UNIVERSITY HEIDELBERG

Internship Report

Binaural Sound Localization in Spiking Neural Networks

Author:
Laura KRIENER

Supervisor:
Thomas PFEIL

July 30, 2014

Contents

1 Introduction

1.1 Motivation

In 1948 Lloyd Jeffress proposed a simple model for binaural sound localization using time differences between signals received at both ears. On neuromorphic hardware a spiking neural network inspired by this model could be used as a sensory system in robotics and a real-time interface for artificial cochleas.

The aim of this internship was to implement the network in PyNN [1], to simulate its behaviour using NEST [2] and to investigate the effects of different synaptic parameters. Parameters such as the synaptic weights were tuned to achieve a good performance of the network.

Since the network should be able to run on the Spikey chip, integrate-and-fire neurons with conductance-based synapses which are closest to the models in hardware were used. Regarding the limited neuron count of the Spikey chip the influence of the networks size on its performance was investigated.

1.2 Sound localization in biology

Auditory systems in mammalian brains use mostly two cues for localization of a sound signal: the *interaural time difference* (ITD) and *interaural level difference* (ILD). If the sound is a tone of only one frequency f with the amplitude a_0 and the source of the sound is positioned on either side of the head, the sound waves $a(t)$ and $a'(t)$ arriving at the ears can be described as

$$a(t) = a_0 \cdot \sin(2\pi f t) \qquad \text{(left ear)}$$
$$a'(t) = (a_0 - \Delta a_0) \cdot \sin(2\pi f(t + \Delta t)) \qquad \text{(right ear)}$$

where Δt describes the ITD and Δa_0 the ILD.

In the following, we focus on the processing of ITDs in auditory systems. The sound arriving at the ears as varying air pressure has to be transformed into spike coded information. This is done in the cochlea, where the inner hair cells trigger action potentials in the auditory nerve fibres. The resulting spike signals are phase-locked. This means that an action potential is created with the highest probability at the maximum amplitude $a(t)$ of the incoming signal. Additionally each fibre has a characteristic frequency at which it is most effective. Therefore, the temporal information of the incoming signal is encoded in the spike train itself, whereas the frequency information is spatially encoded by different auditory nerve fibres with characteristic frequencies [4].

These spike signals are transmitted to the medial superior olive, where the ITD-information of the signals from both ears are analysed. A model for the network used to analyse the ITDs was proposed by Jeffress [3] in 1948 and was confirmed by many following studies.

1.3 Jeffress Model

The network shown in Figure 1 consists of two lines of neurons (*ITD-detectors*) that receive inputs from both ears. Each neuron in a detector receives its input signals over delay lines (nerve fibres of different lengths). The lengths of these delay lines vary for the different neurons in the detector. By that the ITD-detector generates a spatial coding of ITDs by compensating the ITD with the delays added by the delay lines. This uses that a neuron will spike with a higher rate if it receives two input spiketrains coincidently than if it receives two input spiketrains that are shifted in time.

As an example we assume that the source of the sound is on the right side of the head. The signal from the right ear therefore arrives earlier at the point Y than the signal from the left ear at X. Because the pathway from Y to neuron 2 is longer than from X to this neurons, neuron 2 receives the signals from both ears approximately at the same time and spikes with a higher rate than the other neurons which receive the signals at two different times. If the source of the sound is on the left side if the head, one of the neurons with higher indices will spike with the highest rate.

2 Methods and results

Earlier studies [6] have shown that phase-locking as observed in the cochlea and auditory nerves can be implemented on neuromorphic hardware. The network used in this work models only the processing of ITDs in the medial superior olive. It is assumed that the network receives phase-locked spiketrains as input. For the used neuron parameters see Table 1 in the appendix.

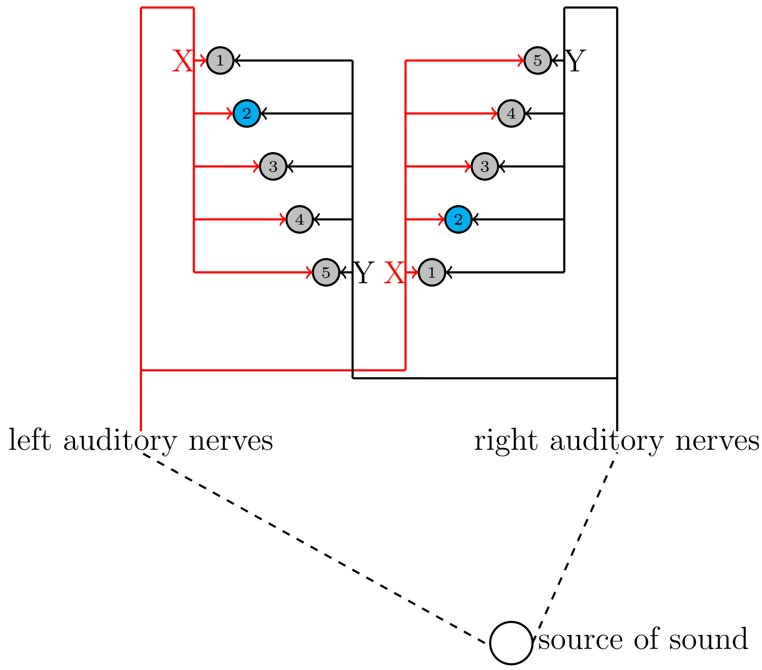


Figure 1: Jeffress-Model with sound on the right side of the head. Most active neurons are coloured blue. For neuron 2 the delay lines compensate the ITD

