Analytical and Clinical Validation of Digital Language Assessments

Jessica Robin [1], Mengdan Xu [1], William Simpson [1,2], Jekaterina Novikova [1]

(1) Winterlight Labs, Toronto, ON, Canada, (2) Department of Psychiatry and Behavioural Neuroscience, McMaster University, Hamilton, ON, Canada

Background

Digital tools offer new possibilities for cognitive assessment that may be more sensitive to cognitive changes and less burdensome to patients.[1,2] These novel technologies require both analytical and clinical forms of validation to ensure they are fit for purpose.[3,4] As described in the V3 framework, analytical validation verifies that a measure is accurately measuring the outcome of interest.[3] Clinical validation serves to test the relationship of a given outcome measure with a clinical condition or symptom. In this study, we evaluate the analytical validity (i.e. how accurate are the automated scores?) and clinical validity (i.e. are the scores sensitive to clinical differences?) of digital language assessments in older adults. To accomplish this goal, we test the properties of automated versions of standard assessments and their outcome scores from four language tasks.

Methods

- The Winterlight App provides a range of digital language assessments including standard neuropsychological language tests such as: picture description, object naming, phonemic fluency and semantic fluency.
- In each task, participants are guided through the task and prompted to make verbal responses, describing a picture, naming objects displayed on a screen, or naming as many words as possible in a minute that fit into a certain category (i.e. animals or words that start with the letter F).
- Verbal responses are recorded by the app, transcribed and analyzed, generating >500 variables that describe the acoustic and linguistic characteristics of the speech recording.
- For each task, a standard score is generated reflecting performance on the task following standard scoring practices.
- For analytical validation, two trained human raters manually scored 150-200 recordings each made by healthy older adults (MoCA scores >= 26) for each of the speech tasks.
- Pearson correlations were computed between the raters and the automated scores, and between the two human raters, for comparison.
- For clinical validation, scores on each task were compared between groups of healthy older adults (MoCA scores >= 26, N = 43) and those with cognitive impairment due to mild cognitive impairment (MCI) or early Alzheimer's disease (AD) (N = 22) using linear regression models with factors of group, age, sex and years of education.

Figure 1: Clinical validation of task scores for detecting language changes in MCI/early AD

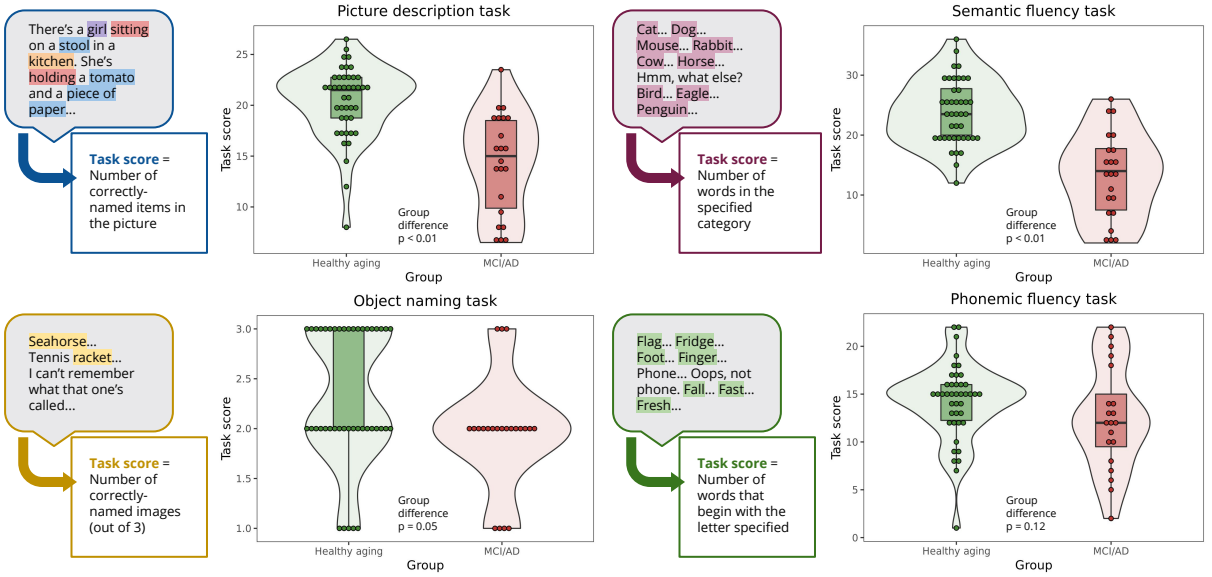


Table 1: Analytical validation to assess agreement between manual and automated scoring

