Chorowski, attention based models for speech recognition

*Has good performance on sequence to sequence network of image captioning.*

Iteratively process their input by selecting relevant content at every step.

Performance degraded quickly with longer concatenated utterances.

**General Framework**

An attention based recurrent sequence generator is a recurrent neural network that stochastically generates an output sequence from an input x. In practice, x is often processed by an encoder which outputs a sequential input representation more suitable for the attention mechanism to work with.

LSTM and GRUs are typically used as a recurrent activation