

Oct. 13, 1959

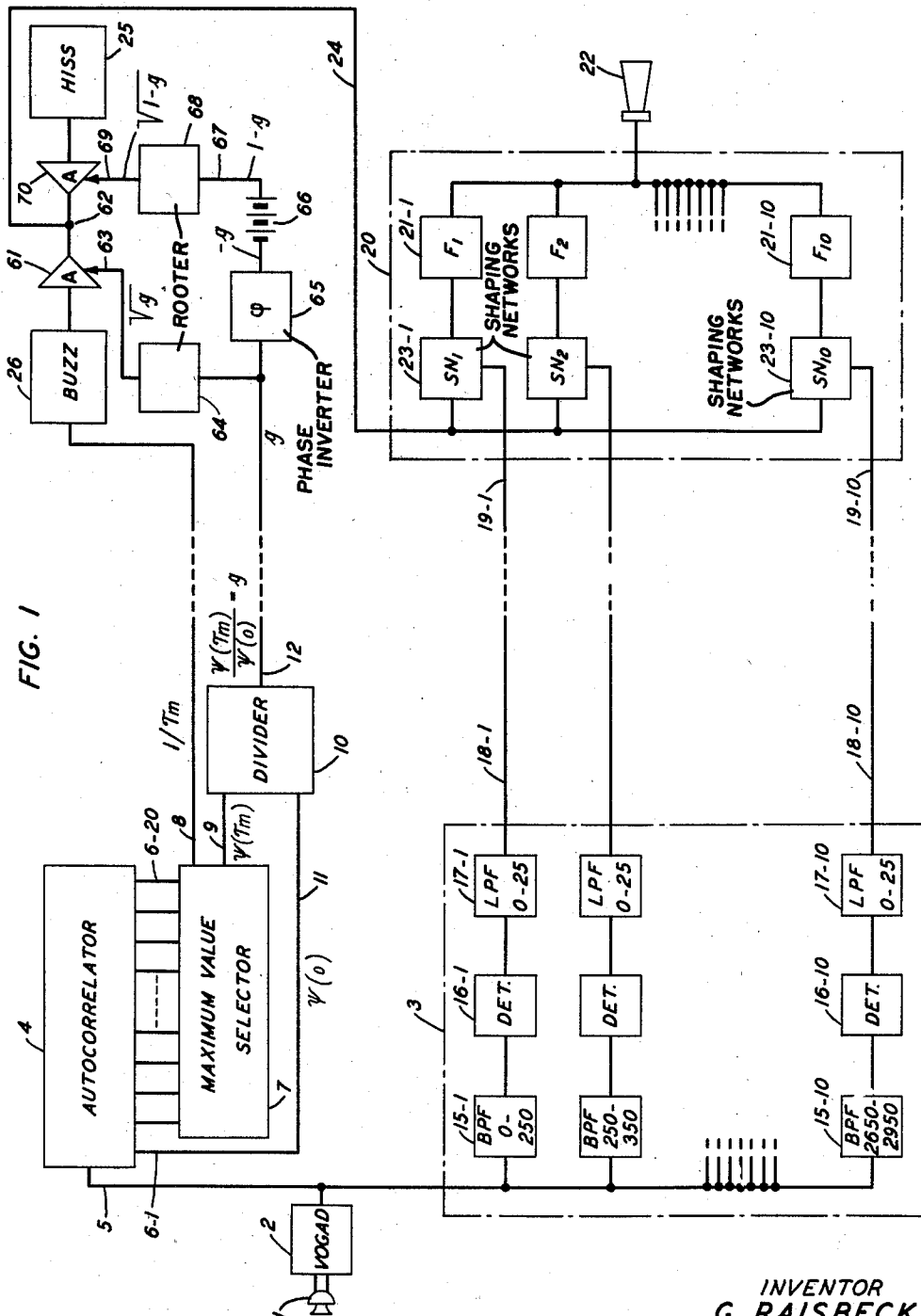
G. RAISBECK

2,908,761

VOICE PITCH DETERMINATION

Filed Oct. 20, 1954

3 Sheets-Sheet 1



INVENTOR
G. RAISBECK
BY *Harry C. Hart*
ATTORNEY

Oct. 13, 1959

G. RAISBECK

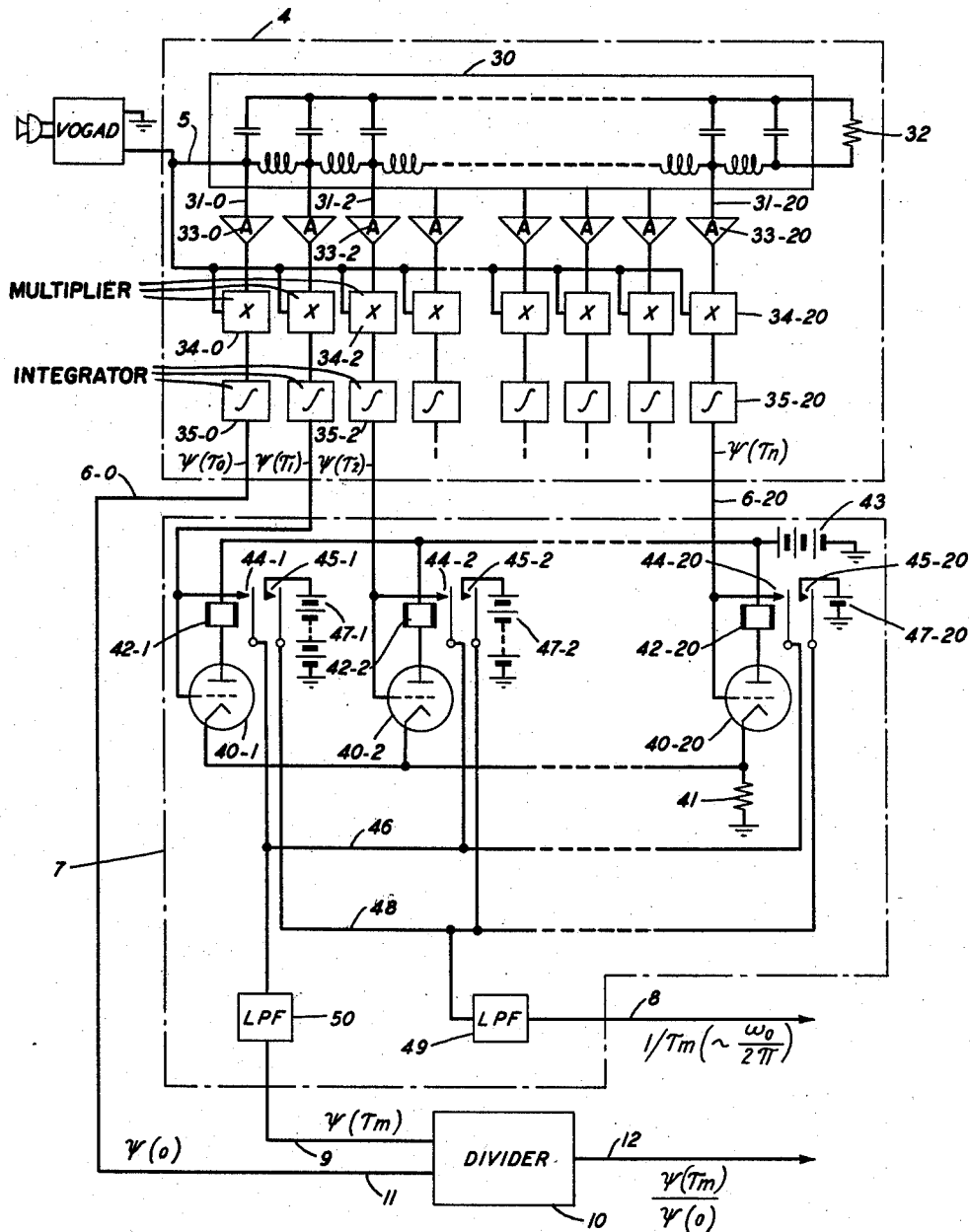
2,908,761

VOICE PITCH DETERMINATION

Filed Oct. 20, 1954

3 Sheets-Sheet 2

FIG. 2



INVENTOR
G. RAISBECK
BY
Harry C. Hart
ATTORNEY

Oct. 13, 1959

G. RAISBECK

2,908,761

VOICE PITCH DETERMINATION

Filed Oct. 20, 1954

3 Sheets-Sheet 3

FIG. 3

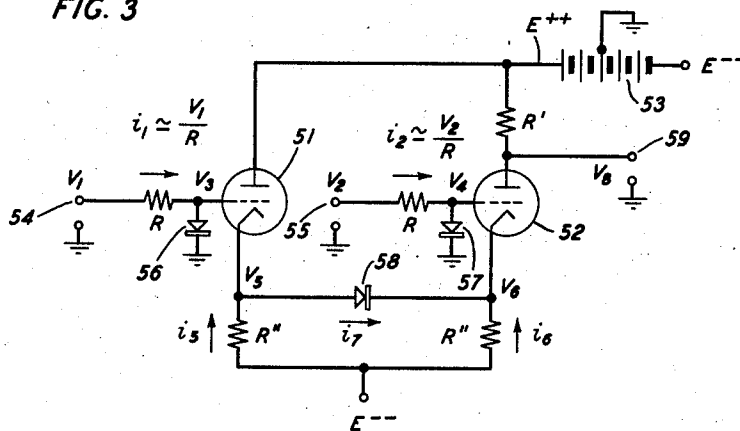
