X. PROCESS ANALYSIS AND SYNTHESIS

Dr. M. V. Cerrillo
Prof. P. Elias
Prof. R. M. Fano
Prof. E. A. Guillemin
Prof. Y. W. Lee
Prof. C. E. Shannon
R. E. Wernikoff

[This is a continuation of the discussion that was presented in the Quarterly Progress Report of July 15, 1957.]

3.41 Methods for Increasing the Spectral Probability

Two methods suggest themselves by means of which we can increase the probability that certain spectral lines are generated only by the direct effect of the meter and then select the lines that have this property.

The first method consists of comparing the extremal distribution spectra of one channel at different time intervals (without going to another channel), say, for example, after one, two, or three successive sampling intervals that have a duration equal to the time-length of the bar and then selecting the coincident lines that are common to each interval. We can realize this method by re-recording the music in the channel on a tape recorder and playing it back with two, three, four, . . . , N playback heads that are displaced by a distance equal to the bar duration multiplied by the speed of the tape. In this way, we construct new spectral functions $R_{k,n\Delta}(t)$, $r_{k,n\Delta}(t)$, $E_{k,n\Delta}(t)$, and $e_{k,n\Delta}(t)$, $n = 1, 2, 3, N$, for each head. The index $n\Delta$ indicates a head displaced by n units. Let us now construct the nonempty spectral function

$$\mathscr{R}_k = \bigcap_{n=1}^{N} R_{k,n\Delta}(t); \quad k = \text{channel index}$$

(See Fig. X-1.) It can be seen that the lines of $\mathscr{R}_k(t)$ have a high probability of belonging to the quantitative elements of the meter because of the resulting periodic character of these new lines. This operation of intersection produces, at the same time, the required selection of lines which are more closely associated to the rhythm. This selective operation can be realized by a group of coincidence circuits. A similar operation, with a different interpretation, produces other new spectral functions

$$\mathscr{E}_k = \bigcap_{n=1}^{N} E_{k,n\Delta}(t); \; e_k = \bigcap_{n=1}^{N} e_{k,n\Delta}(t); \text{ and } \rho_k = \bigcap_{n=1}^{N} r_{k,n\Delta}(t)$$

Although this method is simple, it is not practical because it depends upon the bar duration – which has not yet been determined.

The second method consists of simultaneous comparison of the isomorphic spectral functions for different channels. For example, we can form new spectral functions that have a higher probability of belonging to the meter:

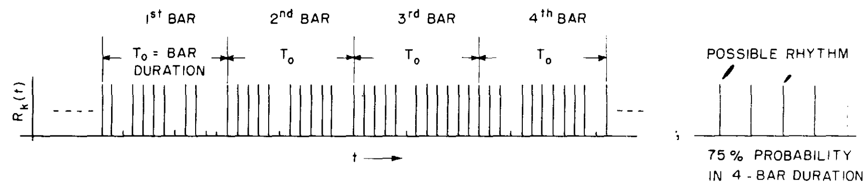

Fig. X-1. Schematic illustration of the first method. Rhythm is extracted by comparing the $\mathscr{R}_k(t)$ spectra for the same channel at different bar intervals of the meter. Similar comparison with $E_k(t)$ produces the corresponding emphasis. By using the $r_k(t)$ and $e_k(t)$ spectra simultaneously the probability of the rhythm can be increased almost to 1 in four bars. The accent would come from $\mathscr{R}_k(t)$ and $\mathscr{E}_k(t)$.

