UF COLLEGE OF EDUCATION

Challenges for Using Automatic Speech Recognition Systems to Detect Children's Reading Errors

Jaiden Magnan, Liliana Lusvardi, Olivia Schwartz, Alejandra Casillas, Walter Leite
Virtual Learning Lab: vll@coe.ufl.edu

CENTER FOR UNDERGRADUATE RESEARCH UF

Introduction

- Underperformance in reading impacts emotional and academic areas, including science and math.
- Schools lack resources for one-on-one tutoring; reading apps can help but currently lack tailored feedback.
- No existing apps provide real-time, individualized feedback on reading errors.
- Current Automatic speech recognition (ASR) research focuses on typical development, with little focus on error patterns in struggling readers.
- This project aims to develop ASR-based models for detecting reading errors and creating a reading app intervention with just-in-time feedback.

Methods

- Two pre-trained ASR models are used: Whisper (transformer-based encoder-decoder) and XLSR-53 (WAV2VEC architecture with CTC loss).
- Models are tested for accuracy in detecting reading errors.
- Audio recordings of first to third graders' reading practice.
- ASR - generated transcripts are being used to train and evaluate models and app feedback system.

Whisper

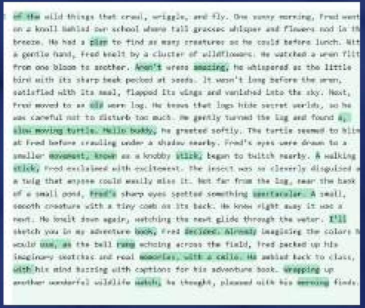
- Whisper is Open AI's ASR model that was trained on 680,000 hours of audio from a variety of different languages
- Whisper offers five models: base, tiny, small, medium, and large. Tiny is the most efficient but least accurate, while large is the least efficient but most accurate.

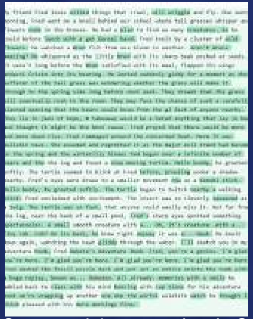
Wav2Vec

- Wav2Vec is a model developed by Facebook AI that converts speech to text.
- The XSL-53 is designed to handle multilingual scenarios, likely helping with decoding children's speech.

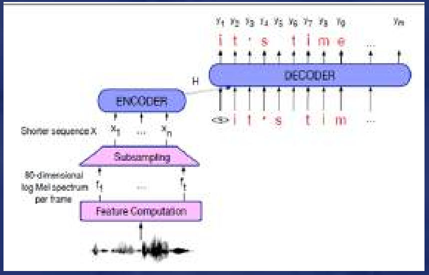
Limitations

- Children's recordings had background noise, which impacted the accuracy of the ASR models
- The whisper model had issues generating the correct punctuation, which caused the accuracy to drop.


of the wild things that crawl, wriggle, and fly. One sunny morning, Fred went on a knoll behind our school where tall grasses whisper and flowers nod in the breeze. He had a plan to find as many creatures as he could before lunch. With a gentle hand, Fred knelt by a cluster of wildflowers. He watched a wren flit from one bloom to another. Aren't wrens amazing, he whispered as the little bird with its sharp beak pecked at seeds. It wasn't long before the wren, satisfied with its meal, flapped its wings and vanished into the sky. Next, Fred moved to an old worn log. He knows that logs hide secret worlds, so he was careful not to disturb too much. He gently turned the log and found a, slow moving turtle. Hello buddy, he greeted softly. The turtle seemed to blink at Fred before crawling under a shadow nearby. Fred's eyes were drawn to a smaller movement, known as a knobby stick, began to twitch nearby. A walking stick, Fred exclaimed with excitement. The insect was so cleverly disguised as a twig that anyone could easily miss it. Not far from the log, near the bank of a small pond, Fred's sharp eyes spotted something spectacular. A small, smooth creature with a tiny comb on its back. He knew right away it was a newt. He knelt down again, watching the newt glide through the water. I'll sketch you in my adventure book, Fred decided. Already imagining the colors he would use, as the bell rang echoing across the field, Fred packed up his imaginary sketches and real memories, with a smile. He ambled back to class, with his mind buzzing with captions for his adventure book. Wrapping up another wonderful wildlife watch, he thought, pleased with his morning finds.

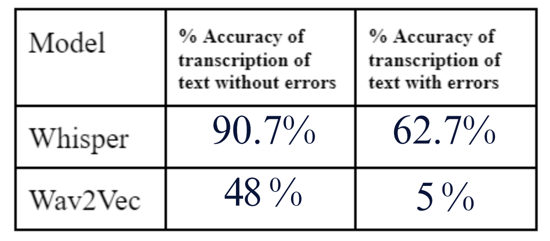

[illegible]

Above are audio transcriptions from Whisper. The left side displays the results from the recording with no errors, and the right side shows it with errors. The green highlights show when either the model accurately identified a mispronunciation or generated nonsensical text because it could not parse input.


y_1 y_2 y_3 y_4 y_5 y_6 y_7 y_8 y_9 y_m
i t ' s t i m e ...
DECODER
H
ENCODER
<s> i t ' s t i m ...
Shorter sequence X x_1 ... x_n
Subsampling
80-dimensional log Mel spectrum per frame f_1 ... f_T
Feature Computation

ufdatastudio.com/posts/2023-08-20-ASR-Blog/

Results



Model	% Accuracy of transcription of text without errors	% Accuracy of transcription of text with errors
Whisper	90.7%	62.7%
Wav2Vec	48 %	5 %

References