Automated detection of unfilled pauses in speech of healthy and brain-damaged individuals

Ossewaarde, Roelant; Jonkers, Roel; Jalvingh, Fedor; Bastiaanse, Yvonne

Automated detection of unfilled pauses in speech of healthy and brain-damaged individuals

Roelant Ossewaarde[a,b], Roel Jonkers[a], Fedor Jalvingh[a,c], Roelien Bastiaanse[a]

[a]Center for Language and Cognition, University of Groningen; [b]Institute for ICT, HU University of Applied Science, Utrecht; [c]St. Marienhospital Vechta, Geriatric Clinic Vechta, Germany

1. Introduction

Pauses in speech may be categorized on the basis of their length. Some authors claim that there are two categories (short and long pauses) (Baken & Orlikoff, 2000), others claim that there are three (Campione & Véronis, 2002), or even more.

Pause lengths may be affected in speakers with aphasia. Individuals with dementia probably caused by Alzheimer's disease (AD) or Parkinson's disease (PD) interrupt speech longer and more frequently. One infrequent form of dementia, non-fluent primary progressive aphasia (PPA-NF), is even defined as causing speech with an unusual interruption pattern ("hesitant and labored speech").

Although human listeners can often easily distinguish pathological speech from healthy speech, it is unclear yet how software can detect the relevant patterns. The research question in this study is: how can software measure the statistical parameters that characterize the disfluent speech of PPA-NF/AD/PD patients in connected conversational speech?

2. Methods

We used speech data 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). A total of nine spontaneous conversations at three different moments in time were held with participants from different groups:

(1) Non-brain-damaged individuals (n=7).
(2) patients with a clinical diagnosis of a form of dementia:
    (a) Probable Alzheimer's disease (n=9).
    (b) Non-fluent primary progressive aphasia (PPA-NF, n=2)
    (c) Semantic dementia (PPA-SD, n=1)
    (d) Parkinson's disease (n=6).
    (e) Behavioral fronto-temporal dementia (n=4).
    (f) Parkinson's disease with minor cognitive impairment (n=4).
    (g) Parkinson's Disease with dementia (n=3).

The average conversation length was 5m47s (± 2m30s). The 22 hours of speech were automatically analyzed for speech and pauses using our own R-implementation of the Voice Activity Detection algorithm proposed by Ramırez, Segura, Benıtez, De La Torre, and Rubio (2004)

CONTACT Roelant Ossewaarde. Email: r.a.ossewaarde@hu.nl

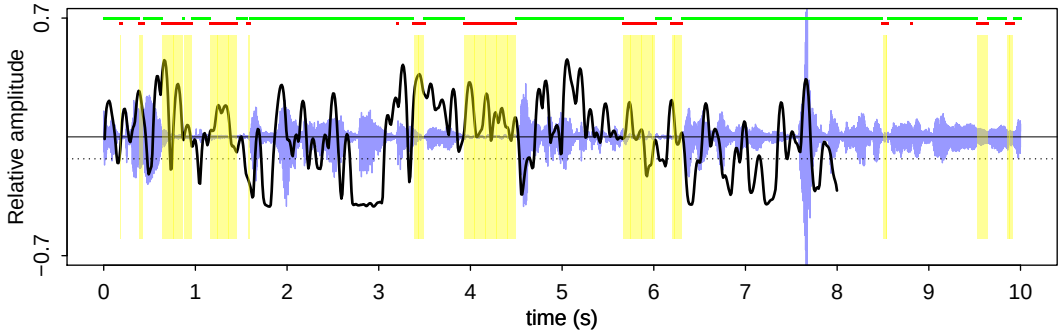

Figure 1.: A waveform of a segment from a speaker with Parkinson's disease, annotated with results of the VAD-algorithm.

to detect the acoustic envelope, with a custom decision procedure to capture the different types of pauses of the speaker, cf. Figure 1.

