、全てP<0.01の水準で有意差がみられた(図8)。

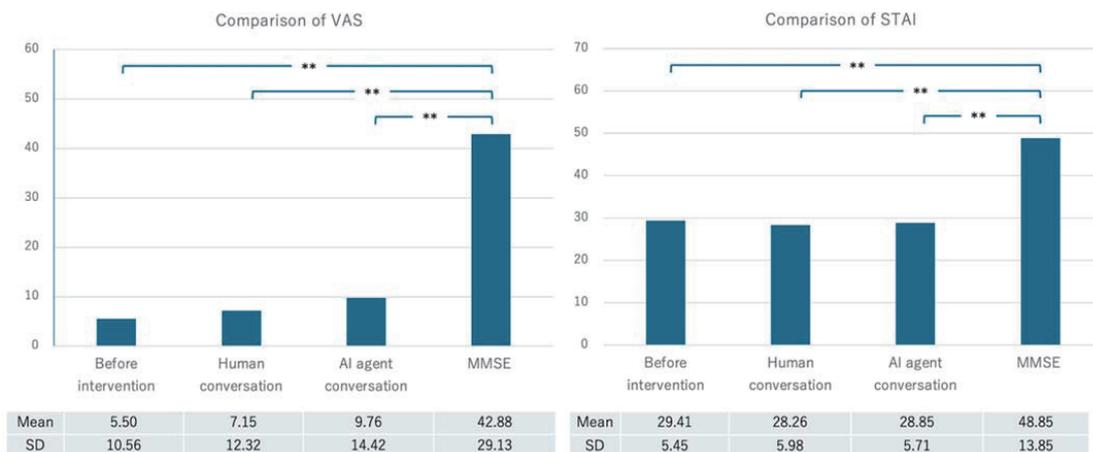


図8 VASとSTAIによる平均値の差の比較[44]

6. まとめと今後の展望

本研究では、地域在住高齢者を対象に、自然発話を用いた認知機能推定手法を開発し、検査者として AI エージェントを導入することで心理的負担への影響を検証し、早期の認知機能低下の発見と介入判断を支援するシステムの構築を目指した。自然な会話内容による認知評価の実現に向けて、臨床経験を有する心理師の協力のもと、初対面の高齢者にも対応可能な30項目の半構造化面接項目を作成した。さらに、開発項目に基づく自然発話と MMSE 実施時の心理的負担を比較する実験を実施した。

MMSE スコアに基づき、中等度認知症(≤20点)、軽度認知症(21~23点)、MCI(24~27点)の3群を対象とする自然言語処理モデルを構築し、その分類精度を評価した。また、認知評価中の心理的負担に与える影響を検討し、MMSE が日常会話と比較して有意に負担が大きいことを確認した一方で、AI エージェントとの対話は人間との会話と比較して有意な差が見られず、今後の認知機能のモニタリングや早期介入を促進する手段として期待される。以上の結果から、本研究の成果は、医療現場や居宅環境における認知機能評価手法の改良と、AI 技術の実用化に向けた重要な一歩となると考えられる。

7. 謝 辞

本研究の遂行にあたり、多大なるご指導とご支援を賜りましたすべての方々に、心より深く感謝申し上げます。まず、研究の構想から論文執筆に至るまで一貫して的確なご助言と温かいご支援を賜りました指導教員・陳先生に、深甚なる謝意を表します。また、本研究の基盤を整えてくださった高齢社会総合研究機構の飯島先生、医療法人社団つくし会の新田先生、東京大学医学部附属病院の亀山先生に厚く御礼申し上げます。質問項目の開発に際しては、東京大学医学部附属病院の石川様をはじめ、多くの心理師の皆様にご協力をいただきました。さらに、現場での実証実験の実施にご尽力くださった、やがわデイサービスの坂本所長およびスタッフの皆様にも深く感謝いたします。そして、本研究にご参加いただいたすべての実験協力者の皆様に、心より御礼申し上げます。皆様のご協力によって得られた貴重なデータが、本研究成果の基盤となりました。なお、研究期間中にご逝去された参加者の方もおられ、成果をご報告できなかったことが悔やまれますが、いただいたご協力を無駄にすることなく、今後の社会実装を通じてヘルスケア領域への貢献に努めてまいります。

