

Rate coding versus temporal order coding: a theoretical approach

Jacques Gautrais *, Simon Thorpe

Centre de Recherche Cerveau et Cognition UMR 5549, 133 route de Narbonne, 31062 Toulouse, France

Abstract

The belief that neurones transmit information in the form of a firing rate code is almost universal. However, we argue that at least in some situations, the efficiency of a coding strategy based on rate coding is surprisingly poor. A simple mathematical analysis reveals that, due to the stochastic nature of spike generation, even transmitting the simplest signals reliably would require either: (1) excessively long observation periods incompatible with the speed of sensory processing or (2) excessively large numbers of redundant neurones, incompatible with the anatomical constraints imposed by sensory pathways. We argue that such problems may be avoided by using alternative temporal codes which rely on the asynchrony of firing across a population of afferent neurones. © 1998 Elsevier Science Ireland Ltd. All rights reserved.

Keywords: Poisson process; PDP models; Rate coding

1. Introduction

For decades, neurophysiologists have characterised neural activity by the firing rate: PSTHs, tuning curves and even more recent stimulus-reconstruction methods all rely on stimulus-dependent changes in firing rate. Furthermore, the rate coding hypothesis has undoubtedly shaped the development of the ideas underlying artificial neural networks and PDP models.

However, recent studies that have looked at the speed with which sensory systems can process

information (Thorpe et al., 1996) pose serious problems for traditional rate coding schemes. For instance, face selective neurones in primate inferotemporal cortex can respond only 80–100 ms after stimulus onset (Oram and Perrett, 1992). The fact that the first 5 ms of their responses is already selective suggests that the selectivity can be produced by essentially feed-forward propagation of information from the retina to IT via the LGN, V1, V2 and V4, a processing sequence involving roughly ten synaptic stages. On average, this leaves no more than 10 ms between two successive stages of synaptic activation, i.e. 10 ms for synaptic transmission, PSP conduction and

* Corresponding author. E-mail: thorpe@cerco.ups-tlse.fr

integration, spike generation and spike conduction. We have argued that this period is too short to allow rates to be determined accurately, because few neurones will fire more than one spike in this time (Thorpe and Imbert, 1989).

2. Rate coding of analog values

In this section, we will examine the efficiency of rate coding using a very simple Poisson process for generating spikes. It is clear this is only a very approximate model for the firing of real neurones. However, we believe that the general points that can be drawn will apply to any scheme relying solely on counting the number of spikes.

From the efferent neurone's point of view, the less time there is to evaluate afferent spike rate, the more inaccurate this evaluation. If we assume that spike generation is a Poisson process, the crude statistical interval of confidence on the coded frequency F given an observed frequency $f = \text{spikes}/\Delta T$ is:

$$\frac{1}{2\Delta T} \chi^2_{2 \cdot \Delta T \cdot f; \alpha/2} \leq F \leq \frac{1}{2\Delta T} \chi^2_{2 \cdot (\Delta T \cdot f + 1); 1 - \alpha/2} \quad (1)$$

In that case, an efferent neurone receiving a mean of one spike per 10 ms (i.e. an observed frequency of 100 Hz) could perform an evaluation of the true frequency of its afferent as illustrated in Fig. 1. When the time window is reduced to 10 ms, all one can say is that the true frequency lies in the range [5–474 Hz] with 90% chance of being correct.

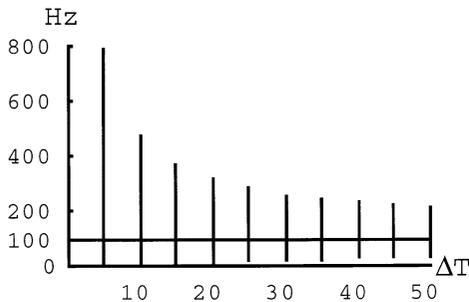


Fig. 1. Confidence interval (90%) on the true frequency of a Poisson process as a function of the time window of evaluation, and given an observed frequency at 100 Hz.

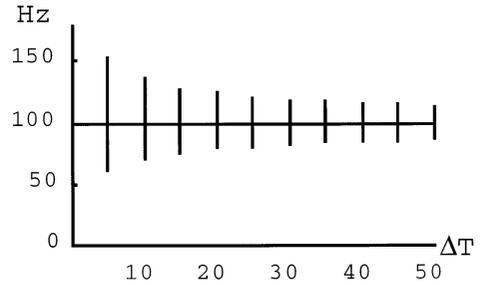


Fig. 2. Confidence interval (90%) on the true frequency of 30 redundant Poisson processes as a function of the time window of evaluation, and given an observed frequency at 100 Hz.

