

アバターとの対話によるマルチモーダル情報を伴った 早期認知症の検出

田中 宏季^{1,a)} 足立 浩祥^{2,b)} 浮田 宗伯^{1,c)} 池田 学³ 敷井 裕光³ 工藤 喬^{2,d)} 中村 哲^{1,e)}

概要：

本稿では、早期認知症を検出するための新たなアルゴリズムを提案する。我々は、アバターによる質問機能をもつ認知症検出システムを開発した。本システムを開発するに当たり、既存で使用されているミニメンタルステート検査 (MMSE) やウェクスラーの記憶検査などの認知症を検出するための各種検査を参考とした。実験的評価として、18名 (9名の早期認知症者と9名の健康人) のマルチモーダル情報を伴ったデータを取録し、質問応答中の視聴覚特徴を抽出した。これらの特徴量を入力とした機械学習のアルゴリズムを構築し、ROC 曲線の AUC において、0.94%となった。これより、本システムが早期認知症を検出するツールとして有効であることを示した。

キーワード： 認知症, 音声対話システム, アバター

Automatic Detection of Early Stage of Dementia by Multimodal Information through Interaction of Computer Avatars

TANAKA HIROKI^{1,a)} ADACHI HIROYOSHI^{2,b)} UKITA NORIMICHI^{1,c)} IKEDA MANABU³ KAZUI HIROAKI³
KUDO TAKASHI^{2,d)} NAKAMURA SATOSHI^{1,e)}

Abstract:

This paper proposes a new approach to detecting very early stage of dementia automatically. We develop a computer avatar with spoken dialog functionalities that produces natural spoken queries referring to Mini Mental State Examination, Wechsler Memory Scale-Revised and other related questions. Multimodal interactive data of spoken dialogues from 18 participants (9 dementias and 9 healthy controls) are recorded, and audiovisual features are extracted. We confirm that the machine learning algorithm can classify into two groups with 0.94 detection performance as measured by areas under ROC curve. It is found that our system has possibilities to detect very early stage of dementia through spoken dialog with our computer avatars.

Keywords: Dementia, spoken dialogue, avatars

¹ 奈良先端科学技術大学院大学 情報科学研究科
Nara Institute of Science and Technology, Ikoma-shi, Nara,
Japan

² 大阪大学 保健センター
Osaka University, Toyonaka, Osaka, Japan

³ 大阪大学 医学系研究科
Osaka University, Suita, Osaka, Japan

a) hiroki-tan@is.naist.jp

b) hadachi@psy.med.osaka-u.ac.jp

c) ukita@is.naist.jp

d) kudo@psy.med.osaka-u.ac.jp

e) s-nakamura@is.naist.jp

1. はじめに

認知症とは、記憶、思考、および行動の領域においてその能力が低下することである [1]。認知症の早期発見は幾つかの点で重要であり、特に患者および家族が将来の計画を立てられることや、外部機関の支援を求めることが早期にできることが理由として挙げられる。加えて、今後認知症の療育法が発見され、利用可能となった場合、早期発見はより重要性を増すと考えられる [2]。よって、認知症の早

期検出, 特に認知症の非常に早い段階を見つけることは必要性の高いテーマになっている [3]. 現在, 認知症を検出するためのツールとして, 臨床心理士などが実施する各種神経心理検査に勝る, 検出力の高い信頼性のあるものは十分に開発されていないと言える. また, 神経心理検査は, 患者にとって大変不安であり, ストレスを生じるものである. 特に早期認知症検出においては, 補足的な検査である血液検査や脳画像検査なども要求される [4]. これらの検査は侵襲的であり, 患者への負担が強く不快感を与え, 相当の時間とコストを要する. これらのことより, 非侵襲な低コストでの早期認知症を検出する方法が強く求められている.

これまでに, 非侵襲および低コストな手法として, 音声および言語の情報から認知症の早期発見を目指した手法がいくつか提案されている [5], [6], [7]. これらの手法においては, ほとんどが対話的なデータを用いたものでなく, 写真の叙述, あるいは神経心理検査中の一方的な発話を分析したものである. 一方で, 人間の行動情報をコンピュータのアバターを用いて対話的に検出する手法が近年提案されてきている [8].

本論文では, 早期認知症を検出するための新たなアプローチを提案する. 我々は音声対話システムのアバターを利用し, アバターからの対話的な質問と患者の応答から早期認知症を検出する手法の実現を目指す. アバターの質問に関しては, 既存のミニメンタルステート検査 (MMSE) [12] などの各種神経心理検査を参考としている.

2. 関連研究

幾つかの関連研究では, 音声および言語情報から認知症の早期検出および診断を目指し, その有効性が示されている. 以下にそれらをまとめる.