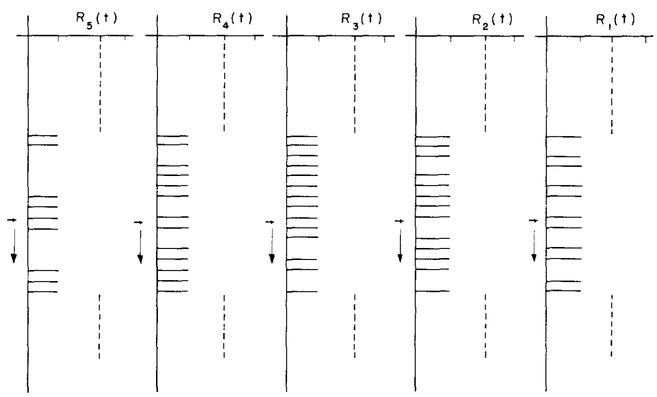

Fig. X-2. Diagrammatic representation of the maximum distribution spectral functions $R_k(t)$ of each channel at the same time interval. Note that the boolean union of these functions tends to produce, with increasing probability, the total number of notes played in a given time interval. The intersection produces a spectral function with increased probability of generating the quantitative elements of the meter.

$$\mathscr{R} = \bigcap_{k=1}^{5} R_k; \quad \mathscr{E} = \bigcap_{k=1}^{5} E_k; \quad \rho = \bigcap_{k=1}^{5} r_k; \quad \xi = \bigcap_{k=1}^{5} e_k$$

In this report we follow the second method. It allows a rapid determination of the bar duration Δ, and offers other advantages that will be appreciated in the discussion that follows. A schematic representation of the method is illustrated in Fig. X-2. We shall limit ourselves to the basic points and omit circuit details, which are not needed for the general understanding of the process of rhythm extraction in music.

3.42 Number of Notes Played in a Given Time Interval

For a time signature that is called for in the script, the number of notes played in a bar may be equal to the upper number of the time signature or it may be a multiple of this number. For example, in three-quarter time there are 3 quarter notes or 6 eighth notes or 12 sixteenth notes, and so forth. Other combinations of notes that are not multiples of the upper number of the meter signature may also appear because of silence intervals inserted in place of notes. For example,

a. One quarter note only. The meter is then completed with 2 one-quarter silence intervals.
b. One quarter note, 3 eighth notes, and 1 eighth silence interval.
c. One half note and 1 quarter note, and so forth.

These examples illustrate the point that it is difficult to talk about the number of notes played in a given interval. To dodge this difficulty we can proceed as follows. Suppose that a given musical composition has the time signature $\frac{3}{4}$. Then suppose that, looking at the script, we find that the notes of smaller value are, for example, sixteenth notes. Then, we must make our count in terms of 12 sixteenth notes per bar. Silence intervals will be included in terms of their sixteenth-note value. In any other time interval than the meter bar, we count the number of notes in terms of the number of notes per bar according to our convention. Under this assumption our counting per bar now becomes a multiple of the upper number indicated in the time signature. The value of this factor may be different for different channels.

With increasing probability, the number of notes in a given interval, counted as above, is of the order of

$$N_q = \left(\bigcup_{k=1}^{5} R_k\right) \cup \left(\bigcup_{k=1}^{5} r_k\right)$$

where the index q indicates the multiple of note-counting. N_q is now independent of the number of channels. Now, although the probability is high that N_q represents the number of notes played in a given interval, it is not necessarily equal to one. To raise the probability of counting almost to one, we use relaxation oscillators that are synchronized with the spectral function N_q. This means that several oscillators, each of which has a frequency that is a multiple of q, would stay in synchronism with the musical rhythm. See Fig. X-3. Any of these oscillators would be enough to yield the note-counting. In fact only one is needed. In practice, two are used because one oscillator may get out of step. The incoming signal spectral function N_q is processed as indicated in Fig. X-3. The output of the oscillators produces spectral functions $\eta_1, \eta_2, \ldots, \eta_q$ that now have almost a probability of one of representing the note-counting.

3.43 Bar Duration Δ and Time Signature

It is a general rule of composition that each bar begins with a note to which we give a stronger emphasis than to the other notes in the same bar, but this is not necessarily so. For example, it may begin with a silence interval or by the proper emphasis of a certain instrument or of a harmonic group, and so on. Nevertheless the cyclic emphasis exists and it is enough to determine the bar duration as the period of this recurrent pattern, particularly so because of the channel decomposition. In general, there are strong, weak, and normal accents on the notes in a bar. Consequently, the bar can be determined as the period of the cyclic emphatic elements contained in $E_k(t)$ and $e_k(t)$. In the determination of the bar length channel 2 — and in some music, channel 1 — has a strong bearing on the determination of the bar length. This is because channel 2 contains the sounds coming from most of the accompanying musical instruments.

