

Input text

"Labas rytas"

**Speaker
embedding**

N-dim vector

**ACOUSTIC
MODEL**

**Mel-
spectrogram**

VOCODER

**Audio
waveform**