2.1 Network model

The networks used in this work are similar to the Jeffress-model, although some small modifications were made (see Figure 2):

The two ITD-detectors in the Jeffress model yield the same results independently. Therefore it is sufficiant to use only one of them and by that reduce the number of needed neurons. That is important, because the Spikey chip is limited to 384 neurons.

Also the way of producing the delays to compensate the ITD is different. Here, the delays are not produced by different length of axonal delays but by shifting the signals in time with different delay times d before feeding them to the neurons. Note that only the input signals of the right ear are changed. This does not change the functionality of the network, because delaying the left signal and moving the right signal forward in time are equivalent:

For an IDT of Δt the neuron with the delay $d = -\Delta t$ receives both input signals coherently and fires with the highest rate.

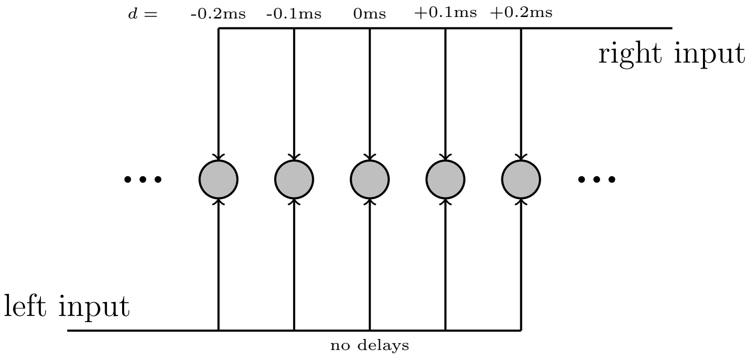


Figure 2: Example of an ITD-detector. In contrast to the Jeffres modell only the right input signals are delayed.

2.2 Effect of synaptic weights on firing rate

In [8] it is shown that the firing rate of a neuron that receives two periodic, phase-locked input spike-trains (cycle duration T) is highest if there is no time difference between the signals and lowest if the time difference is approximately $\frac{T}{2}$. This behaviour is crucial for the functionality of the network described above.

We investigated the effect of the synaptic weights that connect the inputs to the neurons in the ITD-detector on that behaviour. Since all neurons in the ITD-detector have the same parameters, we regarded only one neuron and the effect of the synaptic weight on its spiking behaviour (for different Δt). We found that only for a certain range of synaptic weights the neuron behaves like described in [8]. If the weight is too low, the neuron will not spike at all and the spike rate is not depending on Δt of the input signals. If the weight is higher than a certain threshold, the neuron will spike for every incoming spike and the spike rate is no longer depending on the coincidence of spike arrivals, too.

Within these borders the weight still effects the maximal and minimal firing rate of the neuron as well as the strength of the rates dependancy on Δt: Exemplarily we measured the firing rate r of a neuron that receives two periodic input spike-trains with a frequency of $f = 50\,\text{Hz}$. This rate r was measured for different time differences Δt between the two input signals (see Figure 3).

For a low weight of $w = 0.026\,\mu\text{S}$ the neuron fires only for small Δt and the rate is zero for larger Δt. For a higher weight of $w = 0.029\,\mu\text{S}$ it fires for all time