Now let us consider the following boolean intersections:

$$R_k \cap E_k; \qquad r_k \cap e_k; \qquad k = 1, 2, 3, 4, 5$$

This operation suggests at once that the result of it would tend to produce a pattern whose repetition period might be the same as the bar duration, particularly for k = 2. Since the sets R_k and r_k are disjointed and so, also, are the sets E_k and e_k, then a better description of a set of spectral functions containing the bar duration and its strong, weak, and normal emphasis is clearly given by the resulting spectral function

$$T_k = \left(R_k \cap E_k\right) \cup \left(r_k \cap e_k\right)$$

particularly for channel 2. As a matter of fact, when the rhythm of a musical composition is based completely on the meter, then T_k would contain the rhythmic elements with a fairly good degree of probability. To increase this probability (except, perhaps, in the presence of syncopation), we can use the device consisting of cathode follower, free-running multivibrator, differentiator, and clipper that is shown in Fig. X-3. In this way, we generate an output spectral function τ_k. The spectral function τ_k has almost a probability of one of representing the qualitative elements of the rhythm.

Now consider the boolean operations:

$$C_{k,\nu} = \tau_k \cap \eta_\nu; \qquad k = 1, 2, 3, \ldots, 5; \qquad \nu = 1, 2, \ldots, q$$

If there was not the possibility of error in this representation, then $C_{k,\nu}$ would be independent of ν. But since the actual errors are small, $C_{k,\nu}$ are practically the same for different values of ν when k is held constant. The effect of these errors can be eliminated by a proper combination of $C_{k,\nu}$, for example, by using the common part of each $C_{k,\nu}$. This common part is represented by the operation

$$\mathscr{C}_k = \bigcap_{\nu=1}^{q} C_{k,\nu}$$

It is now clear that a simple arrangement of coincidence circuits will perform this operation and therefore generate a spectral function $\mathscr{C}_k$ for which the bar duration is obviously well defined, particularly for k = 2.

3.44 The Rhythm Structure or Mosaic

The discussion in section 3.43 showed that the function $\mathscr{C}_k$ would produce the fundamental mosaic of the musical rhythm with a period of repetition equal to the bar duration. Our concern, now, is the extraction of the corresponding mosaic when the rhythm period covers several cycles of the bar, as may be the case in symphonic and older forms of music.

We have the feeling that in such cases the pattern can be found by operations of union, intersection, and difference, performed with the set $\mathscr{C}_k$, k = 1, 2, 3, 4, 5, although it is necessary to check this assertion by means of experimental work.

3.45 Injection of the Accent in the Mosaic Pattern of the Rhythm

The spectral function $\mathscr{C}_k$ is simply a series of lines of equal amplitude. If $\mathscr{C}_k$ is used as a signal and is passed through the voice coil of a loud-speaker, we would hear a succession of periodic tappings of equal intensity. This tapping is, of course, in synchronism with the music but will not show the required emphasis of the qualitative elements in the rhythm. We must introduce into this tapping the corresponding musical accent. To bring this about we can modulate τ_k, which has lines of equal amplitude, with the functions $E_k(t)$ and $e_k(t)$. The modulation is performed as follows. A line in

