Comparing longitudinal changes in speech-based digital measures in cognitively healthy, possible cognitive impairment, and MCI/AD individuals

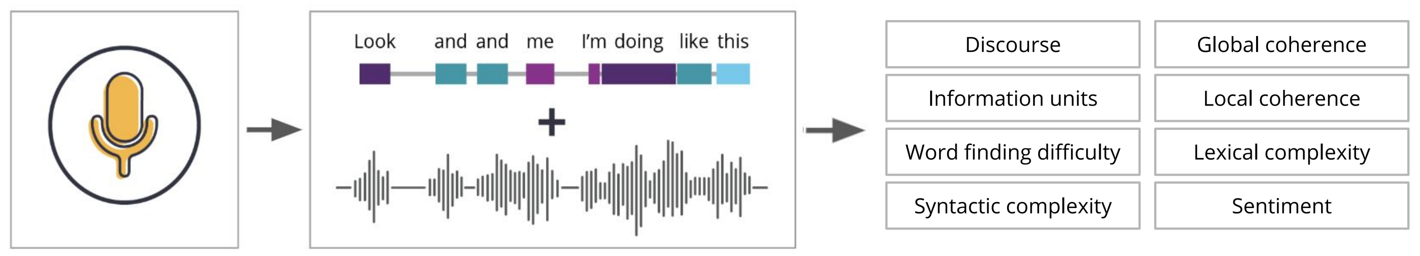
Jessica Robin[1], Mengdan Xu[1], Liam D. Kaufman[1], William Simpson[1,2]
(1) Winterlight Labs, Toronto, ON, Canada
(2) Department of Psychiatry and Behavioural Neuroscience, McMaster University, Hamilton, ON, Canada

Background

- In Alzheimer's disease (AD), changes to speech and language differentiate individuals with AD from healthy controls and may precede clinical diagnosis[1,2,3,4]
- Speech-based digital biomarkers may be able to detect early signs of mild cognitive impairment (MCI) using a brief, naturalistic and objective speech assessment
- Speech-based digital biomarkers may offer more sensitive tools for tracking disease progression in MCI and AD
- The objectives of this study are to determine:
  - If speech-based digital measures can distinguish cognitively healthy older adults from those with possible cognitive impairment based on cognitive screening measures, and those with diagnoses of MCI or AD
  - How well speech-based digital measures can measure change over time in these groups

Methods

- 130 community-dwelling older adults were recruited for this study
- Participants completed a tablet-based speech assessment and the Montreal Cognitive Assessment (MoCA)[5] at Baseline and 6 months
- Participants were divided into three groups:
  - Cognitively healthy (MoCA ≥ 26 at baseline and at 6 months; n = 18, mean age = 66.2 yrs, 61.1% female)
  - Possible cognitive impairment group (MoCA < 26 at both timepoints; n = 19, mean age = 79.6 yrs, 78.9% female)
  - MCI/AD group (clinician dx; n = 17, mean age = 77.0 yrs, 47% female)
- Speech samples were recorded, transcribed and analyzed to produce 8 aggregate scores pertaining to different aspects of speech and language, chosen for their previous association to AD[3]
- Two-way mixed ANOVAs were used to compare language scores across groups and assess change over time

