

Docket #: S23-435

Multimodal machine learning for improved decoding of silent speech

Silent speech interfaces (SSIs) offer a non-invasive alternative to brain-computer interfaces for silent verbal communication. However, available SSIs have limited accuracy. Stanford researchers have therefore developed a new multimodal algorithm for decoding silent, attempted, or imagined speech.

Researchers developed a new algorithm in which many different data modalities (audio, EMG, neural microelectrode arrays, etc.) are encoded via artificial neural networks. Innovative formulations of contrastive loss functions encode each data modality into a unified latent representation. This unified approach allows for the decoding of each individual data modality with superior efficacy and represents a leap forward in the field of multimodal machine learning for speech decoding.

Stage of Development

Prototype: achieves 12.2% word error rate on silent EMG and 3.7% word error rate on vocal EMG (significantly superior to state-of-the-art)

Applications

- Decoding and synthesis of text and audio from various speech forms, including verbalized, silent, attempted, and imagined speech
- Communication devices for individuals with speech impediments and other conditions that impede speech
- Consumer devices for communicating via subvocalization
- New interfaces for conversational AI powered by silent speech

Advantages

- Superior accuracy over existing silent speech interfaces

Publications

- Tyler Benster, Guy Wilson, et al. (2024). [A Cross-Modal Approach to Silent Speech with LLM-Enhanced Recognition](#). arXiv 2403.05583.

Innovators

- Shaul Druckmann
- Tyler Benster
- Guy Wilson
- Reshef Elisha

Licensing Contact

Evan Elder

Senior Licensing Associate

[Email](#)