Third IBM Medical Symposium - October 1961

Endicott, New York

MEDICAL DIAGNOSIS USING A DIGITAL COMPUTER

And

ANALOG COMPUTER TECHNICS IN STUDY OF CONTROL MECHANISMS IN THE CIRCULATORY SYSTEM

By

Homer R. Warner, M. D., Ph. D.
Established Investigator, Amer. Heart Assn.
Director, Cardiovascular Lab.
Latter-day Saints Hospital

I would like to discuss some of the practical problems that have arisen in applying this approach to medical diagnosis and make some suggestions as to how these may be evaluated.

First, let us consider the problems that arise relative to the assumptions of independence of the various diseases; that is, the assumption that a patient may have only one of these diseases. As an example, consider the following fact: The incidence of mild cyanosis in patients with atrial septal defect alone is 2%, while the incidence of mild cyanosis in isolated pulmonary stenosis alone is 1% or less. However, it is evident, since the incidence of mild cyanosis in patients having both defects is 20%, that one cannot deduce anything about the probability of the combination pulmonary stenosis and atrial septal defect in the presence of cyanosis from the statistics regarding the individual lesions. The combinations of these lesions produces an entirely different pathological pattern and this must be treated as a separate entity. On the other hand some compromise must be made, because there is essentially an endless spectrum of diseases and if each variation were treated as a separate entity, an unmanageable number of diseases would result and collection of enough data on the incidence of symptoms in any one disease to allow a reliable calculation of probability to be made would be impossible. The number of disease categories into which the population should be divided for most efficient use of this approach to disease detection from symptom analysis must be decided by answering the following question: "Does each disease entity occur frequently enough to provide significant statistical data on the occurrence of each symptom in this entity?" If it does not, severe restrictions will be placed on the accuracy with which the diagnosis can be made should it be included as a separate entity. Fortunately this very fact means that the disease will not appear very often in the population being evaluated and the error introduced in evaluating a large number of cases will be very small. On the other hand, even a rare disease may be accurately diagnosed if the symptom pattern is sufficiently distinctive.

Next, let us consider the problem of independence of symptoms. x_1 and x_2 are two symptoms and are considered independent of one another if the probability of x_1 being present in disease y_k is not influenced by the presence of symptom x_2.

$$P_{x_1/y_k} = P_{x_1/x_2, y_k} \qquad \text{Equation II}$$

In words, the probability of x_1 in disease y_k is the same as the probability of symptom x_1 in that fraction of the patients with disease y_k who happen to have symptom x_2 as well. Unfortunately at the present time we cannot evaluate our data matrix for independence by this test for each of the symptoms because much of the data was gathered from the literature and it is impossible to determine the coincidence of symptoms from the data as it was presented. With the accumulation of sufficient experience, however, it will be possible to carry out this analysis for each symptom from our own data with the help of the computer.

It does not follow that if two symptoms are not completely independent of one another, as measured by this criteria, that each is not capable of contributing information of value pertaining to the diagnosis of a particular patient. Such symptoms, however, would have to be handled in a special fashion as the assumption of independence is inherent in the equation presented here.

Let us now turn to the problem of evaluating the usefulness of a given symptom in diagnosing any disease. The information content (I) of the message "symptom x_1 is present or absent in this patient" may be defined as the logarithm of the ratio of the probability that symptom x_1 is present or absent in this patient who has disease y_1 to the probability that symptom x_1 is present or absent in any patient from this population.

$$I = \ell n \left(\frac{P_{x_1/y_1}}{P_{x_1}} \right)$$

Now this information can be either positive or negative since the probability of a particular symptom in a given disease may be greater or less than the incidence of that symptom in the group of diseases under study. However, if the information content is defined as the absolute value of the logarithm of this ratio, a number is obtained which is independent of the sign of the measure.

The average information content ($\bar{I}$) of a given symptom for a set of k diseases can be obtained from

$$\bar{I} = \frac{1}{k} \sum_{\text{all } k} \left| \ell n \, P_{x_1/y_k} - \ell n \, P_{x_1} \right| \qquad \text{Equation 12}$$

where P_{x1} is given by equation 5. Now since

$$\frac{P_{y_k/x_1}}{P_{y_k}} = \frac{P_{x_1/y_k}}{P_{x_1}} = e^{\bar{I}} \qquad \text{Equation 13}$$

follows from equations 5 and 11, the term $e^{\bar{I}}$ is the average factor by which the calculated probability of a disease is changed by finding symptom x_1 to be present or absent in the patient to be diagnosed. This term then is a direct measure of the average value of that particular symptom in diagnosing this group of diseases.

