Detecting Dementia from Face in Human-Agent Interaction

Hiroki Tanaka
NAIST, Japan
hiroki-tan@is.naist.jp

Hiroyoshi Adachi
Osaka University, Japan

Hiroaki Kazui
Kochi Medical School, Japan

Manabu Ikeda
Osaka University, Japan

Takashi Kudo
Osaka University, Japan

Satoshi Nakamura
NAIST, Japan

ABSTRACT

This paper proposes an approach to automatically detect dementia from a human face. Although some works have detected dementia from speech and language attributes, there are few studies focusing on facial expression in dementia patients. We recorded the human-agent interaction data of spoken dialogues from 24 participants (12 with dementia and 12 without) and extracted the face features. Our objective was to classify dementia by L1 regularized logistic regression. The facial features and the L1 logistic regression were then used to classify the participants into two groups with 0.82 detection performance, as measured by the areas under the receiver operating characteristic curve. We also identified various contributing features, such as action units, eye gaze, and lip activity. These results demonstrate that our system has the potential to detect dementia from the face through spoken dialog systems and as such, can be of assistance to health care workers.

ICMI '19 Adjunct, October 14–18, 2019, Suzhou, China

ACM ISBN 978-1-4503-6937-4/19/10…$15.00
https://doi.org/10.1145/3351529.3360654

CCS CONCEPTS

• **Applied computing** → *Health care information systems*;

KEYWORDS

Dementia, Alzheimer’s disease, facial expression, eye, agents

ACM Reference Format:

Hiroki Tanaka, Hiroyoshi Adachi, Hiroaki Kazui, Manabu Ikeda, Takashi Kudo, and Satoshi Nakamura. 2019. Detecting Dementia from Face in Human-Agent Interaction. In *Adjunct of the 2019 International Conference on Multimodal Interaction (ICMI ’19 Adjunct), October 14–18, 2019, Suzhou, China.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3351529.3360654

INTRODUCTION

Dementia is a neurodegenerative disorder that presents itself in various types (e.g., Alzheimer’s disease, primary progressive aphasia, normal pressure hydrocephalus, and dementia with Lewy bodies). Its early diagnosis is critical because this enables the patient and family to plan for the future and identify outside sources of assistance. As potentially useful and proven treatments become available, early diagnosis will become increasingly important. However, dementia’s early detection is challenging, especially in its most early stages [17]. In order to detect dementia, patients undergo a series of cognitive tests and assessments. In the case of early-stage detection, complementary tests include the analysis of samples of cerebrospinal fluid from the brain, a magnetic resonance brain imaging (MRI) test, and a blood test. These are relatively expensive and require a significant amount of time and effort. Thus, there is increasing need for additional cost-effective tools that enable the identification of people with dementia in preclinical or early clinical stages [8]. Speech and language cues extracted from conversation may be indicative of underlying cognitive processes (e.g., word retrieval, semantic difficulties, and attention deficits). These features are thus potentially useful for the detection of dementia [7, 9, 15, 16, 22]. There is also a study that suggests using gaze following delay to detect the early stage of dementia [6]. These multimodal findings could be incorporated into dementia detection through human-agent interaction [13, 18, 21].

However, few papers have examined the facial expressions of dementia. One such study investigated the facial expressions exhibited in people with Alzheimer’s disease [3] and found that some types of dementia were characterized by fewer facial expressions [4, 14]. Additionally, we need to consider the fact that saving appearance responses (e.g., smiling) are found in Alzheimer’s patients [11], specifically in the early stage of the disease [18]. In the present study, we examine the potential of multi-dimensional facial features for dementia detection. We describe our data collection in human-agent interaction along with algorithms for automatically detecting dementia.

Figure 1: Agent system for data collection from elderly participants. Sub: What's the date today? (Q1)

AGENT SYSTEM

As a computer avatar, we used an MMDAgent on a Windows application. This system is a Japanese spoken dialogue system that integrates speech recognition, dialogue management, text-to-speech, and behavior generation. The system was adapted for elderly users by adding subtitles, producing a lower pitch, and using a slower speaking rate. The development process was carried out in consultation with a professional psychiatrist. In this paper, we analyzed a part of the system where it asks three fixed queries [18] (see Figure 1): Q1) What's the date today?, Q2) Tell me something interesting about yourself, Q3) How did you come here today?

