# Entropy measures of signal in the presence of noise: evidence for 'byte' versus 'bit' processing in the nervous system

W.R. Klemm and C.J. Sherry

*Department of Veterinary Anatomy and Department of Biology, Texas A & M University, College Station (Texas 77843, USA), 5 February 1980*

*Summary.* A method for detecting signal in the presence of noise in a highly specific way is described. Using action potential interval data from 12 neurons in rat cerebellum, we have demonstrated that the sequential ordering of spike intervals contains both noise and signal. We have identified and quantified the magnitude of relative entropy (uncertainty) for specified sets of interval patterns, ranging in length from 3–5 successive intervals. Some of these sets have exceptionally low entropy and thus seem to be especially meaningful as a set ('word') to the brain.

Entropy, usually expressed as the number of binary digits or bits needed to encode a symbol or higher order group of symbols uniquely, is a measure of the weighted average information (uncertainity) involved with individual alternative events[1]. We have recently demonstrated that entropy, calculated on the basis of our technique which describes and preserves the serial order of sets of 3–5 adjacent intervals[2], is highly correlated with the variability of an underlying biological signal (neuronal action potential intervals). These entropy values correlated with the median interval duration, the 15–85% range of interval durations, and the skewness of the underlying probability density function of interval durations[3]. Measures of the variability of entropy also correlated with the number of interval patterns whose probability of occurrence changed in a statistically significant manner in response to experimental manipulations (i.e. ethanol injections)[4].

These experiences with the entropy of serially ordered spike train interval patterns revealed to us the desirability of comparing entropy ('H' values) for interval patterns of different lengths (2–5 intervals). However, such values cannot be compared directly unless they are transformed onto the same numerical scale. To accomplish this, we expressed entropy, in relative terms, as the percentage of the theoretical maximum possible entropy for a given set of adjacent intervals.

In this present analysis, we have been able to show that entropy does not always vary predictably with the number or duration of intervals in given pattern sets. This in turn indicates that certain pattern sets are more meaningful to the system (i.e. have less 'uncertainty'). Thus, this approach seems to allow one to distinguish the existence of signal in the presence of noise. Moreover, such analysis may provide an essential first step in elucidating the manner in which information is processed in the brain.

*Recording methods.* The recording methodology has been described in detail elsewhere[5]. Briefly, extracellular action potentials were recorded from the cerebellar cortex of paralyzed and artificially respired rats via tungsten microelectrodes. The action potentials were superimposed on a storage oscilloscope to demonstrate that they were obtained from a single neuron. These potentials were then converted to standard pulses and tape recorded. Data from 12 neurons are used in the present analysis.

*Data reduction methods.* The basic reduction techniques (see figure 1) have been described in detail[2]. Briefly, intervals between adjacent action potentials were sequentially stored in a computer memory; intervals were measured to the nearest 0.1 msec and rounded to the nearest msec. The 1st interval is compared to the next sequential interval and a $-$, 0, or $+$ is written in the computer memory if the 1st interval is smaller than, equal to, or larger than, the 2nd interval, respectively. 2nd and 3rd intervals are compared similarly, etc. These symbols ($-$, 0, $+$) are stored sequentially in the memory and are then arranged into a series of transition matrices. For example, the digram (2-symbol) matrix specifies the probability that a $-$ is followed by a $-$, by a $+$, or by a 0, etc., as shown in figure 1. Note that this describes the sequential relationship of 3 interspike intervals. The matrices for trigrams (3 symbols) and quadgrams (4 symbols) are generated in the same manner and specify the sequential relationships of 4 and 5 intervals respectively.

*Information theory methods.* The entropy value ('H') for each single symbol and for each set of symbols was calculated following the procedure of Shannon[1]. Strictly speaking, the entropy of a given symbol or symbol set is a 'fractional' entropy, being a portion of the total entropy associated with all sets of symbols of the same length (diagrams, trigrams, etc.).

