Natural speech as a digital biomarker in preclinical Alzheimer's disease: Usability and reliability of a remote burst speech assessment

Rosanne L. van den Berg, MA[1,2,3], Casper de Boer, PhD[1,2], Marissa D. Zwan, PhD[1,2], Jessica Robin, PhD[4], Bill Simpson, PhD[4], John E Harrison, PhD[1,5], Roos J. Jutten, PhD[6], Mark A. Dubbelman, PhD[1,2], Frederik Barkhof, MD, PhD[7], Lyduine E. Collij, PhD[7], Argonde C. van Harten, MD, PhD[1,2], Arenda Mank, MD[1,2], Lior A. Kroeze[1,2], Charlotte E. Teunissen, PhD[8], Elsmarieke M. van de Giessen[7], Wiesje M. van der Flier, PhD[1,2] and Sietske A.M. Sikkes, PhD[1,2,3]

Background

- **Language difficulties** are often reported as one of the earliest symptoms in Alzheimer's disease (AD)[1-3] , but evidence is inconclusive about the sensitivity of traditional neuropsychological language tests for early detection[4-6].
- Fine-grained speech analysis offers the potential for capturing subtle cognitive deficits in **early-stage AD**.,[7,8]
- As such, digital recordings of natural speech are a promising **digital biomarker**, particularly when measured in **bursts** (repeated short measurements).[9,10]

Aim

Evaluate the **usability** of a tablet-based speech assessment and examine the **test-retest reliability** of acoustic speech features and its **association with amyloid-beta (Aβ) pathology** (+/-) in cognitively normal (CN) adults.

Methods

- Participants: 50 cognitively normal (CN) Dutch-speaking adults from the Alzheimer Center Amsterdam (Table 1).
- Test-retest design within 2-3 week interval, in an at-home setting.
- Measures: 1) Winterlight Speech Assessment on a tablet, implemented in a burst-design (Figure 1), 2) System usability scale (SUS): 5-point Likert scale to evaluate usability of the speech assessment.
- Acoustic speech features were extracted from speech recordings (e.g. silent pauses, fundamental frequency, jitter, shimmer).

Table 1. Participant characteristics

	Amyloid-beta positive (N=23)	Amyloid-beta negative (N=27)
Demographics		
Age, years, mean ± SD	69.61 ± 6.34	67.30 ± 6.03
Female gender, n (%)	13 (56.5)	16 (59.3)
Education, years, mean ± SD	15.22 ± 4.61	15.30 ± 2.97
Own tablet (iOS), n (%)	6 (26.1)	15 (55.5)
Amyloid-beta biomarkers		
Cerebrospinal Fluid, n (%)	3 (13.0)	2 (7.4)
Positron Emission Tomography, n (%)	20 (87.0)	25 (92.6)
Cognitive measures		
MMSE, mean ± SD*	28.83 ± 1.07	29.52 ± 0.70
CDR, mean ± SD	0 ± 0	0 ± 0

