

# TOWARDS A NEURAL MODEL OF TIMING

Guido BUGMANN

School of Computing, University of Plymouth, Plymouth PL4 8AA, UK

<http://www.tech.plym.ac.uk/soc/staff/GuidBugm/Bugmann.html>

[gbugmann@soc.plym.ac.uk](mailto:gbugmann@soc.plym.ac.uk)

## **Abstract:**

This paper describes a neural model of interval timing, which reproduces the duration discrimination experiments of Wearden (1992).

The model comprises 3 layers of neural units. The units in the first layer represent clusters of neurons with probabilistic internal feedback that maintains self-sustained (short-term memory) activity for a random time. The unit in the second layer is a spiking neuron that fires as long as a sufficient number of input clusters are active. The unit in the third layer detects the offset of firing in the previous layer by producing a short burst of spikes.

Analysis and simulation of the model shows spikes produced at random times with a distribution determined by the number of units in the first layer, their survival time constant, and the threshold of the unit in layer 2. Interval times can be learned with any of these parameters but lead to different Weber law relations. A variable threshold in layer 2 predicts S-shaped Weber curves, a variable number of units in layer 1 leads to a saturation of the Weber curve (decreasing Weber fraction) and a variable time constant in layer 1 causes a linear Weber curve.

## **1. Introduction.**

Timing is an important function of biological systems. There are two forms of timing: phase timing and interval timing. Phase timing is the ability to time-mark events within seasonal or circadian cycles and is not discussed in this paper. Interval timing is the ability to predict when an expected event is to occur after some event or action. Car drivers can predict when a given red light becomes green. Interval timing also allows to detect unexpected events or the non-occurrence of expected events. Parachute users know exactly how much time it takes for a parachute to open.

Animal interval timing capabilities are tested in classical conditioning or reinforcement experiments using intervals of up to 10 minutes. For instance, pigeons can be trained to jump on a perch, stay there for a given time and then to jump off (Jasselette, Lejeune and Wearden, 1990). In

humans, temporal duration discrimination tasks (Wearden, 1992) or temporal order recognition tasks are used (Milner, Corsi & Leonard, 1991). Intervals are generally smaller than a second to prevent the use of counting strategies.

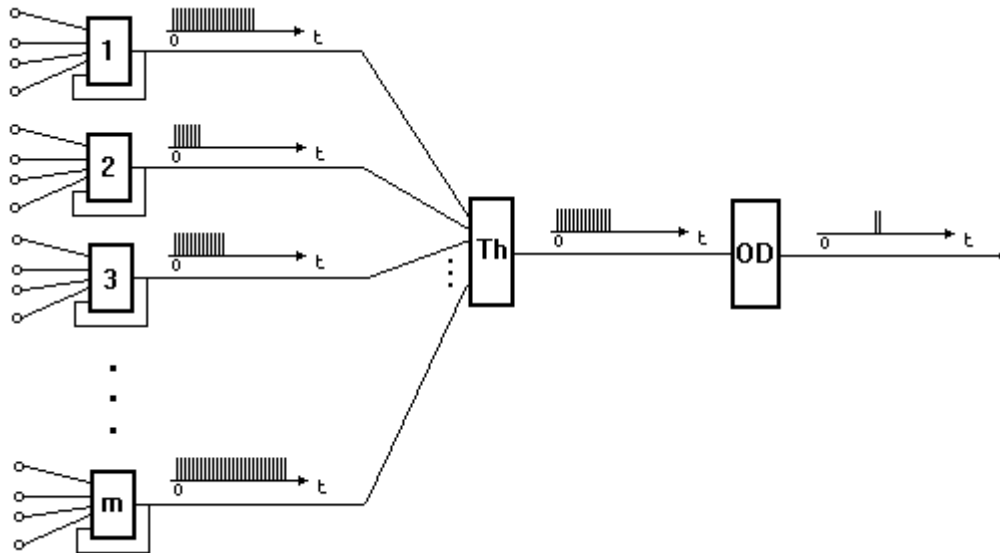
The biological substrate of interval timing is still not well understood. A large number of experimental data are reviewed in Gibbon, Malapani, Dale and Gallistel (1997) and these point to the involvement of almost every brain area including hippocampus, cerebellum, basal ganglia, cortex and frontal cortex.