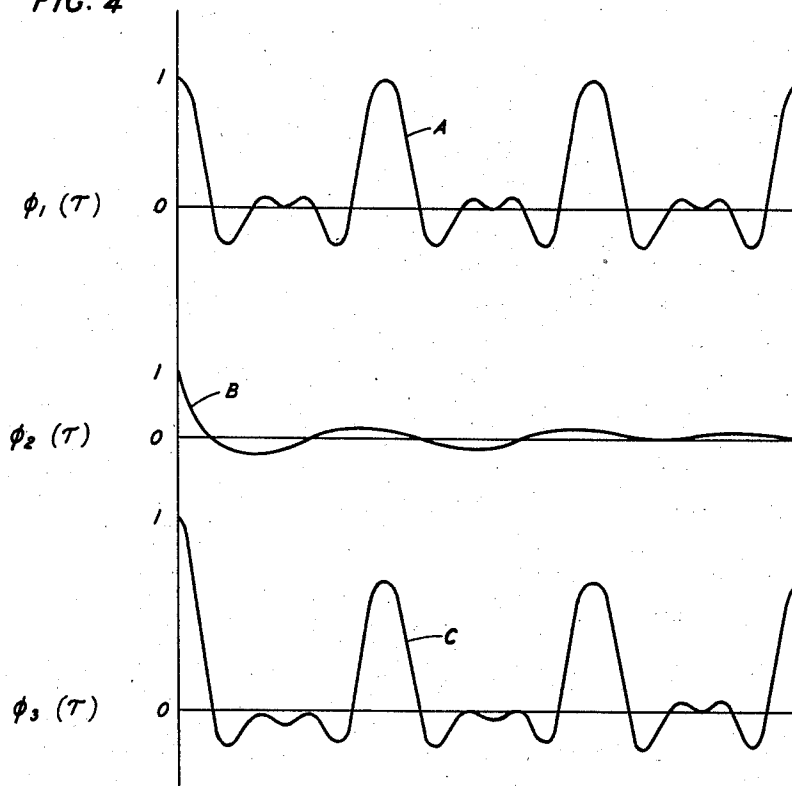


FIG. 4



INVENTOR
G. RAISBECK
BY Harry C. Hart
ATTORNEY

1

2,908,761

VOICE PITCH DETERMINATION

Gordon Raisbeck, Bernards Township, Somerset County, N.J., assignor to Bell Telephone Laboratories, Incorporated, New York, N.Y., a corporation of New York

Application October 20, 1954, Serial No. 463,467

16 Claims. (Cl. 179—15.55)

This invention relates to the transmission of speech over narrow band media by vocoder techniques. Its principal object is to improve the accuracy and naturalness of the reproduced speech.

