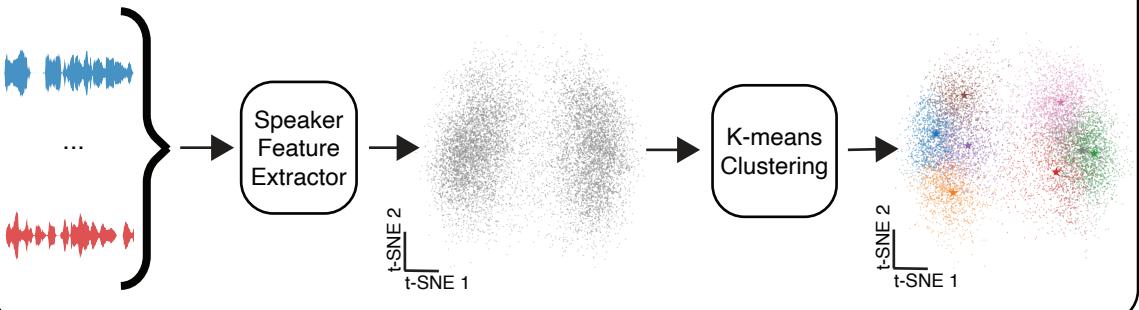
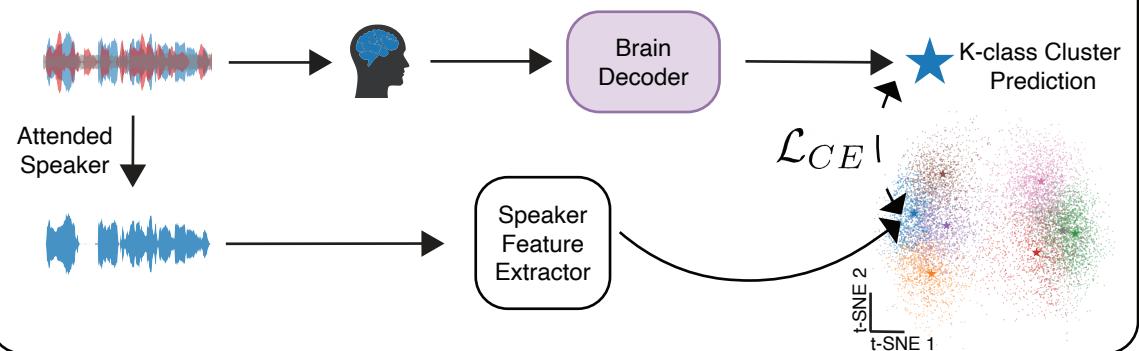


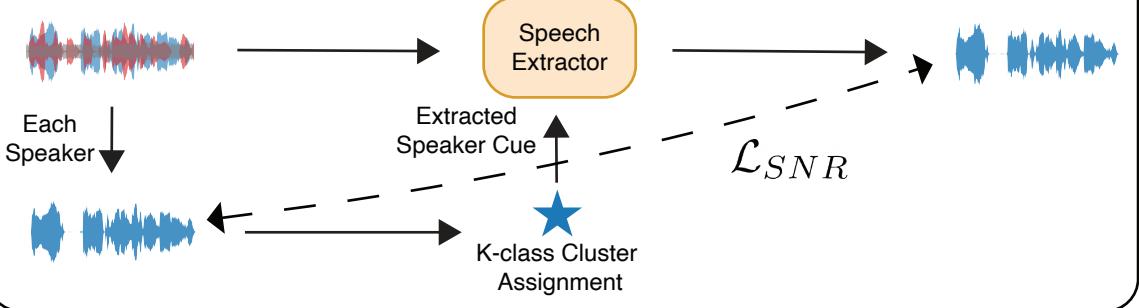
Step 1 - Speech Clustering



Step 2 - Neural Decoder Training



Step 3 - Speaker Extractor Training



Inference

