Interspeech 2018
2-6 September 2018, Hyderabad

Detection of dementia from responses to atypical questions asked by embodied conversational agents

Tsuyoki Ujiro[1], *Hiroki Tanaka*[1] *Hiroyoshi Adachi*[2] *Hiroaki Kazui*[3] *Manabu Ikeda*[4] *Takashi Kudo*[2] *Satoshi Nakamura*[1]

[1]Graduate School of Information Science, Nara Institute of Science and Technology, Japan
[2]Health and Counseling Center, Osaka University, Japan
[3]Department of Neuropsychiatry, Kochi Medical School, Japan
[4]Graduate School of Medicine, Osaka University, Japan

ujiro.tsuyoki.uq5@is.naist.jp

Abstract

Detection of dementia requires examinations, such as blood tests and functional magnetic resonance imaging (fMRI), that can be very stressful for the patient. Previous studies proposed screenings for easy detection of dementia that utilized acoustic and language information derived from conversations between patients and medical staff. Although these studies demonstrated effectiveness in automatically detecting dementia, the tasks used were created based on neuropsychological tests. The effect of habituation on this limited variety of tasks might have a negative impact on routine dementia screening. We propose a method to detect dementia using responses to more atypical questions asked by embodied conversational agents. Through consultations with neuropsychologists, we created a total of 13 questions. The embodied conversational agent obtained answers to these questions from 24 participants (12 dementia and 12 non-dementia). We recorded their responses and extracted speech and language features. We classified the two groups (dementia/non-dementia) by a machine learning algorithm (support vector machines and logistic regression) using the extracted features. The results showed a 0.95 detection performance in the area under the curve of the receiver operating characteristic (AUROC). This result demonstrates that our system using atypical questions can detect dementia.

Index Terms: Dementia, atypical questions, speech feature, language feature, embodied conversational agents

1. Introduction

According to a survey by the Ministry of Health, Labour and Welfare, Japan's aging rate is 27.5% and the number of elderly is increasing. Japan's growing elderly population has led to an increase in the number of dementia patients and is considered a serious social issue. Dementia causes deterioration in memory, thinking, and behavior, thus significantly interfering with the ability to function at work or perform ordinary daily activities [1]. There are various types of dementia such as Alzheimer's disease (AD), normal pressure hydrocephalus (NPH), and dementia with Lewy Bodies.

Early diagnosis of dementia is obviously critical for patients and families to plan for the future and identify outside sources of assistance. In addition, as potentially useful and proven treatments become available, early diagnosis will become increasingly important [2].

It is challenging to detect dementia in its early stages [3]. Many medical institutions diagnose dementia based on the Diagnostic and Statistical Manual of Mental Disorders (DSM)-IV-TR [4] or DSM-V [5]. Early detection combines neuropsychological examinations, blood tests, and brain image scans such as functional magnetic resonance imaging (fMRI) [6]. These examinations are invasive, adding to the patient's psychological burden by inducing anxiety and stress. In addition, considerable time and cost are required to complete this series of tests.

Therefore, a non-invasive and easy detection method is required. Some methods for the non-invasive and easy detection of dementia have been proposed. These studies use speech and language information to detect or diagnose dementia [1], [7], [8]. Most analyze utterances about photographs spoken during neuropsychological examinations (e.g., in conversations with medical staff).

A detection method during interaction with agents has also been proposed [9]. Previous studies attempted to detect dementia using the patient's voice as well as language and image information derived from three questions taken from a neuropsychological examination. However, these exam questions might be remembered by elderly people. To detect early stage dementia, it is necessary to monitor the elderly over a long period.

On the other hand, it is well known that dementia is associated with memory disorders [5]. Studies on memory related to dementia have been conducted [10], [11], [12], [13] showing, for example, that patients with dementia are likely to remember long-term events, but forget more recent occurrences.

In this paper, we propose a new approach to detect dementia using embodied conversational agents and an atypical question set that addresses several degrees of memory.

