Proc. 3rd Conf. on Artificial Neural Networks (ICANN'93, Amsterdam), Gielen S. and Kappen B. (eds), Springer, London, p.165-168.

A Model for Latencies in the Visual System

Guido Bugmann, John G. Taylor
Centre for Neural Networks, King's College London, London WC2R 2LS, United Kingdom
Fax (+44) 71 873 2017, Phone: (+44) 71 873 2234, email: gbugmann@oak.cc.kcl.ac.uk

Abstract

The visual system is modeled as a multilayer network of coincidence detecting neurons switching to a state of sustained firing after the production of their first spike. Inhibition from the next layer stops the sustained firing only when the information has been used (coincidence detected). Simulations of such a self-timed network show the propagation of a firing probability wave from layer to layer. The latency of the activity onset in each layer is determined by the level of the sustained activity and by the initial jitter of onset times in the first layer. Theoretical analyses confirm the observed relations. Our results indicate that the layer-to-layer latencies in the visual system may be mainly caused by the jitter in onset time in retinal ganglion cells.

1 Introduction

There are conflicting data on latencies in the visual system. Measurements based on electrical stimulations consistently indicate latencies of less than 2 ms per neuronal relay [1]. However, when visual stimulation is used, an average latency of 10 ms per neuronal relay is a generaly accepted value [2]. It is observed, for instance, between layer 4 and layers 2-3 in the striate cortex [3]. What is the cause for these longer latencies with visual input ? It is assumed that electrical stimulations produce highly synchronized input spikes [4] which cause unnaturally short latencies. With visual stimulation, input spikes arrive asynchronously and it has been proposed that they must be integrated for appropriate intervals [3]. However, temporal integration is incompatible with the observed irregularity of the spike trains [5]. On the contrary, coincidence detection based on very short integration times, or possibly a synaptic saturation mechanism [6], seems required. Coincidence detection corresponds to a multiplicative, AND-type, function of the neuron which has been postulated for neurons in MST [7]. If we assume neurons operating as coincidence detectors, and thereby discard the temporal integration hypothesis, how can the long latencies be explained ?

In the present paper we describe a simple model of information processing in the visual system, based on coincidence detection, which can reconcile the above mentioned observations, producing long delays with visual stimulation and short ones with electrical stimulations. We make the following assumptions:

1) Neurons in the visual cortex are part of a pyramidal multilayer network where each neuron receives spikes from m distinct neurons in the preceding layer.

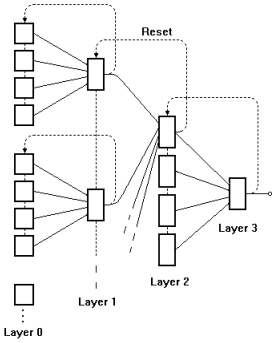
2) Neurons act as coincidence detectors, firing only if all *m* inputs provide a spike within a given time window.

3) Neurons receive excitatory feedback from local-circuit neurons and stay in a state of persistent activity ("ON" state) after producing their first spike, as sugested by physiological observations [4].

4) Neurons in the self-sustained ON-state are silenced by inhibitory feedback from their target neurons in the next layer. This feature ensures that neurons fire only during the minimum necessary time and prevents the loss of information from one layer to the next. Similar feedback inhibition schemes are used in models of speech production [8], olfactory recognition [9] and visual search [10].
Inhibitory feedback is usually assumed to be a local process [4], with a delay determined by the local circuitry. In our model, this would cause frequent failures in propagating the information, in contradiction to the systematic response of neurons to optimal visual stimuli [11]. However, it is not excluded for inhibitory feedback to have a dual role: i) controlling the local level of activity, in which case inhibition would be most apparent in pathological cases like electrical stimulation, ii) disabling a selection of neurons for computational reasons, in which case it would be determined by long-range projections.

