An Experimental Polytonic Signaling System

By C. A. LOVELL, J. H. McGUIGAN and O. J. MURPHY

An experimental high-speed signaling system capable of signaling at the rate of 100 decimal digits per second over a variety of existing Bell System transmission facilities is described. This system uses several frequencies in combinations as its code elements. Because the manner of discriminating between different frequencies is significantly different from present-day practice, the system has been called polytonic *to distinguish it from existing multifrequency systems. Signal separation in this system depends not on the steady-state response of filters, but on the transient response of simple reactive networks. Discrimination is enhanced by the use of rather precise timing in the signal generating mechanism and somewhat less precise timing in the receiver. A second system, capable of signaling at the rate of 300 decimal digits per second over a more restricted range of transmission facilities, is also briefly discussed.*

1.0 INTRODUCTION

The need in future telephone systems for a signaling method of significantly higher speed that that provided by the familiar dial or even by the intertoll multifrequency system[1] has led to a research program in high speed signaling. The program has resulted in laboratory models of several experimental systems based on different principles. This paper describes one of these systems, based on the concept of orthogonal functions. Several different frequencies are used, in combinations, to transmit signal information. The means of generating and recognizing these frequencies represent a departure from present practice; in order to distinguish this experimental system from the multifrequency system, the term *polytonic* was coined. It is hoped that this paper provides an interesting illustration of the way a mathematical conception may be transformed, through successive stages, to a physical system.

---

[1] C. A. Dahlbom, A. W. Horton and D. L. Moody, Application of Multifrequency Pulsing in Switching, Trans. A.I.E.E., **68,** pp. 392–396, 1949.

Telephone circuits fall naturally into two classes, local and toll. The bandwidths of these two classes of facilities and the types of distortion which they exhibit are so different that it was found expedient to build two separate embodiments of the polytonic system. One was designed to work in the presence of the characteristic forms of distortion found in the toll plant, while the other was built to take advantage of the higher speed permitted by the nature of the local plant.

Most of the following discussion applies equally well to both signaling systems. Throughout the body of the paper the main emphasis will be placed on the toll system; except for occasional mention, the local system will be treated in a separate section. The discussion will begin with the theoretical background for the signaling method under consideration and proceed to a detailed examination of the signals representing the code elements. Problems encountered in the design of practical systems will then be taken up, including certain arbitrary requirements of coding and signal content. This will be followed by a description of some of the circuitry evolved to meet the requirements and an operational description of the toll-signaling system which was built to assess the capabilities of this approach to the signaling problem. A brief account of tests of this system over certain Bell System toll facilities will be given. The discussion will conclude with some remarks about circuitry devised for the local-signaling embodiment which was capable of achieving a substantially higher signaling speed.

2.0 THEORETICAL BACKGROUND

The basis for the signaling systems about to be described lies in the concept of orthogonality as applied to certain classes of mathematical functions. A set of normalized time-functions, $\Phi_1(t)$, $\Phi_2(t) \cdots \Phi_n(t)$ are said to be orthogonal over the time-interval T if

$$\begin{aligned} \int_0^T \Phi_i(t) \cdot \Phi_j(t)\, dt &= 0 \qquad \text{for } i \neq j \\ &= 1 \qquad \text{for } i = j \end{aligned} \tag{1}$$

Sinusoids of various frequencies provide a familiar example of one such class of functions; the period of orthogonality, T, may be any value which embraces an integral number of periods of each of the frequencies under consideration. It was recognized that direct use of this property could be made in separating, at the receiving point, signals made up of combinations of such functions, without employing the steady-state selective properties of reactive filter networks.

As an illustration of how such a signaling system might be mechanized, consider a group of oscillators whose frequencies are multiples of some common base frequency. The period of this frequency then defines one possible value of the period of orthogonality, T, which will satisfy the requirements for all frequencies in the group. Now suppose that some coding procedure is set up which involves sending a tone-spurt of one or more of these frequencies over some common transmission medium to a distant point to represent each of a number of desired signal elements or characters. If the tone-spurts have a duration equal to or greater than the period T, it will be possible to distinguish, at the receiving end, which frequencies are present, by application of the above principle. This may be accomplished by having a group of oscillators at the receiving end whose frequencies are identical with those at the sending end and providing a number of parallel channels for the received signal, equal to the number of oscillators. In each of these channels, the received signal is multiplied, in the mathematical sense, with the aid of suitable circuitry, by the output of the corresponding oscillator and the resulting product integrated for a period of duration T. An examination of the several channel integrators, at the expiration of the period T, reveals a "unit" output from those whose local frequency corresponds to a component of the received signal, while those corresponding to frequency components not present in the received signal show "zero" output. The terms "unit" and "zero" are employed here to represent the ideal case; in practice, there is found, in a properly designed circuit, a substantial value of output representing correspondence and little or no output for a lack of correspondence. After the integrator outputs have been examined, which is done virtually instantaneously on a sampling basis, the electrical integrators are reset to zero, thus clearing the receiver for the reception and recognition of the next tone-spurt from the sending end.

The system just outlined represents a mechanization of the classical concept of orthogonality for signaling purposes. It has been suggested[2] that an extension of the concept of orthogonality can be used to define a new set of properties which seems, in some respects, more useful than those defined by (1). Consider a set of normalized functions $\theta_1(t)$, $\theta_2(t)$ $\cdots$ $\theta_n(t)$ which, over the time interval T, satisfy the following conditions:

$$\begin{aligned} \int_0^T \theta_i(\lambda)\cdot\theta_j(T-\lambda)\,d\lambda &= 0 \qquad \text{for } i \neq j \\ &= 1 \qquad \text{for } i = j \end{aligned} \tag{2}$$

---

[2] This suggestion was made by R. B. Blackman of Bell Telephone Laboratories.

The practical importance, in the generation and reception of electrical signals, of functions which satisfy (2) is that the defining integral describes the transient response of a network; that is to say, if $\theta_i(t)$ is the response of a network to an impulse time-function, then the integral represents the response of the network when it is acted upon by an applied signal of the form $\theta_j(t)$. Thus the inherent factors which govern the instantaneous transient response of passive electrical networks may be utilized to perform the required mathematical processes of multiplication and integration without recourse to special functional circuits for the purpose.

