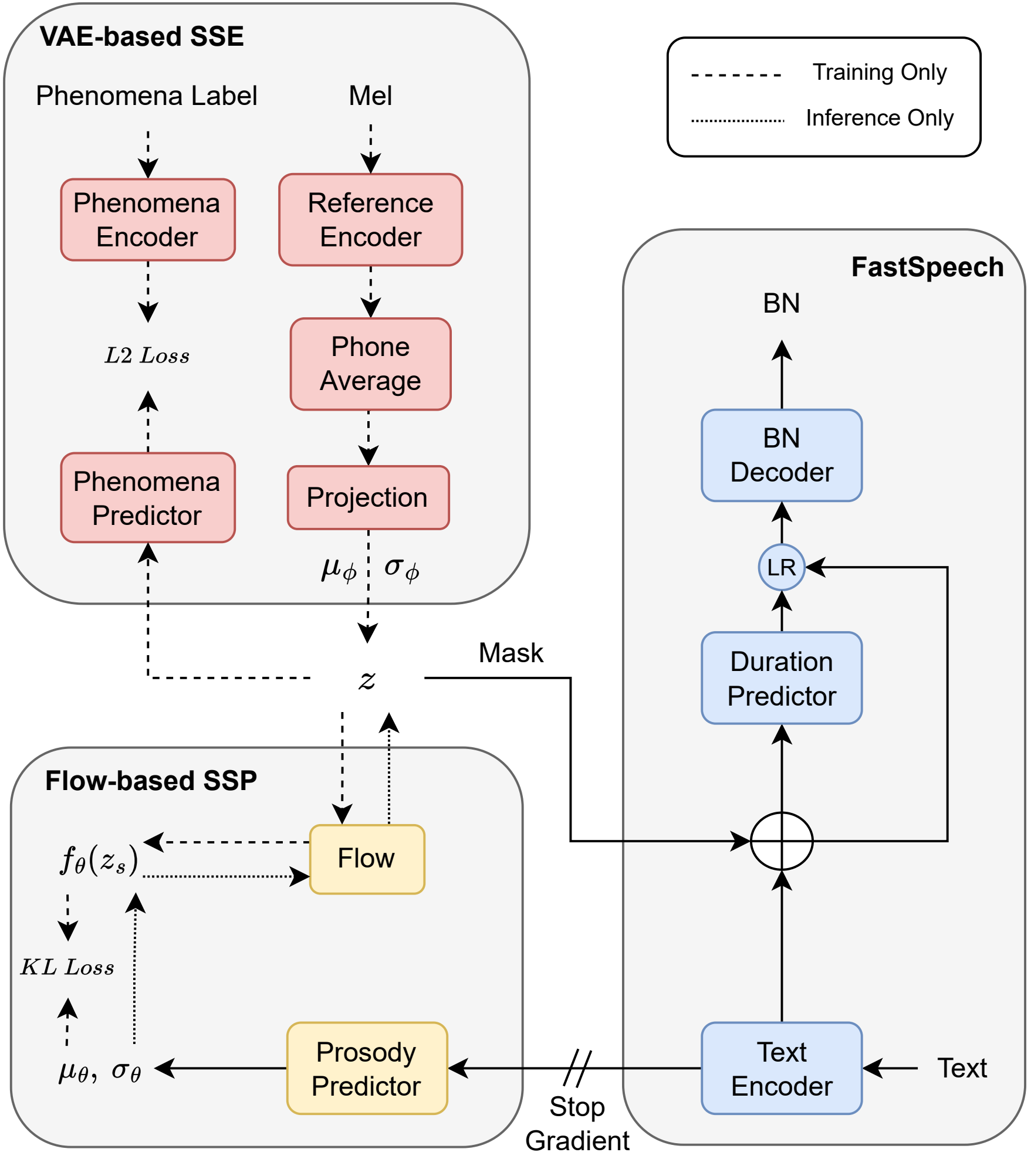
VAE-based SSE
Phenomena Label
Mel
Phenomena Encoder
Reference Encoder
L2 Loss
Phone Average
Phenomena Predictor
Projection
μ_ϕ σ_ϕ
Training Only
Inference Only
FastSpeech
BN
BN Decoder
LR
Duration Predictor
Mask
z
Flow-based SSP
$f_\theta(z_s)$
Flow
KL Loss
μ_θ, σ_θ
Prosody Predictor
Stop Gradient
Text Encoder
Text