One way to tackle this uncertainty would be to calculate a firing rate across a population of neurones. Let us suppose that we have an array of 30 redundant neurones, each generating spikes according to the same Poisson process. Due to the independence of the interspike interval under the Poisson regime, the estimation interval simply becomes:

$$\frac{1}{2 \cdot n \cdot \Delta T} \chi^2_{2 \cdot n \cdot \Delta T \cdot f; \alpha/2} \leq F \leq \frac{1}{2n \cdot \Delta T} \chi^2_{2 \cdot (n \cdot \Delta T \cdot f + 1); 1 - \alpha/2}$$

since observing one Poisson process during ΔT ms is equivalent to observing n such processes during $\Delta T/n$ ms. How much more precise would the estimation be? As one can see from Fig. 2, the estimate on 10 ms would now be [72–135 Hz], i.e. $\sim 100 \pm 30$ Hz.

In fact, to obtain an estimate as precise as 100 ± 10 Hz in 10 ms would need no less than 281 redundant and independent neurones. This is clearly a very inefficient way to code one single analog value. Furthermore, recent data suggests that the amount of redundancy between neighbouring neurones in the visual system may be surprisingly small (Gawne et al., 1996).

One can also ask whether the observed frequency provides an efficient estimator of the coded frequency. Unfortunately, even this is not so clear. The statistical theory of Poisson processes shows that the estimated frequency calculated as the mean number of spikes per time unit

tends towards the true frequency. However, this is under the assumption that all frequencies are equally probable and it is clear that this is not the case in neural systems. Typically, the firing rate distribution of real neurones is heavily biased towards low frequencies.

Suppose that we have a very simple distribution of discharge frequencies of the afferents following an negative exponential law such as:

$$P[f] = \frac{1}{\bar{f}} e^{-f/\bar{f}} = \tau \cdot e^{-f \cdot \tau} \quad (3)$$

Fixing the mean frequency \bar{f} (or the corresponding mean period τ) would give a distribution of the frequencies like the one shown in Fig. 3.

In this distribution, where the mean frequency has been set at 30 Hz, roughly half of the frequencies are less than 20 Hz, 90% are less than 70 Hz, and only 3.5 % reach beyond 100 Hz. In fact, recent experimental data suggest that the probability distribution in the visual cortex may be shifted even further to the left, since the mean frequency is only ~ 4 Hz in the anaesthetised cat, and not much more than 18 Hz in an awake monkey (Baddeley et al., 1997).

From the efferent point of view, the estimation problem becomes: given this distribution of the true frequencies and the fact that they involve Poisson processes, what would the most likely frequency given a certain number of spikes in ΔT ?

This is the a posteriori conditional probability distribution that can be calculated from the Bayes formula:

$$\begin{aligned} P[f|n] &= \frac{P[n|f]}{P[n]} P[f] \\ &= \frac{(f \cdot \Delta T)^n}{n!} e^{-f \cdot \Delta T} \cdot \tau \cdot e^{-f \cdot \tau} \\ &= \frac{\Delta T^n}{\tau \cdot (\Delta T + \tau)^{n+1}} \\ &= (\Delta T + \tau) \frac{[f \cdot (\Delta T + \tau)]^n}{n!} e^{-f \cdot (\Delta T + \tau)} \quad (4) \end{aligned}$$

Fig. 4 illustrates this result and shows some examples of conditional probability distributions for different time windows and observed frequencies. It is clear that although the estimation becomes more accurate with longer observation windows, with short windows the observed frequency is only a poor estimator of the underlying frequency. For instance, if the efferent neurone receives one spike in the first 10 ms, the most likely afferent frequency is actually 23 Hz and not 100 Hz as the observed frequency would indicate. Even with a time window of 100 ms in which the efferent receives ten spikes (so that the observed frequency is still 100 Hz), the most probable frequency is still only 75 Hz. In fact, to obtain a most probable coded frequency of 100 Hz using a 10 ms time window, the neurone should receive at least four spikes, i.e. an observed frequency of 400 Hz. Thus, the use of rate coding with short time windows is clearly a very inefficient way to transmit information, since the observed frequency will always tend to overestimate of the underlying frequency.

