Identifying Dementia Patients based on Behavioral Markers in Human-Avatar Interaction

◎ Tsuyoki Ujiro, Hiroki Tanaka, Hiroyoshi Adachi,
Hiroaki Kazui, Manabu Ikeda, Takashi Kudo, Satoshi Nakamura

Early detection of dementia

- Dementia is one of the major causes of disability

Around 50 million people have dementia

- There is no treatment currently available to cure dementia or to alter its progressive course

Dementia detection tools

Neuropsychological tests

- MMSE [Folstein et al. 1975]
  One of the frequently used tests for detecting dementia
- WMS-R [Wechsler 1997]
  Designed to measure different memory functions

Medical experts are necessary for these tests
Patients need to go to hospital

▶▶▶ **Much easier tool is needed**

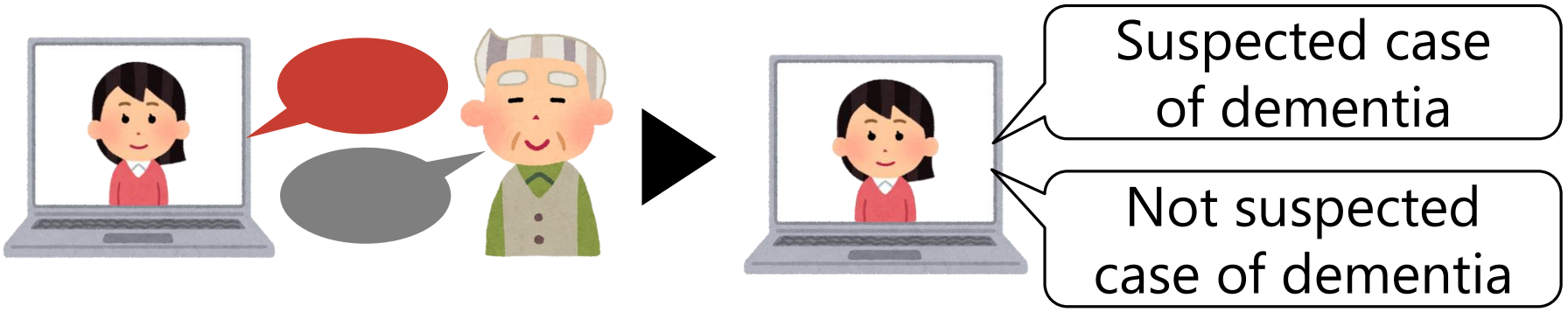
We focus on speech and language of dementia
[Roark et al. 2011], [Aramaki et al. 2016]

