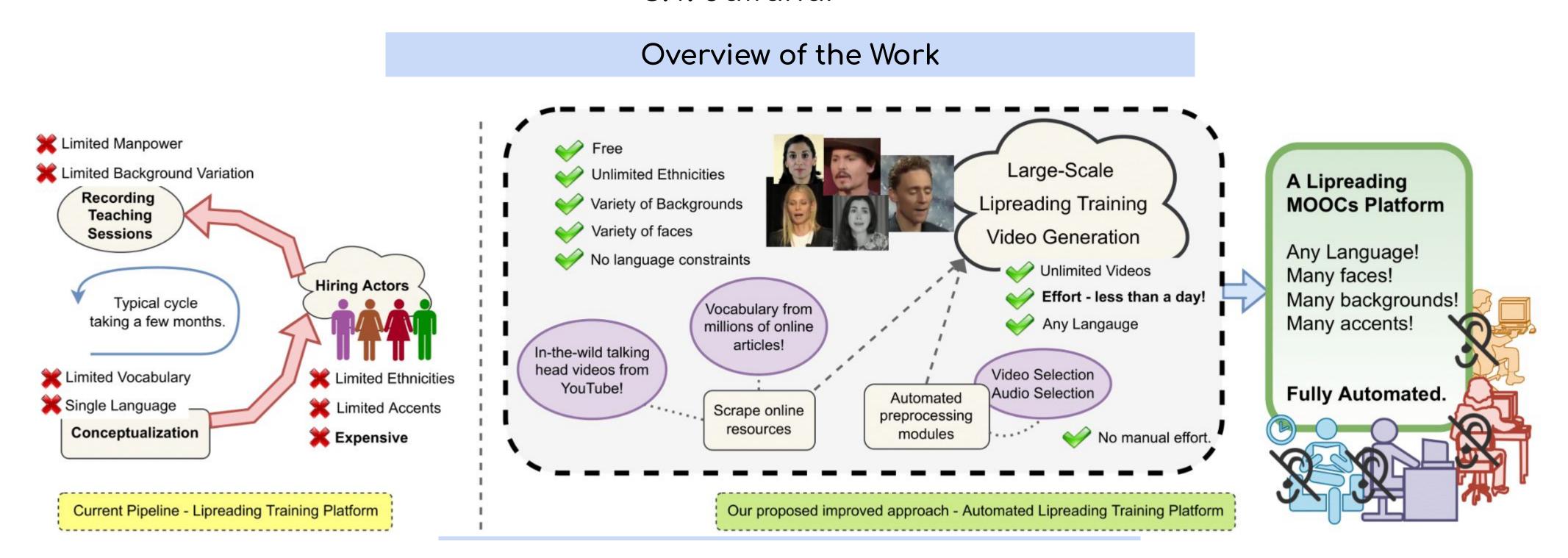


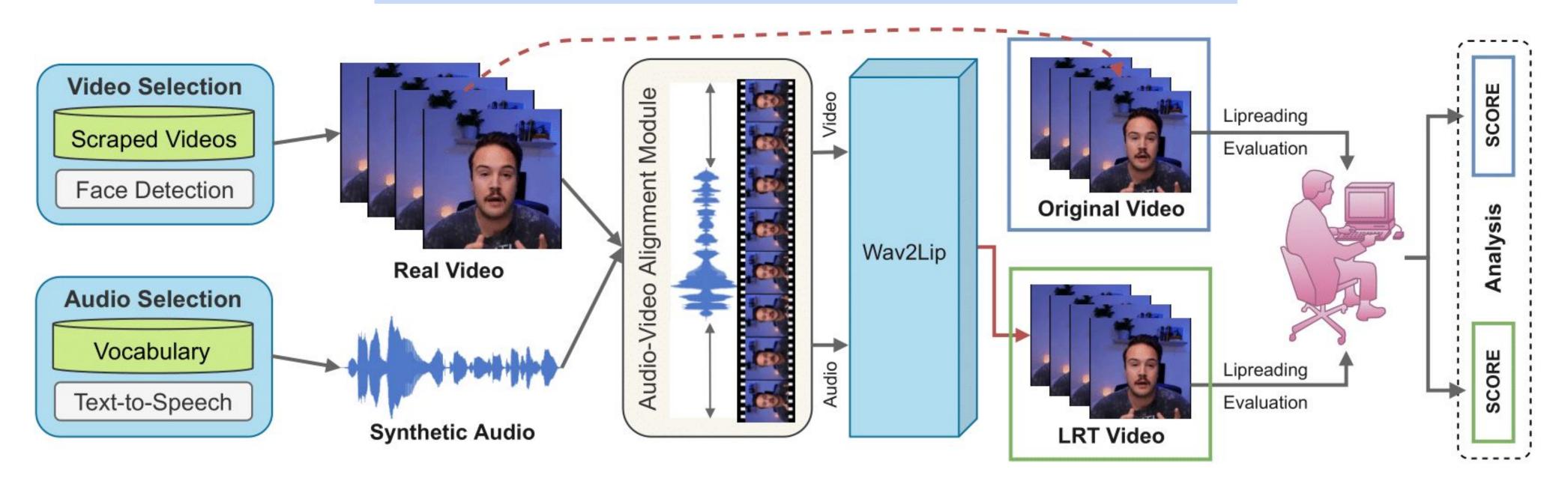
## Towards MOOCs for Lipreading: Using Synthetic Talking Heads to Train Humans in Lipreading at Scale