It seems clear, therefore, that rate coding provides only a very poor estimate of an absolute analog value unless a very long observation period is available.

3. Comparisons between two analog values

Neural processing is often aimed at detecting differences in activation rather than absolute values: contrast rather than pure luminance, edges rather than areas, etc. Would a rate coding strategy be more efficient for such a purpose?

Let us take two populations *A* and *B*, each composed of *N* neurones, which send spikes to a

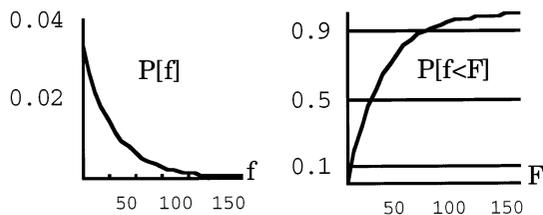


Fig. 3. Distribution of the true frequencies given a mean frequency set to 30 Hz.

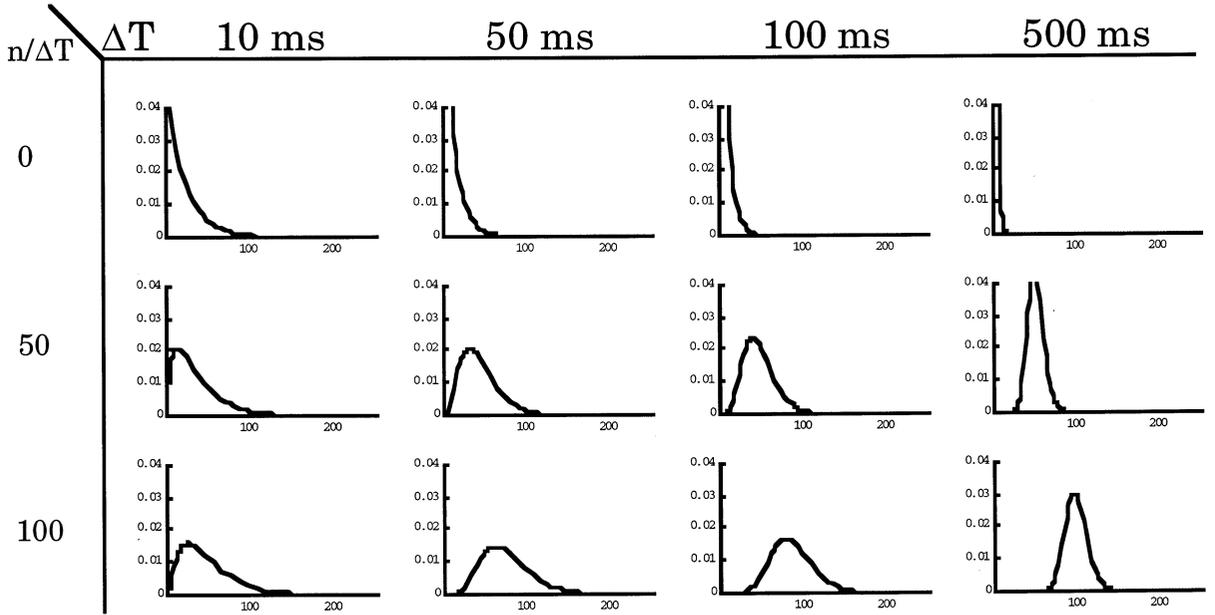


Fig. 4. Conditional probability distribution of the true frequency that gave an observed frequency of $n/\Delta T$ spikes.

second-level neurone D that has to react on the basis of the difference of activity between the two afferent sub-populations. For instance, the efferent neurone D could test whether population A is more active than population B . With a very simple architecture such that each neurone in A sends one positive weight to D , whereas neurones in B have negative weights of the same absolute value, the decision by D will just depend on the relative number of spikes received from A and B .

The decision rule would simply be:

- if there are more A spikes than B spikes, decide A is greater than B ,
- if there are as many A spikes as B spikes, decide randomly between both,
- or else, decide B is greater than A .

Let us take again the Poisson hypothesis, and say that all neurones of A (and respectively for B) emit spikes according to a Poisson process of same parameter FA (FB in the case of B).

We can calculate the error rate α that gives the probability to decide B is greater than A , whereas this is false: this is the probability that the B population emits more spikes than A when $FA > FB$ (with a small correction for the equality case).