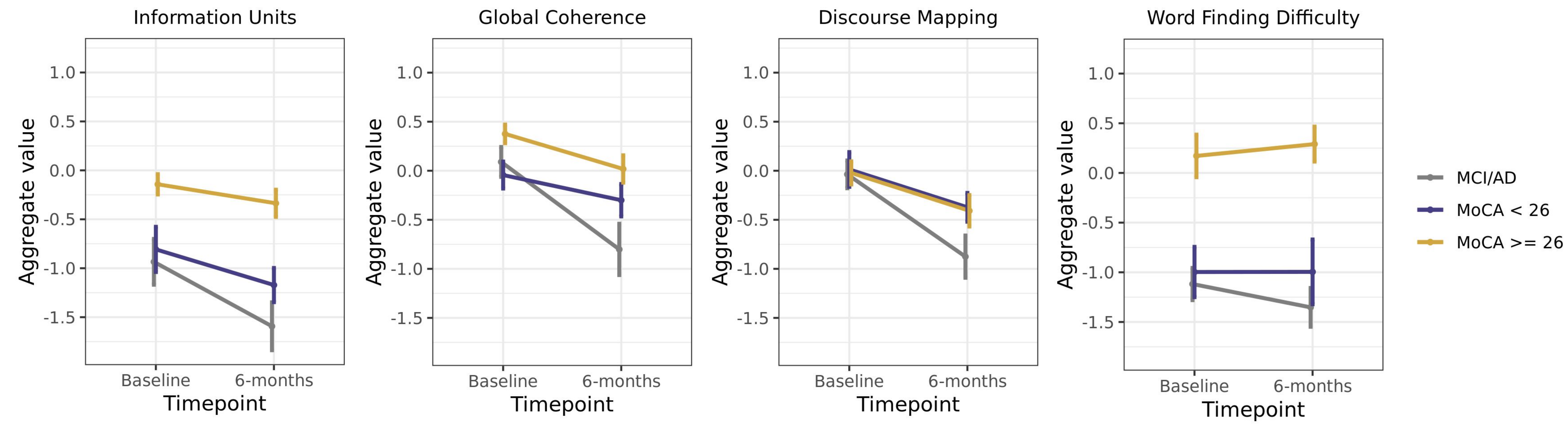


Speech, collected through iPad app

500+ automatically extracted features from the linguistic and acoustic domains

Features grouped into 8 aggregate scores representing different aspects of speech and language

Figure 1: Group differences and change over 6 months in selected speech aggregates



Results

Speech Aggregate	Effect of group (p-value)	Effect of time (p-value)	Interaction of group x time (p-value)
Information Units	**< 0.001**	**< 0.001**	0.15
Global Coherence	**0.02**	**< 0.001**	0.06
Discourse Mapping	0.27	**< 0.001**	0.20
Word Finding Difficulty	**< 0.001**	0.74	0.46
Local Coherence	**0.03**	0.07	0.27
Lexical Complexity	**0.02**	**< 0.001**	0.21
Syntactic Complexity	**< 0.001**	**0.03**	0.07
Sentiment	**< 0.001**	**< 0.001**	0.23
MoCA Scores	**< 0.001**	0.56	0.36

- Seven of the eight speech aggregate scores showed significant effects of group, suggesting that the groups can be distinguished based on multiple aspects of speech and language
- For all aggregates except sentiment, the cognitively healthy (MoCA > 25) group had the highest scores, consistent with predictions
- Six of the eight speech aggregates showed significant effects of time, with three (information units, global coherence, discourse mapping) showing declines in scores over 6 months
- No language composite had a significant interaction of group x time, though several showed trends of steeper decline in MCI/AD
- There was no significant effect of time or group x time interaction on MoCA scores

Conclusions

- This study demonstrates that speech-based biomarkers are sensitive to detect differences in individuals based on cognitive status and MCI/AD diagnosis
- A number of speech aggregates showed significant decline over a 6-month period, unlike MoCA scores
- Scores reflecting information content and coherence of speech both differentiated the groups and showed decline in a 6 month period
- No speech measure had a significant interaction between group and time, which may be due to the timescale of follow up or the small sample sizes
- Ongoing work with larger samples and longer study periods will continue to examine which aspects of speech and language are most sensitive to cognitive status and disease progression and validate novel digital biomarkers
- Digital speech assessments represent promising tools for characterizing early cognitive decline and monitoring change over time

References

1. Fraser KC, Meltzer JA, Rudzicz F. Linguistic Features Identify Alzheimer's Disease in Narrative Speech. JAD. 2015 Oct 15;49(2):407–22.
2. Boschi V, Catricalà E, Consonni M, Chesi C, Moro A, Cappa SF. Connected Speech in Neurodegenerative Language Disorders: A Review. Front Psychol. 2017 Mar;8.
3. Slegers, A., Filiou, R.-P., Montembeault, M. & Brambati, S. M. Connected Speech Features from Picture Description in Alzheimer's Disease: A Systematic Review. J. Alzheimers. Dis. 65, 519–542 (2018).
4. Berisha V, Wang S, LaCross A, Liss J. Tracking Discourse Complexity Preceding Alzheimer's Disease Diagnosis: A Case Study Comparing the Press Conferences of Presidents Ronald Reagan and George Herbert Walker Bush. JAD. 2015 Mar 24;45(3):959–63.
5. Nasreddine ZS, Phillips NA, BÃ©dirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: A Brief Screening Tool For Mild Cognitive Impairment: MOCA: A BRIEF SCREENING TOOL FOR MCI. Journal of the American Geriatrics Society. 2005 Apr;53(4):695–9.

