Comparing Automatic Speech Recognition Word Error Rates for Spoken and Sign Supported Speech Lectures

Kayla Vodacek, **Abraham Glasser**, Dr. Raja Kushalnagar

Rochester Institute of Technology, Rochester, New York

R·I·T | NATIONAL TECHNICAL INSTITUTE FOR THE DEAF

Introduction

Using Automatic Speech Recognition (ASR) in Simultaneous Communication (SC, "sim-com") classrooms

- What is sim-com?
  - Sign supported speech / speech supported sign
  - Grammar differences
- Linguistic features
  - Aligning onset of words
  - Code-blending

ASR

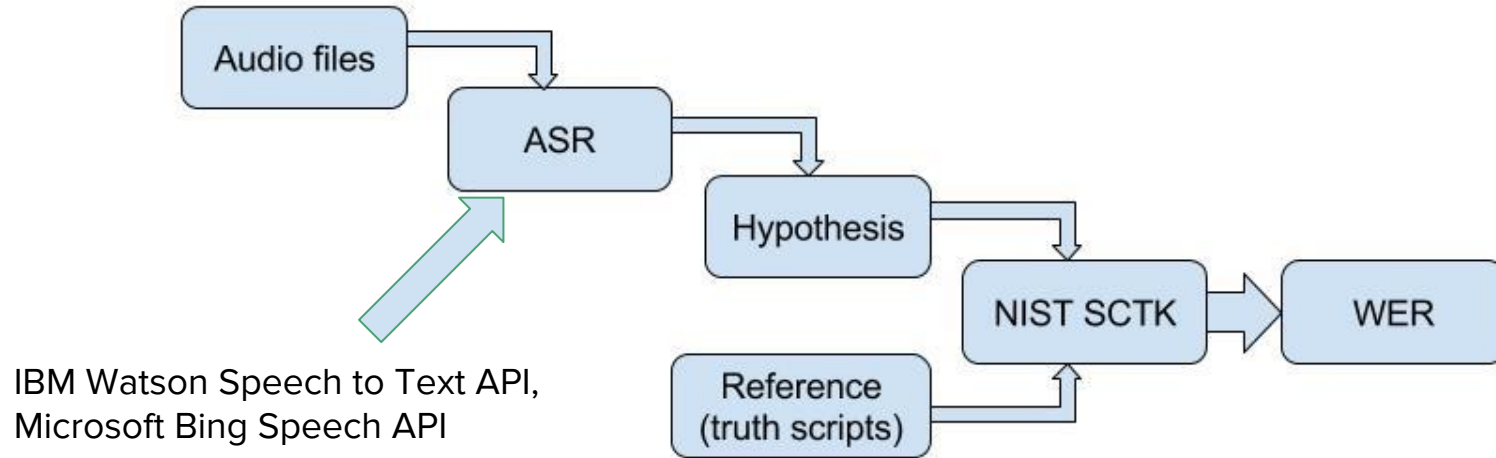- Voice recognition
- Accuracy
- Types of errors
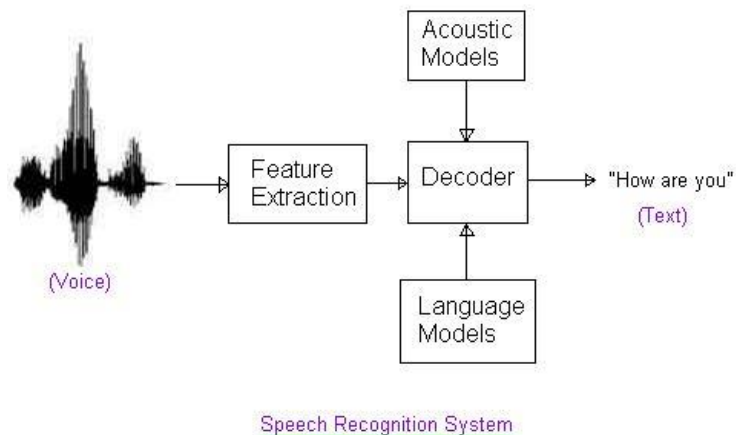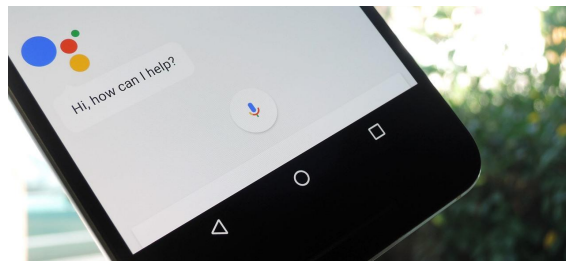
Research Questions

1. What is the effectiveness of an ASR for a spoken versus SC transcripts for a single speaker?

2. What are linguistic features of SC that may be affecting the ASR results?

Method

- Record audio of spoken only (non SC)
    - 10 lectures (each between 10 and 11 minutes)
    - Adapted from a history class on open online courses from Yale
- Record audio and video of SC
    - Same 10 lectures and presenter