As an example, let us compare the average information content $(\bar{I})$ of three symptoms. Symptom 17 is a systolic murmur loudest in the fourth interspace. P_{x17} is .52, which means it occurs in 52% of all patients in this population, and $(\bar{I})$ is 1.61. Stated another way, on the average the presence of this symptom in a patient decreases or increases the calculated probability that the patient has any given disease by a factor of 5.0 over the probability that the patient had that disease before it was determined whether or not the patient had symptom 17. Symptom 42 is a holosystolic murmur loudest in the fourth interspace and has an incidence of 36%. This is a refinement of the description of the murmur defined as symptom 17 and has an information content of 2.46. The value of a symptom must be judged, however, not only on its information content but also on the reliability with which it can be detected. Our experience to date indicates that in spite of the fact that the more detailed description lowers reliability, this description of a murmur is more useful if a phonocardiogram is used to provide an objective record.

Finally, let us consider a vague symptom such as "fatigues easily" which occurs in approximately half of the patients. $\bar{I}$ for this symptom is only 0.59. Since this symptom is also difficult to evaluate it is doubtful that it should be included in the data matrix. I suggest the following criteria for evaluation of symptoms to be used for diagnosis:

1) A symptom should be one whose presence or absence can be accurately recognized.

2) A symptom should be independent of other symptoms in any given disease.

3) A symptom should have an average information content greater than 1.0.

The purpose of the next section of this paper is to present an approach to the study of regulation and control of the circulation centered around the use of an analog computer. The problem will be discussed in five parts: 1) analysis of the circulation as a complex, closed-loop system, 2) the use of the computer as an aid in performing experiments, 3) the computer as a tool for reducing data to the desired form for analysis, 4) the use of an analog computer in testing an hypothesis or mathematical model, and 5) the combined use of analog and digital computers.

The block diagram in figure 1 represents the circulation as a closed-loop with two symmetrical halves. Each half consists of a distensible reservoir (the large veins and atrium), a pump which has two states, diastole and systole, a transmission line which is the large arteries, and a source of resistance to run-off from the large arteries which is located at the level of the small arteries and arterioles. These elements are connected to the other half of the circuit to form a complete closed-loop. Not only is each element effect by the element just ahead and just behind, but its performance is also influenced by events taking place at remote parts of the circulation. For instance, information regarding pressure in the large arteries is sent to the central nervous system which in turn modified flow and resistance to flow. However, the circulatory system, even when deprived of this nervous control, has the ability to return to an equilibrium state following a transient disturbance. This we might call the "auto-regulation" of the circulation. It is important to first understand this phenomenon since any nervous control of the circulation must be simply superimposed upon this basic "auto-regulation" phenomenon.

A set of eight equations is used to describe each half of the circulation. The two halves differ only in the values of the parameters.

$$\text{Equation 1} \quad - \quad V_1 = V_1\,(t=o) + \int (F_1 - F_2)\,dt$$

states that the volume of the atrium and large veins is equal to its initial volume plus the difference between flow into the atrium (F_1) and flow out of the atrium (F_2) integrated with respect to time.

The pressure (P_1) in the left atrium and pulmonary veins is treated as a power function of the volume divided by the capacitance of that chamber, in order to account for the well-known convexity toward the volume axis of the volume pressure curve of veins. This is expressed in

$$\text{Equation 2 - } \quad P_1 = \frac{V_1^{\,n}}{C_1} \, .$$

Flow (F_2) out of the left atrium and into the left ventricle is zero during ventricular systole and during diastole is equal to the pressure gradient across the valve divided by the resistance to flow (R_1) minus an inertia term which depends on the rate of change of flow as shown in

$$\text{Equation 3 - } \quad F_2 = \frac{(P_1 - P_2)}{R_1} - L_1 \frac{dF_2}{dt} \qquad \text{(diastole)}$$

$$F_2 = 0 \, . \qquad \text{(systole)}$$

Pressure in the left ventricle (P_{2d}) during diastole is again expressed as a power function of volume divided by the diastolic capacitance of the ventricle C_{2d}.

$$\text{Equation 4 - } \quad P_{2d} = \frac{V_2^{\,m}}{C_{2d}}$$

The volume of the ventricle (V_2) may be expressed as some initial volume plus the integral of flow in minus flow out as shown in

$$\text{Equation 5 - } \quad V_2 = V_{2(t=o)} + \int (F_2 - F_3)dt \, .$$

Flow out of the ventricle (F_3) is zero during diastole and during systole depends upon the volume of the ventricle divided by its systolic capacitance (C_{2s}), on R_2, the frictional forces which limit the rate of contraction and also on the pressure in the aorta which will depend on the volume of the aorta (V_3) and the aortic capacitance (C_3), and on an inertia term proportional to the rate of change of flow. This is expressed in Equation 6.

