Henry Ford Hospital Medical Journal

Volume 12 | Number 3 Article 6

9-1964

Some Experiences And Comments About Bio-medical Data Processing

W. R. McCrum

Recommended Citation

McCrum, W. R. (1964) "Some Experiences And Comments About Bio-medical Data Processing," *Henry Ford Hospital Medical Bulletin* : Vol. 12 : No. 3 , 333-341.
Available at: https://scholarlycommons.henryford.com/hfhmedjournal/vol12/iss3/6

Henry Ford Hosp. Med. Bull.
Vol. 12, September, 1964

SOME EXPERIENCES AND COMMENTS ABOUT BIO-MEDICAL DATA PROCESSING

W. R. McCrum, Ph.D.*

For the past year we have had a digital computer system in operation in our department and have been using it primarily to solve some problems in the analysis of the electroencephalograms of monkeys and humans. We have been asked to write our views about the use of such a system in medical and biological research in general. There have been many articles and editorials in the medical journals regarding this area. These have ranged from reports of a specific and limited application of a particular computer program to pure speculation about automating the practice of medicine. Our excuse for adding to this literature is our hope that by illustrating some general ideas about data processing with specific examples from our experience we can make more understandable the few examples of success, and the many examples of failure in the use of these "high speed electronic idiots" in bio-medical research.

The first point to stress in a discussion such as this is that the computer does not solve any problems; it merely provides certain arithmetical and logical operations on data that is fed to it. It can make no guarantee as to the validity of the problem definition, nor as to the interpretation of the results of the computations it performs. The only success in using computers in scientific problem solving has come about by employing a method called "systems analysis". Systems analysis is the design of a complete set of methodical operations, both physical and mental, which when properly and completely performed will provide a specific solution to a specific problem. This set of operations demands first of all a specific and limited definition of the problem to be solved. When this has been done the mathematical or statistical model can then be constructed that will permit the solution of this problem. The next step is defining the kinds of data that must be used and the methods by which it will be gathered and processed into numbers that can be understood by a computor. Following this is the design of a computer "system" that can provide the most efficient and economical operation upon the data. The final stage will be the formulation of the proper display of the output from the computer and the procedures by which these computed results will be fitted to the model in such a fashion that a proper solution to the specified problem will be obtained.

---

*Division of Neurology.

The most critical and the most difficult step is the first one, namely, that of the definition of the problem that is to be solved. By intuitive and associative mental processes the human mind quickly becomes aware of relationships that exist in his experience. In research, whether it be clinical or basic, the object is to place exact measurements on these relationships in so far as is possible. This is usually very difficult to do because, although these relationships do exist and we are aware of them, they are quite elusive to precise definition let alone measurement. Since the computer has no intuitive or associative processes but operates only in a determinate fashion, we must ourselves define precisely these relationships we wish to resolve.

Let us take a "for instance" with regards to the matter of problem definition. For some 30 years the recording of brain waves has been used with modest success in determining the presence or absence of organic brain lesions. It has been assumed, and I emphasize the word assumed, that this amount of information is only a small fraction of the total information carried by brain waves. This assumption carries a further assumption that the main source of these brain waves is in the neurophysiological processes that are involved with mentation. Accepting these assumptions, it was desired to know if there were changes in the brain waves that could be related to the changes in an animal's ability to perform a learned task under various environmental conditions. Intuitively such a relationship did exist, but we could not present this intuition to a computer. First of all, then, we had to define precisely what we meant by brain waves or more simply the EEG. The EEG is simply the continuous recording in time of the potential difference between a set of points on the surface of the head. In clinical electroencephalography the EEG is defined only as to its frequencies; that is, except for an intuitive recognition that extremes in amplitude do add somewhat to the interpretation of a record, the major portion of the information is carried in the frequency pattern and more importantly in the relation of the frequency patterns of various pairs of electrodes covering the entire head. In clinical EEG the frequency bandwidth is defined as existing between ½ cycle per second up to and including 70 cycles per second. This in turn is divided into 4 smaller bandwidths: ½ - 3 cps. called delta activity, 4 - 7 cps. or theta activity, 8 - 13 cps. or the alpha rhythm, and above 14 cps. or beta frequencies. This definition was historically determined by the limitations of the amplifiers and recording equipment available at the time, and has never been formally changed although it is now known that the frequency bandwidth extends up to several hundred cycles per second as well as down to extremely low frequencies. Despite its inadequacies, this clinical definition of the EEG must be accepted as a starting point both because of its simplicity and also for the lack of any real evidence that other parameters of the recording do carry significant information. Classical methods of waveform analysis (auto and cross-correlation, power spectrum determination and Fourier analysis) were first considered. All of these were considered unsuitable for our particular problem; they become unwieldly when the recording sample is longer than a few seconds, they require certain assumptions that in the case of brain waves are quite suspect, and they have great limitations on accuracy above a very narrow bandwidth. More important perhaps is the fact that to use them to analyze the entire recording would be very expensive, several thousand dollars. The entire record had to be analyzed, for although we assummed the information was in the frequency pattern, we made no assumptions as to when in time the information was present.

