Sensor channels S = 6Frequency resolution Spectrograms F = 8T = 16Temporal resolution CNN-regression model Output tensor (x, y) FC (2), Linear c) Regression FC (E = 64), ReLu function FC (D = 196), ReLu Flatten 2 x 2 MaxPool, stride 2 BatchNormalization Conv 3 x 3 Conv (C = 60), ReLu 2 x 2 MaxPool, stride 2 b) Feature BatchNormalization extraction 3 x 3 Conv (B = 55), ReLu Conv., 2 x 2 MaxPool, stride 2

BatchNormalization
3 x 3 Conv (A = 50), ReLu

Input tensor (F x T x S)

Conv

a) Input tensor