With \bar{a} being the mean number of spikes from A , (and respectively for B), we have

$$\begin{aligned}
 \alpha &= P[B > A] + \frac{1}{2} P[B = A] \\
 &= \sum_{a=0}^{\infty} \left(P[B > A] \cdot P[A = a] \right. \\
 &\quad \left. + \frac{1}{2} P[B = A] \cdot P[A = a] \right) \\
 &= \sum_{a=0}^{\infty} \left(\left(\sum_{b=a+1}^{\infty} P[B = b] \cdot P[A = a] \right) \right. \\
 &\quad \left. + \frac{1}{2} P[B = a] \cdot P[A = a] \right) \quad (5) \\
 &= \sum_{a=0}^{\infty} \left(\left(\sum_{b=a+1}^{\infty} e^{-\bar{b}} \frac{\bar{b}^b}{b!} e^{-\bar{a}} \frac{\bar{a}^a}{a!} \right) \right. \\
 &\quad \left. + \frac{1}{2} e^{-\bar{b}} \frac{\bar{b}^a}{a!} e^{-\bar{a}} \frac{\bar{a}^a}{a!} \right) \\
 &= e^{-(\bar{a} + \bar{b})} \sum_{a=0}^{\infty} \left(\frac{\bar{a}^a}{a!} \left(\sum_{b=a+1}^{\infty} \frac{\bar{b}^b}{b!} \right) + \frac{\bar{b}^a}{2a!} \right)
 \end{aligned}$$

As Fig. 5 shows, the error on our decision will of course depend on the gap between FA and FB , the number of neurones and the time window

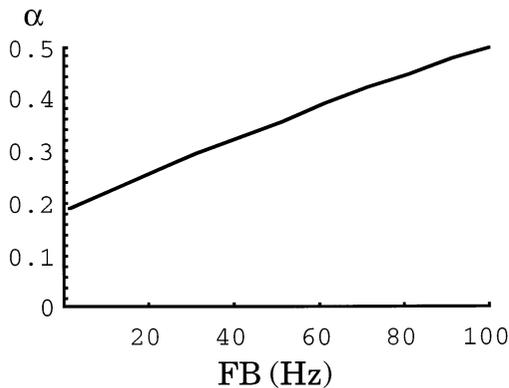


Fig. 5. Error on the decision as a function of FB , given $FA = 100$ Hz, one neurone per population, and a 10 ms time window.

allocated for the decision (i.e. \bar{a} and \bar{b}), so we must specify these parameters to evaluate α .

Let take the case where $FA = 100$ Hz and $FB = 75$ Hz. If we need to decide whether A or B is more active on the basis of the number of spikes received in 10 ms from one neurone in A and one in B , we would have $\alpha = 0.43$, i.e. only slightly above chance. Of course, with larger time windows, the probability of making an error decreases, as shown in Fig. 6. However, we would have to wait at least 760 ms in order to reduce the error rate to 5%. This is far too long to be compatible with the speed of visual processing.

Here again, redundancy across a population of neurones could be used to save the rate code

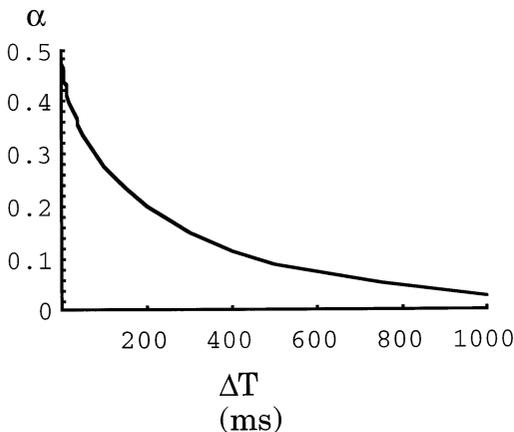


Fig. 6. Decrease of α as the time window increases, for $FA = 100$ Hz, $FB = 75$ Hz and one neurone per population.

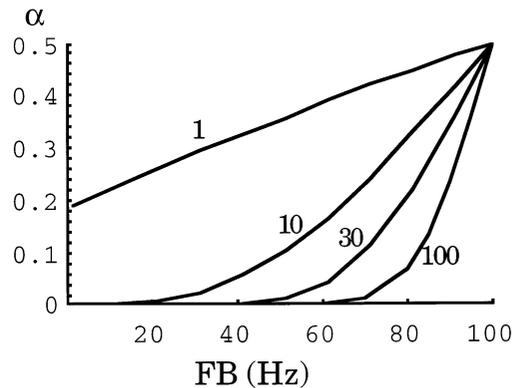


Fig. 7. Error rate as a function of FB for increasing numbers of neurones per population with a time window reduced to 10 ms and with $FA = 100$ Hz.

hypothesis. Instead of waiting 760 ms to be sure of our decision, we could make a decision after only 10 ms, but this would necessitate at least 76 neurones in each population.