Orimaye ら [9] は認知症の一種であるアルツハイマー病の診断ツール開発を目指し, DementiaBank データセット *1 の音声書き起こしから, 多数の言語特徴を抽出した. 計 242 サンプル (アルツハイマー病および健康人) を用いて, 5 つの異なる機械学習のアルゴリズムを使用し, 結果としてサポートベクターマシン (SVM) により 2 グループを 74% の正解率で正しく分類することができたと報告している. Konig ら [10] は 4 種類の検査中のデータを収録し, 健康人, 軽度認知障害 (MCI), アルツハイマー病, に被験者をグループ分けした. 音声特徴を抽出し, 健康人と MCI で 79% の分類正解率, 健康人とアルツハイマー病で 89% の分類正解率を得ている. 近年の最も高い検出率を持つ研究として, Fraser ら [6] は DementiaBank データセットの書き起こし文と音声から音声言語特徴を抽出し, 92% の正解率で健康人とアルツハイマー病を自動識別することを達成

した.

我々は音声対話システムのアバターを用い, アバターからの対話的な質問と患者の応答から早期認知症を検出する手法を提案する.

3. 音声対話システム

我々は, 音声対話システムのアバターとして MMDA-agent*2 を使用した. 本システムは, 日本語の音声認識, 対話管理, 音声合成および行動生成を含んでいる. MMDAgent は Windows PC で動作する. 我々は高齢者 (ユーザ) が理解しやすい様に, アバターの発話速度を下げ, 発話内容に字幕を付与した. 親しみを持って対話することが可能になると考えたため, リアルなキャラクターよりも, アバターはアニメの女性キャラクターを使用した. システムが発話を認識した数秒後, アバターは頷きの行動を行う. 頷きの動作に関しては MikuMikudance を用いて生成した *3.

我々は, 早期認知症を検出するため 6 種類の対話の課題を用意した. 以下にそれぞれの概略についてまとめる (図 1 も参照).

- (a) 自己紹介: アバターが自己紹介し, ユーザの名前と年齢を尋ねる. 自己紹介はユーザが対話を緊張感なく行えるために導入した.
- (b) 視線計測: システムが PC ディスプレイに小さな動く黒い点を表示する. ユーザはその動く黒い点を目で追うように指示される.
- (c) 読み課題: システムが画面にウェクスラーの記憶検査 [7], [11] 中の一文を表示する. ユーザはその文を声に出して読む.
- (d) 固定質問: MMSE などを参考にし, アバターが 3 種類の固定質問を尋ねる. 質問は, 1) 今日は何月何日ですか, 2) これまでで楽しかった思い出について話してください, 3) 今日はどうやってここまで来ましたか, で構成されている.
- (e) ランダム質問: 計 13 問の質問セットから, ランダムに 5 質問を選びアバターが尋ねる.
- (f) リテリング: システムがウェクスラーの記憶検査中の一文を読み, その後ユーザに記憶した文を繰り返して言ってもらう.

アバターとの会話中において, システムはユーザの音声と動画像をそれぞれマイクと内蔵カメラで収録する. システムは, クローズド質問および簡単に答えられる質問に対しては 10 秒間, 難しい質問に対しては 15 秒間の待ち時間を検出し, 次の質問に自動的に進むようにしている. 大学院生が各課題に対して負荷度をアンケートにより答え, その結果より, 簡単な課題から順番に行うように順序を設定した. 全課題を終えるのに必要な時間は, 10-15 分程度で

*1 <https://talkbank.org/DementiaBank/>

*2 <http://www.mmdagent.jp/>

*3 <http://www.geocities.jp/higuchuu4/>

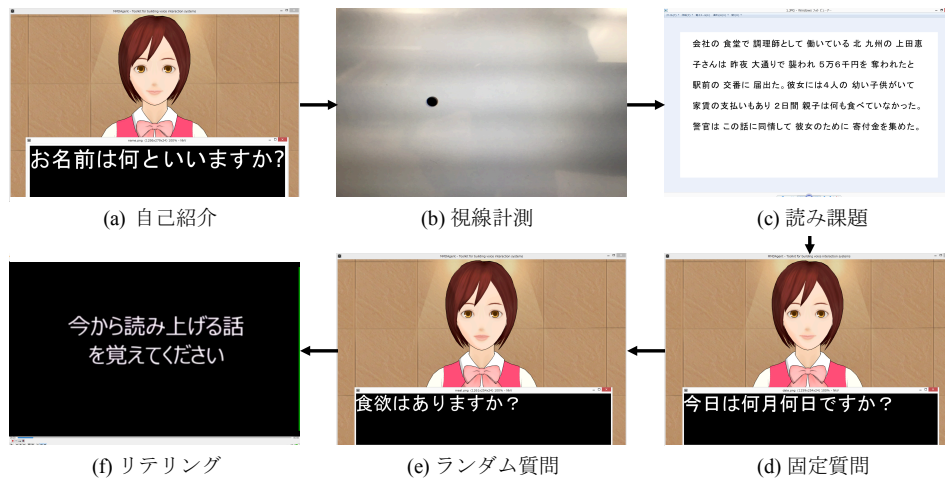


図 1 対話の課題例.