The EEG had to be defined in as simple a fashion as possible that would permit evaluating the pattern throughout the entire recording. We would then define a base line through the recording and would measure the length of time it took for each wave to cross this base line. We called this our "zero crossing". We had then a precise but arbitrary measurement of a frequency pattern of the EEG. We would record the EEG while the animal was at rest and also while he was perfoming different kinds of problems and would compare the measurement of the EEG pattern with these various behavioral conditions.

The word compare used in the last sentence is the word which must be defined precisely by the mathematical model. This mathematical or statistical model must be designed so it will measure the amount of relationship between the EEG and the periods of the test conditions.

Without a precise definition of the problem we could not construct a mathematical model, and without the mathematical model we could not talk to the computer; therefore, the construction of the mathematical model stands with equal importance alongside the definition of the problem. For a definition of the mathematical model let us quote Dr. Grove C. Nooney.[1] "Let us think of a mathematical model as a collection of mathematical sentences containing mathematical quantities, operations and relations together with definitions. This collection is to reflect, in some sense, the real situation." This last sentence needs further elaboration. We tend to forget that in biology and medicine our knowledge of the reality of the systems with which we work is extremely limited, and in the main we work with idealized systems that have been hypothesized from a statistical sampling of many small parts of a highly complex and integrated system. Let us take a simple example of some mathematical models that are used in medicine every day. A small sample of blood is drawn and cell counts are made along with certain chemical determinations. Results of these determinations in themselves give no information about the character of the total blood supply of the patient; however, when they are fitted to the stochastic (probabilistic) model that has been developed by making such determinations thousands of times one can predict within very narrow limits the values of any other sample of blood drawn from the same patient. Further than this, the values of this particular sample of blood can be fitted to another mathematical model that places the limits within which the various components can vary in disease and non-disease states. With continued and successful use of the system over the years it is forgotten that the usefulness of these determinations lies in the accuracy of the mathematical models. Not all models in medicine and biology are as successful. This is particularly true regarding models of the nervous system. For example let us look at the model of the dynamics of the nerve cell membrane proposed by Hodgkin and Huxley.

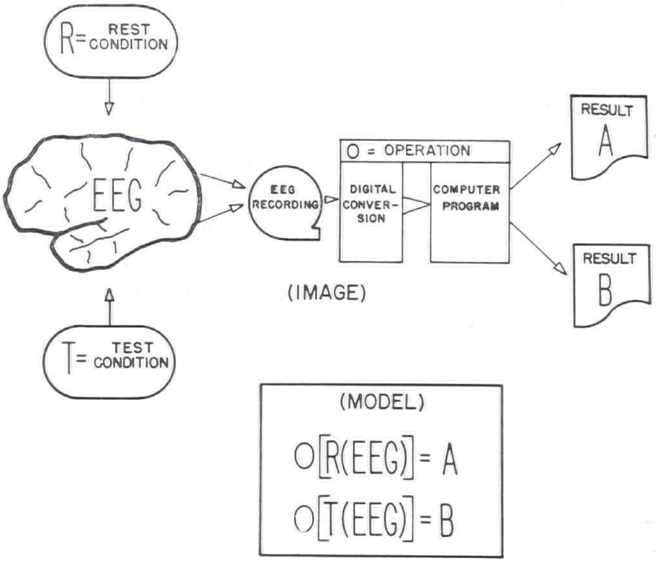
$$E_m = 58\text{mV} \cdot \log \frac{P_k K_i + P_{na} Na_i + P_{cl} C_o}{P_k K_o + P_{na} Na_o + P_{cl} Cl_i}$$

