Automatic Lip Reading

Jack Pay

July 4, 2023

Summary

The proposed software will detect and read the lips of a person, interpreting the shape of the lips as words and sentences. The software will use Computer Vision to detect and measure a person's lip movement and motion. A Machine Learning model will be trained to interpret these movements to calculate the most likely word being spoken and fit these into a sentence that is valid syntactically and semantically.

Aims

The following section covers the aims of the project.

- Detect and localise to a person's lips and record lip movement across frames
- Model should be made that is able to interpret lip movement as phonemes or words
- Need some form of natural language processing in order to put together phonemes or words into valid sentences to display
- System should have a User Interface to display the current video feed as well as the words being spoken

Plan

- Use Mediapipe to localise and detect user lips and convert lip movement to valid data of some form
- Train a model to use the lips data to recognise and interpret lip movement as either phonemes or words
- String together words and phonemes to create valid words and sentences
- Create a user interface to display video feed and words being interpreted

References

- Tool to capture user movement: `https://mediapipe.dev/`
- Dataset of lip reading: `https://www.robots.ox.ac.uk/~vgg/data/lip_reading/lrw1.html`
- Dataset of lip reading 2: `https://www.robots.ox.ac.uk/~vgg/data/lip_reading/lrs2.html`
- Similar work, lip reading using machine learning: `https://khazit.github.io/Lip2Word/`
- Similar work, lip reading via phonemes: `https://ietresearch.onlinelibrary.wiley.com/doi/full/10.1049/cit2.12131`

Tools

- Pytorch
- Mediapipe