We devise a cascade GAN approach to generate talking face video, which is robust to different face shapes, view angles, facial characteristics, and noisy audio