# Assignment 1

**COSC345 Cool Group:** Oliver Kenning (6594104), Poppy Schlaadt (6057450), Padraig Carnegie (6927257), Laura Yoo (5045964)

**Due date:** 28th of July

**What data sets you chose:**

- ASL to Text: From "Kaggle", the first dataset consists of images of sign language gestures that correspond to characters, which can be used to convert ASL to text.
- Text to Speech: Also from "Kaggle", this dataset includes text normalisation, which can then be transformed into speech.
- Text to Speech: The LJ Speech dataset from "Papers with Code" is designed for text to speech use, and consists of audio recordings and corresponding text.


As a group, we decided to use a dataset of sign language to text, and a dataset of text to speech. We chose these datasets as they were common, simple to understand, and primarily, gave us the idea for our application, a sign language to speech app. This is something we all feel interested in, and motivated to create. By combining these two datasets, we hope to be able to make something new, that doesn't exist in the market, yet has potential to revolutionise how individuals with hearing impairments and non-sign language users communicate effortlessly and inclusively.

**Why the two data sets go together:**

These data sets can be combined in order to translate American Sign Language into text, and then this text can be converted into speech. By putting these two datasets together, we allow for translation/communication from ASL, to text, to speech. Therefore 2 datasets give us the ability to communicate in 3 different formats. This is all in an attempt to create seamless/less challenging ways for those who are deaf or hard of hearing to interact with the hearing world and vice versa. We have two options for the text to speech dataset because we want to be able to test both options and see which one performs better.

**What you are going to build:**

We are going to build an app which takes in gestures as input and outputs the gestures to speech. The Application will utilise the user's camera to detect ASL which it is then able to translate into text and then speech which can be read aloud to the user. We want to build this application because it both satisfies the requirements of the assignment by combining two unrelated datasets to make an app for the betterment of the user/society as well as being an application that we are interested in and excited to create. The goal of building this application is to create a more accessible and interactive tool for people who use sign language and those who do not understand sign language.

**Features:**

**REQUIRED FEATURES**

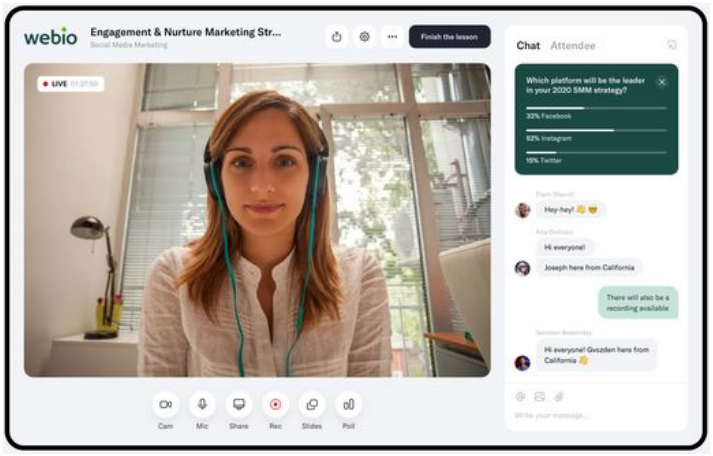
- ASL to Speech
    - Fundamentally, the app should convert the gestures into text and then output them into speech. This is the basis of the program.
- Speech to Text
    - The app shall contain the option to record the voice of a user and convert it into text, so a conversation can be had.
- Subtitles
    - The app shall feature subtitles which appear. The subtitles will be different colours for the user that is using the sign language and the user who is speaking.

**ADDITIONAL FEATURES**

- Record transcript
    - There should be an option for the transcript to be recorded (optionally), and is saved to a .txt file.
- Add gestures
    - The user should have the option to add custom gestures, in case it is not included in the dataset.
- Change voice
    - The app should include an option for the user to change or customise the voice.
- Emotion detector
    - Records the emotion of the face of the signer and then uses this data in order to create the voice.
- Dictionary Mode
    - There should be a list of every single gesture that the user can access.
    - The user should be able to bookmark gestures so they can learn it later (could also link to a video).