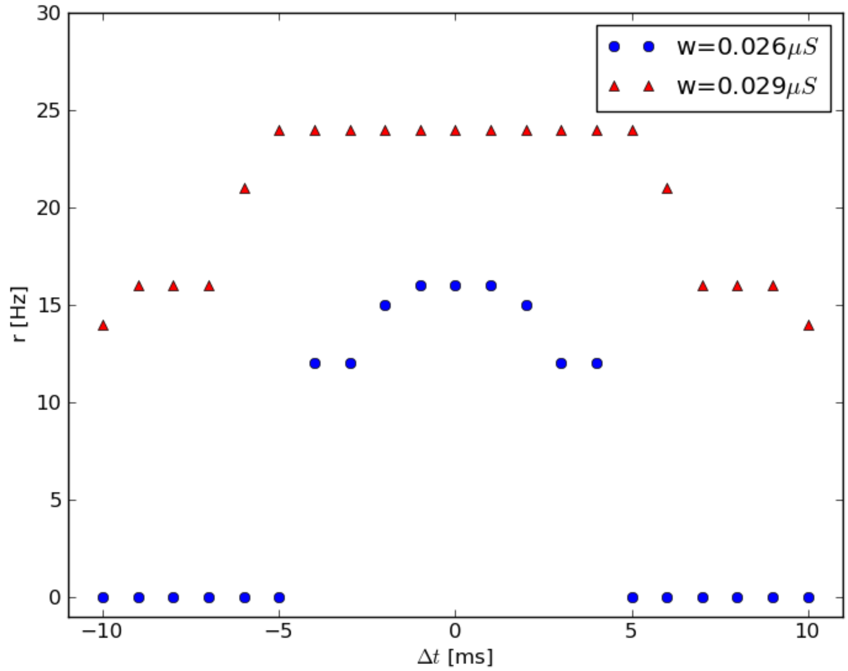
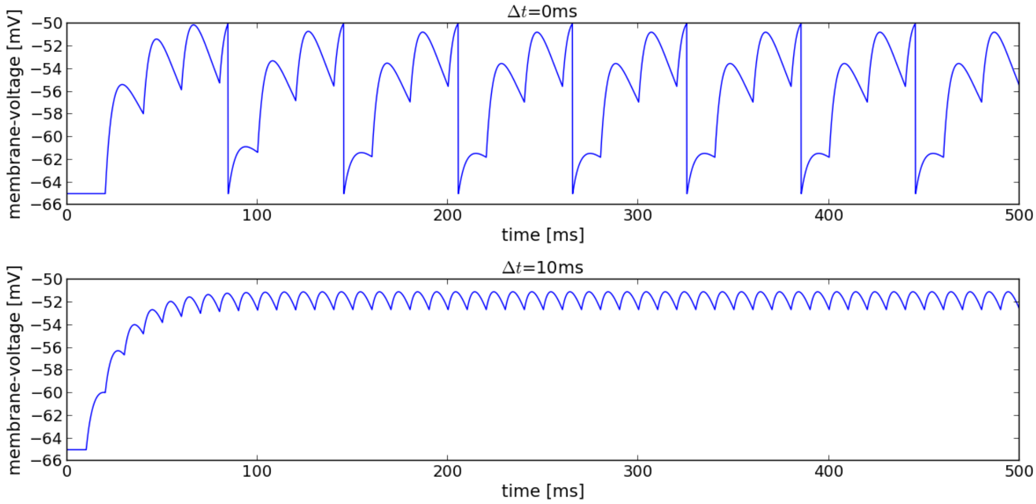


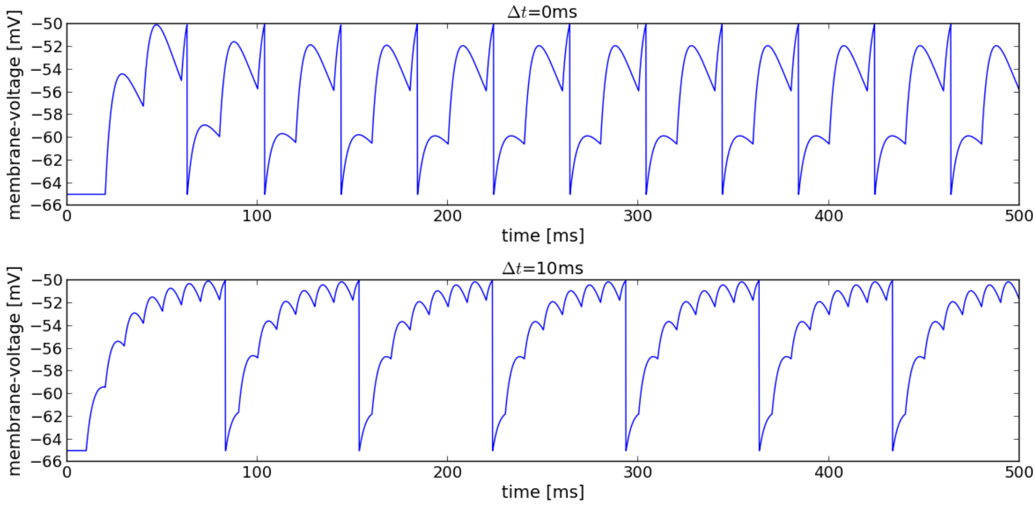
Figure 3: Firing rates r of a neuron with two periodic inputs over the time difference Δt between the inputs

differences but with different rates. Also the maximum rate is higher for the higher weight.

The reasons for that can be understood by regarding the membrane potential of the neuron over time (Figure 4) For the low weights and a maximal time difference of $\Delta t = \frac{T}{2}$ the neuron does not spike at all, because the post-synaptic potentials (PSPs) are small and the membrane potential stays below the firing threshold. For $\Delta t = 0$ two spikes arrive at the same time and their PSPs are integrated such that the membrane reaches the threshold and the neuron fires. After spiking the potential is reset and the integration of the PSPs starts again. Therefore the neuron fires periodically. For the higher weight this behaviour is observed for all values of Δt, because a higher weight causes bigger PSPs. For small Δt the threshold is reached after a shorter time than for large Δt. Therefore the firing rate grows when Δt increases.



