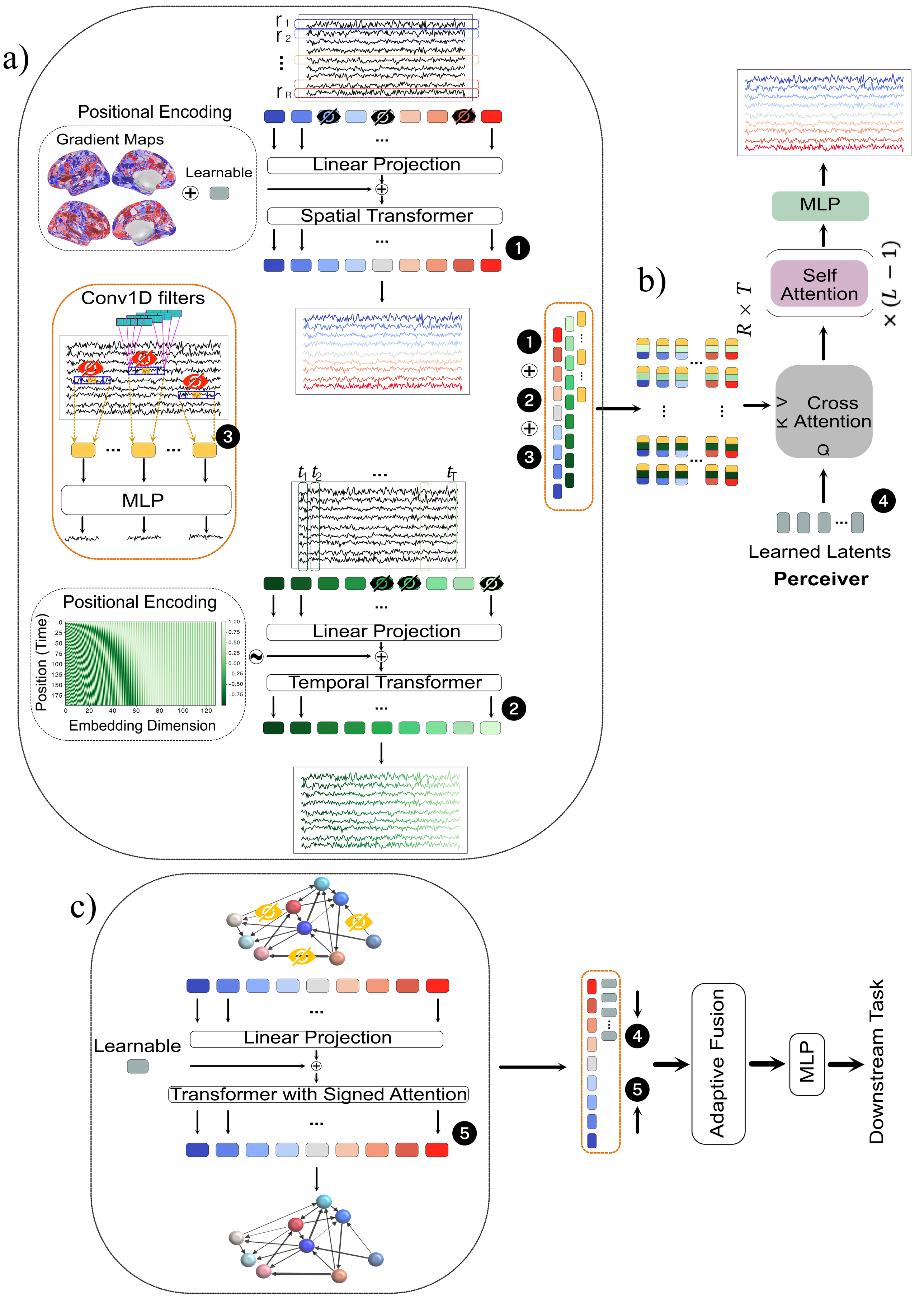
a)
Positional Encoding
Gradient Maps
Learnable
Linear Projection
Spatial Transformer
Conv1D filters
MLP
Positional Encoding
Position (Time)
Embedding Dimension
Linear Projection
Temporal Transformer
b)
MLP
Self Attention
Cross Attention
Q
K, V
R ×
(L − 1) ×
Learned Latents
Perceiver
c)
Learnable
Linear Projection
Transformer with Signed Attention
Adaptive Fusion
MLP
Downstream Task