2. Related Work

Some previous studies effectively used speech and language information to detect dementia, as summarized below.

Roark et al. [8] recorded responses during the Wechsler memory recall task [14] and classified the speech into non-dementia and mild cognitive impairment (MCI) using complexity of the syntactic tree and speech features. The results showed that using multiple complementary measures outperformed the single value of a neuropsychological examination, proving that linguistic and speech information are useful for early detection of dementia.

In addition, Aramaki et al. [7] examined the relationship between cognitive ability and language ability for suspected MCI. The Hierarchic Dementia Scale-Revised (HDS-R) [15] was used as a marker to identify people with MCI. They directed the participants to write and speak, and measured their language ability. Analysis of the writing showed no significant

 10.21437/Interspeech.2018-1514

difference between the non-dementia group and people with MCI. However, the speech analysis did show that the vocabulary size was greater in those with suspected MCI than the non-dementia participants.

Another prior study attempted to detect dementia through dialogues with a computer agent. Tanaka et al. [9] prepared questions by referring to the Mini-Mental State Examination (MMSE) [16] and other neuropsychological examinations in an attempt to detect dementia. Language features, speech features, and facial expressions were extracted from the responses to these questions to classify non-dementia and dementia. As a result, the area under the curve of the receiver operating characteristic (AUROC) was 0.94 when using the linear kernel in a support vector machine (SVM).

Bahman et al.'s study [17] also prepared questions in the same form as a questionnaire created by a psychiatrist to detect dementia from conversations with an agent. Using speech features and linguistic features, they classified between neurodegenerative dementia and functional memory disorder with about 91% accuracy, thus demonstrating the effectiveness of detecting dementia from responses to agent questions.

In this paper, we propose a method to detect dementia from the responses of elderly people to atypical questions using an agent system of spoken dialogue.

3. Data Collection

3.1. Embodied conversational agents

We used the agent system created by Tanaka et al. [9] for data acquisition using a regular laptop, the Microsoft Surface Pro 3. This system uses the MMDAgent[1] as an embodied conversational agent to decrease the agent's speech rate and subtitle the content of each utterance, so that the elderly (user) can easily understand it. In addition, six kinds of tasks were prepared to detect dementia (see Figure 1). The summary of these tasks is as follows:

(a) Self-introduction: The agent introduces herself and asks the user's name and age.

(b) Gaze tracking: Display a small black dot on the screen and move it. The user is instructed to follow the point.

(c) Reading task: Display sentences on the screen and instruct the user to read out loud. The text is derived from the WMS-R [14].

(d) Fixed questions: Ask the user three questions prepared based on a neuropsychological examination such as the MMSE [16].

(e) Random questions: Ask the user five randomly selected questions from 13 prepared questions.

(f) Retelling: The agent reads out a sentence used in the WMS-R [14] and directs the user to remember it. After reading aloud, the user is asked to recall the sentence.

3.2. Atypical question

Our focus is on the random questions provided by an agent system. We prepared 13 types of atypical questions rather than questions derived from a neuropsychological examination. We prepared these questions based on consultations with neuropsychologists. The prepared question set is shown in Table 1. The agent randomly chooses and asks the elderly participants five questions from these 13 questions. During the conversation with the agent, we recorded the user/participant's voice and video with a microphone and built-in camera. If the user/participant was silent for more than 15 seconds, the system automatically shifted to the next question.

[1] http://www.mmdagent.jp/

Table 1: *Prepared random question set.*

	Contents
Q1	Please tell me about your family.
Q2	Please tell me something that you feel is stressful in your life.
Q3	What is your hobby?
Q4	What is your favorite song?
Q5	Please tell me about Yujiro Ishihara.
Q6	Please tell me about Shigeo Nagashima.
Q7	Please tell me about Hibari Misora.
Q8	Who is Japan's prime minister?
Q9	What season is it now?
Q10	What year is it?
Q11	Are you left-handed or right-handed?
Q12	Do you sleep well?
Q13	How is your appetite?

