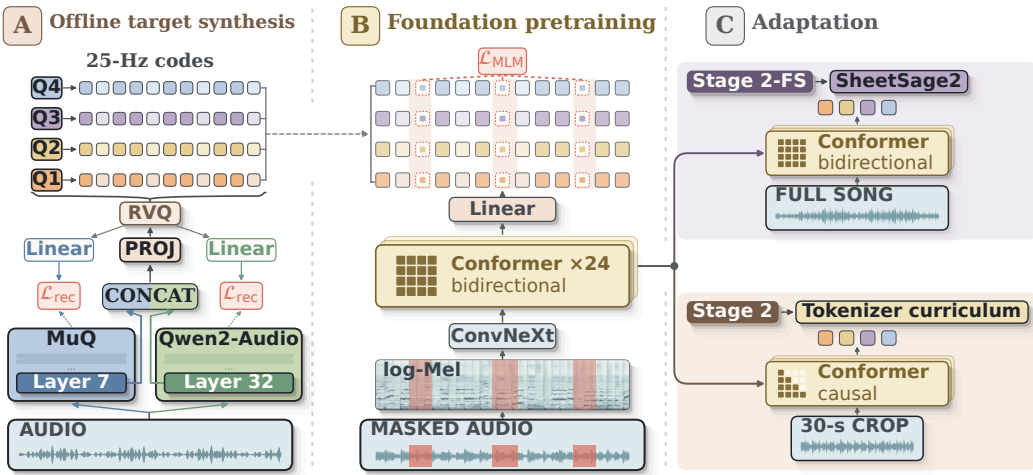
A
Offline target synthesis
25-Hz codes
Q4
Q3
Q2
Q1
RVQ
Linear
PROJ
Linear
$\mathcal{L}_{\text{rec}}$
CONCAT
$\mathcal{L}_{\text{rec}}$
MuQ
Layer 7
Qwen2-Audio
Layer 32
AUDIO
B
Foundation pretraining
$\mathcal{L}_{\text{MLM}}$
Linear
Conformer ×24
bidirectional
ConvNeXt
log-Mel
MASKED AUDIO
C
Adaptation
Stage 2-FS
SheetSage2
Conformer
bidirectional
FULL SONG
Stage 2
Tokenizer curriculum
Conformer
causal
30-s CROP