# Leveraging speech production as a physiological marker of cognitive decline: Demonstration of the role of timing and acoustic changes

Jennifer Sorinas[1], Minh Tam Davide Huynh[1], Kristin Hannesdottir[2], Jessica Robin[3], William Simpson[3], Neva Coello[4], Angelika Caputo[4], Marie-Emmanuelle Riviere[4] and Jelena Curcic[1]

1. Biomedical Research, Novartis, Basel, Switzerland
2. Biomedical Research, Novartis, Cambridge, MA, USA
3. Cambridge Cognition; Toronto, Canada
4. Novartis Pharma AG, Basel, Switzerland

## Background

- Changes to speech patterns have been identified as early signs of Alzheimer's Disease (AD) and have been shown to progress with disease[1,2,3,4]
- The Clinical Dementia Rating (CDR) scale is a dementia severity staging tool based on a semi-structured interview and is frequently used in clinical drug development. The CDR interview is often recorded for quality control purposes.
- The interaction between the rater and participant during the CDR interview may be used to derive speech-based measures with no additional site or participant burden

## Objective

- To leverage machine learning classifiers to examine whether different groups of speech features, derived from natural language processing algorithms, could accurately predict the transition from cognitively unimpaired (CU) to cognitively impaired (CI)

## Methods

- We analyzed audio recordings of the CDR from 85 English-speaking participants at risk of developing AD enrolled in the Alzheimer's Prevention Initiative Generation Program[5]
- Participants were a mean age of 68.7 (SD=4.6) and 66% female
- All participants were cognitively healthy at baseline according to established cut-offs on the Mini Mental State Examination (MMSE), Repeatable Battery for the Assessment of Neuropsychological Status (RBANS) and the CDR
- CDR recordings were processed using the Winterlight speech analysis platform (Figure 1). Samples were diarized and an array of objective speech features were extracted from the participant's speech
- The CDR can be divided into multiple subtasks. For this analysis we focused on two specific subtasks
    - The "recent experience" interview – this section of open-ended discussion between the rater and participant is the longest sample of continuous speech within the CDR interview. Here the rater (who has received details from an informant) asks the participant to recall details of a recent experience
    - The "address repeat" task – this is an example of constrained speech which a participant is required to encode a fictitious address and then repeat it back to the rater
- We hypothesized that linguistic characteristics of speech may carry greater signal in the recent experience task, while acoustic and timing characteristics may be more salient in the address repeat task
- We applied different machine learning strategies to examine which groups of features were predictive of the emergence of clinically significant cognitive impairment
- Model inputs were the longitudinal changes in speech features from baseline to each visit. Models were trained using a 5-fold nested cross-validation strategy with label stratification and subject-wise split
- Overall, 55.3% of the study population transitioned from CU to CI over the course of the study. The time to conversion varied considerably between participants (Figure 2)