In the vocoder transmission system of Dudley Patent 2,151,091, an input speech wave is analyzed to determine its fundamental frequency or pitch and the distribution of amplitudes among a number of frequency sub-bands into which the speech frequency range is divided. The result of this analysis is translated into a number of control currents each of which, over and above the first, represents the speech energy in one sub-band, while the first control current represents its fundamental frequency or pitch. These control currents are transmitted to a synthesizer and are there utilized to build up, from sources of energy in the synthesizer, an artificial speech wave having the characteristic pitch and amplitude-frequency distribution of the original impressed speech. More particularly, the synthesizer apparatus includes a source of periodic energy, commonly known as a buzz source, for reproducing voiced sounds, and a source of aperiodic energy, commonly known as a hiss source, for reproducing unvoiced sounds, and the outputs of these sources are applied, in the alternative, to the several members of a bank of filters, each of which embraces one of the sub-bands into which the voice frequency range is divided. The sub-band control currents derived by the analyzing apparatus control the gain or attenuation in these several sub-band paths. The choice between the buzz and the hiss source is controlled by one characteristic of the pitch control current while the fundamental frequency of the buzz source is controlled by another characteristic of the pitch control current. In other words, the pitch control current tunes the buzz source to the fundamental frequency of the original voice, turns the buzz source on for a voiced sound, indicated by a strong pitch control current, and turns the buzz source off and the hiss source on for an unvoiced sound indicated by a weak pitch control current.

With such apparatus every sound of normal speech is, in the synthesizer apparatus, classed either as a voiced sound or as an unvoiced sound. While such classification is adequate in some cases there are many other cases in which it is inadequate. For example, the sound of the letter Z in the normal English or American pronunciation of the word "azure" partakes equally of the nature of a voiced sound and of an unvoiced sound; so that any classification of such a sound either as voiced or as unvoiced but not both is arbitrary and artificial. Heretofore this artificial classification of all sounds into one or the other of two mutually exclusive categories has been an unfortunate necessity imposed by the lack of any measure of the relative amounts of the original speech energy of the two kinds, periodic or aperiodic.

Another defect of present vocoder systems is that the derivation at the analyzer station of a reliable pitch control current has always presented a difficult problem to the engineer. Many voices are so rich in harmonic com-

2

ponents that the energy of the fundamental component is small in comparison with the harmonic energy, and is therefore difficult to segregate. Under some conditions the energy at the fundamental frequency disappears entirely and resort must be had to indirect measures, such as the intermodulation of adjacent harmonic components, to derive a difference frequency. Aside from the apparatus complexities entailed, such difference frequency is a true measure of the voice pitch only in the case of a steady sound, while variations of frequency and of phase of the intermodulated components in the course of inflection causes such instantaneous frequency to be wholly inadequate.

The present invention approaches the problem of pitch determination in the time domain instead of the frequency domain; i.e., it seizes hold of the fundamental period of the voice instead of its fundamental frequency, and tracks it; i.e., continues to hold it, as it changes. It is characteristic of a periodic wave, no matter how complex that, after a certain time interval known as the period its form is a repetition of what has gone before. In the case of an exactly periodic wave the repetition is exact. In the case of a nearly periodic wave (and syllabic rates in speech are so slow compared with voice frequencies of interest that every voiced speech wave is periodic or nearly periodic) the repetition is inexact and approximate, but nevertheless easily recognizable. Such repetition or near repetition of the waveform in successive periods holds good quite aside from the existence of physical energy at the fundamental frequency.

The invention turns these considerations to account by dividing the voice wave into two paths, delaying the energy in one with respect to that in the other by a controllable amount, comparing the delayed wave with the undelayed wave, varying the amount of delay until a best match is obtained, and noting the corresponding amount of delay, with the recognition that this amount of delay is, identically, the fundamental period. A period control current may then be derived which would serve as well as a pitch control current, the period and the pitch being reciprocals of each other. It is preferred, however, in order that presently known buzz source frequency control apparatus may be employed without change, to derive in the first instance a control current which is reciprocally related to the observed fundamental period of the voice; i.e., it is directly proportional to the voice pitch. This indirectly derived pitch control current is indistinguishable from the pitch control current of the prior art except in respect to its greater reliability, and may be employed in the presently known fashion.

There are in general two physical causes for inexactness of the match between the waveform of any fundamental period and that of the prior period. The first cause has to do with the fact that the length of the period changes from period to period in the course of inflection. Inflection rates in voiced speech sounds are so slow, compared with their fundamental frequencies, that this cause is responsible for only a negligible departure from exactness of match, provided only that the apparatus is capable of holding and tracking the fundamental period through its changes.

The second cause has to do with the fact that no speech sound is wholly a voiced sound but partakes more or less of the nature of an unvoiced sound. In other words, any speech sound may be regarded as comprising the sum of two components of which one is periodic while the other is aperiodic; and the amount by which the best match between the delayed wave and the undelayed wave falls short of perfection is a measure of the aperiodic or unvoiced energy which the sound contains. In accordance with another feature of the invention, therefore, and in addition to determining the delay

for which the best match is obtained, there is also determined the degree to which this best match fails of perfection. From this determination a novel control current is derived. It is transmitted to the synthesizer station where it operates to control the relative amounts of energy supplied, respectively, by the buzz source and the hiss source to the filter bank of the spectrum synthesizer apparatus. Thus the switching operation as between these two sources, which characterizes the prior art, is completely dispensed with. The buzz source and the hiss source remain connected to the filter bank at all times and feed the filters of the bank with their respective energies in the proper amounts under control of this novel control current and therefore in accordance with the distribution of energy in the original speech as between its periodic component and its aperiodic component.

