INFORMATION FLOW AND TEMPORAL CODING

IN PRIMATE PATTERN VISION

J. Heller
Division of Applied Sciences, Harvard University, Cambridge, MA.

J. A. Hertz
Nordita, Blegdamsvej 17, DK-2100 Copenhagen Ø, Denmark.

T. Kjær
Laboratory of Neuropsychology, National Institute of Mental Health, Bethesda MD 20892, USA.

B. Richmond
Laboratory of Neuropsychology, National Institute of Mental Health, Bethesda MD 20892, USA.


1Current address: Jewish Theological Seminary, New York, NY.
2Current address: Nordita, Blegdamsvej 17, DK-2100 Copenhagen Ø, Denmark.
3Reprint requests: Laboratory of Neuropsychology, 49 Convent Drive, Bethesda, MD 20892-4415.

NORDITA - Nordisk Institut for Teoretisk Fysik
Blegdamsvej 17 DK-2100 København Ø Danmark
In press: J. Computational Neuroscience.

Information Flow and Temporal Coding in Primate Pattern Vision

Josha Miller *

John A. Herz

Troels W. Kier

Barry J. Richmond

Abstract

We performed time-matched calculations of the information transmitted about visual patterns by neurons in primary visual and inferior temporal cortices. All measurable information is carried in an effective trans-

1 2 3 4

ning rate, obtained by averaging the neuronal response with a resolution of 1/125 s. The results show that the firing rate of individual neurons is strongly correlated with the firing rate of neurons in the primary visual cortex.

Keywords: 1. introduction

Recent studies have established [16, 15, 12] that neurons in the primary visual cortex convey information about the timing of their action potentials. However, very little is known about just how this information is represented in their responses. In this investigation we try to determine what features of the neuronal response carry information about a stimulus. We ask two questions in particular:

1. At what degree does the information transmission depend on the firing rate of the spikes?

2. What is the time course of the signal? For instance, is the information transmitted in a single short burst, or distributed more evenly over the course of the response?

We address these questions with two basic assumptions: primary visual cortex (V) and inferior temporal cortex (IT).

A visual system may be seen as a communication system. The visual stimulus (a message) is transmitted by the neurons (transmitters) into a set of neural firing patterns (a code). The neurons (acting as a channel) pass that encoded message on to other neurons (reservoirs). Ultimately, the signals are used by other brain areas (the destinations) to determine the nature of the original stimulus. In fact, information about the stimulus is carried by many neurons in the visual system, and, as working as ensembles, they can transmit much more information than any one neuron. However, our goal in this study is to quantify the behavior of the building blocks of the system, individual neurons.

Probing the structure of this neuronal code requires a reliable tool for estimating transmitted information. We use an artificial neural network designed for this task [9, 11]. The input to the network is a representation of the neuronal re-

1. METHODS

1. Data collection

The experiments that yielded the data analyzed here have been reported previously [17, 27]. They were collected from two different visual areas, V1 and IT, in awake monkeys, using standard techniques in macaque monkeys. The times of each action potential and each stimulus were recorded with a resolution of 1 μs.

In total, recorded from V1 in three monkeys, 11 produced enough data for our analyses. All 11 were complete cells, located in the extrastriate visual field. The receptive fields were located in the lower contralateral visual field, 15 degrees in the lower contralateral visual field. The IT neurons were analyzed from 10 neurons, located in the inferior temporal cortex. Of the IT neurons, we analyzed the response of the IT that had the largest amount of stimulus related information.
The stimuli used for the studies were Walsh patterns: two-dimensional black and white 4 x 4 patterns based on Walsh functions, plus their contrast-reversed counterparts, making a total of 32 stimuli (Fig. 1).

3.4. Statistics
Each stimulus is a possible message for the neuron to transmit. The output (or transmitted) information is

\[ I(S; R) = \sum_x P(x) \log_2 \left( \frac{P(x|R)}{P(x)} \right) \]

where \( S \) is the set of stimuli, \( R \) is the set of signals (here the neuronal responses) \( x \), \( P(x|R) \) is the conditional probability of stimulus class \( x \) given an observed response \( x \), and \( P(x) \) is the prior probability of stimulus \( x \). The brackets indicate an average over the signal distribution \( P(x) \).

