Journal Report 3

9/15/19 - 9/22/19

Arnav Bansal
Computer Systems Research Lab
Period 5, Bansal

Daily Log

Monday September 2

Moved beyond small test .wav files to TED talks with the Speech Recognition, Google, and Microsoft APIs. Unfortunately didn't work with the Google and Microsoft APIs as the file size was too big for the free trial. Will need to upgrade in the future, but started with Speech Recognition for the time being. Unfortunately ran into errors like unexpected pauses and empty array files when testing.

Wednesday September 4

Stayed home sick :(

Friday September 6

After corresponding with Arvind on how laughter and applause in the TED talks could be throwing the API off, I brainstormed and researched ways we could overcome this. I looked into audio algorithms that detect laughter and others that detect applause. Considered feasibility of implementing models to detect these sound types and splice them out of the audio files.
Also tried to manually time how long laughter and applause each lasted in each of their instances in the Ken Robinson TED Talk. Seem to be around 2.5 and 4 seconds each, but will need larger sample size to be more confident. Realized that if we find similar reliable numbers, we could consider breaking the TED Talks into 2.5 second segments and splicing out ones that the API doesn't translate instead of implementing models to detect laughter and applause.

Timeline

Date	Goal	Met
September 6th	Finish formatting dataset for initial processing and proof-of-principle model training	Yes, Have dataset with 2,461 entries of TED Talks corresponding to their transcripts and metadata
September 13th	Test various Speech/Audio to Text APIs on this dataset	Yes, Implemented Speech Recognition tests with sample .wav file with Python, Google, and Microsoft APIs.
September 20th	Have initial results for baseline implementations of these APIs on this dataset	Yes, got transcripts from APIs (albeit with errors for laughter and applause)
September 27th	Find high-quality algorithms for laughter and applause detection in audio files, Manually approximate average lengths of each in several TED Talks	
October 4th	Scale API(s) to small subset of the TED Talk dataset and compare transcripts	

Reflection

Running into the errors with laughter and applause was certainly not expected or pleasant, but they allowed me to think about other exceptions or special cases to consider when using the various APIs. It would be interesting to see how Google and Microsoft fare down the line but only after overcoming the hurdles with the more accessible Speech Recognition API. On the bright side, it was great that the API still returned a solid transcript, and algorithms for laughter and applause detection may prove promising. We might also be able to get away with splitting the TED Talks into segments. I'm looking forward to the bigger goal in the short-term of generating transcripts across these APIs with out too much trouble or manual work, allowing for easy comparision as we later move on to summarization.