> TL;DR InfiniteTalk is an audio-driven dubbing framework that can generate long talking videos with synchronized lips, head/body motion, and facial expressions. It works as video-to-video or image-to-video, includes acceleration (TeaCache) and quantization options, and exposes practical flags for controlling length, quality, and VRAM.

What is InfiniteTalk?

InfiniteTalk proposes a sparse-frame video dubbing approach that goes beyond lip-only edits. Given an input video (V2V) or a single image (I2V) plus an audio track, it synthesizes a new video with:

- Lip synchronization to the audio
- Coordinated head movements and body posture
- Facial expressions aligned to speech
- Identity preservation across long durations

Links:

- Project: https://meigen-ai.github.io/InfiniteTalk/
- Paper (tech report): https://arxiv.org/abs/2508.14033
- Code: https://github.com/MeiGen-AI/InfiniteTalk
- Models: https://huggingface.co/MeiGen-AI/InfiniteTalk

Highlights

- Sparse-frame dubbing: edits lips, head, body, and expressions, not just lips
- Infinite-length generation: streaming mode can produce long videos
- Stability vs. prior baselines: reduces hand/body distortions (per repo notes)
- Lip accuracy: improved sync compared to MultiTalk (qualitative claims in README)
- Modes: V2V (mimics original camera motion) and I2V (single image → video)

Notes: For very long clips, the README mentions potential color shift; suggested mitigations include SDEdit for short clips and simple image-to-video camera movement tricks for I2V.

Quick Start (inference)

Environment (abbrev.):

``(bash conda create -n infinitetalk python=3.10 conda activate infinitetalk

PyTorch (CUDA build), flash-attn, core deps

pip install torch torchvision torchaudio --index-url https://download.pytorch.org/whl/cu124 pip install flash,\text{attn}==2.7.4.post1 pip install -r requirements.txt conda install -c conda-forge ffmpeg librosa)``

Models (from README table):

```
huggingface-cli download Wan-AI/Wan2.1-I2V-14B-480P --local-dir ./weights/Wan2.1-I2V-14B-480P
huggingface-cli download TencentGameMate/chinese-wav2vec2-base --local-dir ./weights/chinese-wav2vec2-base
huggingface-cli download MeiGen-AI/InfiniteTalk --local-dir ./weights/InfiniteTalk
```

Run (single-GPU, streaming mode):

```
python generate_infinitetalk.py \
  --ckpt_dir weights/Wan2.1-I2V-14B-480P \
  --wav2vec_dir weights/chinese-wav2vec2-base \
  --infinitetalk_dir weights/InfiniteTalk/single/infinitetalk.safetensors \
  --input_json examples/single_example_image.json \
  --size infinitetalk-480 \
  --sample_steps 40 \
  --mode streaming \
  --motion_frame 9 \
  --save_file infinitetalk_res
```

720p variant: set `--size infinitetalk-720` and adjust compute accordingly.

Practical flags and tips

- Modes:
  - `--mode streaming` for long videos; `--mode clip` for single-chunk output.
- Guidance scales:
  - `--sample_text_guide_scale` (text adherence)
  - `--sample_audio_guide_scale` (lip/body sync) — README suggests ~4 without LoRA, ~2 with LoRA.
- VRAM saving:
  - `--num_persistent_param_in_dit 0` to run with very low VRAM.
  - Quantized weights are available to reduce memory.
- Acceleration:
  - `--use_teacache` and `--teacache_thresh` for TeaCache speed-ups.
- Length:
  - `--max_frame_num` controls duration; default ~40 s (1000 frames). Longer clips may need quality trade-offs.
- V2V vs. I2V:
  - V2V mimics original camera motion (not exact); SDEdit improves accuracy for short clips but may introduce color shift.
  - I2V can run >1 min, but color shift increases; consider simple image→video panning/zooming tricks.

References

- Project: https://meigen-ai.github.io/InfiniteTalk/
- Tech report: https://arxiv.org/abs/2508.14033
- Repo: https://github.com/MeiGen-AI/InfiniteTalk
- Models: https://huggingface.co/MeiGen-AI/InfiniteTalk

  Notes: Claims about stability/lip accuracy and long-duration behavior reflect the project README at publish time. Test with your own assets; performance depends on hardware, settings, and inputs.