$$\text{Equation 6 - } \quad F_3 = \frac{V_2}{R_2 C_{2s}} - \frac{L_2}{R_2} \frac{dF_2}{dt} - \frac{V_3}{R_2 C_3} \qquad \text{(systole)}$$

$$F_3 = 0 \qquad \text{(diastole)}$$

The volume of the aorta depends on the integral of the flow in minus the flow out as shown in

$$\text{Equation 7 - } \quad V_3 = V_{3(t=o)} + \int (F_3 - F_4)dt \, .$$

And finally, equation 8 expresses the flow out of the aorta (F_4) as a function of aortic volume, aortic capacitance, and the resistance to flow out of the arterial bed. The inertia term (L_3) is small but must be included in order to account for the experimental observations. (1)

$$\text{Equation 8 - } \quad F_4 = \frac{V_3}{R_3 C_3} - \frac{L_3}{R_3} \frac{dF_4}{dt}$$

To complete the loop, eight more equations must be written to describe the properties of the large systemic veins and right atrium, the right ventricle, and the pulmonary arterial bed. These equations have the same form as the ones just written.

The most important assumption made in the derivation of these equations is that the pumping action of the ventricles results from a triggering by the electrical event of the ventricular muscle from one passive state to another, represented here as a change from diastolic to systolic capacitance and that the ejection of blood during systole is determined by the volume to which the ventricle filled during the previous diastole, the passive resistive and capacitive properties of the ventricle during systole and impedance to flow into the aorta. These properties are not time dependent but remain constant during the course of systole. In support of this concept is the finding that wave forms generated by the model compare favorably with flow curves recorded with an electromagnetic flowmeter on the ascending aorta of a dog as shown in figure 2.

It is possible with this mathematical model to examine the response of this system to a transient disturbance from its equilibrium state and to compare this response to the response of its biological counterpart, which in this case is the circulatory system of a dog deprived pharmacologically of his autonomic nervous system. Simultaneous solution of this set of equations predicts a rapid return to equilibrium with no overshoot except at high flows following such a disturbance as a Valsalva maneuver in which blood is displaced from the pulmonary circuit into the systemic circuit. This is in agreement with the response observed in the biological preparation.

Once the investigator is satisfied that the equations will represent the observed dynamic inter-relationships in the circulation, he may proceed with a systematic investigation of the role played by each component in determining overall system performance. A new solution may be obtained with each sweep of the oscilloscope and a parameter change is accomplished by simple adjustment of a potentiometer in the analog computer. The ease with which such an analysis of parameter changes may be performed is a distinct advantage of the analog computer for this kind of work.

It is essential that this model, representing only the inter-relationships among the physical properties of the circulation, be expanded to include the nervous reflex elements which play such an important role in controlling these physical properties in the intact animal. The analog computer has also proved valuable in approaching this problem. The remainder of this paper will be a presentation of examples to illustrate the role an analog computer may play in acquiring data from the experimental animal, in reducing the data to a meaningful form for analysis, and finally in simulation of the biological system under study as a means for testing an hypothesis regarding its dynamic characteristics. First, consider the role of an analog computer as an integral part of data acquisition equipment for studying a reflex pathway.

In attempting a quantitative description of the effect of a particular nervous element in the control system, the exact form of the input function must be known. For instance, if a sinusoidal variation in the frequency of stimulation of a particular nerve is desired, a sine wave is generated by a function generator. This is fed to the computer where it is scaled and biased to the desired level. The output voltage is then fed in parallel to a tape recorder to represent the input to the animal and to an analog-to-frequency converter. Such devices are designed to operate at high frequencies. In order to increase the accuracy at low frequencies, pulses from the converter are fed to a counter where the output of the second decade is sampled. This divides the frequency by 100 and thus increases the accuracy of the analog-to-frequency conversion near zero frequency by a factor of 100. The pulse out of the counter then drives the stimulator which in turn is connected to the stimulating electrodes in the animal.