(a) $w = 0.026\mu S$



(b) $w = 0.029\mu S$

Figure 4: Membrane potential for neuron with periodic input for $\Delta t = 0$ and $\Delta t = \frac{T}{2}$ for different weights w

For detecting the ITD of the input signals we need a distinct maximum of the neurons firing rate. Therefore, we choose a rather low synaptic weight, for which the neuron only spikes for small Δt.

2.3 Measuring the phase difference

As described above we could detect Δt by measuring the firing rate of one single neuron with two periodic inputs. But it is easier to use many neurons and add additional delays to the input signals (as described in section 2.1), because that way we only have to find the neuron with the maximal firing rate. The delay of that neuron is equal to $-\Delta t$.

After choosing the weights as in Figure 4(a) we set up an ITD-detector. Figure 5 shows an exemplary response of this ITD-detector to an input signal with $\Delta t = -3$ms. The used detector consisted of 41 neurons and the artificial delays added to the signal (d in Figure 2) ranged between -20 to 20ms. Each neuron in the detector is characterized by the delay d that is added to its right input and therefore each bin in the histogram corresponds to one of the neurons. One can clearly see that the neurons which receive an the input with a delay d close to 3ms, spike with the highest rate, because these delays compensate the given ITD of $\Delta t = -3$ms.

There is a strong threshold behaviour concerning which neurons spike (Figure 5: e.g. $d = -3$ms and $d = 8$ms). This does not seem realistic, since in hardware as well as in biology one has to expect noise on all network components, e.g. spike loss in synapses or in the cochlea. To model this, each spike in the spiketrain was deleted with a probability of 30%.

The same detector as before now produces the response in Figure 6. The comparison of Figure 5 and 6 indicates that the spike loss does not affect the location of the maxima, but it weakens the threshold behaviour by broadening the distribution and it reduces the spike rates.

The asymmetric shape of the maxima and their different height in Figure 6 are due to statistical effects, because the spike loss happens randomly. They could be removed by elongating the simulation time.

Up to now we ignored the fact that there is an other maximum around $d = -17$ms. For more neurons and a lager range of d there are maxima at $d = 23$ and $d = -37$ms as well. They occur because for input signals of $f = 50$Hz (cycle duration $T = 20$ms) ITDs of e.g. $\Delta t = -3$ms and $\Delta t = +17$ms result in the same two spiketrains. In general all $\Delta t'$ of the form

$$\Delta t' = \Delta t + n \cdot T \qquad\qquad n \in \mathbb{Z}$$

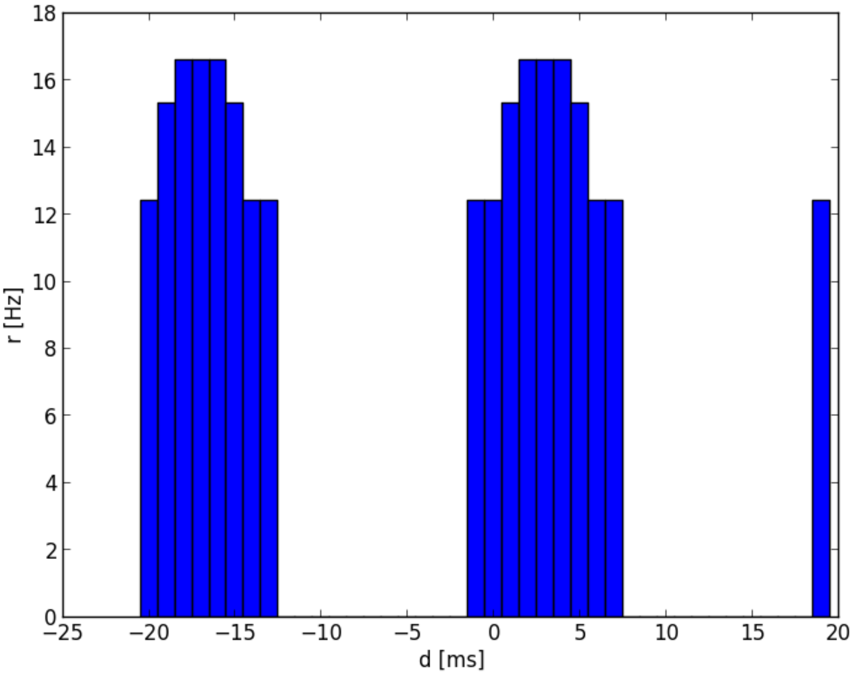


Figure 5: Firing rates r of ITD-detector consisting of 41 neurons with input-frequency 50Hz to ITD of $\Delta t = -3$ms

result in the same spike train as Δt. This shows that the network is only able to detect the phase difference between two signals, but is unable to distinguish between different ITDs causing the same phase difference.

2.4 Measuring the ITD

To distinguish ITDs with the same phase difference we use a network proposed in [5] and [7]: Using more than one detector with different input frequencies will solve the problem, because the same Δt between two signals will result in a different phase difference for each frequency:

For different frequencies the cycle durations T are different and therefore the value of d at which the "wrong" maxima occur, are likely to be different for all detectors. Only the d that corresponds to the ITD is equal for all detectors.