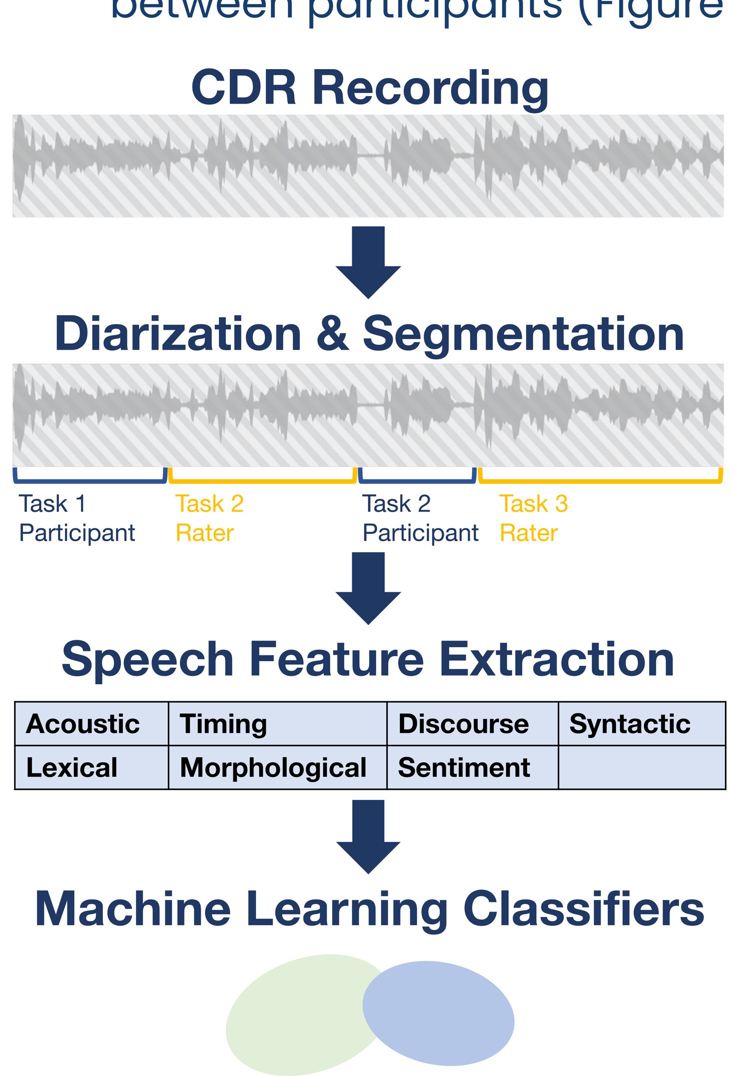



Figure 1: Overview of speech sample segmentation, processing and input into machine learning classifiers

