

Item-specific Vocal Acoustic Markers From the Mini-Mental State Examination for Cognitive Impairment

P3-06

Hyunsun Ham^{1,2}, Seung Jae Choi³, Hee Won Yang⁴, Jeong Lan Kim⁴, Jun-Young Lee², Minwoo Cho^{3,5}, So Yeon Jeon²

1 Interdisciplinary Program in Cognitive Science, Seoul National University, Seoul, Republic of Korea
2 Department of Psychiatry, SMG-SNU Boramae Medical Center, Seoul, Republic of Korea
3 Department of Transdisciplinary Medicine, Seoul National University Hospital, Seoul, Republic of Korea
4 Department of Psychiatry, School of Medicine, Chungnam National University, Daejeon, Republic of Korea
5 Department of Medicine, Seoul National University College of Medicine, Seoul, Republic of Korea

KEY TAKEAWAY

- * VOICE BIOMARKERS OF COGNITIVE IMPAIRMENT ARE ITEM-SPECIFIC AND MAP ONTO THE COGNITIVE DEMANDS OF INDIVIDUAL MMSE ITEMS.
- * TEMPORAL ORIENTATION ALONE, A SINGLE PROMPT REQUIRING ONLY SECONDS OF SPEECH, ACHIEVED AN AUC OF 0.94 FOR AD DETECTION.

INTRODUCTION

- Vocal acoustic features are promising markers of subtle cognitive decline, with diagnostic performance comparable to standard cognitive assessments.
- Standard assessments such as the MMSE primarily rely on response-based scoring, limiting the capture of acoustic information embedded in how patients produce their responses.
- Prior studies have typically analyzed speech at the test level by aggregating features across all items, rather than examining item-level acoustic patterns.
- No study has examined diagnostic classification using acoustic features from individual items within a standardized neuropsychological test, leaving item-specificity largely unknown.

METHODS

- Speech was collected from 927 Korean older adults across four Korean medical centers, including 408 cognitively normal participants, 228 with mild cognitive impairment, and 291 with Alzheimer's disease.
- Voice was recorded across eight MMSE items: temporal orientation, spatial orientation, backward recitation, delayed recall, serial subtraction, naming, repetition, and judgment.
- Twenty-seven acoustic features were extracted from 21,337 utterances, including prosodic features, voice quality features such as jitter, shimmer, and HNR, and spectral features including MFCCs.
- Binary classifiers were trained and evaluated across individual MMSE items, item combinations, and full-test aggregation.
- External validation was performed using a center hold-out design with Konkuk University Hospital as the held-out center (n = 269), and AUCs were estimated with bootstrap 95% confidence intervals using 1,000 iterations.

Table 1. Participant characteristics (mean ± SD; all p < 0.001)

Group	N	Age	Educational years	MMSE
CN	408	66.2 ± 9.5	11.0 ± 4.2	26.8 ± 2.6
MCI	228	75.3 ± 7.6	8.4 ± 4.6	22.2 ± 4.5
AD	291	79.3 ± 7.3	7.0 ± 4.7	15.9 ± 5.3