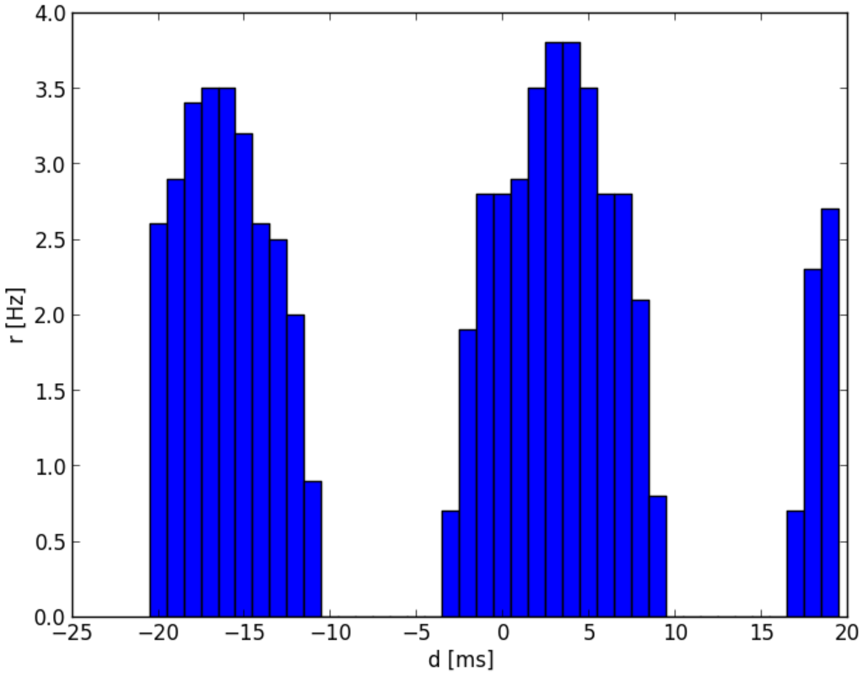


Figure 6: Firing rates r of ITD-detector consisting of 41 neurons with input-frequency 50Hz and 30% spike loss to ITD of $\Delta t = -3$ms

As an example we can look at Figure 7. It shows the response of a network consisting of 4 ITD-detectors with different frequencies to an ITD of $\Delta t = -3$ms. Each detector isolatedly shows ambiguous results by detecting only the phase difference between its inputs, but one maximum per detector is likely to overlay with maxima from all other detectors. This common maximum is around $d = 3$ms and corresponds to the delay that compensates the ITD.

If more then one detector is used, the weights to connect the neurons to their input have to be set very carefully. They should be calibrated in a way that each detector shows clearly distinct maxima and minima and the maxima of all detectors are of approximately the same height. If that is not the case, the result of one detector might overrule the other when they are added together to determine the ITD.

It was found that this is the case, when choosing the weights in a way that each neuron fires with a maximum rate of around two thirds of the lowest input

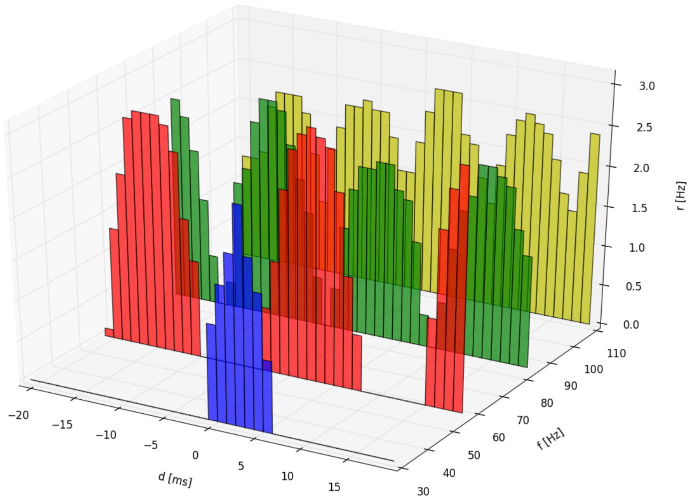


Figure 7: Firing rates r of multiple detectors to ITD of -3ms with 30% spike loss

frequency for no spike loss and around one fifth of the input frequency for a spike loss of 30%.

To determine the ITD from the different detector answers, one can sum the spikes of the neurons that belong to the same delay but different frequencies. The distribution of spike sums over the corresponding delays has one major maximum, which corresponds to the ITD and smaller maxima caused by the other maxima of some of the detectors. To increase the contrast of maximum to background, another population of neurons (*integration layer*) of the same size as one detector was used. All neurons in the detectors, which correspond to the same delay, excite one of the neurons in the integration layer. These neurons do not spike with the same rate as their input, but instead will modulate the result with their non-linear gain function shown in Figure 8. Low input rates are damped more than higher input rates, which causes the smaller maxima to become even smaller. Therefore the difference between the maximum and the background is increased. In Figure 9 the spikes of the integration layer (of the example in Figure 7) are shown on the left, on the right the summed rates of the neurons corresponding to the same delays are plotted. We see that the difference between maximum and background

