Dynamically Multilayered Visual System of the Multifractal Fly

M. S. Baptista,[1] Celso Grebogi,[2] and Roland Köberle[3]

[1]*Institut für Physik Am Neuen Palais 10, Universität Potsdam, D-14469 Potsdam, Germany*
[2]*Institute for Complex Systems, King's College, University of Aberdeen, Aberdeen AB24 3UE, United Kingdom*
[3]*Instituto de Física, Universidade de São Paulo, CP 369, 13560-970 São Carlos, SP, Brasil*

(Received 30 October 2005; published 27 October 2006)

We dynamically analyze our experimental results on the motion sensitive spiking $H1$ neuron of the fly's visual system. We find that the fly uses an alphabet composed of a few letters to encode the information contained in the stimulus. The alphabet dynamics is multifractal both with and without stimulus, though the multifractality increases with the stimulus entropy. This is in sharp contrast to models generating independent spike intervals, whose dynamics is monofractal.

DOI: 10.1103/PhysRevLett.97.178102 PACS numbers: 87.19.−j, 05.45.−a, 89.70.+c

The conflicting demands of variability and reliability require neurons of, e.g., the sensory system to judiciously adapt their dynamics to the statistics of the stimulus [1]. If a spiking neuron has to encode *relevant* features of the stimulus, it has to fire precisely timed spikes depending on the presynaptic input generated by the instantaneous stimuli the organism receives. In this Letter, we present experimental results on the fly visual system that strongly suggest that this encoding takes place on *multilayered sets*, a characteristic of the complex nature of this system. These sets are defined in terms of symbolic sequences of letters, selected from an alphabet according to the size of spike-time intervals. Furthermore, we present strong evidence that the underlying dynamics (UD) on each layer is multifractal, suggesting thus the possibility for a chaoticlike type of encoding. Already in the nonstimulus regime, because of the multifractal dynamics, the UD is highly flexible, ready to adapt its dynamics to the statistical properties of the stimulus to be encoded. Then, in the presence of the stimulus, finely tuned spike times ride on a set whose UD has now an increased multifractality, shaped by the properties of the stimulus.

Multifractal analyses of neural activity have been performed by some authors [2,3], using real brain data: Bershadsky *et al.* [2] use multifractality to analyze long term correlations in a particular region of anesthetized rats, whereas Zheng *et al.* [3] envisage its use as a neurosurgical tool. Silchenko and Hu [4] study the effect of stochastic resonance in an artificial noisy bistable system. By contrast, we analyze neural sensory data by combining elements of information theory and methods of the theory of *dynamical systems*. We use orders of magnitude more data than could be obtained from mammalian brain studies to be able to reveal the existence of a highly nontrivial dynamics.

We establish the underlying dynamical behavior of the neural activity in a prominent example of a spiking neuron: the $H1$ neuron. This neuron is located in the lobula plate of the fly Chrysomya megacephala, and is mainly sensitive to image motion associated with horizontal back-to-front rotations around a vertical axis [5]. This neuron was stimulated by a computer-controlled random, vertical bar pattern with horizontal velocity $v(t)$, new frames being shown every 2 ms [6]. In order to be minimally representative, we selected four types of stimuli $v(t)$: S_0 = no stimulus, S_1 = constant velocity, S_2 = a stimulus generated by an Ornstein-Uhlenbeck process with correlation time τ_c = 20 ms [7], and S_3 = an uncorrelated Gaussian stimulus.

Such a spiking neuron generates a sequence of spike times t_i, $i = 1, 2, 3, \ldots N_s$, where $N_s \sim 10^5$ in our experiments. All the information received is compressed into the sequence of intervals $\Delta t_i = t_{i+1} - t_i$. In order to extract the UD, we classify all the possible intervals, depending on their size, into a discrete set of cardinality N: $\Delta t_i \leq d_1$, $d_1 < \Delta t_i \leq d_2$, $d_2 < \Delta t_i \leq d_3$ etc., where d_j, $j = 1, 2, \ldots, N-1$, are a set of *dividers*, each index j generating one layer. Evidently, if we make this set large enough, we recover the original intervals. The question is this: Can we choose a reasonably small set of layers, without compromising the information contained in the original intervals Δt_i and study their dynamics? In other words, what is the size of the alphabet the fly's $H1$ neuron uses to speak postsynaptically?