The following description will deal in some detail with signaling arrangements exploiting these properties of reactive networks, in which the signal generation and recognition is performed on the basis of the instantaneous responses at particular times rather than on the basis of a steady-state frequency-response, as in more conventional systems.

3.0 POLYTONIC SIGNALS

The elements of the signals used in the polytonic signaling system are damped sine waves, such as that shown in Fig. 1(a), which are characterized by the expression:

$$E_m = E_0 e^{-at/T} \sin m\pi t/T$$

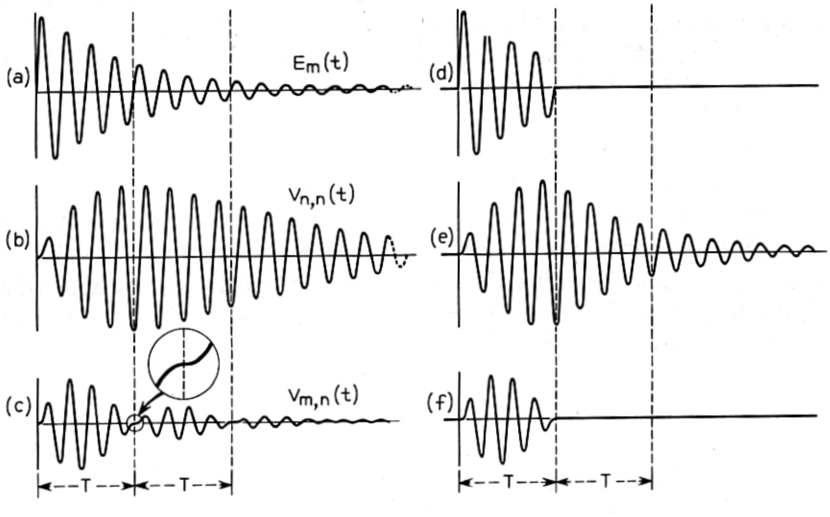

Fig. 1 — Polytonic signal wave forms and network responses.

The frequency variable is expressed in the form shown to emphasize the fact that the frequencies in the set are integral multiples of a common base-frequency, $1/2T$. If the multiplier m is restricted to successive odd integers or successive even integers, as is actually done in the system to be described, the frequency separation between adjacent members of the set is then $1/T$. For any integral value of m, this expression satisfies the conditions of both (1) and (2).

This type of response can be obtained, for example, by applying a step-function of voltage to a series circuit of R, L, and C elements and observing the voltage across the resistive element, which is proportional to the current. Circuits of this kind are, in fact, used to generate the polytonic signals at the transmitting end of the system.

The discriminating devices at the receiving end may be similar simple circuits of R, L, and C elements, matched to the respective generator circuits in both frequency and damping. It is desired that the impulsive voltage transfer-function of the receiving networks match the signals generated by the corresponding transmitting networks. This correspondence may be obtained by taking the voltage across the condenser element, which is proportional to the integral of the current, as the output of the network.

We shall now proceed to determine the response of one of these receiving networks both to a transmitted signal which it matches and to one which it does not match. If we consider the signal

$$E_m(t) = E_0 e^{-at/T} \sin m\pi t/T$$

applied to a network whose voltage transfer function is

$$G_n(t) = K e^{-at/T} \sin n\pi t/T$$

we have, for the response

$$V_{mn}(t) = K E_0 e^{-at/T} \int_0^t \sin (m\pi\lambda/T) \sin [n\pi(t - \lambda)/T] \, d\lambda \qquad (3)$$

When evaluated, this yields

$$V_{mn}(t) = K E_0 \frac{T e^{-at/T}}{\pi(m^2 - n^2)} (m \sin n\pi t/T - n \sin m\pi t/T) \text{ for } m \neq n \qquad (4)$$

and

$$V_{nn}(t) = K E_0 \frac{T e^{-at/T}}{2n\pi} [\sin n\pi t/T - (n\pi t/T) \cos n\pi t/T] \text{ for } m = n \qquad (5)$$

When equations (4) and (5) are evaluated for $t = T$ we find

$$\begin{aligned} V_{mn}(t) &= 0 \\ V_{nn}(t) &= KE_0(-1)^{n+1}(T/2)e^{-a} \end{aligned} \tag{6}$$

If a suitable normalizing factor is applied to the second expression, this pair is seen to fulfill the conditions of (2).

Thus, if a receiving network is excited only by the signal element, or damped sine wave like that shown in Fig. 1(a), to which it is matched, the voltage across the condenser element of the network is given by (5) and will have the form shown in Fig. 1(b); this will hereafter be referred to as a "desired" response. A desired response is seen to consist of an oscillatory voltage which starts with zero amplitude and zero slope and whose successive peak values are contained within an envelope which builds up to a broad maximum and then decays slowly. One of the peaks is centered at time T; by suitable choice of damping constant, ($a = 1$) the maximum value of the envelope may be made to coincide with this peak, and this is the condition illustrated in the figure.

If a receiving network is excited only by a signal element to which it is not matched, its response, in terms of condenser voltage, is given by (4) and depends on the relationship, in frequency spacing, between the frequency-of-match and that at which it is excited. This will be regarded as an "undesired" response; Fig. 1(c) shows the undesired response when the exciting frequency is the next higher frequency in the set. This response will be seen to consist of an oscillating voltage which starts with zero amplitude and zero slope, builds up to a maximum of amplitude, and then returns to zero amplitude and zero slope at time T. Between successive multiples of T, the envelope of the undesired response exhibits maxima of successively smaller amplitude. More remote signal elements produce undesired responses with zeroes at submultiples of T, but integral multiples of T are the only times at which all undesired responses are simultaneously zero in both slope and amplitude. For stimuli of equal strength, the maximum amplitude of the undesired response is always considerably less than that of the desired response.

