Using Automatic Speech Recognition with Artificial Intelligence in the Classroom for Real-Time Captioning

Abstract

At RIT, many Deaf and Hard of Hearing (DHH) students use access services in the classroom, including real-time captioning services like TypeWell and C-Print. Demand for real-time captioning hours has increased from 15,440 hours in 2007 to 24,335 hours in 2016, a 58% increase. Due to increased demand and limited resources, we are exploring the possibility of using real-time automated real-time captioning as a cost-effective alternative. Our ongoing study examines the success and quality of automated real-time captioning through several Automatic Speech Recognition (ASR) services, including Microsoft and IBM, that use cognitive services.

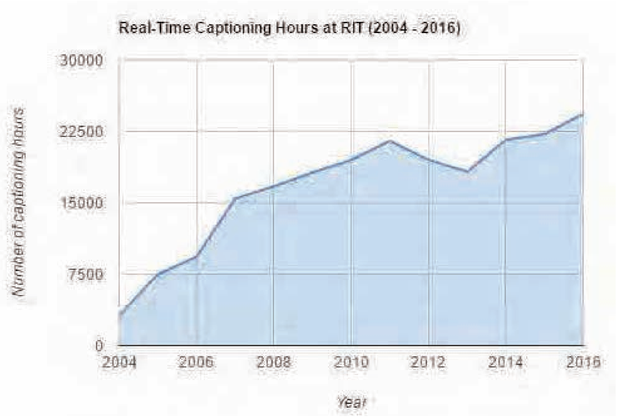
Objective

To enhance DHH students' learning experience in the classroom, it is essential that high-quality captions are provided. The purpose of ASR services is to translate any audible dialogues occurring in the classroom and turn these into real-time captions without significant latency. The students should receive coherent information by watching the caption feed provided from software that use ASR services. We are looking for ASR services that could provide real-time captioning that is functionally equivalent to the current accommodations we have today.

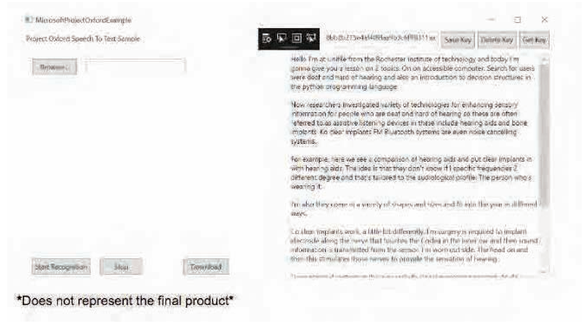
Customer Needs

- Cost-efficient and productive alternative to TypeWell and C-Print
- Functional equivalence
- Clear, correct and readable captions
- Low Word Error Rate (WER)
- Lecture keywords are accurate
- Ease of set up and tear down