ある.

4. 実験的評価

本節では、システムを使った実験的評価について述べる.

4.1 被験者

我々は、計 20 名の被験者を募集した. 各被験者は、事前に十分な説明を受けた上で本実験の参加に同意した. この内、10 名は大阪大学附属病院における認知症グループとして、10 名は奈良先端科学技術大学院大学における健康人グループとして、それぞれ収録に参加した. 認知症グループは早期認知症であると既存の基準に従い、大阪大学附属病院の医師により診断を受けている. 我々は全被験者より年齢と MMSE スコアを取得した. 音声対話システムを使用するために、Surface Pro 3 を使用し、全被験者が課題を完了することができたことを確認した.

表 1 に示す通り、結果として我々は 18 名分のマルチモーダルデータを分析に使用した. ここで 2 名の被験者が分析から除外された. その理由として、1 名の認知症グループの被験者がまだ確定診断を受けていないこと、1 名の健康人グループ被験者の MMSE スコアが 22 点であり、カットオフ値 23 点を下回っていたためである. MMSE スコアは 2 グループで有意に異なっており、また年齢には有意差がないことを確認している.

表 1 被験者情報. 年齢と MMSE スコアの平均と標準偏差を記載している.

| グループ | N | 年齢 | MMSE スコア |
|------|---|------------|------------|
| 認知症 | 9 | 76.4 (7.4) | 21.4 (1.6) |
| 健康人 | 9 | 73.9 (4.5) | 26.4 (5.7) |

4.2 特徴量セット

収録データに対し、音声書き起こしを実施した. 音声、言語、対話、表情特徴量を 3 つの固定質問のユーザ応答から抽出し、全質問での平均値を算出した. 我々は、被験者に依存しない特徴量を先行研究に基づき 10 種類選択した [8].

4.2.1 音声特徴

音声特徴量抽出に関して、我々は Snack sound toolkit *4 を使用した. ここでは基本周波数、パワー、発話速度、声質に関する特徴を抽出した. 基本周波数に関しては、F0 の変動係数、パワーに関しては平均値、発話速度に関しては、発話形態素数/発話時間を算出した. 声質特徴量に関しては、第一倍音と第三フォルマントの振幅差を算出した.

4.2.2 言語特徴

我々は、Mecab*5 を用いて日本語の形態素解析を行った. Mecab の出力より、トークン (形態素) 数、フィルターの割合を得た. タイプトークン比 (TTR) についても [13] に基づき抽出した.

4.2.3 対話特徴

対話特徴量としてポーズを抽出し、これはアバターの質問終了から、ユーザの応答開始までの時間差である.

4.2.4 表情特徴

表情特徴量として、笑顔の割合を抽出した. Facetracker*6 を利用し表情特徴を取得した. 日本人女性の表情画像データベース JAFFE [14] から笑顔の表情特徴を SVM によりモデル化し、ユーザの応答中の動画像から笑顔と判定されたフレームの割合より、笑顔の割合を抽出した.

*4 <http://www.speech.kth.se/snack/>

*5 <http://taku910.github.io/mecab/>

*6 <https://github.com/kylemcDonald/FaceTracker>

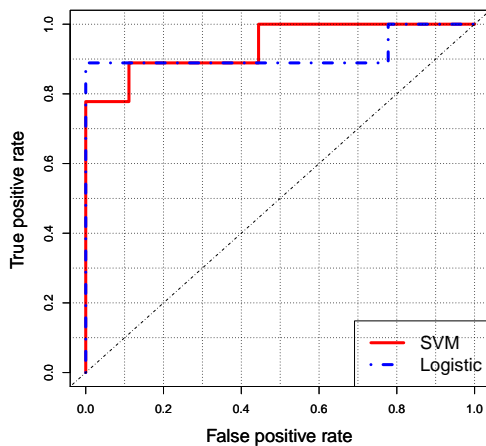


図 2 ROC 曲線. 赤色実線と青色破線は、SVM とロジスティック回帰をそれぞれ示している。

4.3 識別モデル

我々は、2種類の機械学習の識別モデルを使用し、特徴量から早期認知症を検出するアルゴリズムを作成した。識別モデルとして、シグモイドカーネルによる SVM とロジスティック回帰を使用した。識別モデルの入力として全ての特徴量を考慮し、認知症グループと健康人グループを分類するモデルを学習した。分類精度の評価として、Leave-one-participant-out 交差検証から ROC 曲線を描き、ROC 曲線より下の面積の割合 (AUC) を算出した [7]。

