

Acoustic Voice Markers Within Different Languages And Speech Elicitation Tasks: An Exploratory Study

Stephanie Li, MPH^{1,2}, Francesca Marino, PhD¹, Cody Karjadi^{1,2}, Katherine A. Gifford, PsyD^{1,2}, Honghuang Lin, PhD³, Rhoda Au^{1,2}, PhD, Huitong Ding, PhD^{1,2}

¹Anatomy & Neurobiology, Boston University Chobanian & Avedisian School of Medicine, ²Framingham Heart Study, ³University of Massachusetts Chan Medical School

Background:

Human speech represents a potential scalable and non-invasive measure of cognition, but cross-language and task heterogeneity limits translation to clinical settings. This study explores whether **(1) acoustic features differ between two speech corpora** and **(2) the associations between acoustic features and cognitive status differ across a picture description task in English versus a reading-paragraph task in Spanish.**

Methods:

We included two corpora from DementiaBank, Pitt (English) and Ivanova (Spanish), with participants aged ≥ 65 years who had available audio recordings and demographic data. Cognitive status was harmonized across corpora as cognitively intact and cognitively impaired (CI). Speech was recorded during the Cookie Theft picture description task (Pitt) and reading the first two lines of *Don Quixote* (Ivanova). Sixty-five ComParE 2016 low-level descriptor acoustic features were extracted from both datasets using OpenSMILE. First, between-cohort differences in acoustic features were evaluated using age-, sex-, and education-adjusted means. Second, logistic regression models estimated associations between each acoustic feature and cognitive status adjusting for age, sex, education, and cohort with secondary regression models stratified by cohort. All statistical tests were corrected for false discovery rate (FDR).

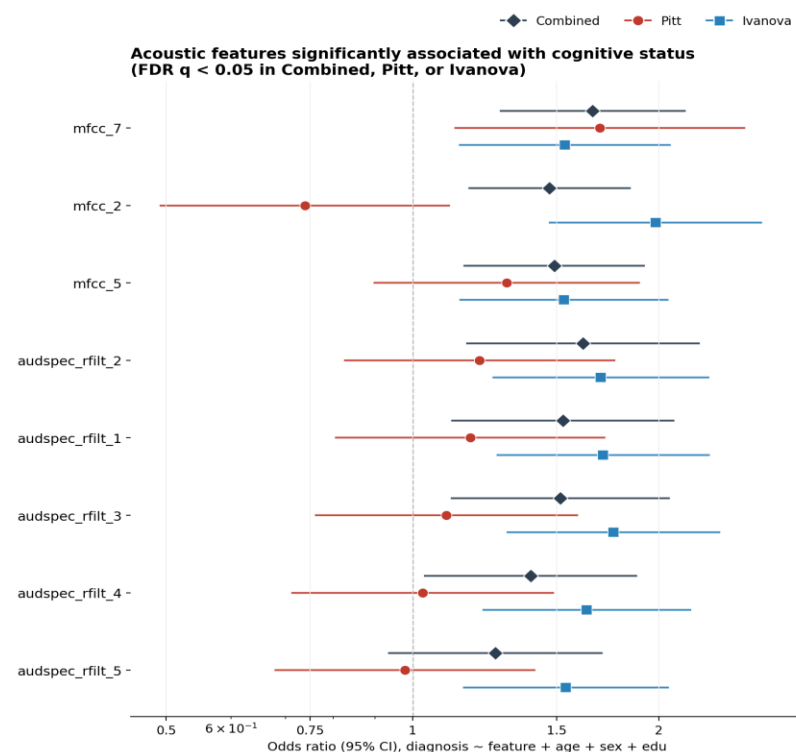
Table 1: Demographics

	Pitt Corpus (N = 196)	Ivanova Corpus (N = 262)
Age, years		
Mean (SD)	73.5 (6.0)	78.3 (7.5)
Sex, N (%)		
Male	63 (33.2)	81 (30.9)
Female	127 (66.8)	181 (69.1)
Education, years		
Mean (SD)	12.5 (3.1)	8.7 (3.8)
Cognitive status, N (%)		
Cognitively intact	49 (25.8)	145 (55.3)
Cognitively impaired	141 (74.2)	117 (44.7)

Table 2: Association of acoustic features with cognitive status across cohorts.