Presented at the Alzheimer's Association International Conference, July 27-31 2020

Contact: jessica@winterlightlabs.com

A Comparison of Clinician Assessment of Speech Versus Automated Speech Analysis in Mild Cognitive Impairment and Alzheimer's Dementia

WINTERLIGHT

UNIVERSITY OF TORONTO

Anthony Yeung, MD[1], Andrea Iaboni, MD DPhil[1,2], Elizabeth Rochon, PhD[1,2], Monica Lavoie, PhD[1], Calvin Santiago, MSc MD[1], Maria Yancheva, MSc[3], Jekaterina Novikova, PhD[3], Mengdan Xu, MSc[3], Jessica Robin, PhD[3], Liam D Kaufman, MSc[3] and Fariya Mostafa, MSc[3]

1. University of Toronto, ON, Canada, 2. Toronto Rehab, University Health Network, ON, Canada, 3. Winterlight Labs, Toronto, ON, Canada

Correspondence: Jessica Robin, PhD <jessica@winterlightlabs.com>

Background

- Language impairment is a core feature of Alzheimer's disease (AD) and other neurodegenerative disorders.[1]
- Prior studies have shown a link between AD symptom severity and declining speech and language capability in picture description tasks.[2]
- Speech and language changes include alterations in speech rate, utterances, frequency of words, word-finding difficulties, and repetitions.[3]
- Despite these pervasive language changes, there is no universally accepted system of terminology used to describe language impairment, and large inter-rater variability can also exist between clinicians.[4]
- In view of current limitations, the role of automated speech analysis is emerging as a novel, and potentially more objective method of assessing language in individuals with neurologic and psychiatric disorders.
- We sought to: (1) define a set of speech and language capability ratings that can be used by clinicians with different areas of specialization, (2) determine if these speech and language ratings are applied consistently in a sample of patients including healthy controls, mild cognitive impairment (MCI), and AD, and (3) use automated speech analysis to identify what acoustic and linguistic variables correlate with clinician ratings of speech and language.

Methods

- Speech samples were obtained via the DementiaBank (DB) dataset through the TalkBank Project, with equal numbers of healthy controls, MCI, and probable AD participants.
- Participants provided a recording of a speech sample which consisted of a verbal description of the Boston Cookie Theft picture.
- The recordings were rated by 5 clinicians (1 geriatric psychiatrist, 1 psychiatry resident, 1 neurology resident, and 2 speech language pathologists) with clinical experience in assessing MCI and AD, according to four characteristics: (1) word-finding difficulty, (2) incoherence, (3) perseveration, and (4) errors in speech; these were rated on a Likert scale (range: 0-3) as being: not present/normal finding, mild, moderate, or severe **(Table 1).**
- Speech recordings were transcribed, and linguistic and acoustic variables were extracted through automated speech analysis using NodeJS and React. Data processing and feature extraction was performed using Python-based standard acoustic and language processing libraries (e.g., spacy) and custom code.
- The correlation between clinician-identified speech characteristics and the acoustic and linguistic variables were then compared using Spearman correlation.
- Exploratory factor analysis (EFA) was then applied to find common factors between variables for each speech characteristic, using R version 3.6.3 and Python version 3.6.