2. Simulations

The simulations are performed with a pyramid of pRAM neurons [12]. These neurons can act as coincidence detectors, like leaky integrate-and-fire neurons [13], and are easier to analyse mathematicaly. The pRAM operates in discrete time-steps Δ t with spike trains defined as sequences of 1's (a spike) and 0's. Each pRAM has m=4 inputs and is set for the coincidence detection mode, firing only when all 4 input spikes are present during a time-stepΔ t. As soon as the first spike is produced, the pRAM is set to fire at each subsequent time step with a probability P_1. This produces a random spike train of frequency f_1=P_1/Δ t characteristic of the ON-state and simulates the effect of local excitatory feedback. Such a random firing is consistent with findings that, at all frequencies, biological neurons fire with near Poisson distributions of interspike intervals [5].
The pyramidal network has 64 neurons in its input layer (Layer 0), divided in 16 groups of 4 neurons connected to 16 neurons in layer 1. These neurons are divided into 4 groups of 4 neurons connected to 4 neurons in layer 2. These 4 neurons are connected to a single neuron in layer 3.



The input neurons are initially set to fire randomly with a frequency $f_0=P_0/\Delta t$. However, as soon as their first spike is produced, their frequency is set to f_1. So, the only role of f_0 is to produce a jitter in the starting time of the neurons in the first layer.
For a neuron in a layer *n+1* to fire there are 2 conditions: 1) all *m* neurons in layer *n* must be in the "ON" state and 2) a coincidence has to occur (coincidence probability $P_c=P_1^m$). In biological neurons, coincidences are defined as spikes arriving in a given time window τ and occur with a probability $P'_c=f_s^m\ \tau^m$ [13] where f_s is the biological sustained rate. If we assume the coincidence probabilities to be equal in both systems, this gives us the relation between the time-steps and frequencies used in simulations and those in biological systems to be $f_1\ \Delta t = f_s\ \tau$.
When a neuron in layer *n+1* has fired, the firing probability of all its *m* input neurons is set to zero. The figures 1 and 2 show the observed firing probabilities at each time-step in two cases, respectively without and with feedback inhibition. In the layers 0, 1 and 2, the feedback seems to reduce the firing probability. Actually, neurons are still firing at f_1 when they are in the "ON" state but they do so at different times in different runs. This is exemplified in figure 3 showing a spike raster of 30 runs in the case with feedback. There are no changes in the last layer which receives no feedback. After the onset time, the firing probability saturates at P_1.

We define the total latencies L_n as the time elapsed from the start of the simulation until a neuron in layer n as reached an average firing probability $P_{1/2}$. We define the relative latencies as $\Delta L_n = L_n - L_{n-1}$.
Figure 4 shows the latencies L_0 and ΔL_n for n=1, 2, 3 in the case of a fixed P_0=0.05 for various P_1. The relative latencies converge all to the same minimum value ΔL_{min} for large values of P_1. The theoretical analysis of the system (to be described in details in a later paper) predicts a $1/P_1^m$-dependence for ΔL_n. This is confirmed by simulations (the curve in figure 4 is: $\Delta L_1 = \Delta L_{min} - 0.8 + 0.8/P_1^4$). Figure 5 shows the minimum latencies ΔL_{min} (determined at P_1=1) for various values of P_0. The curve in the figure is the theoretical prediction $\Delta L_{min} = log(1/m)/log(1-P_0)$.