ここに改めて、関係各位のご厚意に深く感謝申し上げます。

参考文献

1. 厚生労働省, 2022年度の高齢社会白書.
2. 厚生労働省, 健康寿命延伸プラン, 2019年.
3. 厚生労働省, 第7期介護保険事業計画に基づく介護人材の必要数について
<https://www.mhlw.go.jp/stf/houdou/0000207323.html>.
4. 土田耕司. 福祉現場における介護人材不足の背景. 川崎医療短期大学紀要, 2010.
5. 東京都福祉保健基礎調査 東京都福祉保健局, 平成 22 年度の東京都福祉保健基礎調査.
6. 日本神経学会. "認知症疾患診療ガイドライン 2017." 認知症疾患ガイドライン作成委員

- 会, 東京, 医学書院(2017): 1-384.
7. 尹智暎, 大藏倫博, 角田憲治, 辻大士, 鴻田良枝, 三ッ石泰大, ... & 金勳. (2010). 高齢者における認知機能と身体機能の関連性の検討. 体力科学, 59(3), 313-322.
 8. 福田大祐, & 森千鶴. (2014). 18) 初期の認知症者における遂行機能障害と IADL との関連性の検討. 日本看護研究学会雑誌, 37(3), 3_130-3_130.
 9. 神崎恒一. (2018). VI. 加齢に伴う認知機能の低下と認知症. 日本内科学会雑誌, 107(12), 2461-2468.
 10. Scheltens, Philip, et al. "Alzheimer's disease." *The Lancet* 397.10284 (2021): 1577-1590.
 11. 国立大学法人九州大学, 令和5年度老人保健事業推進費等補助金(老人保健健康増進等事業)認知症及び軽度認知障害の有病率調査並びに将来推計に関する研究報告書
 12. William. "Positron emission tomography and magnetic resonance imaging in the diagnosis and prediction of dementia." *Alzheimer's & dementia* 2.1 (2006): 36-42.
 13. 杉下守弘, et al. "MMSE-J (精神状態短時間検査 - 日本版) 原法の妥当性と信頼性." *認知神経科学* 20.2 (2018): 91-110.
 14. 加藤伸司. "改訂長谷川式簡易知能評価スケール (HDS-R) の作成." *老年精神医学雑誌* 2.11(1991): 1339-1347.
 15. Tiberti, C., Sabe, L., Kuzis, G., Garcia Cuerva, A., Leiguarda, R., & Starkstein, S. E. 1998 Prevalence and correlates of the catastrophic reaction in Alzheimer's disease. *Neurology* ,50, 546-548.
 16. Morris, John C. "The Clinical Dementia Rating (CDR) current version and scoring rules." *Neurology* 43.11 (1993): 2412-2412.
 17. 小林敏子, 播口之朗, 西村健・他: 行動観察による痴呆患者の精神状態評価尺度 (NM スケール) および日常生活動作能力評価尺度 (N-ADL) の作成. *臨床精神医学*, 1988, 17 (11): 1653-1668.
 18. H. Kawaguchi and S. Sato. 2002. A study on the evaluation of cognitive ability of demented elderly by others. *Care and Behavioral Sciences for the Elderly* 8, 2 (2002), 37-45. In Japanese.
 19. 大庭輝, et al. "日常会話式認知機能評価 (Conversational Assessment of Neurocognitive Dysfunction; CANDy) の開発と信頼性・妥当性の検討." *老年精神医学雑誌 / 「老年精神医学雑誌」編集委員会 編* 28.4(2017): 379-388.
 20. 内閣府, Society 5.0. https://www8.cao.go.jp/cstp/society5_0/.
 21. Toshiharu Igarashi, Yumi Umeda-Kameyama, Taro Kojima, Masahiro Akishita, Misato Nihei. Assessment of Adjunct Cognitive Functioning through Intake Interviews Integrated with Natural Language Processing Models. *Frontiers in Medicine* 10 (2023): 1145314., Jan 2023.
 22. Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).
 23. 佐藤敏紀, 橋本泰一, and 奥村学. "単語分かち書き用辞書生成システム neologd の運用 - 文書分類を例にして." *研究報告自然言語処理(NL)* 2016.15(2016): 1-14.
 24. Toshiharu Igarashi, Yumi Umeda-Kameyama, Taro Kojima, Masahiro Akishita, Misato Nihei. Questionnaires for the assessment of cognitive function secondary to intake interviews in in-hospital work and Development and Evaluation of a Classification Model Using Acoustic