Language score	Agreement between automated scores and human raters (*r*)	Inter-rater agreement between two human raters (*r*)
Picture description	0.66-0.71	0.76
Object Naming	0.63-0.66	0.92
Semantic Fluency	0.67-0.83	0.96
Phonemic Fluency	0.93-0.96	0.99

Conclusions

This study evaluates the use of digital language assessments and automated scoring for assessing language abilities in older adults. Agreement between automated scores and human scorers was highest for phonemic fluency, and comparable to interrater agreement for picture description. Picture description and semantic fluency scores were the most sensitive to differences between healthy participants and those with MCI or early AD. Overall, a digital version of the picture description task appears to be as reliable as human scoring and the most sensitive to detecting cognitive impairment, supporting the utility of digital assessments to assess cognition. Digital speech assessments can be used remotely, enabling faster, safer and less burdensome screening and monitoring for dementia.

References

(1) Kourtis, L. C., Regele, O. B., Wright, J. M. & Jones, G. B. Digital biomarkers for Alzheimer's disease: the mobile/wearable devices opportunity. npj Digital Med 2, 9 (2019).
(2) Dagum, P. Digital biomarkers of cognitive function. npj Digital Med 1, 10 (2018).
(3) Goldsack, J. C. et al. Verification, analytical validation, and clinical validation (V3): the foundation of determining fit-for-purpose for Biometric Monitoring Technologies (BioMeTs). npj Digit. Med. 3, 55 (2020).
(4) Robin, J. et al. Evaluation of Speech-Based Digital Biomarkers: Review and Recommendations. Digit Biomark 99–108 (2020) doi:10.1159/000510820.

Presented at ISCTM 17th Annual Scientific Meeting. April 6-9, 2021

Variation in speech and language variables based on demographic factors

Jessica Robin [1], Mengdan Xu [1], William Simpson [1,2]

Winterlight Labs, Toronto, ON, Canada, (2) Department of Psychiatry and Behavioural Neuroscience, McMaster University, Hamilton, ON, Canada

Background

Speech is a promising modality for developing digital biomarkers for psychiatric and neurological disorders.[1] Variations in speech and language have been shown to occur in a broad spectrum of neurological and psychiatric indications, making these measures potentially useful for detecting and monitoring disease.[2,3,4] To use speech assessments to accurately measure disease, however, variations in speech and language based on demographic factors including age, sex and education must be understood and accounted for. In this study, we determine the relationship between demographic factors and speech variables based on a normative dataset of older adults, in order to understand variations in these measures that occur independently from disease-related changes.

Methods

- Speech recordings were collected from 164 community-dwelling study volunteers (57 M, 107 F, mean age = 70, range = 50-95, mean years of education = 15, range = 6-26) who were enrolled in a normative data collection study.
- Speech samples were elicited by an open-ended picture description task.
- Picture descriptions were recorded and analyzed using natural language processing tools, generating >500 variables per recording measuring the different acoustic and linguistic characteristics of speech.
- To determine the relationship between demographic factors and speech variables, non-parametric group comparisons based on sex were made using Mann-Whitney U tests, and correlations between age and years of education and speech variables were tested using Spearman rank-order correlations.

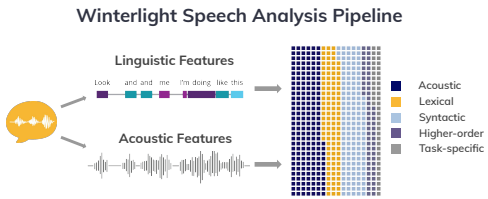


Figure 1: Number of speech variables that differ according to demographic variables

