Emotifying Digital Communication

ECE 539 - Neural Networks
Nick Engebretsen, Caleb Starfeldt, Kwan Man Cheng
nengebretsen@wisc.edu,
kwancheng54@wisc.edu,
cstarfeldt@wisc.edu

1. Overview

Our project's goal is to help people who are deaf or hard of hearing understand conversations better by turning what's said into simple emotes that show feelings and key ideas. We plan to do this by using technology to analyze the tone of voice and words of a spoken sentence. First, we figure out the feeling behind the words, like if someone is happy, sad, or neutral. Then, we change the spoken words into text and pick out an important word to turn into a picture. This way, we create two images that show both the emotion and the main point of what was said.

In short, by classifying our spoken word, we aim to use machine learning to simplify what is said.

2. Background

People with hearing disabilities face many inconveniences in our society. This idea is mainly for people who can't hear well, including those who have always had this challenge or those who have lost their hearing over time. We want to make this technology work with special glasses that show these emotes as a small image, projecting onto something similar to Google Glass or Vision Pro, so it's easy for users to see and understand the conversation in front of them instantly. Our goal is to make talking and sharing feelings easier for everyone, no matter if they can hear or not, by using visuals to fill in the gaps of spoken words.

Lots of previous datasets from others have been made separately. There are emoji datasets(https://www.kaggle.com/datasets/subinium/emojiimage-dataset), RAVDESS Emotional speech audio(https://www.kaggle.com/datasets/uwrfkaggler/ravdess-emotional-speech-audio) that we can make use of in our project. By utilizing those data into our model, we can predict the comprehensive emotion and content for the people with hearing disabilities.

> *Background: Explain a bit background of this project, what have been done before, what is the state of the art outcome (A good place to cite a couple of most relevant works, including URLs, citation to references, etc.). This is where you provide references/sources for your "baseline" project (give the URL if it is a github/Kaggle project)*

3. Statement of Work

- 3-1. Datasets

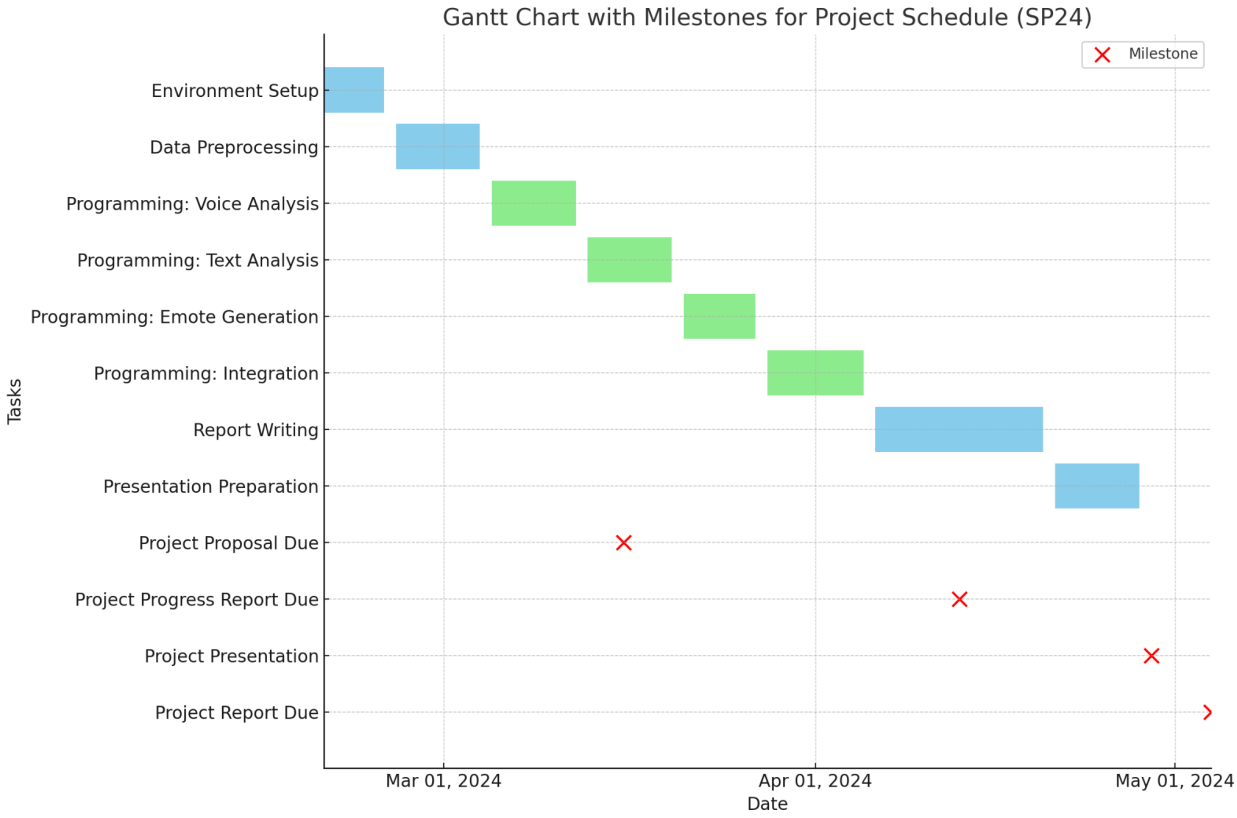
- Source of the dataset.
    - Emoji dataset https://www.kaggle.com/datasets/subinium/emojiimage-dataset
    - RAVDESS Emotional speech audio
    - https://www.kaggle.com/datasets/uwrfkaggler/ravdess-emotional-speech-audio
    - KP20K Dataset
    - https://huggingface.co/datasets/taln-ls2n/kp20k?row=0
- Contents of the dataset.
    - Emoji's pixels and labels
    - Audio's emotion/intensity/Modality
    - Abstracts, key words and their associated titles from scientific articles.

