Speech Processing with Spiking Neural Networks

Background

Artificial neural networks are commonly assigned to three generations. Networks based on binary threshold activation functions are defined to be the first generation. Perceptrons, Hopfield nets and Boltzmann machines fall into this group of networks. The second generation distinguishes itself from the previous generation by the use of continuous output activation functions. Examples for this generation are feedforward and recurrent networks using sigmoidal activation functions. Spiking neural networks form the third generation of artificial neural networks by using neurons, which code information into temporal sequences of short energy pulses [1].

Although they have been shown to solve problems more efficiently than today’s heavily used networks, they still lack practical applications.

Related papers:

- “Which Model to Use for Cortical Spiking Neurons?” [2]
- “Unsupervised Learning of Visual Features through Spike Timing Dependent Plasticity” [3]
- “Evolving Spiking Neural Networks for Audiovisual Information Processing” [4]

Description

The aim of this thesis is to develop a speech processing chain with spiking neurons. This includes the input layer with raw audio input, processing layers as well as the output layer. The developed architecture should demonstrate that this can be realized without the need of any digital algorithm like FFT or MFCC.

Tasks

This student project consists of the following tasks:

- Familiarizing with spiking neural networks and signal processing
- Evaluating different architectures given in the literature
- Developing and implementing an architecture based on the research
- Testing the chain at spoken MNIST and wake word detection tasks
- Documenting the results

References