Table 2: *Age of participants, MMSE score, and educational history (education in years) with mean (M) and standard deviation (SD) values.*

Group	N	Age	MMSE	Education
		M (SD)	M (SD)	M (SD)
Non-dementia	12	74.5 (4.3)	27.5 (1.8)	8.8 (2.6)
Dementia	12	75.9 (7.3)	21.2 (5.1)	13.9 (3.8)

4. Detecting Dementia from Responses to Atypical Questions

4.1. Participants

We recruited a total of 24 participants, whose demographics are shown in Table 2. Each participant agreed to participate in this experiment with sufficient explanation in advance. We obtained informed consent from all participants.

Twelve patients took part in the recording as the dementia group at the Osaka University Medical School Hospital and the remaining 12 served as the non-dementia group at the Nara Institute of Science and Technology. The detailed diagnosis of the dementia patients was nine Alzheimer's disease (AD), one normal pressure hydrocephalus (NPH), one MCI, and one AD+NPH. Per the existing criteria, the dementia group was diagnosed by the psychiatrists of the Osaka University Medical School affiliated hospital based on DSM-IV-TR [4]. We obtained the age, MMSE score, and educational history of all participants. To use the spoken dialogue agent, we confirmed that all participants completed the task.

4.2. Feature set

We transcribed the speech on the recorded data. From this, we extracted speech features and language features for the responses to five questions and calculated their mean value. Finally, we selected 21 features based on previous studies [9].

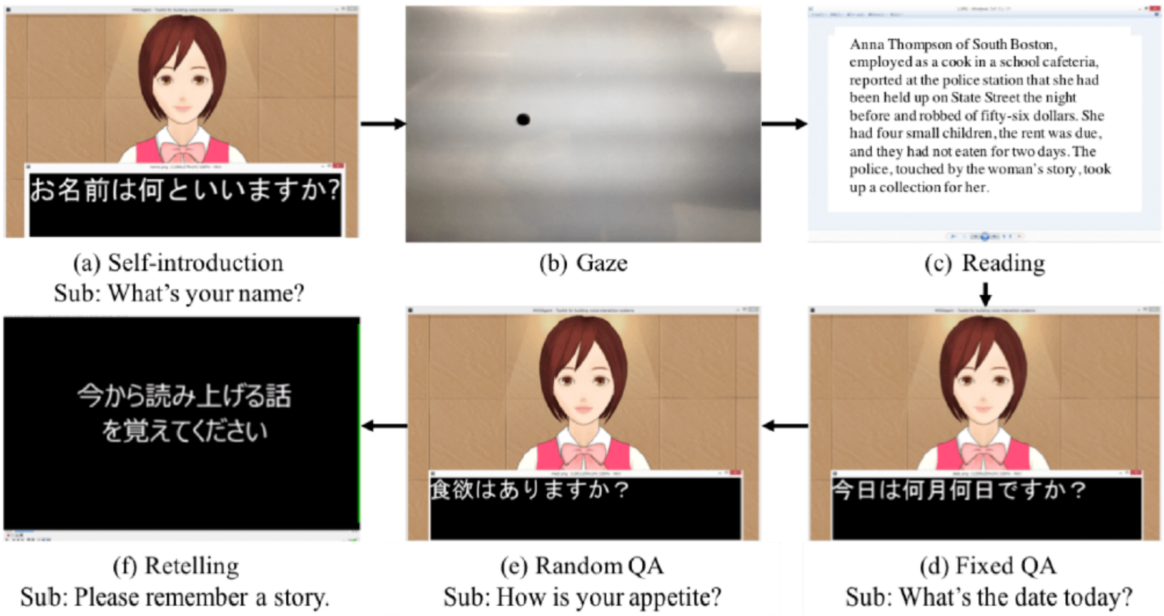

Figure 1: *Embodied conversational agents for data collection from elderly participants.*

4.2.1. Speech features

We used the Snack Sound Toolkit [2] to extract the speech features related to Gap, Pause, fundamental frequency (F0), power, and answer time. Gap is the time difference from the end of the agent's question to the start of the participant's response. Pause is the number of times which exceed one second between the participants' utterances. We also calculated the longest and mean values of Pause (longest value of Pause and mean value of Pause). Answer time is the total time of participant's responses.