Estimating conditional probability for categorical data is a standard estimation problem. We employ a conventional measure of goodness of fit, maximum likelihood. We select the parameters of the model that come closest to predicting our data, using as our cost function the negative log-likelihood

\[ E = -\sum_x P(x) \log_2 P(x|R) \]

where \( P(x|R) \) is the data used to make the fit, \( P(x) \) is the stimulus and \( P(x|R) \) is the response in training set. \( E \) measures the degree to which the data and the fit form of \( P(x|R) \) differ, and we refer to it broadly as the 'fit error,' \( E \), or simply as the error. It can be calculated both for the data used to make the fit (training error) and for independent data (test error).

3.5. Neural network model
A neural network can be trained using backpropagation so that given the input it outputs provide an estimate of the conditional probabilities \( P(x|R) \). Our backpropagation model was very much like the standard backpropagation model [20]. The model is pictured in Figure 2, and described in detail in Kjaer et al. [99].

The error (3) is used as the cost function for the backpropagation algorithm, leading the network to search for the parameters in the fit to \( P(x|R) \) which give the largest log-likelihood. The learning rate \( \eta \) and the moment \( \alpha \) are used to control the speed/accuracy tradeoff of the learning, and we always use \( \eta \) and \( \alpha \) to be between (0.5) and (0.1), depending on the representation and the cell.

Once we have trained the network so that its outputs \( \hat{P}(x|R) \) provide a good estimate of \( P(x|R) \), we can substitute \( \hat{P}(x|R) \) for \( P(x|R) \) in the expression (1) and average over a data set \( \{x; r\} \) to estimate the transmitted information

\[ I_{\text{est}}(S; R) = \sum_x \hat{P}(x|R) \log_2 \left( \frac{\hat{P}(x|R)}{P(x)} \right) \]

with \( P(x) \) estimated as \( \frac{1}{n} \sum_i P(x) \), where \( n \) is the total number of samples.

A neural network can be trained to an arbitrary degree of accuracy as long as the number of samples is finite. However, at some point, the network's learning is based on features of the specific sample, rather than on features of the data in general. We used the 'early stopping' procedure described in Kjaer et al. [99] to control such overfitting. We divided the data into training and test sets. This work was tested on the test set while the training set was used to train the backpropagation algorithm. Training was stopped when the test set error reached a minimum. (3)

In order to obtain the final estimates of information (1) and receiver fitting error (2), the individual estimates for the four test sets were averaged.

Fig. 2. The network. The components of the representation of the stimuli comprise the inputs, the hidden units have Walsh pattern functions, and the output units are normalized exponential functions of a single linear combination of the hidden units and the excitatory and inhibitory inputs (1) for all other cases.

Since our tests were used to determine the network training time, the correct procedures for estimating the transmitted information would be to perform the sample (3) over an entirely different data set, and makes in training too in determining how to stop the training of the network. However, we have found in previous investigations on these data that, within the confidence limits estimated in the above manner, the values of \( I_{\text{est}}(S; R) \) do not depend on whether the sampling in over our training data, our test data, or both new data [11].

4.3. Data limitations
In order to examine the time course of the information flow, we evaluated the response using two different types of time structure. We derived an estimated latency \( \tau \), which was later updated to reflect more accurate calculations, for each cell. The first window begins at \( \tau \) and was \( \tau \) long.
2.5. Representations

Once a time segment of the response had been selected for study, it had to be represented in a form which could be used to test the network. We used all of the features in Table 1 for growing windows of widths 16, 22, 24, 26, 32 and 44. The representations were compared on the basis of their average test set errors. These preliminary results had been obtained, a few representations were used on a wider selection of widths. A few of these representations were then used on sliding windows of widths 16, 26, 32 and 44. All of these features were performed on each cell individually.