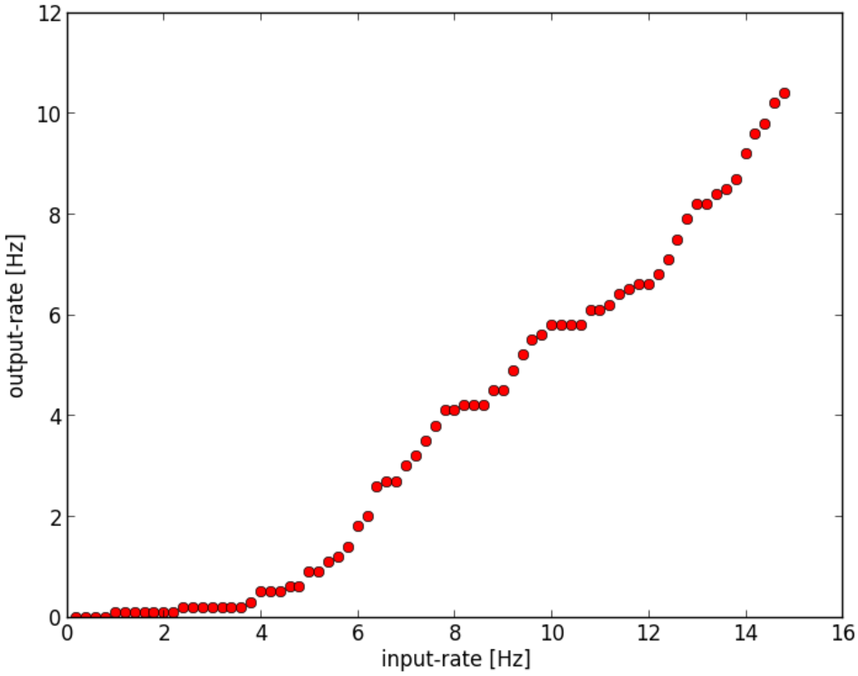


Figure 8: Non-linear gain function produced by a neuron in the integration layer fed with poisson-input of different rates (weight $w = 0.83\mu\text{S}$)

is increased by the integration layer. Additionally, the result could be improved even more by adding lateral inhibition resulting in a winner-take-all network to the integration layer as described in [7].

2.5 Network performance

The results of these networks depend on the choice of weights, number of neurons per detector and the used frequencies. To determine the optimal values for these parameters we defined a measure for the performance of the network.

In order to do so, we compared the ITD to the networks response to that ITD. To define a response value we regard the spike histograms of the neurons in the integration layer (like in Figure 9) and define the response as d_{max} which is the middle of the highest bin. If the maximum consists of more than one bin, the bins

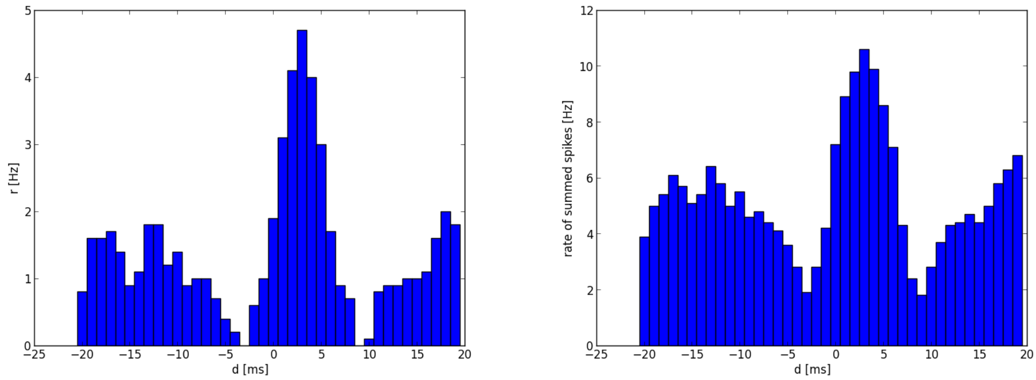


Figure 9: Firing rates r of neurons in the integration layer connected to the detectors with a weight of $w = 0.83\mu$S (left) and the summed spikes of the detectors (right). The integration layer increases the difference between main maximum and background.

of equal height are combined to one larger bin and d_{max} is the middle of that new bin.

