

# Examining the stability of dementia-relevant speech measures in a high-frequency picture description task repeated over 5 days

CAMBRIDGE  
COGNITION

Jessica Robin<sup>1</sup>, Rosanne van den Berg<sup>2</sup>, Mengdan Xu<sup>1</sup>, Michael Spilka<sup>1</sup>, Francesca Cormack<sup>1</sup>, Roos J. Jutten<sup>2</sup>, Casper de Boer<sup>2</sup>, Sietske Sikkes<sup>2</sup>

1. Cambridge Cognition, Cambridge, UK (Winterlight Labs, Inc.) 2. Alzheimer Center Amsterdam, Neurology, Vrije Universiteit Amsterdam, Amsterdam UMC location VUmc, Amsterdam, Netherlands  
Contact: jessica.robin@camcog.com



## Background

- Speech characteristics including speech and pause rate, noun and pronoun use, and information content differ in mild cognitive impairment (MCI) and in dementia due to Alzheimer's disease (AD).<sup>1,2,3</sup>
- The picture description task can be used to elicit naturalistic speech samples to quantify these changes.<sup>4,5</sup>
- Previous work has demonstrated the feasibility and usability of this short, low-burden task in remote, high-frequency testing, and that the reliability of speech features was highest when averaging multiple samples.<sup>1</sup>
- We leveraged data from this 5-day burst testing study in cognitively unimpaired memory clinic patients to compare the reliability of speech measures when pictures were repeated or alternated across days.
- We additionally explored the presence of practice effects across the 5-day bursts and compared these for the repeating and alternating pictures.

## Methods

- 50 cognitively unimpaired Dutch-speaking adults (Table 1) completed a 5-day remote burst testing session on a tablet with two picture description tasks each day.
- In the repeating condition, the same picture stimulus was shown every day; in the alternating condition, a unique picture was shown each day (Figure 1).
- Two to three weeks later, a second burst was completed with identical stimuli to the first burst.
- We analysed the speech recordings from the picture description tasks using the Winterlight speech processing pipeline.
- We assessed the reliability of selected speech measures (Table 2) over the five days of testing using intra-class correlations (ICCs) measuring consistency over the five days of each burst.
- We used linear models with fixed effects of session and random effects of subject to test for practice effects over the five days of each burst, for the repeating and alternating pictures (Bonferroni-corrected alpha-level = 0.0025).



Figure 1. Burst testing design. Speech assessments were completed remotely using a tablet via the Winterlight Assessment (WLA) app.

Table 1. Participant demographics

n	M/F	Mean Age (SD)	Mean YOE (SD)	Mean MMSE (SD)
50	21/29	68.4 (6.2)	15.3 (3.8)	29.2 (1.0)

Note: SD = Standard deviation; YOE = Years of education; MMSE = Mini-Mental State Exam score

Table 2. Selected speech measures

Speech measure	Type
Speech rate	Timing
Pause-to-word ratio	Timing
Total words	Lexical
Pronoun-to-noun ratio	Lexical
Information units (objects)	Information content

## Results: Reliability

- ICCs were in the moderate-to-good range (ICC = 0.53–0.76) for the selected speech measures in the first burst, except for pause-to-word ratio for the repeating picture (ICC = 0.38) and pronoun-to-noun ratio for both pictures (ICC = 0.39–0.48) (Figure 2).
- In the second burst, all ICCs were moderate-to-good (ICC = 0.52–0.82).
- Reliability was comparable for the repeating and alternating pictures, with numerically higher reliability for pronoun-to-noun ratio in the repeating picture condition.

