

52. Classification of Spontaneous Speech of Individuals with Dementia Based on Automatic Prosody Analysis Using SVM

Roelant Ossewaarde^{1,2}, Roel Jonkers¹, Fedor Jalvingh^{1,3} and Roelien Bastiaanse¹

¹ Center for Language and Cognition, University of Groningen, r.a.ossewaarde@rug.nl, ² HU University of Applied Science, Institute for ICT, Utrecht,

³ Albertinen-Krankenhaus Hamburg, Neurologie, Germany

Introduction Brain damage as caused by dementia may affect the prosody, fluency and voice quality of speech. These can be measured in the acoustic signal, and so are easy to compute. We studied whether we can predict dementia, and its type, based on automated analyses of everyday conversations with individuals with Parkinson's Disease, Alzheimer's Disease or Primary Progressive Aphasia.

Methods & participants

We used 22 hours of speech from 9 spontaneous conversations ($\mu=5m47s \pm 2m30s$) spread over three different moments in time. Participants were non-brain-damaged speakers ($n=7$) and patients with a clinical diagnosis of: Probable Alzheimer's disease (AD, $n=9$), Non-fluent primary progressive aphasia (PPA-NF, $n=2$), semantic dementia (PPA-SD, $n=1$), Parkinson's disease (PD, $n=6$). PD and AD typically cause prosodic and voice quality problems; PPA-NF typically causes disfluency instead.

Speech was analyzed using our R-implementation of a Voice Activity Detection algorithm (Ramirez et al., 2004) to detect the acoustic envelope (cf. Fig. 1), with a custom decision procedure to capture the different types of pauses of the speaker (fluency). For prosodic variation, we computed the spread of the local and global pitch trends (Matteson et al., 2013). For voice quality, we measured the harmonicity of the signal with the Cepstral Peak Prominence measure (CPPS, Hillenbrand and Houde 1996).

A Support Vector Machine classifier (Cortes and Vapnik 1995) was used to measure the predictive value of the discovered patterns.

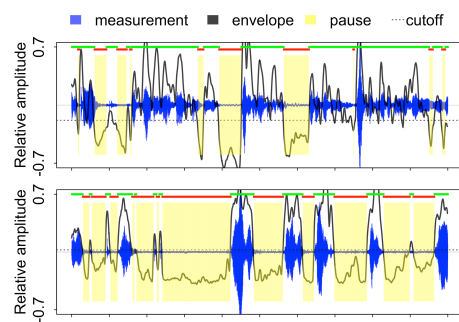
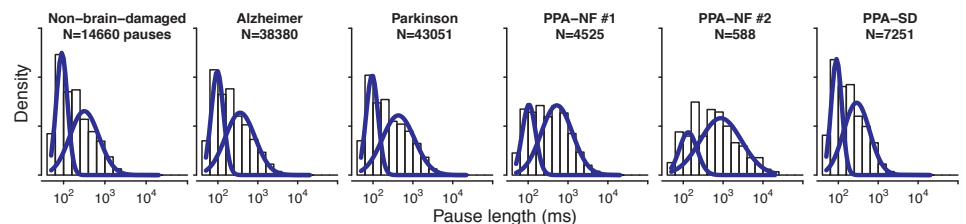


Fig. 1: Output of our pause detection algorithm, exemplified on fragments of NBD (top) and PPA-NF (bottom) speech.

Fig. 2: Distribution of pause durations, with fitted Gaussians overlayed.



Results (1) - Fluency

A sum of 2 distinct Gaussians, each representing a pause category, provides a good fit for the data (cf. Fig. 2). The pause patterns for non-brain-damaged speakers, and those with AD, PD, SD are remarkably similar, with a short pause at about 100 ms and a long pause at around 350 ms. Speakers with the non-fluent kind of primary progressive aphasia lack the distinctive short pause peak, but have relatively more medium-to-long pauses.

Results (2) - Prosody

Groupwise differences of parameters between individuals with and without neurodegenerative diseases were evaluated through a side-by-side comparison using the Wilcoxon rank sum test. At 95%, the tests indicated that the range in segments of NBDs is not different from the range in any of the other participant groups.

Results (3) - Voice Quality

The descriptive statistics of the cepstral measure across the three topics and five groups show that there were no significant differences between mean CPPS scores (where $p < .05$) across groups: $F(6,34) = 0.86, p = 0.53$, or topics: $F(2,68) = 2.26, p = 0.11$, or their interaction: $F(12, 68) = 1.69, p = 0.08$.

Conclusion

The PPA-NF patients in this study pause their speech longer. Both brief and long pauses last longer than those of controls.

The distribution of pauses over these categories is marked in PPA-NF speech: it has more long pauses. This pattern arises after log-normalization of the data.

The classifier can distinguish speech from a dementia type associated with disfluency, but not from types with decreased pitch range or voice quality.

Our follow-up study will include word- and sentence-level variables as well.

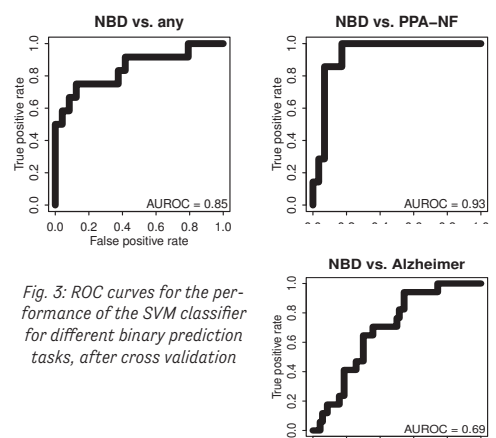


Fig. 3: ROC curves for the performance of the SVM classifier for different binary prediction tasks, after cross validation

References

Baken & Orlikoff (2000), *Clin. Measurement of Speech and Voice*.
Gorno-Tempini et al. (2011) *Neurology* 76 (11).

