Detecting level of impairment in dementia using automatically calculated discourse and contextual features of connected speech

NEW YORK UNIVERSITY

IBM Research

Graham Flick[1,2] & Rachel Ostrand[2]

1. Department of Psychology, New York University; 2. IBM Research

graham.flick@nyu.edu

Introduction

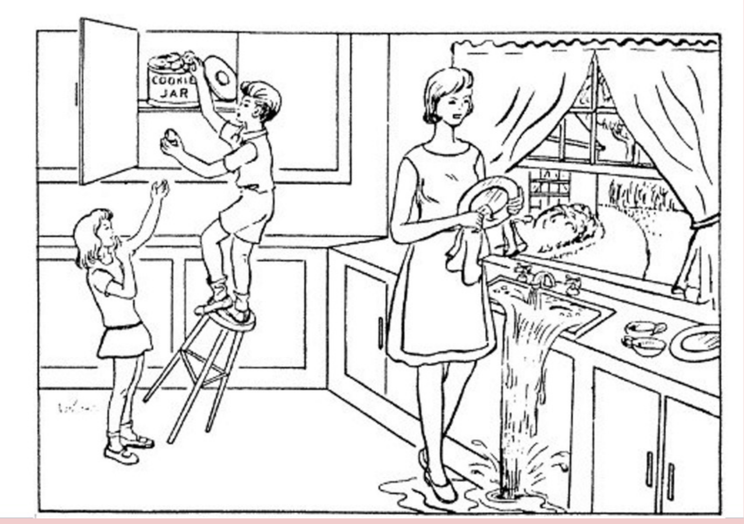
- Existing screening methods for Alzheimer's Disease and other forms are Dementia are costly, require substantial expertise, and may not be sensitive to mild changes in cognitive decline.
- We have demonstrated an alternative in past work, using features of connected speech to predict impairment (Ostrand & Gunstad, 2021, *Journal of Geriatric Psychology & Neurology*).
- However, most prior work on this topic has focused on using simple lexical features of speech, such as part of speech counts, to predict level of cognitive impairment, without taking the larger language context into account.

Research Aims

- **Goal:** Capture more holistic effects of cognitive impairment.
- **Method:** Compute linguistic features that capture sentential or discourse context properties of speech, and measure how well they predict degree of cognitive impairment.
- Linguistic features we investigated:
  - **How surprising words are, given the contexts that they appear in (computed from GPT-2, a large computational language model)**
  - **Where and how filler words (e.g., um) are used in speech**

Dataset & Analysis

- Transcribed picture descriptions from the *Cookie Theft* task from the DementiaBank corpus (Becker et al., 1994, *Arch Neurol*).
- N = 211 transcripts containing at least 50 words were submitted to a Python-based automatic feature extraction pipeline.
- Demographic data (age, sex, years of education) were used to construct baseline models.
- Outcome measure: Mini-Mental State Examination (MMSE) score, used to capture degree of cognitive impairment.



