Final Project GEA Worksheet

Biggest Bird LLC.

Cole Bianchi, Dante Dodds, Kevin Dong, Nathan Litzinger, Andre Mitrik, Efe Sahin

Our Project

Project Chosen

Configurable Speech Censorship

Project Description

Utilizing a speech recognition model to identify a provided subset of flagged words and phrases that the user wishes to be filtered/masked out of an audio outlet. Upon detection, the system would eliminate the audio for the duration, potentially real-time with proper implementation. Also opens up discussion of how organizations/streaming platforms could abuse this to filter out certain messages and content to viewers (skewing political message outreach, sensitive subjects, etc.).

Goals

What might be the goals of the intelligent system? (what we're automating, prerequisites, directives and plan blueprints)

This intelligent system would identify certain flagged words and phrases (transcription phase), timestamps on when those words are said within some kind of audio input. Under some kind of set delay, our system would be reactive to the flags given and then there would be some kind of replacement or masking/censoring for the audio snippet under question.

Sequentially, we can lay out our automated goals as such:

1. Identifying all words (essentially transcribing an audio clip)
2. Finding the words and ensuring to associate when they're spoken to an instance in time to track it
3. Performing some kind of censorship based on what we're flagging/the goal our system

Environment

What is the environment that the system will be adapting within? (conditions of environment, necessary contextual information, inner/outer environment definitions)

Inner environment would be the individual word/phrase definition within the trained model that we will use. Intrinsically, it would also include the list preset words that we are considering a "flag" for our timestep tracking.

Outer environment would include the words and audio adapted from the input data/audio files.

Given the "delay" between the stream of input audio data and the output post-censoring, it gives us time for our automated intelligence to react to the data using an algorithm instead of being constricted to prediction heuristics. This also opens up another layer of complication that we could use in terms of how much weight to certain flagged words and phrases and how to deal with cases where we may want to react differently.

Adaptation

How might the system adapt to the environment given the goals? How might its inner environment change? (what constitutes a "trigger" in our system, what conditions affect events we analyze)

The adaption itself happens from voice to voice to pick up the words based on an individuals' unique dialect, speaking styles, etc. which will probably happen somewhere within the actual training phase when we train the model as opposed to the execution time. It is a primary goal of ours to include a broad training set so that our model does not get oversaturated and become biased towards recognizing only one "style" of voice.

As far as testing is concerned, some things we may need to do include:

- Acquiring many sources of audio data of various people speaking
- Potentially creating various classifications of voices to make training "simpler"
- Figuring a workaround for when some audio inputs may not be "clear"