Note. SD = standard deviation, * indicates p-value < 0.05

Statistical analyses

- Usability: Differences in SUS-scores between groups: Wilcoxon rank sum test.
- Test-retest reliability: Intraclass correlation coefficients (ICC) in cumulative numbers of sessions per subtask, for each feature.
- Aβ-pathology: Differences in acoustic features between Aβ+ and Aβ- groups: linear models (LMs) corrected for age, sex, education and MMSE.
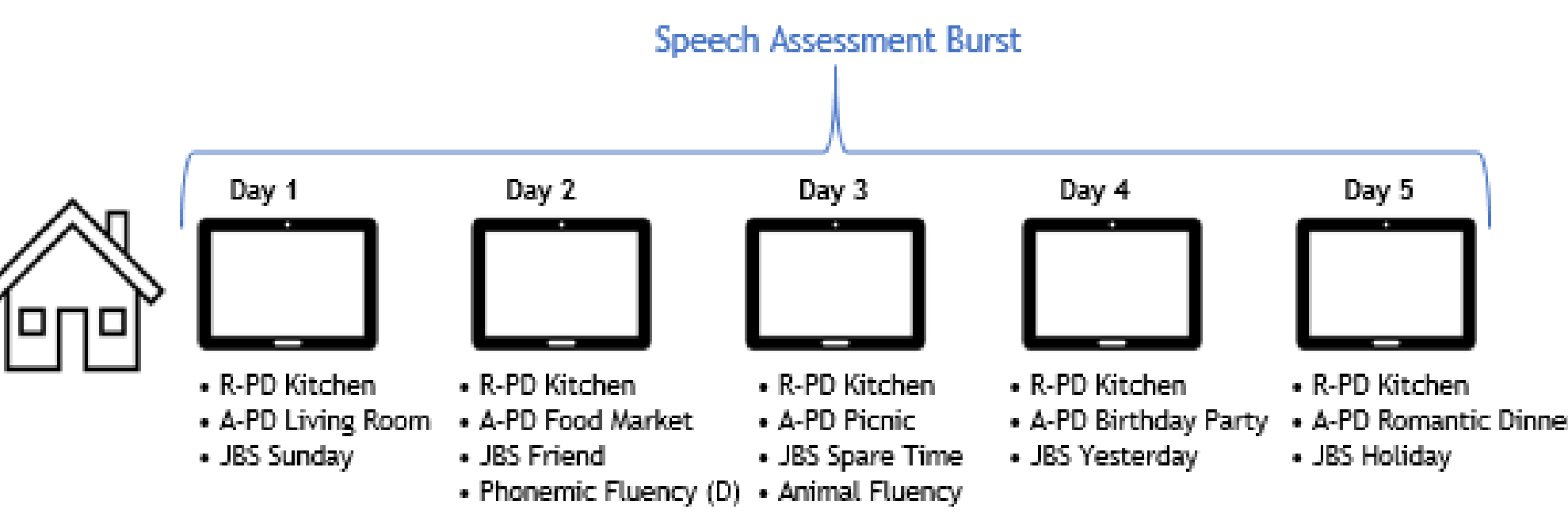
Figure 1. Study procedure of Winterlight speech Assessment implemented in a burst-testing design.
Note. RPD = repetitive picture description; APD = alternating picture description; JBS = journal-based storytelling

