End-to-End Architecture Diagram

1. Audio Encoder
Pretrained EfficientNet from DeepAFX-ST paper.

2. Controller VAE
Input is the audio encoding from previous layer.
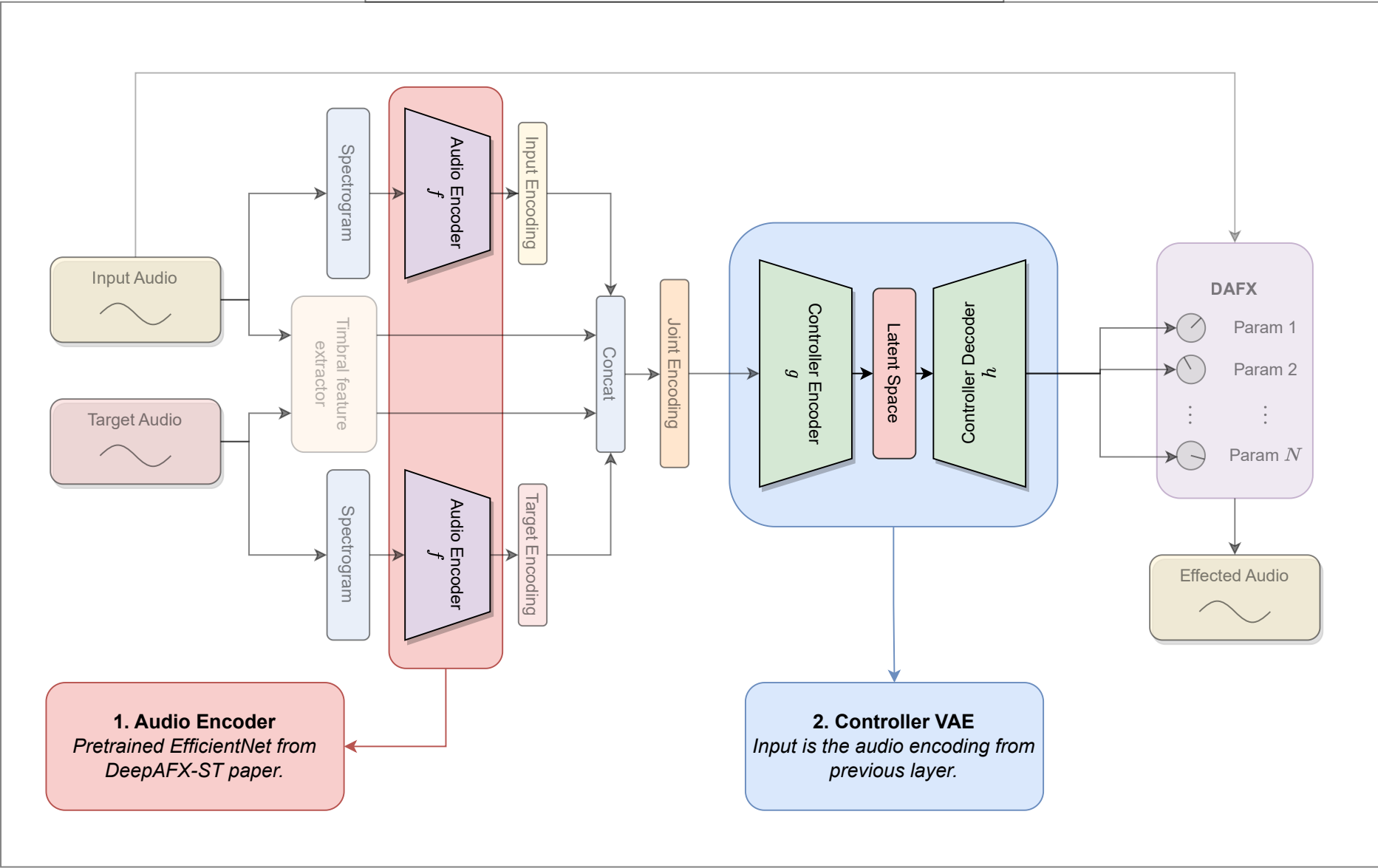
Spectrogram VAE Architecture Diagram

3. Spectrogram VAE
Itention is to train this and use the Spectrogram Encoder to replace the pre-trained Audio Encoder above.