Alexander Terziyski
CART 263

Final Project: Proposal

The goal of my final project is to touch on the combination of word pronunciation and hand gestures. Using the capabilities of p5.speech and MediaPipe Hand gestures, I want to enable users to dynamically manipulate word pronunciation based on the position of their hands on the screen. This combination of speech recognition and hand gesture detection will also be supplemented with some text visualization.

The core concept of the project revolves around the dynamic modification of word pronunciation through hand gestures. As users interact with the application, their hand movements will influence how words are pronounced and visualized on the screen. This concept should encourage the user to experiment with different hand gestures and observe their impact on the pronunciation of words.

From a technical perspective, the project will integrate p5.speech to enable speech recognition and synthesis functionalities, allowing the application to process spoken words and generate corresponding audio output. MediaPipe Hand gestures will be utilized to detect and track the position of the user's hands in real-time, providing data that will dynamically modify word pronunciation and visualization on the screen. JavaScript and p5.js framework will be employed to develop an interactive and visually appealing user interface that effectively communicates the interaction mechanics and encourages exploration. Additionally, there is a possibility of incorporating these elements into a larger visual or auditory experience, such as simulating an orchestra where hand gestures control different musical elements or creating captivating visual effects that respond to both speech and hand movements, offering users a multisensory exploration of language and expression (Although, I’m still debating on this part),

Despite the potential of the project, several challenges need to be addressed, including accurately interpreting hand gestures to translate them into meaningful modifications in word pronunciation. Synchronization between p5.speech and MediaPipe HandPose functionalities,

as well as optimizing performance, will be crucial technical considerations. Furthermore, designing an intuitive user interface that effectively communicates the interaction mechanics and encourages exploration will also be important.