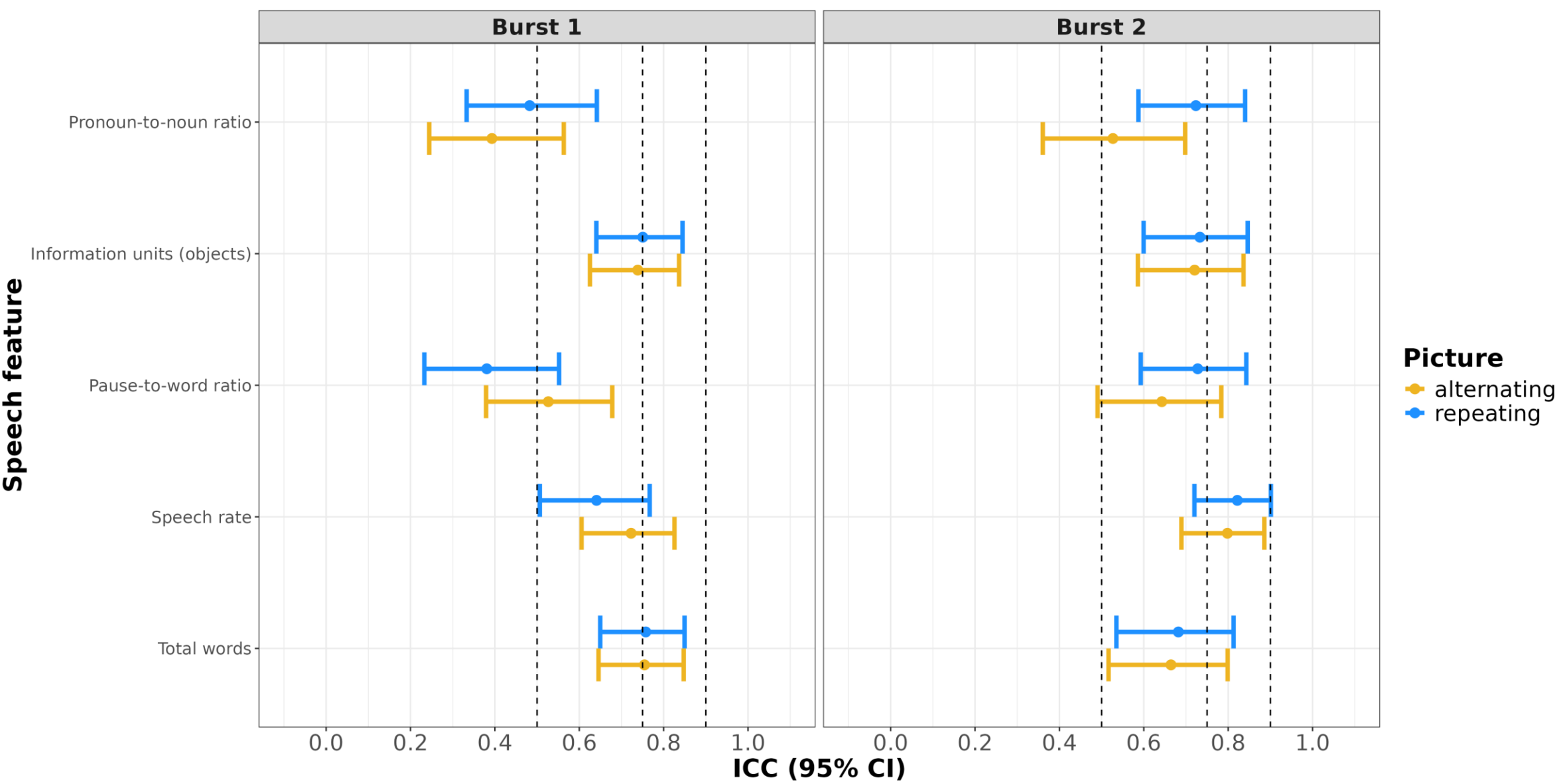


Figure 2. ICC values (± 95% CI) for selected speech features over the first and second 5-day bursts, for alternating and repeating pictures.

## References

- Van den Berg, R. L. et al. Digital remote assessment of speech acoustics in cognitively unimpaired adults: feasibility, reliability and associations with amyloid pathology. *Alz Dis Ther* 16, 176 (2024).
- Robin, J. et al. Automated detection of progressive speech changes in early Alzheimer's disease. *Alz Dis* 15, e12445 (2023).
- Robin, J., Xu, M., Kaufman, L. D. & Simpson, W. Using Digital Speech Assessments to Detect Early Signs of Cognitive Impairment. *Front. Digit. Health* 3, 749758 (2021).
- Slegers, A., Filou, R.-P., Montembeault, M. & Brambati, S. M. Connected Speech Features from Picture Description in Alzheimer's Disease: A Systematic Review. *Journal of Alzheimer's Disease* 65, 519–542 (2018).
- Mueller, K. D., Hermann, B., Mecollari, J. & Turkstra, L. S. Connected speech and language in mild cognitive impairment and Alzheimer's disease: A review of picture description tasks. *Journal of Clinical and Experimental Neuropsychology* 40, 917–939 (2018).

## Results: Practice effects

- Linear models showed significant effects of session ( $p < 0.00001$ ) on total words in both picture conditions in the first 5-day burst, and for the alternating picture in the second burst (Figure 3).
- The number of object information units correctly described significantly increased for the repeating picture in the first burst ( $p < 0.00001$ ), but this was not significant for the alternating picture or for either picture condition in the second burst.
- Other speech measures had no significant linear changes in either burst.