The TIMIT corpus (Garofolo, Lamel, Fisher, Fiscus, & Pallett, 1993) was used to benchmark the performance of the algorithm against other algorithms in speech of non-brain-damaged individuals. The decision procedure was compared to manual annotations of pathological speech obtained from DementiaBank (Becker, Boiler, Lopez, Saxton, & McGonigle, 1994; MacWhinney, 2007).

We modeled the resulting data under the assumption of multimodality. A Support Vector Machine classifier was used to measure the predictive value of the discovered patterns.

3. Results

The results show that the algorithm can detect that the speech-pause pattern in speech of individuals with PPA-NF is different from that of individuals from the other classes. Differences between the other classes are more subtle, and may be statistically significant.

The generating distribution is a sum of multiple distinct Gaussians, each of which represents a pause category. The mean and variance of the Gaussians are clearly distinct for each of the participant categories, cf. Figure 2.

The performance of the classifier beats a baseline ("Zero Rule") strategy that always predicts the majority class.

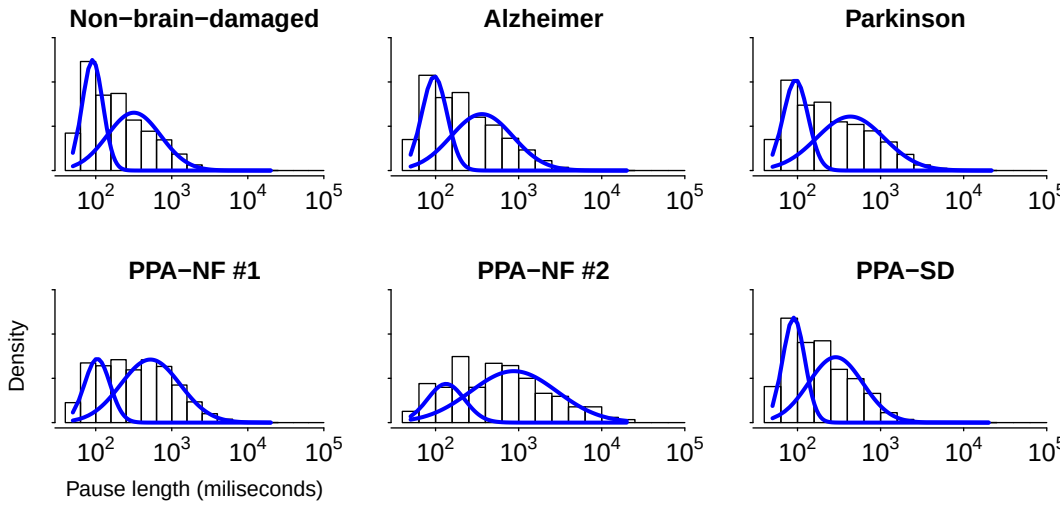

Figure 2.: The distribution of pause lengths as detected by the algorithms. Each bar represents the number of pauses with a given length. Overlayed are Gaussians that model the data as the sum of a two-model process. The two PPA-NF participants show a pattern that is clearly distinct from the other classes.

References

Baken, R. J., & Orlikoff, R. F. (2000). *Clinical measurement of speech and voice*. Cengage Learning.

Becker, J. T., Boiler, 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.

Campione, E., & Véronis, J. (2002). A large-scale multilingual study of silent pause duration. In *Speech prosody 2002, international conference.*

Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., & Pallett, D. S. (1993). Darpa timit acoustic-phonetic continous speech corpus cd-rom. nist speech disc 1-1.1. *NASA STI/Recon technical report n*, *93*.

MacWhinney, B. (2007). The talkbank project. In *Creating and digitizing language corpora* (pp. 163–180). Springer.

Ramırez, J., Segura, J. C., Benıtez, C., De La Torre, A., & Rubio, A. (2004). Efficient voice activity detection algorithms using long-term speech information. *Speech communication*, *42*(3), 271–287.