If the input to the animal is to be a function of some variable elsewhere in the animal, this may be accomplished as follows: the variable is sensed with a transducer and the voltage from the transducer is fed to the computer where the operations are performed in accordance with the known characteristics of the biological transducer or sense organ being simulated. This system then, when placed in parallel with the dog's own biological transducer (for instance, the carotid sinus) acts to amplify any particular characteristic of this organ and may be used to study the effects of changing this organ's parameters on the performance of the while circulatory system. Such a study has been carried out in the case of the carotid sinus. (2)

Once the experiment has been performed and the output of the transducers recorded on magnetic tape, a computer may be used to convert this raw data into a more meaningful form for analysis. As an example of this, the analog computer program used to obtain beat-by-beat stroke volume, heart rate, cardiac output, mean arterial pressure, and "peripheral resistance" from three variables recorded on magnetic tape from the experimental animal, namely, the electrocardiogram, aortic flow, and aortic pressure is shown in figure 3. The electrocardiogram recorded during exercise may be considerably distorted due to electrode movement and muscle artefact. Thus, in order to obtain a trigger pulse at the time of each "R" wave, the electrocardiogram is differentiated. Then rectified and biased in order to obtain a single spike from the R wave and zero voltage during the rest of the heart cycle. This spike triggers a sawtooth generator which in turn triggers two pulse generators in sequence. Each pulse generator operates a relay coil. Following each R wave relay 1 closes first and allows amplifier 4 to charge up to the voltage existing on the output of amplifier 3. This relay then opens and relay 2 closes, setting amplifier 3 back to its initial conditions. Each relay is closed for only 2 milliseconds. Amplifier 3 then begins integrating the constant voltage generating a sawtooth whose final height is proportional to the period of the heart cycle. The initial condition voltage on amplifier 3 is set to compensate for the time lost in relay closure. The output of amplifier 4 then is a step voltage whose height is proportional to the period of the preceeding cardiac cycle. Since the heart rate is proportional to the reciprocal of this voltage, the output of amplifier 4 is divided

into a constant voltage to generate the heart rate signal. Also shown are essentially identical circuits used to derive voltages proportional to the integral of arterial pressure with respect to time over one heart cycle and a voltage proportional to stroke volume. Then with multiplier and divider circuits we may obtain beat-by-beat heart rate, stroke volume, cardiac output, mean arterial pressure, and the ratio of mean arterial pressure to cardiac output, the so-called peripheral resistance. Once the data has been reduced to this form and beat-by-beat values for each of these variables recorded simultaneously, the inter-relationships among them may become evident from the time sequence of events following a step or sine wave variation in speed of the treadmill on which the dog is exercising.

The final and perhaps the most useful role of the analog computer is in testing hypotheses derived to represent the inter-relationships among the component parts of a circulatory reflex loop. As an example, a study of the relationship between heart rate and the frequency of efferent action potentials on the sympathetic nerves going to the heart will be presented. This system is just one link in the heart rate control system but it is essential that each link be characterized in this fashion before the overall control mechanisms can be understood. In this study the computer was used to generate an analog voltage proportional to the input which here is the frequency (f_1) of stimulation of the right preganglionic efferent sympathetic nerves to the heart of a dog anesthetized with nembutal. Both cardiac sympathetic and both vagus nerves are severed. The electrocardiogram and a voltage proportional to f_1 are recorded on magnetic tape using a reel-to-reel transport. Later, the recorded data is reproduced and selected portions of the experimental data recorded on a continuous loop of tape along with the heart rate calculated by the computer from the electrocardiogram as described above. Once this modified data is on a continuous loop it is reproduced over and over again at eight times the original recording speed. The heart rate (or output) is displayed on the oscilloscope and the input (f_1) is fed into the computer on which is programmed the equations representing the hypothetical relationship between heart rate and f_1. The measured and predicted heart rates then are compared on a dual beam oscilloscope or X-Y plotter. The parameters of the equations are adjusted to obtain the best possible

fit of the two curves. The block diagram and set of equations are used to predict heart rate from frequency (f_1) of sympathetic nerve stimulation are shown in figure 4. Equation 1 states that the rate at which noradrenalin concentration (A_o) changes just beyond the sympathetic nerve endings is proportional to the frequency of sympathetic nerve stimulation (f_1) and the number of fibers responding to this stimulus (n) and that this noradrenalin diffuses to the active site on the S. A. node at a rate proportional to the concentration gradient ($A_o - A_1$). A_1 reacts with a substance B in a reversible fashion to form compound AB and the resulting change in heart rate is proportional to AB concentration. Since the substance B is present in limited quantity it establishes the maximum heart rate which can be obtained. These phenomenon are represented by these five equations and are programmed on an analog computer. In the top recording in figure 4 is shown the heart rate recorded in response to a step input in frequency (f_1) of sympathetic nerve stimulation. Below this is shown the predicted heart rate in response to the same step input and at the bottom the predicted and recorded heart rate are superimposed. The equation parameters were manipulated to obtain the best possible fit with the experimental curve. The rate of increase in heart rate depends upon frequency of stimulation. In many dogs, the response to a stimulus of 2 per second is approximately 3/4 the response of a stimulus of 20 per second. Once the optimal equation parameters have been determined for two such step inputs in f_1, the equations will accurately predict the time-course of heart rate resulting from sinusoidal variations in f_1.

