Problem definition

Writing reports can take a lot of effort do type. Can this be done in some other way? Talking out a report is faster than typing it out on a keyboard. In fact, the average typing speed in the English language is between 61-90 words per minute (Cooper, W., 1983). While the speaking speed of an average English speaker is approximately 125-150 words per minute (Feng, J. Sears, A.,2004). So, wouldn't it be nice if we could write reports with our voice? This would increase productivity.

A downside of doing speech-based report making is that some people might be good in writing a language but not necessarily in speaking it. But luckily speech-based recording would actually help with this language problem. If speech-based report making would be used the speech could be automatically translated into the desired language in text form. This would make it very easy to share the report while also making it easy for anyone to "write" or speak it in their desired language. Making report writing very easy.

Idea

We propose "**Speechsy**" as a solution.
This would be a website where you could do the following:

1. The user can record audio or send already recorded audio as an input
2. The audio it will be send to a central server
3. This server would use an AI component to turn the audio into text
4. Then the text would be translated to the desired language
5. And then the text will be sent back to the user on the website as output

If we want to be fancier, we could also try to steam the audio to the server as it is recording to generate the text live as the user is speaking.

Objectives

Our current rough outline for the software is as follows:

- A website interface for recording and sending the audio and receiving the text
    - Flask and html/css + possibly react
- The server hosting
    - Digital ocean server hosting or something else hosting
- The AI that would receive the audio and try to turn it into text
    - Flask receiving the audio and sending it too the model
- Making/building/training the speech AI model
    - Google speech to text (https://cloud.google.com/speech-to-text/docs/libraries)
- Translation feature
    - Neural Model in Python
- Sending the translated text back in a good format
    - Using flask to send the text back and display it nicely

For communications within the team, we have created a server on Discord and a group chat on WhatsApp. We have also all set up accounts on Trello and are in a team together to help with our workflow and make sure everyone is aware of what is being done and what needs doing still.

References

1. Cooper, W. E. (Ed.). (2012). *Cognitive aspects of skilled typewriting*. Springer Science & Business Media.

2. Feng, J., & Sears, A. (2004). Are we speaking slower than we type? exploring the gap between natural speech, typing and speech-based dictation. *ACM SIGACCESS Accessibility and Computing*, (79), 6-9.

3. Speech-to-Text Client Libraries | Cloud Speech-to-Text Documentation. (n.d.). Retrieved October 02, 2020, from https://cloud.google.com/speech-to-text/docs/libraries