Feature	Group	OR	95% CI	p-value	FDR q
mfcc_7	Combined	1.66	1.28-2.15	<0.001	0.01
mfcc_7	Pitt	1.69	1.13-2.54	0.01	0.36
mfcc_7	Ivanova	1.53	1.14-2.06	0.01	0.04
mfcc_2	Combined	1.47	1.17-1.84	<0.001	0.03
mfcc_2	Pitt	0.74	0.49-1.11	0.14	0.99
mfcc_2	Ivanova	1.98	1.47-2.67	<0.001	<0.001
mfcc_5	Combined	1.49	1.16-1.92	0.002	0.04
mfcc_5	Pitt	1.3	0.90-1.89	0.16	0.99
mfcc_5	Ivanova	1.53	1.14-2.05	0.004	0.04
audspec_rfilt_2	Combined	1.62	1.17-2.24	0.004	0.07
audspec_rfilt_2	Pitt	1.21	0.83-1.76	0.33	0.99
audspec_rfilt_2	Ivanova	1.7	1.26-2.30	<0.001	0.01
audspec_rfilt_1	Combined	1.53	1.12-2.08	0.01	0.07
audspec_rfilt_1	Pitt	1.18	0.81-1.72	0.40	0.99
audspec_rfilt_1	Ivanova	1.71	1.27-2.30	<0.001	0.01
audspec_rfilt_3	Combined	1.52	1.12-2.06	0.01	0.10
audspec_rfilt_3	Pitt	1.1	0.76-1.59	0.61	0.99
audspec_rfilt_3	Ivanova	1.76	1.31-2.37	<0.001	0.01
audspec_rfilt_4	Combined	1.39	1.04-1.88	0.03	0.18
audspec_rfilt_4	Pitt	1.03	0.71-1.48	0.88	0.99
audspec_rfilt_4	Ivanova	1.63	1.22-2.18	<0.001	0.01
audspec_rfilt_5	Combined	1.26	0.94-1.70	0.13	0.41
audspec_rfilt_5	Pitt	0.98	0.68-1.41	0.91	0.99
audspec_rfilt_5	Ivanova	1.54	1.16-2.05	0.003	0.03

Fig 1: Forest plot distribution of acoustic features significantly associated with cognitive status across cohorts



Results:

The Pitt corpus included 196 participants (mean age 73.5 ± 6.0 years; 66.3% female; education 12.4 ± 3.1 years; 147 [75%] CI), and the Ivanova corpus comprised 262 participants (mean age 78.3 ± 7.5 years; 69.1% female; education 8.7 ± 3.8 years; 117 [44.7%] CI). 78% of acoustic features differed between cohorts. 44 features had higher mean values in Pitt than Ivanova and 9 had lower values (all $P_{FDR} < 0.05$). In the combined sample, 3 acoustic features were related to cognitive status ($P_{FDR} < 0.05$). When stratified by cohort, 8 acoustic features were associated with cognitive status in Ivanova ($P_{FDR} < 0.05$), of which 3 were associated with higher odds of CI. No features were associated with odds of CI in Pitt ($P_{FDR} < 0.05$).

Conclusion:

This exploratory analysis suggests differences in acoustic features across two cohorts with different speech tasks, spoken language, and clinical characteristics.

Future studies will explore the nuances of task, language, and cognitive status on speech features.

References:

- DementiaBank (Pitt corpus): Becker, J. T., Boller, F., Lopez, O. L., Saxton, J., & McGonigle, K. L. (1994). The natural history of Alzheimer's disease: description of study cohort and accuracy of diagnosis. *Archives of Neurology*, 51(6), 585-594. Note: Please also acknowledge this grant support for the Pitt corpus -- NIA AG03705 and AG05133.
- DementiaBank (general): Lanzi, A. M., Saylor, A. K., Fromm, D., Liu, H., MacWhinney, B., & Cohen, M. (2023). DementiaBank: Theoretical rationale, protocol, and illustrative analyses. *American Journal of Speech-Language Pathology*, 32(2), 426-438. doi.org/10.1044/2022_AJSLP-22-00281
- Ivanova, O., Meilán, J.G., Martínez-Sánchez, F., Martínez-Nicolas, I., Llorente, T.E., & González, N.C. (2022). Discriminating speech traits of Alzheimer's disease assessed through a corpus of reading task for Spanish language. *Computer Speech & Language*. doi: https://doi.org/10.1016/j.csl.2021.101341