An important characteristic of interval timing is its stochastic nature. For instance, pigeons stay on the perch *on average* during the reinforced time. However, actual perching times in individual trials can be shorter or longer. The distribution of perching times is usually centred on the reinforced time interval with a width following a Weber law, i.e. the width is a linear function of the average time.

Current neural models of timing either model average responses of subjects, not the stochastic nature of the biological timing (Grossberg and Merrill, 1992, 1996) or contain a noise term without clear biological basis (Miall, 1996; Church and Broadbent, 1990).

In section 2 of this paper a low level model of a timing mechanism is proposed which is compatible with biological hardware and has an intrinsic stochasticity caused by probabilistic transmission at synapses. In section 3, the relation between Weber law properties of the model and the parameters that are adjusted during time interval learning is analysed. A linear relation is only one of the possibilities. In section 4 an anatomically distributed timing network is proposed and used for the reproduction of experimental data from temporal duration discrimination tasks (Wearden, 1992). In section 5, properties of the model and future work are discussed.

## 2. Neural Network Model

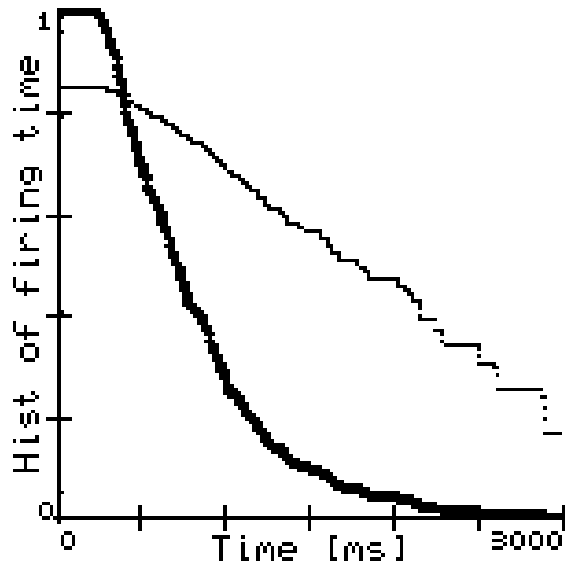


**Figure 1.** 3-layer timing neural network. The  $m$  units in the first layer are clusters of neurons with probabilistic self-sustained activity. The unit in the second layer is a threshold spiking neuron that fires as long as the number of active input clusters exceeds a threshold  $n$ . The unit in the third layer (Offset Detector) detects the offset of the firing in the previous layer and responds to it by producing a short burst of spikes (a neural circuit with this function was described in Bugmann (1997)).

### 2.1 Modelling self-sustained clusters in layer 1.

Persistent firing activity is a form of temporary information storage used in many parts of the brain. The pre-frontal cortex is a well known example (Fuster, 1980) but many other parts of the cortex have the capability of storing information in that way. Cortical neurons are imbedded in an excitatory network that can produce prolonged firing if inhibition is removed (Douglas and Martin, 1991) or, in the worst case, epileptic seizure (Tasker, Peacock and Dudek, 1992).

The behaviour of clusters of neurons with probabilistic excitatory connections between neurons has been simulated in (Bugmann, 1997). As some input spikes fail to produce EPSPs, there is always a moment when a catastrophic series of transmission failures prevents the self-sustained activity from continuing. As shown in Figure 2, the probability of finding the cluster in an active state after all inputs are silenced, is a roughly exponentially decaying function. This is similar to the one obtained by a single neuron with probabilistic self-feedback (curve a, figure 3).



**Figure 2.** *Thick line:* Probability of finding a 100 neurons cluster in an active state. At time 100 ms the input is removed. The plateau lasting until  $t = 300$  ms reflects a minimum time taken for the activity to decay to 5% of active neurons (the threshold used for defining the active self-sustained mode). *Fine line:* The natural logarithm of the thick line (times 0.125). The linear portion corresponds to a decay time constant of 4 seconds. *Parameters:* The same network is used as in figure 5 of Bugmann (1997). Neurons within the cluster are connected to 10% of the other neurons. The synaptic transmission probability is 0.5. The EPSCs are short alpha functions with a maximum at 1 ms of amplitude 1.775 (a.u.). The curves shown depend critically on the seed used for randomly interconnecting neurons.

The decay time constant of the sustained firing depends on many parameters, such as the connectivity, size and duration of EPSPs, and synaptic transmission probability. There has been no systematic investigation of the influence of these parameters. There are initial indications that long decay times (over 1 second) are very sensitive to parameter values and may not be a stable property of such networks.