Fig. 7 shows how the error rate α changes as a function of FB , given a firing rate in A of 100 Hz, and for various numbers of neurones in each population. It can be seen that even with 100 neurones per population, FB must be no greater than 80 Hz to allow a 95% correct decision in 10 ms. Note that in such a case, we would need 200 neurones to encode less than 1 bit of information (i.e. the information sufficient to make a binary choice accurately), and this amount of redundancy needed rises very steeply with smaller differences in firing rate, as shown in Fig. 8.

This theoretical evaluation of the rate coding strategy demonstrates that it is remarkably inefficient, even when we only need to make qualitative judgements about stimuli. In particular, rate coding has the major drawback that increasing either the number of neurones or the amount of time available actually results in a decrease in the amount of information transmitted per spike.

4. Asynchrony: another way to code

In this section, we will consider some of the alternatives to rate coding. Typically, these involve some form of temporal coding, i.e. a code in

which the timing of spikes plays a crucial role. With very short time scales (10 ms or less), time coding and rate coding tend to become confounded since ultimately, any form of temporal code can be described in terms of very rapid changes in firing rate (Rieke et al., 1997). However, the situation becomes much more interesting if we consider the timing of spikes across a population of afferent neurones.

In recent years, many authors have discussed the possibility that neurones could be sensitive to the degree of synchrony between their afferent neurones, but another option is to make use of the fact that afferents will typically fire asynchronously. In the case of the visual system, stimulus dependent asynchrony is well established. The latencies of visual neurones tends to increase progressively as one goes from the retina through to higher order visual areas (Nowak and Bullier, 1997). At the retinal level, the onset latency of ganglion cells depends on both local luminance and contrast (Levick, 1973). LGN cells have latencies that depend on both contrast and spatial frequency (Sestokas and Lehmkuhle, 1986). Also, in V1, Celebrini et al. (1993) reported that the latency of the visual responses for some cells was clearly dependent on stimulus orientation.

This asynchrony is easily explained with a simple leaky integrate-and-fire model with a continuous input, as illustrated in Fig. 9.

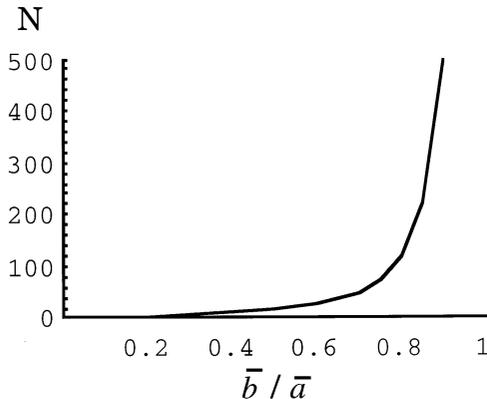


Fig. 8. Number of neurones per population needed to reach 95% decision correctness as a function of \bar{b}/\bar{a} with a time window set at 10 ms.

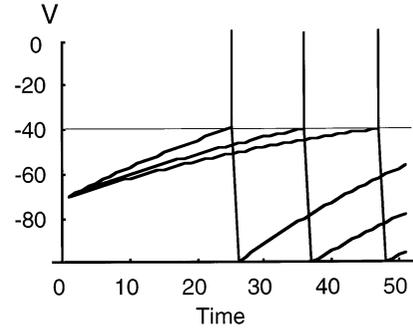


Fig. 9. A simple leaky integrate-and-fire model fed with a continuous input results in spikes whose latency depends on the strength of the input.