Purpose of our study

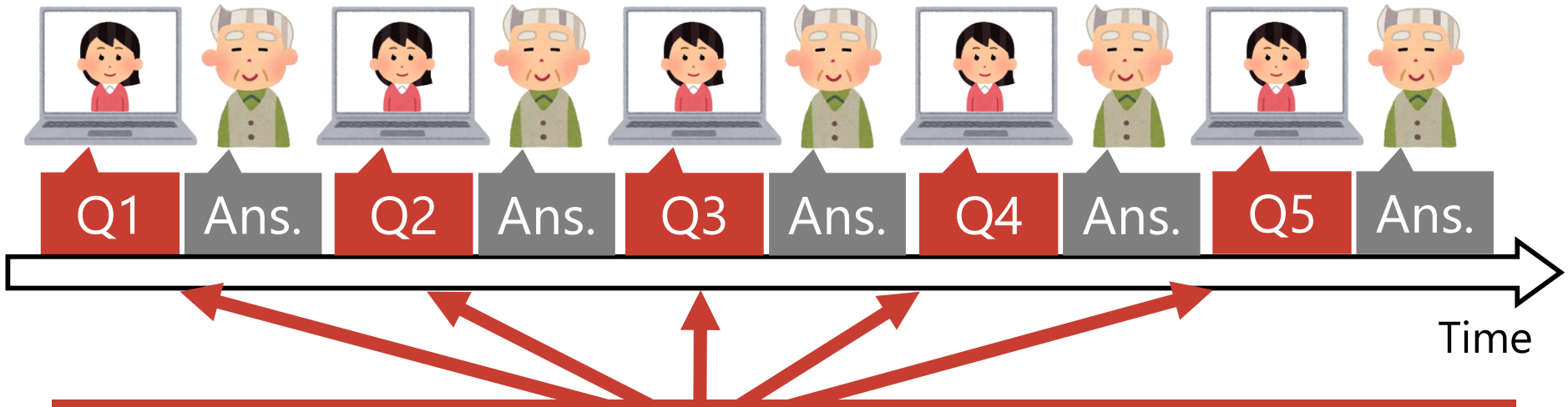
Purpose

- Avatar-based detection
- Language and speech features

Detection of dementia from Q&A using an avatar system



Q&A

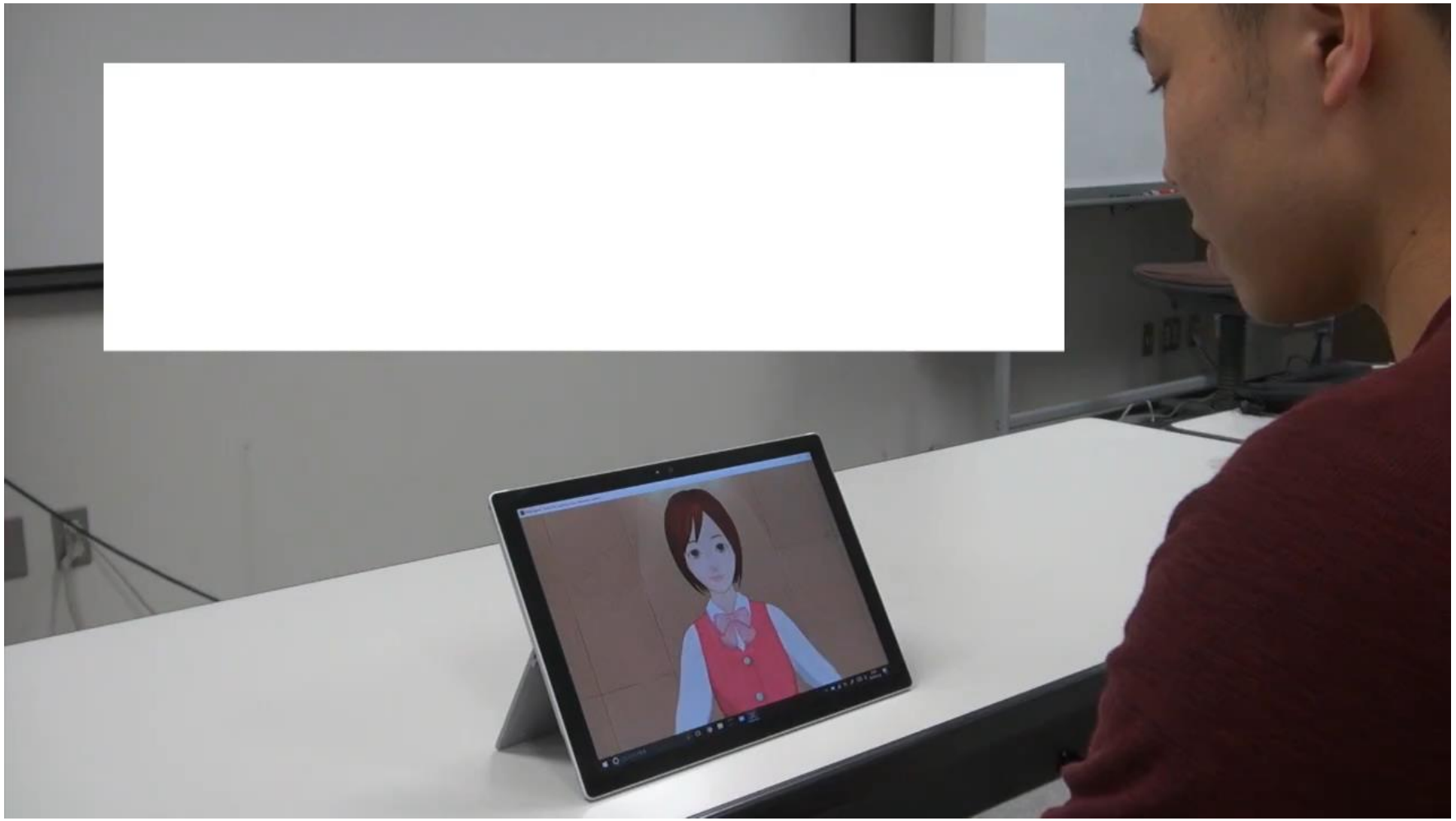


Question examples (A total of 13 questions)

- What's your hobby?
- Please tell me about Yujiro Ishihara.

