

**Oct. 18, 1966**

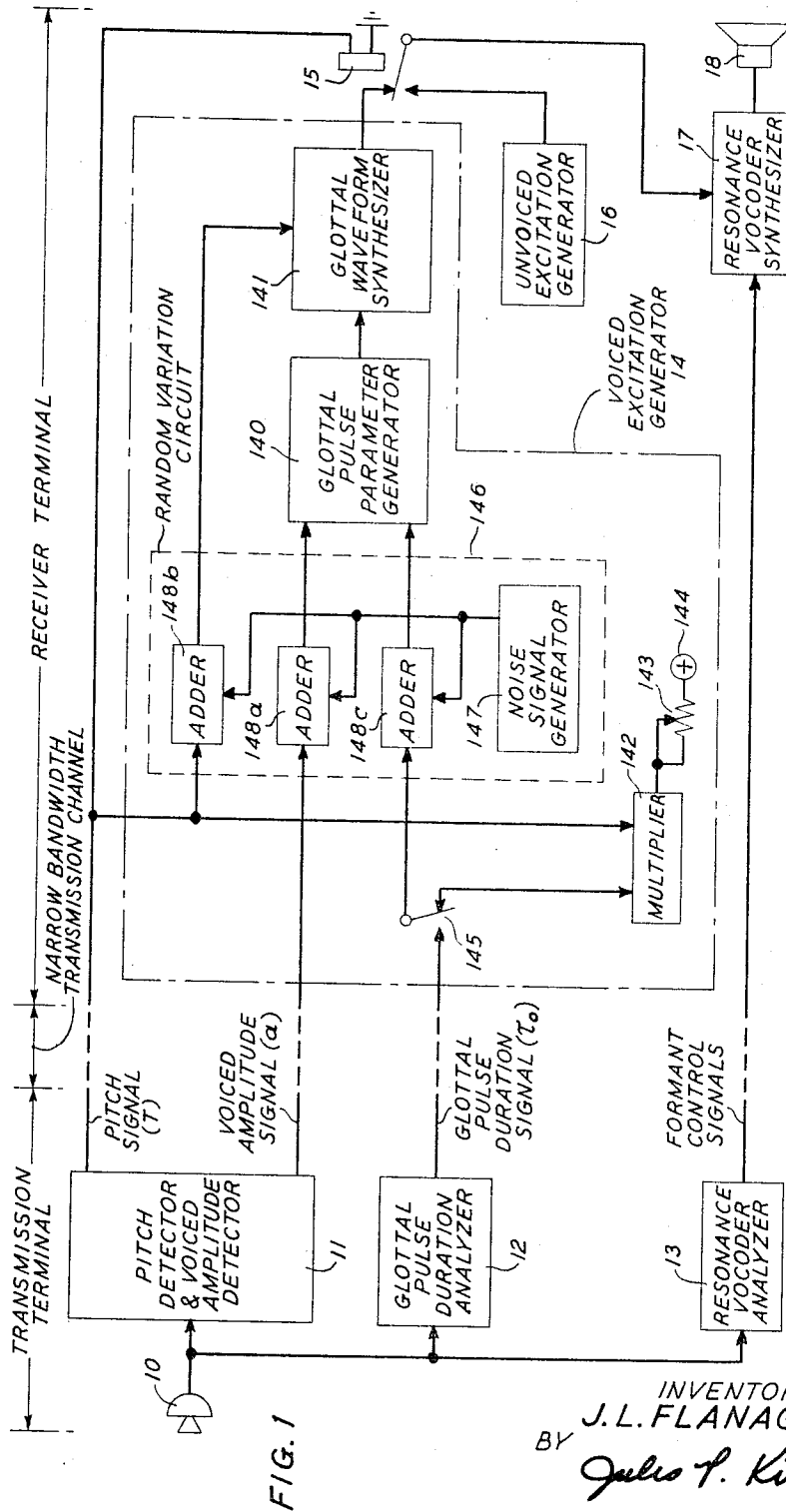
J. L. FLANAGAN

**3,280,266**

# SYNTHESIS OF ARTIFICIAL SPEECH

Filed May 15, 1963

5 Sheets-Sheet 1



Oct. 18, 1966

J. L. FLANAGAN

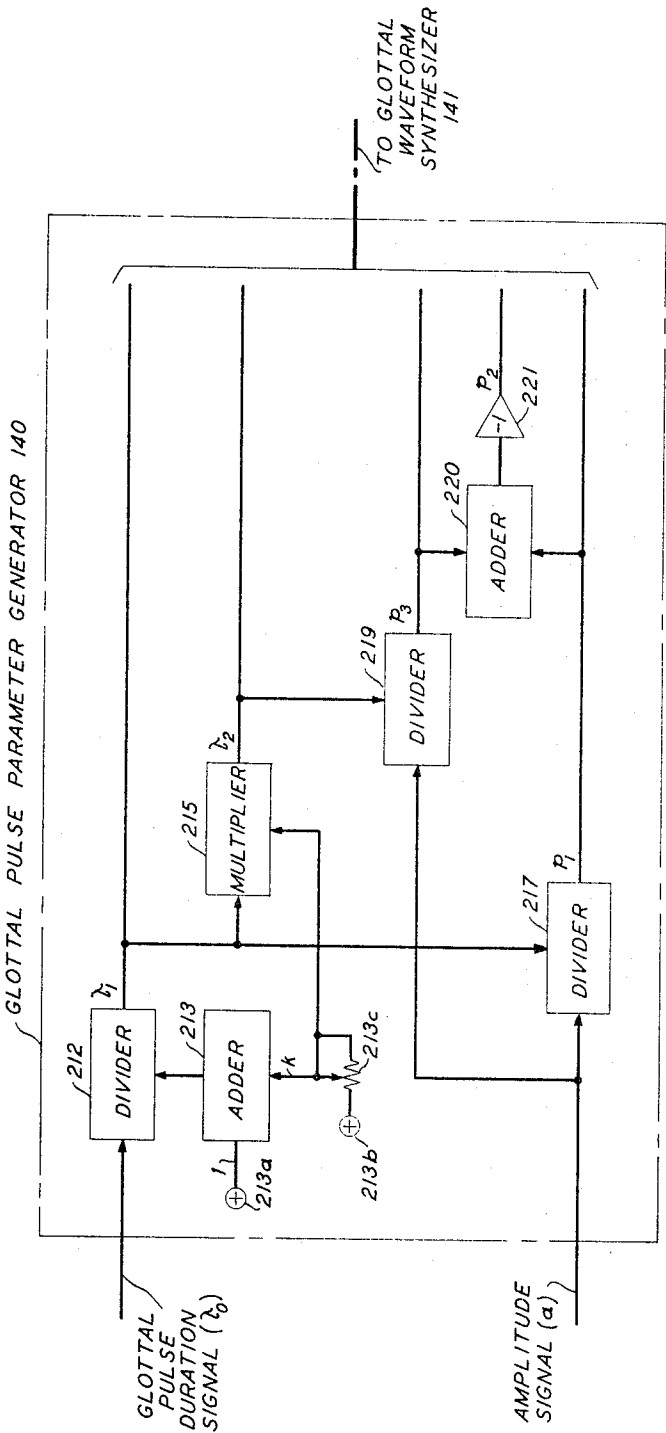
3,280,266

SYNTHESIS OF ARTIFICIAL SPEECH

Filed May 15, 1963

5 Sheets-Sheet 2