Results

- Usability: The average SUS score was 86 (SD = 9.88). The Aβ+ group evaluated usability as good (M=83±10), while the Aβ- group rated usability to be excellent (M=88±9, *p* = 0.019; Figure 2).
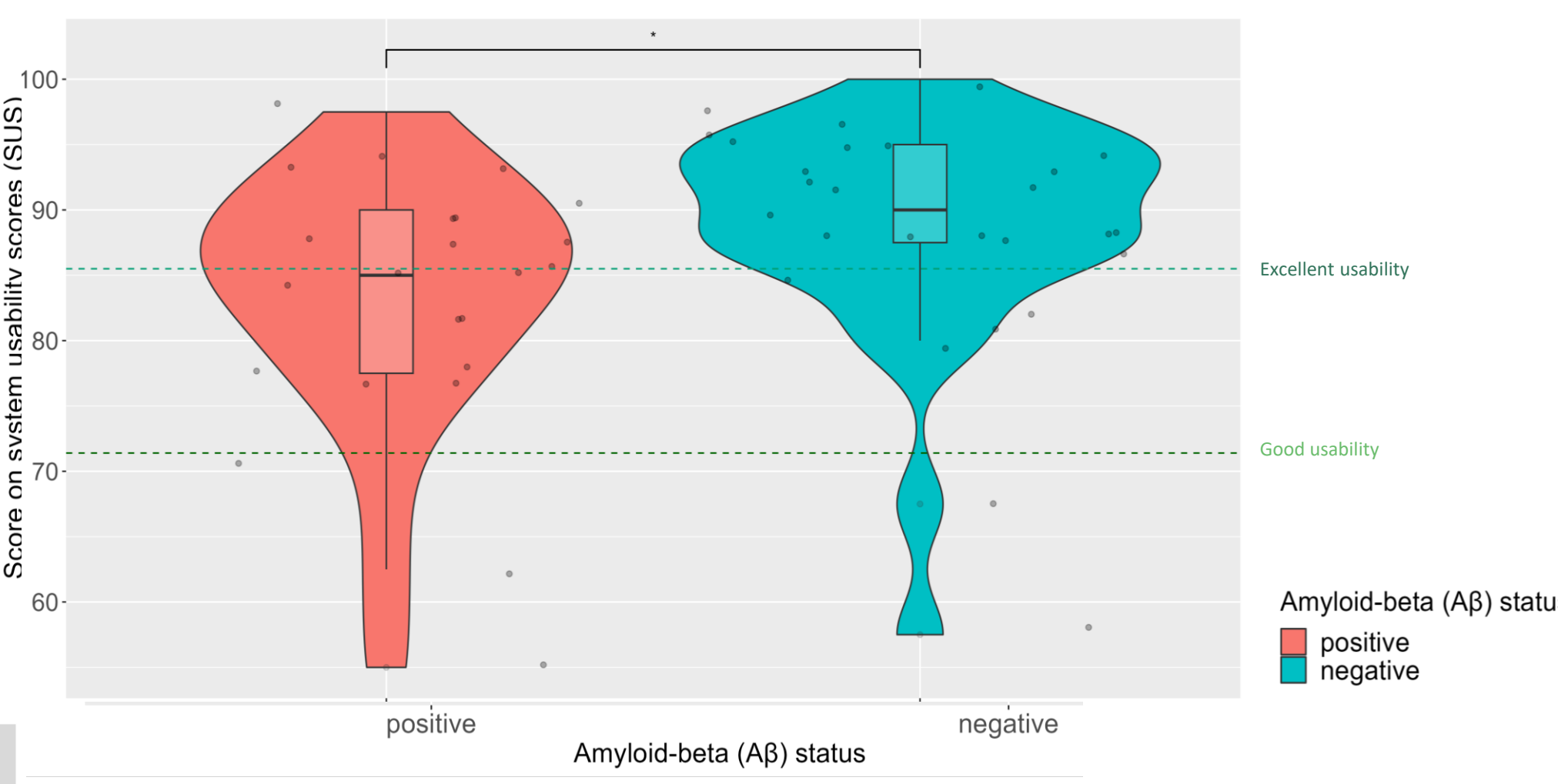
Figure 2. Scores on the system usability scale (SUS) in Aβ+ and Aβ- individuals.
