Measurement of speech parameters in casual speech of dementia patients

Roelant Ossewaarde[1,2], Roel Jonkers[1], Fedor Jalvingh[1,3], and Roelien Bastiaanse[1]

[1]Center for Language and Cognition Groningen (CLCG), Rijksuniversiteit Groningen, Netherlands, [2]HU University of Applied Science, Utrecht, Netherlands, [3]St. Marienhospital - Vechta, Geriatric Clinic Vechta, Germany

Introduction

Individuals with dementia often experience a decline in their ability to use language. Language problems have been reported in individuals with dementia caused by Alzheimer's disease, Parkinson's disease or degeneration of the fronto-temporal area.

Acoustic properties are relatively easy to measure with software, which promises a cost-effective way to analyze larger discourses. We study the usefulness of acoustic features to distinguish the speech of German-speaking controls and patients with dementia caused by (a) Alzheimer's disease, (b) Parkinson's disease or (c) PPA/FTD. Previous studies have shown that each of these types affects speech parameters such as prosody, voice quality and fluency (Schulz 2002; Ma, Whitehill, and Cheung 2010; Rusz et al. 2016; Kato et al. 2013; Peintner et al. 2008).

Prior work on the characteristics of the speech of individuals with dementia is usually based on samples from clinical tests, such as the Western Aphasia Battery or the Wechsler Logical Memory task. Spontaneous day-to-day speech may be different, because participants may show less of their vocal abilities in casual speech than in specifically designed test scenarios. It is unclear to what extent the previously reported speech characteristics are still detectable in casual conversations by software.

The research question in this study is: how useful for classification are acoustic properties measured in spontaneous speech.

Methods

Participant recruitment and data

The speech data used in this study was collected during a larger study of processing of verbs and nouns in speakers with different types of dementia, currently performed by one of the co-authors (FJ). Participant recruitment, data elicitation and manual CLAN-annotation were performed in the context of that study. Spontaneous speech fragments were elicited from German controls (n=7) and patients with a clinical diagnosis of a form of dementia: (probable) Alzheimer's disease (AD, n=9), PPA (n=3), bvFTD (n=4), Parkinson's disease (PD, n=6), PD with MCI (n=4), PD with dementia (n=3). In this study, only data on controls and participants diagnosed with PPA, AD or PD are reported.

For each participant, discourses on three different topics (past, present, future) were elicited. Because the ultimate goal of the larger study is to track the long-time decline of the linguistic system in non-controls, the elicitation of the three topics was repeated three times with non-controls, with about 6 months between each elicitation session.

Narrative sampling

The interviewer asked participants in separate sessions to speak of childhood memories (topic: past), of a typical day in the present (topic: present), and of plans that they might have for the next week, month or year (topic: future).

Elicitation was done in the participant's own environment. This affects recording quality: background noises are present in the signal, such as children playing, telephones ringing and papers being shuffled. Segments of the interviewer giving instructions or asking questions were removed, but only if they are of significant length and truly interrupt the flow of the discourse of the participant. This was judged by the researcher (RO). The resulting discourses (μ=6m47s; σ=3m30s) are of sufficient length that minor utterances by the interviewer to move the discourse along (*hmm-hmm*, *oh yeah*, etc.) do not significantly impact the data analysis of the speaker's voice characteristics.

Acoustic feature extraction

Audio recordings were analyzed for voice activity using an unsupervised learning framework (Ying et al. 2011), and for pitch using an automatic pitch extraction algorithm implemented in REAPER[1].

Results of the analyses were stored in a database and then read in by R-scripts for further statistical analysis. Reaper's output (in Hz) was translated to pitch interval (in cents, $\mathbb{P}^0$) as proposed by Matteson, Olness, & Caplow (2013).

The following variables were used for analysis:

1. Fluency
   1. Pause length
   2. Pause frequency
2. Phonation
   1. Duration of speech
3. Voice quality
   1. Jitter
   2. Shimmer
4. Prosody
   1. Pitch level
      1. mean, median, maximum and minimum $\mathbb{P}^0$
   2. Pitch range
      1. SD
      2. Four standard deviations around the mean (SD4)
      3. Max-min $\mathbb{P}^0$
      4. The difference between the 95th and the 5th percentile (HDI, 90% span)
      5. The difference between the 16th and the 84th percentile (HDI 68% span)
      6. Skewness and kurtosis

Machine learning

We trained a generalized linear multilevel model using R and Stan (McElreath, 2016; R Core Team, 2017) and evaluated its performance. Evaluation was conducted using five-fold cross-validation over the set of fragments. In each of the five folds, the parameters of the model were first learned in a

---

[1] David Talkin, https://github.com/google/REAPER.

training phase using 80% of the data, and then applied to the held-out data to predict the participant's diagnosis. This procedure is repeated for each of the five folds, with accuracy being the average performance on the test data across all folds.

