TikTok Parent Unveils PolyVoice, Speech-to-Speech Translation with Language Models

ByteDance, TikTok's parent company, is stepping up in the speech-to-speech translation (S2S) game with its newly proposed PolyVoice – a language model-based framework.

Announced in a research paper on June 13, 2023, the China-based tech company introduces a decoder-only model to enable direct translation, diverging from the traditional encoder-decoder modeling, which remains prevalent in speech modeling.

As noted in the Slator Interpreting Services and Technology Report, published in late 2022, research and development activity in S2S translation is booming. Meta has contributed to data collection through the release of a large-scale multilingual corpus. Rival tech giant Google has been active in technological development, as demonstrated by the release of its fully unsupervised Translatotron3 model.