Demand for Real-Time Captioning Services

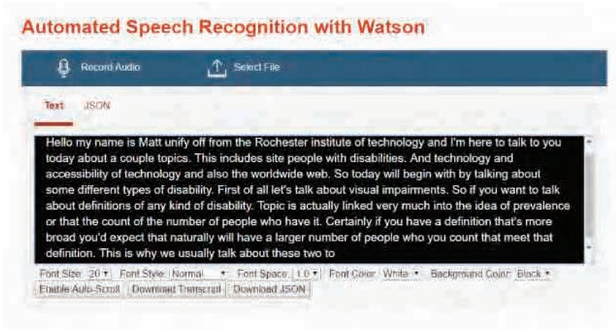


Microsoft Transcript Example

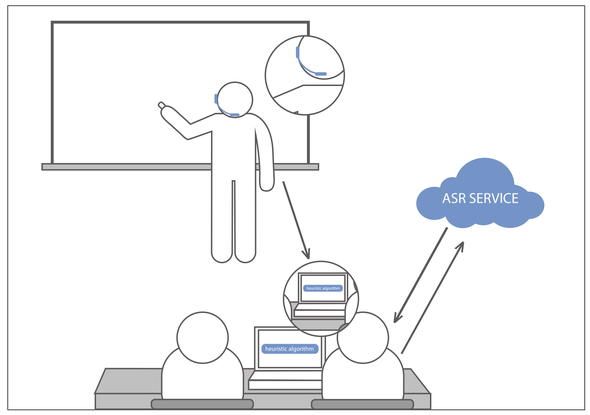


Does not represent the final product

IBM Watson Transcript Example



Functional Overview

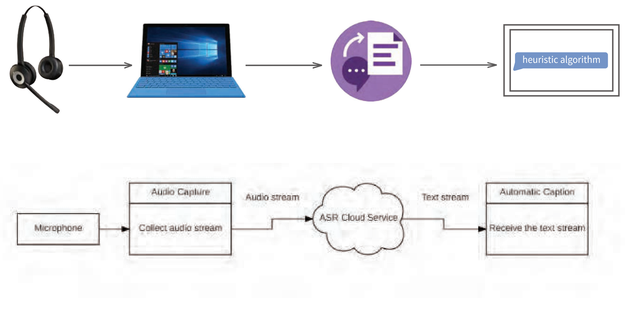


Comparative Analysis

Word Error Rate (WER)	Microsoft Summer Result (50 MB)	Microsoft Fall Result (500 MB)	Watson Summer Result (50 MB)	Watson Fall Result (500 MB)
With Punctuation	Total error = 36.1% Percent Correct = 70.2%	Total error = 17.5% Percent Correct = 85.6%	Total error = 17.1% Percent Correct = 86.9%	Total Error = 19.8% Percent Correct = 83.3%
Without Punctuation	Total Error = 30.0% Percent Correct = 76.3%	Total error = 17.5% Percent Correct = 86.5%	Total error = 13.2% Percent Correct = 91.1%	Total Error = 12.9% Percent Correct = 90.3%

Preliminary Data

System Overview



Methodology

1) Collect and record audio samples through classroom instructor's microphone in a classroom environment or sound-proof booth.

2) Feed live audio from lectures to ASR services and receive live transcription.

3) Send audio recordings to CaptionFirst, a Communication Access Real-Time Transcription (CART) service, for baseline transcripts.

4) Compare classroom/sound booth transcripts with the baseline transcript.

5) Receive Word Error Rate (WER) results from transcript evaluations.

Qualitative Result

From our preliminary findings, real-time captioning generated from ASR is not functionally equivalent to TypeWell and C-Print. Ideally, automatic captions would have a WER of 5% or less, with accurate keywords from classroom lectures. There are some variables that could have impacted the quality of recordings and live audio such as microphone quality, clarity of speech, background noise, and classroom reverberations. ASR has its own challenges such as word variation, speech speed rate, context dependency, and more. Cognitive services is a relatively new field with many players like IBM, Microsoft, Google, Amazon, and Baidu. With their investment in artificial intelligence, achieving functional equivalence with real-time automatic captions could become a reality in the near future.

- We would like to express our gratitude to the following for their contribution to this project -

- Brian Trager
- Emily Prud'hommeaux
- James DeCaro
- Lisa Elliott
- Matt Huenerfauth
- Michael Stinson
- Patrick Smith
- SungYoung Kim

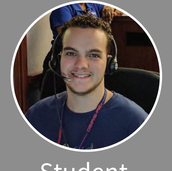
R·I·T
National Technical Institute for the Deaf

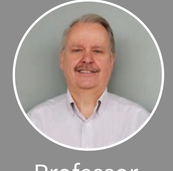
Microsoft

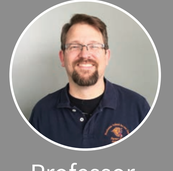
IBM Watson


Student
Mary Rose Weber


Student
Ka Bo Cheung


Student
Abraham Glasser


Professor
Gary Behm


Professor
Chris Campbell