

A Multimodal Benchmark of Speech, Gaze, and Sketches for Detecting Alzheimer’s disease and related dementias

Leticia Pinto-Alva Leslie Moreno Gwen Bradforth
Riley Ashford Cecily Chung Maja Matarić Jesse Thomason
University of Southern California, Los Angeles, CA, USA
pintoalv@usc.edu

Abstract

Over six million Americans are living with Alzheimer’s disease or another form of dementia, and the rates are projected to increase to nearly 12 million by 2050. The development of accessible non-invasive screening tools is of paramount importance for individuals, family members, and caregivers. The majority of dementia detection studies primarily rely on datasets such as the Pitt Corpus and the ADReSS Challenge, which focus on audio and transcription data. We introduce a novel multimodal dataset that includes speech audio and transcripts, eye gaze, and pen pressure across several tasks used for standard dementia screening. Our data comprise 41 individuals, of whom 27 had normal cognitive function and 14 were diagnosed with Alzheimer’s Disease. By extracting features from our multimodal dataset, we facilitate the development of both unimodal and multimodal studies, enabling more comprehensive examinations of dementia presentation.

1 Introduction

Dementia, including Alzheimer’s Disease, is an umbrella term for loss of memory and other cognitive functions to a degree that significantly disrupts daily life. Alarming, one in three seniors in the US ultimately succumbs to Alzheimer’s Disease or another form of dementia ([Association](#)). In many instances, dementia is only detected in its advanced stages, often mistaken for typical age-related behavior. Early diagnosis has become crucial for enhancing treatment outcomes and providing more effective patient support. However, computer science researchers face substantial challenges when dealing with sensitive health information, with frequently outdated or insufficient patient data for constructing robust models.

We established a research collaboration with the University of Southern California (USC) Alzheimer Disease Research Center (ADRC) to

develop a multimodal dataset of speech, gaze, and pen pressure from 41 participants on three common dementia screening tasks. The participants comprise two distinct groups: those with dementia (AD) and those with normal cognitive function (NC). This dataset aims to assist computing research into detecting Alzheimer’s disease.

2 Dementia Screening Tasks

A variety of neuropsychological tests have been proposed for detecting dementia, including the Cookie Theft Picture Description Test, Yes/No Questions ([Kertesz, 2007](#)), Read Given Sentences, Word Fluency, and Name the Item That... as demonstrated by ([Becker et al., 1994a](#)). We considered three such tests: The Cookie Theft Picture, The Clock Drawing Test, and The Word Fluency Test.

The Cookie Theft Picture, originally designed for use with aphasic patients in the Boston Diagnostic Aphasia Examination ([Goodglass et al., 2001](#)), has proven to be a valuable tool for dementia detection as well ([Lira et al., 2014](#)). A key element of this task is the ability to retrieve appropriate lexical items. As explained by ([Nicholas and Brookshire, 1993](#)), one of the most important variables is information units, "words that are intelligible in context and accurately convey information relevant to eliciting the stimulus." Information units can be extracted from the image description.

The Clock Drawing Test ([Pinto and Peters, 2009](#)) is employed in clinical trials to screen for cognitive disorders and serves as a frequently used tool for detecting early signs of dementia in the elderly. This task includes the drawing of an analog clock from memory and indicating "10 minutes past 11 o’ clock".

The Word Fluency Test ([Becker et al., 1994b](#)) is a timed assessment in which participants are asked to name words of a particular category, such as animals or foods, within one minute. Next, the

task is made more difficult by requiring to name words within a category but only those that start with letters F or S.

3 Data Collection Protocol

Inspired by the importance of rich information in screening for dementia detection, we recorded user speech, eye gaze, and pen pressure during the tasks, as relevant.

We used a Surface Laptop Studio, Intel Core i7 with 32 GB RAM, 1TB SSD, NVIDIA GeForce RTX, the eyetracker is a Tobii Pro X3-120 and a Surface Slim Pen 2 to collect the data. We recorded participant speech using the computer camera and audio, and collected participant eye movement data through the Tobii Pro SDK, and recorded pen pressure using the InkCanvas class.

For each study participant, we first explained the study goals, tasks objectives, and obtained consent. Then we set up the computer and eye tracker. Next, we started recording the participant's speech, calibrated the eye tracker to their eyes and sitting position, and began the first screening task. The Cookie Theft picture was shown on the computer screen. The task took two minutes or until the participant said they were done. The participant was asked to describe the picture in as much detail as they could.