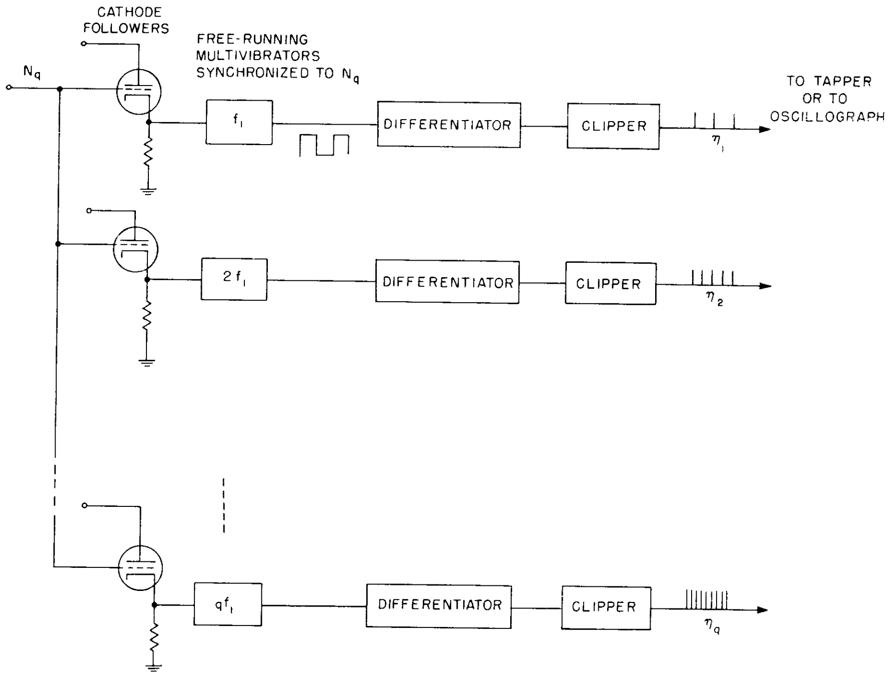

Fig. X-3. The new spectral functions $\eta_1, \eta_2, \ldots, \eta_q$ have a probability almost equal to one of yielding the note-counting. One line of operations would be enough, say η_1. The other spectral functions $\eta_2, \ldots, \eta_q$ have a pulse-repetition rate equal to an even multiple of η_1.

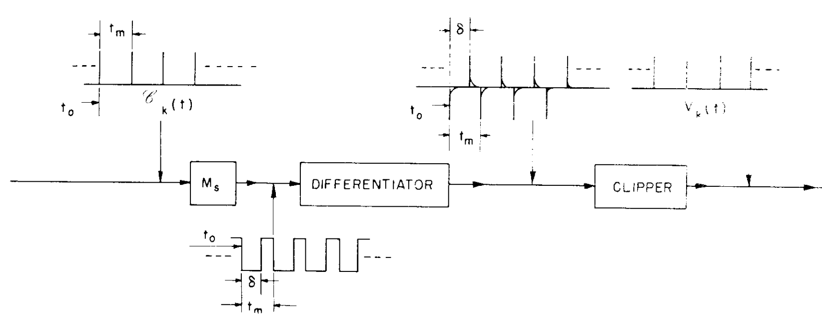

Fig. X-4. Delay correction for the rhythm spectral functions $\mathscr{C}_k(t)$ or $V_k(t)$. M_s is a monostable multivibrator with adjustable delay. The clipper reshapes the spectral function.

τ_k is increased in proportion to the height of the coincident line of $E_k(t)$, and a line in τ_k is decreased in proportion to the height of the coincident line in e_k. Since pulse-modulation circuits are well known we omit the descriptive details here.

Let us denote by $V_k(t)$ the modulated spectral function. If $V_k(t)$ is passed as a signal through the voice coil of a loud-speaker, we shall hear periodic tappings whose intensity changes follow the strong (coming from $E_k(t)$), medium (coming from T_k itself), and weak (coming from $e_k(t)$) accents of the musical rhythm associated with the meter.

3.46 Rephasing of the $\mathscr{C}_k(t)$ Spectral Function

The lines of the spectral function $\mathscr{C}_k(t)$ – or $V_k(t)$ – form a time distribution which is in synchronism with the meter of the music, because of the way in which these spectral functions were constructed. But $\mathscr{C}_k(t)$ – or $V_k(t)$ – is not in phase with the meter. In other words, if we listen simultaneously to the musical composition and to the tapping produced by the function $\mathscr{C}_k(t)$, we notice that the tapping sounds are somewhat delayed from the actual meter beat that came from the music. The delay is caused basically by the time required to perform the smoothing operation described in section 3.32 (Quarterly Progress Report, July 15, 1957, p. 96). To correct this phase difference, we can pass the spectral function $\mathscr{C}_k(t)$ through a delaying circuit to delay this function further, until it comes in phase with the bar immediately behind the bar from which the tapping was extracted. The delaying circuit can be a monostable multivibrator followed by differentiator and clipper circuits.