We choose our dividers so as to minimize information loss or maximize Shannon's entropy [8]. For a given set of $N-1$ dividers, we convert the sequence of spike intervals Δt_i into a sequence of words of length L composed of N letters. Notice that an L-letter word may comprise a very long time interval. Now count up all the different L-letter words showing up in an experiment, get their probabilities P_k, and compute the average entropy (per letter) of a word of length L with N possible letters of the experimental sequence $H(L, N) = -\frac{1}{L}\sum_k P_k \log_2(P_k)$.

Figure 1(a) shows the entropy for $L = 10$ with only one divider, for the four different data sets. The entropy shows a maximum at $[d_1(S_0), d_1(S_1), d_1(S_2), d_1(S_3)] = [23, 6, 5, 5]$ bins. We now construct a uniquely defined *generating* [9] partition of our time intervals; i.e., our alphabet has to be able to allow a one-to-one encoding of all sequences of intervals. This will yield the partition with $N = 2 \times 2^l = 2, 4, 8, \ldots$. In order to be generating, all the

 © 2006 The American Physical Society

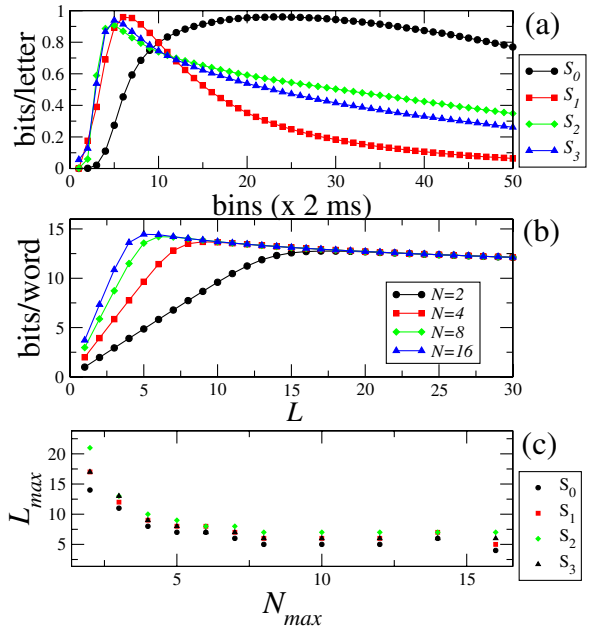

FIG. 1 (color online). (a) Entropy vs divider for a two-letter code, for the four data sets: S_0, S_1, S_2, S_3. The entropy maximum gives the best divider d_1. (b) Word entropy ($H_{\max}(L, N) \times L$) vs the length of words L for the data S_1. Note that for each N there is a value of L for which the entropy reaches its maximum. We refer to these N and L values as $N_{\max}$ and $L_{\max}$. Each curve corresponds to a different number of letters. (c) Scaling law that relates $N_{\max}$ with $L_{\max}$ for which the maximum of the entropy is achieved, for all the data sets.

dividers of a particular layer J must already be contained in the previous layers $j = 1, 2, \ldots, J-1$. With this *consistency requirement* in mind, we construct Table I in two steps. First, maximizing the entropy $H_2(N = 2 \times 2^l)$ for $N = 2, 4, 8, \ldots$, we get the respective dividers in Table I. Second, we turn to the remaining values of N, searching for dividers satisfying our consistency condition. This is indeed possible, if we fix, e.g., the bold faced dividers and perform a constrained maximization to obtain the remaining ones. We obtain the entropy per letter $H_2(N = 3, 5, 6, 7, \ldots)$, which equals to within 5% the entropy obtained with an unconstrained maximization, which yields completely different dividers. We conclude that $H_2(N)$, for $N = 2 \times 2^l = 2, 4, 8, \ldots$, is a generating partition. Is it unique? If we perform the same procedure for other partitions, e.g., $N = 3 \times 2^l = 3, 6, 12, \ldots$, the first step yields the same value for the unconstrained entropy maxima $H_3(N = 3, 6, 12, \ldots)$. In the second step, to obtain the remaining dividers, the constrained maximizations yield smaller entropies $H_3(N = 2, 4, 5, 7, 8, \ldots)$ by as much as 10% as compared to the unconstrained ones, as we show in Fig. 2. Here, $H_3(N = 2)$ means the following: (i) maximize $H_3(N = 3)$ to get $[d_0, d_1]$; (ii) maximize $H_3(N = 2)$, using either d_0 or d_1, whichever gives the larger entropy. The only exception here is one point corresponding to $N = 3$. The same is also true for other partitions, starting with prime numbers 5 and 7, indicating thus that within errors the binary partition is the only generating one.