In the remaining parts of the paper a cluster will be modelled as a single unit with a self-feedback connection having a probability  $\alpha$  of transmitting the spike and causing another output spike. The probability  $P(t)$  of finding that unit in an active state after  $k$  feedback loops (of duration  $\Delta t$  each) is given by:

$$P(t = k\Delta t) = \alpha^k = e^{-t/\tau} \quad (1)$$

$$\tau = \frac{\Delta t}{\ln(1/\alpha)}$$

The probability of finding the self-feedback unit in a active state decreases exponentially with time after the initiation of the activity (curve a, figure 3). In the next sections we will see how units with such properties can be used to design a timer. The question of how realistic they are will be reconsidered later, once it has been shown that the net can usefully emulate biological timers.

## 2.2 Integrating self-sustained clusters in layer 2.

By feeding  $m$  self-feedback units into a threshold unit **Th** (figure 1), a short-term memory can be built that stays active as long as the number of active inputs exceeds the threshold  $n$ . (curve b, figure 3). The probability  $M(m,n,t)$  of finding the threshold unit in an active state is given by Bugmann and Taylor (1997):

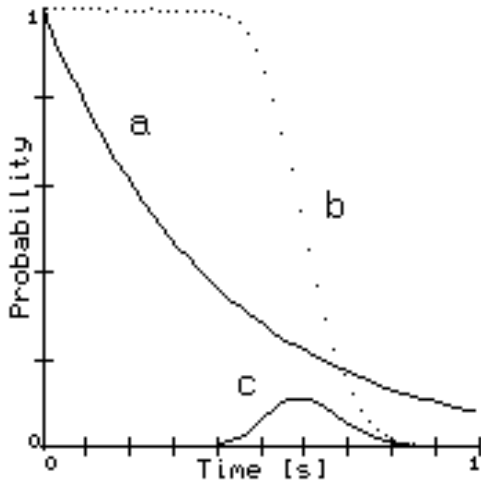
$$M(m,n,t) = \sum_{i=n}^m P(i,t) \quad (2)$$

where  $P(i,t)$  is the probability that the number of active units has decayed to exactly  $i$  at time  $t$ :

$$P(i,t) = C_i^m e^{-\frac{i \cdot t}{\tau}} (1 - e^{-\frac{t}{\tau}})^{m-i} \quad (3)$$

## 2.3 Detecting the offset of the threshold unit in layer 3.

By detecting the end of the short-term memory activity with a unit that produces a burst of spikes (Bugmann, 1997), a stochastic timer can be built that has a distribution of firing times given by  $P(n,t)$  (curve c, figure 3).



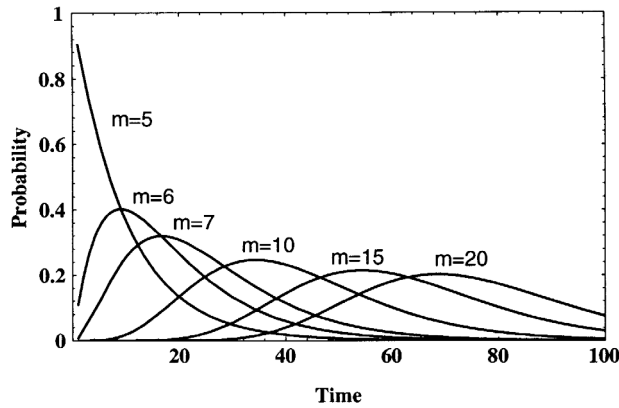
**Figure 3.** Examples of firing probabilities in the three layers of the network in figure 1. *Curve a:* Probability of finding a probabilistic self-feedback unit in layer 1 in an active state. *Curve b:* Probability of finding the unit in layer 2 in an active state. *Curve c:* Probability of the unit in layer 3 producing a spike.

The most probable firing time of the timer is  $t_{\max}$  given by the maximum of  $P(n,t)$ :

$$t_{\max} = \tau \ln\left(\frac{m}{n}\right) \quad (4)$$

The most probable firing time is determined by three parameters: The number of self-feedback units  $m$ , the threshold  $n$  of the integrating unit **T** and the decay time constant  $\tau$  of the self-feedback units. The value of  $\ln(m/n)$  can be much smaller or much larger than 1, hence the most probable firing time can be smaller or larger than the decay time of the self-feedback units, as illustrated in figure 4. This makes the model very flexible.

