



Phoneme Pronunciation Project

Alex Vu
Machine Learning Class 2025-26(Dr. Santos)
nam.vu@concordiahanoi.org

Project Goal

Learning Application to Help Elementary School-Aged Students Learn to Pronounce English Words

This project aims to create an application that can be used by students in elementary school (specifically Grades 1-2) who are native speakers of another language learn and practice English pronunciation. Furthermore, this project also aims to aid teachers in collecting the data from student's practice and pronunciation, identifying words that they are struggling with.

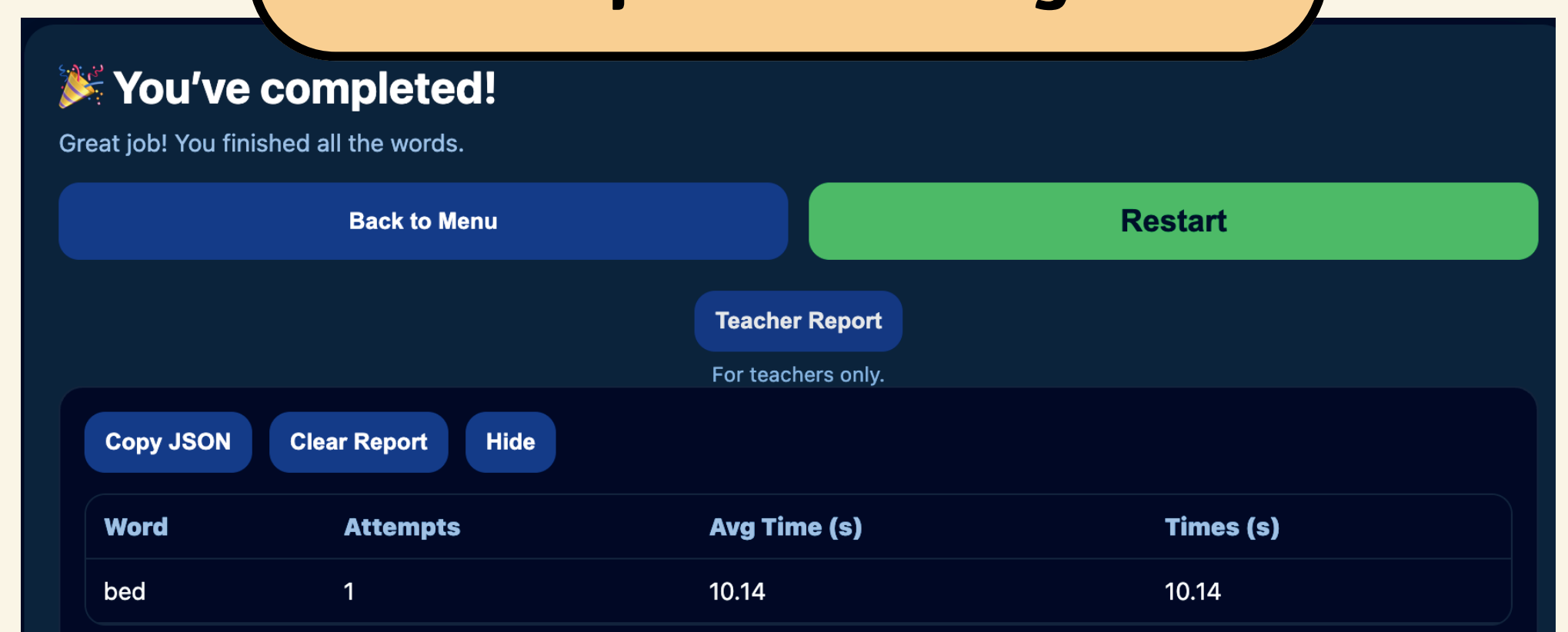
Model

The model used in this project is vitouphy/wav2vec2-xls-r-300m-timit-phoneme, based on Meta AI's Wav2Vec2-XLS-R architecture. It was fine-tuned on the TIMIT dataset to perform phoneme-level recognition of American English speech. The model converts raw audio input into phoneme predictions using a CTC decoding mechanism. In this system, it serves as the core engine for detecting and evaluating children's pronunciation. This allows the platform to provide accurate, real-time feedback for phoneme learning and speech practice.

Sample Recording Page



Sample Result Page



The project currently has three main webpages. The first webpage is a menu for selecting words, the second the recording page, where the word is displayed and the microphone automatically starts recording. This is to minimize distractions for younger students, who tend to play with the buttons. Finally, the result page displays two buttons, with a teacher report that is locked to ensure students don't see and get pressured by their data.

Conclusion and Future Work

Possible future work includes assessment mode, which teachers can use to assess the skills of students, the ability for the project to collect the data for a student over time and stores it, and incorporation of AI to help analyze the trends in data over time, such as phonemes that a student is repeatedly struggling with.

Demo