FIG. 2A



Oct. 18, 1966

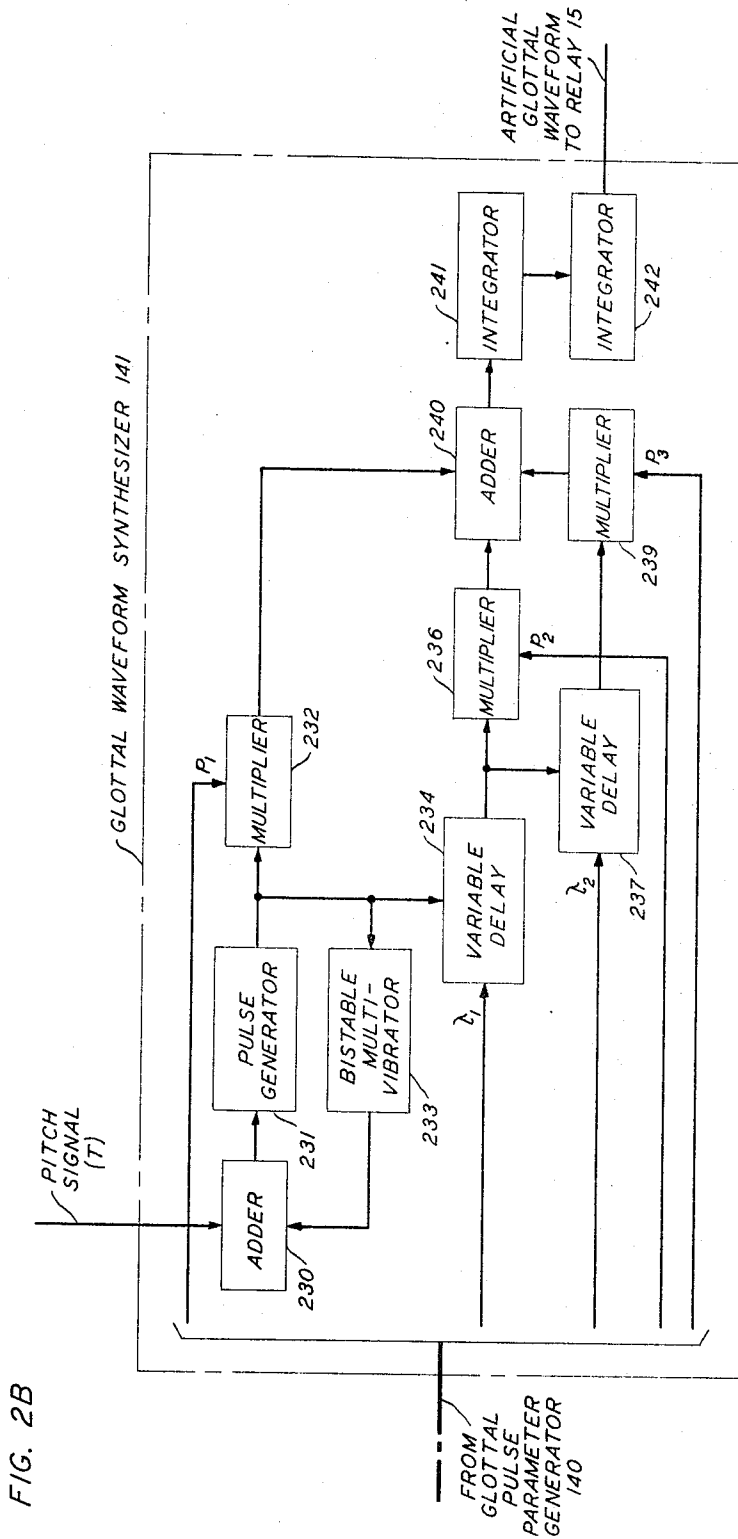
J. L. FLANAGAN

3,280,266

SYNTHESIS OF ARTIFICIAL SPEECH

Filed May 15, 1963

5 Sheets-Sheet 3



Oct. 18, 1966

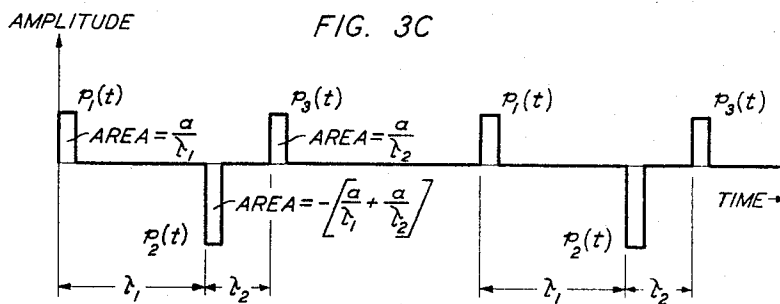
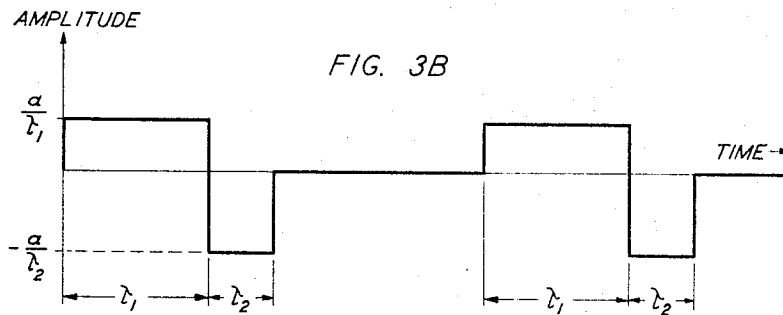
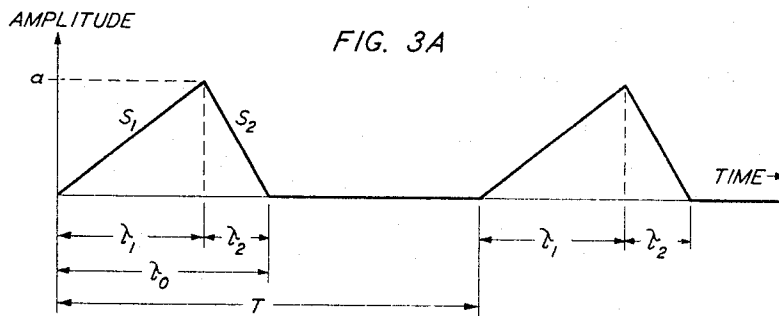
J. L. FLANAGAN