2.6. Latency

A visual system response can change its firing rate earlier in response to same stimuli than to others, or the firing rate might not change at all. Latency is often defined as the delay until the cell's first response to stimulus. However, we are particularly interested in representations that could be used to differentiate same stimuli. Therefore we defined an information latency as the time since enough change had taken place in the response that stimulus could be distinguished at a threshold level. To determine this threshold we first computed the best error and transmitted information obtained when the network was trained on spontaneous (non-stimulus-related) activity, using a 16 ms-wide window which was swept from 100 ms before the stimulus to 12 ms after it. The windows were spaced at 4 ms intervals. Within each window, the data were sorted in classes and re-ranked using only the event code. For each cell, we were able to obtain both the error and information values, with their means and standard deviations, for 20 window positions (Fig. 3). The information was dropped below 0.05 bits, and its distribution shows a tail which extends up to about 0.02 bits. The distribution of the test error has a mean of 3.002) and a standard deviation of 0.284 (0.310). The standard deviation is 1.53% of the total variance (0.048) and the test information is 0.057 (0.030).

Fig. 3. The distribution of the test error and test information. The error bars indicate significant differences among stimuli to. To determine the information latency, we performed this calculation of the mean 50 errors on the 10-ms-wide time window was moved along the response period. When we reached the point where the current fell below our criterion, we set the latency to the mean of the distribution of the test information for the next time interval. The deviation and information for the next time interval were found by subtracting the information of the current time interval from the information of the next time interval. This procedure defined the latency values and used for growing windows in all of our subsequent analyses.

We also established a criterion for a firing latency, as follows. We computed the SPKs in each window across all trials with stimuli and smoothed the resulting curve with a Gaussian (σ = 1.5 ms) to produce a spike density output. We defined the firing latency-based latency as the first time at which this smoothed spike density exceeded the value of the previous mean of the peak.
3.2. Comparison of representations

All of the representations were tested on a set of windows of durations 16, 32, 64, 128 and 256 ms, of which had their edges aligned as in Figs. 6C and 6D. Each representation was tested on each window size for each cell. For each window size in each brain region, the best error value (Fig. 5) and the information for a given representation were averaged across all cells and recorded (Fig. 7). For small window widths, the cost code (c) did as well as any other representation. For wider windows, it usually carried a large proportion of the information conveyed in the best representation. On average, this fraction was 78% ± 23% in VI and 89% ± 33% in IT.

The binary code (b), preferred fewer than all the other representations. The proportion of components of this code (b) preferred better, but still not as well as other representations. The binary code was compared with truncated Gaussian kernels and was found to be a lower rate (every 5 ms or so) in some studies with wider Gaussian kernels, larger values of K produced increasingly better codes for high and low thresholds. When the principal components of the Gabor fits were entered into the binary code, they were used differences in error.
The instantaneous rate of information transmission was estimated as a function of time by sliding a window along the response, as described above under latency. However, while each window used in the latency analysis was identified by its starting point, each window used in measuring instantaneous information rate was identified by its center point, so features found using sliding windows of different widths could be compared and aligned.

For each window position, we estimated the information using two different indices: the constant (c), and the combined score (R(t)) consisting of the constant together with up to three principal components of the Gaussian-accorded (p = 5) data.

We first used a 20 ms sliding window. Combining the principal components with the constant did not yield higher information rates than the constant for either IT or V1 data. The mean (±SEM) relative rate for a 20 ms sliding window. This reveals that one window, the spike count (c), is a statistical feature that is the result of combining the information in the data over the 15 ms or 24 ms time scale. These additional features with smaller resolution, the PCs, would have reflected that fact and carried additional information. Thus, a 20 ms response could be used to estimate an instantaneous host rate as the information measured in a time window centered at the time the question divided by the window duration. There are times and intervals in the rate, as can be seen in Figure 1A and C cells in Figure 10. The peak host rates observed in different cells range from 0.2 to about 30 bits/s. We experimented with three shorter sliding windows (12 ms), but the host values were higher and more variable, reflecting the relative lack of spikes in such short time intervals.