Figure 1(c) shows that the number of letters needed to maximize the entropy decreases with increasing word length L, due to correlations manifesting themselves and undersampling effects. Since these put a limit on the usable size of the alphabet, we address the UD using just a few letters. We notice, however, that even for N smaller than $N_{\max}$, e.g., seven dividers for experiments S_2 and S_3, we obtain essentially the information the original spike trains convey about the stimulus [8]. Figure 3 shows the information per letter I_A for S_2 [10], divided by the information the original spike trains conveys about the stimulus, using the dividers of Table I. That is, for $N = 2, 3, 4, 5, \ldots$, we use the dividers [5], [3, 5], [3, 5, 11], [3, 4, 5, 11], ...

Since we are going to extrapolate below to large L, we focus on sequences of words of length L containing four letters, using the best dividers of Table I. Only fine details of our results depend on this choice.

We now encode each of these L-letter words in a one-to-one map into a real number W_i, $0 \le W_i \le 1$, whose frequencies we count in order to compute their probabilities p_i. The structure of the space of sequences $\mathcal{C}(W)$ for a given N may now be uncovered by computing the generalized dimensions D_q [11,12]. These are logarithmic ratios between the probabilities p_i and their physical occupation ϵ, which in our space $\mathcal{C}$ is give by $\epsilon = N^{-L}$. The index q can be thought of as a filter: a larger q enhances this ratio

TABLE I. The best time dividers (rows 3, 4, etc.) in units of bins, which generate the entropies H_2 for all data sets, considering different number of letters N (second row). Since $H(L = 1, N) \ge H(L > 1, N)$, we calculate these dividers fixing $L = 1$.

	S_0								S_1							S_2								S_3						
N	2	3	4	5	6	7	8		2	3	4	5	6	7	8	2	3	4	5	6	7	8		2	3	4	5	6	7	8
d_1	23	9	9	**9**	**9**	**9**	6		6	**6**	4	**4**	**4**	**4**	3	5	3	3	**3**	**3**	**3**	2		5	**5**	4	3	**3**	**3**	3
d_2		**23**	23	**23**	**23**	14	9			10	6	**6**	**6**	5	4		**5**	5	4	**4**	**4**	3			9	5	**4**	**4**	**4**	4
d_3			49	**49**	34	**23**	14				10	**10**	8	**6**	5			11	**5**	**5**	**5**	4				9	**5**	**5**	**5**	5
d_4				71	**49**	**34**	23					15	**10**	**8**	6				**11**	**11**	7	5						**9**	**9**	6	6
d_5					**71**	**49**	34						**15**	**10**	8					34	**11**	7							21	**9**	7
d_6						**71**	49							**15**	10						**34**	11								**21**	9
d_7							71								15							34									21