3,280,266

SYNTHESIS OF ARTIFICIAL SPEECH

Filed May 15, 1963

5 Sheets-Sheet 4



Oct. 18, 1966

J. L. FLANAGAN

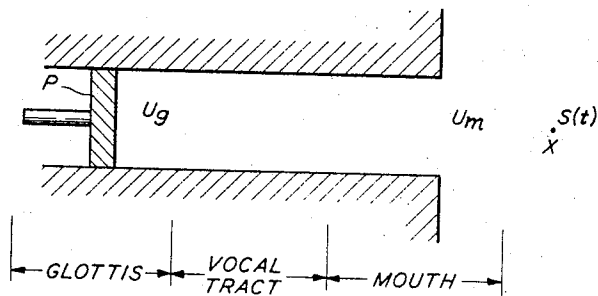
3,280,266

SYNTHESIS OF ARTIFICIAL SPEECH

Filed May 15, 1963

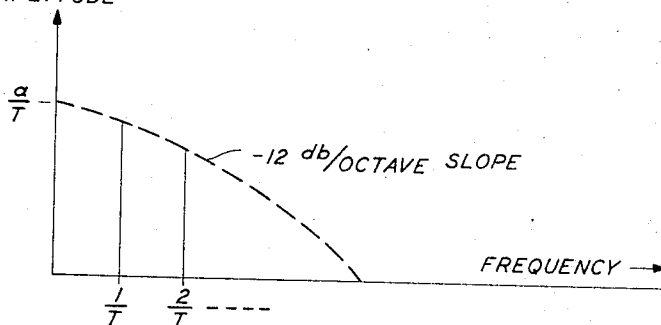
5 Sheets-Sheet 5

FIG. 4A



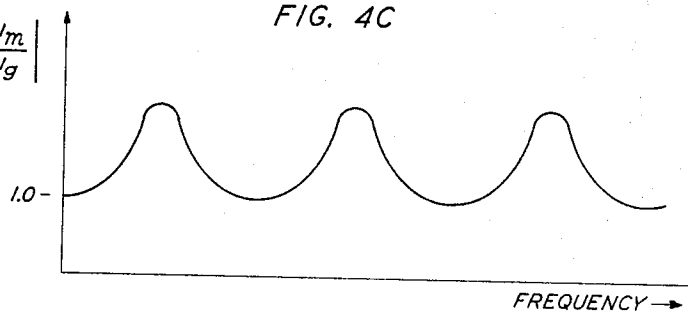
AMPLITUDE

FIG. 4B



$\left| \frac{u_m}{u_g} \right|$

FIG. 4C



1

3,280,266

## SYNTHESIS OF ARTIFICIAL SPEECH

James L. Flanagan, Warren Township, Somerset County, N.J., assignor to Bell Telephone Laboratories, Incorporated, New York, N.Y., a corporation of New York  
Filed May 15, 1963, Ser. No. 280,620  
5 Claims. (Cl. 179-15.55)

This invention relates to the synthesis of complex waves and, in particular, to the synthesis of natural sounding speech waves.

Conventional speech communications systems, for example, telephone systems, convey human speech by transmitting an electrical facsimile of the acoustic waveform produced by human talkers. It has been recognized, however, that facsimile transmission of the speech waveform is a relatively inefficient way to transmit speech information, because the amount of information contained in a typical speech wave may be transmitted over a communication channel of substantially narrower bandwidth than that required for facsimile transmission of the speech waveform. A number of arrangements for reducing or compressing the bandwidth necessary for transmission of speech information have been proposed, in which selected information bearing characteristics of speech are represented by narrow bandwidth control signals having a collective bandwidth that is substantially narrower than that of the speech waveform. In a typical bandwidth reducing arrangement, the narrow bandwidth control signals are obtained at a transmitter terminal by analyzing an incoming speech wave to determine the selected information bearing characteristics, and after transmitting the control signals to a distant receiver terminal over a relatively narrow bandwidth transmission channel, an artificial speech wave is synthesized from the control signals to provide a replica of the original speech wave.

These speech bandwidth reducing or compressing arrangements, which are often referred to as "vocoders" or "vocoder" systems, are relatively efficient from the standpoint of transmitting information, both in terms of the relatively small amount of transmission channel bandwidth that is required for the control signals and in terms of the relatively good intelligibility of the artificial speech wave that is reproduced from the control signals. From the standpoint of subjective speech quality, however, the artificial speech reproduced from the control signals does not sound as natural as speech that has been transmitted by a facsimile waveform transmission system.

It has been determined that one of the important factors in the perception of natural sounding speech is the presence of small irregularities in various speech parameters. These irregularities produce small fluctuations in the speech waveform and therefore they are preserved in facsimile transmission systems, but in vocoder systems these naturally occurring irregularities tend to become obscured during analysis and synthesis operations. Several prior vocoder systems have attempted to improve the quality of artificial vocoder speech either by preserving these irregularities during analysis and transmission, as shown in M. R. Schroeder, United States Patent No. 3,030,450, issued April 17, 1962, where a portion of the original speech wave is transmitted together with the narrow bandwidth control signals; or by introducing irregularities into the artificial speech wave during the synthesis operation, as shown in the copending application of J. L. Flanagan, Serial No. 257,947, filed February 12, 1963, where certain variations in selected speech parameters are derived from the transmitted control signals and reproduced in the artificial speech wave.

A general object of the present invention is the im-

2

provement of vocoder speech quality by introducing selected irregularities into the artificial speech wave during the synthesis operation.