The width of the distribution of firing times is determined by the same three parameters  $\tau$ ,  $m$  and  $n$ . As a rule of thumb, the larger the ratio  $m/n$ , the narrower the distribution and the more precise the timer as illustrated in Bugmann and Taylor (1997).



**Figure 4.** Distribution of firing times of a timer with a fixed threshold  $n=5$  in layer 2 and a variable number  $m$  of units in layer 1. The decay time of units in layer 1 is  $\tau = 50$ . We may note the similitude between these curves and the spectral functions postulated in Grossberg and Merrill (1992). Figure reprinted with permission from Bugmann and Taylor (1997).

### 3. Weber fraction

#### 3.1 Weber law

In most timing experiments, the width  $\sigma$  of the distribution of timed response is a linear function of the average time  $t_{av}$  as given by following equation known as the generalised Weber law (Jasselette, Lejeune and Wearden, 1990):

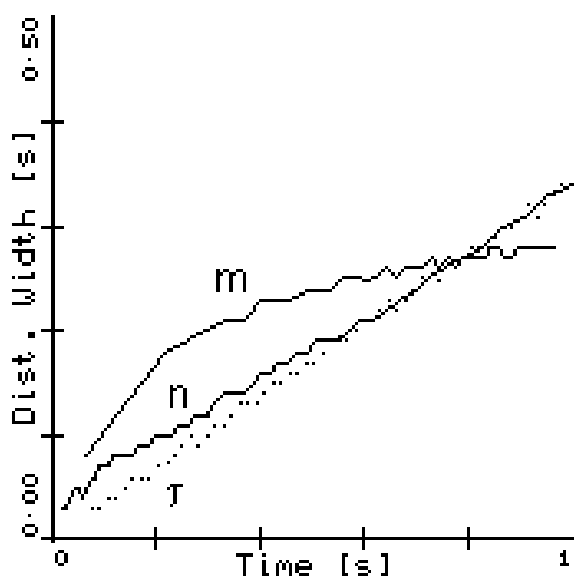
$$\sigma = c \cdot t_{av} + d \quad (5)$$

If the distribution is Gaussian, the width  $\sigma$  is the standard deviation and  $t_{av} = t_{max}$ . In our model, the distribution is not Gaussian (as illustrated in figure 4) although it may become close to a Gaussian for large values of  $m$ . In this section, the value of  $\sigma$  is approximated by the width at half height of the distribution (3).

#### 3.2 Weber fraction and time interval learning

If a subject used our network model to learn time intervals, he would have the option to modify either  $m$ ,  $n$ , or  $\tau$  to adjust its internal timer. Modifying  $m$  means connecting or disconnecting new units into the threshold neuron, or using all-or-none modifiable weights. Modifying  $n$  means changing uniformly the weights from an unchanged number of units. Modifying  $\tau$  means modifying some internal property in a cluster, represented by the transmission probability in the feedback connection of the model unit.

Numerical estimations using (3) and coarse theoretical analysis show that the relation between the width of the distribution and the average time depends on the variable parameter that is used to set the average time. Figure 5 shows a linear Weber relation when  $\tau$  is learnt, a slightly S-shaped curve when  $n$  is learnt and a saturating curve when  $m$  is the variable. Usually, straight lines are fitted to experimental data. It may be worth looking at these more closely, as they may give a hint about the parameter that is being modified during learning. However, more precise theoretical analysis is required to produce an analytical expression to adjust to experimental curves.



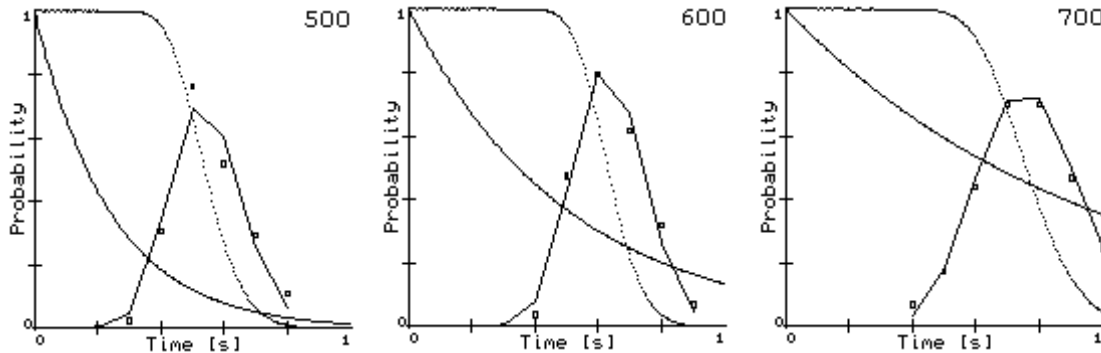
**Figure 5.** Width of the distribution of firing times of the timer against the average firing time. Both values were calculated using (3) for  $m$  variable,  $n$  variable or  $\tau$  variable. The two fixed parameters were  $m = 70$ ,  $n = 15$ ,  $\tau = 0.5$  sec. The average firing time has been obtained from numerical integration of equation (3).