In this way, we obtain a new spectral function $\overline{V}_k(t)$ which is in synchronism and in phase with the original meter of the music. A schematic representation of this process is indicated in Fig. X-4.

3.50 Hold and Stop of Tapping

In section 3.42 we introduced relaxation oscillators that are controlled by the spectral function N_q. The oscillators are used for two purposes: to increase the probability that the rhythmic distribution function N_q corresponds to the rhythm of the meter; and to hold the beat when the music partially stops during the one- or two-bar duration when such a partial stop is called for in the script. But we also want to stop the tapping automatically when the music finally ends. (We consider the music "ended" when it stops for four or more bars.) This condition is not satisfied by the previous arrangement because no device was introduced to stop the oscillations conveniently. We want to describe a method to achieve this last objective automatically. This is the purpose of the envelope function $\mu_k(t)$, which we have not used thus far. (See sec. 3.32, Quarterly Progress Report, July 15, 1957.)

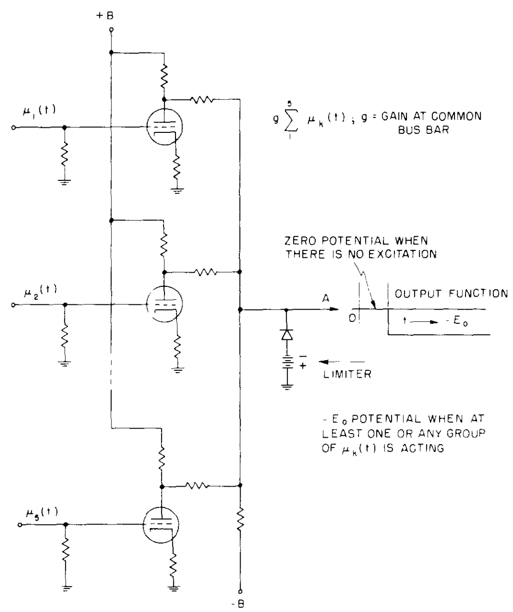

Fig. X-5. Mixing of $\mu_k(t)$ and limitation of output to almost constant potential E_o when signals are acting, and to zero potential in the absence of all signals. Note the effect of the condenser in holding E_o.

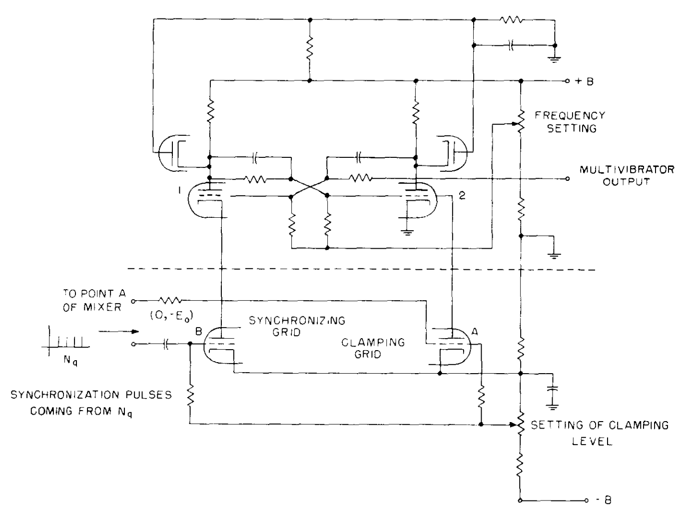

Fig. X-6. Tubes A and B are so biased that tubes 1 and 2 are both cut off when the potential E_o (here, E_o because of the type of circuit used) is applied to grid A. Tube B conducts slightly instantaneously at each pulse in N_q. When the mixer output is zero (no signal), tube A conducts and automatically clamps the grid of tube 2.