The expected response is $d_{expect} = -\Delta t$, since $-\Delta t$ compensates the ITD. The distance δ between expectation and result is defined as

$$\delta = \left| d_{expect} - d_{max} \right| = \left| \Delta t + d_{max} \right| .$$

To derive the performance from this distance one has to distinguish two different cases: If the bin containing d_{expect} does not have the higest rate, the performance is better the closer d_{expect} is to d_{max}. But if the bin of maximal rate contains d_{expect}, it does not matter any more how close d_{expect} is to the middle of the bin (d_{max}), it only matters, that it is inside the correct bin. Instead it is important how wide the maximum bin is. Therefore the performance should correspond to the width of the maximum bin b (if the maximum is a combined bin, b is the sum over the width of those bins): If the correct value is in a maximum with a small width, the performance is better than if it is in a wide maximum. Because of that the performance value p is defined as

$$p = \begin{cases} \delta & \text{if } \delta \geq \frac{b}{2} \\ \frac{b}{2} & \text{if } \delta < \frac{b}{2} \end{cases}$$

The dependency of performance on the number of neurons and the number of detectors are shown in Figure 10. The performance was determined by testing the network with 10 random ITDs and averaging over the measured performances. The used frequencies were equally distributed over the interval between 20 and 120Hz (except for 2 detectors, we used 30Hz and 110Hz for). We see that the performance improves if more neurons are used, because the time resolution ($\propto b$) and therefore the minimal value of p ($p_{min} = \frac{b}{2}$) is smaller. But after a certain number of neurons (here around 50 neurons) it does not improve noticeably. We assume that this is because of the relatively large time constants, that prevent a higher time resolution.

One can also see that the performance is not improved if more than 3 frequencies are used. The reason is that the different frequencies are only needed to resolve the ambiguity, but if the ambiguity is already resolved by 3 frequencies, the others do not improve the result.

The performance is impaired, if the used frequencies are multiples of each other. As an example we used 6 detectors with the frequencies 20, 40, 60, 80, 100 and 120Hz. In Figure 11 we see that this leads to ambiguous responses of the network and therefore a bad performance.

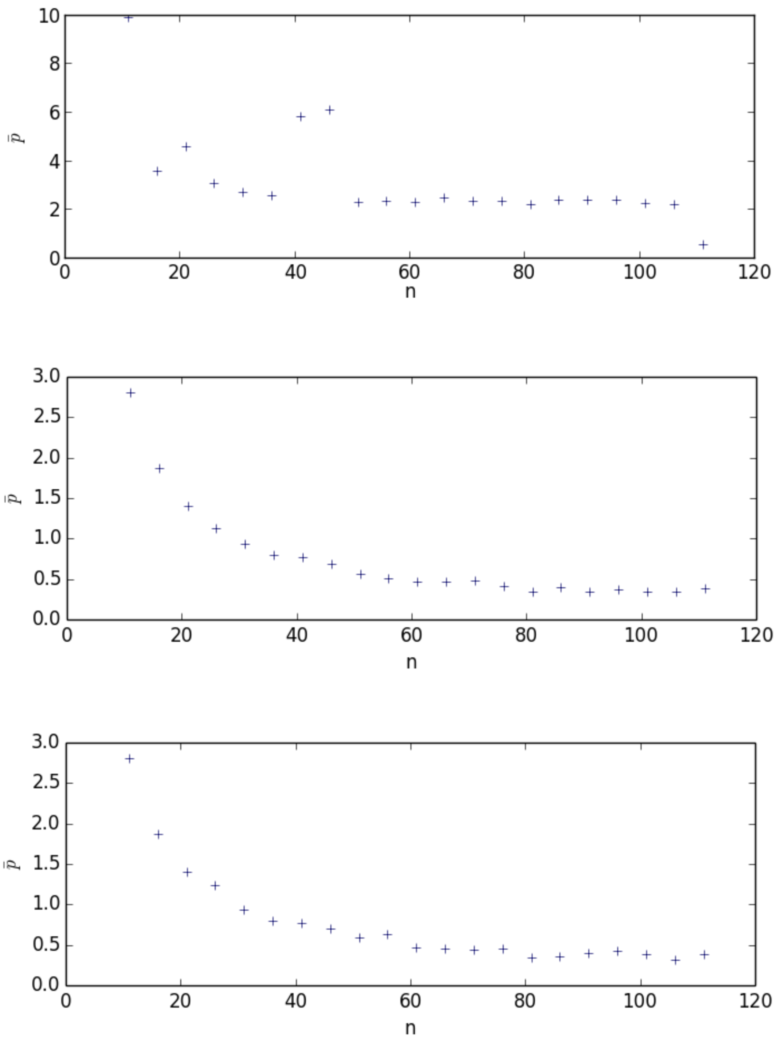


Figure 10: Mean performance $\bar{p}$ over number of neurons per detector n using 2, 3 and 4 detectors

2.6 Scalability of the parameters

The results presented before were obtained by using the standard neuron parameters in PyNN (see Table 1 in the appendix). By adjusting the neuron parameters as well as the synaptic paramters we can change the time scale on which our network works:

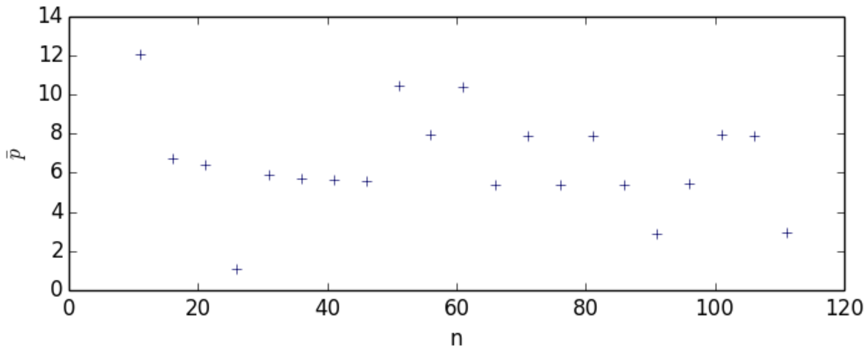


Figure 11: Mean performance $\bar{p}$ using 6 frequencies (multiples of each other)