The system records videos and speech during the user's response to the questions. The frame rate of the video recording was 25. If the user was silent for more than 15 seconds, the system automatically shifted to the next question.

FACIAL AND EYE FEATURE EXTRACTION

For the facial feature extraction, we used video clips between the end of an agent's question and the beginning of its next question for each of the three questions. We first computed the facial attributes of 2D facial landmarks (mouth region: numbers 48 to 67), face pose (x, y, and z), gaze angles (x and y) [2], and intensity of action units (AUs) (specifically: 1, 2, 4, 5, 6, 7, 9, 10, 12, 14, 15, 17, 20, 23, 25, 26, 28, and 45), based on Openface [5]. We also obtained the user's response time to the agent's question [21] by lip activity (talking face). We utilized the lip activity detection proposed in [12], which used optical flow and entropy. We extended this approach to 2D facial landmark points around the mouth region. For each landmark detected in a video frame X_t, the entropy of the moving directions of pixels around the mouth was used to measure the lip movement. Between two mouth regions in two consecutive frames X_t and X_{t+1}, we computed the entropy of the pixel directions of the lip region as follows:

$$Entropy(X_t, X_{t+1}) = -\sum_i P(\alpha_i)\log(P(\alpha_i)), \tag{1}$$

where $P(\alpha_i)$ is the probability that a random pixel chosen from the lip region will have the direction α_i (radian). For each face detected in the segment X (being the sequence of N frames $(X_t, ..., X_{t+N-1})$), the lip activity measurement is calculated as

$$M_v(X) = \frac{1}{N-1}\sum_{i=1}^{N-1} Entropy(X_{t+i-1}, X_{t+i}). \tag{2}$$

Then, the decision of talking face is made by comparing $M_v(X)$ to a given threshold. The threshold in this study was set to 1.7, and we empirically determined N as 10. We extracted the response time as duration until first talking face. All features were split into before and after the response time (Pre- and Post-). Mean and standard deviation (SD) values of the features were then extracted.

CLASSIFYING DEMENTIA

12 dementia outpatients (mean age: 75.9 (SD: 7.3)) and 12 non-dementia (mean age: 74.5 (SD: 4.3)) were recorded. 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 patients were diagnosed by psychiatrists at a hospital affiliated with the Osaka University Medical School on the basis of DSM-IV-TR [1]. We obtained the age, the Mini-mental state examination (MMSE) score [20], and educational history of all participants. To use the agent system, we confirmed that all participants completed the task. Participant demographics are shown in Table 1.

Table 1: Age of participants, MMSE score, and educational history (education in years) with mean (M) and standard deviation (SD) values.

	N	Age	MMSE	Education
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)

We used a machine learning algorithm for differentiating dementia from non-dementia. L1 regularized logistic regression was used as a classifier. Because it is a binary classifier, it is well suited to the classification task of dementia vs. non-dementia. The feature values were normalized to the mean of 0 and SD of 1. We evaluated the classification performance with nested leave-one-participant-out cross-validation [21] and plotted the ROC curve with the areas under it (AUC). In the ROC curve, we plotted the true and false positive rates.

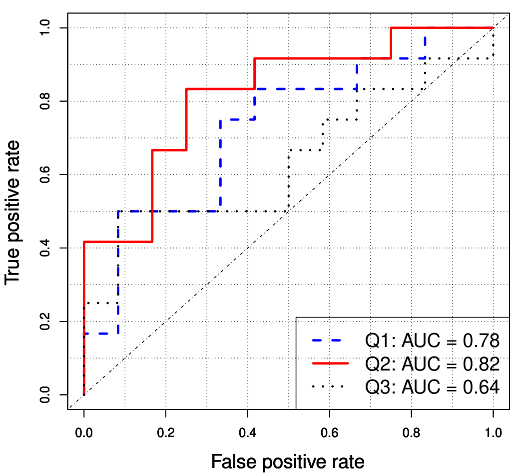

Figure 2: ROC curves for true and false positive rates. Blue, red, and black lines show questions 1, 2, and 3, respectively.