This sort of leaky integration mechanism can be modelled as follows:

$$\frac{dV}{dt} = -\text{Leaks} + \text{Inputs} = -\frac{1}{\tau}(V(t) - R) + I \quad (6)$$

With $V(0) = R$, we would obtain:

$$V(t) = R + I\tau(1 - e^{-t/\tau}) \quad (7)$$

With a threshold T and $I_s = T/\tau$, this leads to a latency that depends on input:

$$\begin{aligned} \text{Latency}(I) &= -\tau \cdot \text{Ln} \left(1 - \frac{T}{\tau I} \right) \\ &= \tau \cdot \text{Ln} \left(\frac{I}{I - I_s} \right) \end{aligned} \quad (8)$$

As a result, such neurones can be thought of as performing a form of analog-to-delay conversion. With a retinal array of such units, a luminance profile presented on the retina can be encoded in the relative latencies of firing, even under conditions where each cell only emits one spike (Fig. 10).

Could this asynchrony be used by neurones at later stages? A coincidence detection model would lead to second-level leaky neurones that respond to iso-luminant areas because of the better synchrony of their inputs (Burgi and Pun, 1994; Opara and Wörgötter, 1996). Alternatively, asynchrony can also be used directly as a way to encode stimulus information (Thorpe, 1990).

5. Rank order coding

One simple way of using asynchrony is to use the order in which the neurones spike as a code. In this case, the exact latency at which a neurone fires is not critical—only the rank order of each neurone is important (Thorpe and Gautrais, 1997; 1998). Such a scheme offers a number of advantages.

Firstly, a code based on the order will be more robust to noisy temporal jitter of each spike than a pure temporal code that must rely on temporal precision, especially when decoded by coincidence detectors. Note that trial to trial variations in the latency of individual neurones are not necessarily problematic, since the critical feature is the order of firing—any fluctuations which affect the whole population will have no effect on the order.

Secondly, the order in which neurones fire is fully invariant with respect to changes in both the contrast and the overall luminance of the image on the retina, as illustrated in Fig. 11.

The fact that invariance to contrast and luminance variations is directly embedded in the code

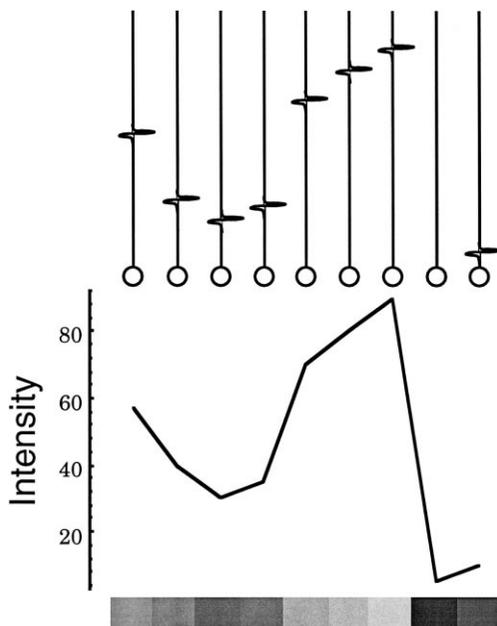


Fig. 10. A luminance profile can be encoded in the relative latency of spikes emitted by the receptors.

is a very useful property which fits with the known characteristics of the human visual system, but is difficult to obtain in more conventional coding schemes without using relatively complex normalisation procedures.

A third advantage of rank order coding is that it can be used to transmit a large amount of information very rapidly—much more efficiently than with conventional rate coding schemes. Consider a simple scheme where n afferent neurones can each emit only one spike. In the case of a rank order code, there are factorial n possible arrangements, which, if we assume that each arrangement is equally likely, means that the maximum amount of information that can be transmitted is $\log_2(n!)$ bits. Contrast this with a population rate coding strategy which essentially just counts the number of spikes generated. In this case there are just $n + 1$ possible states (if we include the case where none of the neurones fires) leading to an upper limit of $\log_2(n + 1)$ bits. With 15 afferent neurones, an optimal count code can convey a maximum of 4 bits whereas the rank order code can convey up to 40 bits. Note that in the case of a count code, every additional neurone can only provide $\log_2(n + 1)/n$ bits of extra information. In contrast, when using an order code, the same neurone can add up to $\log_2(n!)/n$ bits of information.

Put simply, with a count code, the more neurones you have, the less information can be carried by each. In contrast, with rank order coding, the amount of information that can be transmitted by each neurone increases with the number of neurones available (Fig. 12).

