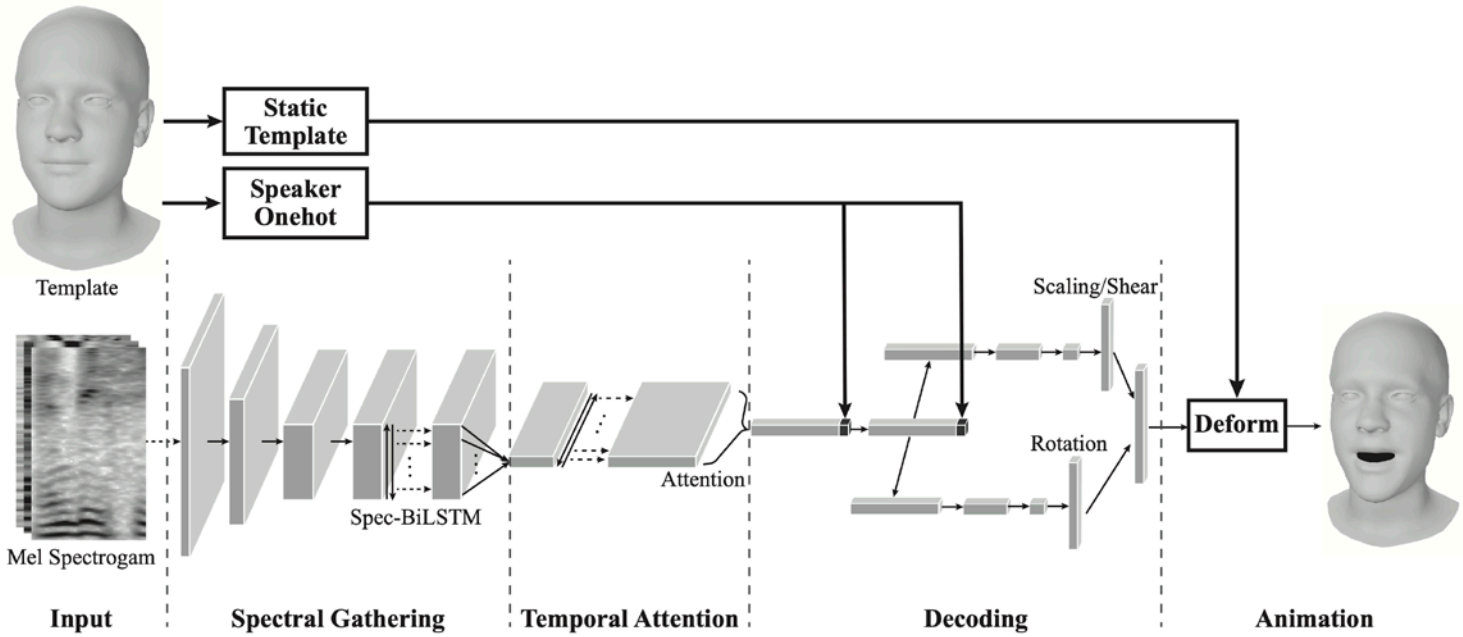
Speech-Driven Facial Animation with Spectral Gathering and Temporal Attention

Yujin CHAI, Yanlin WENG, Lvdi WANG, Kun ZHOU

Frontiers of Computer Science, DOI: 10.1007/s11704-020-0133-7

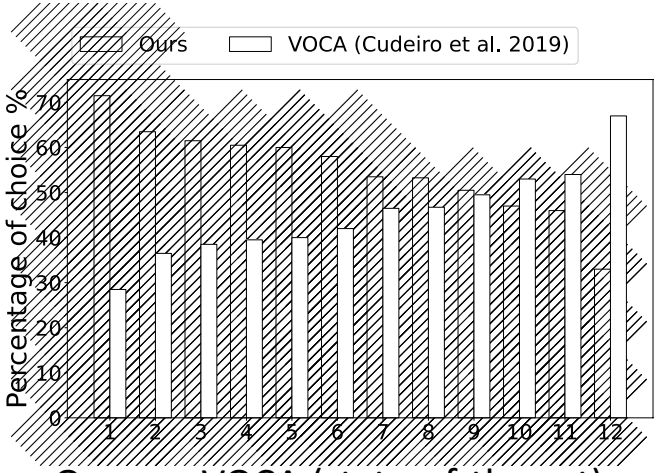
Problems & Ideas

- Two challenges for speech-driven facial animation:
  - Robust audio processing;
  - Effortless motion retargeting.
- Ideas:
  - A robust and lightweight speech encoder tailored to our task:
    - ***Spectral-dimensional*** Bi-LSTM network;
    - ***Temporal*** attention mechanism.
  - **Deformation gradients** as a representation of facial motion:
    - Represent nuanced local motions;
    - Easily applied to new 3D avatars with different topologies.



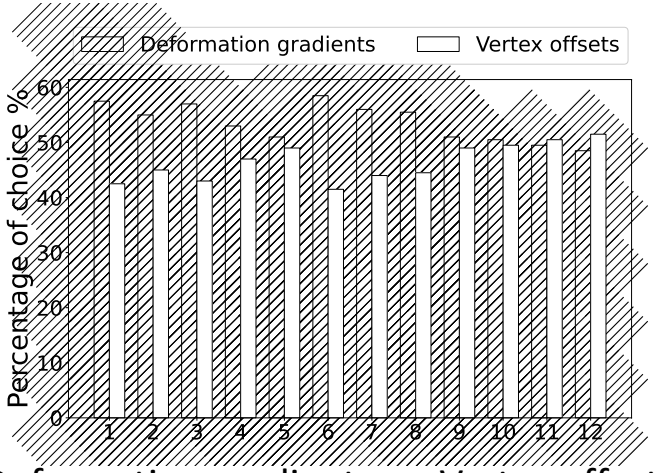
Experimental results

- User study



Ours vs. VOCA (state-of-the-art)

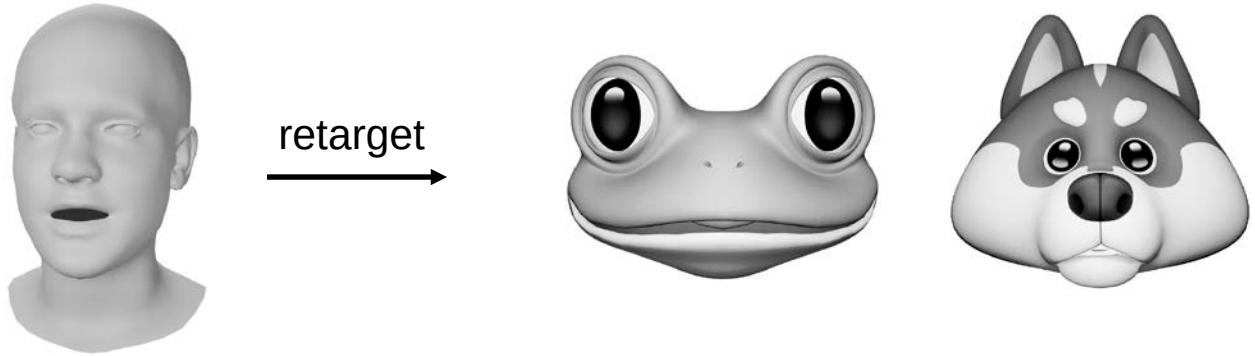
Users prefer ours (54.86%±10.02%)



Deformation gradients vs. Vertex offsets

Users prefer former (53.58%±3.42%)

- Topology-independent retargeting



More results available at https://chaiyujin.github.io/sdfa