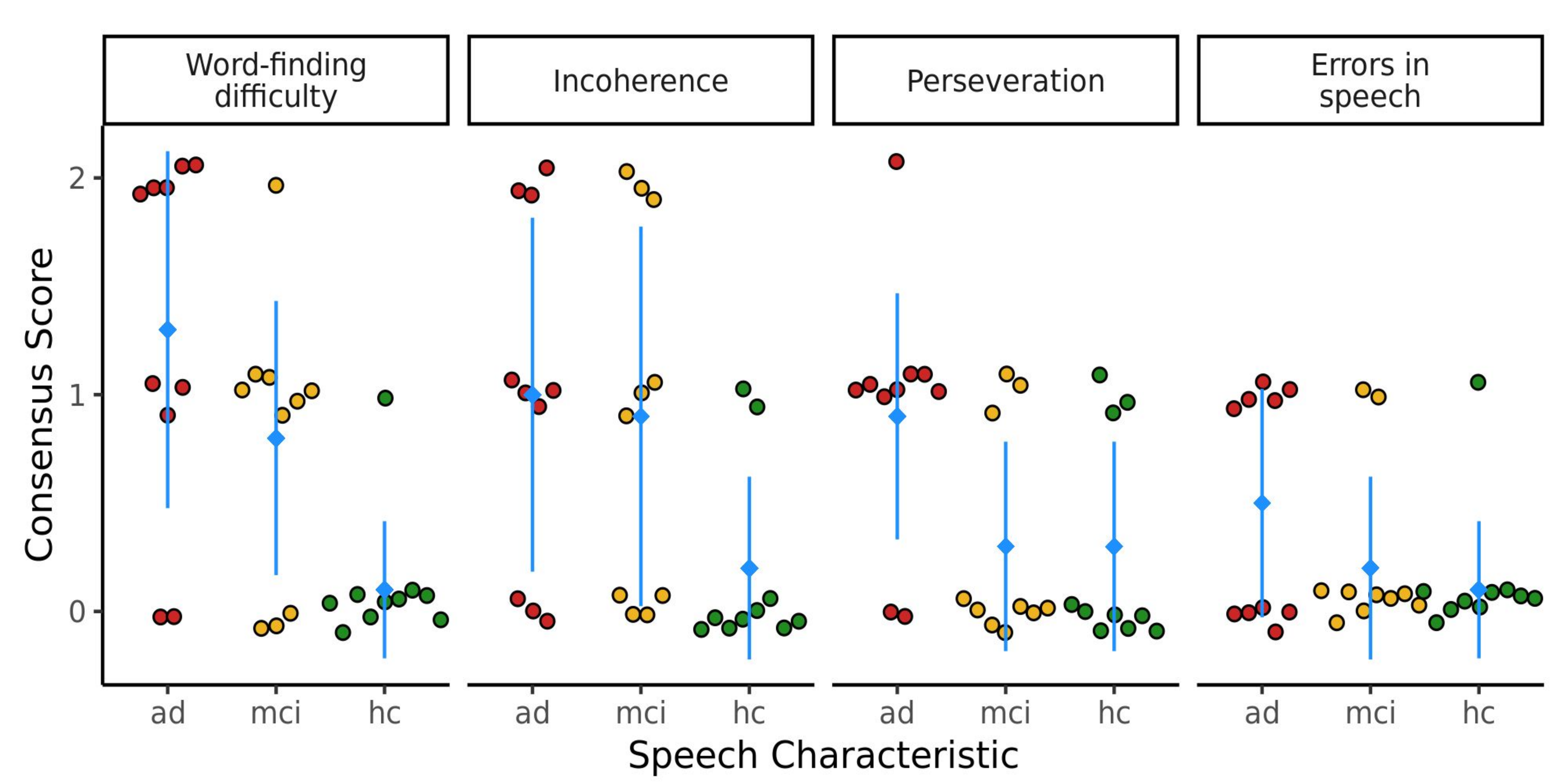
Table 1: Clinician Consensus Table of Speech Characteristics

Speech and Language Characteristic	Clinical Features
Word-finding difficulty	• Reduction in content words, circumlocution, false starts, pauses while searching for words, fluency (rate, phrase length, amount of hesitation), revisions (repetitions of complete words or phrases/elaborations), indefinite terms (fillers).
Incoherence	• Disorganized speech, derailment or sudden topic shifts, tangentiality, flight of ideas, or word salad.
Perseveration	• Repetition of word or phrase even after the stimulus for the behavior (word or phrase) has been taken away; persistence of behavior (word or phrase) despite repeated failure; intrusions (i.e., Inappropriate repetition of prior responses after intervening stimuli).
Errors in Speech	• Phonetic errors (omissions, additions, substitutions, distortions), stuttering, sequences of phonemic approximation.

Table 2: Participant demographics by diagnostic group

	Controls (n=10)	MCI (n=10)	AD (n=10)
Age at visit, mean (SD), y	61.2 (9.7)	69.9 (5.9)	64.0 (11.0)
Female (%)	50	50	50
MMSE, mean (SD)	29 (0.9)	24 (2.0)	18 (1.6)
Education, mean (SD), y	14.2 (2.3)	14.0 (1.9)	13.8 (2.2)

Figure 1: Consensus Clinician Ratings for Each Speech Characteristic



Distribution of the consensus clinician ratings for each speech characteristic, by diagnosis group. The mean consensus rating for each group is indicated with a blue diamond and whiskers indicate the standard deviation. For all ratings, a rating of 3 = severe, 2 = moderate, 1 = mild, and 0 = no presence or a normal finding of that speech characteristic.