3.51 The oscillations of the relaxation oscillators described in section 3.42 must be stopped when every one of the envelope functions $\mu_k(t)$, k = 1, 2, 3, 4, 5 is zero for a time longer than 4 or more bar durations. The oscillators must keep on running when at least one or any group of $\mu_k(t)$ is not zero, or when all are zero, during a time interval that is less than a 3-bar duration. To attain this purpose, the five envelope functions $\mu_k(t)$ are first added by means of a mixer and then are clipped by a limiter. The mixer-limiter arrangement is designed so that the common bus-bar potential to ground is zero in the absence of every one of the excitation functions $\mu_k(t)$. When one or more functions $\mu_k(t)$ are acting, the common bus-bar potential to ground has a constant value of E_o independently of the number of acting envelope functions. The limiter was introduced to perform this last operation. The value of E_o is selected to be somewhat smaller than the value produced by the smaller $\mu_k(t)$ function when it acts alone. See Fig. X-5. Now, the oscillators must stop, or keep on running, if there is, or is not, a bus-bar potential E_o.

A condenser of appropriate capacitance value may be connected in parallel with the bus bar. The effect of this condenser is to hold the bus-bar potential for a short time in the absence of $\mu_k(t)$ signals.

Note that when the music stops the envelope functions do not go to zero instantaneously. They die out in accordance with the time constants of the smoothing circuit, as in Fig. X-2. Therefore, the oscillators hold for short musical stops, but will stop after several bar durations when the music ends.

3.52 There are, of course, several types of relaxation oscillators that we can use for our purpose. One type is a simple saw-tooth generator tube; for example, a double-grid thyratron. The spectral function N_q acts in one grid, and the E_o acts as bias in the other. Because of the effect of the temperature characteristic of gas tubes, they were found unreliable for use in rhythm extraction. It was found that a somewhat complicated free-running multivibrator is preferable, particularly the one represented in Fig. X-6 because we can force the circuit to start all of the oscillations with the same plate and thus avoid trouble in rephasing the meter when the music stops for two bars. Fig. X-6 shows the general wiring diagram of a convenient multivibrator, and indicates its control for starting and synchronizing the multivibrator, and its clamping action in the absence of signal.

3.6 Sequence of the Final Operations

Three operations are required to extract the rhythm of a musical composition. The first deals with the operations in each channel in order to generate the set of spectral distribution functions associated with each channel. The second deals with the proper manipulation of the spectral distribution functions of the different channels which tends