It is well known that human speech sounds are produced by excitation of the human vocal tract, the type of sound produced depending upon the manner in which the vocal tract is excited. Thus the voiced portions of a speech wave are produced by exciting the vocal tract with quasi-periodic puffs of air which are released into the vocal tract from the lungs by the glottis or vocal folds, while the unvoiced portions of a speech wave are produced by the turbulent passage of air from the lungs through constrictions in the vocal tract. Analysis of the waveform of the glottal puffs by several investigators has revealed irregularities in the parameters of the glottal waveform; for example, see R. L. Miller, Nature of the Vocal Cord Wave, vol. 31, Journal of the Acoustical Society of America, page 667 (1959); J. L. Flanagan, Some Properties of the Glottal Sound Source, vol. 1, Journal of Speech and Hearing Research, page 99 (1958); and P. Lieberman, Perturbations in Vocal Pitch, vol. 33, Journal of the Acoustical Society of America, page 597 (1961).

As described in detail in the above mentioned articles, a single period of the glottal waveform typically comprises a nonzero portion and a zero portion respectively corresponding to open and closed positions of the glottis. The nonzero portion is frequently triangular in shape and exhibits apparent random variations in both duration and amplitude from period to period. Further, the length of each period appears to fluctuate in a random manner from period to period, and, as observed by Lieberman, for some speech sounds the length of each period tends to alternate between relatively long and relatively short durations.

The present invention provides for the introduction of small irregularities in an artificial vocoder speech wave by generating from certain of the transmitted narrow bandwidth control signals an artificial glottal waveform signal containing the above mentioned variations in duration, amplitude and period. This artificial glottal waveform signal is then employed with the remaining narrow bandwidth control signals to synthesize an artificial speech wave which is a natural sounding replica of the original speech wave.

An important feature of this invention is that the generation of the artificial glottal waveform does not require the derivation or transmission of additional control signals from the original speech wave, other than those ordinarily obtained, hence the improvement in vocoder speech quality achieved by this invention does not decrease vocoder bandwidth efficiency.

The invention will be more fully understood from the following detailed description of illustrative embodiments thereof, taken in connection with the appended drawings, in which:

FIG. 1 is a block diagram showing a complete bandwidth compression system embodying the principles of this invention;

FIGS. 2A and 2B are block diagrams showing in detail certain components of the system illustrated in FIG. 1;

FIGS. 3A, 3B, and 3C are graphs of assistance in explaining certain features of the present invention; and

FIGS. 4A, 4B, and 4C are drawings of assistance in explaining certain other features of the present invention.

### Theoretical considerations

The glottal waveform that excites the vocal tract to produce the voiced sounds of human speech is characterized by several distinctive features. As shown in FIG. 3A, the glottal waveform is quasi-periodic with a period T, and each period of the glottal waveform

3

typically comprises a nonzero portion having a triangular shape with a base or duration  $\tau_0$ , sides  $S_1$  and  $S_2$ , and an amplitude or altitude  $a$ . In general, the nonzero triangular portion is not symmetrical, that is, in general, the sides  $S_1$  and  $S_2$  are not equal, and the degree of asymmetry may be expressed in terms of a ratio  $k$  of either the sides  $S_1$  and  $S_2$  of the triangle, or the corresponding segments,  $\tau_1$  and  $\tau_2$ , of the base  $\tau_0$ , where

$$k = \frac{\tau_2}{\tau_1} \quad (1a)$$

and

$$\tau_0 = \tau_1 + \tau_2 \quad (1b)$$

Each of the parameters  $T$ ,  $\tau_0$ , and  $a$  may vary not only from sound to sound, but also from period to period. It is the latter variations which are random in nature, which constitute an important source of the irregularities in the speech waveform and therefore contribute to speech naturalness. However, in bandwidth compression systems in which one or more of the above parameters is transmitted in terms of one or more corresponding narrow bandwidth control signals, an individual control signal typically represents only variations from sound to sound, since in the course of analyzing a speech wave to measure a particular parameter the value of the parameter is ordinarily averaged over several periods, thereby obscuring the random, period to period fluctuations or irregularities. The present invention restores these irregularities during synthesis of an artificial speech wave by constructing from the transmitted control signals an artificial glottal waveform in which each of the above mentioned glottal waveform parameters is controlled and made to vary randomly from period to period in a manner approximating the random fluctuations of these parameters in the original glottal waveform. By utilizing the artificial glottal waveform generated by this invention to synthesize an artificial speech wave, the artificial speech wave is made to contain irregularities similar to those found in the original speech wave, thereby improving the naturalness of the artificial speech wave.

#### Apparatus

Turning now to FIG. 1, this drawing illustrates a complete vocoder or bandwidth compression system embodying the principles of this invention. In FIG. 1, as well as in FIGS. 2A and 2B, signal paths between various circuit elements are shown by single lines in order to avoid unnecessary complexity. It will be obvious to those skilled in the art at what points one or more wire pairs or other complete circuits may be required to practice this invention.

In FIG. 1, a speech wave from source 10 at the transmitter terminal is applied in parallel to elements 11, 12, and 13, where source 10 may be a conventional microphone of any desired variety. Element 11 analyzes the incoming speech wave from source 10 to derive a pitch signal whose magnitude is representative of both the period  $T$ , of voiced portions of the speech wave and the nonperiodically of unvoiced portions of the speech wave, and a voiced amplitude signal which is representative of the amplitude of the fundamental frequency component of the speech wave. As pointed out in the appendix below, the amplitude of the fundamental component of the speech waveform is a reasonable measure of the amplitude  $a$  of the nonzero portion of the glottal waveform that produced the speech wave, hence the voiced amplitude signal obtained by element 11 is also indicative of the amplitude of the nonzero portion of the glottal waveform. A suitable pitch detector and voiced amplitude detector is illustrated and described in the copending application of E. E. David, Jr. et al., Serial No. 235,703, filed November 6, 1962, now United States Patent No. 3,190,963, issued June 22, 1965.

Element 12, a so-called glottal pulse duration analyzer, derives from the speech wave a narrow bandwidth control signal indicative of the duration  $\tau_0$  of the nonzero

4

portion of each period of the glottal waveform. If desired, this element may be identical in structure to the glottal duty cycle detector described in M. V. Mathews et al., United States Patent No. 3,083,266, issued March 26, 1963.

Element 13 may be any one of a variety of vocoder analyzers for deriving from the speech wave a group of control signals representative of additional speech parameters. For example, analyzer 13 may be a resonance vocoder analyzer of the type described by E. E. David, Jr. in "Signal Theory in Speech Transmission," vol. CT-3 I.R.E. Transactions on Circuit Theory, page 232 (1956), in which case the group of control signals is representative of the frequencies of selected formants or peaks in the speech amplitude spectrum.

