High-functioning individuals may perform in the normal range on cognitive tests despite significant existing pathology. Previous research has identified changes in voice features in AD and MCI. Here, we aim to detect cognitive effort in the context of a verbal cognitive task, using automated voice analysis and machine learning. The goal is to develop a novel voice biomarker as a means to detect high functioning patients with existing AD pathology.

## Method

- 2,868 participants aged 17-86 years (M=34.5, SD=12.3) completed a web-based implementation of verbal digit-span backwards, scored using automated speech recognition (ASR) on the Cambridge Cognition Neurovocalix platform (Figure 1).
- Testing was carried out in the participant’s home on their own devices. Raw audio data, scored responses and participant demographics were recorded. Participant characteristics are shown in Table 1.
- Participants’ responses were scored using ASR. The task terminated when participants responded incorrectly on three occasions at a span length. Participants were excluded if they were unable to hear a span of three. Maximum working memory span ranged from 3 to 8 items.
- We calculated cognitive load with respect to each participant’s maximum span. Responses were categorised as “high load” if they were >0.6 of their maximum span.
- Data were divided into training (60%) and test (20%) and validate (20%) datasets, with different participants in each set to minimise learning of individual voices. Training and test were used in model building and hyper-parameter tuning, respectively. Validation was held out for final model evaluation (Table 2).
- Audio features were extracted from each response. Feature vectors were normalised with respect to each participant, expressing within-subjects differences in vocal features across trials of varying load. Only correct responses were included in the analysis.
- Five different models (logistic regression, naive Bayes, support vector machine, random forest and gradient boost) were trained to predict the cognitive load score based on acoustic features.
- Once trained, models were validated on a dataset consisting of utterances of a single span (four) to ensure the model was not fitting to features related to span length.

## Results

- Results of the different classification models lengths are shown in Table 3 and ROC curves are presented in Figure 2.
- The best performing models are Random Forest and Gradient boosting. These are both ensemble models which are robust to high-dimensional datasets with correlated and redundant features.

## Conclusions

- We have demonstrated the ability to derive measures of cognitive load from voice data collected on participants’ own devices and in their own homes.
- These data suggest that automatically administered and scored verbal cognitive tests can be used to generate both reliable measures of performance and useful vocal features.
- Cognitive effort shows potential as a novel digital biomarker.
- Future work will aim to replicate these findings in patients with neurodegenerative disease, and examine the potential of these digital biomarkers in increasing sensitivity to the presence of neurodegenerative pathology.