- Measure WER and Word Accuracy for all ASR transcripts
    - NIST (National Institute of Standards and Technology)
        - Speech Recognition Scoring Toolkit (SCTK) Version 2.4.0
- Calculate the average WER/Word Accuracy for non SC and SC

- Watch videos and identify linguistic characteristics of SC

Illustration of process

Audio files → ASR → Hypothesis → NIST SCTK → WER

Reference (truth scripts) → NIST SCTK

IBM Watson Speech to Text API, Microsoft Bing Speech API

Sample texts for comparison

SPOKEN

Today we're talking about loyalists and in doing so we're going to be talking about the range of responses to the building crisis in British America. So basically over the course of the first few lectures here we've looked at the logic behind some of British policy. We look at the logic of resistance

Very good transcription, broken up almost correctly

Original

Today we're talking about Loyalists, and in doing so we're going to be talking about the range of responses to the building crisis in British America. So basically over the course of the first few lectures here, we've looked at the logic behind some British policy; we've looked at the logic of resistance.

SIM COM

Today we're talking about loyalists and in doing so we're going to be talking about the range of responses in the building crisis in British America so basically over the course of **first. Few** lectures here we've looked at. The logic behind some of the British policies. We look at the logic of resistance.

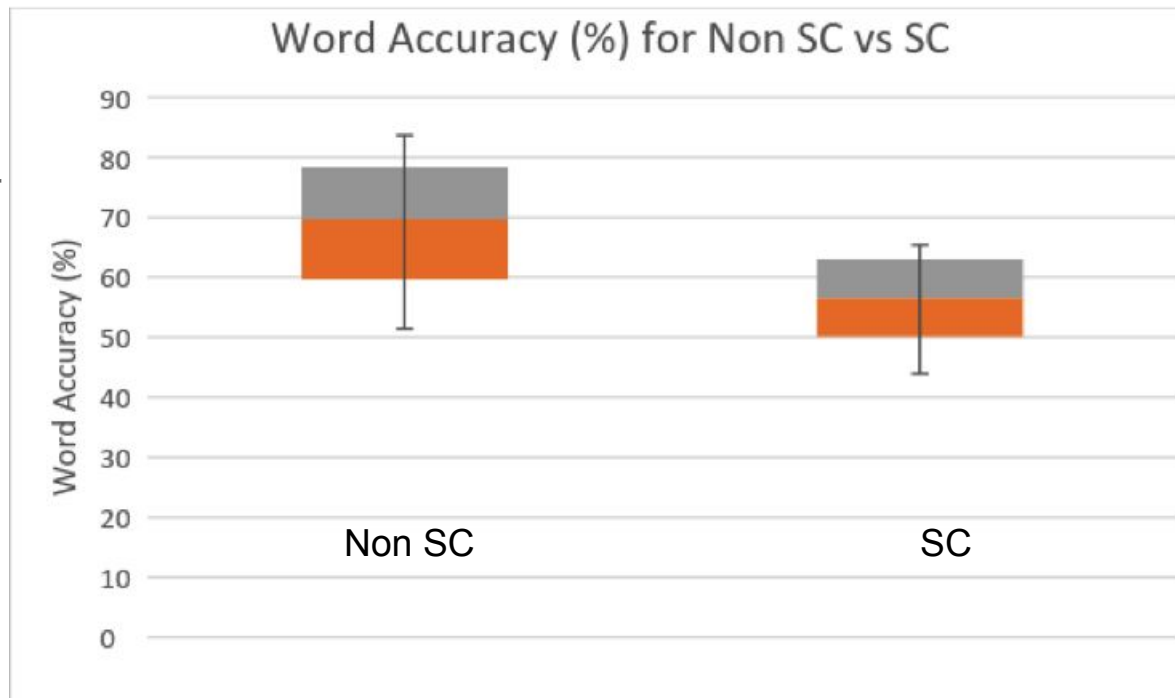Pretty accurate for what was spoken, some words missing, broken up incorrectly

Results

RQ1 (Effectiveness)
- Non SC performed better than SC
- Average word accuracy 50% better for non SC

RQ2 (Linguistic features)
- Nearly every word was spoken, but with much longer pauses

Word Accuracy (%) for Non SC vs SC

Conclusions

- Linguistic features
    - Delay in English production

- ASR language models are based on standard English
    - SC model alters the oral information
    - ASR does not do well with pauses

- Neither average is about 80% threshold
    - Shows that ASR still needs to be improved for long lectures rather than voice commands

Summary and Future Work

- This was a preliminary study

- ASR is continuously being developed and improved

- Further testing necessary
  - Classroom environment w/ students
  - Student perception
  - More SC users (different experience & styles)
  - Different subject other than history