#### 4. Modelling Interval Discrimination Experiments

The model has been used to reproduce the *stochastic* behaviour of subjects in a number of experiments (Bugmann and Bapi, submitted): 1) Using only the two first layers, the decay of the working memory retention performance observed in monkeys by Fuster et al. (1981) could be reproduced with a threshold unit receiving inputs from 3 units in layer 1. These units had a decay time constant of approximately 76 seconds. 2) Using only the first layer, the recency discrimination data of Milner et al. (1991) could be reproduced using a single-unit with decay time of 60 seconds as short-term memory.

In this section it is shown how the full 3-layer model can reproduce the duration discrimination task data of Wearden (1992). This will require over 50 input units with decay times of the order of 500 ms. This time constant is much smaller than those mentioned above, which are probably characteristic of self-sustained activity in the prefrontal cortex.

In the form of duration discrimination task which is modeled here, subjects are presented with a tone of standard duration of either 500 ms, 600 ms or 700 ms (Wearden, 1992). Subjects then listen to test tones of various durations and determine if their duration is equal to the standard. The fractions of "yes" responses for each standard duration are shown in figure 6. It can be seen that subjects sometimes believe that a tone lasting 800 ms is lasting 500 ms. How can this be explained ?



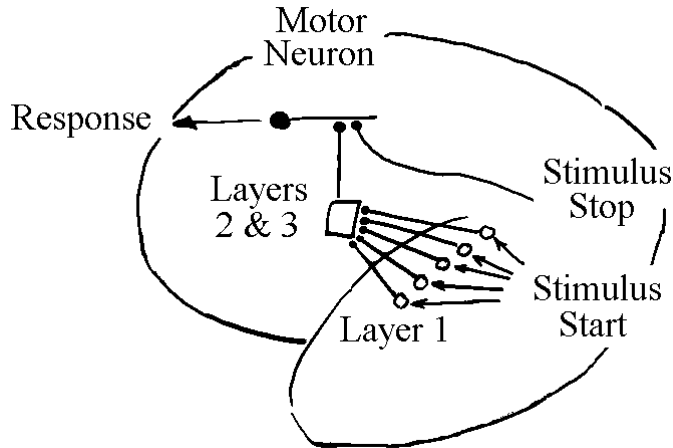
**Figure 6.** *Square symbols:* Probability that a subject judges a tone of the indicated duration as equal to the standard of 500 ms (left figure), 600 ms (middle figure) and 700 ms (right figure). *Bell shaped full line:* Best fit by our model. *Exponentially decaying full line:* Behaviour of an input unit with the best fit parameters. *Dotted line:* Behaviour of the threshold unit with the best fit parameters (given in table 2).

A standard model used by Wearden (1992) to fit his data assumes that the subject replies "yes" if the difference between a noisy version of the standard  $s^*$  and the test tone  $t$  is smaller than a threshold  $b \cdot t$  proportional to the duration of the test tone (equation 6)

$$|t - s^*| < b \cdot t \quad (6)$$

This is a purely phenomenological approach. In figure 7, a neural network model is proposed that has a function similar to (6), except for a constant threshold as justified below. In this model it is assumed that the onset of the test tone initiates a self-sustained firing activity in a number of independent clusters probably located in a sensory area of the cortex. During learning (modelling of which is not attempted here) the output of these clusters is adequately connected to a threshold unit and an offset detector, both possibly located in a frontal area of the cortex. The output of this timer activates a neuron in a motor area which initiates the behavioural response.





