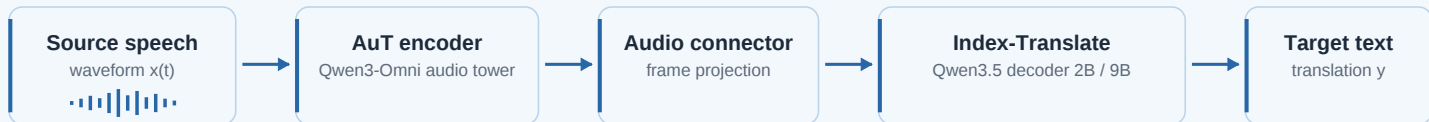


Index-Echo: from speech translation to speech generation

A Speech-to-text translation (S2TT)

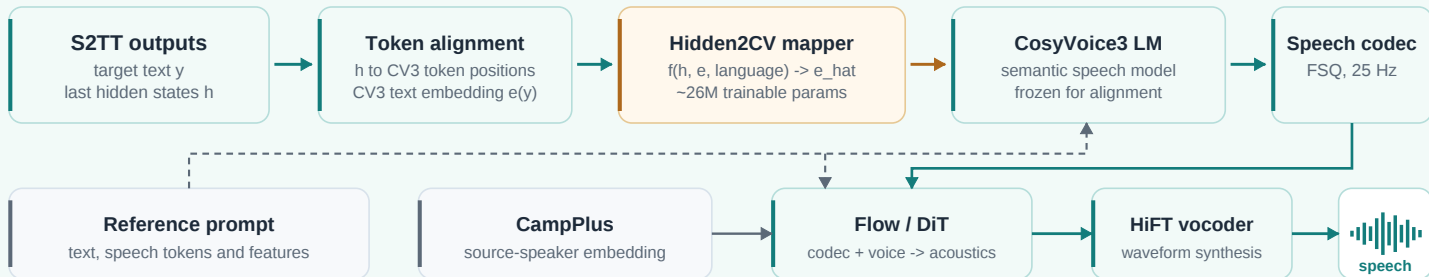
end-to-end supervised



S2TT training updates the speech encoder, connector, and translation decoder from speech-text supervision.

B Speech-to-speech translation (S2ST)

source-voice conditioning



Source speech also supplies the reference prompt and the CampPlus voice condition; the translated-text slot is replaced by $\hat{e}$.

TRAINING SEQUENCE