The avatar asks randomly five questions from 13 questions

Demo video



Data collection

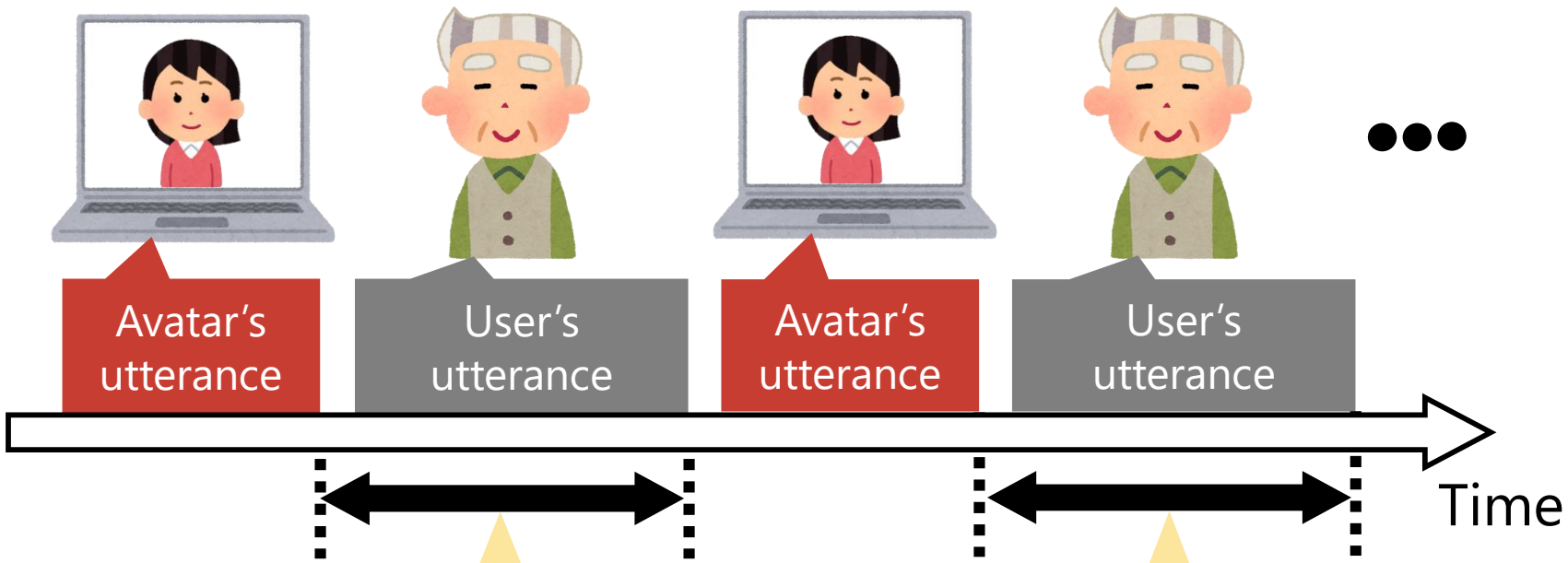
- Diagnosis is based on DSM-IV-TR

Group	N	Age	MMSE
		mean (SD)	mean (SD)
Non-dementia	12	74.5 (4.3)	27.5 (1.8)
Dementia	12	75.9 (7.6)	21.2 (5.1)

Dementia

- Alzheimer's disease (AD): 9
- Normal pressure hydrocephalus (NPH): 1
- Mild cognitive impairment (MCI): 1
- AD+NPH: 1

Analysis of the user's utterances

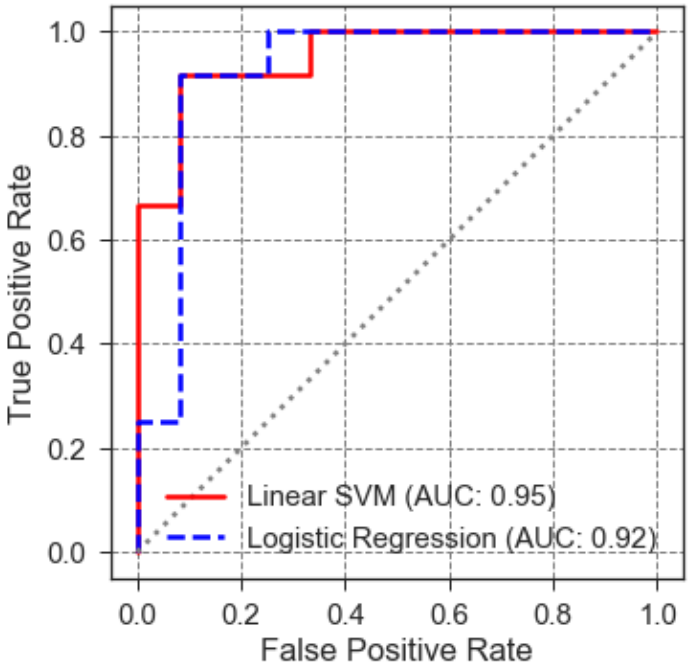


Feature extraction from five questions

- Speech features:
  Gap, Pause, Fundamental frequency (f0), Power, Answer time
- Language features:
  Speech Rate, Tokens, Fillers, Part of Speech