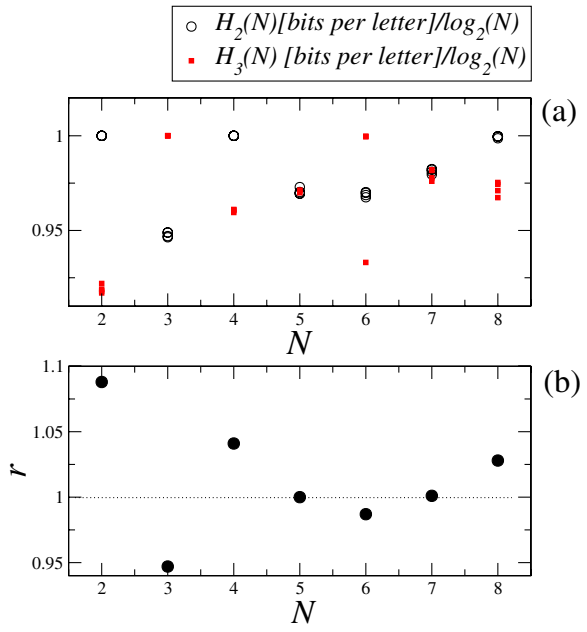

FIG. 2 (color online). (a) Entropy per letter normalized by $\log_2(N)$ for all data sets. Circles represent the entropy $H_2(N)/\log_2(N)$ and squares the entropy $H_3(N)/\log_2(N)$ for all data sets. (b) Ratio of $r = \langle H_2\rangle/\langle H_3\rangle$, where $\langle\cdots\rangle$ indicates the average over $S_0, \ldots, S_3$.

for large probabilities, whereas a negative one emphasizes the smaller probabilities:

$$D_q = \lim_{L\to\infty}\left\{\left[\ln\left(\sum_i p_i^q\right)\right]\Big/[(1-q)L\ln(N)]\right\}. \quad (1)$$

An equivalent quantity is f_α, the Legendre transform of $(1-q)D_q$. The index α measures the possible local fractal dimensions [13] in our space $\mathcal{C}$, occurring with singularity strength f_α. This is the global dimension of the set of points that locally scales with singularity strength α. They are given by $\alpha = \frac{d[(q-1)D_q]}{dq}$ and $f_\alpha = q\alpha - (q-1)D_q$ [14].

In Figs. 4(a) and 4(b) we show the spectra of the symbolic sequences in $\mathcal{C}(W)$ with their error [14], which allow us to draw the following conclusions. The $H1$ neuron has a multifractal character, exhibiting the existence of an infinite number of dimensions α with densities f_α. This is in sharp contrast to a memoryless, uncorrelated spike train with a Poissonian or similar probability distribution. In fact, any distribution with independent increments yields a nontrivial f_α for suboptimal dividers. Yet, since optimal dividers can always be chosen to yield a uniform distribution, it is really monofractal for the optimal ones with $D_q = \log_2(N)$. For example, for a Poissonian distribution with $N = 2$, we set $e^{-d_1(\text{Poisson})\lambda} = 1/2$, where λ is the spike rate to obtain $p_0 = p_1 = 1/2$. Real spike trains by contrast are multifractal even for the best dividers for all types of stimuli. The fractal dimensions are roughly the same for all the data sets, since $D_0 = f_1 = 1$. This means that the neuron's dynamics has continuous support on $\mathcal{C}$, the probability measure being distributed without "holes." The spectra's shape—$f_{\alpha_{\min}} = 0$ and $f_{\alpha_{\max}} > 0$ (except for S_3)—indicates that $\mathcal{C}$ has at least two scales and two main

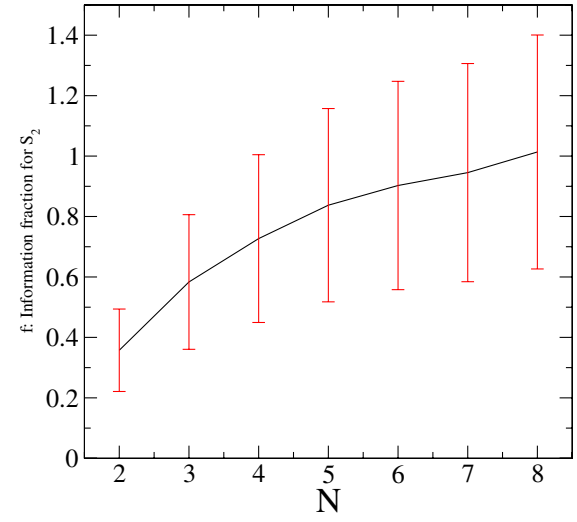

FIG. 3 (color online). Information per letter I_A the alphabet conveys about the stimulus for S_2, divided by the average information per spike I_S: $f = I_A/I_S$. The computation of the noise entropy H_{Noise}, where $I_A = H - H_{\text{Noise}}$, introduces the large error bars. Our data are extrapolations to $L \to \infty$ [8].

