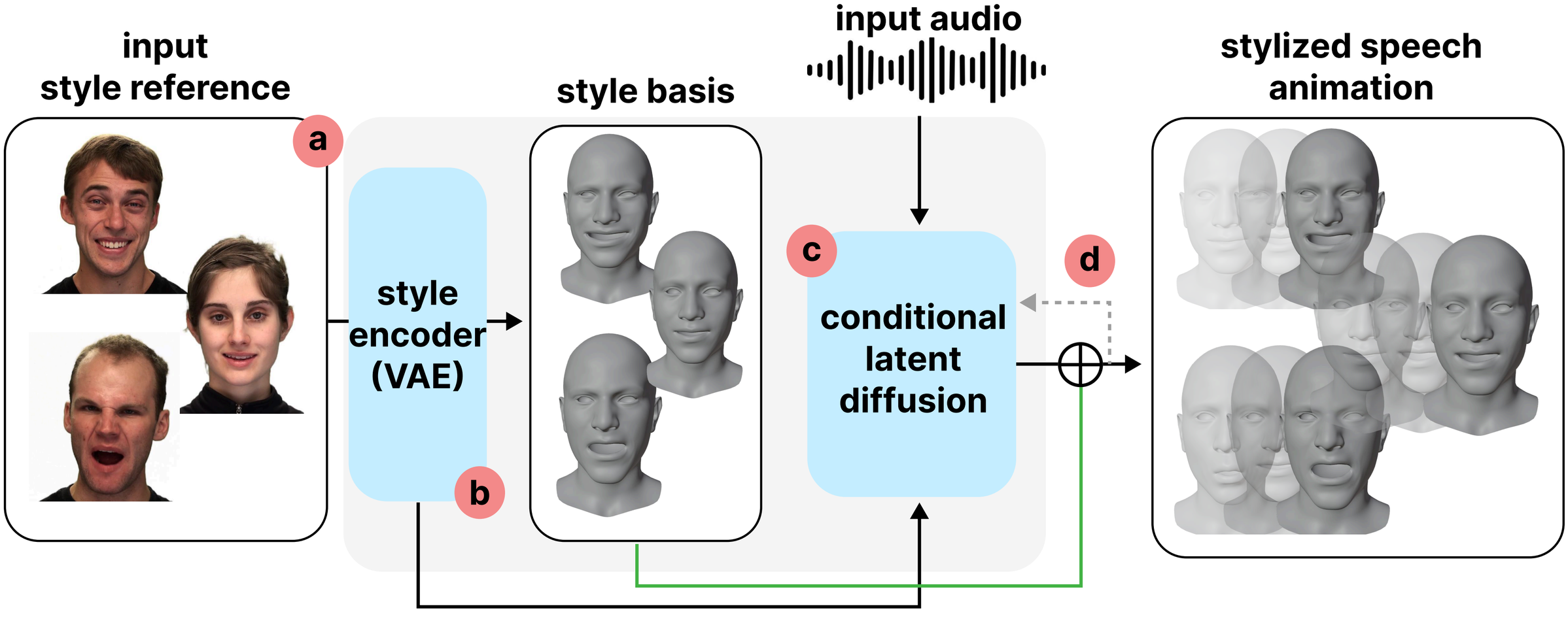
input
style reference
a
style basis
input audio
stylized speech
animation
style
encoder
(VAE)
b
c
conditional
latent
diffusion
d