The total 'H' value for all symbol sets within the same group (digram, etc.) was calculated using the following formulas:

$$H_D = -\sum_{1}^{9} P_i \log_2 P_i$$

$$H_Q = -\sum_{1}^{81} P_{ijkl} \log_2 P_{ijkl}$$

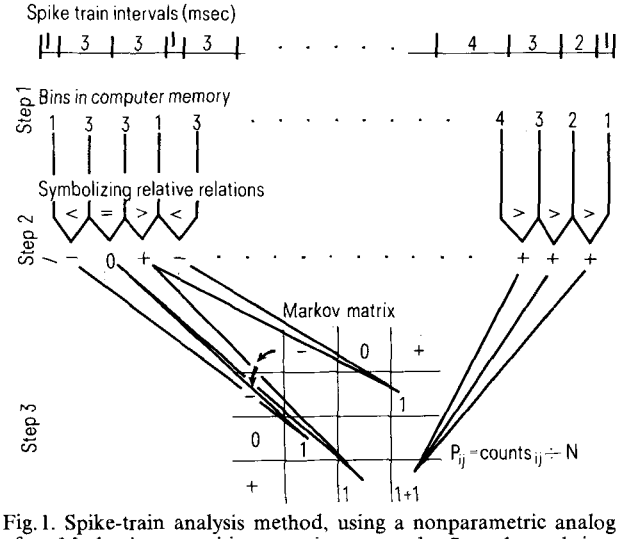



Fig. 1. Spike-train analysis method, using a nonparametric analog of a Markovian transition matrix approach. Step 1: real-time intervals are stored sequentially in the computer memory. Step 2: each interval is compared with successors in terms of being $<$, $=$, or $>$. The corresponding symbols ($-$, 0, $+$) are stored in the memory, with sequential order maintained. Step 3: the symbols in the memory are tallied into a matrix, counting in the simplest case, e.g., how many times a given symbol is followed by another ($-$ by $-$, $-$ by 0, etc.). For the 3rd order (trigram) matrix the computer counts, e.g., how many times a given pair of symbols is followed by another $--$ by $-$, $--$ by 0, etc.); for the 4th order (quadgram) matrix the computer counts how many times a given triplet of symbols is followed by another ($---$ by $-$, $---$ by 0, etc.). For all possible patterns (9 digrams, 27 trigrams, 81 quadgrams), the incidence probabilities are calculated.

where the subscripts have the same meaning as in the formula below.
The formulas for determing the 'H' value for individual symbols and symbol sets are:

$$H_i = -P_i \log_2 P_i$$
$$H_{ij} = -P_{ij} \log_2 P_{ij}$$
$$H_{ijk} = -P_{ijk} \log_2 P_{ijk}$$
$$H_{ijkl} = -P_{ijkl} \log_2 P_{ijkl}$$

where the probability of occurrence of each symbol or symbol set is $P_i \ldots P_{ijkl}$.
The basic idea of subsequent operations is to calculate the theoretical maximum total entropy for each group of symbol sets (digrams, etc.). Then we obtain a theoretical average for each individual symbol set. These average values thus serve as a frame of reference for evaluating experimentally observed entropy values, which can be expressed for a given set of symbols in terms of the fraction of entropy (percent maximum entropy) that is actually present in a symbol set compared with the average amount that is theoretically possible. Also, by summing and averaging the percent maximum entropy of these individual symbol sets, we obtain an expression of the percent maximum entropy for an entire symbol group. This scheme permits one to compare individual symbol-set values and group values on the same relative scale, even though the original data are derived from clusters of symbols that differ in number.
Entropy is maximum with large absolute values of H (-sign disregarded). With maximum entropy, the probability of occurrence of all events (i.e., symbols or sets of symbols) is equal; in other words, with such equal probability there is a maximum uncertainty or disorder. Calculation of the theoretical maximum entropy for digrams, trigrams, and quadgrams is as follows:

$$H\,max_D = \log_2 N_D$$
$$\vdots$$
$$H\,max_Q = \log_2 N_Q$$

where $N_D$ is the number of digrams (9) and $H\,max_D$ is the theoretical maximum entropy for digrams, etc.
The average amount of maximum entropy for each defined set of symbols in a digram, etc., is calculated simply as:

$$\bar{X}\,H\,max_{ij} = H\,max_D/N_D$$
$$\vdots$$
$$\bar{X}\,H\,max_{ijkl} = H\,max_Q/N_Q$$