**Figure 7.** Timing circuit proposed for the duration discrimination task. Description of its operation in the text.

The comparison between the learnt standard  $s$  and the test  $t$  is done as follows: It is assumed that the timer has learnt the standard duration  $s$  and its distribution of firing times is centred on the standard time interval. It is also assumed that there are sensory neurons that detect the offset of the test stimulus and activate directly the *same* "motor" neuron as the timer. By setting the input weights of the "motor" neuron to sufficiently low values, it acts as a coincidence detector and fires only if two input spikes arrive each from one of the two sources within a given time-window (Bugmann, 1997). As the time window is essentially determined by the duration of the EPSPs generated by input spikes and is not related to the duration of the test tone, it has to be assumed that our model has a constant threshold  $b'$ :

$$|t - s^*| < b' \quad (7)$$

To calculate the subject's response probabilities our timer is used as the noisy standard  $s^*$ . The timer has a distribution of firing times that allows test tones of various durations to be sometimes judged as equal to the standard. Using equation (3) the probability that the timer fires between  $t_1 = t - b'$  and  $t_2 = t + b'$  is calculated. This gives the probability  $P_{y,t}$  that the subject replies "yes" for a test stimulus of duration  $t$ :

$$P_{y,t} = M(m,n,t_1) - M(m,n,t_2) \quad (8)$$

This value is brought as close as possible to the experimental data by adjusting the parameters of the model. For comparison, a best fit using equation (6) has also been attempted. The results are given in table 1.

Standard [ms]	$m$	$n$	$\tau$ [sec]	$b$	Tot. Sq. Error
500	62	30	0.685	0.1945	0.0102
600	70	16	0.399	0.187	0.0018
700	63	29	0.9073	0.189	0.0057

**Table 1.** Values for the best fit parameters for the model  $|t-s^*| < b \cdot t$ . The total square error is the sum of the squares of the differences between experimental probabilities of "yes" answer and those predicted by the model, for the seven points shown in figure 6 for each experiment.

Standard [ms]	$m$	$n$	$\tau$ [sec]	$b'$ [sec]	Tot. Sq. Error
500	54	6	0.238	0.101	0.0177
600	83	24	0.5	0.115	0.0096
700	65	30	0.9598	0.1487	0.0033

**Table 2.** Values for the best fit parameters for the model  $|t-s^*| < b'$ . Corresponding graphs shown in figure 6.

The parameters giving the best fits for equation (7) are given in table 2. The quality of the fit is comparable to that obtained with equation (6). The increase in the value of  $b'$  for longer standards reflects the rather constant value of  $b$  in table 1. However, for consistency with the model in figure 7, the values of  $b'$  and  $\tau$  should not depend on the standard. Therefore a fit to the data has then been attempted by using the same values of  $b'$  and  $\tau$  for the three curves. After some trial and error, it was found that  $b' = 0.115$  sec and  $\tau = 0.5$  sec gave good results. Finally it has been determined if the subjects could have learned the standard by modifying only one parameter, either  $m$  or  $n$ . The results are given in table 3. The fit for  $m$  variable is slightly better but probably not significantly, considering that experimental data have errors. An indication for that is the significant better fit obtained with all models to the data for  $s = 600$  ms.

Standard [ms]	$m$	$n$	$\tau$ [sec]	$b$ [sec]	Tot. Sq. Error
500	83   43	29   15	0.5	0.115	0.067   0.033
600	83   51	24   15	0.5	0.115	0.010   0.022
700	83   66	19   15	0.5	0.115	0.053   0.043

**Table 3.** Values for the best fit parameters for the model  $|t-s^*| < b'$  with  $b'$  and  $\tau$  fixed. Two cases are shown on the same table,  $m$  fixed or  $n$  fixed (in these two cases only one free parameter was adjusted).

## 5. Concluding comments

Overall the results indicate that the model has the potential for explaining the experimental data. In particular, it provides a well defined source for the variability of the subjects responses.

Regarding time learning, the fit to the data of this series of experiments equally well supports several learning scenarios. However, the model provides an additional constraint in the form of a relation between behaviour of the Weber curve and the time interval learning mechanisms. This can not be exploited with these data, but data from other experiments may deserve re-examining in this light. More theoretical work is also needed.

The value of  $b'$  of approximately 115 ms found for the window of coincidence detection lies close to the upper limit of what is biologically realistic. It requires EPSPs of long duration, which can be produced by inputs to distal synapses of pyramidal cells, possibly in cortical layer I.

