FastPitch Architecture

Parallel Generation
(Non-autoregressive)

Pitch
Embedding

Transformer
Decoder
(Self-Attention)

Mel-spectrogram
Output

Pitch
Predictor

Transformer
Encoder
(Self-Attention)

Length
Regulator

Explicit Duration
and Pitch Control

Character
Embedding

Duration
Predictor

Text Input
(Phonemes)

Duration
Targets
(Training)

Pitch
Targets
(Training)

INPUT

PREDICTION

DECODER

OUTPUT