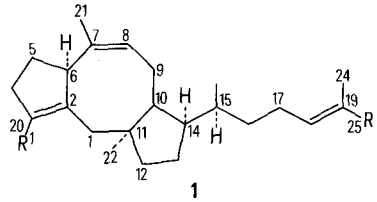
Note, that for the theoretical case $\bar{X}$ H max has the same value for each symbol set in a digram, etc.
With real data, one can calculate percent maximum entropy for each observed specific set of symbols as follows:

$$\%\,H\,max_{ij} = (H_{ij} - \bar{X}\,H\,max_{ij})/\bar{X}\,H\,max_{ij}$$
$$\vdots$$
$$\%\,H\,max_{ijkl} = (H_{ijkl} - \bar{X}\,H\,max_{ijkl})/\bar{X}\,H\,max_{ijkl})$$

where $H_{ij}$ is the amount of experimentally observed entropy associated with each specific diagram and $\%H\,max_{ij}$ is the percent maximum entropy for specific digrams. This calculation permits one to place entropy measures for digrams, trigrams, and quadgrams on the same relative scale, thus enabling direct, meaningful comparisons (see below).
Then for experimentally observed data, one can calculate an average percent maximum entropy for digrams, etc., as follows:

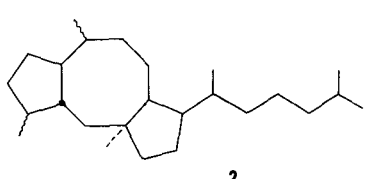
$$\bar{X}\,\%\,H\,max_D = \sum_1^{N_D} (\%\,H\,max_{ij})/N_D$$
$$\vdots$$
$$\bar{X}\,\%\,H\,max_Q = \sum_1^{N_Q} (\%\,H\,max_{ijkl})/N_Q$$

These calculations can be performed on absolute % H max $_{ij}$ values or can be expressed as the algebraic average by taking into account whether the observed values were greater or less than the theoretical average. Thus, this measure allows an overall evaluation of the gross amount of entropy present in a given signal.
*Maximum entropy for symbol groups.* The average entropy (observed or the theoretical maximum) per set of symbols in a group decreases as the string of symbols increases from digram to quadgram (table 1). This results from the fact

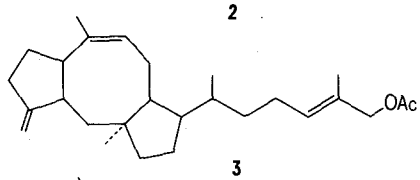


1

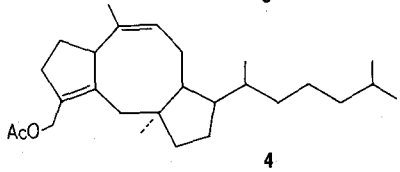
**a** R = $R_1$ = $CH_2$–OH
**b** R = $R_1$ = $CH_2$–OAc
**c** R = $CH_2$–OAc $R_1$ = $CH_2$–OH
**d** R = $CH_2$–OH $R_1$ = $CH_2$–OAc
**e** R = $R_1$ = CH = O
**f** R = $CH_2$–OH $R_1$ = CH = O
**g** R = $CH_2$–OAc $R_1$ = CH = O
**h** R = $CO_2Me$ $R_1$ = $CH_2$–OH



2




3




4

Table 1. Maximum entropy for symbol groups

| Symbol groups | Average entropy | | Total entropy | |
|---|---|---|---|---|
| | Observed* | Theoretical ($\bar{X}$ H $max_D$) | Observed* | Theoretical (H $max_D$) |
| Diagrams | 0.242 | 0.352 | 2.18 | 3.17 |
| Trigrams | 0.119 | 0.176 | 3.21 | 4.75 |
| Quadgrams | 0.052 | 0.078 | 4.21 | 6.32 |

* Averaged across the H values for 12 neurons.

that the total entropy is distributed over so many more different symbol sets (81) in the quadgrams than in the digrams (9). The important observation is that total entropy increases from digrams to quadgrams, reflecting the greater uncertainty involved in ordering 4 symbols than 2. It seems also pertinent that all symbol groups had less observed entropy (on the average, 32%) than was theoretically expected (table 1). This is a strong suggestion that ordering is present in the symbol groups and that the biological signal is not random and independent.

