Abstract

Voice Biomarkers in Cognitive Impairment: A Stratified Approach to Linguistic Health Applications

Luwen Cao, Lei Feng, Yilin Wu, Zhiming Bao
DigiPen Institute of Technology
National University of Singapore

Keywords: Voice biomarkers, cognitive impairment, health applications, acoustic analysis, population differences

Background and Motivation: Voice analysis represents a promising intersection of linguistic research and clinical applications, offering non-invasive approaches to health monitoring that could reduce diagnostic costs and improve accessibility. However, traditional voice biomarker research has employed uniform analytical approaches that may not capture the full complexity of linguistic variation, potentially limiting diagnostic accuracy in diverse populations.
Objectives: This study investigates differential patterns in voice parameters among individuals with cognitive impairment using stratified analytical frameworks. We hypothesized that linguistic markers of cognitive decline would manifest through distinct pathways, necessitating more nuanced approaches to voice-based health assessment.
Methods: We conducted a cross-sectional analysis of 150 participants (75 controls, 38 amnestic MCI, 37 non-amnestic MCI) with balanced demographic distribution. Voice recordings were obtained using standardized protocols in controlled acoustic environments. Acoustic parameters were extracted from sustained vowel phonations. Statistical analysis employed assumption-based test selection followed by pre-planned stratified analyses to examine population-specific patterns.
Results: Combined analysis revealed major voice parameters significantly differentiated cognitive groups. However, stratified analysis uncovered a remarkable finding: **complete divergence between demographic subgroups with zero overlapping significant variables**. One subgroup demonstrated 12 significant biomarkers (75% of total findings) predominantly involving voice instability measures with large effect sizes (η^2 up to 0.229). The other subgroup showed only 1 significant biomarker—median pitch elevation in non-amnestic MCI (η^2=0.083, p=0.043)—suggesting fundamentally different pathophysiological mechanisms.
Linguistic Implications: The stratified patterns reveal that cognitive impairment affects linguistic production through distinct pathways: neuromotor control disruption in one population versus fundamental frequency modulation in another. These finding challenges traditional approaches to voice-based linguistic analysis and demonstrates

the critical importance of considering demographic factors in language-health applications.

Innovation and Applications: Our findings represent a paradigm shift in voice biomarker research, demonstrating that effective clinical applications require population-specific diagnostic algorithms. The complete absence of overlap between demographic subgroups suggests that mixed-population training data may systematically miss critical linguistic indicators.

Theoretical Contributions: This work bridges theoretical linguistics and clinical applications by demonstrating how demographic factors influence the linguistic manifestation of cognitive pathology.

Practical Applications: Results inform the development of stratified voice analysis tools for cognitive screening, potentially improving diagnostic accuracy and reducing bias in health applications. The pitch elevation pattern offers a novel early detection marker for non-amnestic cognitive impairment in specific populations.

Future Directions: This research opens new avenues for investigating population-specific linguistic biomarkers across various health conditions, supporting the development of personalized approaches to voice-based health monitoring and the integration of demographic considerations into computational linguistics applications.

Conclusion: Voice biomarkers for cognitive assessment exhibit complete population-specific patterns, challenging current theoretical frameworks and practical applications. This finding necessitates fundamental changes in how linguistic health applications are conceptualized, developed, and clinically implemented, while highlighting the critical importance of stratified analysis in linguistic research.