

Lip Sync Assignment Solution

[prepared by: Shobika NKS]

1.Problem Statement:

The assignment requires developing a solution that maps audio inputs to corresponding lip movements in video content, ensuring high accuracy and minimal latency

2.Approach:

Convert text to speech using Edge TTS, then generate a realistic lip-synced video with Wav2Lip using a GAN-based model for natural lip movements. The process ensures high accuracy and temporal consistency in the output.

3. Workflow:

The lip-syncing project uses a clever mix of Edge TTS and Wav2Lip models, with a bit of magic from Generative Adversarial Networks (GANs) to create natural and realistic lip movements. It all starts with converting text into speech using the `edge_tts` library. By selecting the `en-IN-NeerjaNeural` voice model, the generated audio sounds clear, natural, and perfectly suited for the assignment. Once the audio file (`unique_audio_edge_tts.wav`) is ready, the Wav2Lip model steps in. This model is fantastic for lip-syncing because it not only syncs the lips accurately but also ensures smooth transitions between frames, giving the video a natural feel. It takes an input image (`input1.jpeg`) and the generated audio, then uses a pre-trained GANcheckpoint (`wav2lip_gan.pth`) to create a lip-synced video (`lip_synced_unique_output.mp4`). The real star here is the GAN architecture, which works in a unique way: the Generator creates frames with synchronized lip movements, while the Discriminator plays the critic, ensuring the movements match the audio perfectly. This back-and-forth between the Generator and Discriminator keeps improving the model's output, resulting in a super-realistic video. Finally, a simple Python script runs the whole process, checking if the audio file exists and generating the output video. This method blends advanced TTS and GAN-based lip-syncing technologies to produce a polished and professional-looking video, showcasing a practical and efficient solution to the lip-syncing assignment.