6. Conclusions

The idea that neurones transmit information in the form of a rate code is extremely entrenched. There have been numerous other suggestions over the years (Perkel and Bullock, 1968), but they have done little to overturn the overwhelming popularity of the rate coding hypothesis. Even today, with more and more researchers interested in the possibility that temporal synchrony might play an important role in neural computation

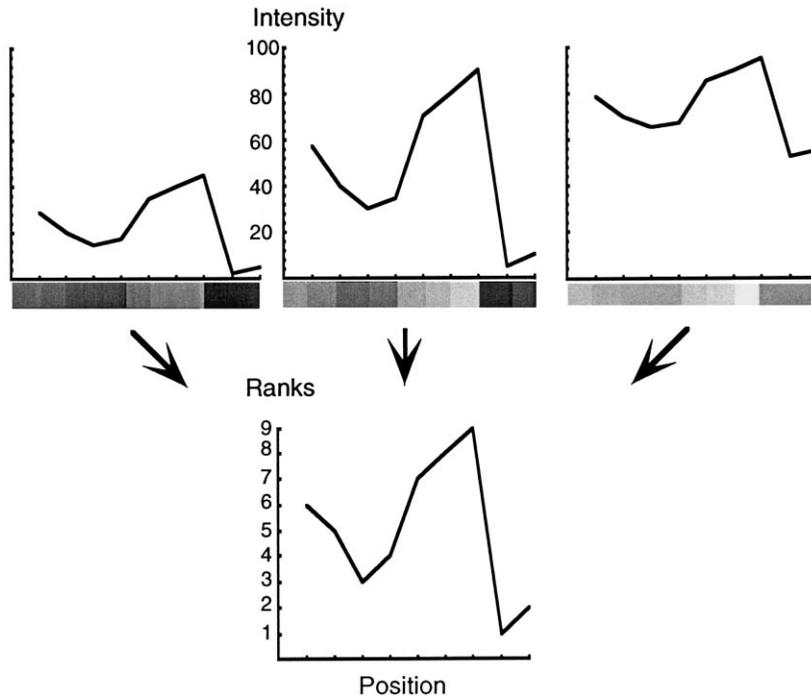


Fig. 11. Order coding is independent of contrast and mean luminance.

(Abeles, 1991, Singer and Gray, 1995), most people still consider that firing rate is the underlying code. Recently, however, it has been shown that because of activity-dependent synaptic depression, many central synapses are effectively unable to signal changes in maintained firing rates over about 30 spikes per second (Abbott et al., 1997). There are even some data from intracellular

recordings from visual cortical neurones that suggest that transmission of trains of geniculo-cortical spikes is highly unreliable—while the first spike in a pair produces a highly reliable response, the second spike often fails completely unless the interpulse interval is around 100 ms (Stratford et al., 1996). If confirmed, such results imply that all frequency related information over approximately ten spikes/second is effectively suppressed at the level of geniculo-cortical synapse.

In this paper, we have provided additional theoretical arguments that raise serious doubts about the view that rapid sensory processing could be achieved using rate coding. The mathematical arguments used were mainly based on a simple Poisson model for spike generation, but similar conclusions would apply to more complex stochastic models of spike generation. For example, the Poisson model ignores the existence of a refractory period which would prevent very closely separated spikes in the same neurone. However, this is hardly a problem in the situations analysed here, because we are specifically

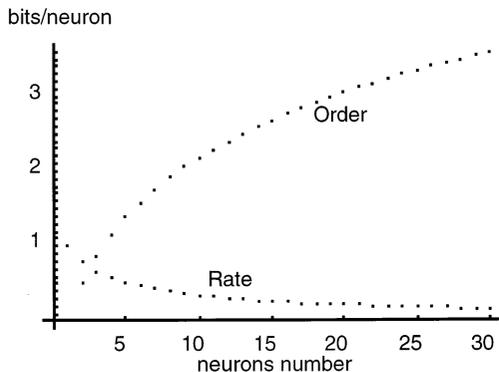


Fig. 12. Mean information per neurone for a count code and an order code.

interested in very short observation periods (< 10 ms) where the probability of having two spikes from the same neurone is very small.

Fortunately, there are alternatives to classical rate coding. It is clear, for example, that using the asynchrony inherent in sensory processing and in particular the order in which cells fire opens a whole range of largely unexplored possibilities (Thorpe and Gautrais, 1997, van Rullen et al., 1998).

Of course, there are many situations where the use of rate coding may well be the best choice for the nervous system. For example, in the motor system, the amount of force developed by a muscle will presumably depend on the total amount of neurotransmitter released at the neuromuscular junction, which in turn will directly depend in a fairly simple way on the firing rate of the motor neurones. It seems likely that a similar situation would exist in many brain structures. However, we would argue that at least when processing has to occur very rapidly, other coding strategies may be required.