It is well known that a quantitative measure of the degree to which any wave matches any other wave is found in the cross-correlation function of these two waves. When one of the waves in question is a delayed replica of the other, the function in question is known as the auto-correlation function. The present system employs auto-correlation techniques for the determination of the fundamental period.

The invention will be fully apprehended from the following detailed description of a preferred illustrative embodiment thereof taken in connection with the appended drawings in which:

Fig. 1 is a block schematic diagram showing a vocoder transmission system embodying the invention;

Fig. 2 is a block schematic diagram showing apparatus for deriving a speech period control current and an energy distribution control current;

Fig. 3 is a schematic circuit diagram showing a divider for use in the combination of Fig. 2; and

Fig. 4 is a set of curves which are referred to in the explanation of the invention.

Referring now to the drawings and in particular to Fig. 1, speech currents, which may originate in a telephone instrument 1 are first passed through a unit which acts to hold them to an approximately constant level suitable for the operations of the remainder of the apparatus. It may be a voice-operated gain adjusting device 2 now commonly known as a "Vogad." The output of the vogad is applied in parallel to apparatus shown in the upper part of the figure which derives the novel period control current and the novel energy distribution current, and to conventional spectrum analyzer apparatus shown in a broken line box 3 in the lower part of the figure. These energy paths, like all others subsequently to be described, are shown by single lines in the drawing merely in order to avoid complexity. It will be obvious to those skilled in the art at what points wire pairs or other complete circuits may be required in practice.

The novel apparatus shown in the upper left-hand part of the figure is described below in connection with Fig. 2 which shows its details. For the present it suffices to note that it comprises an autocorrelator having a single input point and a number of output points 6-1 to 6-20, a maximum value selector 7 having a number of input points 6-2 to 6-20 and two output points 8, 9, and a divider 10 having two input points 9, 11 and a single output point 12. The number of output points 6 of the autocorrelator 4 exceeds by one the number of input points of the maximum value selector 7, and each of these, other than the first, is connected to one such input point. The first output point 6-1 of the autocorrelator 4 is connected directly to one input point 11 of the divider 10 while the other input point of the divider 10 is furnished by one of the two output points 9 of the maximum value selector 7. The remaining output point 8 of the maximum value selector 7

carries a signal which may be utilized at the receiver station without further change.

The autocorrelator 4 measures the autocorrelation $\psi(\tau)$ of the input signal for various discretely different amounts of relay τ , including zero, and delivers on each of its output points 6 a signal which is proportional to the autocorrelation for one such delay value. The maximum value selector 7 picks out the greatest among these, $\psi(\tau_m)$ from among all the others and supplies it as one input to the divider 10. At the same time the maximum value selector 7 acts to tag or identify the delay value τ_m for which this greatest autocorrelation was obtained and to deliver an additional identifying signal for use without further change at the synthesizer station. As a matter of convenience the identifying signal is made inversely proportional to the identified delay and therefore directly proportional to its reciprocal. Because the delay for which the autocorrelation is greatest is equal to the fundamental period of the speech, its reciprocal, $1/\tau_m$, is equal to the pitch frequency.

The conventional spectrum analyzer 3 shown in the broken box in the lower part of Fig. 1 comprises a bank of bandpass filters 15 connected in parallel, each of which is proportioned to pass a preassigned sub-band of the voice frequency band of interest, while together they pass the entire band. For the sake of illustration, ten such filters are indicated, the first two and the last being shown. Each such filter 15 is followed by a detector 16 which in turn is followed by a low-pass filter 17. The control current output of each of these several low-pass filters 17 is thus a measure of the voice energy in that sub-band to which such low-pass filter is connected.

These spectrum control currents are transmitted by any desired means, indicated by conductors 18, 19, to a receiver station which comprises conventional synthesizing apparatus shown in the broken line box 20 at the lower right-hand part of the figure. This synthesizer comprises a number of filters 21 having their output terminals connected in parallel to a reproducer 22. These several filters are proportioned to exhibit transmission characteristics like those of the several analyzing filters 15. Shaping networks 23 precede the several filters 21. Each shaping network 23 is supplied, by way of a conductor 24, with locally generated energy from a hiss source 25 and a buzz source 26, while its transmission is modulated by the control current derived at the analyzer station by a corresponding one of the filters 15, 17 and transmitted over the intervening channels 18, 19. Aside from the fact that the locally generated energy supplied to the several shaping networks 23 is continuously a mixture of the outputs of a buzz source 26 and a hiss source 25 in varying proportions, this spectrum reconstruction apparatus is conventional.

Turning now to Fig. 2, the autocorrelator 4 itself is a variant of one shown in Bennett et al. Patent 2,676,206. It comprises a delay device 30 such as an electromagnetic transmission line. The latter may comprise, for example, forty similar sections of series inductance and shunt capacitances having twenty evenly spaced taps 31; i.e., one located at every other section. The line may be terminated in well known fashion for no reflection by a resistive load 32. In terms of propagation time along the transmission line, the spacing between each tap 31 and the next one may be 500 microseconds, the total delay for all 40 sections being thus 10,000 microseconds or 20 milliseconds.

It is well known that the autocorrelation of a wave for any particular value of delay τ is obtainable by multiplying the delayed wave by the undelayed wave and integrating the product. To this end each of the output taps 31 of the delay line 30 is connected, by way of a buffer amplifier 33, to one input point of a multiplier 34, while the undelayed wave is supplied from the input terminal 5 of the line to the remaining input

points of all of these multipliers 34 in parallel. The products of the delayed signal by the undelayed signal thus formed by each of these multipliers for each particular value of delay τ , as represented by the locations of the several taps 31 along the transmission line 30, is now averaged in the time domain by an integrator 35. The output of each such integrator is thus the autocorrelation $\psi(\tau)$ of the signal for a particular delay τ . In general, the outputs of the several integrators 35 differ from each other in magnitude, some one being greater than the others. This greatest signal is a measure of the best match between the undelayed signal and one of the delayed signals, and the location along the transmission line 30 of the tap 31 on which it appears is a measure of the corresponding delay τ_m .