This example illustrates the way in which an analog computer may be used to analyze the dynamic relationship between input and output for one component of a reflex arc. Such studies must be performed on each element of a nervous control loop or reflex before it can be integrated into the general model of the circulation.

Two points in regard to this approach should be emphasized. First, when an equation is found which will describe a system, its value must be judged on the basis of its ability to describe the system under all circumstances. The equation should have a minimum number of parameters and each of the parameters should be sensitive to changes in a particular

system characteristic. If these two criteria be satisfied, the question of uniqueness of the equation is of no concern. The second point deserving emphasis is the fact that valuable information may be obtained each time an equation fails to predict the behavior of a system. Such a failure means that existing concepts regarding this system's performance are inadequate to account for the observed facts since the equation being tested was derived from these concepts. Thus a modification of the prevailing concepts is necessary and new concepts must be sought. This "negative information" may be very valuable.

The very nature of the analog computer lends itself to this sort of theoretical physiology. Often when a mathematical model fails to simulate its biological counterpart, valuable insight into the type of modification of the theory necessary to improve its correspondence to the biological system will be gained by careful consideration based on knowledge of the electrical analog system. In other words, the electrical system provides just one more physical realm from which to derive insight regarding the dynamic characteristics of the biological system under study.

The examples presented illustrate the way in which an analog computer may be used as an integral part of a laboratory engaged in the study of physiologic control, emphasizing its role in initial analysis of the overall system, as a tool for obtaining experimental data and for reducing this data to the desired form for analysis, and as a means for solving sets of equations set up to represent hypotheses regarding input-output relationships. The accuracy with which an analog computer will perform these tasks will vary from 0.1 to 2%, depending upon the nature and complexity of the system being studied. It is apparent, however, that the accuracy of the analog computer does not determine the limits of accuracy of solution for most problems encountered in physiology today. Instead, these limits are set by the accuracy with which the original measurements are made and the necessity of simplifying assumptions incorporated into any mathematical model derived to represent a physiological system at the level at which measurement of system variables is now possible.

Although the analog computer is well suited for this type of simulation work because of the ease with which it solves differential equations and the direct way in which the solutions are displayed for the investigator, it has certain limitations. Chief among these is its inability to store information upon which to base complex decisions and its inability to handle accurately calculations involving small differences in large numbers. With an effective analog-to-digital and digital-to-analog conversion system it would be possible to utilize a small digital computer to advantage in simulation work such as that described above. The experimental data would first be recorded on analog tape. If called for by the nature of the problem, the data could then be filtered, scaled, biased, or integrated in an analog computer before being converted to digital form. The digital computer could then be used to perform certain logical decisions and based on this, modify the data in a specified fashion adding control information derived from judgements made on the data. This modified data with the control information would then be converted back to analog form and recorded on an analog tape loop. The control information might be a series of pulses appropriately spaced in time to control the closure of relays in the analog computer.

The analog computer could then be used for the initial attempts at simulation which often involve long integrations which are done only relatively slowly in the digital computer. More elaborate and realistic models would be possible due to the operations already performed on the data by the digital computer and the presence of control pulses inserted at proper places in the data based on decisions made by the digital computer. The digital computer could also be used to carry out calculations of the interrelationship of any two variables and the results printed out. This relationship could then be utilized in the analog computation by programming the appropriate curve on a function generator.

After the analog computer has been used to determine the general form of the model and the range of parameters to be expected, the general model could be programmed on the digital computer. The logical decisions used by the analog computer operator in optimizing this solution by visualizing the comparison of the predicted and recorded output on an oscilloscope could then

be formalized and programmed into the digital computer so that it would, by a series of decisions based on the results obtained from each solution, find the optimal parameters to satisfy a least squares fit or other appropriate statistical measure between the predicted and recorded data. Such an interplay between analog and digital computers might offer the optimum solution to the complex problem of simulation of biological control systems, since it takes advantage of the ability of the analog computer to rapidly solve differential equations and display the results to the investigator and the ability of the digital computer to store information and make logical decisions.