For the Clock Drawing Test, the participant is told to first draw an analog clock from memory with no specific time indicated on the clock. Next, they were asked to draw another clock, this time indicating "10 minutes past 11 o'clock". During this portion of the task, we recorded the pressure on the pen, the drawing coordinates over time, and the final drawings. The Clock Drawing task is not time-limited so the participants touch the finish button whenever they were done.

For the Word Fluency task, the participants were asked to name animals, foods, and words that start with the letters F or S. Each of these four subtasks had a one-minute duration, and participants were instructed to allow the full minute to elapse, even if they could not recall more words.

4 Dataset Overview and Next Steps

Using our system, we have completed our first round of data collection for the three dementia diagnosis tasks. The data consist of 41 participants, 24 self-declaring as female and 17 as male, all aged 50 or older, with the largest group being between

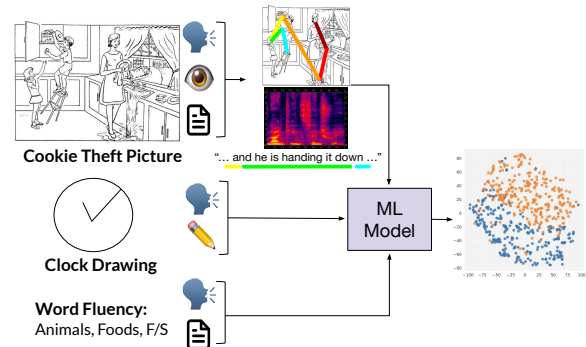


Figure 1: Our work considers three tasks: The Cookie Theft Picture (eye gaze, speech, and transcripts), Clock Drawing (Pencil information, speech), Word Fluency (Speech and transcripts). Features from different modalities will be extracted for downstream ML models.

75 and 79 years old. Of the 41 study participants, 27 are in the control group and 13 have a dementia diagnosis. Our sample includes individuals who self-declared as White (31), African American (4), White and Hispanic (2), White and Native American (1), and Asian descent (3).

Our next research steps include data annotation and modeling for the purpose of classification into Normal Cognitive (NC) and Dementia (AD) groups. Our work will include an extensive unimodal and multi-modal models, as well as minimum meaningful data per task necessary for screening, aiming to reduce the amount of time and tests so as to minimize the burden on patients and their caregivers and family members.

Acknowledgements

This study was made possible thanks to the generous support of our grant from the National Institutes of Health, 5-P30-AG-073105-02. We extend our gratitude to our Institutional Review Board (IRB) under the reference APP-22-06050 for their oversight and guidance throughout the study. We extend our deep appreciation to our collaborators at the USC Alzheimer Disease Research Center (ADRC), whose partnership enabled our research efforts. Our heartfelt thanks go to the participants who generously volunteered their time and made this work possible.

References

Alzheimer's Association. Alzheimer's and dementia. https://www.alz.org/alzheimer_s_dementia. [Accessed 20-10-2023].

- J. T. Becker, F. Boller, O. L. Lopez, J. Saxton, and K. L. McGonigle. 1994a. The natural history of alzheimer's disease: description of study cohort and accuracy of diagnosis. *Archives of Neurology*, 51(6):585–594. Grant Support: NIA AG03705 and AG05133.
- James T Becker, François Boiler, Oscar L Lopez, Judith Saxton, and Karen L McGonigle. 1994b. The natural history of alzheimer's disease: description of study cohort and accuracy of diagnosis. *Archives of neurology*, 51(6):585–594.
- Harold Goodglass, Edith Kaplan, and Sandra Weintraub. 2001. *BDAE: The Boston diagnostic aphasia examination*. Lippincott Williams & Wilkins Philadelphia, PA.
- Andrew Kertesz. 2007. Western aphasia battery—revised.
- Juliana Onofre de Lira, Thaís Soares Cianciarullo Minett, Paulo Henrique Ferreira Bertolucci, and Karin Zazo Ortiz. 2014. Analysis of word number and content in discourse of patients with mild to moderate alzheimer's disease. *Dementia & neuropsychologia*, 8:260–265.
- Linda E Nicholas and Robert H Brookshire. 1993. A system for quantifying the informativeness and efficiency of the connected speech of adults with aphasia. *Journal of Speech, Language, and Hearing Research*, 36(2):338–350.
- Elisabete Pinto and Ruth Peters. 2009. Literature review of the clock drawing test as a tool for cognitive screening. *Dementia and geriatric cognitive disorders*, 27(3):201–213.