4.4 結果

図 2 は ROC 曲線を示しており、2つの識別モデルから、それぞれ 0.94 (SVM)、0.91 (ロジスティック回帰) の AUC を得た。これらの結果は関連研究と比較しても高い認知症の検出率であることが確認される [5], [6], [7]。

5. まとめ

本研究では、音声対話システムのアバターを用い、対話的な質問と患者の応答から早期認知症を検出する手法を提案した。計 18 名の早期認知症者と健康人により本システムの効果を実験的に検証し、結果として、AUC で 0.94 の検出精度を達成することができた。これより本システムが早期認知症を検出するツールとして有効であることが示された。今後はより大規模な実験を行い、音声認識を組み込んだシステムの実用化に向けて取り組む予定である。

謝辞 本研究は、Yanmar Innovation Lab. 2112 の支援を受けて行われた。

参考文献

[1] McKhann, G. M., Knopman, D. S., Chertkow, H., Hyman, B. T., Jack, C. R., Kawas, C. H., Klunk, W. E., Koroshetz, W. J., Manly, J. J., Mayeux, R. et al.: The diagnosis of dementia due to Alzheimer's disease: Recommendations from the National Institute on

Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease, *Alzheimer's & Dementia*, Vol. 7, No. 3, pp. 263–269 (2011).

[2] Santacruz, K. S. and Swagerty, D.: Early diagnosis of dementia, *Am Fam Physician*, Vol. 63, No. 4, pp. 703–713 (2001).

[3] Taler, V. and Phillips, N. A.: Language performance in Alzheimer's disease and mild cognitive impairment: a comparative review, *Journal of Clinical and Experimental Neuropsychology*, Vol. 30, No. 5, pp. 501–556 (2008).

[4] Laske, C., Sohrabi, H. R., Frost, S. M., López-de Ipiña, K., Garrard, P., Buscema, M., Dauwels, J., Soekadar, S. R., Mueller, S., Linnemann, C. et al.: Innovative diagnostic tools for early detection of Alzheimer's disease, *Alzheimer's & Dementia*, Vol. 11, No. 5, pp. 561–578 (2015).

[5] Lehr, M., Prud'hommeaux, E. T., Shafran, I. and Roark, B.: Fully Automated Neuropsychological Assessment for Detecting Mild Cognitive Impairment., *INTERSPEECH*, pp. 1039–1042 (2012).

[6] Fraser, K. C., Meltzer, J. A. and Rudzicz, F.: Linguistic Features Identify Alzheimer's Disease in Narrative Speech, *Journal of Alzheimer's Disease*, Vol. 49, No. 2, pp. 407–422 (2015).

[7] Roark, B., Mitchell, M., Hosom, J.-P., Hollingshead, K. and Kaye, J.: Spoken language derived measures for detecting mild cognitive impairment, *Audio, Speech, and Language Processing, IEEE Transactions on*, Vol. 19, No. 7, pp. 2081–2090 (2011).

[8] Tanaka, H., Sakti, S., Neubig, G., Toda, T., Negoro, H., Iwasaka, H. and Nakamura, S.: Automated Social Skills Trainer, *Proceedings of the 20th International Conference on Intelligent User Interfaces*, ACM, pp. 17–27 (2015).

[9] Orimaye, S. O., Wong, J. S.-M. and Golden, K. J.: Learning predictive linguistic features for Alzheimer's disease and related dementias using verbal utterances, *Proceedings of the 1st Workshop on Computational Linguistics and Clinical Psychology (CLPsych)*, pp. 78–87 (2014).

[10] König, A., Satt, A., Sorin, A., Hoory, R., Toledo-Ronen, O., Derreumaux, A., Manera, V., Verhey, F., Aalten, P., Robert, P. H. et al.: Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease, *Alzheimer's & Dementia: Diagnosis, Assessment & Disease Monitoring*, Vol. 1, No. 1, pp. 112–124 (2015).

[11] Wechsler, D.: *WAIS-III: Administration and scoring manual: Wechsler adult intelligence scale*, Psychological Corporation (1997).

[12] Folstein, M. F., Folstein, S. E. and McHugh, P. R.: "Mini-mental state": a practical method for grading the cognitive state of patients for the clinician, *Journal of psychiatric research*, Vol. 12, No. 3, pp. 189–198 (1975).

[13] Bucks, R., Singh, S., Cuerden, J. M. and Wilcock, G. K.: Analysis of spontaneous, conversational speech in dementia of Alzheimer type: Evaluation of an objective technique for analysing lexical performance, *Aphasiology*, Vol. 14, No. 1, pp. 71–91 (2000).

[14] Lyons, M., Akamatsu, S., Kamachi, M. and Gyoba, J.: Coding facial expressions with Gabor wavelets, *Proceedings of Third IEEE International Conference on Automatic Face and Gesture Recognition*, IEEE, pp. 200–205 (1998).