*Percent maximum entropy for symbol sets (figure 2).* These results clearly disclose the advantage of expressing all data on a similar percentage scale. Individual entropy values for the various symbol sets, averaged over these same 12 neurons have been calculated for a previous purpose[6], and when these are evaluated in terms of percent maximum entropy (figure 1), striking effects are noted. For example, there is little quantitative difference in the entropy for the digram $-+$ ($-0.525$), for the trigram $-+-$ ($-0.458$), and for the quadgrams $-+--$ and $-+-+$ ($-0.254$ and $-0.374$). But in terms of percent of maximum entropy, the differences are drastic: the $-+$ digram, at 49.1%, increases almost threefold with the trigram value being 160.2%; the increases are 5–7 fold for the quadgrams (225.6% and 379.4%).

The clustering of sets containing zeros around $-90\%$ is striking. Many of these sets have 2 or 3 zeros in the set. Recall that these are not theoretical values, where one might expect low entropy because so many of the same symbols are in the set; thus, their predominance of less entropy than the theoretical average maximum for the group (0%) could be provocative.

Many other observations of seemingly unique entropy distribution could be made. For example, the quadgrams $----$ and $++++$ have very low entropy, while both the digrams and trigrams which might be thought of as subsets ($--$, $---$, and $++$, $+++$) all had greater entropy; if they were subsets, the entropy should be less, because fewer symbols are required to complete the 'word'. Similar comments apply to the quadgram $---+$.

Some sets of symbols have distinctly different values, even though they should be the same if random processes were operating, because they contain the same number of identical symbols in the set (compare $--++$ and $++--$ with $-+-+$ and $+-+-$).

One also expects on theoretical grounds that disorder values should increase in the order of digrams, trigrams, and quadgrams. However, some quadgrams have less disorder than trigrams and digrams (compare $----$ and $++++$ with $+++$, $-+-$, and $+-+$, or with $++$ and $--$). Note that one of the least disordered symbol sets is a quadgram without zeros; $---+$; many quadgrams of similar values have only 1 zero.

The individual symbol sets, which seem to be the most and the least meaningful to the system, are clearly distinguished. Those with the least amount of disorder often contain 2 or more zeros, indicating that these have an unusual amount of importance to the system (note that the only other laboratory reporting analyses of this kind has chosen to ignore the zero case[7]).

We have not attempted to make any probabilistic inferences with conventional statistics but rather have focused on the development of a practical, descriptive approach for

Table 2. Average % maximum entropy (± SD) for symbol groups

| Symbol groups | Algebraic average* | Average of sets with same direction of change | |
|---|---|---|---|
| | | Less than 0% | More than 0% |
| Digrams | −31.3± 63 | 83.3± 7.9 (5)** | 33.7± 18.1 (4) |
| Trigrams | −32.4± 89 | 80.8± 21.2 (20) | 105.8± 52.5 (7) |
| Quadgrams | −12.9± 123 | 82.6± 18.3 (67) | 210.0± 101.7 (14) |

* Averaged over all digrams, etc. ** Number of symbol sets.

