Project Plan -

1st part -Sign Language(Video- Text)

ASLLD/ RWTH-PHOENIX-Weather 2014T(have corresponding Labels), Normalising the frames- then comes the Sign Language Recognition model - Spatio temporal data(CNN-LSTM), 3D CNN, Transformers for SLT , hand and body pose detection,training the model to map visual representations to txt

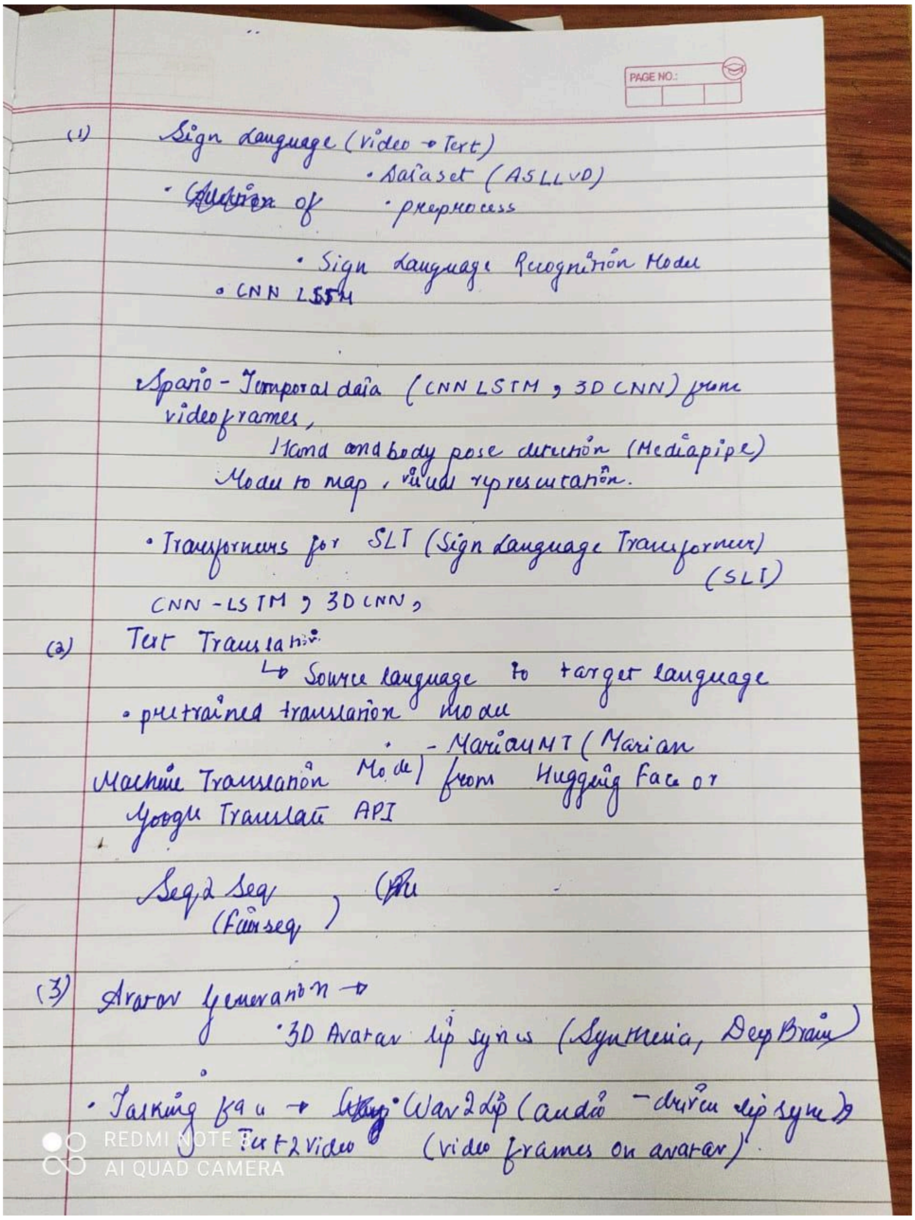
2nd part - (Text Translation)-

Translation model , API can be used for translation

3rd part (Text to video)

Avatar that lip syncs

https://pytorch.org/hub/snakers4_silero-models_stt/
https://ai4bharat.iitm.ac.in/areas/tts
https://github.com/neonbjb/tortoise-tts
https://github.com/aI4Bharat/IndicOOV



PAGE NO.:

(1) Sign Language (Video → Text)

- Creation of
  - Dataset (ASLLVD)
  - preprocess
  - Sign Language Recognition Model
- CNN LSTM

- Spatio - Temporal data (CNN LSTM, 3D CNN) from videoframes,
  Hand and body pose detection (Mediapipe)
  Model to map, visual representation.

- Transformers for SLT (Sign Language Transformer) (SLT)

CNN - LSTM, 3D CNN,

(2) Text Translation
↳ Source language to target language

- pretrained translation model
  - MarianMT (Marian Machine Translation Model) from Hugging Face or Google Translate API

Seq2 Seq (Fairseq)

(3) Avatar generation →
- 3D Avatar lip syncs (Synthesia, DeepBrain)
- Talking face → Wav2Lip (audio - driven lip sync), Text2Video (video frames on avatar).

https://pytorch.org/hub/snakers4_silero-models_stt/
https://ai4bharat.iitm.ac.in/areas/tts
https://github.com/neonbjb/tortoise-tts
https://github.com/aI4Bharat/IndicOOV