

Dementia and Technology

Can Gaze–Speech Coupling in Reading Help Clinicians Detect Cognitive Decline Remotely? A. Laghai, S. Shafiyani, N. Thomas, M. Kunz, K. Fraser, B. Wallace, R. Goubran, F. Knoefel. *Gerontechnology* 25(s)

Purpose Early detection of cognitive decline is essential, yet access to specialist assessment is limited outside major centres. There is a growing need for a simple, scalable, remote screening tool that does not require specialists. This study explores the use of eye tracking and speech to support a technology-enabled remote screening tool. Cognitive decline affects both connected speech (e.g., reduced fluency [1]) and eye movements (e.g., longer fixations and reduced word skipping [2]), yet these modalities have largely been studied in isolation. No existing work has examined how these behaviours unfold during naturalistic oral reading, a task requiring coordinated visual and linguistic processing. This study jointly analyzes eye-tracking and speech during a standardized reading task. By indexing on-screen words and aligning fixation behaviour to speech, the strength and timing of gaze–speech coupling can be quantified. These results assess whether individuals with cognitive impairment demonstrate reduced gaze–speech coupling compared to healthy controls, and whether these metrics have potential as differentiating features for remote cognitive screening tools. **Methods** 15 healthy controls (HC) and 15 with clinically diagnosed cognitive decline (CD) completed three visits, spaced six months apart, yielding 84 total measurements. Gaze and speech data were collected following the procedure described by [3], and fixations were classified using the I-VT algorithm [4]. Speech was recorded and transcribed using WhisperX to obtain word-level timestamps. Participants read aloud a 64-word standardized passage across two sequential slides. Each word was assigned a region of interest and a unique index corresponding to its order in the story. For each trial, two synchronized time-series were generated: a gaze index, marking the index of the word currently fixated upon, and a speech index, marking the first time each content word was spoken. These traces represented visual and spoken progression through the reading task. Because the slide transition introduced a task boundary requiring participants to press a button, correlation metrics were computed separately for slide 1 and the entire passage. Slide 1 is focused upon as it captures natural reading behavior before the task interruption. Full-passage results followed similar patterns with smaller effect sizes. **Results and Discussion** Gaze–speech coupling was quantified using zRMSE and zMAE (magnitude of error between the fixated and spoken word indices), Pearson and Spearman correlations (how closely the gaze and speech move together over time), and peak cross-correlation (strongest alignment while allowing gaze to lead or lag). Error metrics were lower in the cognitive decline group than in healthy controls (zRMSE: CD 0.78 vs HC 0.99, $p = 0.002$; zMAE: CD 0.57 vs HC 0.76, $p = 0.003$). Pearson and Spearman correlations were higher in the cognitive decline group (Pearson: CD 0.62 vs HC 0.46, $p = 0.009$; Spearman: CD 0.59 vs HC 0.42, $p = 0.007$). Peak cross-correlation was also higher in the cognitive decline group (Cross-Correlation: CD 0.71 vs HC 0.58, $p = 0.002$). These findings show that CD group participants maintained tighter gaze–speech alignment than healthy controls, with smaller errors and stronger correlations across all coupling metrics. This pattern is consistent with a more conservative word-by-word reading strategy, in contrast to the more anticipatory gaze behaviour seen in healthy readers. These coupling metrics may therefore be used as differentiating features in a portable screening tool, enabling remote detection. Critically, this approach can be designed around webcam-based eye tracking, enabling deployment on laptops (and other personal devices). For real-world adoption, robustness must be established across recording devices, with consistent performance under variable at-home conditions. Even with these requirements, the approach has an advantage over digitized neuropsychological tests that infer cognition from taps/clicks/keystrokes, as it supports a more natural reading task, reducing variability due to technological proficiency.

References

- [1] Xu L, Chen K, Mueller KD, Liss J, Berisha V. Articulatory precision from connected speech as a marker of cognitive decline in Alzheimer’s disease risk-enriched cohorts. *J Alzheimers Dis* 2025;103:476–86. <https://doi.org/10.1177/13872877241300149>.
- [2] Opwonya J, Ku B, Lee KH, Kim JI, Kim JU. Eye movement changes as an indicator of mild cognitive impairment. *Front Neurosci* 2023;17:1171417. <https://doi.org/10.3389/fnins.2023.1171417>.
- [3] Kunz M, Syed A, Fraser KC, Wallace B, Goubran R, Knoefel F, et al. Reducing Fixation Error Due to Natural Head Movement in a Webcam-Based Eye-Tracking Method. 2023 IEEE Sensors Applications Symposium (SAS), 2023, p. 1–6. <https://doi.org/10.1109/SAS58821.2023.10253976>.
- [4] Boulay E, Wallace B, Fraser KC, Kunz M, Goubran R, Knoefel F, et al. Improving Synchronization of Eye Fixation and Saccade Measurements with Speech Recognition for Cognitive Assessment. 2024 IEEE Sensors Applications Symposium (SAS), 2024, p. 1–6. <https://doi.org/10.1109/SAS60918.2024.10636646>.

Keywords: Eye-Tracking, Speech, Biomarker, Dementia

Affiliation: Bruyère Health Research Institute, Ottawa, ON, K1N 5C9. Email: Fknoefel@bruyere.org

Acknowledgement: This work was supported by the National Research Council of Canada Aging in Place Program, AGE-WELL NCE Inc. and the AGE-WELL National Innovation Hub Program, and Bruyère Health.