Deep Laterally Recurrent Spiking Neural Networks for Speech Enhancement

Dr Julie Wall, j.wall@uel.ac.uk

1. Motivation
Compared to human performance ASR systems perform badly in noisy environments
- Traditional signal processing approaches to noise reduction only partially successful
- Lack of spatial information in the encoding of sound e.g. single channel GSM encoding
- Is there some way we can make use of the rich spectro-temporal information to separate speech from noise?

2. Biological Inspiration
What inspiration can we take from the biology for solving these problems?
- Cochlea is basically an FFT that digitises sound to spiking stimulus
- Known since early ’80s that cochlear nucleus is comprised of tonotopically organised lateral inhibitory neurons
- The neurons compete across the frequency bands

3. Spiking Neurons
Temporal code
- Individual neurons make decisions about firing based on the timing of the stimuli they receive from other neurons
- Excitation carries information
- Inhibition routes and synchronises stimuli

4. Lateral Recurrent Inhibition
Causes synchrony
- Substantial biological evidence for synchronous activity being crucial to sensory processing systems such as vision and audition
- Lateral inhibitory connectivity binds stimuli together using synchronous (and preferably near-synchronous states)

5. Laterally Recurrent Layer
Abbott’s idea can be extended…
- Symmetrical connectivity parameterised by two parameters:
  - Connection Length
  - Neighbourhood Radius

6. Speech Enhancement Pipeline

7. Conclusions
Some issues still remain to be resolved:
- Individual neurons make decisions about firing based on the timing of the stimuli they receive from other neurons
- Excitation carries information
- Inhibition routes and synchronises stimuli

Spiking Lateral inhibitory networks can be used to pre-process spectrograms for deep learning-based ASR systems

This work has been carried out in collaboration with Dr Cornelius Glackin, Intelligent Voice Ltd. neil.glackin@intelligentvoice.com