For F0, we calculated the coefficient of variation, mean, maximum, median, minimum, standard deviation, and range values. For power, we calculated the mean and standard deviation values.

4.2.2. Language features

From the transcription, we performed a Japanese part of speech analysis using the MeCab toolkit [3]. Based on MeCab's output, we calculated speech rate and the number of tokens (Tokens), fillers (Fillers), nouns (Nouns), verbs (Verbs), adjectives (Adj.), and adverbs (Adv.).

4.3. Classification model

We constructed a classification model using two machine learning algorithms from the extracted features. We built the logistic regression and linear kernel support vector machine (SVM) as classifiers. We normalized each feature to mean value of zero and standard deviation value of one, and trained a model to classify the dementia and non-dementia groups. For evaluation of the classification model, based on Roark et al.'s [8] and Tanaka et al.'s [9] studies, we drew the ROC curves from leave-one-participant-out cross validation and calculated the area under the curve (AUC) of the ROC. Here, we used one participant for tuning the model's parameters and one for the test.

[2]http://www.speech.kth.se/snack/

[3]http://taku910.github.io/mecab/

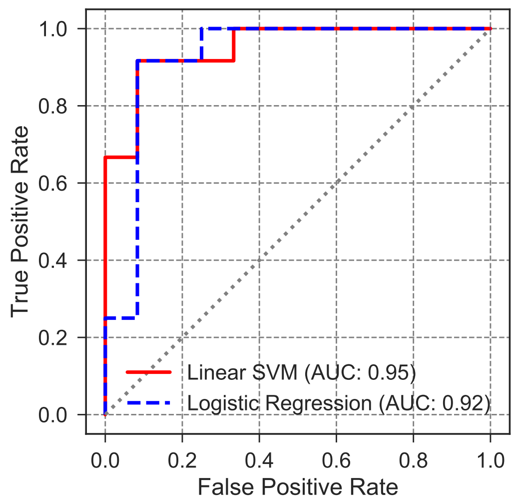

Figure 2: *ROC curve for the linear SVM and logistic regression. The red solid line is the ROC curve of the linear SVM. The blue dotted line shows the ROC curve of the logistic regression.*

4.4. Results

Figure 2 shows the ROC curve results of the two models. We obtained an AUC of 0.95 (the linear support vector machine) and 0.92 (the logistic regression), respectively. These results show that the detection performance is higher than the previous study [9]. We also performed an experiment with only AD patients, which means that we removed three patients from the 12 people with dementia, and obtained an AUC of 0.89.

For post-hoc analysis, we also examined the weight of each

Table 3: *AUC and unweighted accuracies of input variations.*

Input (model)	AUC	Unweighted accuracies
Gap only (SVM)	0.69	0.63
MMSE scores (SVM)	0.85	0.83
Speech features (SVM)	0.85	0.83
Speech features (Logistic)	0.90	0.83
All w/o MMSE (Logistic)	0.92	0.92
All w/o MMSE (SVM)	0.95	0.92

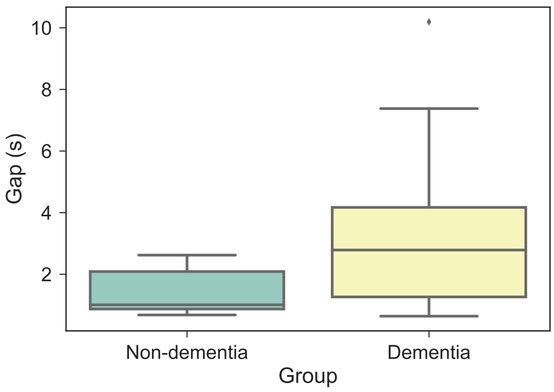

Figure 3: *Gap boxplots of non-dementia and dementia patients.*

feature in the logistic regression for each training set combination. For each training set, we ranked the features' absolute values and measured the sum values of the entire training set. The top five features were: 1) Gap, 2) F0 range, 3) F0 max, 4) mean value of Pause, and 5) Verbs. Gap had the highest feature weight in the logistic regression. Previous studies found Gap to be the most effective attribute for detecting early stage dementia [9]. Our results support these findings.