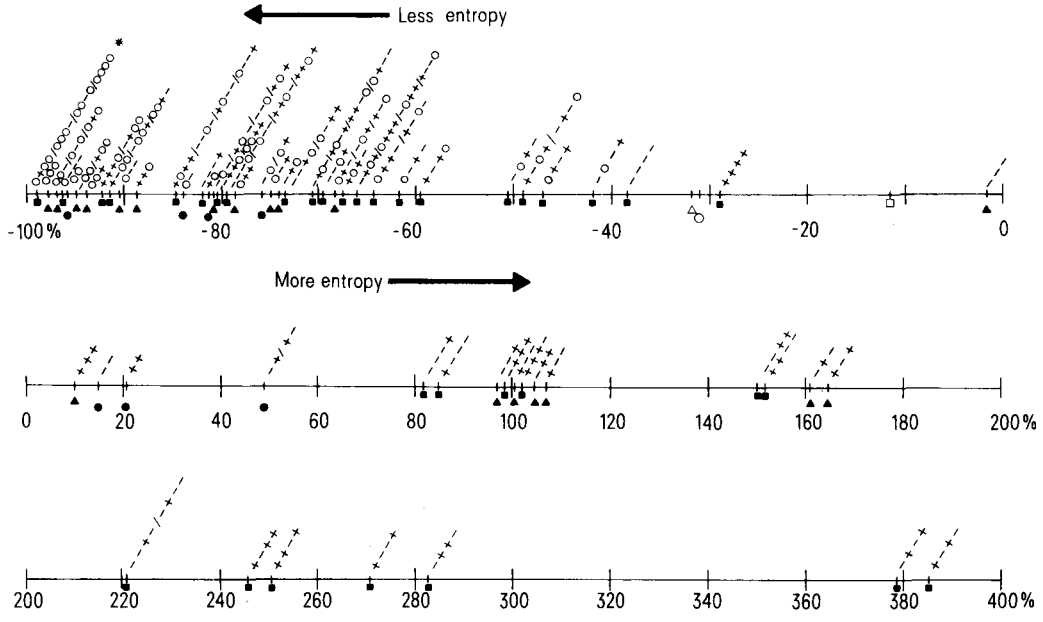



Fig. 2. Diagrammatic illustration of the interrelations among several kinds of average percent maximum entropy data: 1. the % maximum entropy for each set of symbols (solid symbols: digrams are circles, trigrams are diamonds, and quadgrams are squares, (% $H\,max_{ij}$) (not shown are 5 other symbol sets that contain 3 zeros and 10 sets that contain 2 zeros); and 2. the signed average percent maximum entropy for entire groups of symbol sets (digrams, etc.) (open symbols). At 0%, the entropy for a given symbol set is the same as the theoretical maximum average for its group (digram etc.).

identifying potentially meaningful interval-pattern 'bytes'. The physiological significance, i.e. the information carrying value, of interval-pattern bytes must be determined empirically by reference to some external event or parameter.

The apparent unpredictability of the entropy of specified symbol sets is a clear indication that certain of these sets have a special meaning for the system; such clusters of intervals may well be recognized in the system as a 'word' Of special note are the clusters $----$ and $++++$, because simple cases of this kind are readily subject to experimental verification.

The presence of large entropy values of certain symbol sets seems to document the widely held belief that there is much noise in a spike train. Nonetheless, there appears to be 'signal' in the presence of that noise, and our technique seems able to distinguish between the two in a highly specific way.

The main advantage of our relative entropy scheme is that it permits rational development of hypotheses. Out of the vast array of possibilities we are able to order symbol sets of different length and content in a rational way that permits systematic development of hypotheses for empirical testing. For example, we might postulate that symbol sets with the most relative entropy are less biologically meaningful than sets with the least entropy.

*Average % maximum entropy for symbol groups.* Consistent with the observations on average maximum entropy, the average % maximum entropy increases from digrams to quadgrams. However, when sign is taken into account (table 2), it is clear from the overall negative averages that many of the symbol sets in a group had substantially less entropy than the theoretical maximum. Note that the least indication of disorder (entropy) was in the quadgrams. While the large SDs preclude statistically significant differences among the groups, they do serve to emphasize the fact that there was a wide variation in the percentage of maximum entropy for individual symbol sets within a given group. It seems reasonable to suspect that the spike train is not random and independent, but in fact contains 'information'.

