

Week 4

Monday, October 06, 2025 2:47 PM

Dataset Storage Issues

- Dataset has large storage requirements
 - Advice on most optimal way to handle this

Manifest Structure for Training/Validation Data

- Working on getting a baseline test ready but struggling on what order AV-Hubert expects the manifest to be in
- To my knowledge all data processed and ready for training for a singular volunteer
 - 25 fps video
 - 16kHz audio (.wav)
 - Isolated lips and stored as singular frames (25 per second)
 - Txt files with transcribed audio

Work on smart blurring

- Working on a table that builds on the table taken from your paper (VISEME DEFINITIONS COMPARISON FOR VISUAL-ONLY SPEECH RECOGNITION)
 - Sort phonemes into groups and point to corresponding viseme
 - Outline the ROI's for each viseme
 - Part of the mouth and how it will be acting
 - Hope to map this visually in the future to aid the implementation when the time comes
- Looked into the Montreal Forced Aligner

Work on interpolation and Downsample

- Would be very interested in speaking to Professor Kokaram or Professor Pitie
 - Get a better idea of what tools are available for use in the neighborhood of interpolation
 - Need to come up with good technique to evaluate the quality of interpolation
 - Looked into some maths to keep in mind for the shift that will occur once the interpolated video is downsampled