These several autocorrelation outputs $\psi(\tau_1)$, $\psi(\tau_2)$, $\psi(\tau_3)$, and so forth, other than the first one $\psi(\tau_0)$, are connected to the several input terminals of a maximum value selector 7. The latter may be a variant of one shown in Davis et al. Patent 2,646,465, and may comprise a plurality of amplifiers, e.g., vacuum tubes 40 of which all the cathodes are connected together and by way of a load resistor 41 to ground, while their anodes are connected by way of individual relay windings 42 to the positive terminal of a source 43 whose negative terminal is grounded. The several input terminals of the selector 7 are provided by the control grids of these amplifiers 40, which must of course be adjusted in their steady potentials by appropriate bias circuits not shown. Each relay 42 is provided with two pairs of contacts 44, 45, shown in their open positions. In each case the inner contact pair 44 acts, when pulled up, to apply the input signal directly to a first common bus 46, while the second contact pair 45 acts, when pulled up, to apply the potential of an individual battery 47 to a second common bus 48. The potential of each of the several batteries 47 differs from that of every other battery 47, and these potentials are selected in accordance with a pattern to be described subsequently.

In operation, let it be assumed that the match of the undelayed signal with the signal delayed by transmission over the line to the third tap 31—2 is more perfect than the match of the undelayed signal with the delayed signal as it appears at any other tap 31. In this case the signal applied to the grid of the second tube 40—2 of the selector 7 exceeds in magnitude the signal applied to the grid of any other tube 40 of the array. While conduction may well start in others of these tubes, the greater anode current of the second tube, flowing to ground by way of the cathode resistor 41, establishes a voltage drop across this resistor sufficient to raise the potentials of the cathodes of all of these tubes 40 to levels such that the conduction of all tubes other than the second is negligible. Since the second tube 40—2 is the only one which conducts in a significant amount, the contacts of the second relay 42—2 are pulled up and no others. By the closing of the inner contacts 44—2 a signal proportional to the autocorrelation $\psi(\tau_2)$ for the second delay value τ_2 is applied directly to the first common bus 46. By the closing of the outer contacts 45—2 the potential of the second battery 47—2 is applied directly to the second common bus 48.

Thus, under each of n different possible conditions, where n is the number of different values of delay τ which the autocorrelator 4 provides, only one of the tubes 40 conducts significantly, and one and only one relay contact pair 44, 45 is closed, thus placing on the first common bus 46 the magnitude of the autocorrelation $\psi(\tau)$ for a particular delay and, on the second common bus 48, the potential of a particular battery 47 which is uniquely associated with that delay. Furthermore the autocorrelation $\psi(\tau)$ and the corresponding delay τ thus selected are the maximum autocorrelation, $\psi(\tau_m)$ and the corresponding delay τ_m .

In order that conventional receiver apparatus may be employed with a minimum of change, the potentials of the several batteries 47 are selected in inverse proportion to the several values of delay τ with which they are associated. This is indicated on the output lead of the second common bus 48 by a legend showing that the signal thereon is the reciprocal of that delay τ_m for which the autocorrelation is a maximum. From what has been stated above this is evidently proportional to the fundamental frequency or pitch of the speech.

By employing a very large number of delay line taps 31 and a correspondingly large number of recognizing circuits in the maximum value selector 7 the variations of this signal may be made substantially continuous. With a smaller number, e.g., twenty of such circuits as contemplated, this output signal naturally contains somewhat abrupt transitions. These, however, may readily be eliminated by the interposition of a low-pass filter 49.

Similarly the changes, as operation proceeds, from recognition of one autocorrelation signal as the greatest to recognition of another one as the greatest, is reflected in abrupt transitions in the magnitude of the signal on the first common bus 46. As before, these transitions may in principle be reduced as far as desired by the employment of a large number of delay line taps 31 in the autocorrelator 4 and a correspondingly large number of recognizing circuits in the selector. In practice, however, it is sufficient to interpose a low-pass filter 50.

The filtered signal from the first common bus 46 is applied to one of the two input points 9 of a divider 10 while the autocorrelation $\psi(\tau_0)$ of the speech wave for zero delay is applied to the other input point 11 of this divider. The divider 10 itself may be any unit which forms on its output lead 12 the quotient of a signal applied to its upper input lead 9 divided by another signal applied to its lower input lead 11. Various constructions are possible, a suitable one being shown in Fig. 3 which carries out a logarithmic division operation by electronic means.

That is to say it carries out the following operation:

(1) It takes the logarithm of the signal applied to its upper input lead;

(2) It takes the logarithm of the signal applied to its lower input lead;

(3) It subtracts the second result from the first result; and, finally,

(4) It takes the co-logarithm of the result of operation No. 3.

The first principle which is turned to account in the divider of Fig. 3 is the well established relation which holds for a germanium or silicon diode rectifier, namely, that the current through it is an exponential function of the voltage across it. More exactly and as expounded in standard texts, such as "Crystal Rectifiers," by H. C. Torrey and C. A. Whitmer, Rad. Labs. M.I.T. Series, volume 15, especially at page 82

$$i = i_0 \left(e^{\frac{qV}{kT}} - 1 \right) \quad (1)$$

where

V is the applied voltage,

e is the base of Napierian logarithms,

T is the absolute temperature,

k is Boltzmann's constant,

q is the electronic charge,

i_0 is a constant; i.e., the current which flows in the absence of any applied voltage V .

