Early development of a unified, speech and language composite to assess the clinical severity of Frontotemporal Lobar Degeneration (FTLD)

Aparna Balagopalan, MSc [2,3], Liam D Kaufman, MSc [2], Jekaterina Novikova, PhD [2], Omer Siddiqui, MSc [4], Robert Paul, MD, PhD [4], Michael Ward, PhD [4], William Simpson, PhD [1,2]

(1) Department of Psychiatry and Behavioural Neuroscience, McMaster University, Hamilton, ON, Canada, (2) Winterlight Labs, Toronto, ON, Canada, (3) Department of Computer Science, University of Toronto, Toronto, ON, Canada, (4) Alector Pharmaceuticals, South San Francisco, CA, USA.

Background

Frontotemporal Lobar Degeneration (FTLD) is a progressive neurodegenerative disorder that presents clinically as heterogeneous phenotypes, including bvFTD and Primary Progressive Aphasia (PPA).
The variable presentation of FTLD makes the development of measures for diagnosis, clinical staging and longitudinal tracking of disease progression challenging *(1-3)*. The pronounced language deficits observed in PPA suggest that a detailed, computational analysis of speech may be a suitable approach for developing such a metric. Previously, our group found the analysis of speech could be used to distinguish between PPA subtypes with up to 95% accuracy *(4)*. Speech and language variables identified using these approaches could be used to derive a 'speech assay' of FTLD that may serve as an accurate marker of clinical severity. The goal of this study was to evaluate the validity and test-retest reliability of such a metric in a proof-of-concept study.

Methods

- To create a candidate list of speech and language features, we leveraged data from an ongoing, prospective observational study of FTLD participants (n=34), a normative cohort of healthy adults aged 50-95 (n=135), and AphasiaBank (n=115); a publically available dataset of participants with post-stroke aphasias.
- All datasets contained recordings of picture description tasks, using the Cookie Theft or equivalent images. For AphasiaBank, only Wernicke's and Broca's Aphasia participants were selected
- Using Winterlight's analysis platform, ~550 acoustic and linguistic features were extracted from each speech sample
- Candidate features were selected using a 3-step process:
  - FTLD and normative dataset participants were compared. Features with statistical differences between groups were selected (List A)
  - Using AphasiaBank, correlations were computed between individual features and available Western Aphasia Battery (WAB) subscores for Spontaneous Speech Fluency + Information Content, Repetition, Object Naming and Word Fluency. Features with a correlation >0.5 were selected (List B)
  - Features in List A and List B were compared. If a feature appeared in both lists, it was selected for the speech and language composite.
- The speech and language composite was generated as follows:
  - Individual features were z-scored against the normative dataset
  - Polarity in z-scored features was adjusted so that an increase in all features indicated improved performance
  - z-scored features were averaged to produce a single composite score
- The composite was validated against the longitudinal FTLD dataset where test retest reliability (intra-class correlation) was evaluated. Longitudinal change over time was examined graphically.
- To evaluate construct validity, the composite score was generated for healthy, FLTD and Alzheimer's disease (AD) participant groups. Group means and distributions were assessed.

Figure 1: Full FTLD sample longitudinal trajectory for the language composite score (A). Individual trajectories for randomly selected bvFTD (B) and PPA (C) participants.