Dr. Warner is an Established Investigator for the American Heart Association. This project was supported by grant No. H-3607 from the U. S. P. H. Service and by a grant from the Utah Heart Association.

References

1) Warner, H. R. Circulation Research 5:79, 1957

2) Warner, H. R. Circulation Research 6:35, 1958

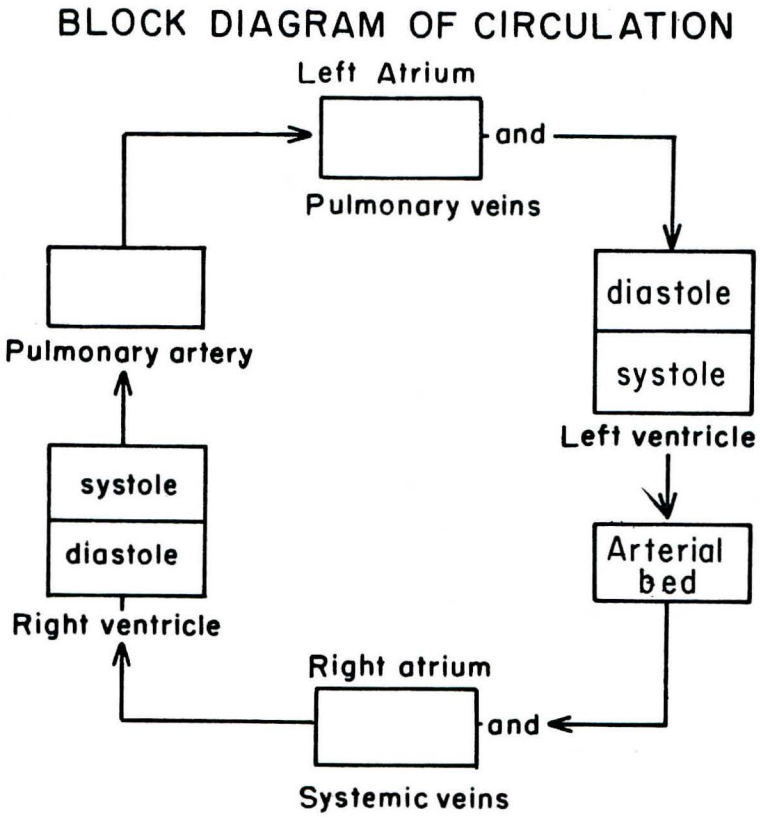


Figure 1

Schematic block diagram of the circulation described mathematically by the equations in the text.

Figure 2

Comparison of the time-course of flow in the ascending aorta predicted by the equations (smooth line) and the flow recorded from the ascending aorta of the dog. The theory does not account for the small reversal of flow at the end of systole seen on the recorded flow curve.