- Features. *Sensors* 23.11 (2023): 5346. Feb 2023.
25. Laranjo, Liliana, et al. "Conversational agents in healthcare: a systematic review." *Journal of the American Medical Informatics Association* 25.9 (2018): 1248-1258.
 26. Luxton, David D. "Ethical implications of conversational agents in global public health." *Bulletin of the World Health Organization* 98.4 (2020): 285.
 27. Walker, Gareth, et al. "Characterising spoken responses to an intelligent virtual agent by persons with mild cognitive impairment." *Clinical Linguistics & Phonetics* 35.3 (2021): 237-252.
 28. Wolters, Maria Klara, Fiona Kelly, and Jonathan Kilgour. "Designing a spoken dialogue interface to an intelligent cognitive assistant for people with dementia." *Health Informatics Journal* 22.4 (2016): 854-866.
 29. NTT Communications "Started Free Trial of Brain Health Check Toll-Free Dial." [(accessed on 26 April 2024)]. Available online: <https://www.ntt.com/about-us/press-releases/news/article/2022/0921.html>.
 30. Hikino Junta, Nakano Yukiko, and Yasuda Kiyoshi. "Communication Support for Dementia Patients Using a Conversation Agent." *Proceedings of the 73rd National Conference* 2011.1 (2011): 195-196.
 31. Igarashi, Toshiharu, et al. "Eliciting a User's Preferences by the Self-Disclosure of Socially Assistive Robots in Local Households of Older Adults to Facilitate Verbal Human-Robot Interaction." *International Journal of Environmental Research and Public Health*.
 32. Kobayashi Tooru, et al. "Study on a System for Detecting Signs of Dementia Using a Social Media Mediating Robot." *Journal of the Institute of Electronics, Information and Communication Engineers D* 105.10 (2022): 533-545.
 33. Bickmore, Timothy W., et al. "Response to a relational agent by hospital patients with depressive symptoms." *Interacting with Computers* 22.4 (2010): 289-298.
 34. 3D Character Creation Software VRoid Studio [(accessed on 26 April 2024)]. Available online: <https://vroid.com/studio>.
 35. GitHub pixiv three-vm [(accessed on 26 April 2024)]. Available online: <https://github.com/pixiv/three-vm>.
 36. Igarashi, Toshiharu, et al. "Detailed Analysis of Responses from Older Adults through Natural Speech: Comparison of Questions by AI Agents and Humans." *International Journal of Environmental Research and Public Health* 21.9 (2024): 1170.
 37. Web Speech API - MDN Web Docs [(accessed on 26 April 2024)]. Available online: https://developer.mozilla.org/ja/docs/Web/API/Web_Speech_API.
 38. AI Service for Generating Voice and Facial Motion Koemotion [(accessed on 26 April 2024)]. Available online: <https://rinna.co.jp/news/2023/06/20230612.html>.
 39. <https://openai.com/index/introducing-chatgpt-and-whisper-apis/>
 40. Igarashi, Toshiharu, et al. "Qualitative Analysis of Responses in Estimating Older Adults Cognitive Functioning in Spontaneous Speech: Comparison of Questions Asked by AI Agents and Humans." *Healthcare*. Vol. 12. No. 21. MDPI, 2024.
 41. Spielberger, Charles D., et al. "The state-trait anxiety inventory." *Revista Interamericana de Psicologia/Interamerican journal of psychology* 5.3 & 4 (1971).

42. 肥田野直, et al. “新版 STAI マニュアル— State-Trait Anxiety Inventory-Form JYZ 実務教育出版.” Hidano, T., Fukuhara, M., Iwawaki, M., Soga, S., & Spielberger, CD (2000).
43. Huskisson, Edward C. “Measurement of pain.” *The lancet* 304.7889 (1974): 1127-1131.
44. Toshiharu Igarashi, Katsuya Iijima, Kunio Nitta, Yu Chen. Estimation of Cognitive Functioning of the Elderly by AI Agents: A Comparative Analysis of the Effects of Psychological Burden of Intervention. *Healthcare*. Vol.12, No.18, p.1821, Jul 2024.
45. Toshiharu Igarashi, Estimating Cognitive Function Using Spontaneous Speech in Older People Living in the Community, Pages:105, Published:2024-09-25, ISBN:978-99993-2-042-9, ELIVAbooks.