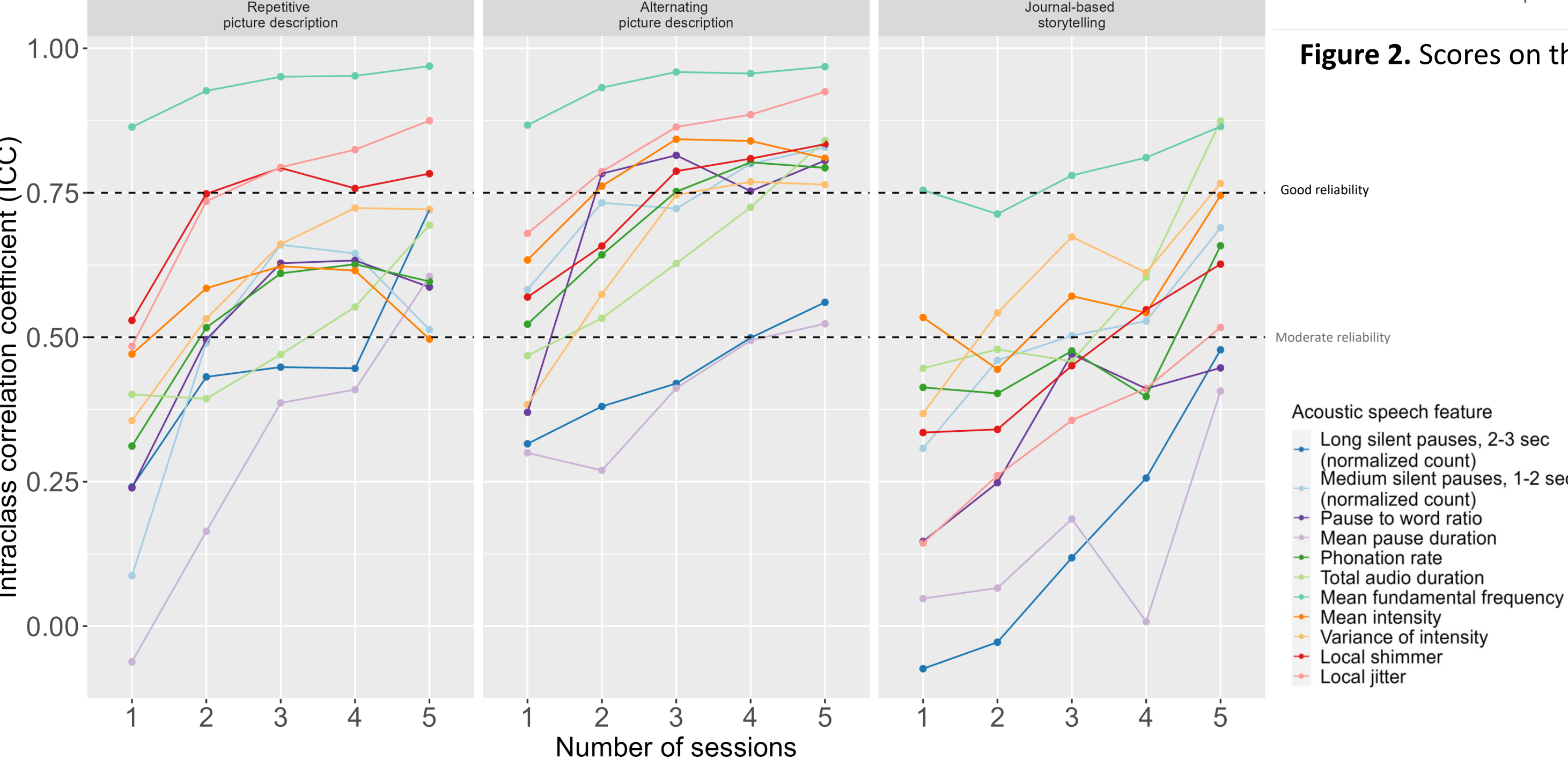
Figure 3. Intraclass correlation coefficients for test-retest reliabilities (2-3 week interval) for acoustic speech features in cumulative numbers of burst of repetitive-PD, alternating-PD and journal-based storytelling.