**How you are going to build it:**

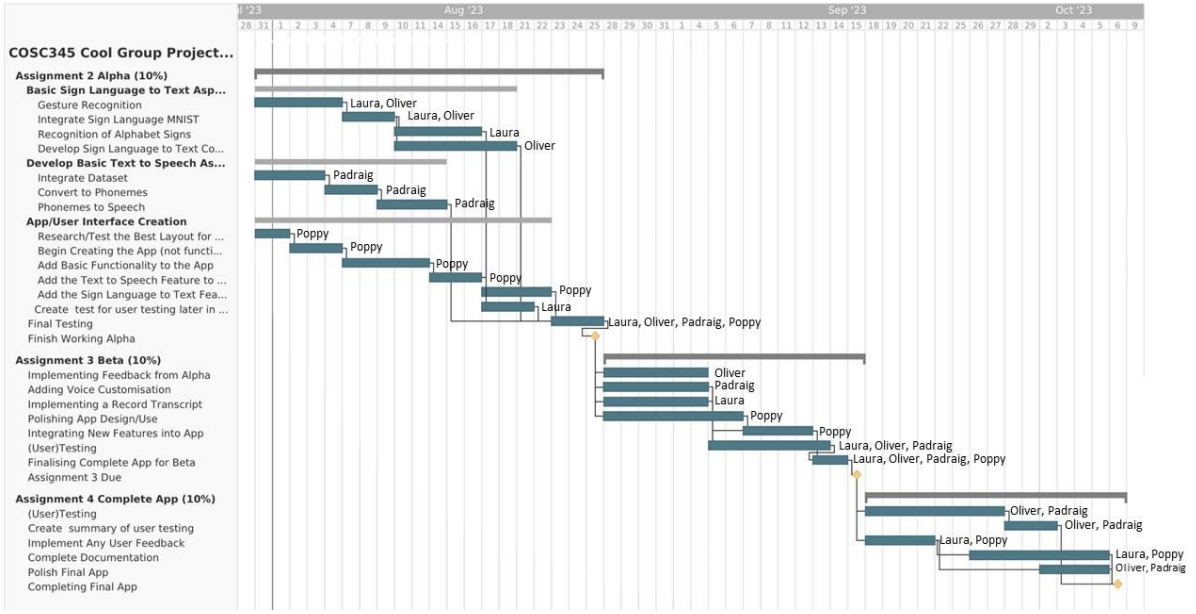
- We will program this in C++
- The application will be split into multiple parts:
    - The Sign Language to text component
        - Which will use Tensorflow Keras to implement Convolutional Neural Network model to build image recognition model
        - With the use of Tensorflow object detection API, will only detect hand in the frame and crop the space for Sign Language to be translated to text
    - The Text to Speech component
        - Which will convert text to phonemes then to speech
    - General Application
        - QT for the GUI
        - Below is an example of the layout we are planning for our app. We want the camera view to take up most of the space and have different sections/areas where additional features are shown.

- 


**Potential Risks**

- A member drops out/emergency
    - 10% chance
    - We will resolve this by trying to distribute the tasks evenly, so if a member can no longer participate in the project we will not be extremely hindered
- Societal Problem (e.g. global pandemic, natural disaster etc)
    - 2% chance
    - We will put measures in place for remote work, such as group chats or video calls
- Internet Failure
    - 5% chance
    - We will aim to set deadlines for important criteria a few days before it is actually required, so that if an unexpected event such as internet failure occurs there will be time to resolve this issue over the coming days.
- Data failure
    - 5% chance
    - By committing often in order to use GitHub as a backup, it should mitigate this, as only the data since the last commit should effectively be lost in this scenario.
- Burnout/loss of motivation
    - 30% chance
    - Other team members will offer support and tasks will ideally be distributed so each member gets time to rest and take breaks.
- Underestimation of Time
    - 50% chance
    - Other team members will help out with the task and when planning the schedule we took this into consideration and added a buffer to each task to accommodate for the high likelihood of underestimation of the time it takes to complete tasks.

