Exploiting Longitudinal Speech Data via Voice Assistant Systems for Early Detection of Cognitive Decline

Kristin Qi[1], Youxiang Zhu[1], Caroline Summerour[2], John A. Batsis[2], Xiaohui Liang[1]
[1]University of Massachusetts Boston; [2]University of North Carolina Chapel Hill

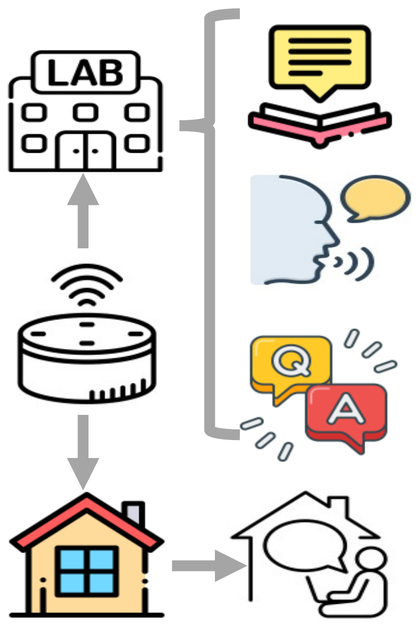
Introduction

Cognitive decline, particularly dementia, impacts an estimated 55 million people worldwide. Mild Cognitive Impairment (MCI) is an intermediate stage between normal cognition and dementia, and early detection of MCI is crucial. Speech impairments are a symptom of cognitive decline, and the analysis of spontaneous speech enables the development of cost-effective methods to assess cognitive decline.

- We employ voice assistant systems (VAS), such as Amazon Alexa, to passively collect speech data from the elderly.
- This approach allows us to study speech and language patterns for long-term monitoring of cognitive decline.

Collect Data via Voice Assistant Systems

Collect longitudinal speech session data at three-month intervals over **18 months** (2022–2024) using two approaches**: <u>In-lab and At-home sessions</u>**. The **in-lab** data is collected following specific protocols, while the **at-home** data consists of free speech.



Command-Reading Task: 90 participants (*30 healthy controls, 30 MCI, and 30* Alzheimer's) read a list of 34 commands designed for Alexa*, e.g. "What is the weather outside", "Alexa, remember my daughter's birthday is June first."*

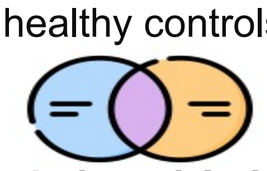
Command-Generation Task: 35 participants (*15 healthy controls and 20 MCI*) were provided 34 intent keywords, such as "*play music*", "*add a reminder*" and participants generate commands based on these intents.

Answer 18 Cognitive Questions: We selected questions from telephone-based cognitive assessment interviews, and the same 35 participants responded to questions via Alexa*, e.g. "Name words beginning with 'F' in one minute."*

Free Speech: 15 of the same 35 participants interacted with Alexa daily without any restrictions, *e.g. "Alexa, what time is it", "Alexa, play I heart radio."*

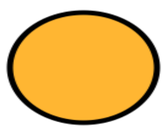
Feature-based Approach

The **<u>mixed-effect model</u>** aims to compute **fixed effects** on differences between MCI and healthy controls, and **random effects** to account for individual variability.



In-Lab and At-Home Shared Features

Acoustic Features	**Pre-trained:** ○ Wav2Vec2 ○ HuBERT ○ Whisper ○ WavLM **Handcrafted:** ○ Pauses, Pitch, Intensity ○ Speech Rate ○ eGeMAPS ○ ComParE-2016
Linguistic Features	**Pre-trained:** ○ BERT ○ RoBERTa ○ GPT-2 **Handcrafted:** ○ Part-Of-Speech ○ Syntactic Complexity ○ Vocabulary Richness ○ Semantic Similarity



At-Home Unique Features

Domain Features	18 Domain Categories: ○ reminders ○ device control ○ IoT, weather, etc.
Text Metric	Perplexity scores
Semantic Diversity	Variety in commands via embedding distances
Daily Routines	Usage peaks at what time each day
Sentiment Analysis	Jubilant commands (*e.g., "give jokes"*), imperative (*e.g., "stop"*), or neutral

Demographics

- **Age**
- **Gender**
- **Education**
- **Income**

Recursive Feature Elimination
Feature Importance (SHAP)

Behavioral Profile Embeddings

Mixed-effects model

Linear discriminant analysis (LDA) score

LDA score:

- ✓ Summarizes effects of features to classify groups
- ✓ Shows trajectories of score changes

LLM-based Approach

The **<u>qualitative approach</u>** employs large language models (LLM) to identify changes in Alexa daily command usage in these steps:

Alexa commands in a predefined time span (*e.g. quarterly*).

