

Input: Spectrogram 300×400 (3sec \times 4000Hz)



Convolution 1: 16 filters of 16×12



Max-Pooling 1: 1×2



Convolution 2: 24 filters of 12×8



Max-Pooling 2: 1×2



Convolution 3: 32 filters of 7×5



Max-Pooling 3: 1×2



LSTM: bi-directional, 128×2



Dense Layer: length 64



Dropout: length 64



Soft-max: length 4



Output: 4 posterior probabilities