When the signal applied to a receiving network contains both desired and undesired components, the response is much more complex in appearance than those shown in Figs. 1(b) and 1(c) which represent excitation by single signal elements. The response is, in fact, a linear combination of components of the two classes described, each of which retains its individual properties. Thus if the voltage outputs of the receiving networks are sampled instantaneously, at time T, all networks which have found their counterpart signal element in the complex received signal will exhibit an output representing the peak value of an oscillatory wave of corresponding frequency, while those which have not found their counterpart will show, ideally, no output at this instant even though they may have shown a considerable oscillatory voltage before this time and may show an appreciable voltage at a later time.

4.0 CONSIDERATIONS IN THE DESIGN OF A PRACTICAL SYSTEM

The design of a practical signaling system involves considerations not covered by the theory of the signaling method used. Some of these considerations are entirely under the designer's control, while others are influenced by the external conditions under which the system must work. The external influences will be discussed first, and then the design parameters which the designer can choose freely.

4.1 Timing

At the outset the mathematical theory considered the response of receiving networks to signals sent out by transmitting networks and found that at a particular instant the desired response was large and the undesired response was zero. Mathematically it is easy to examine a function at a particular instant; from a circuit designer's viewpoint it is usually easy to sample a network response at a particular instant. However, where there is interposed between the transmitter and the receiver a tra[illegible]ine facility having an unknown delay and a restricted transmission [illegible] it is not easy to know the particular instant at which to take the sample. The starting time, from which the interval must be measured, may not be very sharply defined at the receiving end, and yet a fairly high degree of precision in timing is obviously necessary. If, for example, an error of one-quarter period of the oscillation frequency is made in taking the instantaneous sample of a desired response, a zero output will be obtained instead of the legitimate peak value. Extending the sampling period, to avoid this difficulty, would bring in appreciable energy from the undesired responses, which are zero only at the critical instant. There are still other difficulties, related to timing, which will be taken up later in connection with the design of a specific system, but enough has already been said to point up the desirability of some artifice which would retain the advantages of the orthogonal approach to the signaling problem while either relaxing the stringency of the timing requirements or shifting them from the receiver to some other part of the system, such as the transmitter, where they might be easier to meet.

Several proposals were considered which effected the desired easing of the timing requirement in greater or lesser degree, but the one finally adopted did, in effect, shift the timing mechanism to the transmitter. This arrangement produced a generated signal which no longer perfectly fulfilled the postulates of orthogonal theory, but whose deviations from perfection were found to be small and relatively unimportant in their effect on the ability of the signal receiver to discriminate between signal elements. The justification for, and utility of, this modification will become apparent upon further examination of the behavior of the networks of the receiver.

At the time T, or any multiple thereof, it may be seen that there is no energy stored in the receiving networks exhibiting undesired responses: the zero of voltage means that there is no energy in the condensers (neglecting dielectric polarization) while the zero of slope means that there is no current flowing and hence no energy in the coils. If, then, instead of persisting indefinitely, as was previously assumed, the energizing signal is removed at this time, no further energy will be introduced into the networks. Hence no voltage will appear across the condensers and the undesired response will remain zero thereafter. A network exhibiting a desired response, on the other hand, has considerable stored energy; its condenser voltage, indeed, is at a peak at this critical time. If the source of energy is removed at time T, an oscillating voltage will continue to appear across the condenser, decaying in amplitude somewhat more rapidly than was formerly the case, as the stored energy is dissipated and not replaced. Thus, by removing the source of energy at t[illegible] T it is theoretically possible to obtain infinite discrimination, in th[illegible]sence of noise, between desired and undesired responses *at any time thereafter*. The effect of removing the source of energy may conveniently be achieved by abruptly stopping the decaying oscillations of the transmitting network at time T after their initiation, as shown in Fig. 1(d), instead of allowing them to persist indefinitely. The behavior of the receiving circuits in the cases of wanted and unwanted responses are shown in Figs. 1(e) and 1(f). It is relatively easy to accomplish this timing job since the initiating pulse for the transmitting networks has a steep wave front which defines the starting time precisely.

The timing requirement at the receiver is then merely that the sampling of the network condenser-voltage occur *after* the expiration of a time T measured from the starting time of the signal, and a fairly generous margin for time uncertainty can be provided. The sampling period itself may now be extended, since the undesired responses will remain zero instead of building up again after the null point, and can advantageously be chosen to be of such length that one or more peaks of desired response are embraced, thus avoiding the possibility of missing a single peak through mis-timing an instantaneous sample.

The means employed for arresting the oscillations of the transmitting networks at the critical time, and thus producing a signal of controlled length, will be taken up in detail in the section devoted to the design of the signal transmitter for a practical system.

4.2 Transmission Facilities

Any signaling system intended for telephone use must be capable of working over the wide variety of transmission facilities existing in the telephone plant. These facilities all fall short of perfect transmission in a variety of ways which do not seriously impair the intelligibility of speech but which do distort signals of all kinds. Non-human signal receivers are limited in complexity both by cost considerations and by their designer's limited skill, so they are less tolerant of distortion than is the ear.

Telephone circuits, as indicated in the introductory section, fall naturally into two classes, local and toll. Subscriber lines and local trunks usually have moderate attenuation and low noise, and their bandwidth restriction appears as a gradual roll-off accompanied by little phase distortion in the useful transmission band. Toll trunks, on the other hand, may have a wide range of both attenuation and noise. Bandwidth is usually less than on local trunks, although carrier circuits may show greater bandwidth than some small-gauge local trunks. Both voice-frequency and carrier toll trunks have sharper cutoffs than local trunks, with attendant phase distortion. In long loaded and repeatered circuits the phase distortion near the edges of the band accumulates to such an extent that it is manifest as an appreciable difference in transmission time, spoken of as delay distortion. Carrier circuits exhibit a characteristic kind of distortion caused by lack of synchronism between the carrier oscillators at the two ends of the circuit. This type of distortion is held within limits which make it unnoticeable to the ear but which may still render an arbitrary signal shape unrecognizable. The consequences of these various types of distortion seriously complicate the receiver design problem.