3. Discussion

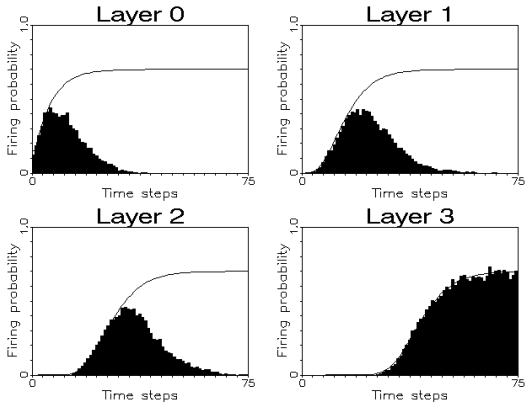
In our model we have not included transmission delays, synaptic delays or background potential effects [14]. Therefore the model provides a lower bound to latencies and is purely related to the performed computation. It shows two components to the computational latencies: 1) The initial jitter or time necessary for all m input neurons to be in the ON state. This depends on P_0 in the model. 2) The time for a coincidence to occur which depends on P_1.
The initial jitter, corresponding physiologically to the fluctuations in onset times in retinal ganglion cells [15], determines the absolute minimum in per-layer-computation time ΔL_{min}. The gain in per-layer computation time due to the level of sustained activity P_1 becomes smaller and smaller as P_1 is increased. With a value P_1=0.6 - 0.8 most of the gain is already made (Fig. 4). Assuming a realistic biological time-window for coincidences of the order of 2ms, sustained frequencies of 300 - 400 Hz (as observed in biological neurons) would be sufficient for the computation time to approach its smallest possible value. As ΔL_{min} increases with the number of inputs, the spread of latencies of neurons in a same layer [3] could reflect differences in fan-in.
Preliminary investigations show that this model may also explain the psychometric curves obtained in visual masking experiments.

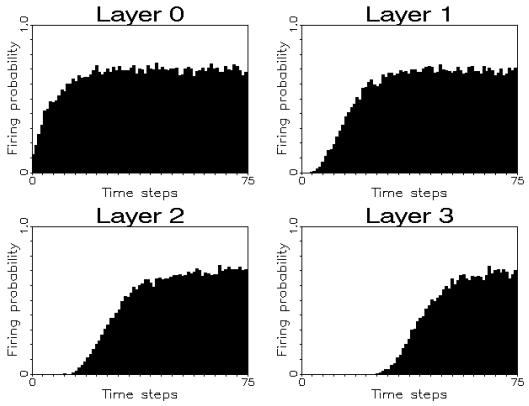
4. Conclusion

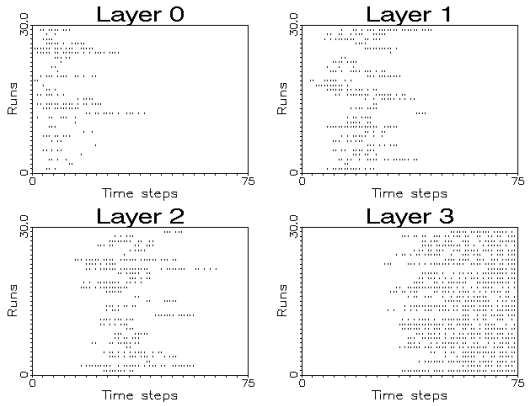
If the hypothesis of this model are valid, it indicates that the visual system uses maximum firing rates probably near to the optimum with regards to computation speed. The main limiting factor would therefore be the noise at the level of the sensory receptors which determines the latencies in all subsequent processing layers. Noise refers here to jitter in onset times, not to fluctuations in firing levels. A prediction of this model is that latencies in the visual system can be manipulated by modifying the onset-time jitter in retinal ganglion cells.

Aknowledgements

This work has benefitted from discussions with M.D. Plumbley and from the SERC grant GR/H22495.







Figures 1,2,3. Top Left (1): temporal evolution of the firing probability in the case without reset by the inhibitory feedback. Middle left (2): same with reset. Bottom left (3): spike raster of 30 runs in the case with reset. Each dot indicates the time of occurence of a spike. Simulation parameters: P_0 = 0.15, P_1 = 0.7.

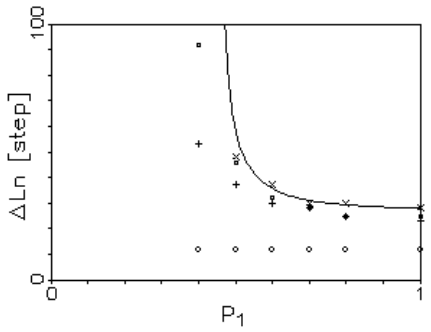


Figure 4. Relative latencies ΔL_n in dependence of the sustained firing probability P_1. Symbols: circle: L_0, cross: ΔL_1, square : ΔL_2, x: ΔL_3.

