

Stage 1

MERT

Codeword Embeddings 🔥

Music MLM loss

Acoustic MLM loss

Transformer Encoder

1D convolution feature extractor 🔥

100-hour Multi-Cultural Audio Data
(20% Music4All)

Learning Rate Re-Warming

5e-4

5e-5

10% Warm-up

Stage 2

MERT

Codeword Embeddings 🔥

Music MLM loss

Acoustic MLM loss

Transformer Encoder

1D convolution feature extractor 🔥

650-hour Multi-Cultural Audio Data

Learning Rate Re-Warming

5e-5

5e-6

1% Warm-up