Table 3 shows the detection performance result on the AUROC curve. In this table, we calculated the case of only the Gap feature and MMSE scores in addition to the two models shown in Figure 2. We confirmed that our proposed method (especially for All w/o MMSE (SVM)) has the highest AUC and accuracy. In addition, we confirmed that detection performance of only speech features is higher than using MMSE scores.

Finally, we calculated the mean value of Gap as shown in Figure 3. Figure 3 shows that the patients with dementia have larger Gap values than the non-dementia participants. We calculated the p-value of the Mann-Whitney U test (two-tailed), and confirmed a significant difference between dementia and non-dementia (p-value = 0.04) as well as the Cohen's d (d = 0.98).

The Gap values for each question are shown in Figure 4. As evident in this figure, Gap tends to be larger in the dementia patients than the non-dementia participants. There is a particularly large difference for questions Q8 and Q9. These questions were (Q8) Who is the Prime minister? and (Q9) What season is it now? In general, disorientation such as forgetting a temporal event is said to indicate disorders associated with early stage dementia. This was reflected in Q9.

Closed-ended questions such as Q11, Q12, and Q13 show no difference between the two groups.

Questions Q5, Q6, and Q7 concern stories about past events. It appears that Gap differences do not occur in questions concerned with memory of the participants' younger lives.

Generally, memory disorder progresses in order of recent

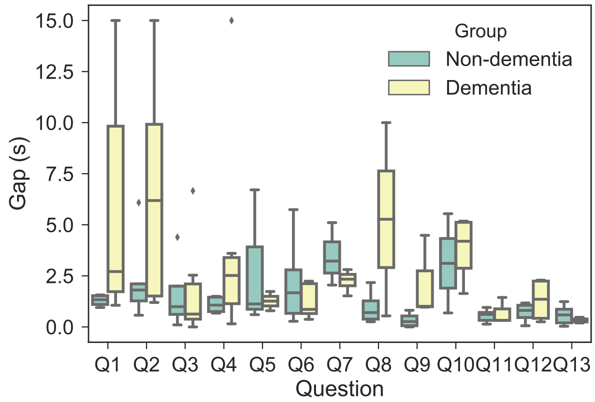

Figure 4: *Gap boxplots of non-dementia and dementia patients for each question.*

storage to immediate storage to remote storage [13]. It is commonly understood that the progress of dementia in its early stage has not advanced to disorders of remote memory. It therefore might be useful to measure the degree of dementia progress by examining the range of memory storage.

5. Conclusions

In this paper, we proposed a method to detect dementia from responses to atypical questions using embodied conversational agents. The effect of this method was experimentally tested with 24 dementia patients and non-dementia participants. We obtained a detection accuracy of 0.95 in the AUROC curve. We demonstrated that dementia can be detected even when using atypical questions.

We plan to use this system to automate early dementia detection. However, we need to address several problems in order to put this system to practical use. First, we need to automate the transcription process; we thus plan to use automatic speech recognition (ASR). Second, we did not include features that can be extracted from video. There are a few studies on detecting dementia using facial expressions and eye movements [18]. We will therefore consider extracting a variety of features from video. Third, we raised the possibility of measuring the degree of dementia from range of memory. To confirm this, we need to prepare a question set that considers varied degrees of event terms and measure the relationship between the degree of dementia and terms of the questions. Fourth, we need to consider effects of educational history, part of speech, and difference of five selected questions in the future.

6. Acknowledgements

Part of this work was supported by the Collaborative Research Project with Suntory Global Innovation Center Ltd. as well as JSPS KAKENHI Grant Numbers JP17H06101 and JP18K11437.