4.3 Codes

On a two-amplitude or on-off basis, at least four frequencies are required to provide a distinct combination for each of the ten decimal digits. If a fifth frequency is provided, the familiar two-out-of-five code with its self-checking feature can be used. In a polytonic signaling system operating over transmission facilities of limited bandwidth, the use of the five-frequency code entails a slight loss of signaling speed: since the bandwidth must be divided into five channels rather than four, the frequencies are closer together and therefore the interval T is longer. In addition, the fifth frequency requires extra equipment for its implementation. These costs were felt to be small compared to the advantage of the self-checking code, so the two-out-of-five code was chosen.

4.4 *Frequency Choice*

One of the first decisions to be made in designing a system of this kind is to choose the frequencies which will be used. The frequencies chosen for the experimental toll signaling system were 500, 700, 900, 1,100 and 1,300 cps. These frequencies lie comfortably within the pass-band of all transmission facilities over which the system might be required to operate.

4.5 *Signaling Speed*

Since T is the reciprocal of the frequency separation, the frequencies chosen give a value for T of 5 milliseconds. It was decided to take another 5 milliseconds, after the time T for each digit, to allow for detection, network deenergization, and other system operations which were required between digits. Each digit thus occupied a time of 10 milliseconds, and the signaling rate was 100 digits per second.

4.6 *Mode of Operation*

The signaling speed of one hundred decimal digits per second is much higher than the rate at which a human operator can supply the information to be transmitted. To make efficient use of this speed, therefore, the input information must be accumulated in some kind of temporary storage mechanism, or register, from which it can be spilled at the rate for which the transmitter is designed. One signal transmitter of a high-speed system would send out completed input information derived from any of a number of registers, so that the time required to set up the information, in individual instances, would not constitute a limiting factor. Similarly, one receiver might accept messages from any one of a number of transmitters, thus further exploiting whatever inherent speed capabilities the system might possess.

If the self-checking features of the code, discussed in an earlier section, are to show their full advantage, the system must be arranged to send the same signal information again, upon demand, whenever an implausible received combination causes rejection of the information received. The easiest way to do this, and the one adopted for this system, is to send the complete signal information or "message" repetitively, whether repetition is needed or not, until an acknowledgment signal is obtained from the receiver that a complete message, plausible in all digits, has been received. Reception of this signal, at the transmitting end, would then release the transmitter.

Because of the repetitive character of the signal transmissions, and the possibility of a signal receiver "coming in" on a line from a transmitter in the middle of the message, some internal indication is required in the message itself to mark the end of one iteration and the beginning of the next; this subject is discussed further in Section 5.22.

For the system to be described, the message is defined to consist of any arbitrary succession of eight decimal digits, representing a typical subscriber's number with office code, line number, and party designation. The desired message is completely set up before any part of it is transmitted; and after transmission is begun it is sent, digit by digit, repetitively, with internal indication of sequence beginning, until the receiver acknowledges reception or until, in the case of laboratory tests, it is desired to send some other message.

5.0 DESIGN FEATURES OF A PRACTICAL SYSTEM

5.1 The Signal Transmitter

A block diagram of a signal transmitter conforming to the mode of operation outlined in Section 4.6 is shown in Fig. 2. It begins with a source of frequency which determines the rate at which the digit signals are produced, and which is shown, for convenience, as a multivibrator. The onset of one part of the cycle of this multivibrator is used for control of the stepping process, as shown. The other part, which must be of length *T*, times the action of the keying circuit to shock the tuned circuits and thus produce the damped sine wave signal elements. The stepper-distributor conveys the pulses originating in the keying circuit to the digit selector switches in succession. It is represented schematically as a rotating switch and might indeed be such a switch in slow-speed systems. In the system herein described, however, it was made up of a gas-tube stepping chain actuating relatively high-speed mercury-contact relays.

The digit selection switches are also shown schematically as rotary switches and they were thus constituted in the experimental system, two-deck switches serving to convey the distributed pulses to any selected combination of two out of the five tuned circuits. In a working system these switches would probably be replaced by aggregations of relays under control of push-button keysets, or some other source of signal information.

The tuned circuits are five in number and are used in pairs to provide the two-out-of-five code elements required. A small resistance was placed in the common lead of all five circuits and the voltage drop across it used as the signal output.

The keying circuit is perhaps the item of greatest interest in the signal transmitter. Figs. 3 and 4 show circuits implementing two methods, which will be described, for generating the controlled-length polytonic signals previously discussed. Both methods have been used and neither has a clear-cut advantage over the other.

The keying circuit has two duties to perform: it must initiate shock-oscillations in selected tuned circuits and it must arrest these oscillations after an integral number of half-cycles occupying the period T. It is also

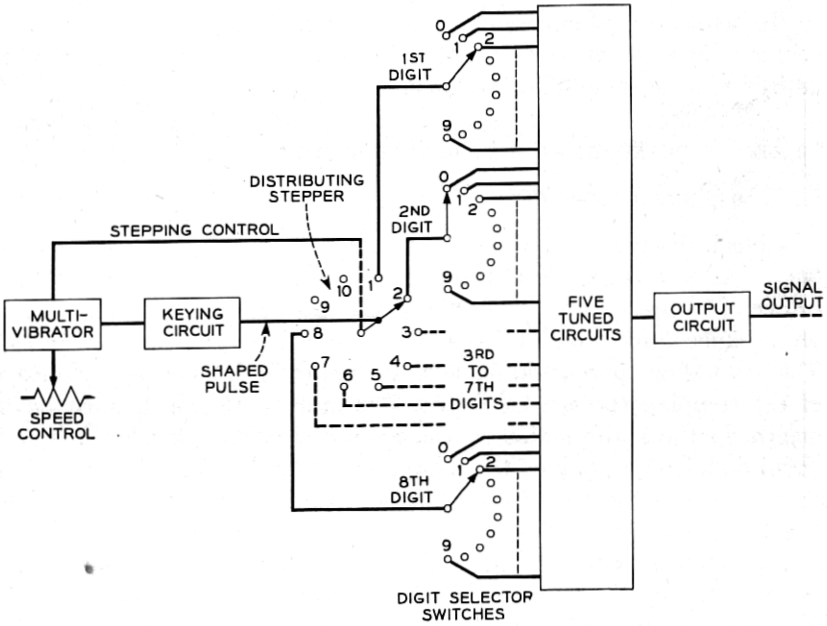

Fig. 2 — Schematic diagram of signal transmitter.

highly desirable, to facilitate the generation of a succession of signals, that the shock excitation voltage supplied by the keying circuit have the same value near the end of a digit period as it had just before the steep wave front occurred, so that a complete cycle is traversed for each signal generated.