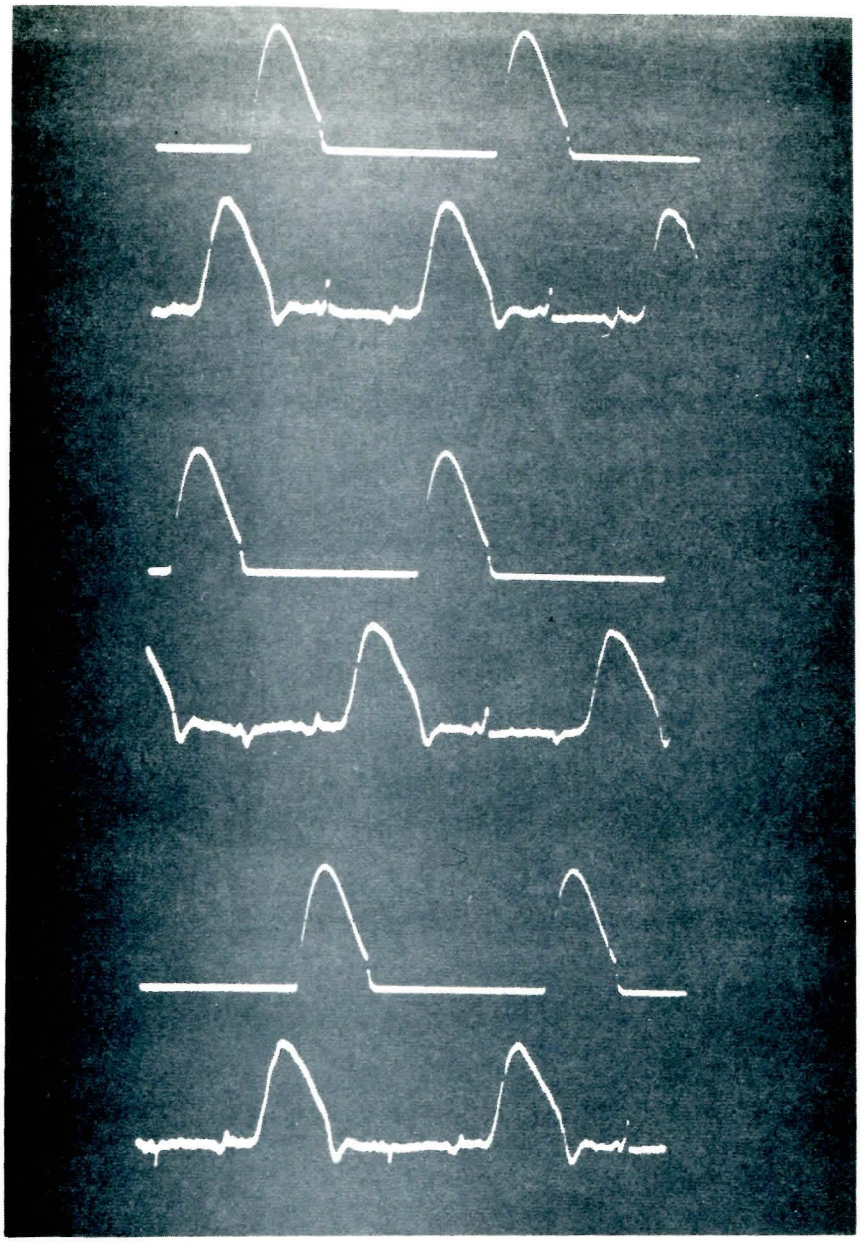


Figure 3

Diagram illustrating the analog computer program for deriving heart rate stroke volume, cardiac output, mean arterial pressure, and resistance beat-by-beat from a tape recording of the electrocardiogram, flow in the ascending aorta, and aortic pressure.