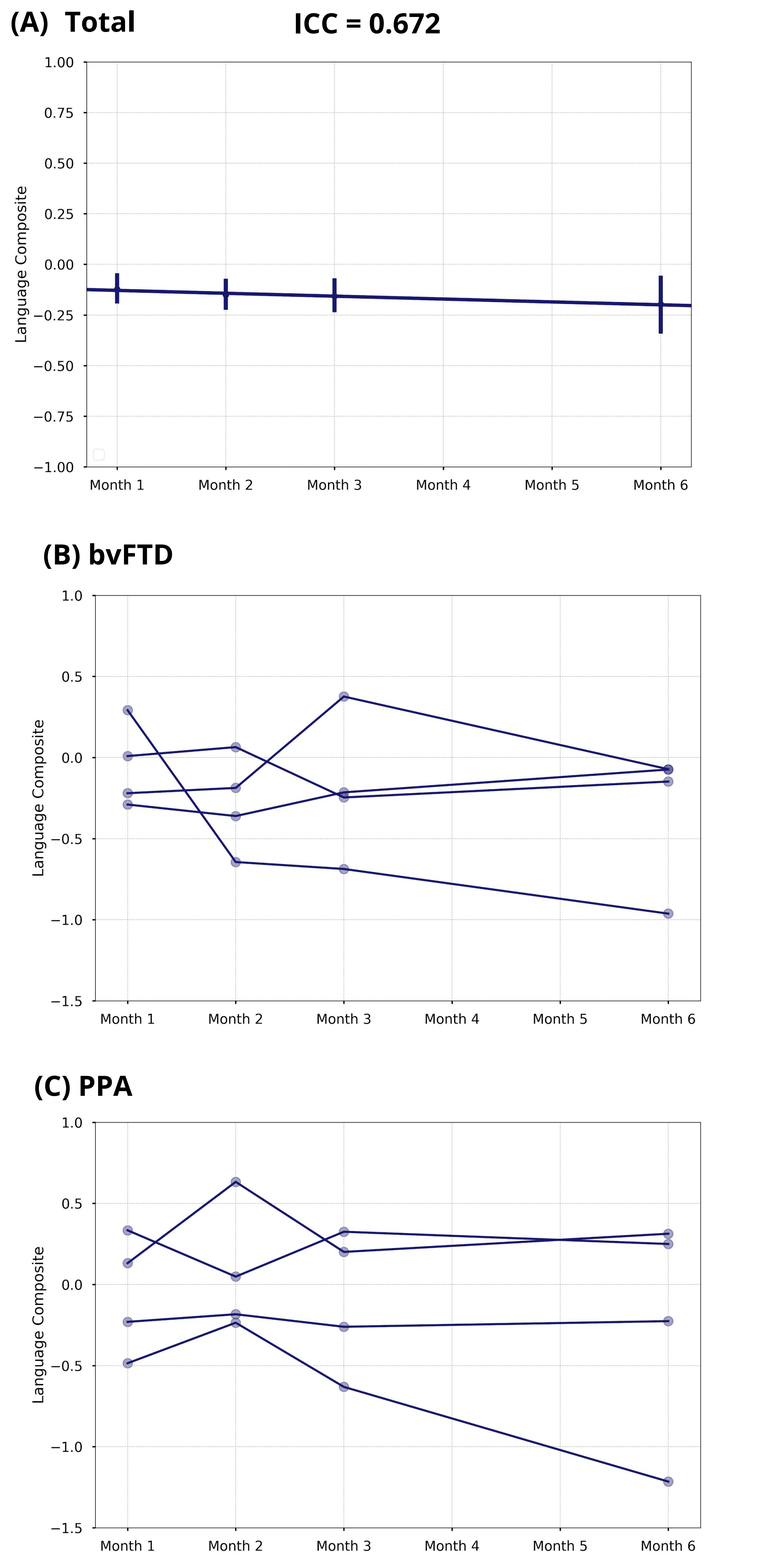
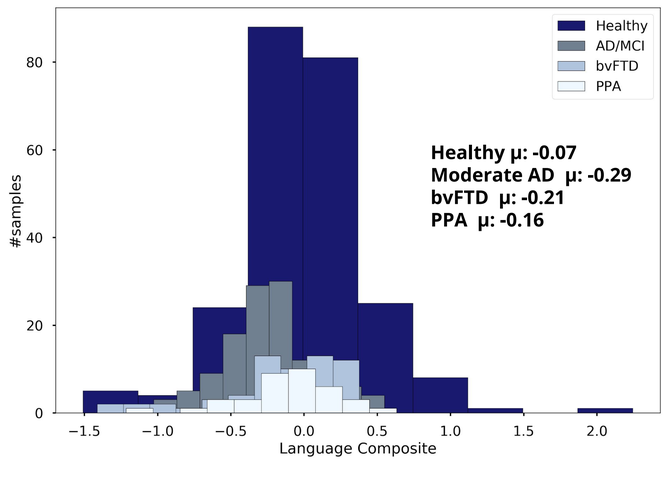


Figure 2: Distribution of language composite scores for Healthy (HC), Alzheimer's Disease (AD), bvFTD and PPA individuals



Results

Feature Category	N Sig	Examples
Acoustic	12	MFCCs, spectral composition
Lexical Complexity and Richness	12	Frequency, Parts of Speech
Word Finding Difficulty	3	Speech rate, NID words
Syntactic Complexity	32	Speech production rules

WAB Subscore	Spearman (rho) with Language Composite
Spontaneous Speech - Info Content	0.36
Spontaneous Speech - Fluency	0.29
Repetition	0.37
Word Fluency	0.67

- The proposed language composite showed good test-retest reliability, ICC= 0.672
- Longitudinal trajectories for the complete bvFTD/PPA sample suggest a slow and fairly stable decline (Figure 1A)
- Longitudinal trajectories over a 6 month time period however, showed higher heterogeneity (Figure 1 B,C)
- Preliminary examination of construct validity found that bvFTD and PPA participants in our prospective cohort were similar in their language performance to those with AD. Both AD and bvFTD/PPA scored lower than HC, though the standard deviation of scores for each group was broad (Figure 2)

Conclusions

In this preliminary, proof-of-concept study, we show that it is feasible to generate a stable speech and language composite score from a detailed analysis of speech data. At the group level, this composite appears to track broadly with the expected clinical trajectory of FTLD and does appears to capture some disease heterogeneity. Further, more extensive analysis of construct validity using larger sample sizes is required to determine the clinical utility and diagnostic sensitivity of this metric.

References

(1) Miyagawa, T. et al. Use of the CDR® plus NACC FTLD in mild FTLD: Data from the ARTFL/LEFFTDS consortium. Alzheimers. Dement. (2019).
(2) Knopman, D. S. et al. Development of methodology for conducting clinical trials in frontotemporal lobar degeneration. Brain 131, 2957–2968 (2008).
(3) Mioshi, E., Hsieh, S., Savage, S., Hornberger, M. & Hodges, J. R. Clinical staging and disease progression in frontotemporal dementia. Neurology 74, 1591–1597 (2010).
(4) Fraser, K. C. et al. Automated classification of primary progressive aphasia subtypes from narrative speech transcripts. Cortex 55, 43–60 (2014)

Presented at Clinical Trials for Alzheimer's Disease (CTAD), San Diego, CA, USA. Dec 4-7, 2019