The bandwidths of the control signals obtained by elements 11, 12, and 13 are relatively narrow, so that these control signals may be transmitted to a receiver terminal over a transmission channel of substantially narrower bandwidth than that required for facsimile transmission of the incoming speech wave from source 10. In FIG. 1, a suitable narrow bandwidth transmission channel is indicated by broken lines connecting the transmitter terminal to the receiver terminal.

At the receiver terminal, the pitch, voiced amplitude, and glottal pulse duration signals are applied to voiced excitation generator 14 of this invention, while the control signals from vocoder analyzer 13 are applied to resonance vocoder synthesizer 17, which may be of a construction complementary to that of analyzer 13. In addition, the pitch signal is applied to the control terminal of a switching element 15, which may be any one of a number of well-known devices such as a relay for directing a signal from either voiced excitation generator 14 or unvoiced excitation generator 16 to synthesizer 17. Excitation generator 14, which is described in detail below, provides an artificial glottal waveform for synthesizing natural sounding voiced portions of the artificial speech wave produced by synthesizer 17, while generator 16, which may be a conventional hiss or noise source, provides a random waveform for synthesizing unvoiced portions of the artificial speech wave produced by synthesizer 17. When a voiced sound is present in the original speech wave, the magnitude of the pitch signal is greater than a predetermined level necessary to energize relay 15, and relay 15 passes the artificial glottal waveform from generator 14 to synthesizer 17. However, when an unvoiced sound is present in the original speech wave, the magnitude of the pitch signal falls below the predetermined level, and relay 15 is thereby de-energized so that its back contact is closed to connect generator 16 to synthesizer 17.

Before proceeding to a detailed description of voiced excitation generator 14, it is convenient at this point to distinguish between three possible situations in the construction of an artificial glottal waveform. In the first situation, the duration of the nonzero portion,  $\tau_0$ , of each period of the glottal waveform is measured at the transmitter terminal, as shown by element 12 in FIG. 1, and transmitted via a control signal to the receiver terminal. In the second situation,  $\tau_0$  is made a function of  $T$ , for example, the ratio of  $\tau_0$  to  $T$  is a constant, denoted  $d$ ,

$$d = \frac{\tau_0}{T} \quad (2)$$

In the second situation, it is not necessary to measure  $\tau_0$  provided that  $T$  is measured, hence element 12 in FIG. 1 may be eliminated in the event that the artificial waveform is to be constructed so that  $\tau_0$  is a function of  $T$ . The third situation, a simplification of the first, also eliminates the transmission of a  $\tau_0$  signal by simply choosing an appropriate constant value for  $\tau_0$  and generating a corresponding constant value signal at the receiver terminal. Implementation of this alternative is not shown in the drawing, but is believed to be obvious to those skilled in the art.

Turning now to generator 14 of FIG. 1, the voiced amplitude signal is applied to random variation circuit 146; the pitch signal is applied simultaneously to one of the input terminals of multiplier 142 and to random variation circuit 146; and the glottal pulse duration signal is applied to one of the two contacts of switch 145. Within voiced excitation generator 14, the other input terminal of multiplier 142 is supplied through potentiometer 143 with a signal from constant energy source 144, for example, a battery, representative of a selected constant ratio  $d$ , as specified in Equation 2, so that the product signal developed at the output terminal of multiplier 142 is proportional to a value of  $\tau_0$ , which is dependent upon the instantaneous magnitude of  $T$ , since

$$T \times d = T \times \frac{\tau_0}{T} = \tau_0 \quad (3)$$

The output terminal of multiplier 142 is connected to the other contact of switch 145, hence the position in which switch 145 is placed depends upon which one of the two above mentioned variations of the artificial glottal waveform it is desired to construct.

To each of the applied pitch, glottal amplitude and glottal pulse duration signals, random variation circuit 146 adds a noise signal in order to reproduce the kind of random variation observed in these parameters of the original glottal waveform. A suitable noise signal is produced by noise signal generator 147, which may be of any well-known variety, and the noise signal is added to each of the pitch, amplitude and duration signals in adders 148a, 148b, and 148c, respectively. From adders 148a and 148c the amplitude and duration signals are passed to parameter generator 140, while from adder 148b the pitch signal is passed to waveform synthesizer 141.

As described in detail below, parameter generator 140 derives from the glottal amplitude,  $a$ , and glottal duration,  $\tau_0$ , a group of so-called parameter signals representing the desired characteristics of an artificial glottal waveform, and from these parameter signals synthesizer 141 constructs an artificial glottal waveform. The derivation of the parameter signals specifying the desired characteristics of the artificial glottal waveform is based upon the following considerations.

Referring back to FIG. 3A, it is observed that one of the characteristics of the nonzero portion of each period is the length of the segments  $\tau_1$  and  $\tau_2$  of the duration  $\tau_0$ . However, only the total duration,  $\tau_0$ , of the nonzero portion of the glottal waveform is specified by the glottal duration signal derived by element 12 at the transmitter terminal. Hence, in order to obtain values for  $\tau_1$  and  $\tau_2$ , it is first necessary to select a desired degree of asymmetry, as expressed by the ratio in Equation 1a, and from Equation 1b it is then possible to obtain values for  $\tau_1$  and  $\tau_2$ . Thus

$$k = \frac{\tau_2}{\tau_1} \quad (4)$$

$$\tau_2 = k \cdot \tau_1$$

and

$$\begin{aligned} \tau_0 &= \tau_1 + \tau_2 \\ &= \tau_1 + (k \cdot \tau_1) \\ &= (1+k) \tau_1 \end{aligned}$$

hence

$$\tau_1 = \frac{\tau_0}{(1+k)} \quad (5)$$

Having obtained values for  $\tau_1$  and  $\tau_2$  from  $\tau_0$  and a predetermined  $k$ , as given by Equations 4 and 5, it is taken necessary to form a group of signals from which a triangular waveform may be constructed. A suitable group of signals is shown in FIG. 3C, where each of the pulses  $p_1(\tau)$ ,  $p_2(\tau)$ ,  $p_3(\tau)$  in each period represents an impulse function whose leading edge respectively occurs at the beginning of  $\tau_1$ , the end of  $\tau_1$  and beginning of  $\tau_2$ , and

the end of  $\tau_2$ . In addition, each of these impulse functions is made to have a selected area so that double integration of each impulse function will produce the triangle waveform shown in FIG. 3A.