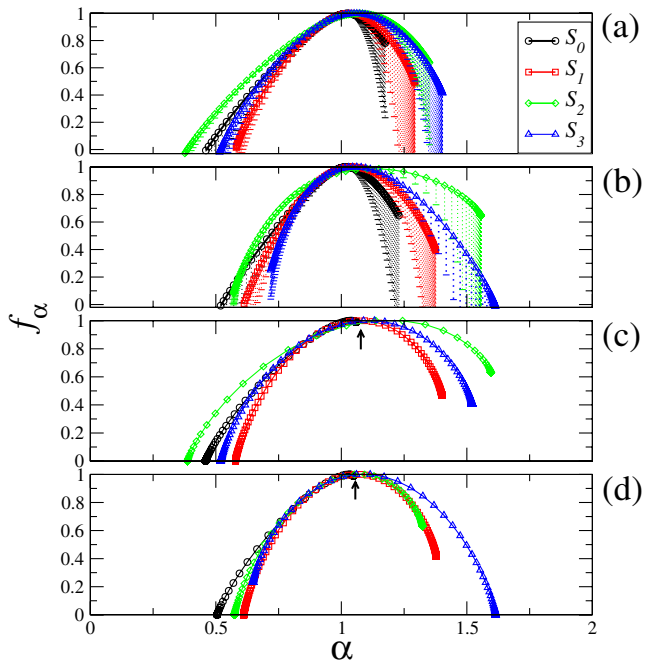

FIG. 4 (color online). f_α spectra using Table I: (a) two letters and (b) four letters. To avoid cluttering, we show only one-sided error bars as dashed lines. Spectra analogous to (a) and (b), using Eq. (2): (c) for two letters with $b_{S_0;S_1;S_2;S_3} = [30; 1.55; 2.0; 1.44]$ and (d) for four letters with $b_{S_0;S_1;S_2;S_3} = [30; 1.45; 2.0; 1.2]$. The arrows indicate the end point $\alpha_{\max}$ of the curve for S_0.

components: high probability sets—hot spots, localized in small portions of the symbolic space $\mathcal{C}$ with density $f_{\alpha_{\min}}$, and another low probability set—cold spots, spread out all over $\mathcal{C}$ with density $f_{\alpha_{\max}}$. In the set associated with S_0, the number of cold spots very much dominates—$f_{\alpha_{\max}} \sim \alpha_{\max} \sim 1$—the hot ones, implying one dominant scale. For the other data sets, the scales are comparable. Therefore, as the dynamics of the stimulus becomes faster as we go from $\tau_c = \infty$ to $\tau_c = 0$, the suppressed scale emerges.

Given the f_α spectra of our data, what is the simplest set with this spectrum? To address this question, we construct a probabilistic two-scale set with the following rule: an interval of length unity is divided into $b + 1$ equal pieces, such that one piece receives p_0 of the original measure and the remaining b receives bp_1, with $p_0 + bp_1 = 1$. Iterating this process self-similarly yields a set of dimensions:

$$D_q = [\ln(p_0^q + bp_1^q)]/[(1 - q)\ln(1 + b)], \qquad (2)$$

with $p_0 = (1 + b)^{-\alpha_{\min}}$ and $p_1 = (1 - p_0)/b$, where b is adjusted to produce the correct value for $f_{\alpha_{\max}}$ in Figs. 4(a) and 4(b). The resulting f_α spectra are shown in Figs. 4(c) and 4(d). In Fig. 4(d), to reproduce the spectra for S_3 in Fig. 4(b), we make $p_0 = (1 + b)^{-\alpha_{\max}}$.