Results

Results are compared to a baseline ("Zero Rule") strategy that always predicts the majority class. The classifier is considered informative if it performs better than the baseline strategy. Machine learning results suggest that the proposed model is superior to the baseline standard of predicting the majority class, measured as the area under curve (AUROC), cf. figure 1. Individual univariate Wilcoxon rank sum tests, adapted with Benjamini-Hochberg correction for false discovery rate (Benjamini and Hochberg 1995), show that SD patients have significantly shorter pauses than controls, and PD patients have significantly lower values for voice quality parameters than controls.

Discussion

Post-hoc analyses show that most influence in the model comes from fluency and voice quality variables, while prosody variables contribute the least. Monopitch has frequently been associated with dementia speech, but the role of pitch is very limited in the model for this convenience sample. A possible explanation is that casual spontaneous speech invites less pitch variation, both in controls and in patients.

References

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. Journal of the Royal Statistical Society. Series B (Methodological), 289–300.

Gorno-Tempini, M. L., Dronkers, N. F., Rankin, K. P., Ogar, J. M., Phengrasamy, L., Rosen, H. J., … Miller, B. L. (2004). Cognition and anatomy in three variants of primary progressive aphasia. Annals of Neurology, 55(3), 335–46.

Kato, S., Endo, H., Homma, A., Sakuma, T., & Watanabe, K. (2013). Early detection of cognitive impairment in the elderly based on bayesian mining using speech prosody and cerebral blood flow activation. Conf Proc IEEE Eng Med Biol Soc, 2013, 5813–6.

Ma, J. K.-Y., Whitehill, T., & Cheung, K. S.-K. (2010). Dysprosody and stimulus effects in Cantonese speakers with Parkinson's disease. Int J Lang Commun Disord, 45(6), 645–55.

Matteson, S. E., Olness, G. S., & Caplow, N. J. (2013). Toward a quantitative account of pitch distribution in spontaneous narrative: method and validation. The Journal of the Acoustical Society of America, 133(5), 2953–71.

McElreath, R. (2016). Rethinking: Statistical rethinking book package.

Mesulam, M. M. (2001). Primary progressive aphasia. Ann Neurol, 49(4), 425–32.

Neary, D., Snowden, J. S., Gustafson, L., Passant, U., Stuss, D., Black, S., … Benson, D. F. (1998). Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria. Neurology, 51(6), 1546–54.

Peintner, B., Jarrold, W., Vergyriy, D., Richey, C., Tempini, M. L. G., & Ogar, J. (2008). Learning diagnostic models using speech and language measures. Conference Proceedings : Annual

International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference, 2008, 4648–51.
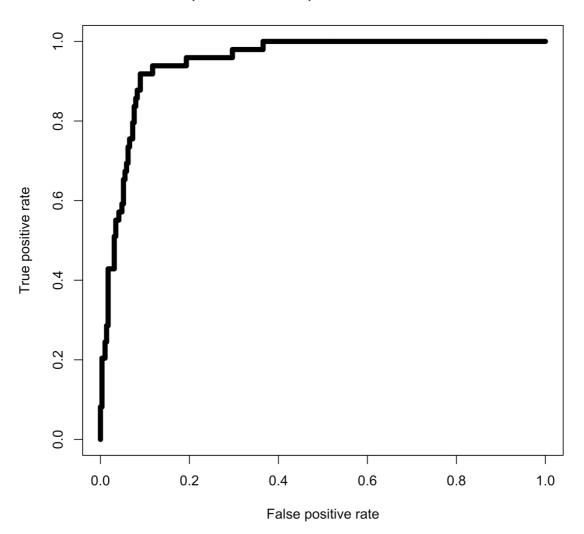
R Core Team. (2017). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from https://www.R-project.org/

Rusz, J., Tykalová, T., Klempíř, J., Čmejla, R., & Ružička, E. (2016). Effects of dopaminergic replacement therapy on motor speech disorders in Parkinson's disease: Longitudinal follow-up study on previously untreated patients. J Neural Transm (Vienna), 123(4), 379–87.

Schulz, G. M. (2002). The effects of speech therapy and pharmacological treatments on voice and speech in parkinson s disease: A review of the literature. Curr Med Chem, 9(14), 1359–66.

Ying, D., Yan, Y., Dang, J., & Soong, F. K. (2011). Voice activity detection based on an unsupervised learning framework. IEEE Transactions on Audio, Speech, and Language Processing, 19(8), 2624–2633.

Figures

ROC curve (AUROC = 0.68) for controls versus dementia

Figure 1. Area under Receiver Operating Characteristics for the binary classification of controls versus dementia.