As explained in E. A. Guillemin, "Introductory Circuit Theory," p. 198 (1953), the integral of an impulse function is a step function, and the integral of a step function is a ramp function. Since a triangular function may be formed from three adjacent ramp functions, the first and third of which have positive slopes and the second of which has a negative slope, the triangular glottal waveform in FIG. 3A may be constructed from double integration of each of the impulse functions in FIG. 3C, with the asymmetry of the triangular glottal waveform being determined by the areas of the various impulse functions. Thus, by making the area of the impulse functions, denoted  $p_1(t)$ ,  $p_2(t)$ , and  $p_3(t)$  in each period in FIG. 3C, respectively, equal to

$$p_1(t) = \frac{a}{\tau_1} \quad (6a)$$

$$p_2(t) = -\left[\frac{a}{\tau_1} + \frac{a}{\tau_2}\right] \quad (6b)$$

and

$$p_3(t) = \frac{a}{\tau_2} \quad (6c)$$

a single integration of the impulse functions produces the step function waveform shown in FIG. 3B, in which the amplitude and duration of the first step waveform are  $a/\tau_1$  and  $\tau_1$ , respectively, and the amplitude and duration of the second waveform are  $-a/\tau_2$  and  $\tau_2$ , respectively. A second integration, this time performed upon the step waveform shown in FIG. 3B, produces the triangular glottal waveform shown in FIG. 3A.

Summarizing the above discussion, the pulse parameters that must be obtained by parameter generator 140 are

$$\tau_1, \tau_2, \frac{a}{\tau_1}, \frac{a}{\tau_2}, \text{ and } -\left[\frac{a}{\tau_1} + \frac{a}{\tau_2}\right]$$

as defined in Equations 5, 4, 6a, 6c, and 6b. Signals representative of these parameters are derived from the incoming amplitude and duration signals,  $a$  and  $\tau_0$ , in the following manner.

Turning now to FIG. 2A, the glottal pulse duration signals denoted  $\tau_0$ , which may be supplied from either glottal pulse duration analyzer 12 or multiplier 142, are applied to the dividend terminal of a conventional divider 212 that develops at its quotient terminal a signal proportional to the pulse segment parameter  $\tau_1$ . This is accomplished by applying to the divisor terminal of divider 212 a signal from adder 213 which is representative of  $(1+k)$ , where  $k$  is a desired degree of asymmetry. Adder 213 develops a signal proportional to  $(1+k)$  from energy source 213a, which supplies a signal representative of unity, and from energy source 213b, which in conjunction with potentiometer 213c, supplies a signal representative of a desired value of  $k$ . The  $\tau_1$  signal appearing at the output terminal of divider 212 is delivered to glottal waveform synthesizer 141, and also to conventional multiplier 215 and conventional divider 217 for use in developing other glottal pulse parameter signals.

In multiplier 215, the  $\tau_1$  signal is multiplied by the  $k$  signal from potentiometer 213c to obtain a product signal indicative of the pulse parameter  $\tau_2$ , since from Equation 4

$$\tau_2 = k \cdot \tau_1 = \frac{\tau_2}{\tau_1} \cdot \tau_1 = \tau_2$$

The  $\tau_2$  signal developed by multiplier 215 is also sent to glottal waveform synthesizer 141, and in addition the  $\tau_2$  signal is applied to the divisor terminal of conventional divider 219.

For the derivation of the three signals  $p_1$ ,  $p_2$ , and  $p_3$

specified by Equations 6a, 6b, and 6c, respectively, the incoming voiced amplitude signal, denoted  $a$ , from detector 11 is applied in parallel to the dividend terminal of divider 217 and to the dividend terminal of divider 219.

Divider 217 develops from the  $\tau_1$  signal from divider 212 and the incoming amplitude signal,  $a$ , a quotient signal proportional to  $p_1$ , as defined in Equation 6a. The  $p_1$  signal is sent to glottal waveform synthesizer 141 as a glottal pulse parameter signal and to adder 220 for the development of a  $p_2$  parameter signal. Divider 219 develops from the applied  $\tau_2$  and  $a$  signals a quotient signal indicative of  $p^3$  in accordance with Equation 6c, and the  $p_3$  signal is sent to glottal waveform synthesizer 141 and applied to adder 220. Adder 220 combines the  $p_1$  and  $p_3$  signals to form a sum signal proportional to

$$\left[ \frac{a}{\tau_1} + \frac{a}{\tau_2} \right]$$

and by passing this sum signal through a polarity reversing device 221, for example, a well-known minus one amplifier, there is produced at the output terminal of device 221 a signal representing  $p_2$  as specified by Equation 6b.

Referring now to FIG. 2B, this drawing shows apparatus for constructing an artificial glottal waveform from the glottal pulse parameter signals supplied by generator 140. Within glottal waveform synthesizer 141 the pitch signal transmitted from detector 11 is caused to represent a glottal period that alternates between relatively short and relatively long durations according to a preselected pattern. This alternation in period duration is provided by alternately increasing and decreasing the magnitude of the pitch signal through the application of alternate positive and negative signals to adder 230 during each glottal period. This is accomplished by passing the pitch signal from adder 230 to a pulse generator 231, for example, a relaxation oscillator of any well-known variety, which produces a sequence or train of uniform pulses, for example, unit area pulses, with a fundamental frequency or period determined by the magnitude of the pitch signal from adder 230. The pulses from generator 231 are employed to cause conventional bistable multivibrator 233 to alternate between its two stable states, one state producing a pulse with a first predetermined amplitude and the other state producing a pulse with a second predetermined amplitude. If desired, the predetermined amplitudes of the two pulses produced by multivibrator 233 may be chosen to cause the pitch signal developed at the output terminal of adder 230 to alternate between representing relatively long and relatively short periods according to the statistics reported in the previously mentioned Lieberman article.

From pulse generator 231 the train of periodic pulses is delivered to multiplier 232 and variable delay element 234. Multiplier 232, which may be of any well-known variety, is also supplied with the  $p_1$  parameter signal from generator 140, so that the product signal developed at the output terminal of multiplier 232 is a train of marker pulses with areas proportional to the instantaneous value of  $p_1$  and with leading edges coinciding with the beginning of each period of the gottal waveform, similar to the impulse function denoted  $p_1(t)$  in FIG. 3C.

The other two trains of marker pulses, corresponding to the impulse functions denoted  $p_2(t)$  and  $p_3(t)$  in FIG. 3C, are generated by the respective pairs of series-connected elements 234, 236, and 237, 239. Elements 234 and 237 are conventional variable delay elements, for example, phantastron delay devices, each of which in response to an incoming pulse generates a delayed pulse at a variable instant of time following the application of the incoming pulse. In the apparatus of FIG. 2B, the amount of time which elapses between application of an incoming pulse from generator 231 to multivibrator 234 and generation of a delayed pulse by multivibrator 234 is controlled by the  $\tau_1$  parameter signal from parameter

generator 140. Thus, the leading edge of each of the pulses produced by multivibrator 234 follows the leading edge of a corresponding earlier pulse from generator 231 at an interval determined by the instantaneous value of the  $\tau_1$  signal. Similarly, by applying the delayed pulses from multivibrator 234 to multivibrator 236, and by controlling the delay time of multivibrator 237 with  $\tau_2$  parameter signal from parameter signal generator 140, the leading edge of each of the pulses produced by multivibrator 237 follows the leading edge of a corresponding earlier pulse from multivibrator 234 at an interval determined by the instantaneous value of the  $\tau_2$  signal.