From this it follows that the voltage is proportional to the logarithm of the current; i.e.,

$$V \simeq \frac{kT}{q} \log \frac{i}{i_0} \quad (2)$$

These relations, known for some years to hold exactly in principle, have more recently been found to hold to

a very good approximation in practice, especially with rectifiers of the P-N junction variety. See, for example, a note by Goucher, Pearson, Sparks, Teal and Shockley, published in the Physical Review for 1951, volume 81, No. 4, page 637.

These properties of P-N junction rectifier diodes are turned to account in the construction of the divider circuit of Fig. 3 which comprises two vacuum tube triodes 51, 52, the anode of each being connected to the positive terminal E⁺⁺ of an operating potential source 53 and the cathode of each being connected by way of a resistor R'' to its negative terminal E⁻. These resistors R'' are of sufficient magnitude, as compared with the internal resistances of the tubes 51, 52, to provide substantial cathode follower action. One input signal V₁, which is to serve as the dividend appears at a first input terminal 54 and is applied by way of a resistor of magnitude R to the control grid of the left-hand tube 51. A second input signal V₂ which is to serve as the divisor appears at another input terminal and is applied by way of another resistor, also of magnitude R, to the control grid of the right-hand tube 52. Each of these control grids is returned, by way of a P-N junction rectifier diode 56, 57 to a suitable point of intermediate potential of the source 53, here shown as ground. With these connections cathode current flows to each of the tubes 51, 52 in an amount sufficient to hold its control grid, in the absence of an input signal, at cutoff, while the input rectifiers 56, 57 are operated in their forward directions, so that their resistances are small compared with the resistors R to which they are connected. A third P-N junction rectifier diode 58 interconnects the two cathodes, and an output signal is derived across a resistor R' in series with the anode of the right-hand tube 52; i.e., between the right-hand tube anode and ground.

The input diode current is in each case, therefore, proportional to the input signal voltage; i.e.,

$$i_1 \approx \frac{V_1}{R} \quad (3)$$

$$i_2 \approx \frac{V_2}{R} \quad (4)$$

Hence, from Equation 2 the grid voltages applied to the two tubes 51, 52 are given by

$$V_3 = \frac{kT}{q} \log \frac{i_1}{i_0} \quad (5)$$

and

$$V_4 = \frac{kT}{q} \log \frac{i_2}{i_0'} \quad (6)$$

where i_0 and i_0' are the currents which flow in the first diode 56 and in the second diode 57, respectively, when large reverse voltages are applied.

By virtue of the cathode follower action the cathode potentials are in turn given by

$$V_5 \approx V_3 \approx \frac{kT}{q} \log \frac{i_1}{i_0} \quad (7)$$

$$V_6 \approx V_4 \approx \frac{kT}{q} \log \frac{i_2}{i_0'} \quad (8)$$

From elementary circuit considerations, the voltage drop across the third diode is equal to

$$V_5 - V_6$$

From Equation 1 the current through the third diode 58 is an exponential function of the voltage drop across it; i.e.,

$$i_7 = i_0'' e^{\frac{q}{kT}(V_5 - V_6)} \quad (9)$$

where i_0'' is the current which flows in the third diode 58 in the absence of an applied voltage. But when Equation 8 is subtracted from Equation 7 as required by the ex-

ponent in Equation 9 and the logarithm of both sides is taken, there results

$$i_7 = \frac{i_0'' i_0'}{i_0} \frac{i_1}{i_2} = \frac{i_0'' i_0'}{i_0} \frac{V_1}{V_2} \quad (10)$$

that is to say, the current through the third diode 58 is proportional to the quotient of the dividend signal by the divisor signal.

It remains to extract an output proportional to this third diode current without upsetting the operation of the circuit. This may be done by the inclusion of the load resistor in series with the anode of the right-hand tube 52. Its magnitude R' is preferably substantially smaller than that of the other resistors discussed heretofore. With this arrangement the voltage V₈ is evidently equal to the positive potential of the operating source 53, reduced by the voltage drop across this load resistor; i.e.,

$$V_8 = E^{++} - R'(i_6 + i_7) \quad (11)$$

From elementary circuit considerations, taking the magnitudes of currents and voltages as indicated on the drawing into account, this in turn is equal to

$$V_8 \approx E^{++} - \frac{R'}{R''} E^{-} - \left[\frac{R' i_0'' i_0'}{i_0} \right] \frac{V_1}{V_2} \quad (12)$$

In Equation 12 the first and second terms are constants and may be eliminated by standard well known means, while the third term is directly proportional to the desired quotient.

Hence, when the maximum autocorrelation signal $\psi(\tau_m)$ is applied to the left-hand input terminal 54 of Fig. 3 (the upper or dividend terminal of the divider 10 of Figs. 1 and 2) and the correlation signal $\psi(\tau_0)$ for zero delay is applied to the right-hand input terminal 55 of Fig. 3 (the lower or divisor terminal of the divider 10 of Figs. 1 and 2), the required quotient appears at the output terminal 59 of the divider (the terminal 12 of Figs. 1 and 2), namely the ratio of the maximum correlation $\psi(\tau_m)$ to the zero delay correlation $\psi(\tau_0)$.

The significance of these relations will readily be understood from a consideration of Fig. 4, wherein the curve A shows a representative correlation curve, normalized to unity value at zero delay, for a fully periodic signal of a certain average amount of complexity. This fully periodic signal may be termed $f_1(t)$ and its autocorrelation, curve A, may be termed $\phi_1(\tau)$. Curve A shows four maxima of equal heights located at delays of 0, τ , 2τ , and 3τ . The time function represented thus extends over at least three full periods, each one being exactly like the last. The fact that the curve A has the value unity for zero delay expresses the truism that the time function $f_1(t)$ is exactly like itself in the absence of delay. The fact that maxima occur in the autocorrelation curve A at intervals of τ reflects the fact that each full period of the time function $f_1(t)$ resembles its predecessor and that the delayed signal finds its best match with the undelayed signal for discrete values of the delay equal to τ , 2τ , 3τ and so on. The fact that these successive maxima rise to the same height as the maximum at zero delay reflects the fact that the successive periods of the time function $f_1(t)$ are exactly alike, so that for these values of delay the match is perfect.