- 3-2. Method
    - The project should contain two machine learning models. The first one focuses on recognizing the speaker's emotion from audio. The second one is used for extract the keyword from the text.
    - For the speaker emotion recognition model, the team who wrote the paper A proposal for Multimodal Emotion Recognition using aural transformers and Action Units on RAVDESS dataset evaluated a pre-trained xlsr-Wav2Vec2.0 transformer using two transfer-learning techniques. They implemented a multi-layer perceptron on the top of the model to meet the maximum accuracy. The article will greatly inspire our work in the process.
    - For the keyword extraction model, we will use the KP20 dataset. It contains enormous articles with their corresponding keywords. Since the input sentences in our application are likely to be simpler than those in scientific publications, the training result should be relatively better than those complicated cases. We will refer to articles like UCPhrase: Unsupervised Context-aware Quality Phrase Tagging to train this model.
- 3-3. Outcome and Performance Evaluation

For our project, we're looking to create a system capable of clearly conveying the meaning of a sentence through a set of distinct emotes. Our benchmark for success is straightforward—if a person who is unable to hear can accurately interpret the sentence based only on these emotes, without any additional context, then we've achieved our objective. In addition to the core communication function, we're also focusing on transmitting the emotional undertones of the sentence, as understanding sentiment is vital in human interactions. Lastly, the efficiency of our code is paramount. Speed is a critical factor in real-time communication, and we aim to optimize our system to deliver both rapid and accurate results.

Quantitatively, we aim to measure our success in classification and achieve a value above 75%. We also aim to perform cost analysis on the system we designed to determine if it is the most efficient way to go about what we are doing and improve it to a reasonable point before the presentation date such that it can be used in real time.

4. Project Plan



5. References

- https://www.core77.com/posts/112319/Three-Different-Design-Approaches-to-Projecting-Data-onto-an-Eyeglass-Lens
- https://www.hackster.io/news/diy-google-glass-for-under-10-1abea579ccf5
- https://www.digikey.com/en/products/detail/sparkfun-electronics/DEV-11113/5140820?utm_adgroup=&utm_source=google&utm_medium=cpc&utm_campaign=PMax%20Shopping_Product_Low%20ROAS%20Categories&utm_term=&utm_content=&utm_id=go_cmp-20243063506_adg-_ad-__dev-c_ext-_prd-5140820_sig-CjwKCAjwte-vBhBFEiwAQSv_xXzpcRvs1ONsU0W7rWwyC6fpkBlwUOU1C_Dco7gGsZw1IDHApQA9ORoCkhsQAvD_BwE&gad_source=1&gclid=CjwKCAjwte-vBhBFEiwAQSv_xXzpcRvs1ONsU0W7rWwyC6fpkBlwUOU1C_Dco7gGsZw1IDHApQA9ORoCkhsQAvD_BwE
- A proposal for Multimodal Emotion Recognition using aural transformers and Action Units on RAVDESS dataset https://paperswithcode.com/paper/a-proposal-for-multimodal-emotion-recognition
- UCPhrase: Unsupervised Context-aware Quality Phrase Tagging https://paperswithcode.com/paper/ucphrase-unsupervised-context-aware-quality
-

6. Contributions

- A brief statement (no more than 50 words) from each team member detailing their contributions to the proposal preparation.

- Kwan Man Cheng: I discussed the project design with my teammates, found datasets, and contributed to the shared document.
- Nick Engebretsen: I participated in the design and contributed to the shared document.
- Caleb Starfeldt: I discussed the topic of the project with my teammates and contributed to the shared document.