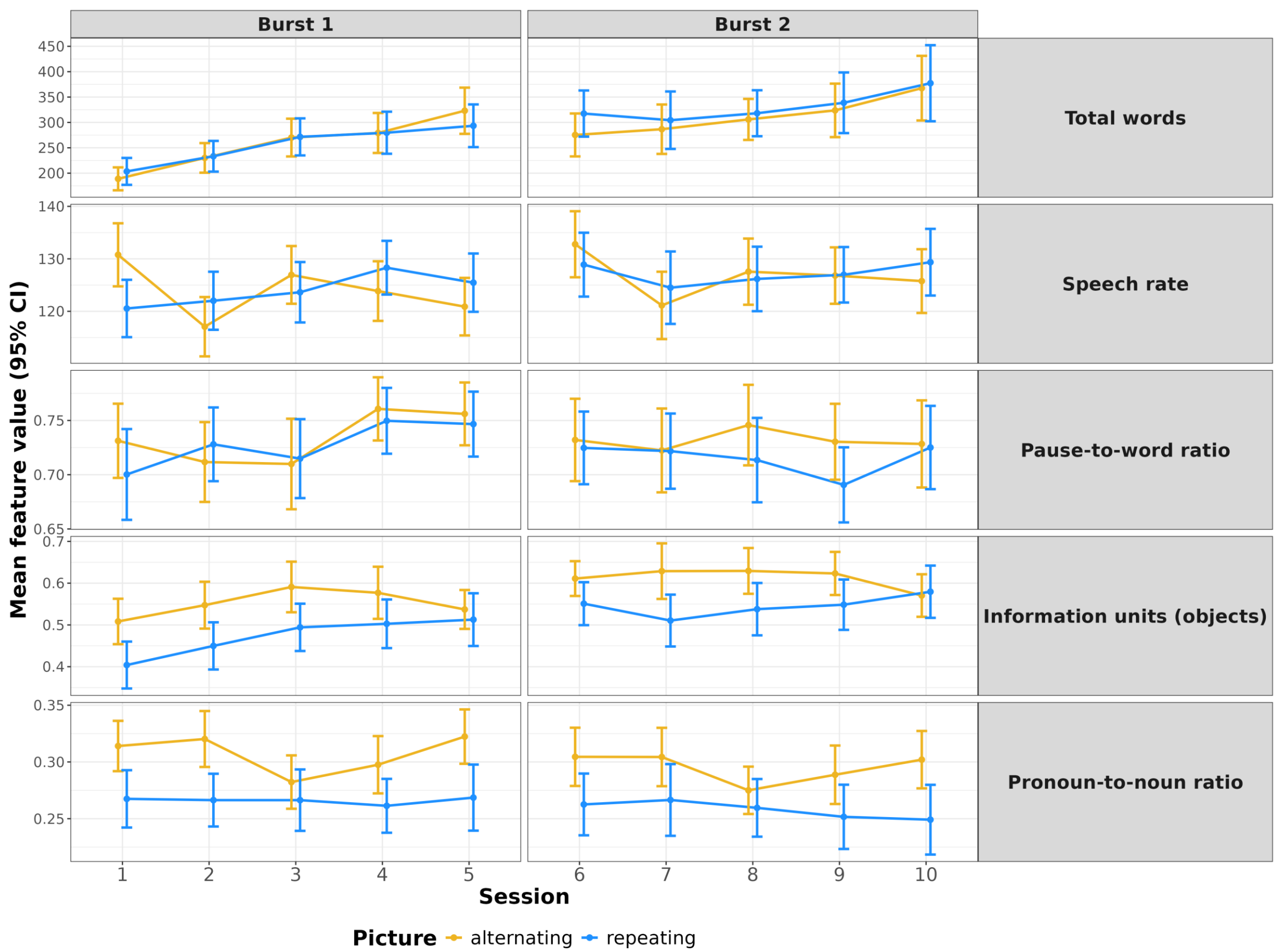


Figure 3. Mean values (± 95% CI) for selected speech features over the first and second 5-day bursts, for alternating and repeating pictures.

## Conclusions

- Selected speech measures, quantifying the rate of speech and pausing, use of nouns and pronouns and the information content described, appeared to be stable and reliable in high-frequency, remote administration of the picture description task.
- Notably, the number of words produced increased with repeated task administration, across both bursts and both picture conditions.
- Increased familiarity with the task may entail longer descriptions, regardless of the stimuli presented. It is therefore important to control for total words produced when analysing longitudinal speech data.
- Repeating the same picture stimulus at high-frequency (daily) led to increases in the number of objects correctly named in the picture, likely reflecting a learning/practice effect. This effect was not present for the alternating picture, or in the second burst, and appeared most prominent in the first 3 days of testing.
- High-frequency administration of the picture description task should therefore alternate picture stimuli to reduce learning effects and/or include pre-study familiarization sessions to reduce learning effects.
- Reliability estimates were in the moderate-to-good range for selected speech measures, and did not show consistent differences between repeating and alternating pictures.
- This research demonstrates that picture description is an ecologically valid, low burden task suitable for remote, higher-frequency, longitudinal testing of speech and language and highlights conditions that may lead to learning effects.
- Future work will probe whether practice/learning effects differ in those at risk for MCI and AD.

## Acknowledgments

The authors thank all participants for their participation. This research project was funded by the PPP Allowance made available by Health - Holland, Top Sector Life Sciences & Health, to stimulate public-private partnerships (#LSHM20084-SGF).

SCAN HERE