Notice that b jumps to lower values once a stimulus is turned on. This means that for no stimulus the dynamics distributes the measure rather uniformly into 31 equal intervals, where 30 of them receive p_1 of the measure and 1 receives p_0 with $p_0 > p_1$. At the rth iteration of the set, 30 intervals will contain p_1^r of the measure (cold spots), 1 interval will contain p_0^r of the measure (hot spots), i.e., virtually all of the measure, and the rest will receive combinations proportional to $p_0^{r-k}p_1^k$, $0 \le k \le r$. When stimulated, the number of pieces drops dramatically to ~ 1, the hot spots approximately balancing the cold ones. The stimuli thus reshape the probability landscape, which becomes more and more structured as the stimulus entropy increases.

In conclusion, the *underlying dynamics* for the $H1$ neuron can be extracted by a finite-sized alphabet with about four letters. Analyzing sequences written in this alphabet allows us to exhibit the multifractal character of the sequence space. Here the stimulus shapes the landscape from mainly uniform to highly structured as the stimuli become increasingly dynamically variable. But this reshaping is played out on different layers, which are dynamically linked—a behavior typical of complex systems. The tools developed here for the analysis of our data are of general applicability. They reveal a fascinating complexity in the dynamics of the fly's visual system, even in the basic aspects of the spike generating dynamics, and they provide a means for their manipulation and control. It remains to be seen whether the layered structure uncovered here has a counterpart in the fly's optical information processing system.

The authors acknowledge FAPESP (M. S. B., R. K., and C. G.) and the A. von Humboldt Foundation (M. S. B.) for partial support.

[1] A. L. Fairhall, G. D. Lewen, W. Bialek, and R. R. de Ruyter van Steveninck, Nature (London) **412**, 787 (2001).
[2] A. Bershadsky, E. Dremencov, D. Fukayama, and G. Yadid, Europhys. Lett. **58**, 306 (2002).
[3] Y. Zheng, J. Gao, J. C. Sanchez, J. C. Principe, and M. S. Okun, Phys. Lett. A **344**, 253 (2005).
[4] A. Silchenko and C. K. Hu, Phys. Rev. E **63**, 041105 (2001).
[5] K. Hausen, in *Photoreception and Vision in Invertebrates*, edited by M. A. Ali (Plenum, New York, 1984), pp. 523–559.
[6] The experiments were performed in *DipteraLab*, Instituto de Física, São Carlos, Brasil. Flies, immobilized with wax, viewed a Tektronix 608 monitor from a distance of 12 cm, the setup being similar to that of Ref. [1]. The light intensity corresponds roughly to that seen by a fly at dusk.
[7] The fly's reaction time is of the order of 20 ms and is thus a behaviorally relevant time scale. Here we discretize time in bins of 2 ms, which is roughly the refractory period of the $H1$ neuron.
[8] The maximization of Shannon's mutual information yields identical results. For details on how to extract the Shannon information and the *noise entropy* from spike trains, see, e.g., S. P. Strong, R. Köberle, R. R. de Ruyter van Steveninck, and W. Bialek, Phys. Rev. Lett. **80**, 197 (1998).
[9] J. Plumecoq and M. Lefranc, Physica (Amsterdam) **144D**, 231 (2000).
[10] S_3 gives virtually identical results.
[11] A. Renyi, *Probability Theory* (North-Holland, Amsterdam, 1971); P. Grassberger and I. Procaccia, Phys. Rev. Lett. **50**, 346 (1983); H. G. E. Hentschel and I. Procaccia, Physica (Amsterdam) **8D**, 435 (1983).
[12] Note that since no metric characterization of the space $\mathcal{C}(W)$ is performed, it is irrelevant the way we encode words into real numbers to calculate the generalized dimensions.
[13] P. Grassberger and I. Procaccia, Physica (Amsterdam) **9D**, 189 (1983); T. C. Halsey, M. H. Jensen, L. P. Kadanoff, I. Procaccia, and B. I. Shraiman, Phys. Rev. A **33**, 1141 (1986).
[14] D_q can be reliably computed, if there are linear regimes, when plotting the numerator vs the denominator of Eq. (1). The q values were selected so as to produce D_q values with a variance of no larger than 5% relative to the absolute value (this is obtained from the residual of the linear fitting), which produces an error in $f_\alpha \ge 0$ of no more than 10% relative to the absolute value, except for some low probability sets afflicted with larger errors.