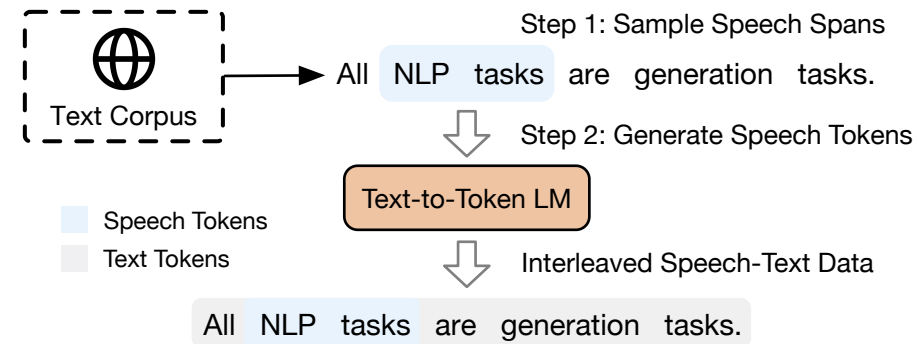


(a) Train a Text-to-Token Model using TTS data



(b) Construct Interleaved Speech-Text Data From Text Corpus

