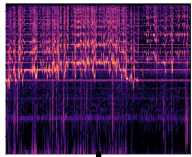


RAW AUDIO

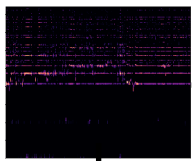


FEATURE EXTRACTION

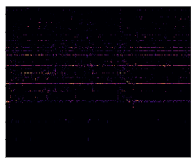
STFT



HARMONIC



SALIENCE



(1025, 22011, 3)

3D CONVOLUTION LAYERS

DENSE LAYERS



⋮



(337, 7332, 3, 64)



(109, 2441, 2, 32)



(35, 812, 1, 16)



(1, 100)



1

TARGET
TONIC