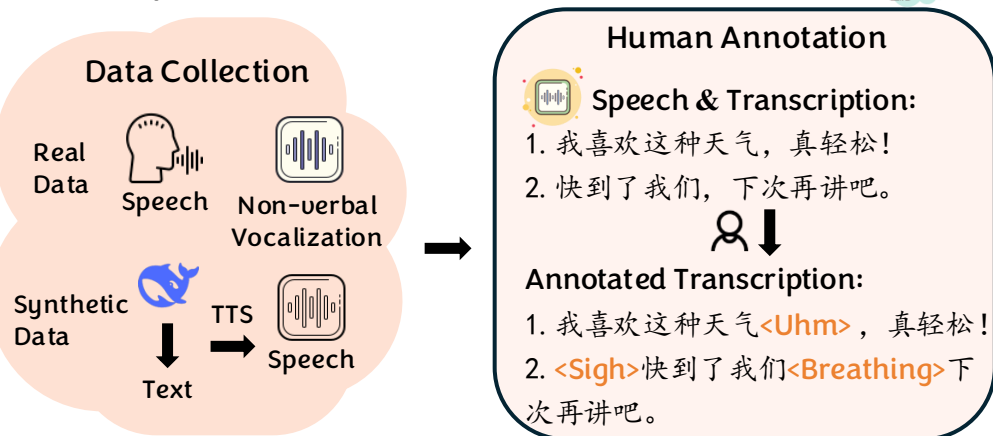
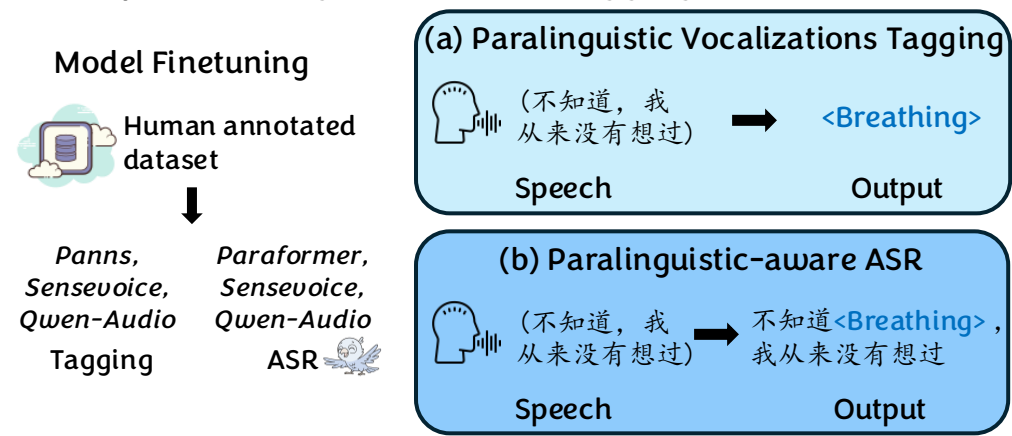


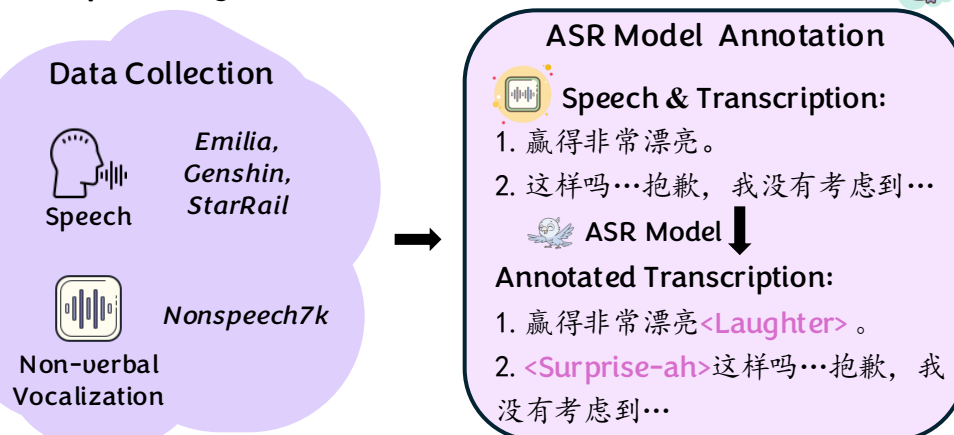
Step 1: Word-level Human Annotation Dataset



Step 2: Paralinguistic-aware Tagging & ASR Model



Step 3: Large-scale Automatic Annotation Dataset



Step 4: Human-like TTS