One method of arresting the transmitter-network oscillations in an almost transient-free manner at the end of an integral number of half-cycles is to introduce, in series with the circuit, a resistor having a value equal to or somewhat greater than the critical damping resistance. This resistor must be removed, of course, before the next signal-producing shock is applied. If such a damping resistor is introduced at time T, it is then possible to return the exciting voltage to the value from whence it started, by means of an oppositely directed step, without occasioning any further oscillations. An exponentially-damped transient will, however, ensue.

A more nearly transient-free signal may be obtained by employing a shocking pulse having a steeply rising wave front, followed by a smooth decay which will reduce its amplitude to a negligible value at time T. Use of a pulse such as this, instead of the ideal step-function, hitherto contemplated as the exciting voltage, will introduce some anomalous frequency components into the spectrum of the transmitted signal, but these have been found to have little influence on the behavior of the receiving networks.

A keying circuit providing the features of a shaped pulse for excitation

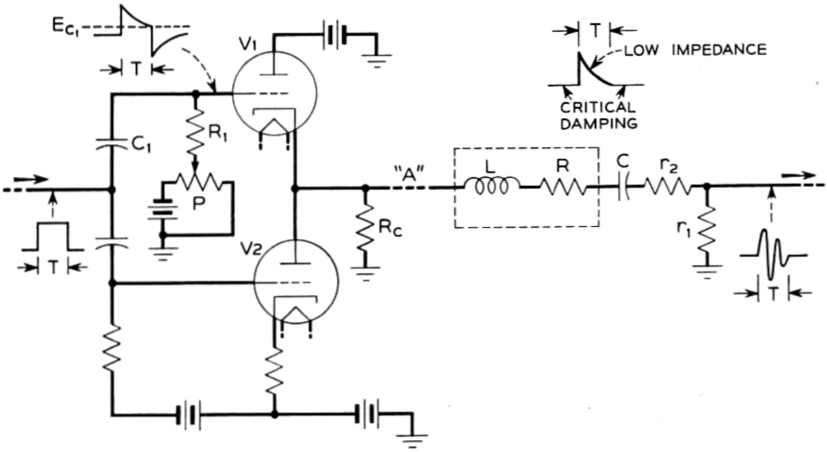

Fig. 3 — A method for generating a controlled-length polytonic signal.

and critical-resistance damping for arresting the oscillations is shown in Fig. 3.

In this arrangement a rectangular voltage pulse of the critical length is applied to tubes V1 and V2. Both tubes are normally cut off. Elements R_1 and C_1 partially differentiate the applied signal at the grid of V1, which is a cathode follower, causing the cathode voltage to rise suddenly and then subside toward ground; the output impedance of the tube, at the cathode, is low during this time, owing to the ordinary dynamic feedback action of the tube. Tube *V2* is also brought into conduction for the period *T* through coupling elements of long time-constant; this tube supplies the current required to provide vigorous dynamic action in tube *V1* as its cathode approaches ground potential and also provides a conductive path for the negative lobes of the oscillatory current in the tuned circuit. The resistor R_c, having a value approximately that of the critical damping resistance for the tuned circuit, is connected from the cathode of *V1* to ground but plays little part in the action just described, since it is shunted by the low dynamic impedance of *V1*. By properly proportioning the circuit elements and adjusting the grid voltage of *V1* by means of control *P*, the cathode of *V1* may be made to arrive at a voltage of approximately zero with respect to ground at the expiration of the period *T*. The trailing edge of the input pulse then restores both tubes to the cutoff condition, the impedance from cathode to ground of *V1* is that of R_c, and the tuned circuit is critically damped by this high impedance. During the period *T* the total resistance in series with the tuned circuit is the sum of that of the cathode-follower driver, that of the physical coil itself and that of the two resistors r_1 and r_2. These two latter resistors are ordinarily both small; r_1 serves as a means of obtaining an output voltage proportional to the oscillatory current while r_2 serves as a padding re-

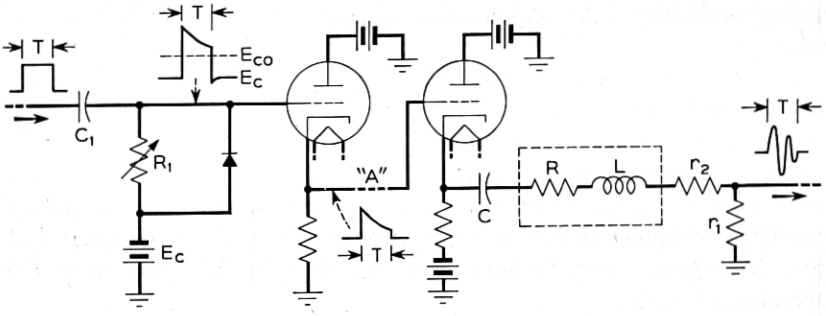

Fig. 4 — Another method for generating a controlled-length polytonic signal.

sistor to adjust the total to the value required for the prescribed damping constant. The distributing stepper and digit-selector switches are interposed at the point marked A.

Another way of terminating the transmitter-network oscillations in a transient-free manner is to apply a second step-function of correct magnitude and polarity to the tuned circuit at the exact completion of any integral number of half-cycles of the oscillation, so that a new oscillatory wave, due to the second shock, will exactly cancel the unwanted remaining portion of the original wave. The number of half-cycles to be chosen is, in this instance, the number occurring within the time interval T. If an odd number of half-cycles have occurred, the terminating shock must have the same polarity as the initial shock, whereas if an even number have occurred a terminating shock of opposite polarity is required. In any case, the magnitude of the terminating shock must be less than that of the initiating shock in exact proportion to the decay of the oscillations. A circuit working on this basis, for an even number of half-cycles within the period T, is shown in Fig. 4.

