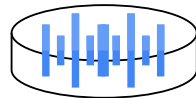


corpus: UltraSuite/TIMIT/TORGO

wav2vec2-xls-r-1b

ARPABET phonemes **Forced**



raw audio

Feature Extractor

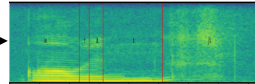
Transformer Layer

FC Layer

CTC
Loss

AO R IH N JH

Alignment
Algorithm



aligned
phonetic
transcription

wav2vec 2.0

projection

phoneme
sequence