**Problem statement:** Can speech features be used to accurately detect cognitive stress?

**Data collection:** The Stroop task

- It is difficult to discriminate between cognitive load and stress\(^{11}\).
- The Stroop task is a standard psychological task to establish cognitive load\(^{12}\).
- The task was run in two trials:

  - Unstressed trial
    - Relaxed, with no time limit or financial repercussions.
  - Stressed trial
    - Time constraint of 1.5 seconds and a financial penalty of $0.75 for each wrong answer.

**Validation:** Was it stressful?

- Speakers self-reported their stress levels during both trials.

  - 23 speakers reported an increase of stress.
  - 4 reported no change.

**Feature extraction**

Speech processing was performed automatically with Praat\(^{16}\).
- Utterance boundaries were time-aligned with transcripts using silence detection.
- Pitch and intensity features were extracted at the word level.

**Classification experiments**

We explored classification across gender vs. by gender, resulting in three data subsets: female and male (F,M), female (F), and male (M). Two classification experiments were performed against each dataset using Python and Scikit-Learn\(^{16}\).
- The first randomly split the data into 80% for training and 20% for testing, tuning parameters on the training set before evaluating on the held-out test set.
- The second used leave-one-subject-out cross-validation across all subjects to see how the model performed on never before seen speakers. This was reported as the average across all subjects.

**Results**

- **Accuracy on 20% held-out test**
  - Consistent improvement over the majority class baseline.
  - Illustrates challenge of combining speech data across genders.

- **Accuracy on LOSOCV**
  - Most algorithms perform much worse, but random forest maintains similar accuracy.

The 3 best and worst subjects in increase/decrease % over MCB Across approaches the LOSOCV models excel and struggle with the same speakers.

**Future work**

- Aggregate data from multimodal sensors, so the system can handle shortcomings in any one sensor for each individual (findings presented at FAVASP 2019).
- Explore personalized, adaptive models that use the speaker's differences to its advantage, rather than removing them through standardization.
- Expand feature set to include spectral and formant features.
- Study detecting other cognitive/affective states through non-invasive sensors, such as focus levels or emotions.

**Bibliography**


**Acknowledgment**

Thanks to fellow researchers Brandon John, Taylor King, and Kristin Sarneczky.

This work was supported by a RIT College of Computing and Information Sciences Kareda Endowed Chair Fund, Health Information Technology Strategic Initiative Grants.