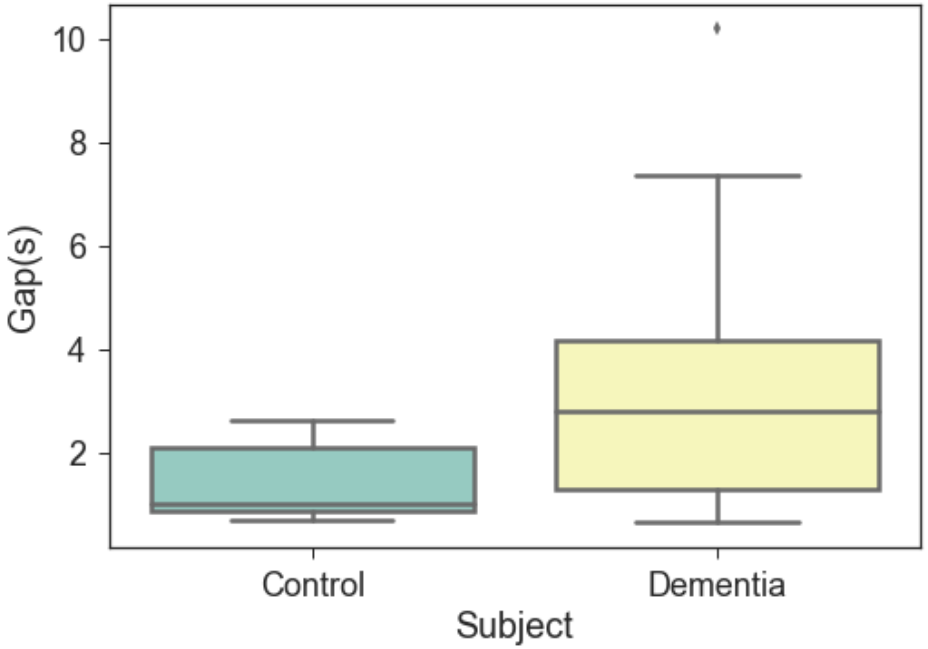
Detection performance

- Normalized the features
  - Mean: 0, SD: 1
- Classification model
  - Linear SVM
  - Logistic regression
- Model evaluation
  - Leave-one-participant-out
  - ROC curve



	Accuracy [%]
MMSE	83
Our system	92

“Gap” is the most important feature



- **The dementia group is slow to respond to questions**
- There was a significant difference between the two groups

  Mann-Whitney’s U test: $p < 0.05 \ \ (n = 24)$
  Effect size: $Cohen's\ d = 0.98$

Other important features

- Range of fundamental frequency (= pitch of voice)

  Non-dementia group < **Dementia group**

- Number of verb usage

  Non-dementia group < **Dementia group**

- Time between utterances

  Non-dementia group < **Dementia group**

Summary

Purpose	Easy detection tools of dementia ● Avatar-based detection ● Language and speech features
Proposed method	Analysis of multiple features used while asking questions by an avatar system
Evaluation	About 92% detection performance
Future work	Analysis of image features

Reference

- B. Roark, M. Mitchell, J. –P. Hosom, K. Hollingshead, and J. Kaye, "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.
- 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.
- 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, 2017.
- 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.
- D. Wechsler, WAIS-III, Wechsler adult intelligence scale: Administration and scoring manual. Psychological Corporation, 1997.

Easier tool: speech and language

- Previous studies show possibilities of detecting dementia from speech and language features in human-human interaction

[Roark et al. 2011], [Aramaki et al. 2016]

Limitation

- These studies were based on MMSE and WMS-R
- Same interaction patterns and questions
  - ▶ Not suitable for daily use

Feature extraction

Speech features

Answer time, **Gap**, **Power** (mean, SD),
Pause (count, mean time, max time),
fundamental frequency (f0)
(coefficient of variation, mean, max, median, min, SD, range)

Language features

Speech rate, **Tokens**, **Fillers**,
POS (noun, verb, adjective, adverb)