Presented at Alzheimer's Association International Conference (AAIC) Virtual Conference, July 29, 2020

Results

- The participants demographics/characteristics are described in **Table 2**.
- Clinician rating agreement was high in three of the four speech characteristics (word-finding difficulty: ICC = 0.92, $p < 0.001$; incoherence: ICC = 0.91, $p < 0.001$; perseveration: ICC = 0.88, $p < 0.001$).
- Speech ratings scores were highest (most impairment) in the probable AD group, followed by MCI and controls. Greater impairments in word-finding difficulty and incoherence were more frequent in AD and MCI.
- For **word-finding difficulty**, variables with the highest correlations to clinician ratings were related to the rate of speech, word duration and length and the number of unfilled pauses. Greater severity of word-finding difficulty was associated with slower speech, shorter words and increased pauses.
- For **incoherence**, the variables with the highest correlations were a mix of syntactic, acoustic and lexical variables, reflecting the use of past tense verb phrases, slower speech rate, and words with higher estimated age of acquisition.
- For **perseveration**, variables with the highest correlations were related to the complexity of speech and vocabulary. Greater severity of perseveration was associated with increased repetitiveness of speech, decreased vocabulary richness, and decreased semantic similarity. A large number of acoustic variables also correlated with perseveration.
- For **errors in speech**, the variables with the highest correlations with the consensus clinician ratings included measures relating to the complexity of speech and vocabulary, use of subordinate clauses, and word length.
- EFA showed that between 1 to 4 factors were found to explain each characteristic (data not shown)

Conclusions

- In this exploratory study, variables extracted through automated acoustic and linguistic analysis of MCI and AD speech were strongly correlated to speech and language characteristics rated by clinicians.
- We were able to demonstrate that commonly used clinical terms such as word-finding difficulty, incoherence, perseveration, and errors in speech, can be correlated to features identified through automated speech analysis.
- Strengths of the study include utilizing clinician ratings to provide an objective, understandable, and rational approach to defining speech changes in AD and MCI.
- Limitations include a small sample size and short speech recording sample based on the Cookie Theft task.
- Our work proposes a standardized approach to investigating speech on both a clinical and pathophysiological level. Potential future applications of this method includes the wide scale deployment of speech analysis in resource-limited or remote settings.

References

(1) Forbes-McKay, K., Shanks, M. F., & Venneri, A. (2013). Profiling spontaneous speech decline in Alzheimer's disease: a longitudinal study. Acta Neuropsychiatrica, 25(6), 320-327.
(2) Kavé, G., & Dassa, A. (2018). Severity of Alzheimer's disease and language features in picture descriptions. Aphasiology, 32(1), 27-40.
(3) Slegers, A., Filiou, R. P., Montembeault, M., & Brambati, S. M. (2018). Connected speech features from picture description in Alzheimer's disease: A systematic review. Journal of Alzheimer's Disease, 65(2), 519-542.
(4) Wilson, S. M., Eriksson, D. K., Schneck, S. M., & Lucanie, J. M. (2018). A quick aphasia battery for efficient, reliable, and multidimensional assessment of language function. PloS one, 13(2).

Quality comparison of remote vs. in-person digital speech assessment for dementia

Jessica Robin[1], Jekaterina Novikova[1], Sasha Sirotkin[1], Maria Yancheva[1], Liam D. Kaufman[1], William Simpson[1,2]

(1) Winterlight Labs, Toronto, ON, Canada

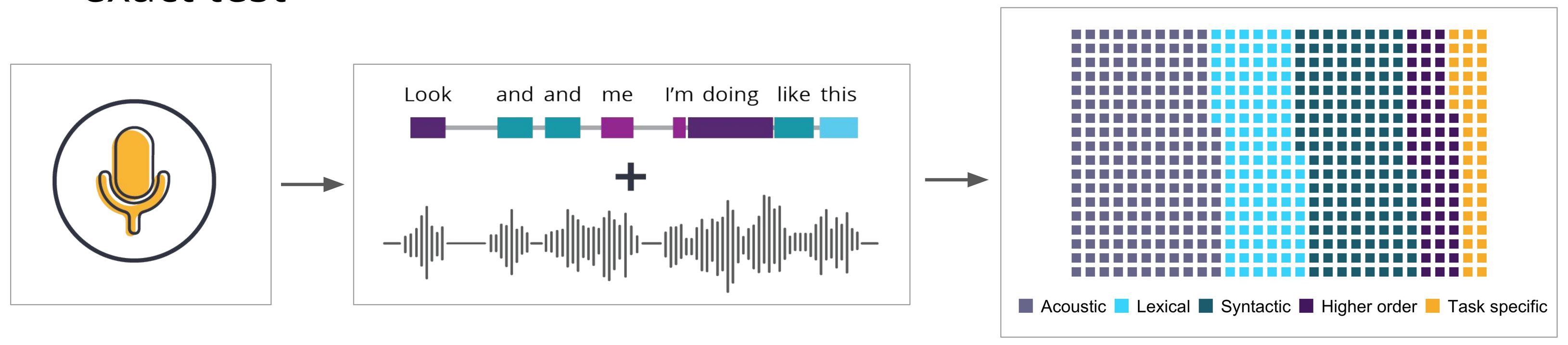
(2) Department of Psychiatry and Behavioural Neuroscience, McMaster University, Hamilton, ON, Canada

