

## Automated Acoustic and Lexical-Semantic Analysis of Digitized Speech Samples in Aging and Neurodegenerative Disease

M. Grossman\*#, N. Nevler, S. Cho, M. Liberman  
Departments of Neurology and Linguistics, University of Pennsylvania

Analyses of semi-structured speech samples are highly informative but notoriously challenging because of laborious analyses, inconsistent inter-rater reliability and ambiguous interpretation of target material. Important insights into speech have been fostered by analyses of very large datasets of healthy speakers, and these insights were possible only with the development of automated analytic techniques. We have been adapting these automated methods for analyzing digitized speech samples in a targeted manner that is suitable for understanding the neurobiology of speech and monitoring disease-modifying treatment trials of neurologically-impaired patients. We illustrate analyses of digitized speech during picture descriptions in sex-matched healthy seniors (n=37; mean  $\pm$ S.D. age: 68.5  $\pm$ 7.8) compared to younger adults (n=84; mean  $\pm$ S.D. age: 20.4  $\pm$ 3.7), and characterize speech in non-fluent/agrammatic variant primary progressive aphasia (naPPA, n=25) and semantic variant PPA (svPPA, n=33). Analyzed acoustic properties include fundamental frequency (f0) range, total speech time, speech segment length, pause length, pause number, and pause rate; analyzed lexical properties include part of speech analysis and analysis of semantic properties of nouns. We found that healthy seniors produce significantly shorter speech segments (p=.009), more frequent pauses (p=.03) and longer pause lengths (p=.001) relative to younger speakers, even though total speech time is matched across groups. The groups did not differ in their f0 range. Comparisons of lexical properties revealed that seniors produce fewer prepositions (p=.001) and determiners (p=.002), but greater numbers of interjections (p=.04), pronouns (p=.001) and verbs (p=.005). Seniors produce more familiar nouns (p=.003), but otherwise match younger speakers in concreteness, ambiguity, frequency, and age of acquisition. These findings are consistent with age-associated differences in speech organization, where senior picture descriptions consist of shorter, grammatically simpler utterances with a smaller number of more familiar concrete nouns. Findings such as these are essential for interpreting changes in speech seen in neurodegenerative diseases such as PPA. Patients with naPPA have distinct acoustic and lexical properties of speech compared to healthy seniors, including significantly restricted f0 range, as well as more nouns, fewer verbs and more speech errors. This is consistent with their effortful, agrammatic speech and apraxia of speech. By comparison, patients with svPPA have distinct lexical and semantic properties of speech compared to healthy seniors, including significantly more pronouns, consistent with their word-finding difficulties, and nouns that are significantly more abstract, familiar and ambiguous, consistent with their semantic deficit of visual feature knowledge of concrete object concepts.

Supported by NIH (AG054519, AG017586, Alzheimer's Association, and Institute on Aging at the University of Pennsylvania)