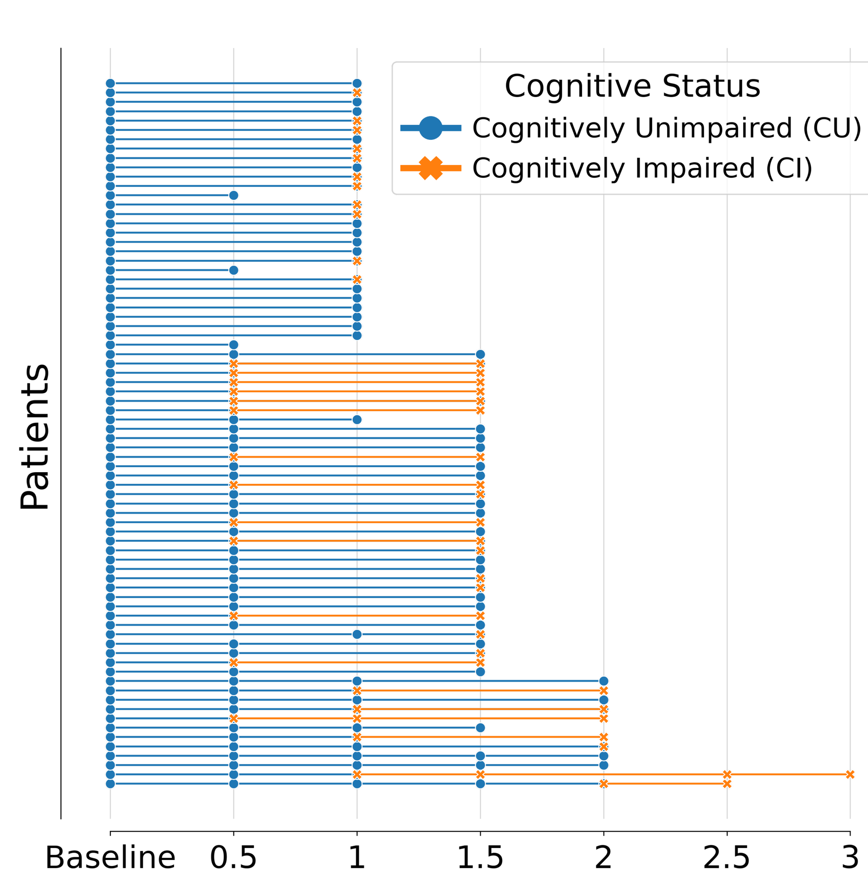
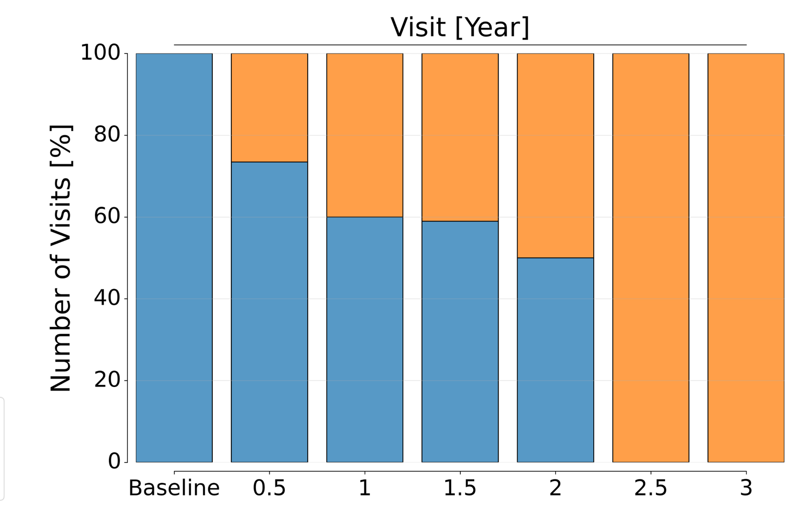
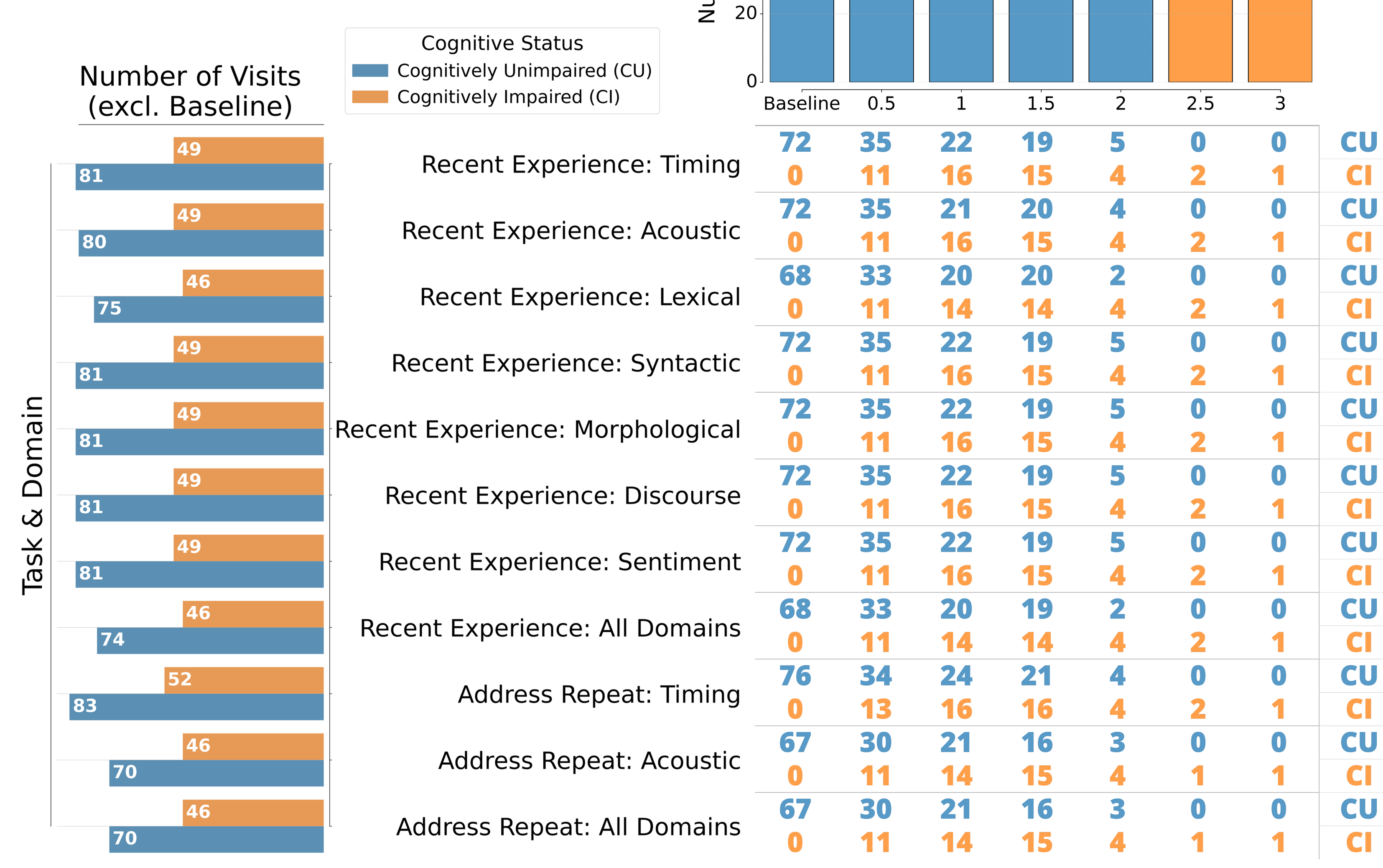