Background

- Advancements in digital health technologies have the potential to enable remote patient assessment and monitoring[1,2,3]
- Remote testing lowers the burden on patients and caregivers, and may enable more frequent and naturalistic assessment
- Digital speech assessments are an example of a digital health tool that can be remotely administered and offers insight into neurological and psychiatric health[4,5,6]
- In this study, we compared the quality of speech assessments administered at home, with the help of a caregiver, to those administered in a clinical setting, in two samples of individuals with dementia (Alzheimer's Disease and Frontotemporal Dementia)

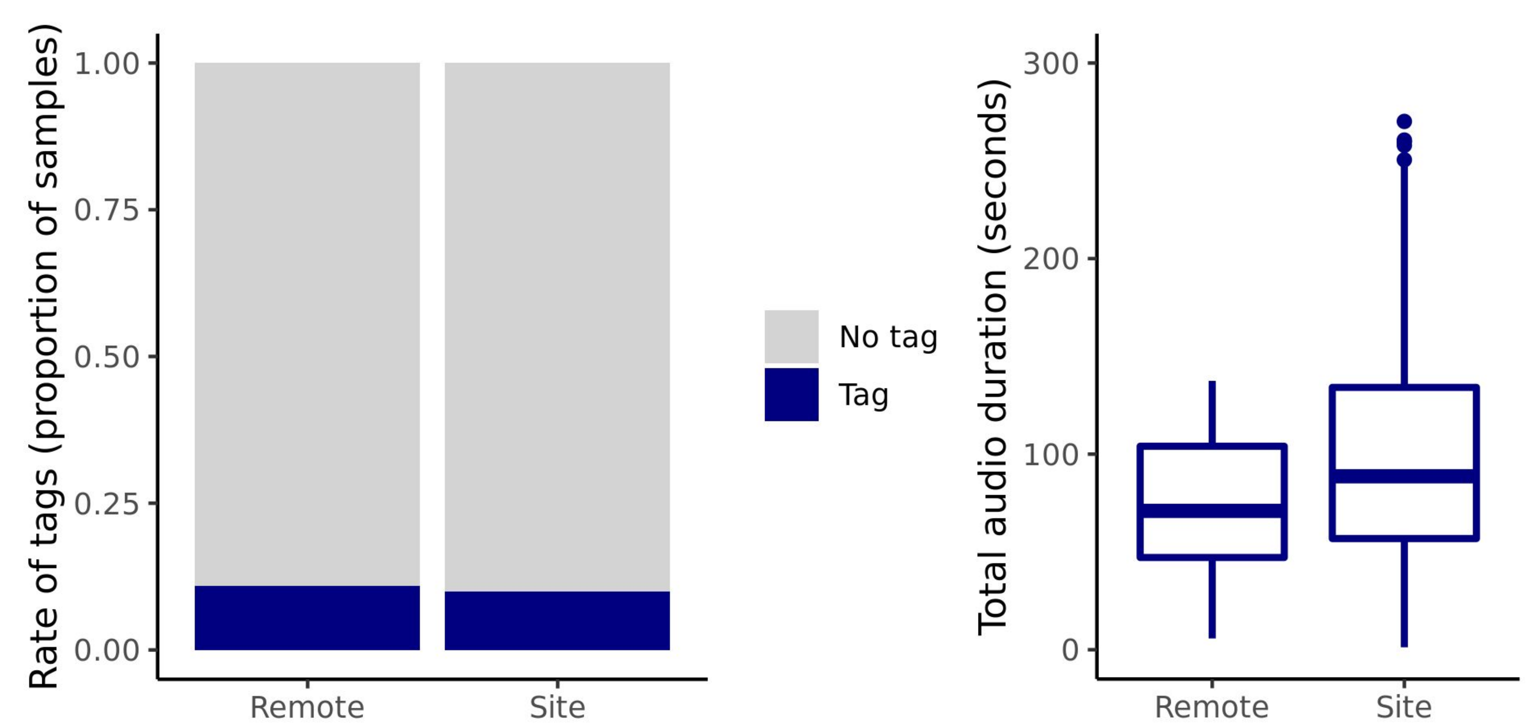
Methods

- Speech samples were examined from two ongoing studies:
  - A clinical trial for Alzheimer's Disease, with speech assessments conducted in a clinical setting, by a trained rater