RESULTS

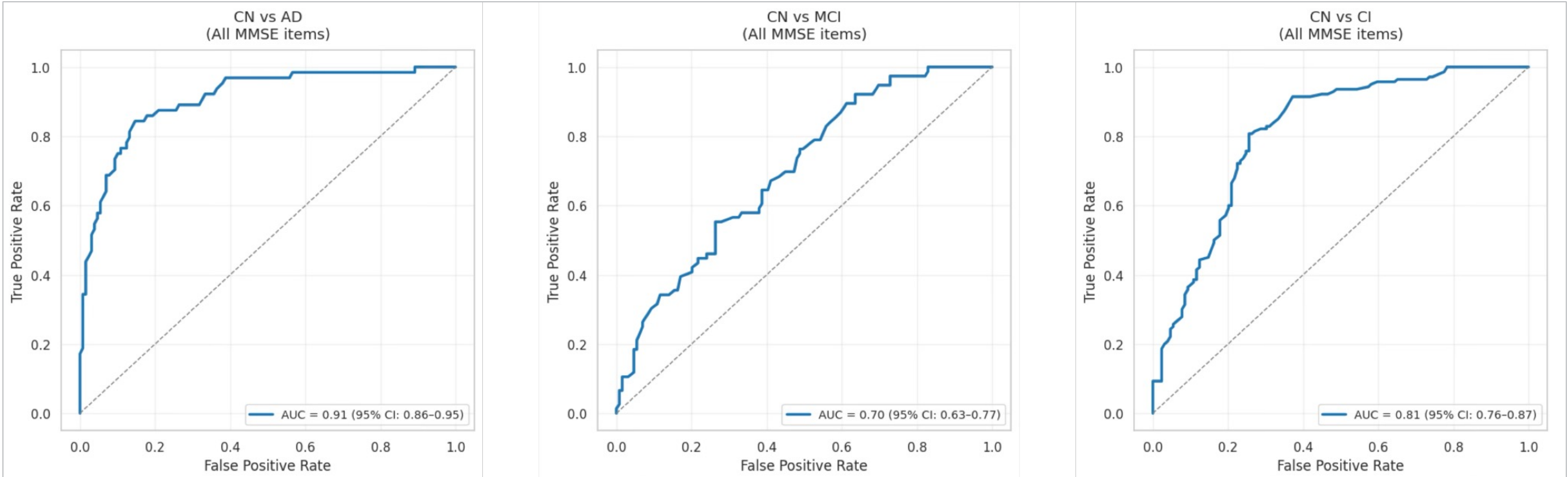


Fig 1. ROC Curves for features aggregated across all MMSE items

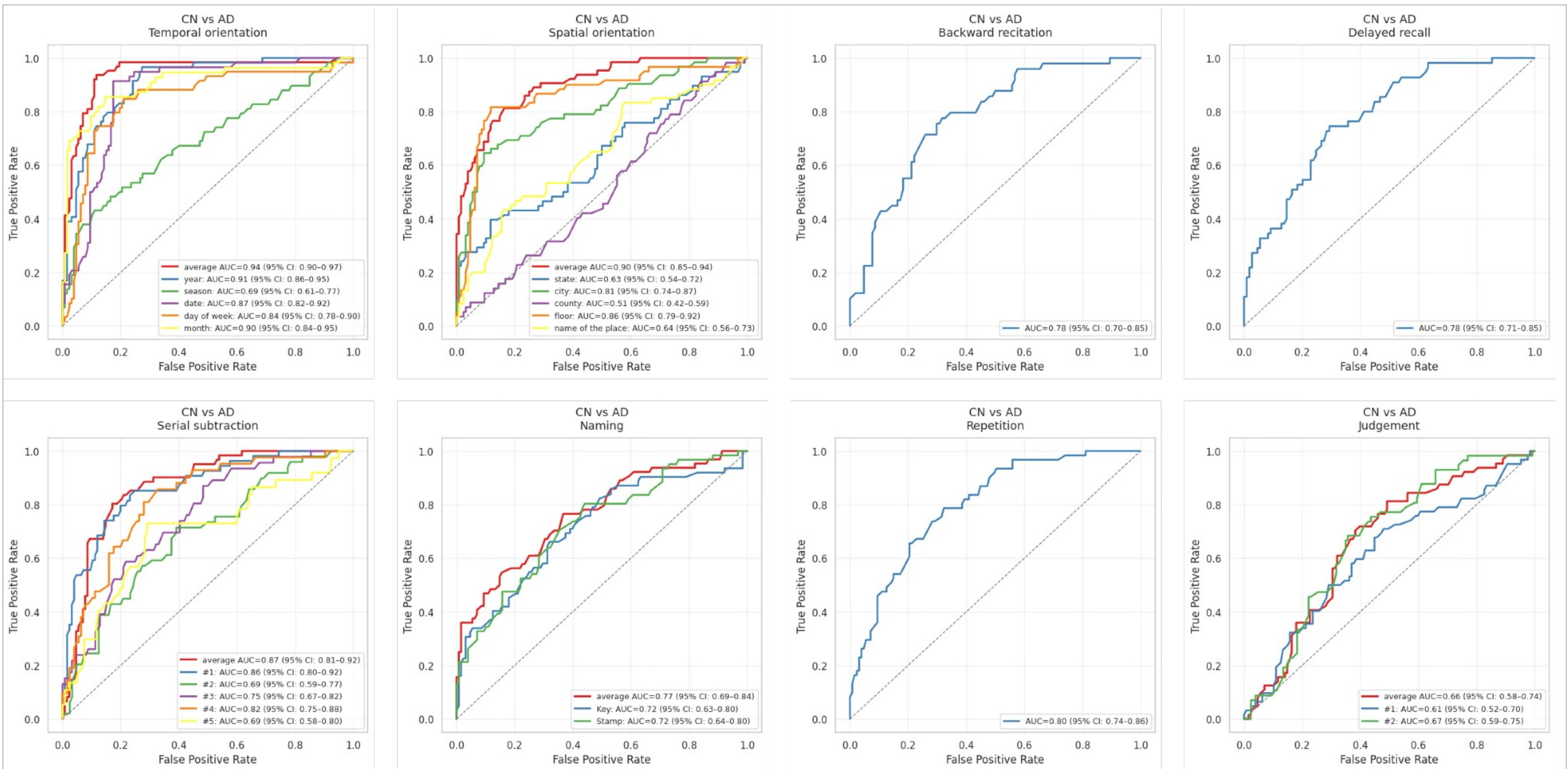


Fig 2. Item-level ROC across eight MMSE items (CN vs AD)

- Aggregation across all MMSE items yielded Random Forest AUCs of 0.91 for CN vs AD, 0.70 for CN vs MCI, and 0.81 for CN vs CI.
- Temporal orientation showed the most robust discrimination across diagnostic contrasts, with AUCs of 0.94 for CN vs AD, 0.73 for CN vs MCI, and 0.86 for CN vs CI, followed by spatial orientation and serial subtraction.
- Delayed recall was particularly informative for MCI classification, achieving an AUC of 0.68.