Figure 2: Depiction of CU vs. CI transitions for each participant over the observational period of the study

Figure 3: Distribution of CU and CI visits available for classifier training across tasks, domains, and timepoints. Top figure: distribution of visits across timepoints, stratified by cognitive status. Left figure: distribution of visits by task and domain. Center: number of visits per task/domain, timepoint, and cognitive status.




| Task | Domain | Model | AUC | Weighted F1 | Balanced Accuracy |
|---|---|---|---|---|---|
| Recent experience | Timing | RF | 0.52 (0.03) | 0.53 (0.05) | 0.52 (0.03) |
| | Acoustic | QDA | 0.51 (0.07) | 0.54 (0.08) | 0.51 (0.07) |
| | Lexical | RF | 0.49 (0.05) | 0.51 (0.06) | 0.49 (0.05) |
| | Syntactic | LDA | 0.49 (0.13) | 0.52 (0.12) | 0.49 (0.13) |
| | Morphological | QDA | 0.53 (0.04) | 0.56 (0.04) | 0.53 (0.04) |
| | Discourse | LDA | 0.57 (0.05) | 0.58 (0.08) | 0.57 (0.05) |
| | Sentiment | LDA | 0.55 (0.07) | 0.57 (0.07) | 0.55 (0.07) |
| Address repeat | Timing | RF | 0.59 (0.03) | 0.61 (0.03) | 0.59 (0.03) |
| | Acoustic | QDA | 0.57 (0.08) | 0.59 (0.07) | 0.57 (0.08) |
| Recent experience | All domains | RF | 0.50 (0.06) | 0.51 (0.08) | 0.50 (0.06) |
| Address repeat | All domains | QDA | 0.52 (0.07) | 0.52 (0.07) | 0.52 (0.07) |

Table 1: Classification metrics for the best performing classification models for each task-domain pairing. Reported are mean and standard deviation values from all cross-validation runs.
RF = Random Forest, QDA = Quadratic Discriminant Analysis, LDA = Linear Discriminant Analysis

## Results

- In total, 61.6% of participant visits were annotated as CU and 38.4% as CI (Figure 3)
- The performance of our evaluated machine learning classifiers was modest (Table 1). AUC values exceeded chance, but did not surpass 0.65
- Timing and acoustic features obtained from the structured address repeat task were most predictive
- Discourse and sentiment features from the free-speech recent experience task showed similar but slightly lower predictive power relative to the address repeat task

## Conclusions

- These results highlight that we have the necessary analytical tools to characterize the power of speech features as phenotypic characteristics of the transition from CU to CI
- Application of these machine learning classifiers is limited by the amount of CU vs CI training data per participant as well as distribution of CU and CI diagnoses per visit
- Future studies examining individual candidate features and theory driven composite scores are ongoing and will further clarify the role of speech biomarkers in early CI

## References

1. Robin, J., Xu, M., Kaufman, L. D. & Simpson, W. Using Digital Speech Assessments to Detect Early Signs of Cognitive Impairment. Front. Digit. Health 3, 749758 (2021).

2. Eyigoz, E., Mathur, S., Santamaria, M., Cecchi, G. & Naylor, M. Linguistic markers predict onset of Alzheimer's disease. EClinicalMedicine 28, 100583 (2020).

3. Amini, S. et al. Automated detection of mild cognitive impairment and dementia from voice recordings: A natural language processing approach. Alzheimers Dement. alz.12721 (2022) doi:10.1002/alz.12721.

4. König, A. et al. Automatic speech analysis for the assessment of patients with predementia and Alzheimer's disease. Alzheimers Dement. Diagn. Assess. Dis. Monit. 1, 112–124 (2015).

5. Lopez, Cristina Lopez, et al. The Alzheimer's Prevention Initiative Generation Program: study design of two randomized controlled trials for individuals at risk for clinical onset of Alzheimer's disease. Alzheimer's & Dementia: Translational Research & Clinical Interventions 5 (2019): 216-227.

NOVARTIS
CAMBRIDGE COGNITION