Figure 2 represents the ROC curves for three questions. An ROC curve along a diagonal shows a random classification model. Q2 outperformed the other questions (AUC: 0.82), and it shows a high detection performance using face. The feature from Q3 was difficult to classify dementia, which is consistent with the previous finding [18]. We accordingly analyzed feature weight. With regard to Q1 and Q2, the following features were highly weighted on all data in L1 regularized logistic regression (highest values in bold font, descending order): [*Q*1] **Post-AU09**, Pre-AU02, Pre-AU05, Post-AU17, Pre-AU09 (SD), Pre-AU12, Post-AU23 (SD), Post-gaze (x), Response time, [*Q*2] **Post-AU17 (SD)**, Pre-AU14 (SD), Post-AU45 (SD), Post-AU10 (SD), Pre-AU04 (SD), Pre-AU12 (SD), and Post-AU09.

CONCLUSION

Our study demonstrated that dementia can be correctly classified from face with 0.82 of the AUC. Because the algorithm utilized in this paper is not novel, other features, such as the features generated by deep models should be considered. In addition, the generalizability of this system to other mental diseases need to be explored. In future work, we will integrate this face attribute into multimodal detection and dialog systems [10] as well as implement personalized feedback [19] that informs medical staff or caregivers of the outcome.

ACKNOWLEDGEMENTS

Part of this study was supported by the Suntory Global Innovation Center.

REFERENCES

[1] [n. d.]. *Diagnostic and Statistical Manual of Mental Disorders, Fourth Edition: DSM-IV-TR*.

[2] Tagduda Ait Challal and Ouriel Grynszpan. 2018. What Gaze Tells Us About Personality. In *Proceedings of the 6th International Conference on Human-Agent Interaction (HAI '18)*. ACM, New York, NY, USA, 129–137. https://doi.org/10.1145/3284432.3284455

[3] K. Asplund, L. Jansson, and A. Norberg. 1995. Facial expressions of patients with dementia: a comparison of two methods of interpretation. *Int Psychogeriatr* 7, 4 (1995), 527–534.

[4] Kenneth Asplund, Astrid Norberg, Rolf Adolfsson, and Howard M Waxman. 1991. Facial expressions in severely demented patients—a stimulus–response study of four patients with dementia of the Alzheimer type. *International Journal of Geriatric Psychiatry* 6, 8 (1991), 599–606.

[5] Tadas Baltrusaitis, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. 2018. OpenFace 2.0: Facial Behavior Analysis Toolkit. *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)* (2018), 59–66.

[6] T. Endo, N. Ukita, H. Tanaka, N. Hagita, S. Nakamura, H. Adachi, M. Ikeda, H. Kazui, and T. Kudo. 2017. Initial response time measurement in eye movement for dementia screening test. In *2017 Fifteenth IAPR International Conference on Machine Vision Applications (MVA)*. 262–265. https://doi.org/10.23919/MVA.2017.7986851

[7] Kathleen C. Fraser, Jed A. Meltzer, and Frank Rudzicz. 2015. Linguistic Features Identify Alzheimer's Disease in Narrative Speech. *Journal of Alzheimer's disease : JAD* 49 2 (2015), 407–22.

[8] D. S. Geldmacher. 2002. Cost-effective recognition and diagnosis of dementia. *Semin Neurol* 22, 1 (Mar 2002), 63–70.

[9] S. Kato, H. Endo, A. Homma, T. Sakuma, and K. Watanabe. 2013. Early detection of cognitive impairment in the elderly based on Bayesian mining using speech prosody and cerebral blood flow activation. In *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. 5813–5816. https://doi.org/10.1109/EMBC.2013.6610873

[10] Yoshino Koichiro, Murase Yukitoshi, Lubis Nurul, Sugiyama Kyoshiro, Tanaka Hiroki, Sakti Sakriani, Takamichi Shinnosuke, and Nakamura Satoshi. 2019. Spoken Dialogue Robot for Watching Daily Life of Elderly People. In *International Workshop on Spoken Dialogue System Technology*.

