Automatic Speech Recognition (ASR) software can be used to produce real-time captions and transcripts of lectures, namely in a classroom setting. In this preliminary study, a performance check is done on how well ASR recognizes speech based on texts of various reading levels, according to the Flesch-Kincaid (F-K) readability test. The texts are read aloud with natural human speech, then the audio recordings are sent to ASR systems for transcription. The Word Error Rate (WER) of the resulting transcripts are calculated by comparing them against the original scripts. A correlation analysis between the WER and F-K measures showed that there was a weak relationship between ASR accuracy and readability levels for the ASR services offered by IBM Watson and Microsoft Azure.

Methods

**Technology Used**
- Online sources with information about American History
- IBM Watson Speech to Text API through IBM Bluemix
- Microsoft Bing Speech API through Microsoft Azure
- Flesch-Kincaid Readability Ease Measure
- National Institute of Standards and Technology Word Error Rate tool

**Procedure**
- Adapted scripts to various F-K scores
- Read scripts aloud in natural human speech
- Sent audio files to ASR systems for transcription
- Performed WER analysis on the results

Results

1. Watson
   - Hints that there is a trend
   - As F-K scores increase (less complicated vocabulary and fewer words in a sentence), Percent Total Error decreases
   - R squared is low, suggests a very weak relationship
   - P-Value is approximately 0.6, means that the relationship is not statistically significant

2. Azure
   - Suggests a trend
   - Same as Watson, but stronger: the R squared value is higher by a factor of 10
   - P-value is approximately 0.067, which is fairly close to the alpha level of 0.05 but it is still not statistically significant

Conclusions

Based on the data, there seems to be a relationship between F-K Scores and ASR Word Error Rate. However, the current data is not strong or statistically significant enough to act as evidence.

There are other factors outside ASR WER that affect its understandability. Watson has a higher accuracy and less error, but when it outputs its transcript, there is a lack of punctuation and formatting. The Azure ASR has periods, question marks, numerals, and paragraph line breaks, while Watson only has periods and looks like raw text.

References


Funding

This work has been generously supported by an NSF REU Site Grant (#1460894) and by AccessComputing of the University of Washington (UW).