The Interplay of Speech And Lip Movements

Presenters: K R Prajwal, Rudrabha M*, Vinay P N, C V Jawahar*
Contact: prajwalrenukanand@gmail.com

Abstract

The strong connection between speech and lip movements allows us to understand and predict one from the other. In this demo video, we extensively explore this close relationship through three state-of-the-art efforts.

We start with the practically important task of lip-syncing arbitrary talking face videos to the desired target speech. We demonstrate two models for this task, each addressing a significant limitation of the previous works. The first model LipGAN works for any identity, voice, and language. More importantly, it is the first model to generate lip-synced face images that can be seamlessly pasted back into the video frame. We follow this up with a second model, Wav2Lip, that uses a significantly more accurate lip-sync discriminator. Consequently, we can lip-sync arbitrary videos in the wild with unprecedented accuracy and quality. Our models make several real-world applications possible. In this video, we show three such applications that can have the most substantial impact: (i) correcting lip movements in (automatically) dubbed movies and lecture videos, (iii) lip-syncing translated press conferences, interviews, and speeches, and (ii) lip-syncing CGI characters in animated films and gaming videos to any speech.

Our efforts are “bi-directional”, i.e., similar to how we can accurately generate lip movements from speech, we also try the reverse problem of generating speech solely from the lip movements. We explore this question from a data-driven learning perspective, *“How accurately can we infer an individual’s speech style and content from his/her lip movements?”* We demonstrate that we can generate accurate speech from silent lip videos in large vocabulary, unconstrained settings for the first time. In addition to single-speaker lip to speech, we are the first to demonstrate results on the highly challenging task of “word-level multi-speaker lip to speech”. Our effort leads to applications such as (i) generating a voice for people suffering from aphonia, and (ii) generating speech in case of a loss of signal.

All our models, code, and datasets are released publicly. We also have an interactive lip-sync demo, which allows you to try out your own video and audio samples. Please visit our website for all the details.

Demonstration Website

For further information, please visit
sites.google.com/view/interplay-of-speech-and-lips