Curve B of Fig. 4 shows the normalized autocorrelation $\phi_2(\tau)$ of an aperiodic signal $f_2(t)$, such as noise. It, too, has the value unity for zero delay, which reflects the truism that even a noise signal is an exact replica of itself. However, and as distinguished from the curve A, the curve B falls steeply and soon tapers off to zero, performing oscillations of small amplitude and of random distribution within a short time.

When to a periodic signal $f_1(t)$ there is added an aperiodic signal $f_2(t)$, an almost periodic signal $f_3(t)$ results. In general, any portion of a speech wave is such an almost periodic function and may therefore be regarded as composed of two parts in additive combination,

the first part being a periodic function $f_1(t)$ and the second being an aperiodic function $f_2(t)$.

While it is not exactly true that the autocorrelation $\varphi_3(t)$ of the sum of two signals is equal to the sum of the autocorrelations of the signals individually, it is sufficiently nearly true to permit a graphic representation such as that of curve C of Fig. 4, which shows the autocorrelation $\varphi_3(\tau)$ of the time function or sum signal $f_3(t)$.

The mathematical formulation of the unnormalized autocorrelation of a time function $f_1(t)$ is well known and is stated in Bennett et al. Patent 2,676,206 and in Lee et al. Patent 2,643,819, as well as elsewhere. For the purposes of these patents and in other circumstances in which comparison of autocorrelations is not needed, normalization is unnecessary. In the present situation, on the other hand, normalization of the autocorrelation to the value unity at zero delay is of assistance in the actualization of further operations of the apparatus.

The normalized autocorrelation of a time function $f(t)$ is given by the expression

$$\varphi(\tau) = \frac{\lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T}^0 f(t)f(t+\tau) dt}{\lim_{T \rightarrow \infty} \frac{1}{T} \int_{-T}^0 [f(t)]^2 dt} \quad (13)$$

It is evidently independent of the magnitude of $f(t)$, being dependent only on its form.

When for the general time function $f(t)$ there is substituted in Equation 13 the specific time function of interest, namely

$$f_3(t) = f_1(t) + f_2(t) \quad (14)$$

there results

$$\varphi_3(\tau) = \frac{\langle f_1 \rangle^2 \varphi_1(\tau) + \langle f_2 \rangle^2 \varphi_2(\tau)}{\langle f_1 + f_2 \rangle^2} + \frac{[\langle f_1 + f_2 \rangle^2 - \langle f_1 \rangle^2 - \langle f_2 \rangle^2] \varphi_{12}(\tau)}{\langle f_1 + f_2 \rangle^2} \quad (15)$$

where $\langle \rangle$ indicates the rms value and φ_{12} is the crosscorrelation of f_1 and f_2 .

To a good approximation,

$$\langle f_1 + f_2 \rangle^2 = \langle f_1 \rangle^2 + \langle f_2 \rangle^2 \quad (16)$$

for the reason that f_1 is purely periodic and f_2 is pure noise. Also, under these circumstances $\varphi_{12}(\tau)$ is small. Hence, to a good approximation,

$$\varphi_3(\tau) \approx \frac{\langle f_1 \rangle^2}{\langle f_1 \rangle^2 + \langle f_2 \rangle^2} \varphi_1(\tau) + \frac{\langle f_2 \rangle^2}{\langle f_1 \rangle^2 + \langle f_2 \rangle^2} \varphi_2(\tau) \quad (17)$$

From the foregoing it is plain that, taking the undelayed autocorrelation as having the value unity, the successive maxima of curve C rise to or very close to a height above the axis which is proportional to the ratio of the periodic energy in the speech to its entire energy. Looked at in another way, these maxima fall below the value unity by a distance which is proportional to the ratio of the aperiodic energy of the speech to its entire energy. Hence a measure of the amplitudes of these maxima, as compared with the amplitude of the autocorrelation curve for zero delay constitutes a measure of the relative proportions of buzz energy and hiss energy, respectively, required for proper and realistic synthesis of artificial speech. This justifies the transmission of the output of the divider 10, derived as aforesaid, to the receiver station without further change.

Returning now to Fig. 1, the pitch control signal $1/\tau_m$ is applied to the buzz source 26 to adjust its oscillation frequency in well known fashion and as shown, for example, in Reisz Patent 2,522,539. The energy ratio signal, namely, the divider output signal as it appears on the divider output terminal 12, is likewise transmitted

to the receiver station where it is employed to control the relative amounts of buzz and hiss furnished to the spectrum reconstructor in the following fashion.

The output of the buzz source 26 is delivered by way of a variable gain amplifier 61 to a combination point 62. This amplifier 61 may be of a well known construction such as to furnish a gain proportional to the magnitude of a gain control signal applied to its control terminal 63. The incoming energy distribution signal, termed $g(t)$ for short, is passed through a router 64 and applied to this control terminal 63. Hence the buzz component of the energy applied to the combination point 62 is proportional to the square root of g , and so to the amplitudes of the periodic components of the speech.

The signal $g(t)$ is also passed through a phase inverter 65 which converts it to $-g(t)$ and then through a steady potential source 66 of one volt connected in series. Thus the signal on the lead 67 beyond the source is equal to $1-g(t)$. This in turn is passed through a router 68 and applied to the gain control terminal 69 of a second variable gain amplifier 70 which is connected in tandem with the hiss source 25. At this point the gain control signal has the form

$$\sqrt{1-g(t)}$$

Thus the energy which appears at the combination point 62 comprises an additive mixture of periodic energy from the buzz source 26 and aperiodic energy from the hiss source 25 in proportions as called for by the energy distribution of the voice and as controlled by the output of the divider 10. The periodic energy of the buzz source 26 is tuned to the required pitch frequency as stated above by the first of the two outputs of the maximum value selector 7. This additive combination is now fed to the Shaping Networks 23 in parallel, where the slowly varying spectrum control currents operate to control the magnitudes of the several frequency sub-bands which collectively constitute the synthesized signal.