Assuming that the distance between two ears is around 20cm, the maximal ITD is Ca $\Delta t = 0.5$ms. This shows, that the time resolution of the networks used before is around 10 times too large. To adjust our network we reduce the membrane time constant and the synaptic time constant by a factor of 10. If faster neurons are used, the input has to have a higher frequency and the weights connecting the input to the neurons have to be changed accordingly. After changing these parameters, the network works as before, but with the same number of neurons the range of detectable Δt is 10 times smaller.

Since the Spikey chip has 10^3 to 10^4 faster time constants than the simulation, the resolution of the hardware network will be up to 10^4 times smaller. This introduces a problem, because the number of required neurons for an appropriate ITD-range would be increased by that factor. On Spikey only 384 neurons are available.

3 Conclusion

This study shows that, if the neuron parameters, the input frequencies and the number of neurons are set appropriately, the detection of ITDs using networks similar to the Jeffress model works very well. The time constants are scalable and the network presumably would also work on the Spikey chip, but only a very small range of Δt would be detectable. Also we would need very high input frequencies.

This kind of network on neuromorphic hardware is therefore not suitable for real-time interfaces of sensors. The high input frequencies might be achieved by utilizing an ultra sound sensor. The small time resolution, which leads to huge neuron counts being required, is more difficult to evade. Adding an additional preprocessing software layer between sensor and network could solve this problem by scaling down the ITD produced by the sensor. This would lower the required number of neurons significantly and needs to be investigated further.

The bachelorthesis following this internship will concentrate on testing the network on the Spikey chip and investigating the performance as well as the effect of synaptic weights.

Appendix

Table of parameters used in simulations if not stated otherwise:

Parameter	Value	Description
V_{rest}	$-65.0\,\mathrm{V}$	Resting potential of membrane
C_m	$1.0\,\mathrm{nF}$	Capacity of membrane
τ_{mem}	$20.0\,\mathrm{ms}$	Membrane time constant
τ_{refrac}	$0.0\,\mathrm{ms}$	Duration of refractory period
τ_{syn_E}	$5.0\,\mathrm{ms}$	Decay time of excitatory synaptic conductance
τ_{syn_I}	$5.0\,\mathrm{ms}$	Decay time of inhibitory synaptic conductance
E_{rev_E}	$0.0\,\mathrm{mV}$	Reversal potential for excitatory input
E_{rev_I}	$-70.0\,\mathrm{mV}$	Reversal potential for inhibitoy input
V_{thresh}	$-50.0\,\mathrm{mV}$	Spike threshold
V_{reset}	$-65.0\,\mathrm{mV}$	Reset potential after spike
i_{offset}	$0.0\,\mathrm{nA}$	Offset current

Table 1: Parameters used in simulations

References

[1] A. P. Davison, D. Brüderle, J. Eppler, J. Kremkow, E. Muller, D. Pecevski, L. Perrinet, and P. Yger. PyNN: a common interface for neuronal network simulators. *Front. Neuroinform.*, 2(11), 2008.

[2] Marc-Oliver Gewaltig and Markus Diesmann. Nest (neural simulation tool). *Scholarpedia*, 2(4):1430, 2007.

[3] Lloyd A Jeffress. A place theory of sound localization. *Journal of comparative and physiological psychology*, 41(1):35, 1948.

[4] Jindong Liu, David Perez-Gonzalez, Adrian Rees, Harry Erwin, and Stefan Wermter. A biologically inspired spiking neural network model of the auditory midbrain for sound source localisation. *Neurocomputing*, 74(1):129–139, 2010.

[5] Jindong Liu, David Perez-Gonzalez, Adrian Rees, Harry Erwin, and Stefan Wermter. A biologically inspired spiking neural network model of the auditory midbrain for sound source localisation. *Neurocomputing*, 74(1):129–139, 2010.

[6] Thomas Pfeil, Anne-Christine Scherzer, Johannes Schemmel, and Karlheinz Meier. Neuromorphic learning towards nano second precision. In *Neural Networks (IJCNN), The 2013 International Joint Conference on*, pages 1–5. IEEE, 2013.

[7] C Schauer, Th Zahn, P Paschke, and H Gross. Binaural sound localization in an artificial neural network. In *Acoustics, Speech, and Signal Processing, 2000. ICASSP'00. Proceedings. 2000 IEEE International Conference on*, volume 2, pages II865–II868. IEEE, 2000.

[8] Anne-Christine Scherzer. Phase-locking on neuromorphic hardware. Bachelor thesis (German), University of Heidelberg, HD-KIP 10-43, 2013.