  - An observational study of Frontotemporal Dementia, with speech assessments conducted at home, with the assistance of a caregiver trained on administering the assessment
- Speech assessments included picture description, phonemic and semantic fluency tasks
- In total, 575 speech samples from clinical sites (AD clinical trial) were compared to 574 speech samples from the remote study (observational FTD study)
- All speech samples were manually transcribed by trained transcriptionists
- As part of the transcription process, samples are tagged for any possible quality issues (see Table 1)
- The rates of all tags and each tag type were compared across samples from the clinical sites and remote sites using Fisher's exact test



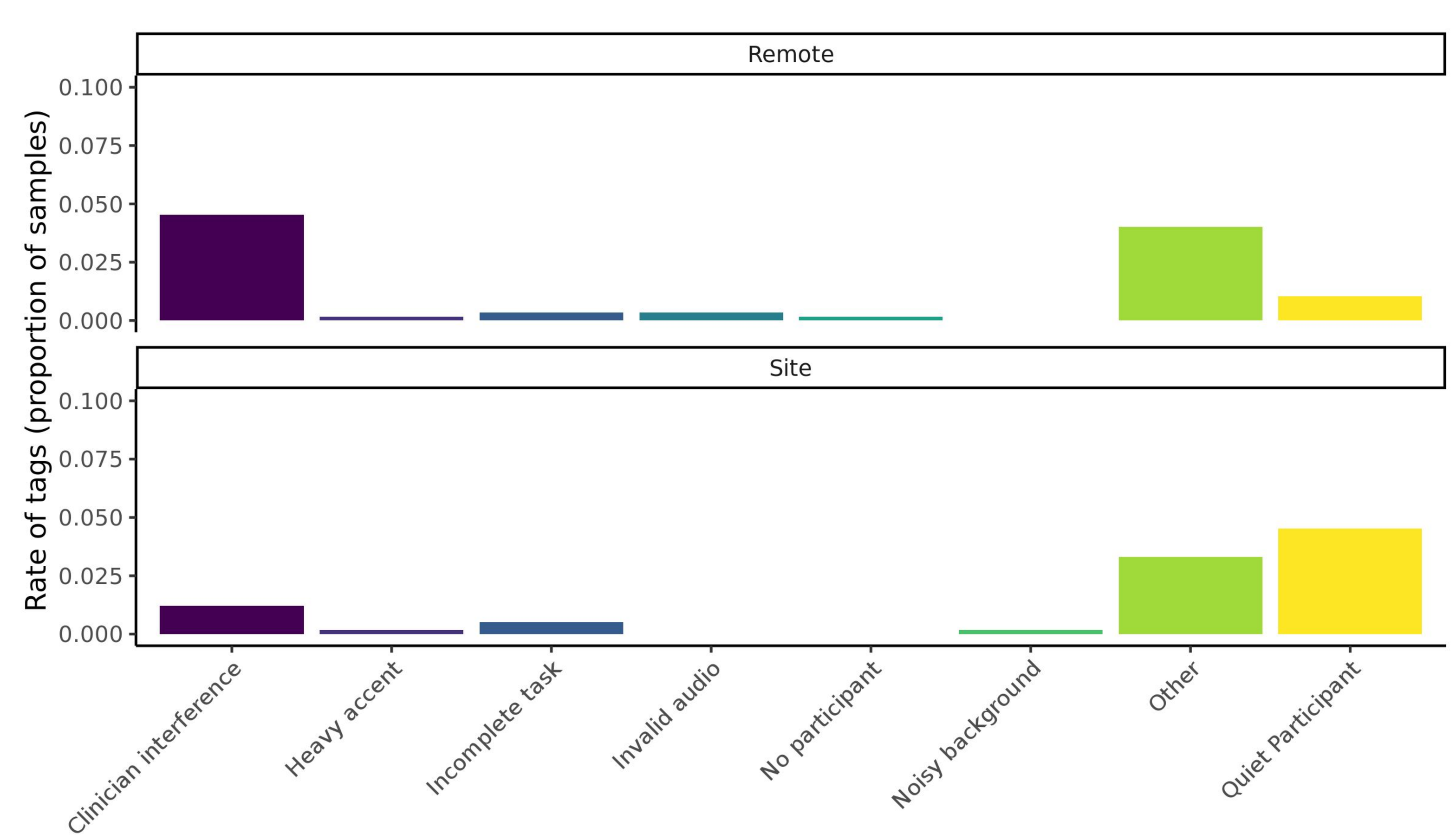
Speech assessment | **Transcription of speech samples** | **Feature extraction**