Here the rectangular input pulse is applied to a cathode-follower through the coupling circuit C_1R_1 whose time-constant is adjusted to match the decay rate of the tuned circuit oscillations and which thereby produces the desired ratio between leading-edge step and trailing-edge step. The diode shunting R_1 serves to discharge C_1 quickly at the termination of the input pulse, which is supplied from a low impedance source. The output of the cathode-follower stage is, then, a voltage pulse of the correct shape which begins and ends at essentially ground potential. This pulse is conveyed to another cathode follower which is arranged to draw a substantial current, and therefore to exhibit a low output impedance, at all times; this tube drives the tuned circuits.

The distributing stepper and digit-selector switches are interposed at the point marked A, as before.

This type of signal generation was employed for the second embodiment of the polytonic system, which was built to attain still higher signaling speeds over local-type facilities. A different set of frequencies from those listed in section 4.4 were chosen for this case. These frequencies, 800, 1,200, 1,600, 2,000 and 2,400 cps, provided an even number of half-waves within the period T.

5.2 The Signal Receiver

5.21 Envelope Detector

It was desired to build the system around a detector which did not have severe timing requirements and which would be capable of withstanding some distortion of the received signals. The detector devised to meet these requirements based its discrimination essentially on the envelope of the tuned circuit response. In the case of undistorted received signals, after time T the envelope of the desired response is a decaying exponential of appreciable magnitude, while the envelope of the undesired response is zero. It was found experimentally that when the received signals were subjected to the kind of distortion produced by transmission over a carrier system with non-synchronous carrier oscillators, the envelope of the desired response was substantially unchanged while the envelopes of undesired responses, though no longer zero, were still much smaller than those of desired responses. When the signals were transmitted over lines with delay distortion, so that one component arrived a time-interval Δ later than the other, the envelopes of undesired responses were observed to be substantially zero after time $(T + \Delta)$, measured from the arrival of the earlier component, while again the envelopes of the desired responses were little changed.

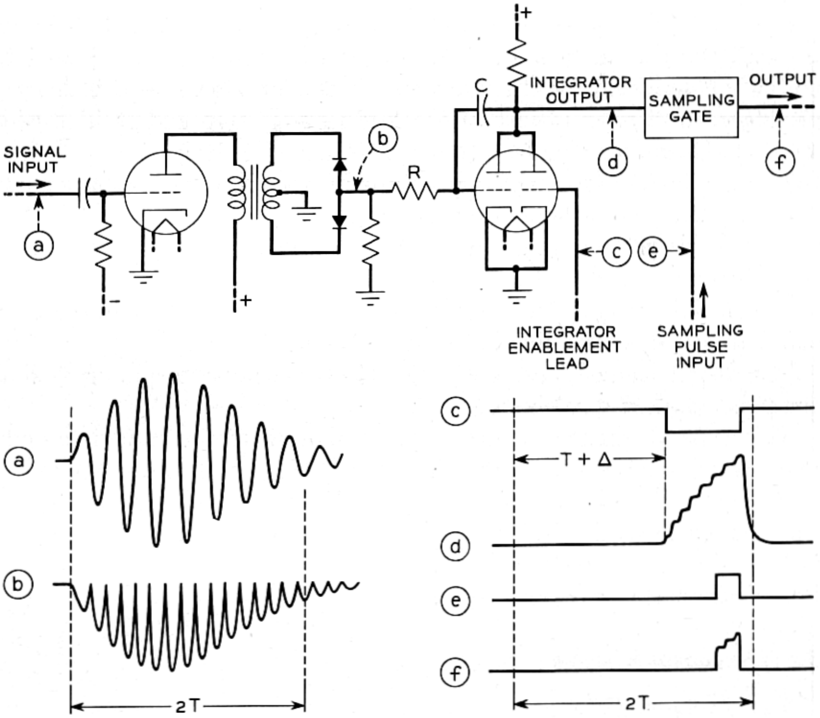

Fig. 5 — Circuit and wave-shapes of receiver channel integrator for envelope detection.

A common method of deriving the envelope of an oscillating signal is to rectify the signal and filter out the carrier and higher frequency components. This method is not feasible here because these envelopes contain significant frequency components at frequencies comparable to the carrier. The approach which was actually used is illustrated in Fig. 5, and consisted in full-wave rectifying the network responses to eliminate their oscillatory character and then integrating over a short period to eliminate the fluctuating nature. This integration should not be confused with that mentioned in the discussion of orthogonal functions. Its only purpose is to make the signal channel outputs monotonic, and it need not, therefore, be mathematically exact. The output voltage from a receiving network which, for a desired response, will have the form shown in line A, is applied to the input of the detector circuit. Full-wave rectification produces the wave shape as shown at B. Elements R and C, together with the left-hand half of the double-triode constitute an integrator of the Blumlein type. The plate of this tube, however, is connected to the plate of the right-hand triode which acts as a clamp during the early part of the digit interval, thereby preventing any substantial integrating action. At time $T + \Delta$ after the beginning of the input signal, the clamp tube is cut off by a negative pulse, derived from a timer circuit and applied to the Integrator Enablement lead; integration then proceeds as indicated in line D of the figure. The integrator output is sent to a sampling gate which blocks it except during a short period when a sampling pulse, also derived from the timer circuit, occurs. The sampling procedure is employed to present the signals from all five channels to a two-out-of-five code-checking circuit at the same instant; without this synchronizing feature the finite slopes of the integrator outputs might cause them to reach threshold value at different times and thereby confuse the checking circuit.

The wave-shapes shown in Fig. 5 refer to a desired response; in the case of an undesired response the input signal is substantially zero during the integration period and no output is produced.

With the signal parameters chosen, the total digit interval was $2T$ or 10 milliseconds. Of this time, 5 milliseconds was allotted for the action of the receiving networks, 1.3 milliseconds for Δ, the margin for timing error and transmission delay-distortion, and 3.2 milliseconds for the integration period. This latter includes the one millisecond occupied by the sampling pulse. Thus 0.5 millisecond remains for resetting the integrators to their initial condition, through the action of the clamp tubes, before the expiration of the 10 millisecond period.