When the averages are tallied separately for those symbol sets that are below or above 0% (table 2), other significant facts become evident. For example, most of the symbol sets have markedly less entropy than the overall group average; note that the averages are about the same for digrams, trigrams and quadgrams and that the SDs in each case are distinctly smaller than those of symbol sets with positive entropy averages.

Thus, we believe that these results provide a basis for several rather profound conclusions about trains of nerve impulses: 1. that the train of intervals is not random and independent, 2. that the signal contains a mixture of noise and of 'information', and 3. that some groups of intervals seem to have a meaning for the system as a 'byte' of successive intervals.

1 C.E. Shannon, Bell Syst. tech. J. *27*, 379 (1948).
2 C.J. Sherry and T.J. Marczynski, Int. J. Neurosci. *5*, 171 (1973).
3 C.J. Sherry and W.R. Klemm, Int. J. Neurosci. *11*, 109 (1980).
4 C.J. Sherry and W.R. Klemm, Prog. Neuro-psychopharmac., in press (1981).
5 J.A. Mikeska and W.R. Klemm, Res. Comm. psych. psychiat. Behav. *4*, 457 (1979).
6 C.J. Sherry and W.R. Klemm, submitted.
7 S. Brudno and T.J. Marczynski, Brain Res. *125*, 65 (1977).

## Effect of menstrual stress on serum lipid levels

Shobha A. Choksi and A.S. Sane[1]

*Department of Biochemistry, B.J. Medical College, Ahmedabad (India), 28 January 1980*

*Summary.* The effect of menstrual stress on serum lipid levels has been investigated in 28 healthy unmarried student nurses aged between 19 and 25 years, with histories of regular 26-30 day menstrual cycles. There were definite and similar patterns of changes in serum cholesterol, phospholipids and triglycerides during different phases of the menstrual cycle.

It is now well recognized that many women become distressed in response to their menstrual cycle. In most cases symptoms appear to peak within 3-4 days prior to the onset of menstrual flow and often disappear when the flow starts[3]. This has led to the coinage of the term premenstrual tension[4]. The term 'paramenstrum' includes 4 days prior to the onset of menstruation and 4 days of menstruation. The other days in the cycle comprise the intermenstrum[5]. Results of the serum lipid studies performed during the menstrual cycle vary and are somewhat contradictory[6-10]. Our earlier work on preoperative stress and serum cholesterol levels led us to undertake the present investigation so as to understand the correlation, if any, between menstrual stress and serum lipids[11].

*Materials and methods.* 28 unmarried healthy female student nurses aged 19-25 years with histories of regular 26-30 day menstrual cycles were selected as subjects for menstrual stress investigation. They did not take any medication or hormone preparation for 3 months before the study. For each subject the serum total cholesterol, phospholipids and triglycerides were estimated on 5 ocassions during the menstrual cycle, viz. 1. 8-9 days after mid-cycle (this time has been designated as end-cycle), 2. the day of menstruation (i.e. 1st or 2nd day), 3. the day after menstruation has ceased (usually 5th or 6th day after starting) when estrogen and progesterone concentrations are generally low, 4. at mid-cycle (assessed as 14 days before the predicted date of menstruation) when estrogen concentrations are normally at their highest, and 5. at end-cycle when the progesterone peak is expected. On all occasions fasting blood samples were collected in the morning between 8.30 and 9.00 a.m. Serum total cholesterol, phospholipids and triglycerides were estimated according to the methods of Zak[12], Connerty et al[13], and Carlson[14], respectively. Means and standard errors were calculated and results were tested for statistical significance by means of a paired t-test.

*Results.* The results recorded in the tables clearly indicate a definite and similar pattern of changes in serum levels of total cholesterol, phospholipids and triglycerides on different days of the menstrual cycle. Though the fluctuations in serum lipid fractions were of lower magnitude, they were statistically significant. The salient features of the observed fluctuations in levels of serum total cholesterol, phospholipids and triglycerides were: relatively low values at end-cycle (premenstrual tension), an increase on the 1st or 2nd