For 20 ms sliding windows, all of the cells showed one or two periods early in the response, followed by the peak of the instantaneous information rate, when two principal components were necessary for the optimal representation (D). However, the differences between the results with the principal components and those based on the spike count were more normal for most IT cells and some V1 cells. The 20 ms sliding window shows greater discrepancies between the information measured by the count compared to that carried by the window with the principal components added, particularly at the points of highest information (though not in all cells). In those cells, however, the contribution of principal components was needed (Fig. 9 and Fig. 10, patches C and D).

The count in the 24 ms sliding window rarely carried more information than the count in the 20 ms sliding window, however, for most V1 cells, at the time of peak information rate, it actually carried less information than the count in the 24 ms sliding window. However, for some V1 cells, the time of peak information rate, it actually carries less information than the count in the 24 ms sliding window.
4. Discussion

These results extend considerably our understanding of temporal coding of information about visual patterns in primates. Previously, it had been established that neurones in V1 and IT code some of this information in a more-or-less delayed way in terms of the timing of their spikes [38, 14, 19]. Reliable estimates of the relationships of this information are now possible [38, 39], but very little was known about how the patterns were coded in time. Our results permit us to determine the temporal resolution with which the ravens transmitted by single neurones need to be read and how the information they convey varies in time, the course of the response. These findings, in turn, place constraints on possible codes these neurones use.

5. The temporal resolution limit

Characteristic temporal resolution limits of the order of 25 ms (V1) and 30 ms (IT) emerge from our results at several points. The first piece of evidence comes from the comparison of the many different representations of the same response. In both V1 and IT, the neurones that showed a 25 ms time window for V1 data were not as sharp as those for IT data. This means that Fourier components of the responses which were smaller than 1250/27 = 27 Hz but 27 Hz were present in IT neurones, whereas the corresponding component leads to a suppression of variations in the response shorter than 48 ms. The test for a small number of principal components introduces additional low-pass filtering, so rapid variations in the response are suppressed even further than in both these optimal representations, and we can be sure that the information is coded over the cutoff frequencies of 27 Hz (V1) or 10 Hz (IT). Nevertheless, the fact
that the current Gibson house produced a significantly (almost significantly) better fit than the wilcoxon for V1 (and vice versa for IT) implies a difference in the temporal resolution scales of the two networks.

Nearly the same trade-offs emerge from the comparison of the performance of different representations in different time windows. The codes which are optimal for the maximal window, encompassing the entire 100 ms of the response, have much of their advantage over other representations in the shorter time windows. In particular, it is evident in Fig. 6 that for windows of 20 ms in V1 and 32 ms or less in IT, the spine count yields a fit virtually as good as the optimal code and conveys essentially all the information it does. Thus, no improvement would occur by analyzing changes in firing rate or timescales shorter than these. On the other hand, for windows wider than these, information is lost if the temporal variation of the firing rate within the window is not taken into account. This places the temporal resolution limits between 26 and 22 ms for V1 and between 32 and 44 ms for IT.

One can see the same transition in the more detailed cumulative information curves (Figs. 9 and 10). At short times, the spine count and the optimal representation transmit essentially the same information, but at later times the curves for the two codes diverge, revealing that temporal variation of the response within the window conveys information. The characteristic times are once again consistent, with the 35 and 37 ms identified in the two areas above. These calculations show that at some times in the windows longer than these values, it is necessary to take into account the temporal variation of the firing rate in these time windows to extract all the information from the signal. The sliding window curves are, of course, identical to the cumulative information curves. The characteristic times are once again consistent, with the 35 and 37 ms identified in the two areas above.

The cumulative information curves tell the same story. They are almost monotonically characterized by a sharp rise at the very beginning of the response, contrasting with the flat peak in the sliding-window curves. Most of the relevant information is transmitted in the first period of 20-40 ms (V1) or 100 ms (IT). The shape of the envelope after this point is almost noiseless, and the transmitted information grows monotonically to a somewhat higher value (25% higher on average) during the entire 200 ms of the response. All of the information calculated in nonoverlapping sliding window segments after 100 ms, the subject of the sliding window curves, would match the cumulative curves. This is not the case. A good deal of the information carried in the sliding window after the first 60 ms is not exhaustible. On the other hand, some of it is new; otherwise, the cumulative curves would not continue to rise.