5.22 Start Circuit

It was decided, as a matter of system design, that the transmitter should begin transmitting the desired number as soon as it was set up, without determining whether the receiver was ready to accept the information. The receiver, on the other hand, would begin to consider information received over the line at an unspecified time depending on the completion of its handling of the previous number. This design philosophy leads to the need for a device to which the name *start circuit* has been given. The receiver may find itself beginning to consider an incoming number at a time when some other digit than the first is arriving; it must be able to recognize in some way which digit in the train is the first digit of the number, and to wait until that digit arrives before beginning its operations.

The first digit is marked at the transmitter by the artifice of preceding it by a blank interval: the digits follow each other at regular intervals except that between the last digit of one iteration of the number and the first digit of the next iteration there is left an interval two digit-periods in length during which no signal is sent out. The blank interval is recognized in the receiver by means of an *R-C* timing circuit: a condenser is connected through a resistor to a high voltage and tends to charge up to the supply voltage. As each digit arrives, the condenser is discharged to zero and begins to charge again. There thus exists across the condenser a sawtooth voltage wave, which is monitored by a threshold circuit. As long as the digits arrive at regular intervals, the condenser voltage never rises beyond a certain maximum value which is less than the firing point of the threshold circuit. However, when the blank interval occurs, the condenser voltage is permitted to rise to about three times its normal maximum value. In so doing, it passes the firing point of the threshold circuit, which delivers a signal used to indicate that the next digit will be the first digit of the transmitted number. Whenever the receiver considers a new number, it waits until a blank interval occurs in the train of digits before beginning its operations.

In addition to recognizing the beginning of the number, it is necessary to recognize the arrival of each digit in turn. The sudden drop in voltage across the start circuit condenser is used to mark the beginning of each digit. This drop marks a well defined instant from which various necessary time intervals previously mentioned, can be measured. It will be recognized that the relationship of this instant to the beginning of the digit is not highly precise. The start circuit is necessarily an amplitude-sensitive device, since it must distinguish between the presence and absence of signals. The sensitivity must be high enough so that the start circuit recognizes weak signals, but not so high as to cause false responses to noise on the line. Signals of the wave form shown in Fig. 1(a) or 1(d) have a definite starting point defined by a discontinuity in the derivative, but this discontinuity will tend to become obscured by transmission over any practical medium. The initial value of the derivative is different for different signaling frequencies, so that the time required to build up to any arbitrarily defined threshold level is not exactly known. It is these considerations, as discussed above in Section 4.1, that makes it impossible to use a detector based on the instantaneous value of the oscillating response of the receiving network. The envelope detector just described, however, is much less time-sensitive and can easily tolerate the lack of precision of the start circuit.

5.23 Receiver Block Diagram

A block diagram of the signaling receiver is given in Fig. 6. It is not proposed to describe the contents of each block in detail, but the sequence of operations in the handling of a received number will be traced out, and the functions of each block will be discussed.

Before describing the block diagram, it is well to point out that the system under discussion was built for purposes of laboratory test rather than practical use and that therefore it contains some features not necessary for plant service and lacks others that would be required. In particular, since in the laboratory it was always known what number was being transmitted, it was possible to include a comparator — a relay circuit which at the end of each receiver operation made a comparison between the transmitted number and the received number and indicated whether the number was correctly registered. Again, since received signals always arrived over the same pair of wires, no circuit was built to

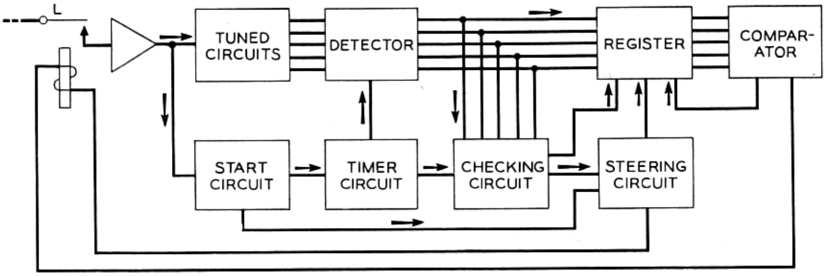

Fig. 6 — Block diagram of signal receiver.

perform the function of connecting the receiver to any one of a number of pairs to receive signals from a variety of sources. In a practical system it would be necessary, as previously mentioned, to send a "dismiss" signal back over the line to cause the transmitter to stop sending when the number was successfully received; this feature was not included here because it was felt to constitute a fairly simple design problem and one that had no special bearing on the particular research problem of interpreting and handling polytonic signals.

The block diagram of the system falls readily into two major divisions. The main detecting and registering functions are performed in the boxes across the top of the diagram, while the lower boxes perform auxiliary control functions.

The sequence of operations begins with the closing of the line relay (L), which connects the line to the input of the receiving amplifier. The L relay, in this laboratory version of the receiver, takes the place of the more elaborate connector circuit which would be required in a practical system with a large number of incoming lines.

When the L relay closes, the received signals, after amplification by the receiving amplifier, are fed to the receiving tuned circuits and to the start circuit. The L relay will close at a random point in the train of signals; one function of the start circuit is to make the register unresponsive to received signals until the blank interval in the train has occurred, thus ensuring that registration of a received number will begin with the first digit of a transmitted sequence.

Another function of the start circuit is to provide a timing reference to mark the beginning of each digit in the received sequence. Starting from this reference, the timer circuit measures off the interval $T + \Delta$ mentioned above and then measures off the integrate period which begins at time $T + \Delta$ and lasts just over three milliseconds.

When a received digit composed of two code elements is applied to all the five receiving tuned circuits, there appear on the five tuned-circuit output leads two desired responses and three undesired responses. These responses are considered, during the integrate period determined by the timer, by the five detectors. At the end of the integrate period there appear pulses on the output leads of the two detectors which were considering desired responses while, ideally, no pulses appear on the other three leads. These pulses are fed to the register circuit.

Until the blank interval occurs in the train of received signals, the register circuit does not respond to the pulses from the detectors. When a blank interval does occur, the start circuit primes the steering circuit, which enables the first section of the register. The first digit of the transmitted number is then stored.