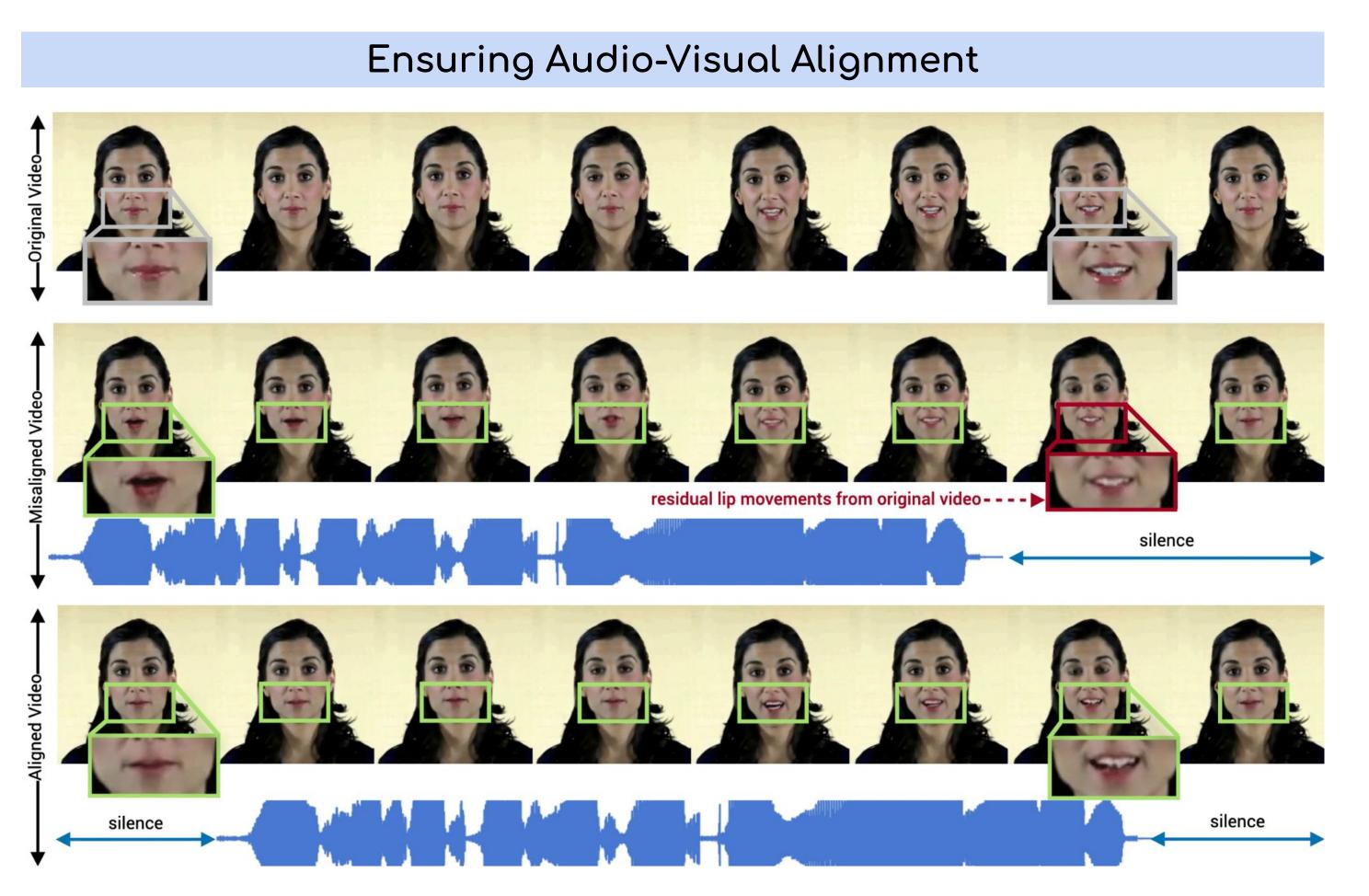


Aditya Agarwal\*, Bipasha Sen\*, Rudrabha Mukhopadhyay, Vinay Namboodiri, C.V. Jawahar

