Audio signal Mel Spectrogram Latent space trajectory Translated video frames (face decoder) Translated video frames (number decoder) Translated video frames (face decoder-high resolution)