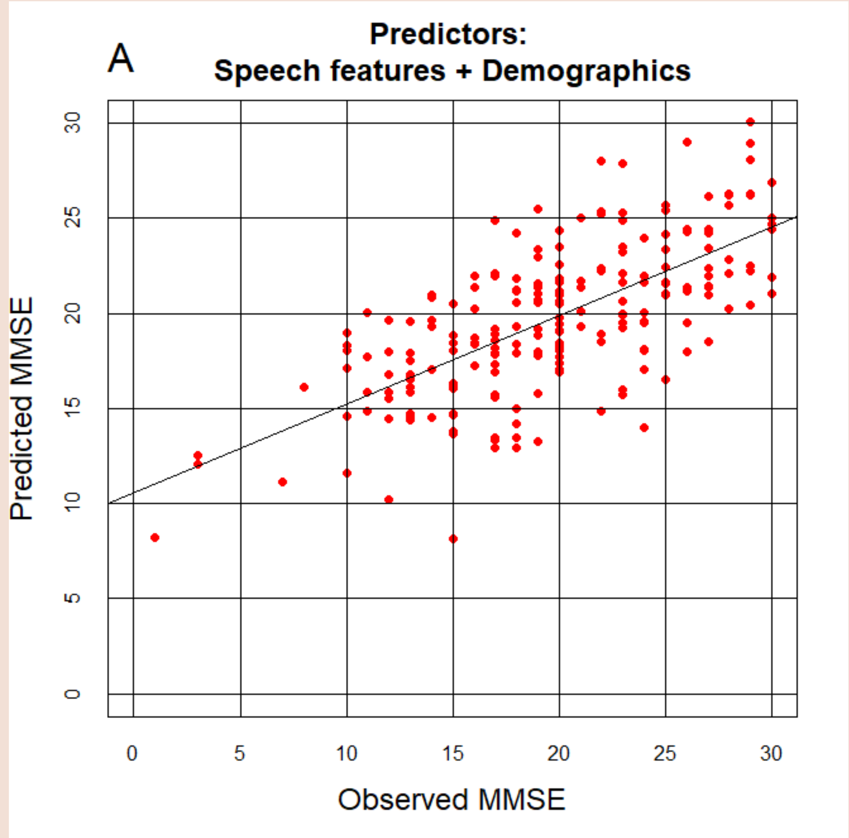
Automatically Extracted Speech Features

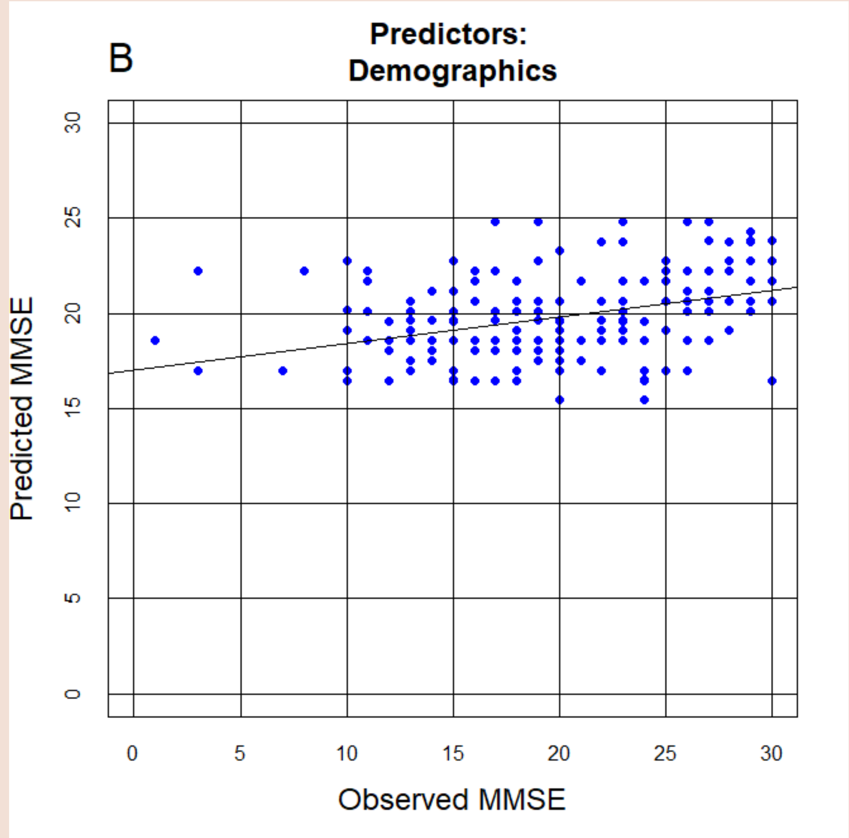
- **Lexical features (12):** Total number of words; proportion of: filler words, empty words, definite articles, indefinite articles, pronouns, nouns, verbs, and content words; median lexical frequency, type-token ratio, and Honoré's statistic (a measure of lexical richness).
- **Context-sensitive features (5):**
- **Linguistic surprisal** (estimated using the GPT-2 large language model): how predictable is each word, given the context of the prior 12 words that were spoken. This is a measure of how surprising/unexpected a particular word is within its discourse.
  - Median, Interquartile range of surprisal of all the words in the transcript
  - Surprisal of words immediately following fillers (*I um walked*; surprisal of *walked*)
  - Distance to the next content word after a filler (*I went um to the store = 3*)
  - Frequency of the next content word following a filler (*I um walked*: frequency of *walked*)
- **Multiple linear regression models** were built with all speech predictors entered jointly, with or without demographic variables, and model comparison was used to assess whether the speech features added explanatory power when predicting MMSE scores.

Results: Model Comparison

The full model, which used all speech and demographic features as predictors (**Fig. A**):

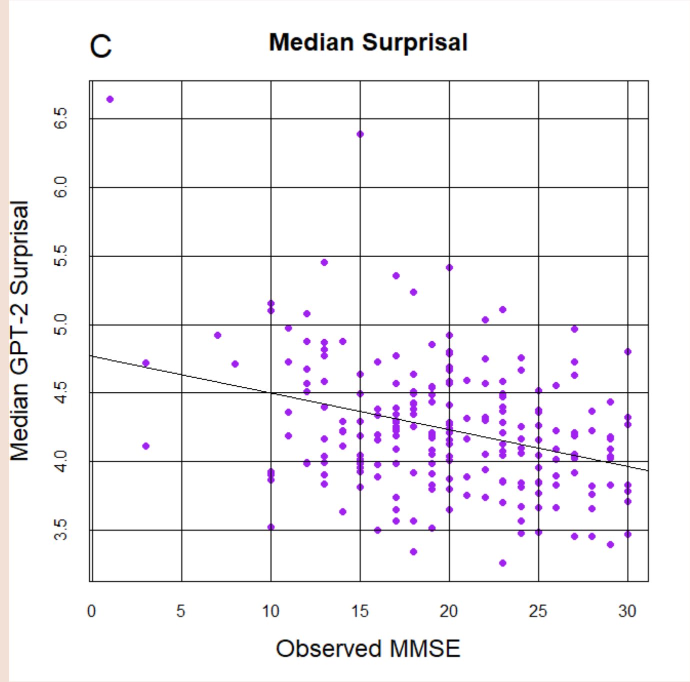
- significantly predicted MMSE scores (Adjusted $R^2 = 0.41$, $F_{20,193} = 8.39$, $p < 0.001$)
- explained significantly more variance than did **demographic variables alone** ($F_{17,190} = 6.89$, $p < 0.001$; **Fig. B**)
- was significantly improved when adding surprisal features (median & IQR), but not filler features, to the set of lexical predictors ($F_{5,190} = 5.619$, $p < 0.001$)