Figures 1 & 2: Comparison of overall incidence of tags and sample durations



- Overall rate of any tag did not differ significantly ($p > 0.6$) between speech samples collected at Remote (10.9%) and Clinical Sites (9.9%)
- Speech samples for picture description tasks were significantly longer in duration ($p < 0.001$) at Clinical Sites (mean duration = 101 seconds) compared to Remote samples (mean duration = 73 seconds)

Figure 3: Comparison of tag types across samples



- Clinician (caregiver) interference was more frequent for Remote samples ($p < 0.01$)
- Quiet participants were more frequent in the Site samples ($p < 0.001$)
- All other tags were equal or infrequent across samples

Table 1 : Tags for possible quality issues

Transcription tags	Usage
Clinician interference	The person administering the speech assessment interferes with the task by providing assistance, encouragement or other commentary
Heavy accent	The participant has a heavy accent (including dialects and non-native accents)
Incomplete task	The participant did not complete the task or did not follow instructions
Low audio quality	The audio quality of the sample is poor, due to artifacting or distortion
Invalid audio	There is no audio recorded or the audio cannot be used
No participant	There is no participant audible in the sample
Noisy background	There is high amounts of noise in the background either from the other people speaking or from the environment
Overwhelming noise	There is background noise to the point that the participant cannot be heard
Quiet participant	The participant is hard to hear due to low volume, whispering or mumbling
Other	Any other possible quality issue with the audio, not covered by the categories above

Conclusions

This study suggests that remote speech assessments yield recordings of comparable quality to in-person assessments. We found higher, though still low, rates of caregiver interference for remote assessment, which should be monitored and mitigated in future remote assessment. Surprisingly, recordings from clinical sites had higher instances of quiet participants, which could be due to microphone placement. Remote assessments yielded shorter recordings, but this may be due to the different dementia diagnoses across groups. Future work should compare the same participants across both assessment settings.

References

1. Piau A, Wild K, Mattek N, Kaye J. Current State of Digital Biomarker Technologies for Real-Life, Home-Based Monitoring of Cognitive Function for Mild Cognitive Impairment to Mild Alzheimer Disease and Implications for Clinical Care: Systematic Review. J Med Internet Res. 2019 Aug 30;21(8):e12785.
2. Bakker JP, Goldsack JC, Clarke M, Coravos A, Geoghegan C, Godfrey A, et al. A systematic review of feasibility studies promoting the use of mobile technologies in clinical research. npj Digit Med. 2019 Dec;2(1):47.
3. Kourtis LC, Regele OB, Wright JM, Jones GB. Digital biomarkers for Alzheimer's disease: the mobile/wearable devices opportunity. npj Digital Med. 2019 Dec;2(1):9.
4. Fraser KC, Meltzer JA, Rudzicz F. Linguistic Features Identify Alzheimer's Disease in Narrative Speech. JAD. 2015 Oct 15;49(2):407–22.
5. Fraser KC, Meltzer JA, Graham NL, Leonard C, Hirst G, Black SE, et al. Automated classification of primary progressive aphasia subtypes from narrative speech transcripts. Cortex. 2014 Jun;55:43–60.
6. Boschi V, Catricalà E, Consonni M, Chesi C, Moro A, Cappa SF. Connected Speech in Neurodegenerative Language Disorders: A Review. Front Psychol. 2017 Mar;8.

Presented at the Alzheimer's Association International Conference, July 27-31 2020

Contact: jessica@winterlightlabs.com