References

- Abbott, L.F., Varela, J.A., Sen, K., Nelson, S.B., 1997. Synaptic depression and cortical gain control. *Science* 275, 220–224.
- Abeles, M., 1991. *Corticonics. Neural circuits of the cerebral cortex*. Cambridge University Press, London.
- Baddeley, R., Abbott, L.F., Booth, M.C.A., Sengpiel, F., Freeman, T., Wakeman, E.A., Rolls, E.T., 1997. Responses of neurons in primary and inferior temporal visual cortices to natural scenes. *Proc. R. Soc. London* 264, 1775–1783.
- Burgi, P.Y., Pun, T., 1994. Asynchrony in visual analysis: Using the luminance-to-response-latency relationship to improve segmentation. *J. Opt. Soc. Am. A* 11, 1720–1726.
- Celebrini, S., Thorpe, S., Trotter, Y., Imbert, M., 1993. Dynamics of orientation coding in area V1 of the awake primate. *Visual Neurosci.* 10, 811–825.
- Gawne, T.J., Kjaer, T.W., Hertz, J.A., Richmond, B.J., 1996. Adjacent visual cortical complex cells share about 20% of their stimulus-related information. *Cereb Cortex* 6, 482–489.
- Levick, W.R., 1973. Variation in the response latency of cat retinal ganglion cells. *Vision Res.* 13, 837–853.
- Nowak, L.G., Bullier, J., 1997. The timing of information transfer in the visual system. In: J. Kaas, K. Rocklund, A. Peters (Eds.), *Extrastriate cortex in primates*, vol. 13, Plenum, New York, pp. 205–241.
- Opara, R., Wörgötter, F., 1996. Using visual latencies to improve image segmentation. *Neural Comput.* 8, 1493–1520.
- Oram, M.W., Perrett, D.I., 1992. Time course of neural responses discriminating different views of the face and head. *J. Neurophysiol.* 68, 70–84.
- Perkel, D.H., Bullock, T.H., 1968. Neural coding. *Neurosci. Res. Program Bull.* 63, 221–348.
- Rieke, F., Warland, D., Ruyter van Steveninck, R., Bialek, W., 1997. *Spikes: Exploring the neural code*. MIT Press, Cambridge, MA.
- Sestokas, A.K., Lehmkuhle, S., 1986. Visual latency of X- and Y-cells in the dorsal lateral geniculate nucleus of the cat. *Vision Res.* 26, 1041–1054.
- Singer, W., Gray, C.M., 1995. Visual feature integration and the temporal correlation hypothesis. *Annu. Rev. Neurosci.* 18, 555–586.
- Stratford, K.J., Tarczyhorno, K., Martin, K.A.C., Bannister, N.J., Jack, J.J.B., 1996. Excitatory synaptic inputs to spiny stellate cells in cat visual cortex. *Nature* 382, 258–261.
- Thorpe, S., Fize, D., Marlot, C., 1996. Speed of processing in the human visual system. *Nature* 381, 520–522.
- Thorpe, S.J., 1990. Spike arrival times: A highly efficient coding scheme for neural networks. In: R. Eckmiller, G. Hartman, G. Hauske (Eds.), *Parallel processing in neural systems*, North-Holland, Amsterdam, pp. 91–94.
- Thorpe, S.J., Gautrais, J., 1997. Rapid visual processing using spike asynchrony. In: M.C. Mozer, M.I. Jordan, T. Petsche (Eds.), *Advances in Neural Information Processing Systems 9*, MIT Press, Cambridge, MA, pp. 901–907.
- Thorpe, S.J., Gautrais, J., 1998. Rank Order Coding: A new coding scheme for rapid processing in neural networks. In: J. Bower (Ed.), *Computational Neuroscience: Trends in Research 1998*, Plenum Press, New York, pp. 113–118.
- Thorpe, S.J., Imbert, M., 1989. Biological constraints on connectionist models. In: R. Pfeifer, Z. Schreter, F. Fogelman-Soulié, L. Steels (Eds.), *Connectionism in Perspective*, Elsevier, Amsterdam, pp. 63–92.
- van Rullen, R., Gautrais, J., Delorme, A., Thorpe, S., 1998. Face processing using one spike per neurone, *Biosystems* 48, 229–239.