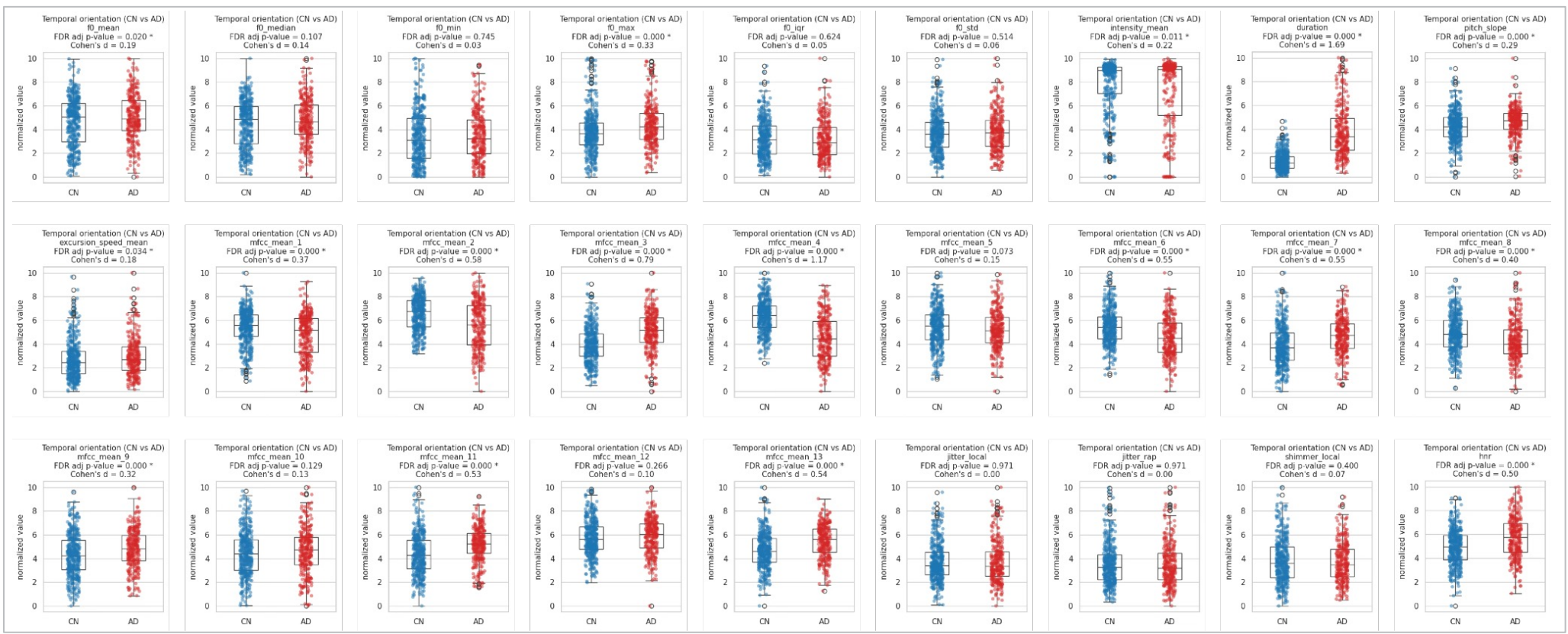


Fig 3. Group-wise acoustic feature differences for temporal orientation (CN vs AD)

- In temporal orientation, response duration showed the largest effect sizes, particularly for CN vs AD (d = 1.69; all adjusted p < 0.001), while mean MFCC 4 showed the strongest spectral separation.
- In delayed recall, spectral features, particularly MFCC 5 and MFCC 4, were the primary contributors to classification, whereas voice quality features such as jitter and shimmer were significant only for CN vs AD.

CONCLUSIONS

- Speech elicited during MMSE administration contains item-specific acoustic markers that reflect the cognitive demands of individual test items.
- Temporal orientation alone achieved diagnostic performance comparable to full-test aggregation, suggesting that a short, targeted voice sample may be sufficient for initial screening.
- Item-level acoustic analysis may enable more efficient and interpretable voice-based screening for cognitive impairment.
- The use of a large multicenter Korean cohort with external validation supports the generalizability of this approach.
- These findings suggest that MMSE-derived acoustic markers may be adaptable to telephone-based assessment and community outreach settings.

REFERENCES

1. Martinez-Nicolás J, et al. Ten years of automatic voice and speech analysis in AD and MCI: a systematic review. Front Psychol. 2021;12:620251.
2. García-Gutiérrez F, et al. Machine learning-based speech analysis in the AD spectrum. Alzheimers Res Ther. 2024;16:26.
3. Heltz C, Engler T, Langer N. Automated analysis of spontaneous speech approximates cognitive domain scores in older adults. NPJ Digit Med. 2026.
4. König A, et al. Automatic speech analysis for assessment of predementia and AD. Alzheimers Dement (Amst). 2015;1:112-124.
5. Yeung A, et al. Correlating NLP and automated speech analysis with clinician assessment in MCI and AD. Alzheimers Res Ther. 2021;13:109.
6. Albert MS, et al. The diagnosis of MCI due to AD: NIA-AA workgroup recommendations. Alzheimers Dement. 2011;7:270-279.
7. Folstein MF, et al. Mini-Mental State: a practical method for grading cognitive state. J Psychiatr Res. 1975;12:189-198.

ACKNOWLEDGEMENTS

This work was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF), funded by the Ministry of Education (RS-2023-00210380 & RS-2026-25476550), and the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (RS-2025-25455095).