The peak transmission times we find in the initial bursts of some of our codes are of the same order as those reported by Edelman and Pigeot in the LGN of the cat [3] and somewhat smaller than those reported by Banki et al. in the H1 motion-sensitive neuron in the monkey [6]. It is an open question whether our cells would maintain such rates in response to rapid changes in stimulus. It is important to note that the H1 cells each represent a large fraction of all visual information carried by the system, while each of the V1 and IT cells we studied is one of many subchannels that convey only a small fraction of that information.

4.2. The time course of information flow

Deeper insight into the nature of the neural code is obtained through our systematic measurements of information flow using sliding and growing time windows. The former reveal a preferred pattern in the instantaneous transmission rate, the peak V1 and IT maxima. There is always an initial peak, peaking very quickly after latency (we have defined the latencies) 20-40 ms window placed over the onset of this peak contains, on the average, half the total information in the measured response as a whole. For V1, the significance of this point is that the superior spatial resolution of the V1 responses is critical for the early discrimination. Once this point passes, the code begins to fade away, typically about 150 ms later, carrying considerable new information. In addition, some information is stored between and after those bands. The null space varies irregularly in time.

The cumulative information curves tell the same story. They are almost monotonically characterized by a sharp rise at the very beginning of the response, contrasting with the flat peak in the sliding-window curves. Most of the relevant information is transmitted in the first period of 20-40 ms (V1) or 100 ms (IT). The shape of the envelope after this point is almost noiseless, and the transmitted information grows monotonically to a somewhat higher value (25% higher on average) during the entire 200 ms of the response. All of the information calculated in nonoverlapping sliding window segments after 100 ms, the same way in the cumulative information curves. The characteristic times are once again consistent, with the 35 and 37 ms identified in the two areas above.

These calculations show that at some times in the windows longer than these values, it is necessary to take into account the temporal variation of the firing rate in these time windows to extract all the information from the signal. The sliding window curves are, of course, identical to the cumulative information curves. The characteristic times are once again consistent, with the 35 and 37 ms identified in the two areas above.

This means that codes including principal components give better fit and more information during the portion of the response in which the information rate is high. For V1, as much information is transmitted by codes including principal components (Figs. 5, panels C and E) than by the spike count alone (panels A and B), whereas the window is wider than 34 ms. However, at other times, the spike count appears to be an adequate measure of the response even over 64-msec periods. Thus, our full temporal resolved uation figure is relevant only to those relatively brief periods of high transmission rates. A corresponding result is found in IT. Employing more than one principal component of the repzen generall means to better fit and higher information for finer windows only around the times of local maxima in the information rate. At other times, the characteristic temporal resolution is apparently larger.

4.3. Implications for the neural code

These results provide insight into the way these neurons code information. They tell us just what aspects of their spike timing convey information to cells that receive their signals. Of course, we have not determined whether those encoding spikes make use of all the information we measure or to what extent signals from a population of cells may be combined synergistically. Nonetheless, we can place bounds on the code's basic units, typically about 100 ms later, carrying considerable new information. In addition, some information is stored between and after those bands. The null space varies irregularly in time.

The cumulative information curves tell the same story. They are almost monotonically characterized by a sharp rise at the very beginning of the response, contrasting with the flat peak in the sliding-window curves. Most of the relevant information is transmitted in the first period of 20-40 ms (V1) or 100 ms (IT). The shape of the envelope after this point is almost noiseless, and the transmitted information grows monotonically to a somewhat higher value (25% higher on average) during the entire 200 ms of the response. All of the information calculated in nonoverlapping sliding window segments after 100 ms, the same way in the cumulative information curves. The characteristic times are once again consistent, with the 35 and 37 ms identified in the two areas above.