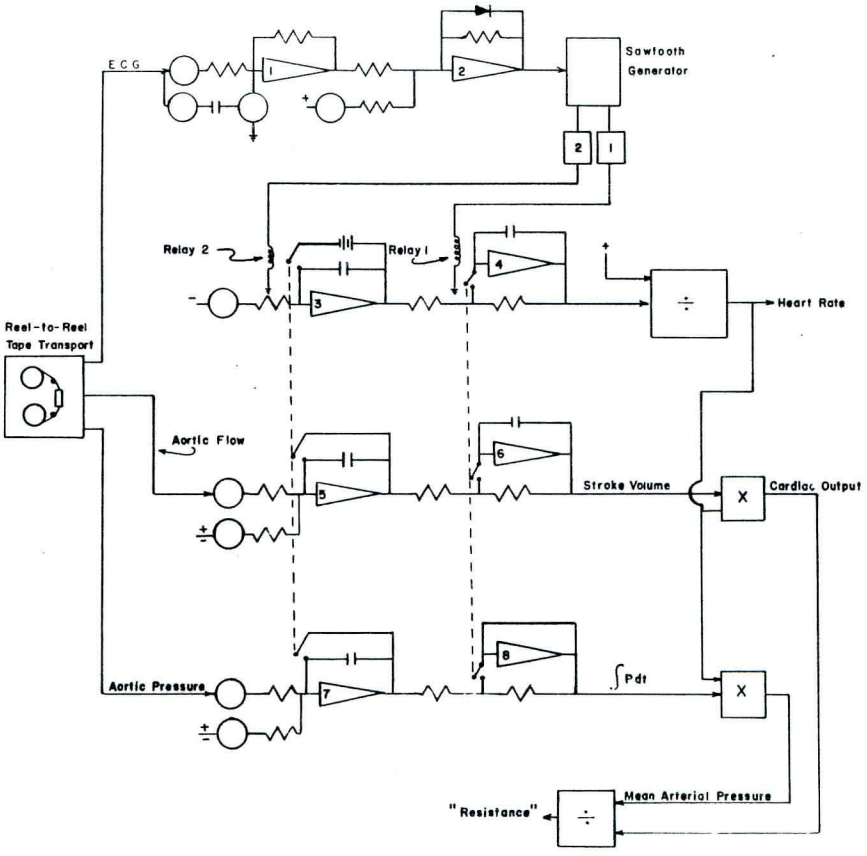
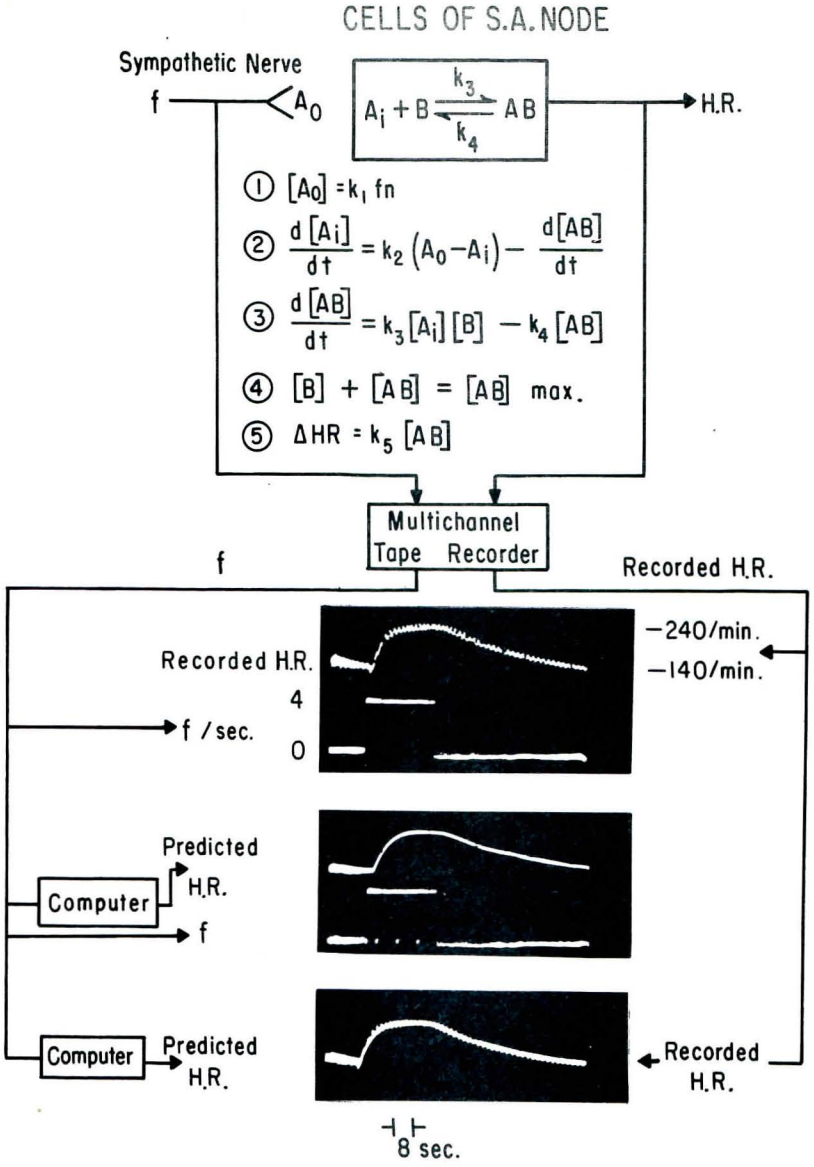

ECG
Sawtooth Generator
Relay 2
Relay 1
Heart Rate
Reel-to-Reel Tape Transport
Aortic Flow
Stroke Volume
Cardiac Output
Aortic Pressure
∫Pdt
"Resistance"
Mean Arterial Pressure

Figure 4

Mathematical model representing the relationship between f_1 the frequency of stimulation of the sympathetic efferent nerves to the heart and heart rate. The time-course of heart rate predicted from a solution of these equations is compared to the recorded heart rate from an experimental animal (see text for definitions).


CELLS OF S.A. NODE
Sympathetic Nerve
f
A_0
$A_i + B \underset{k_4}{\overset{k_3}{\rightleftharpoons}} AB$
H.R.
① $[A_0] = k_1 fn$
② $\frac{d[A_i]}{dt} = k_2 (A_0 - A_i) - \frac{d[AB]}{dt}$
③ $\frac{d[AB]}{dt} = k_3 [A_i][B] - k_4 [AB]$
④ $[B] + [AB] = [AB]$ max.
⑤ $\Delta HR = k_5 [AB]$
Multichannel Tape Recorder
f
Recorded H.R.
Recorded H.R.
4
0
f / sec.
−240/min.
−140/min.
Predicted H.R.
Computer
f
Computer
Predicted H.R.
Recorded H.R.
8 sec.