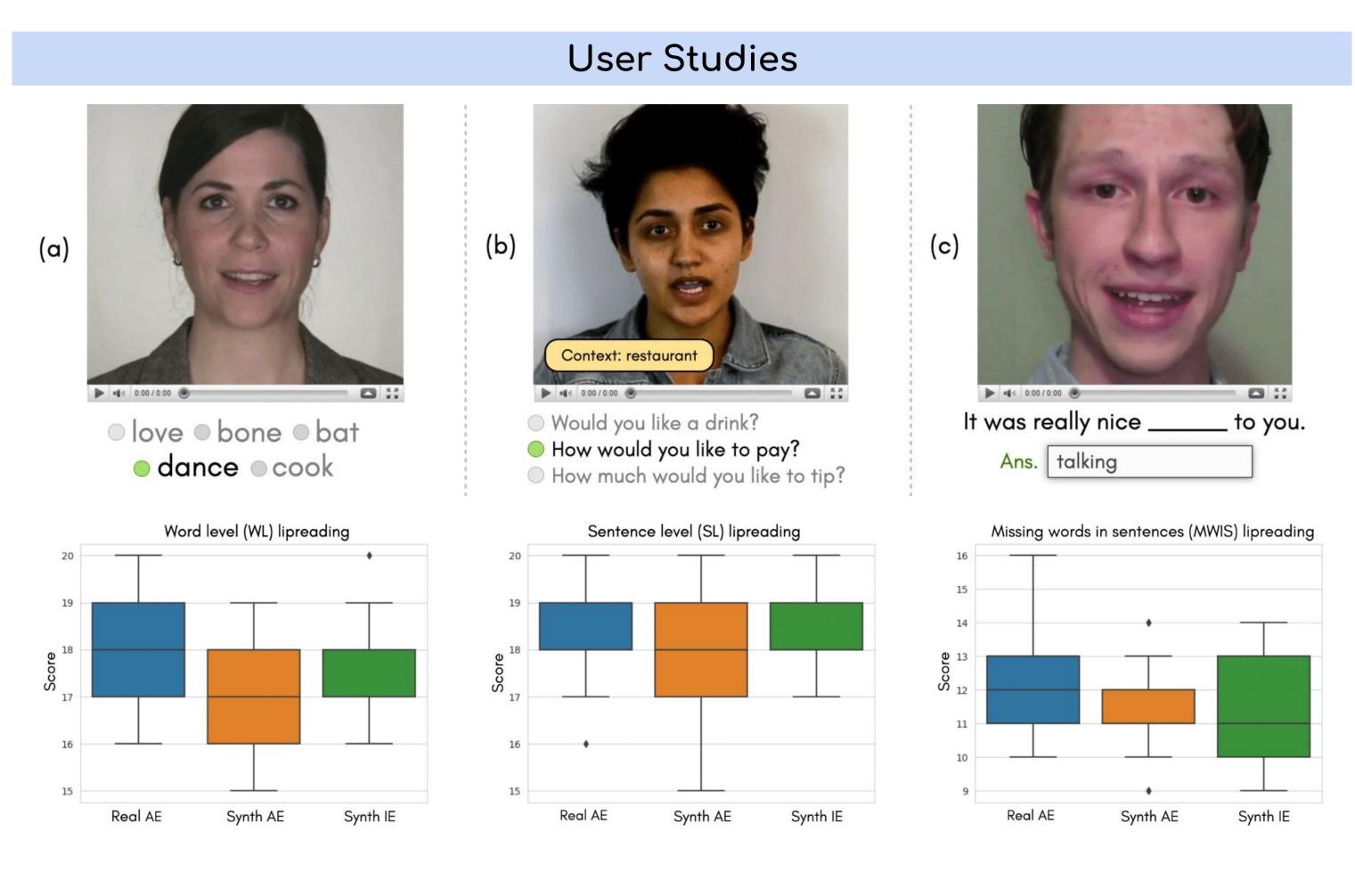


## Pipeline for Generating Synthetic Videos





Wav2Lip modifies the 'Original Video' lip movements according to a given speech utterance. Naively aligning the audio and video can result in residual lip movements. Our audio-video alignment module detects mouth movements in the 'original video', aligns the speech utterance on the region with the mouth movements, and adds silence around the aligned utterance.



We conduct user studies on three lipreading protocols: (a) lipreading isolated words, (b) lipreading sentences with context, and (c) lipreading missing words in a sentence. Our study of 50 participants validated that synthetic videos generated using our pipeline can replace real videos. We also show that users are more comfortable lipreading in their native accent/language.

