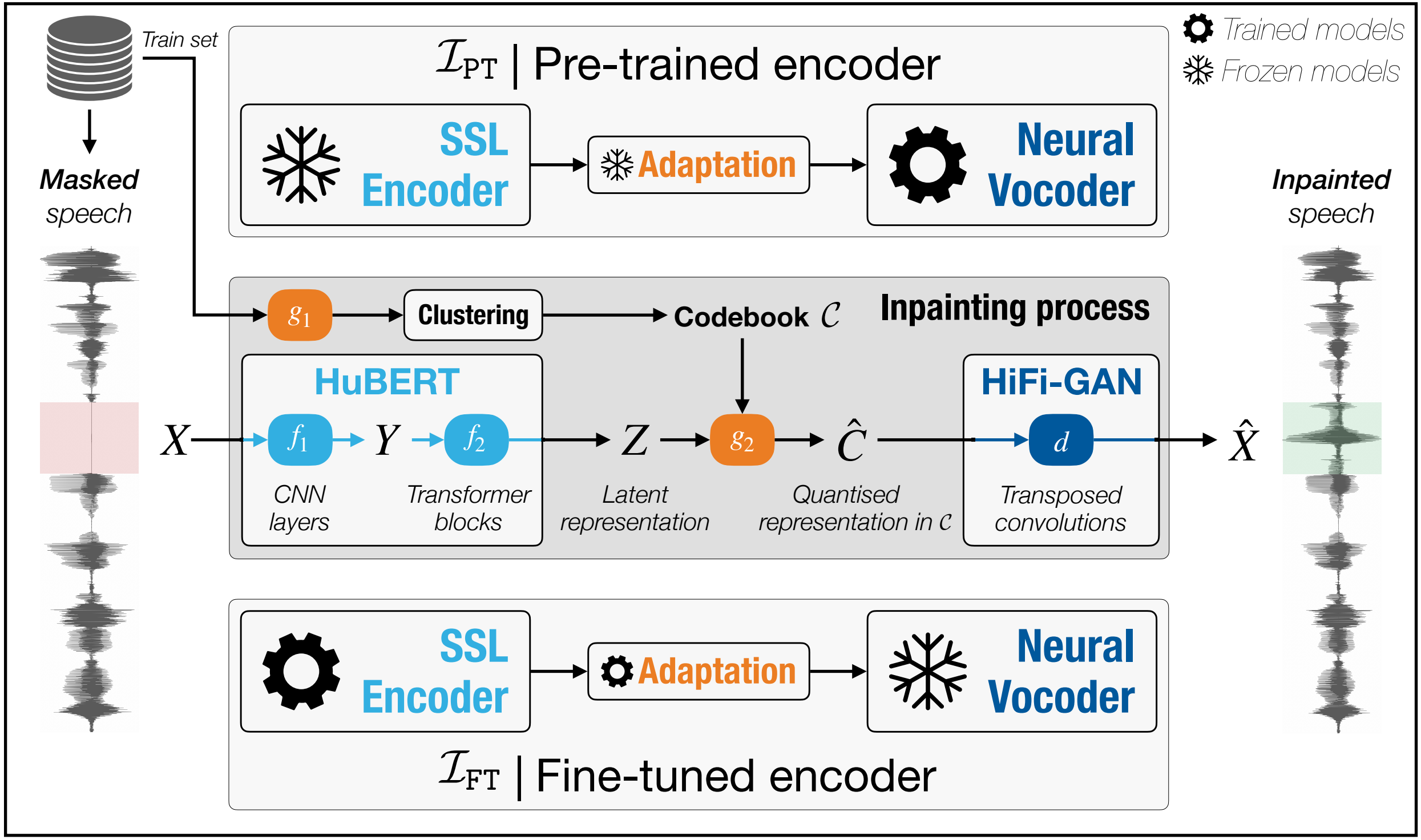
Train set
Trained models
Frozen models
$\mathcal{I}_{\mathrm{PT}}$ | Pre-trained encoder
SSL Encoder
Adaptation
Neural Vocoder
Masked speech
Inpainted speech
g_1
Clustering
Codebook $\mathcal{C}$
Inpainting process
HuBERT
HiFi-GAN
X
f_1
Y
f_2
Z
g_2
$\hat{C}$
d
$\hat{X}$
CNN layers
Transformer blocks
Latent representation
Quantised representation in $\mathcal{C}$
Transposed convolutions
SSL Encoder
Adaptation
Neural Vocoder
$\mathcal{I}_{\mathrm{FT}}$ | Fine-tuned encoder