The outputs of the five detectors are also fed to a checking circuit, which determines whether a desired signal was recognized in exactly two of the five channels. If the digit is plausible, as indicated by two-out-of-five response, the checking circuit advances the steering circuit, so that the next digit can be stored in the next section of the register. If any digit is implausible, as indicated by response in more or less than two channels, the checking circuit does not advance the steering circuit but instead "recycles." That is, it erases all information stored in the register before the arrival of the questionable digit, and it returns the steering circuit to the state that existed before the recognition of the blank interval. The receiver then attempts on a succeeding repetition of the transmitted number to recognize a complete sequence of plausible digits.

For the purpose of simplifying the laboratory set-up, a skeletonized register capable of registering only the first three digits of the number was built. All eight digits, however, are checked. If all digits are plausible, the steering chain delivers a "number complete" signal, which releases the line relay, on the next appearance of the blank interval. The operation of the receiver is then complete.

The storage of the received number in the register represents the final output of the receiver. For laboratory purposes the information was used to operate various data-accumulating message registers as well as the comparator mentioned above. Upon completion of their operations, the register is cleared, and the receiver is then ready to handle another number.

6.0 TESTS

Conditions affecting the successful operation of a signaling system may arise in the transmitter, in the receiver, or in the transmission medium connecting the two. Laboratory tests, in which one condition at a time could be introduced and varied in a controlled manner, were first undertaken to evaluate the capabilities of the system. The results of these tests were so encouraging that arrangements were then made to try the system on a representative variety of toll circuits available on a loop basis out of New York. Good operation was obtained over two-wire voice-frequency circuits up to 175 miles in length and over four-wire voice-frequency circuits up to 350 miles in length. Seven hundred and fifty miles of voice-frequency four-wire circuit caused signaling failure owing to excessive delay-distortion. Two links of K carrier, aggregating 1900 miles, gave satisfactory performance with some restriction in the allowable range of amplitude of signals.

7.0 HIGHER SPEED SIGNALING — LOCAL LINES

The system just described was designed with the objective of tolerating large amounts of certain kinds of distortion. After the construction and testing of this system was completed, investigation was concentrated on the design of a system which would be capable of still higher signaling speed, even though it might be more vulnerable to distortion. Such a system would be useful in the local plant where the most severe types of distortion — carrier shift and delay distortion — do not occur.

In designing a system to signal at higher speed, it is obvious that the time interval T must be shortened, by spacing the signal frequencies further apart. In addition, a further increase of speed can be obtained by shortening the "handling" interval between digits. In the previously described system, it will be remembered, digits were transmitted at intervals of $2T$. This timing allowed an interval T for signal discrimination to appear in the network responses, and another interval T for detection and manipulation of the digit information. By spacing the signal frequencies at intervals of 400 cycles, which shortened T from 5 milliseconds to 2.5 milliseconds, and by reducing the handling time allowance from T to $\frac{1}{3}\,T$, it was possible to raise the signaling speed from 100 digits per second to 300 digits per second and to signal successfully at this speed over local loops. A new type of detector was required for this embodiment. The rest of the system was unchanged in principle from the previous design, so it will not be described.

7.1 Energy Detection

The previously described detector merely observed a voltage manifestation of the energy in the receiving network without making direct use of that energy. After detection of a digit signal, networks exhibiting desired responses still contained substantial amounts of energy, which decayed at a slow rate. With the altered signal parameters used in the higher-speed system, there was a likelihood that the decaying oscillations of a desired response might last into the next digit interval and obscure the discrimination for that digit. For this reason, and also because the available time for signal handling was much shorter, the new detector was devised with the dual objective of recognizing the presence of a desired response quickly and also of terminating as quickly as possible the slowly-decaying oscillations characteristic of a desired response. To accomplish this objective, the energy is used directly: it is made to advertise its own presence and simultaneously dissipates itself as quickly as the network permits. Because of the direct use of the energy, the detector is called an energy detector.

The scheme consists merely of inserting in series with the receiving resonant circuit, at time *T*, a resistance of such value as to make the circuit approximately critically damped. Then practically all the energy in the network is dissipated in the resistance within about one-half cycle of the resonant frequency. A simple circuit to perform this operation is shown in Fig. 7. The critical damping resistance is in the circuit at all times, shunted by a varistor through which steady current flows. As long as the steady current is larger than the peak amplitude of the oscillatory current in the tuned circuit, the varistor remains a low impedance and the circuit behaves as if the damping resistance were not there. When the time for detection arrives, the steady current is removed by driving to cut-off the vacuum tube through which it is supplied. If, at this time, the oscillatory current happens to be flowing in the direction for which the varistor is a low impedance (counter clockwise in the figure) nothing happens; but when the current reverses direction, the energy is dissipated in the damping resistance, producing a voltage pulse which can be used to actuate a storage register. The varistor serves two purposes: not only does it constitute a convenient, fast-acting switch for introducing the damping resistance, but also it ensures that the output pulse will always have the same polarity when it occurs. If the detection is initiated

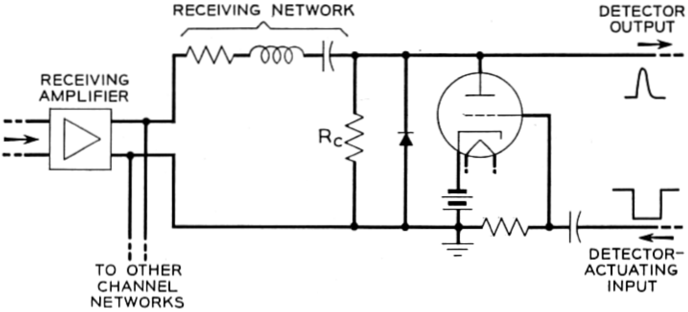

Fig. 7 — Circuit for polytonic signal energy detector.

at a time when the oscillating current is flowing in such a direction as to tend to give an output pulse of the wrong polarity, the output is simply delayed by an amount not greater than one-half period of the oscillation frequency.

8.0 CONCLUSION

A research program in high-speed signaling has resulted in laboratory models of several experimental systems based on different principles. This paper traces one of these principles from its origin in a mathematical concept to its application in two versions of workable physical circuitry. Both versions were found to give satisfactory results in laboratory tests and in limited trials with actual plant circuits.