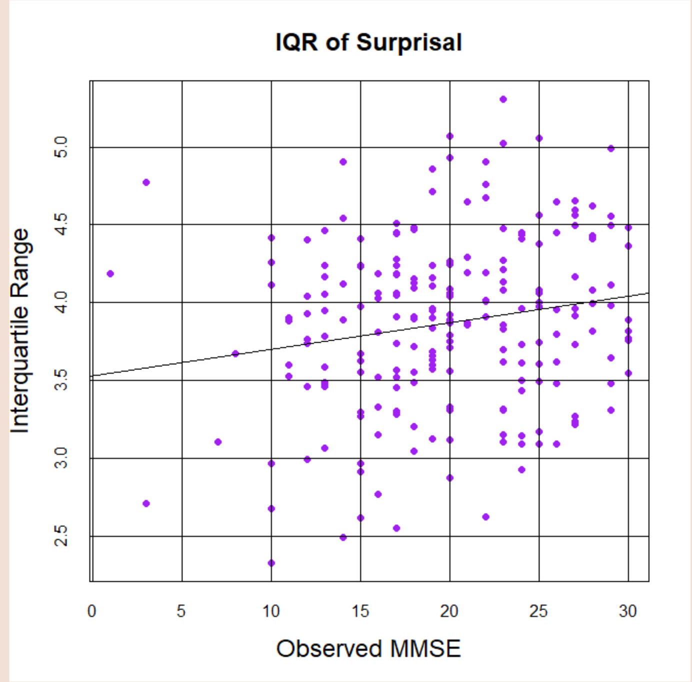


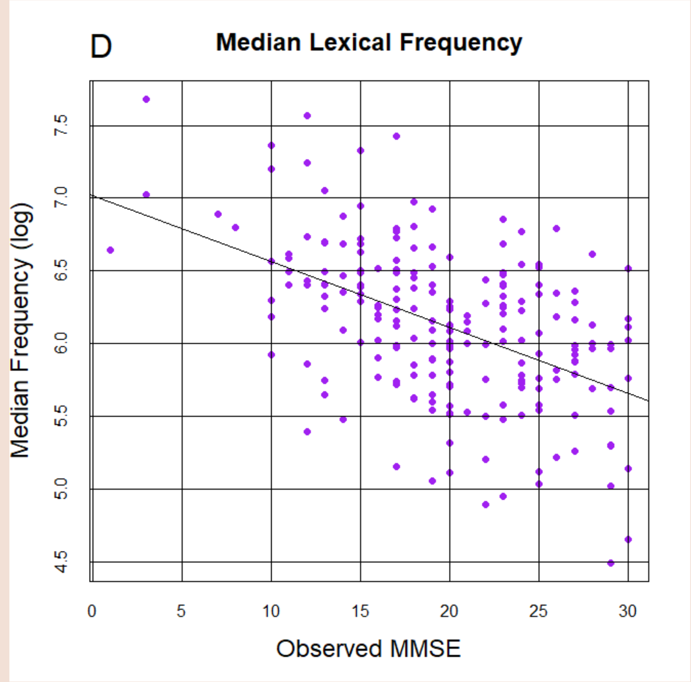


Results: Individual Predictors

- Context-sensitive features: Median word surprisal ($r = -.33$, $p < 0.001$; **Fig. C**) and surprisal interquartile range ($r = 0.18$, $p < 0.02$) were both significantly correlated with MMSE score.
- Lexical features: Median lexical frequency ($r = -0.50$, $p < 0.001$; **Fig. D**), and usage of **definite articles** ($r = 0.31$, $p < 0.001$), **nouns** ($r = 0.26$, $p < 0.001$), and **empty words** ($r = -0.25$, $p < 0.001$) were also individually correlated with MMSE score.







Conclusions

- Lower MMSE scores were associated with speech marked by **more frequent, yet more surprising**, words, increased use of empty words, and fewer definite articles and nouns.
- These results suggest that **computational approaches to estimating lexical predictability** (e.g., GPT-2 and other large language models), may have **value in predicting the degree of cognitive decline** from speech.