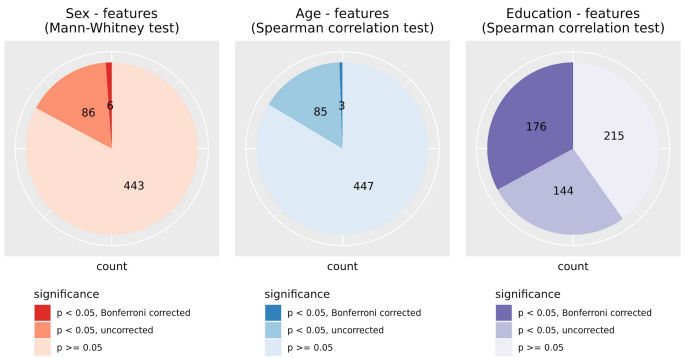


Figure 2: Example speech variables that differ by sex and age

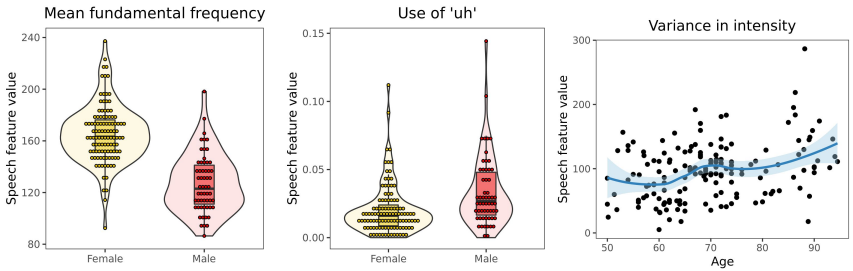
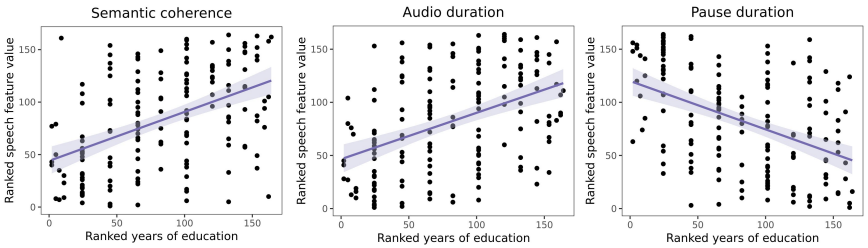


Figure 3: Example Speech variables that differ by education level



Results

- A small number of the >500 acoustic and linguistic variables showed significant associations with sex.
- Variables that differed by sex included acoustic variables such as fundamental frequency, and the use of possessive pronouns and use of the filled pause "uh".
- Similarly, a small number of variables had significant associations with the age of the speaker.
- All variables with significant associations with age were acoustic, including variance and skewness of Mel-frequency cepstral coefficients (MFCCs) and the variance in intensity of the recording, suggesting that older participants have higher variance in their vocal characteristics.
- A higher proportion of speech variables had significant correlations with the years of education of the speaker.
- Variables associated with years of education included the duration of speech, length of pauses, length of utterances and the coherence and graph organization of language.

Conclusions

These results demonstrate that speech and language patterns largely have minimal associations with the age and sex of the speaker in this normative sample, with a few exceptions. There is acoustic variation in speech based on the sex of the speaker reflecting different vocal pitches, and usage of certain word categories (i.e. possessive pronouns, filled pauses) differs by sex. Similarly, there is some variation in acoustic features, including the variance in intensity of the voice, that is associated with the age of the speaker. In contrast, a number of acoustic and linguistic properties of language were associated with the number of years of formal education of the speaker. For example, our results indicate that speakers with more years of education tend to speak longer, with shorter pauses and their speech is more coherent.

Together, these findings indicate that age, sex and especially years of education should be controlled for when analyzing speech and language patterns. These findings have implications for the statistical analysis of novel speech-based biomarkers as exploratory endpoints in clinical trials and explain the variation that might occur in such measures based on the demographic variability of a population.

References

(1) Robin, J. et al. Evaluation of Speech-Based Digital Biomarkers: Review and Recommendations. Digit Biomark 99–108 (2020) doi:10.1159/000510820.
(2) Szatloczki, G., Hoffmann, I., Vincze, V., Kalman, J. & Pakaski, M. Speaking in Alzheimer's Disease, is That an Early Sign? Importance of Changes in Language Abilities in Alzheimer's Disease. Front. Aging Neurosci. 7, (2015).
(3) Poole, M. L., Brodtmann, A., Darby, D. & Vogel, A. P. Motor Speech Phenotypes of Frontotemporal Dementia, Primary Progressive Aphasia, and Progressive Apraxia of Speech. J Speech Lang Hear Res 60, 897–911 (2017).
(4) Low, D. M., Bentley, K. H. & Ghosh, S. S. Automated assessment of psychiatric disorders using speech: A systematic review. Laryngoscope Investigative Otolaryngology 5, 96–116 (2020).

Presented at ISCTM 17th Annual Scientific Meeting. April 6-9, 2021