**A schedule (including Pert or Gantt charts):**

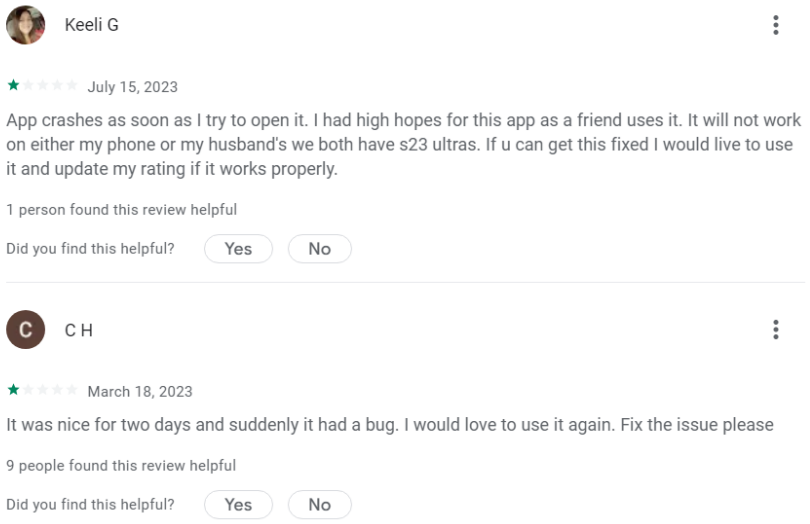



For the Alpha we plan to build a basic working application that contains only the basic and fundamental elements of our final product. This will include a interactable application that is able to translate Sign Language to Speech at a basic level. We will start to begin both debugging and user testing before the Alpha release. For the Beta we plan to use any user feedback to improve the application design/function as well as polishing the existing features and adding any additional features we wish to implement (i.e. voice customisation). We will continue the debugging and user testing over the course of the rest of the project, fixing any bugs and improving the application from user feedback. Then we will mostly focus on improving our application and writing any documentation/reports needed for the final due date. We predict that the Sign Language to Text will be our most challenging feature and so we decided to allocate two people to work on that together. Efficient communication between these two will be important to make sure there is no wasted time or energy in this phase. If any member has completed their scheduled task and has ‘free time’, we have planned that they will spend this time debugging, testing, or documenting to maximise progress.

**What already exists that is similar to your app:**

Hand Talk Translator (Only text/audio to ASL):

The Hand Talk Translator is the most popular app we found that is similar to our app. Hand Talk Translator has more than one million downloads on the google play store, showing the potential market we could be creating our app for. This app has the ability to convert text or speech to ASL or Brazilian Sign Language through AI. This is similar to our app, however this app does not give the ability to convert signs to text or speech. This would make our app different, and potentially more comprehensive and inclusive. The app has a huge amount of reviews, and negative reviews on the app would indicate that some of its users are looking for a better alternative. For instance, many users on the Hand Talk Translator app have reported issues with the app crashing or not opening.




signTranslator (Video to text):
signTranslator is another app that is similar to ours, however this time, signTranslator uses AI to convert video ASL signs to text. This is what we plan to do with our app, however our app will hopefully go a step further and speak the sentence formed from the signs. Video of ASL to text conversion seems rarer than the vice versa, therefore signTranslator seemed like serious competition for our app. Unfortunately, upon testing the app ourselves, we found the ASL video recognition pretty poor, and it struggled to convert most letters from the ASL alphabet. Despite the attempt to convert video ASL signs to text, signTranslator's poor ASL video recognition highlights a potential advantage for our app, which aims to take a step further by converting sign language to both text and speech, offering a more comprehensive and accurate communication solution.

**Evidence that there is customer interest in your app:**
Despite the Hand Talk Translator having many issues and only having one-way translation (text/speech to sign language), it has still gathered many downloads (1M+) and has an overall 4.5 star rating. We believe if we are able to potentially fix the issues of this app, there would be a huge customer interest for it as it would be able to be a better alternative than the already existing application.

There are negative reviews on many similar apps, which show that customers have an interest in the idea, but other applications have an implementation that doesn't appeal to them, or has issues which stops the application from being functional.

Many apps also do not have two way translation, and only do one way or the other. Our goal would then be to create a stable application which offers two way translation in order to create a new product.

**References:**

ASL to Text Dataset Kaggle: https://www.kaggle.com/datasets/datamunge/sign-language-mnist
Text to Speech Dataset Kaggle: https://www.kaggle.com/datasets/google-nlu/text-normalization
Text to Speech Dataset LJ Speech: https://paperswithcode.com/dataset/ljspeech
signTranslator: https://apps.apple.com/us/app/signtranslator/id1592412766
Hand Talk: https://play.google.com/store/apps/details?id=br.com.handtalk&hl=en&gl=US&pli=1
App Example Image:
https://dribbble.com/shots/12215430-Online-Education-Platform-Webinar-page/attachments/3830715?mode=media