

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



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

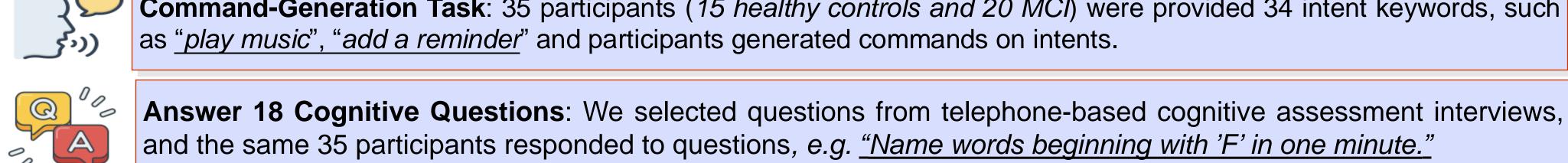
- Mild Cognitive Impairment (MCI) is an intermediate stage between normal cognition and dementia.
- Speech-based diagnosis of MCI is low-cost and non-invasive.
- We employ voice assistant systems (VAS), **Amazon Alexa**, to passively collect speech data from the elderly (Age > 65).
- We propose Cog-TiPRO, the LLM-driven framework to study voice commands for long-term monitoring of cognitive decline.

#### **Collect Data via Voice Assistant Systems**

Collect speech session data at three-month intervals over 18 months (2022–2024) with MoCA clinical test scores. Two approaches: In-lab vs. At-home sessions.



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. "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 generated commands on intents.



and the same 35 participants responded to questions, e.g. "Name words beginning with 'F' in one minute."

Free Speech (At Home): 15 of the same 35 participants interacted with Alexa daily without any restrictions,

## Detect Cognitive Decline via Cog-TiPRO Framework

#### 1. Multimodal Feature Extraction

- Speech features using the HuBERT model
- LLM-driven linguistic aspects (markers & command usage) from voice command transcripts

e.g. "Alexa, what time is it", "Alexa, play I heart radio."

#### 2. Time-series Transformer-based Iterative Prompt Refinement to Detect Cognitive Decline (Cog-TiPRO)

"Let's

think step

by step"

Initial prompt design:

$$P = \{P_{\text{content}}, P_{\text{instruction}}, P_{\text{exemplars}}, P_{\text{output step}}\}$$



Use current prompt to extract linguistic aspects



**Multimodal Feature Extraction** 

LLM | →

Transformer

HuBERT

Audio

Transcripts

Fusion

Acoustic

**Features** 

Linguistic

**Features** 

Detection

MCI

- Aggregate monthly command transcripts
- Fusion with acoustic features

#### Refiner **Iteratively refines prompts:**

- 3-5 iterations 1: Evaluation classification metrics with time-series transformer models
  - 2: Analyze errors (reflection)
  - 3: Give feedback based on errors to produce improved prompts
  - 4: Select the best-performing prompts as linguistic features

## 3. Temporal Modeling for Next **Cognitive Status Detection**

- PatchTST and iTransformer models

Transformer Healthy

Model patterns by using historical data

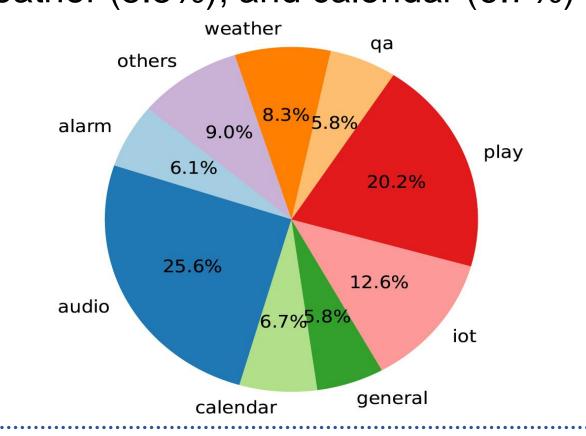
#### **Problem to Solve**

Limited understanding of linguistic aspects in voice commands **Use LLMs' summarization and knowledge** 

Data scarcity and noisy information **Distill MCI-relevant aspects** 

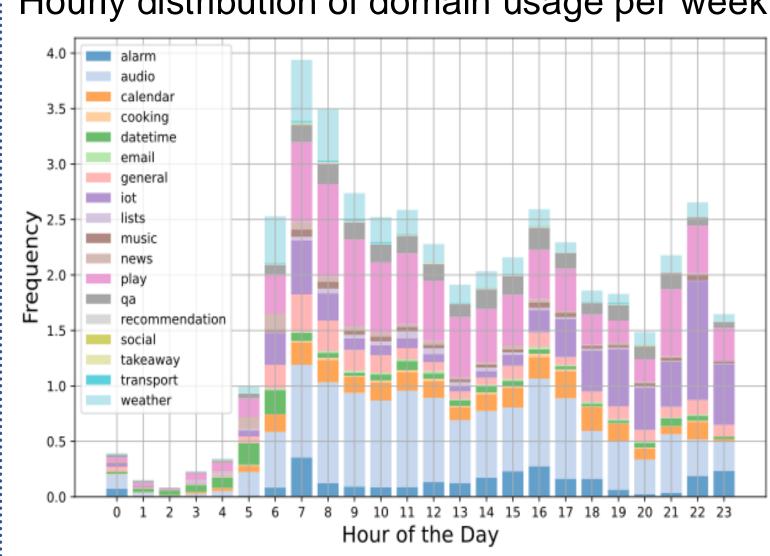
# What voice commands do the elderly like to use?

Top-5 domains used (Age > 65): audio (25.6%), play (20.2%), IoT (12.6%), weather (8.3%), and calendar (6.7%)



## When are voice commands used more often?

Hourly distribution of domain usage per week.



US National Institutes of Health National Institute on Aging R01AG067416