This model is an attempt to explain the electrophysiological mechanism of the membrane potential of a nerve cell. It is a good model as far as it goes but it is by no means a complete model of the real system. It does not admit to the effects of other ions such as calcium, copper and magnesium, nor does it locate the storage place of the extracellular ions (electron microscopists tell us there is no extracellular space in the nervous system.) This model can only be used to stimulate further research into the nature of the real system, but it gives no basis for predicting activity in either single cells or systems of nerve cells. A mathematical model then is a precise statement about a reflection of reality and is used to gain more knowledge about reality. The more simple it is the less useful it is. It can be so simple as to be useless; thus the more of the real situation it includes the better a model it is. In all cases, however, it must be consistent with what is known about the real situation. On the other hand the model can be quite complex and consistent with reality, but unprovable mathematically. If this is the case it is of no more use to us than our naive intuition.

Between reality and the mathematical model there is an area that Dr. Nooney calls "imagery". He states "the world of imagery contains quantities and their known, measured or conjectured relationships, possibly in the form of other kinds of models as graphs, tables or flow charts, . . . the mathematical model then is a translation of a part or all of imagery into formal mathematical terms." Until the advent of computers, research, particularly in the field of biology and medicine, pretty much ended at the construction of the image. The formulation of a mathematical model was rarely attempted. There are several reasons for this, two of the more obvious being that there was little communication between the biologist and the mathematician, and secondly even though the mathematical model could be formulated, the immense computation necessary to conclude this liaision was beyond human effort. With the advent of computers the chasm between the two disciplines is rapidly being bridged and the computations have become a simple matter.

Let us take a look at our EEG problem as an example of the development of the mathematical model. The upper part of Figure 1 shows the image of the system we have designed to analyze the EEG. Below this is the mathematical model that represents this system. As can be seen it is an extremely simple model and really provides little or no information. The "O" or operation we have defined very precisely with a rigidly controlled system for analogue to digital conversion of the recorded EEG and a rigid computor program to perform a precise operation on the converted data. The "R" stands for the rest condition. It is composed of many variables which are not accounted for individually but are lumped together as a single factor. Likewise, the "T" of the test condition represents many variations in the test

FIGURE 1



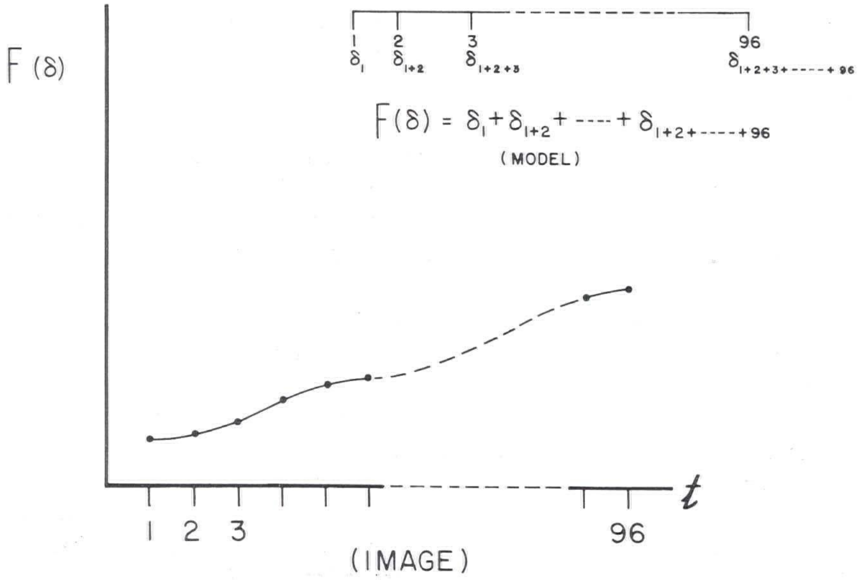


FIGURE 2

condition that are considered as a single factor. The EEG also has many parameters but again these are treated as a single unit. Our only information from this model comes when we test A against B. If it is established that A does not equal B, then we can proceed to develop other models that will provide more detailed information as to how they differ.