The model is still very simple and can be made more realistic by setting some minimum duration to the sustained firing (as in figure 2)(which may cause problems in learning accurately very short time intervals (Jasselette, Lejeune and Wearden, 1990)). Another refinement is to consider stochastic synapses at the "motor" neuron level (which may resemble the use of a stochastic threshold  $b^*$ , as suggested in (Wearden, 1992)). Further, as the properties of the model depend on synaptic characteristics, it may be worth testing it against known effects of drugs on timing (Meck, 1996).

The anatomical layout proposed in figure 7 is very tentative. It reflects the fact that prolonged responses in sensory cortices are rarely longer than a few hundred milliseconds. In contrast, prolonged firing in the frontal cortex lasts easily a few tenth of seconds (Fuster, 1980). As the time constants of the best fits are all smaller than 1 second, it is therefore possible that the units in layer 1 of the model correspond to normal sensory neurons. From masking experiments, it is known that their prolonged firing is interrupted by subsequent sensory inputs (Rolls and Tovee, 1994). In the duration discrimination experiment modelled here it was assumed that the tone *onset* is the input to the layer 1, so that until a new tone is presented, there is no possible interruption of the firing in layer 1 other than due to the stochasticity of the synapses. A prediction of the model is that parasitic tones starting during the test tone may reinitialise the timer and increase the apparent duration of the standard.

An interesting feature of the proposed anatomical layout is the use of sensory circuits that are already in use for processing sensory information. Hence, the biological timing function may be obtained with a very small number of additional neurons.

**Acknowledgement:**

Comments by John Wearden on earlier versions of the manuscript are gratefully acknowledged.

**References:**

Bugmann G. (1997) "Biologically plausible Neural Computation", *Biosystems*, 40, pp. 11-19.

Bugmann G. and Taylor J.G. (1997) "A stochastic short-term memory using a pRAM neuron and its potential applications", in Beale R and Plumbley M.D. (eds) *Recent Advances in Neural Networks*, Ellis Horwood Publishing, To appear

Also available from <http://www.tech.plym.ac.uk/soc/staff/guidbugm/bugmann.htm>

Church R.M. and Broadbent H. (1990) "Alternative representation of time, number and rate", *Cognition*, 37, pp. 55-81.

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

Fuster J.M. (1980) "The Prefrontal Cortex", Raven Press, New York.

Fuster J.M, Bauer R.H. & Jervey J.P. (1981), "Effect of cooling inferotemporal cortex on performance of visual memory tasks" *Experimental Neurology*, 71, 398-409.

Gibbon J., Malapani C., Dale, C.L. and Gallistel C.R. (1997), "Towards a neurobiology of temporal cognition: advances and challenges", *Current Opinion in Neurobiology*, 7, pp. 170-184.

Grossberg S. and Merrill J.W.L. (1992) "A neural model of adaptively timed reinforcement learning and hippocampal dynamics", *Cognitive Brain Research*, 1, pp. 3-38.

Grossberg S and Merrill J.W.L. (1996) "The Hippocampus and Cerebellum in Adaptively Timed Learning, Recognition and Movement", *J. Cognitive Neuroscience*, 8:3, pp. 257-277.

Jasselette P., Lejeune H. and Wearden J.H. (1990) "The perching response and the laws of animal timing", *J. Exp. Psychology: Animal Behavior Process*, 16, pp. 150-161.

Meck W.H. (1996) "Neuropharmacology of timing and time perception", *Cognitive Brain Research*, 3, pp. 227-242.

Miall C. (1996) "Models of Neural timing", in Pastor M.A. and Artieda J. (eds) *Time, Internal Clocks and Movement*, Elsevier Science, Amsterdam, pp. 69-94.

Milner B, Corsi P & Leonard G. (1991), "Frontal-Lobe contribution to recency judgements", *Neuropsychologia*, Vol. 29, no. 6, pp. 601-618.

Rolls E.T. and Tovee M.J. (1994) "Processing speed in the cerebral cortex and the neurophysiology of visual masking", *Proc. Roy. Soc. London B*, 2257, pp 9-15.

Tasker J.G., Peacock W.J. and Dudek F.E. (1992) "Local synaptic circuits and epileptiform activity in slices of neocortex from children with intractable epilepsy", *J. Neurophysiology*, 67:3, pp. 496-507.

Wearden J.H. (1992), "Temporal generalization in humans", *J. of Exp. Psychology: Animal Behaviour Processes*, 18, pp. 134-144.