The areas of the pulses in the respective pulse trains from multivibrators 234 and 237 are made proportional to the parameters  $p_2$  and  $p_3$  by multiplying the respective pulse trains by the  $p_2$  and  $p_3$  parameter signals from generator 140 in multipliers 236 and 239, respectively, thereby developing marker pulses similar to the corresponding impulse functions denoted  $p_2(t)$  and  $p_3(t)$  in FIG. 3C.

The marker pulse trains developed at the output terminals of multipliers 232, 236, and 239 are combined by adder 240 to form a single time sequence of pulses in which the three marker pulses  $p_1$ ,  $p_2$ , and  $p_3$ , occur at successive instants of time in each period,  $T$ . The single sequence of marker pulses formed by adder 240 is passed to series-connected integrators 241 and 242 which performs the double integration required to construct an artificial glottal wave with the shape illustrated in FIG. 3A. Integrators 241 and 242 may be of any suitable construction; for example, they may be either Miller integrators or resistance-capacitance circuits of appropriate design. From the output terminal of integrator 242 the artificial glottal waveform is sent to relay 15, as shown in FIG. 1, so that during voiced portions of the original speech wave the artificial glottal waveform is passed to vocoder synthesizer 17 for reconstruction of a natural sounding artificial speech wave.

#### Appendix

It was previously stated that the voiced amplitude signal derived by detector 11 at the transmitter terminal of the apparatus of FIG. 1, in representing the amplitude of the fundamental frequency component of a speech wave, is also a reasonable measure of the amplitude,  $a$ , of the nonzero portion of the glottal waveform. The relation between the amplitude of the fundamental frequency component of a speech wave and the amplitude of the gottal pulse is demonstrated by the following properties of the mechanism of speech production.

Referring to FIG. 4A, this drawing shows an approximation of the human vocal mechanism in which the vocal tract is represented by a hollow cylinder or tube having at one end a vibrating piston,  $P$ , representing the glottis, and an opening at the other end representing the mouth. The piston  $P$  acts as a source of acoustic volume velocity,  $U_g$ , where  $U_g$  is also the glottal waveform illustrated in FIG. 3A. The amplitude spectrum of the glottal waveform, as shown in FIG. 4B, has an envelope that decreases at twelve decibels per octave in the region of the fundamental glottal frequency,  $1/T$ . In addition, the amplitude of the fundamental glottal frequency component is proportional to  $a/T$ ,  $a$  being the amplitude of the nonzero portion of the glottal waveform.

The glottal waveform passes through the vocal tract and is radiated by the mouth to set up a pressure or speech wave,  $S(t)$  at a point  $x$  in front of the mouth, as shown in FIG. 4A, where the frequency components of  $S(t)$  are also harmonics of the fundamental glottal frequency,  $1/T$ . It is the amplitude of the fundamental frequency component, with frequency  $1/T$ , of the wave  $S(t)$  which is measured by detector 11, hence it is necessary to show that whereas the amplitude of the fundamental gottal component is proportional to  $a/T$ , the amplitude of the fundamental component of  $S(t)$  varies only with  $a$ .

9

It is first observed that because the amplitude of the fundamental component of the glottal wave is proportional to  $d/T$ , an increase in glottal fundamental frequency results in an increase of the fundamental gottal component amplitude at the rate of six decibels per octave. However, as shown in FIG. 4B, the envelope of the glottal spectrum decreases at the rate of twelve decibels per octave, hence the net effect of increasing the fundamental frequency of the glottal wave is to decrease the amplitude of the fundamental component of the glottal wave at the rate of six decibels per octave.

Next, on passing through the vocal tract, it has been observed that low frequencies of the glottal waveform, including the usual values for the fundamental component, are transmitted without change in amplitude. This is illustrated graphically in FIG. 4C by the transmission characteristic of a typical human vocal tract, in which it is noted that the transmission characteristic,

$$|U_m/U_g|$$

is approximately unity for low frequencies.

Finally, on radiating a sound pressure wave from the mouth, the glottal waveform components are increased in amplitude by the frequency characteristic of the mouth radiation, which is proportional to the first power of frequency and has a positive six decibels per octave slope (not shown), thereby offsetting the previously mentioned six decibels per octave decrease in amplitude of the glottal fundamental component with increasing fundamental frequency. Thus the amplitude of the fundamental component of the sound pressure wave radiated from the mouth by the glottal waveform is independent of the fundamental frequency,  $1/T$ , and varies only with changes in the amplitude  $a$  of the nonzero portion of the glottal wave. Measurement of the amplitude of the fundamental component of the speech wave  $S(t)$  therefore provides a reasonable measure of the glottal amplitude as well.

Although this invention has been described in terms of speech communications systems of the type shown in FIG. 1, it is to be understood that applications of the principles of this invention are not limited to this field but include such related fields as automatic speech recognition, speech processing, secrecy systems, and automatic message recording and reproduction. In addition, it is to be understood that the above-described embodiments are merely illustrative of the numerous arrangements which may be devised for the principles of this invention by those skilled in the art without departing from the spirit and scope of the invention.

What is claimed is:

1. A speech communication system that comprises a transmitter terminal including
  - a source of an incoming speech wave,
  - means for deriving from said speech wave first and second narrow bandwidth control signals respectively representative of the fundamental period and the amplitude of the fundamental component of said speech wave,
  - means for obtaining from said speech wave a third narrow bandwidth control signal representative of the duration of the nonzero portion of the glottal waveform characteristic of said speech wave, and
  - means for analyzing said speech wave to obtain a group of narrow bandwidth formant control signals indicative of the locations of selected formants of said speech wave,
  - a narrow bandwidth transmission channel for transmitting said first, second and third control signals and said group of formant control signals from said transmitter terminal to a receiver terminal, and
  - at said receiver terminal,
  - a voiced excitation generator supplied with said first, second, and third control signals for con-

10

structing an artificial periodic glottal waveform in which

each period has a nonzero portion that fluctuates randomly in amplitude and duration from period to period, and

the length of each period alternates between relatively long and relatively short durations according to a predetermined pattern,

an unvoiced excitation generator for generating a random waveform for the synthesis of unvoiced portions of an artificial speech wave,

switching means supplied with said artificial glottal waveform and said random waveform and responsive to said first control signal for transmitting either said glottal waveform or said random waveform according to a predetermined magnitude level of said first control signal, and synthesizer means connected to said switching means and supplied with said group of formant control signals for constructing an artificial speech wave that is a natural sounding replica of said incoming speech wave at said transmitter terminal.