Figure 2 shows the next step in evaluating the information content of the EEG. It includes the graph (or image) of a function of slow-wave activity that we have arbitrarily defined. We repeat such images and models for the other bandwidths we wish to study, i. e. theta, alpha and beta. Figure 3 shows the graph and model of the derivative of this function we have defined, in this case the theta activity. That is it shows the rate of change from alternate periods of rest and performance conditions. Although this is an idealized picture we have attempted to show that it has been found that the derivitive is larger when performance is low. Although such images would seem to show rather clearly a relationship between performance and the EEG activity, establishing the mathematical truth of this relationship can be quite difficult. Figure 4 shows one method by which this EEG-Behavior relationship could be given some mathematical certainty.

The gathering and processing of the data should be one of the simplest of procedures, but all too often it poses some very difficult problems. The computer is never overwhelmed by the quantity of data that is given to it, but is most often stymied by the smallest of inconsistencies in the data. Uniformity of gathering and processing data cannot be over stressed.

Too often the problem is defined after the data has been collected. The need to make it fit a format suitable for the computer may necessitate a considerable loss of information.

More often problems arise in the collection of data by instrumentation. When an instrument (electronic or mechanical) is used between the data source and the computer, it introduces further error into the system. The importance of this error will be largely determined by the degree of understanding of the instrumentation itself.

Turning again to the EEG, it was discovered that the ink written record was at best quite limited in its ability to reproduce the electrical changes that occurred at the electrodes. The inertia and non-linearity of the pen writing system produced a marked distortion of the original signal. The inertialess and linear recording system of magnetic tape and cathode ray oscilloscope provided a much more sensitive and accurate picture. Having captured this continuous signal, it had to be changed to a series of numbers that the computer could understand as being the EEG. This process of analogue to digital conversion is in practical application somewhat difficult for biological signals. Most conversion equipment is designed for high frequency signals and accuracy falls off at low frequencies. With the assistance of the IBM corporation we developed a conversion system that remains accurate in the low frequency range.

CONDITION	B	R	B	IC	B	R	B	IC	B	R	B	IC	B	R	B	IC	B		
ACCURACY		.9		.9		.8		.9		.7		.8		.7		.9			

$$F'(\theta) = \frac{\theta_B}{\theta_P} t$$

(MODEL)

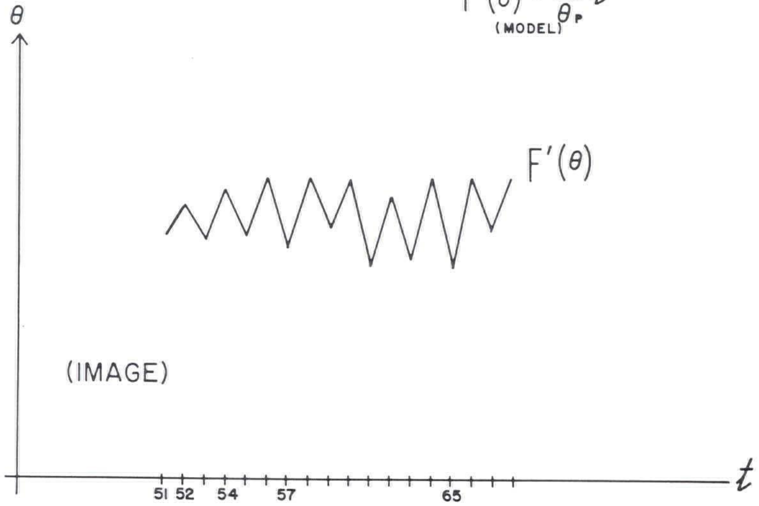
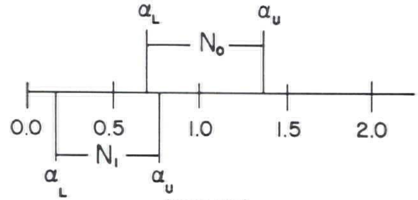


FIGURE 3



(IMAGE)

Let $\alpha = \frac{\theta_R}{\theta_T}$; Then $\underset{\text{NULL HYPOTHESIS}}{N_0}$, $\alpha = \frac{1}{1} = 1.0$

$$\underset{\text{(PROBABILITY)}}{P} - \left[\alpha_L \geqslant 1 \geqslant \alpha_U \right] = 0.99$$

(MODEL)

FIGURE 4

This system also involved fitting a time base to our EEG recording that would identify each change in test conditions. When the system was in operation, one 48-hour experiment would provide a billion numbers for the computer to eschew. Thus the simple job of going from visual examination of an ink written record to an input for the computer involved considerable study and testing of what was thought to be familiar equipment.