These results provide insight into the way these neurons code information. They tell us just what aspects of their spike timing convey information to cells that receive their signals. Of course, we have not determined whether those encoding spikes make use of all the information we measure or to what extent signals from a population of cells may be combined synergistically. Nonetheless, we can place bounds on the code's basic units, typically about 100 ms later, carrying considerable new information. In addition, some information is stored between and after those bands. The null space varies irregularly in time.
the total information in V1 and 19% in IT. However, the rest of the information we have measured is statistically insignificant, and a full description of the neural code for these cells must take the coding of this additional information into account. Nurse et al. [29] have also carried out information measurements on IT cells. They found that when the first 120 ms of the response was removed from the interval analyzed for information about the stimuli, the net information dropped slightly. This reduced amount could be accounted for almost entirely in terms of firing rate alone. Attributing the drop when the initial part of the response was excluded to loss of information about when the spike train started, they suggest that a combination of "static characteristics of the spike train" and the subsequent mean firing rate carried essentially all information in these cells. If the only static characteristic we use is the first spike latency, this combination is exactly our code cSTR, and in our IT cells it accounted for 98% of the total information. In V1, it was not as successful, yielding 64% of the information carried by the optimal code. However, even in IT, the optimal code consisting of the spike count plus principal components did even better than cSTR, and we have learned, moreover, that IT neurons firing more than 32 times carry information.

The first 10 ms of the response of our IT neurons typically contains only 2 spikes, as a single spike can be said to carry a good deal of information. However, apparently it is the presence or absence of these spikes in this period, not their exact timing, that is most important. If "static characteristics" is taken to mean the spike rate in the first, say, 50 or 100 ms, the code proposed by Nurse et al. amounts to a simple combination of two numbers: the spike count in the first 10 ms and the mean firing rate in the following 90 ms. This is in fact the basis for the optimal static characteristics we find. All our results are consistent with the hypothesis that all information about the stimulus is encoded in an effective time-varying firing rate defined by averaging the spike train over a suitable time window. This window should be about as much as the delay window for high information transmission rates, about 25 ms for V1. Elsewhere in the response, longer averaging times, perhaps 100 ms or more, are adequate. The extra degrees of freedom of the response after the initial burst are necessary to give the time rate information seen in the entire response period for most cells. The best code approximates three degrees of freedom by a single number, the post-burst spike count. The temporal code ("early", "middle", and "late" firing rate) proposed by Nurse et al. is only one step beyond this. These are not qualitatively bad approximations, and for the fit of our V1 cells for which the optimal number of principal components was 3 or 4, the latter is essentially equivalent to our optimal model. However, for the remaining 3/4 of our cells, our optimal representation has higher dimensionality, and they require more information.

In V1, the extra information rate after the initial burst is apparently due to new features of the response (characteristic changes in firing frequency), which provide new, independent information. This is evident from the fact that the information carried by the spike count generally remains constant or even decreases little after the first initial rate period (Fig. 9, panels D and E). The extra principal components in the optimal representation are necessary to capture the change that occurs in the nature of the coding.

In contrast, the information carried in the spike count alone in IT neurons generally continues to rise throughout the response period. This finds a natural explanation in a model where the information is carried in a time-independent firing probability. This probability can be estimated better if the spike train is observed for a longer time: simple arguments give an uncertainty in frequency, AF/0.5, where O is the observation time, leading to information a log, which is at least qualitatively consistent with the information seen in most IT cells, including the ones in Fig. 10. However, the agreement does not mean that an identifying firing rate constitutes the code of IT neurons. For them, just as for our V1 cells, better fit and achieved more information is transmitted into the inter-spike intervals that include several principal components, indicating that temporal variations are part of the neural code. The fact that the information in inter-spike intervals employing principal components of the response were optimal does not imply that the cells which receive the response actually perform a principal component decomposition. However, the fact that we could not find any better sets does suggest that downstream neurons can gain more information if their processing can accommodate the features in the temporal structure of the signals that our analysis has identified as carrying information. The minimum temporal resolution we have identified for the initial burst period in V1 neurons matches well with typical cortical membrane time constants. We do not know, nor can we predict, whether the system achieves the integration over longer periods necessary to extract membrane information from later parts of the response. However, this is a problem not just for our code, but for any code based on firing rates averaged over more than 30 ms or so.