- Test-retest reliability: ICCs increased in more-session-bursts compared to one-session measures for all acoustic features in all speech subtasks (Figure 3).

- Aβ-pathology: The Aβ+-group showed more medium pauses in 3- and 4-session-bursts of journal-based storytelling than the Aβ-group (3-session-bursts: β=-0.07, *CI*=-0.12–-0.01, *p*=0.016; 4-session-bursts: β=-0.06, *CI*=-0.11–-0.01, *p*=0.029; Figure 4.). For none of the other number of sessions, subtasks, or acoustic features group differences were found (*p*'s > .05).
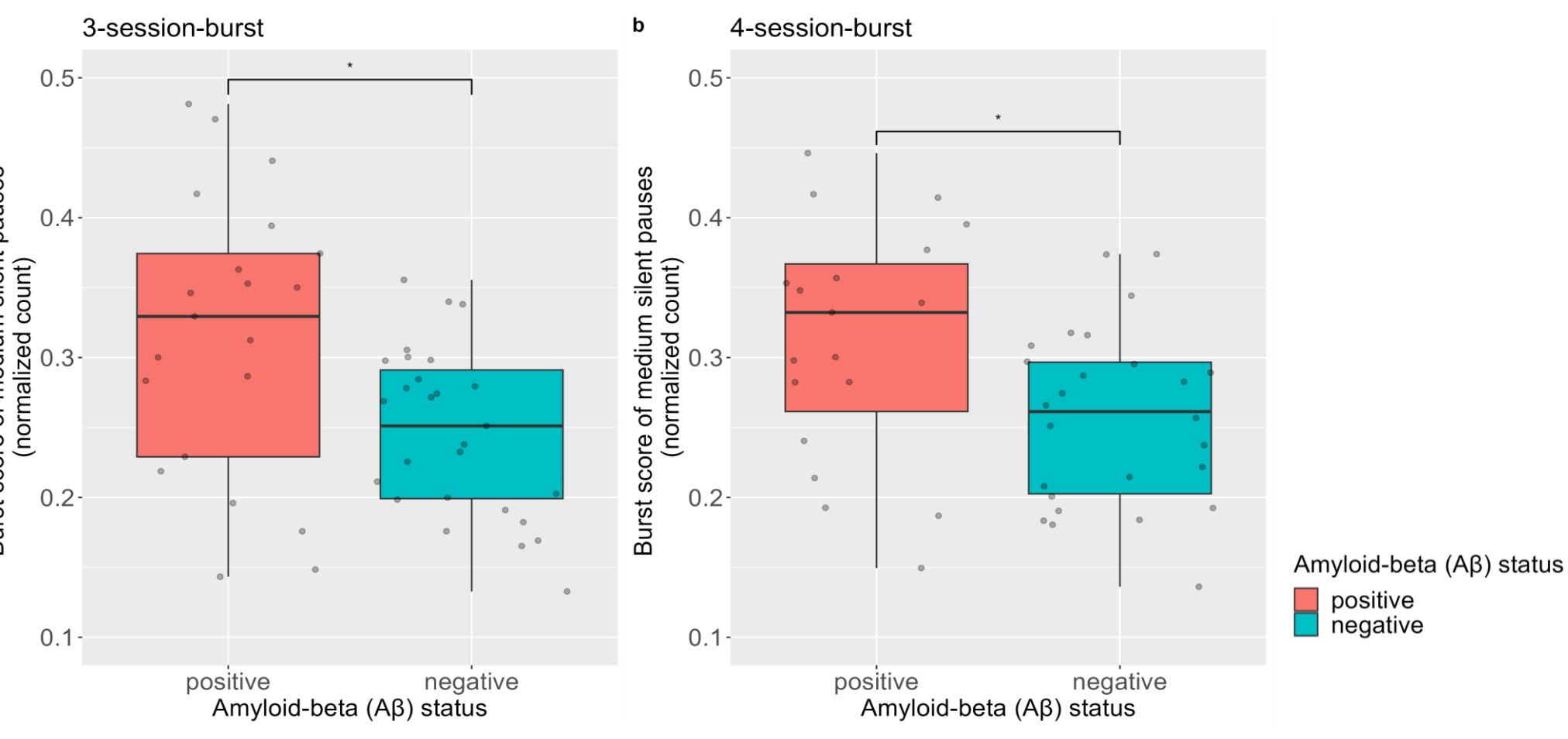
Figure 4. Medium silent pauses (normalized count) in Aβ+ and Aβ- individuals in a) 3-session-bursts and b) 4-session-bursts of journaling question storytelling.

Conclusion

- Usability: Remote burst assessment of speech is feasible in CN older adults.
- Test-retest reliability: Burst assessments enhance test-retest reliability compared to one-session measures
- Aβ-pathology**:** Burst assessments of acoustic speech features are promising to find differences in speech acoustics between Aβ+ and Aβ-negative individuals.
- These results suggest remote burst assessment holds promise for detecting subtle acoustic speech changes in the earliest AD stages.

[1]Valech et al., 2018; [2]Montembeault et al., 2022, [3]Slegers et al., 2018, [4]Papp et al., 2016; [5]Vonk et al., 2020; [6]Mueller et al., 2021, [7]Petti et al. (2020); [8]Martínez-Nicolás (2021), [9]Moore et al (2022), [10]Nicosia et al. (2022)

r.l.vandenberg@amsterdamumc.nl

(1)Alzheimer Center Amsterdam, Neurology, Vrije Universiteit Amsterdam, Amsterdam UMC location VUmc, Amsterdam, Netherlands, (2)Amsterdam Neuroscience, Neurodegeneration, Amsterdam, Netherlands, (3) Department of Clinical, Neuro and Developmental Psychology, VU University, Amsterdam, (4)Winterlight Labs, Toronto, ON, Canada, (5)Metis Cognition Ltd., Kilmington, United Kingdom, (6)Massachusetts General Hospital, Harvard Medical School, Boston, MA, USA, (7)Department of Radiology & Nuclear Medicine, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam UMC, Amsterdam, Netherlands, (8)Neurochemistry Laboratory, Department of Laboratory Medicine, Amsterdam Neuroscience, Vrije Universiteit Amsterdam, Amsterdam UMC, Amsterdam, Netherlands