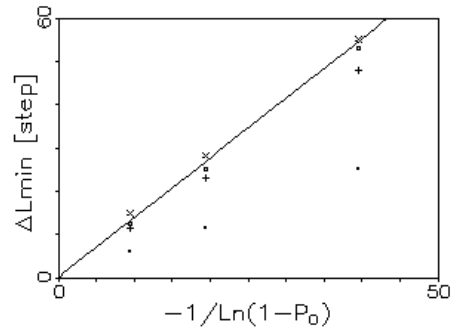


Figure 5. Minimum relative latencies in dependence of P_0. (The jitter in onset time in layer 0 decreases as P_0 increases). Symbols as in Fig. 4 except: dots: L_0.

References

1. Ferster D. and Lindstrom S. (1983) "An intracellular analysis of geniculo-cortical connectivity in area 17 of the cat", em J. Physiol., 342, 181-215.

2. Thorpe S.J. and Imbert M. (1989) "Biological constraints on connectionist modelling", in Pfeifer R. et al. (eds.) "Connectionnism in perspective", Elsevier, 63-92.

3. Maunsell J.H.R. and Gibson J. (1992) "Visual response latencies in striate cortex of the macaque monkey", *J. Neurophysiol.*, 68, 1332-1344.

4. Douglas R.J. and Martin K.A.C. (1991) "A functional microcircuit for cat visual cortex", *J. Physiol.*, 440, 735-769.

5. Softky W.R. and Koch C. (1993) "The highly irregular firing of cortical-cells is inconsistent with temporal integration of random EPSP's", *J. Neurosci.*, 13, 334-350.

6. Bugmann G. (1992) "Multiplying with neurons: Compensation of irregular input spike trains by using time-dependent synaptic efficiencies", *Biol. Cybern.*, 68, 87-92.

7. Verri A., Straforini M. and Torre V. (1992)"Computational aspects motion perception in natural and artificial neural networks", in Orban G.A. and Nagel H.H. (eds.) "Artficial and biological vision systems", Springer, 71-92.

8. Houghton G. (1990) "The problem of serial order: A neural network model for sequence learning and recall", in Dale R., Mellish C. and Zock M. (eds.) "Current research in natural language generation", Academic Press, London, 287-319.

9. Granger R., Ambros-Ingerson J., Staubli U. and Lynch G. (1990) "Memorial operation of multiple, interacting simulated brain structures", in Gluck M.A. and Rumelhart D. (eds) *Neuroscience and Connectionist Theory*, Lawrence Erlenbaum Associates, London, 95-129.

10. Humphrey G.W. and Muller H.J. (1993) "SEarch via Recursive Rejection (SERR): A connectionist model of visual search",*Cognitive Psychology*, 25, 43-110.

11. Newsome W.T., Britten K.H. and Movshon J.A. (1989)"Neuronal correlates of a perceptual decision", *Nature*, 341, 52-54.

12. Gorse D. and Taylor J.G. (1990) "A general model of stochastic neural processing", *Biol. Cybern.*, 63, 299-306.

13. Bugmann G. (1991) "Summation and multiplication: Two distinct operation domains of leaky integrate-and-fire neurons", *Network*, 2, 489-509.

14. Bugmann G. (1992) "The neuronal computation time" in Aleksander I. and Taylor J.G. (eds) "Artificial neural networks II", Elsevier, 861-864.

15. Levick W.R. (1973) "Variation in the response latency of cat retinal ganglion cells", *Vision Res.*, 13, 873-853.

Note: This is a reformated on-line version using larger fonts. The network diagram in section 2 was added for clarity. It is does not figure in the originally published paper.

Guido Bugmann is now at The Centre for Neural and Adaptive Systems, School of Computing, University of Plymouth, Plymouth PL1 5LT, UK.
email: gbugmann@soc.plym.ac.uk,
Home page: http://www.tech.plym.ac.uk/soc/staff/guidbugm/bugmann.html