INVESTIGATING THE USE OF PAUSES FOR SPEECH CLASSIFICATION

Robert Cochran, Andrew Back, Janet Wiles

Introduction

- Investigating the information theoretic properties of pauses for use in real-time speech classification systems
- Evidence has shown that people with neurological diseases like dementia will rely on specific speech patterns more often than usual, e.g. pauses
- Research can be used to help aid the lives of people living with dementia
- Conversations of different age groups are analyzed to find potential insights as to how young and middle aged speakers use pauses and if this can be used for classification
- Novel entropy methods allow for the potential application of entropy to produce low sampled real-time results

Theory

Entropy

Classification based on Shannon entropy: $H(X) = -\sum_{i=1}^{n} P(x_i) \log_b P(x_i)$

Pro: Robust, simple statistical measure

Con: Requires large sample space

Fast Entropy

Uses pre-defined models to reduce the sample space size needed for output [1]

[1] Back, A. D., Angus, D., & Wiles, J. (2018, May 17). Fast Entropy Estimation for Natural Sequences. Brisbane, Queensland, Australia.

Methods/Materials

Calpy

Software that analyses and classifies audio files based on speech behavior. Used to detect speech anomalies.

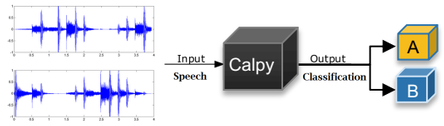

Figure 2: Overview of the Calpy System where A and B are distinct speech behavioral groups (e.g. Young or Middle Aged)

Symbolization

Audio file digitized into array of [0,1]'s where 0=no sound, 1=sound.
Visualization of 2 speaker digitization:

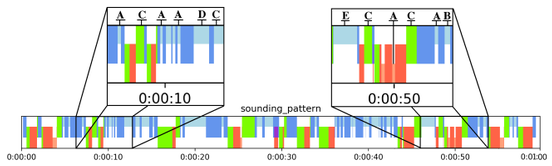

Figure 2: Audio file Pause and Sounding Visualisation

Person A:
- Light Blue – Inner pause
- Blue – Speaking

Person B:
- Light Red – Inner pause
- Red – Speaking

Green – Joint Pause

Symbol Model Example
Symbols: {A, B, C, D, E}
- A: 0 - 50ms
- B: 51 - 100ms
- C: 101 - 200ms
- D: 201 - 400ms
- E: 401 - 1000ms

Symbols are collected and grouped, e.g.

[[ACAADBABAABDCAAAACCD], [ADAADAAAAADDEAAAADAA], … , [ADDAEABAABBBACAEAAACD]]

And become individual entropy values,

[[4.2724593708], [4.2700732917], …, [4.2697115736]]

Results

Age and Pause Use

~95% of pauses under 10ms

Older groups tend to use pauses differently

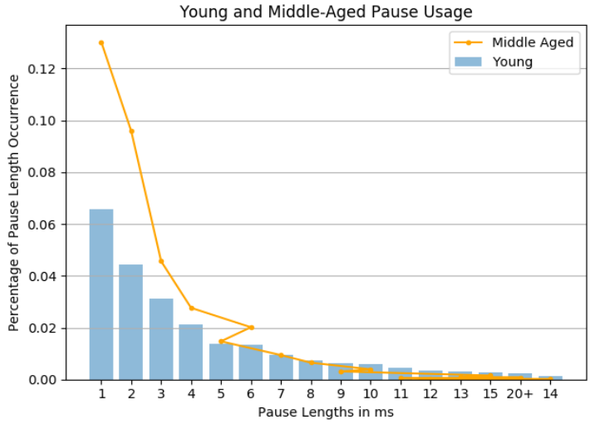

Figure 3: Pause comparison based on Age groups of Young and Middle Aged

Ranked Probability

Symbol models being ranked by count show how symbol usage can differ. This can provide heuristic for classification

Audio Insertion

Entropy values of inserted audio files to other audio helps visualize anomalies in the data

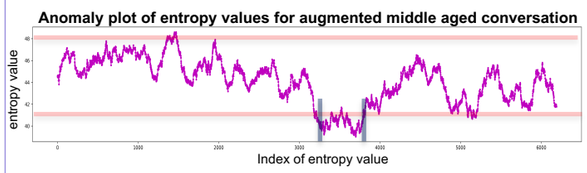

Figure 4: Anomaly plot of middle aged conversation with young conversation inserted

Discussion/Conclusion

Initial results show pauses are a potential candidate for speech behavioral classification, but require further testing. Current results are not suitable as reliable classifier.

Entropy values tended to vary greater for middle aged conversations, meaning middle-ages speakers deliver information less consistently when compared to young speakers (e.g. mix between no change to frequent change in pause use).

Age showed interesting pause characteristics between young and middle aged speakers

Further Work

Further work on different demographics (e.g. gender, language), prosodic speech elements (e.g. pitch, tone, …) to increase effectiveness of the classifier.

Testing on real-time delivery towards Fast Entropy usage.