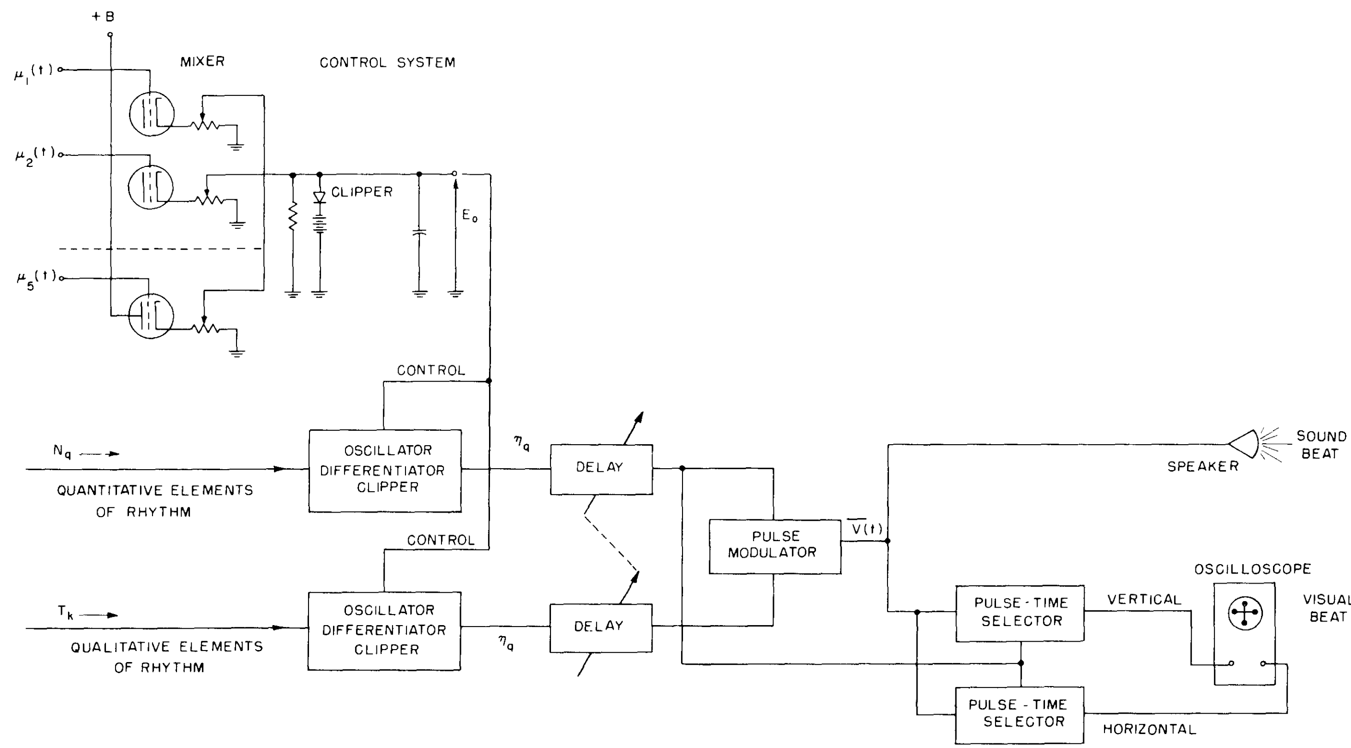

Fig. X-7. Block diagram of the final sections of the system designed to extract the rhythm of a musical composition. Internal circuitry and other details are omitted because they are well-known standard components.

to produce the quantitative and qualitative elements of the rhythm and to increase the certainty that these spectral distribution functions are directly generated by the rhythm. The third deals with the junction of the qualitative and quantitative elements previously extracted in order to accomplish:

a. the synchronous mosaic of the rhythm
b. the proper phasing of the final spectral functions
c. the hold and stop of the mosaic for a few bars in partial stops, or in the final stop of the musical composition
d. auxiliary circuits that are necessary to produce a sound beat or a visual beat on the screen of an oscilloscope.

The first operation was presented in the Quarterly Progress Report of July 15, 1957. In the present report we have mentioned the steps corresponding to the second and third operations. A condensed version of the final procedure is indicated in the block diagram of Fig. X-7. For simplicity of presentation, we have omitted irrelevant circuitry and other details, since the important objective of the report is to indicate the general possibility of extracting the musical rhythm. (The required circuits and components, such as oscillator, differentiator, clipper, delays, and so forth, are so well known that a repetition here is unnecessary.)

3.61 The methods and pertinent operations described in this report are, in fact, more complicated than the actual procedures needed to attain rhythm extraction, particularly in connection with the number of channels. In fact, a machine constructed in the Laboratory to check the ideas presented here uses only one channel, one squaring device, and one phase-correcting network. The multiple-channel presentation was intended to give the reader a better grasp of the problem and to facilitate the understanding of some delicate steps.

3.62 We close this report with two remarks: First, the basic feature of the methods described here is to extract the rhythm of a musical number in the first few, say three or four, bars of the meter, as required by condition b in the Quarterly Progress Report of July 15, 1957, page 94. This is indeed a very strong requirement. It is difficult to satisfy it by other methods of signal separation, particularly by linear processes. Considering such a strong requirement, the method described here is relatively simple. Second, there are possible modifications which would simplify the method described here. A formal and very powerful procedure is the use of group theory, which is, in fact, the proper mathematical approach. The application of the group-theory method is being studied in the general program of our work.

M. V. Cerrillo