7. References

[1] G. M. McKhann, D. S. Knopman, H. Chertkow, B. T. Hyman, C. R. Jack, C. H. Kawas, W. E. Klunk, W. J. Koroshetz, J. J. Manly, and R. Mayeux, "The diagnosis of dementia due to alzheimerfs disease: Recommendations from the national institute on aging-alzheimer's association workgroups on diagnostic

guidelines for alzheimer's disease," *Alzheimer's & dementia: the journal of the Alzheimer's Association*, vol. 7, no. 3, pp. 263–269, 2011.

[2] K. S. Santacruz and D. Swagerty, "Early diagnosis of dementia," *Am Fam Physician*, vol. 63, no. 4, pp. 703–713, 2001.

[3] V. Taler and N. A. Phillips, "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] *Diagnostic and statistical manual of mental disorders (DSM-IV-TR)*. American Psychiatric Association, 2000.

[5] *Diagnostic and statistical manual of mental disorders (DSM-V)*. American Psychiatric Association, 2013.

[6] C. Laske, H. R. Sohrabi, S. M. Frost, K. López-de Ipiña, P. Garrard, M. Buscema, J. Dauwels, S. R. Soekadar, S. Mueller, C. Linnemann *et al.*, "Innovative diagnostic tools for early detection of alzheimer's disease," *Alzheimer's & dementia: the journal of the Alzheimer's Association*, vol. 11, no. 5, pp. 561–578, 2015.

[7] E. Aramaki, S. Shikata, M. Miyabe, and A. Kinoshita, "Vocabulary size in speech may be an early indicator of cognitive impairment," *PloS one*, vol. 11, no. 5, p. e0155195, 2016.

[8] B. Roark, M. Mitchell, J.-P. Hosom, K. Hollingshead, and J. Kaye, "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.

[9] H. Tanaka, H. Adachi, N. Ukita, M. Ikeda, H. Kazui, T. Kudo, and S. Nakamura, "Detecting dementia through interactive computer avatars," *IEEE journal of translational engineering in health and medicine*, vol. 5, pp. 1–11, 2017.

[10] L. R. Squire and S. Zola-Morgan, "Memory: brain systems and behavior," *Trends in neurosciences*, vol. 11, no. 4, pp. 170–175, 1988.

[11] B. J. Knowlton and L. R. Squire, "The learning of categories: Parallel brain systems for item memory and category knowledge," *Science*, vol. 262, no. 5140, pp. 1747–1749, 1993.

[12] L. R. Squire and S. M. Zola, "Structure and function of declarative and nondeclarative memory systems," *Proceedings of the National Academy of Sciences*, vol. 93, no. 24, pp. 13 515–13 522, 1996.

[13] C. E. Hou, B. L. Miller, and J. H. Kramer, "Patterns of autobiographical memory loss in dementia," *International Journal of Geriatric Psychiatry*, vol. 20, no. 9, pp. 809–815, 2005.

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

[15] T. Hosokawa, Y. Yamada, A. Isagoda, and R. Nakamura, "Psychometric equivalence of the hasegawa dementia scale-revised with the mini-mental state examination in stroke patients," *Perceptual and motor skills*, vol. 79, no. 1, pp. 664–666, 1994.

[16] M. F. Folstein, S. E. Folstein, and P. R. McHugh, "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.

[17] B. Mirheidari, D. Blackburn, K. Harkness, T. Walker, A. Venneri, M. Reuber, and H. Christensen, "An avatar-based system for identifying individuals likely to develop dementia," *Proc. Interspeech 2017*, pp. 3147–3151, 2017.

[18] T. Endo, N. Ukita, H. Tanaka, N. Hagita, S. Nakamura, H. Adachi, M. Ikeda, H. Kazui, and T. Kudo, "Initial response time measurement in eye movement for dementia screening test," *Proc Fifteenth IAPR International Conference on Machine Vision Applications (MVA)*, pp. 262–265, 2017.