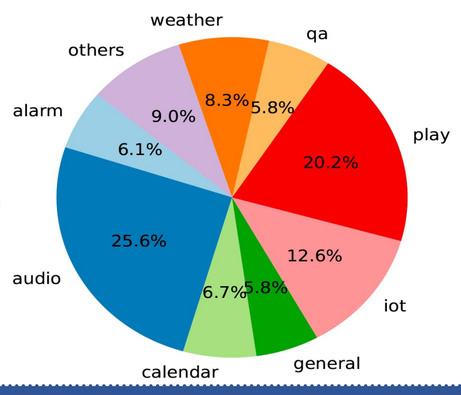
Use LLM to generate summaries based on LLM instructions *(e.g. summarize the daily activities).*

- ✓ **Metric** (Cosine pairwise similarity): Using a semantic similarity model such as the Sentence-BERT to calculate Sim(v_i, v_z), where v_i and v_z are any LLM-based summary.
- ✓ **Objective**: Identify the best instructions that maximize the distance between MCI and health control command usage by:
  $\max(\sum_{v_i \in MCI, v_z \in HC} \text{Sim}(v_i, v_z) - \sum_{v_i, v_z \in same\ label} \text{Sim}(v_i, v_z))$
- ✓ **Tools**: Prompt optimization methods (*e.g., DsPy*).
- ✓ **Expected results**: Discover optimal instructions that can expand the differences in command usage between MCI and healthy controls.

Domain Feature Results

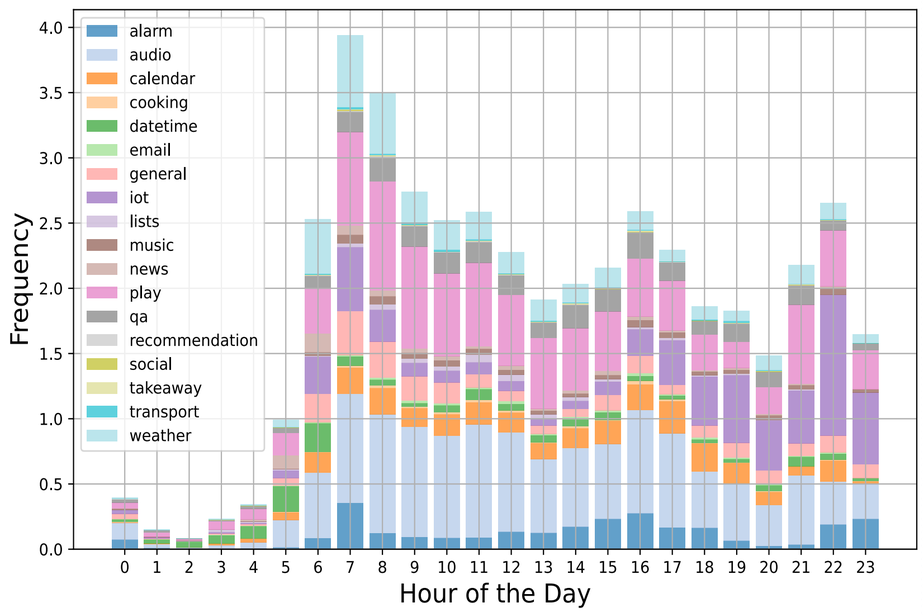
Distribution of domain usage:

Top-5 most used domains are audio (25.6%), play (20.2%) IoT (12.6%), weather (8.3%), and calendar (6.7%)



Command Usage Results

Hourly distribution of domain usage per week:



Check out the study page: https://cogvox.org/
Contact email: yanankristin.qi001@umb.edu

US National Institutes of Health National Institute on Aging R01AG067416

WiML @ NeurIPS 2024