2. Apparatus for reconstructing an artificial speech wave that closely resembles an original speech wave which comprises

a source of a first control signal representative of the fundamental period of voiced portions of said original speech wave,

a source of a second control signal representative of the amplitude of the nonzero portion of each period of the glottal waveform characteristic of said original speech wave,

a source of a third control signal representative of the duration of the nonzero portion of each period of said original speech wave,

a source of a group of formant signals representative of the locations of selected peaks in the amplitude spectrum of said speech wave,

a voiced excitation generator supplied with said first, second and third control signals for generating an artificial glottal waveform that is a replica of the glottal waveform characteristic of said original speech wave, said voiced excitation generator including

means supplied with said first, second and third control signals for introducing random fluctuations into the values represented by each of said control signals, means supplied with said randomly fluctuating second and third control signals for deriving a group of parameter signals specifying selected parameters of the nonzero portion of each period of said glottal waveform, and

means supplied with said randomly fluctuating first control signal and said group of parameter signals for synthesizing an artificial glottal waveform with periods that alternate between relatively long durations and relatively short durations according to a predetermined pattern,

an unvoiced excitation generator for generating a random waveform for the synthesis of unvoiced portions of an artificial speech wave,

switching means supplied with said artificial glottal waveform and said random waveform and responsive to said first control signal for transmitting either said glottal waveform or said random waveform according to a predetermined magnitude level of said first control signal, and

synthesizer means connected to said switching means and supplied with said group of formant signals for constructing an artificial speech wave that is a natural sounding replica of said original speech wave.

3. Apparatus for constructing an artificial glottal waveform that closely approximates the original glottal waveform characteristic of voiced portions of a human speech wave which comprises

11

a source of a first control signal representative of the fundamental period of voiced portions of said speech wave,

a source of a second control signal representative of the amplitude of the nonzero portion of each period of the glottal waveform characteristic of said speech wave,

a source of a third control signal representative of the duration of the nonzero portion of each period of said speech wave,

means supplied with said first, second and third control signals for introducing random fluctuations into the values represented by each of said control signals,

means supplied with said randomly fluctuating second and third control signals for deriving a group of parameter signals specifying selected parameters of the nonzero portion of each period of said glottal waveform, and

means supplied with said randomly fluctuating first control signal and said group of parameter signals for synthesizing an artificial glottal waveform with periods that alternate between relatively long durations and relatively short durations according to a predetermined pattern.

4. Apparatus for generating a periodic waveform with period  $T$ , in which each period contains a triangular, nonzero portion of duration  $\tau_0$  and a zero portion of duration  $T - \tau_0$ , where the duration of the nonzero portion includes two segments  $\tau_1$  and  $\tau_2$  which respectively correspond to the sides  $S_1$  and  $S_2$  of said triangular portion, and the degree of asymmetry of the triangular portion is denoted

$$k = \frac{\tau_2}{\tau_1}$$

which comprises

a source of a first control signal representative of the duration,  $\tau_0$ , of the nonzero portion of each period of said waveform,

a source of a second control signal representative of the degree of asymmetry,  $k$ , of the nonzero portion of each period of said waveform,

a source of a third control signal representative of the amplitude,  $a$ , of the nonzero portion of each period of said waveform,

a source of a fourth control signal having a magnitude substantially equal to unity,

a source of a fifth control signal representative of said period  $T$ ,

means for deriving from said first, second, and fourth control signals a first parameter signal indicative of said segment  $\tau_1$ ,

means for deriving from said second control signal and said first parameter signal a second parameter signal indicative of said segment  $\tau_2$ ,

means supplied with said third control signal and said second parameter signal for deriving a third parameter signal representative of the ratio  $a/\tau_2$ ,

12

means supplied with said third control signal and said first parameter signal for deriving a fourth parameter signal representative of the ratio  $a/\tau_1$ ,

means supplied with said third and fourth parameter signals for obtaining a fifth parameter signal indicative of the sum

$$-\left[\frac{a}{\tau_1} + \frac{a}{\tau_2}\right]$$

and

means supplied with said first, second, third, fourth, and fifth parameter signals and said fifth control signal for constructing said periodic waveform.

5. Apparatus for synthesizing a quasi-periodic waveform in which each period contains both a zero portion and a triangular, nonzero portion, where said nonzero portion has an amplitude  $a$ , a duration  $\tau_0$ , a positive slope  $a/\tau_1$ , a negative slope  $-a/\tau_2$ , and  $\tau_0 = (\tau_1 + \tau_2)$ , which comprises

means for developing a sequence of uniform pulses having a period that alternates between a predetermined relatively short duration and a predetermined relatively long duration,

means supplied with said sequence of uniform pulses for generating a corresponding first train of marker pulses in which the area of each pulse in said first train of marker pulses is proportional to  $a/\tau_1$ ,

means supplied with said sequence of uniform pulses for generating a corresponding second and a corresponding third train of marker pulses, in which the area of each pulse in said second train of marker pulses is proportional to the sum

$$-\left[\frac{a}{\tau_1} + \frac{a}{\tau_2}\right]$$

and each pulse in said second train of marker pulses occurs at a time  $\tau_1$  following a corresponding pulse in said sequence of uniform pulses, and in which the area of each pulse in said third train of marker pulses is proportional to  $a/\tau_2$  and each pulse in said third train of marker pulses occurs at a time  $\tau_0 = (\tau_1 + \tau_2)$  following a corresponding pulse in said sequence of uniform pulses,

means for combining said first, second, and third trains of marker pulses to form a single train of marker pulses, and

means for double integrating said single train of marker pulses.

#### References Cited by the Examiner

##### UNITED STATES PATENTS

|           |        |           |           |
|-----------|--------|-----------|-----------|
| 3,040,450 | 4/1962 | Schroeder | 179—15.55 |
| 3,083,266 | 3/1963 | Mathews   | 179—1     |
| 3,190,963 | 6/1965 | David     | 179—15.55 |

DAVID G. REDINBAUGH, *Primary Examiner*.

S. J. GLASSMAN, *Assistant Examiner*.