Another noteworthy feature of the best representation is that it includes the spike count. It might seem that the first principal component, which corresponds to the largest source of variance among the responses, would duplicate the count. Although the correlation coefficient between the two is low, the form of the firing rate of the first principal component does not resemble the spatial pattern of the second component. The second and principal components always do better than the principal components alone. This suggests that the count conveys information which is not conveyed by the first few principal components.

The importance of the spike count and firing rates averaged over times ranging from 25 ms up to a second suggests another simple hypothesis: that variations in information transmission are due to a direct correspondence of temporal changes in firing rate to high information transmission is achieved when and only when the firing rate averaged over all stimuli, is high. As was the case for other models as tested above, this is not a bad approximation, and it holds surprisingly well for a majority of our cells, but it fails for about a third of the cells (see Fig. 5).

We have also shown that for some cells there are two distinct latencies - one which measures the time until the firing rate rises as a result of stimulus presentation, and the other which measures the time until different stimuli can be distinguished on the basis of the cell's response. For some cells, these two latencies are simultaneous. For others, they are separated by over 10 ms. This means that some cells in the visual system start firing when a new stimulus is presented, but this change in firing rate provides no information about what the stimulus was.

Our study leaves a number of important questions unanswered. A particularly interesting case is just what patterns or features in those patterns can be discriminated on the basis of a neuron's response. In recent work, we have addressed this question with respect to the entire 250 ms response. However, our findings here about the detailed temporal course of the information transmission raise the question of whether information transmission about different kinds of spatial pattern features (for example, low and high spatial frequencies) follows different time courses. Exploration of this aspect would extend our knowledge of the neural code significantly.

Acknowledgments

We would like to thank Dr. Richard Kozanev of the Breakdown University Division of Applied Sciences for his support and many insightful discussions. We would also like to thank Dr. Morris Winkles for his support of our work and for helpful comments on the manuscript.

References

6. Eberlese, N. and Flap, R. (1995) Experience and sup-
mental application of information theory to the effec-
tive interpretation of the data. 1. Experimental results.
Data Qual. Smart Inf., 5, 9-17.
7. Eberline, D., Richmond, B. and Optican, L. M.
(1996) Rate of change temporal pattern in visual
formation in the absence of encoded memory in the
primary visual cortex. Undergraduate thesis, Depart-
ment of Applied Sciences, Harvard University.
9. Hols, J., Kiger, T. W., Herr, J. A. and Richmond,
B.F. Dynamics of information transmission by single
wood, R. J. (1996) Measuring neural network procen-
ding with artificial neural networks. 1. A Journal Ap-
plied Neurosci. 9, 35-56.
Neurosci. 2, 129-144.
of temporal correlates of the visual system. Proceed.
of nonconscious processing during a short-term memory task.
J. Neurosci. 8, 1000-1012.
ning of two-dimensional patterns by single units
in primary visual temporal cortex. I. Information.
15. Richmond, B. J., Optican, L. M. and Flap, R. J.
(1995) Temporal coding of two-dimensional patterns by
single units in primate visual cortex. I. Information.
16. Richmond, B. J., Optican, L. M. and Flap, R. J.
(1996) Temporal coding of two-dimensional patterns by
single units in primate visual temporal cortex. II. Quantifica-
17. Richmond, B. J., Optican, L. M. and Flap, R. J.
(1997) Temporal coding of two-dimensional patterns by
single units in primate visual cortex. I. Quantifica-
coding of two-dimensional patterns by single units
in primate visual temporal cortex. II. Quantification of
coding of two-dimensional patterns by single units
in primate visual temporal cortex. II. Quantification of
coding of two-dimensional patterns by single units
in primate visual temporal cortex. II. Quantification of
coding of two-dimensional patterns by single units
in primate visual temporal cortex. II. Quantification of
coding of two-dimensional patterns by single units
in primate visual temporal cortex. II. Quantification of
coding of two-dimensional patterns by single units
in primate visual temporal cortex. II. Quantification of
coding of two-dimensional patterns by single units
in primate visual temporal cortex. II. Quantification of