Having decided on the definition of the problem and the character of the data to be measured, the design of the computer system was fairly easy. The computer system consists of input hardware, the central processor (computer proper) and the output equipment.

There are several ways of supplying information to the computer: punched cards, paper tape, console typewriter and magnetic tape. Programs are best written, edited and tested by using punched cards. A card reader would be required; however, this would not be able to handle the vast amount of data from just one experiment. Paper tape, likewise, could not handle this particular job. Some rough estimates suggested that one experiment would require several hundreds of miles of paper tape and months to feed it to the computer. Magnetic tape could do the job quite easily. It could feed the information to the computer as fast as the computer could use it, and final operation proved that 60 minutes of raw EEG could be processed by the computer in 70 minutes.

After the computations were complete, the results could be printed on a few punched cards. Since the card punch was already present as part of the card reader, we could use a unit record machine as an off-line printer. This machine is much less expensive than an on-line printer.

In final analysis then we had a computer system that was designed to handle a specific type of problem with the most efficiency and economy. Fortunately, computer systems are now being manufactured that are much more versatile and flexible, so that any one system can be expediently and economically changed to handle any type of problem.

The matter of displaying the results is largely a matter of individual choice. The computer produces only numbers. These numbers can be fitted to the model and the problem is solved. Of course the fitting of these numbers may constitute yet another computer operation. Equally as important as solving the problem specified is the perusal of the results for further ideas. The original problem had to be narrowly defined, and usually this is only part of a recognized but poorly defined broader problem. It would be hoped that some approach to the broader problem might be attained by the proper viewing of the processed data.

Computers can be programmed to drive X-Y plotters that will give graphic or pictorial display of the computations. As an example of this (from another laboratory) the power spectrum analysis of the EEG recording is graphed as a function of area and time, and this is displayed as a map of the brain showing phase relationships of certain EEG frequencies. These are studied for ideas as to how best to produce mathematical models for these phase and areal relationships. This, of course, is not essential to the "systems analysis" but it is part of the dividends that are accrued by using these powerful instruments.

The final test of the system comes from fitting the model. If the mathematical model and its computer program are correct (and these are the most easily proved parts of the system) then the fitting of the results will provide some answers to the original problem. If the results support the original hypothesis the future is bright, but if they don't, the system should be reevaluated before discarding the hypothesis. The results may give clues as to errors in the original definition of the problem or in the method of acquiring the data.

If one has read this far, he must already have arrived at the obvious conclusion that the "analysis of a system" to solve a problem requires a group effort. None of the foregoing steps can be omitted and it requires a considerable knowledge of each step. This is beyond the capabilities of a single individual. This group should contain, besides the initial researcher, a mathematician or statistician (preferably both), a system engineer (or a competent programmer) and if instrumentation is used, an electronics engineer. This group should be formed before any other activity is commenced, and the first effort of the group will be to establish thorough communication among its members. It took literally several days' discussion to define a "wave" to the mutual satisfaction of an electroencephalographer, an electronics engineer and a mathematician.

The foregoing discussion has been aimed at encouraging the use of the computer, but more importantly at the proper use of the computer. If an idea is present, it is not difficult to gather a systems group to develop that idea. Without a great deal of human effort in providing a proper framework within which to operate, the computer is a very grand and expensive and useless tool.

LEGEND

$$F(V) = \sum V_{1} + V_{1+2} + V_{1+2+3} + {*}{*}{*}{*}{*}{*}{*}{*} + V_{1+2+***+96}$$

$V = \text{EEG} = \text{delta}+\text{theta}+\text{alpha}+\text{beta}_1$
$+\text{beta}_2+\text{beta}_3$
delta $= \delta = \frac{1}{2}$ - 3 cps.
theta $= \theta = 4$ - 7 cps.
alpha $= \alpha = 8$ - 13 cps.
$\text{beta}_1 = \beta_1 = 14$ - 25 cps.
$\text{beta}_2 = \beta_2 = 26$ - 35 cps.
$\text{beta}_3 = \beta_3 = 36$ - 50 cps.
B = rest period
R = reversal test
IC = interpolated cue test
P = R + IC
S = left hemisphere
D = right hemisphere

REFERENCE

1. Nooney, Grove C.: Mathematical modeling; Sixth annual IBM symposium on bio-medical data processing, In Press.