[11] Masateru Matsushita, Yusuke Yatabe, Asuka Koyama, Akiko Katsuya, Daisuke Ijichi, Yusuke Miyagawa, Hiroto Ikezaki, Noboru Furukawa, Manabu Ikeda, and Mamoru Hashimoto. 2018. Are saving appearance responses typical communication patterns in Alzheimer's disease? *PLOS ONE* 13, 5 (05 2018), 1–9. https://doi.org/10.1371/journal.pone.0197468

[12] Bendris Meriem, Charlet Delphine, and Chollet Gerard. 2010. Lip Activity Detection for Talking Faces Classification in TV-Content. In *in 3rd International Conference on Machine Vision (ICMV)*. 187–190.

[13] B. Mirheidari, D. Blackburn, T. Walker, M. Reuber, and H. Christensen. 2019. Dementia detection using automatic analysis of conversations. *Computer Speech and Language* 53 (January 2019), 65–79. http://eprints.whiterose.ac.uk/136339/

[14] Ulrich Seidl, Ulrike Lueken, Philipp A. Thomann, Andreas Kruse, and Johannes Schroder. 2012. Facial Expression in Alzheimer's Disease: Impact of Cognitive Deficits and Neuropsychiatric Symptoms. *American Journal of Alzheimer's Disease & Other Dementias* 27, 2 (2012), 100–106. https://doi.org/10.1177/1533317512440495 arXiv:https://doi.org/10.1177/1533317512440495 PMID: 22495337.

[15] Daisaku Shibata, Kaoru Ito, Hiroyuki Nagai, Taro Okahisa, Ayae Kinoshita, and Eiji Aramaki. 2018. Idea density in Japanese for the early detection of dementia based on narrative speech. *PLOS ONE* 13, 12 (12 2018), 1–12. https://doi.org/10.1371/journal.pone.0208418

[16] Daisaku Shibata, Shoko Wakamiya, Kaoru Ito, Mai Miyabe, Ayae Kinoshita, and Eiji Aramaki. 2018. VocabChecker: Measuring Language Abilities for Detecting Early Stage Dementia. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces Companion (IUI '18 Companion)*. ACM, New York, NY, USA, Article 24, 2 pages.

https://doi.org/10.1145/3180308.3180332

[17] Vanessa Taler and Natalie A. Phillips. 2008. Language performance in Alzheimer's disease and mild cognitive impairment: A comparative review. *Journal of Clinical and Experimental Neuropsychology* 30, 5 (2008), 501–556. https://doi.org/10.1080/13803390701550128 arXiv:https://doi.org/10.1080/13803390701550128 PMID: 18569251.

[18] H. Tanaka, H. Adachi, N. Ukita, M. Ikeda, H. Kazui, T. Kudo, and S. Nakamura. 2017. Detecting Dementia Through Interactive Computer Avatars. *IEEE Journal of Translational Engineering in Health and Medicine* 5 (2017), 1–11. https://doi.org/10.1109/JTEHM.2017.2752152

[19] Hiroki Tanaka, Hideki Negoro, Hidemi Iwasaka, and Satoshi Nakamura. 2017. Embodied conversational agents for multimodal automated social skills training in people with autism spectrum disorders. *PLOS ONE* 12, 8 (08 2017), 1–15. https://doi.org/10.1371/journal.pone.0182151

[20] Tom N Tombaugh and Nancy J McIntyre. 1992. The mini-mental state examination: a comprehensive review. *Journal of the American Geriatrics Society* 40, 9 (1992), 922–935.

[21] Tsuyoki Ujiro, Hiroki Tanaka, Hiroyoshi Adachi, Hiroaki Kazui, Manabu Ikeda, Takashi Kudo, and Satoshi Nakamura. 2018. Detection of Dementia from Responses to Atypical Questions Asked by Embodied Conversational Agents. In *Proc. Interspeech 2018*. 1691–1695. https://doi.org/10.21437/Interspeech.2018-1514

[22] Jochen Weiner, Miguel Angrick, Srinivasan Umesh, and Tanja Schultz. 2018. Investigating the Effect of Audio Duration on Dementia Detection Using Acoustic Features. In *Proc. Interspeech 2018*. 2324–2328. https://doi.org/10.21437/Interspeech.2018-57