Because of the nature of the operation of the divider 10 its output is unchanged when its two inputs are increased or reduced in the same ratio. To this extent it supplies an output which is a measure of the relative magnitudes of its two inputs and is otherwise independent of such magnitudes. In other words it carries out a normalizing function in addition to its assigned dividing function. Provided such normalization as between the autocorrelation signal for zero delay and the maximum autocorrelation signal be otherwise obtained, the energy distribution signal $g(t)$ or a near equivalent thereof can readily be derived by a subtraction process.

What is claimed is:

1. In a system for artificial production of speech, the combination which comprises a speech analyzer station having means for deriving from a speech sound a pitch control signal, an energy distribution signal which is representative, for each sound, of its proportional content of periodic energy, and a plurality of spectrum control signals, and a reproducer station having a buzz source, a hiss source and a spectrum synthesizer, means for transmitting all of said control signals to said reproducer station, means for tuning the buzz source under control of the pitch control signal, means controlled by said energy distribution signal for mixing the outputs of said sources in proportions determined by the magnitude of said energy distribution signal, means for applying said mixed outputs to said spectrum synthesizer, and means for controlling said spectrum synthesizer under control of said spectrum control signals.

2. In a system for deriving control signals to control the artificial production of speech, means for analyzing a speech sound, means for deriving from said analysis a pitch control signal representative of the fundamental frequency of said speech sound, means for deriving from said analysis a plurality of spectrum control signals, each

11

representative of the speech energy falling within one of a plurality of frequency sub-bands which collectively embrace the frequency band of said speech sound, means for deriving from said analysis a first measure of the entire energy of said speech sound, means for deriving from said analysis a second measure of the energy of the periodic components of said speech sound, and means for deriving from said two measures an additional signal representative of the distribution of the energy of said speech sound as between its periodic components and its aperiodic components.

3. Apparatus as defined in claim 1 wherein said means for deriving said pitch control signal comprises an autocorrelator and a maximum value selector.

4. Apparatus as defined in claim 1 wherein said means for deriving said pitch control signal comprises means for determining the autocorrelation of said speech sound for various delays, and means for selecting from among said various delays that one for which said autocorrelation as determined is a maximum.

5. In combination with apparatus as defined in claim 4, means for deriving a signal which is substantially inversely proportional to said selected delay.

6. In a system for deriving control signals to control the artificial production of speech, means for analyzing a speech wave, means for deriving from said analysis a pitch control signal representative of the fundamental frequency of said speech wave, means for deriving from said analysis a plurality of spectrum control signals representative of the speech energy falling within a plurality of frequency sub-bands which collectively embrace the frequency band of said speech wave, means for deriving from said analysis a first measure of the entire energy of said speech wave, means for deriving from said analysis a second measure of the energy of the periodic components of said speech wave, and means for deriving from said two measures an additional signal representative of the distribution of the energy of said speech wave as between its periodic components and its aperiodic components, said pitch control signal deriving means comprising means for delaying said wave by each of a plurality of different time lags distributed over a range extending substantially from zero to the longest period of said speech wave, means for individually comparing each delayed wave singly with the original undelayed wave, means for identifying that one of said different time lags for which the delayed wave, most nearly resembles the undelayed wave as determined by said comparison, means for deriving an auxiliary signal proportional to the duration of said identified time lag, and means for reciprocating said auxiliary signal to provide said pitch control signal.

7. Apparatus for deriving a desired control signal indicative of the fundamental frequency of a complex signal wave which comprises means for delaying said wave by each of a plurality of different time lags distributed over a range extending substantially from zero to the longest period of said complex signal wave, means for individually comparing each delayed wave singly with the original undelayed wave, means for identifying that one of said different time lags for which the delayed wave most nearly resembles the undelayed wave as determined by said comparison, means for deriving an auxiliary signal proportional to the duration of said identified time lag, and means for reciprocating said auxiliary signal to provide said desired control signal.

8. In a system for artificial production of speech sounds, a tunable source of periodic energy, a source of aperiodic energy, means for tuning said periodic energy source under control of a pitch control signal, means controlled by an energy distribution signal which is representative, for each sound to be reproduced, of its proportional content of periodic energy for mixing the outputs of said two sources in proportions determined by the

12

magnitude of said energy distribution signal, a plurality of filters having passbands contiguously located on the frequency scale and together embracing the frequency band of said speech sound, means for applying said mixed outputs to all of said filters, means for variably attenuating the energy path of each of said filters under control of a spectrum control signal, and means for reproducing the outputs of all of said filters as an artificial sound.

9. Apparatus for selecting from among a plurality of signals that one which has the greatest value which comprises a plurality of discharge devices each having an anode, a cathode and a control electrode, a first connection extending from all of said anodes to one terminal of an operating potential source, a second connection extending from all of said cathodes to one terminal of a common impedance element, a third connection extending from the other terminal of said common impedance element to the other terminal of said operating potential source, a like plurality of relays each having a winding connected in the anode circuit of one of said discharge devices, each of said relays being provided with two contact pairs, means for applying said input signals individually to the control electrodes of said discharge devices, a first common bus, individual energy paths extending from said first common bus through the first contact pair of each relay to the control electrode of the device in whose anode circuit said relay is connected, a second common bus, a like plurality of auxiliary potential sources, of successively greater potentials, associated respectively with the several relays, and an energy path extending from said second common bus through the second contact pair of each relay to its associated auxiliary source.

10. Apparatus having a first input point, a second input point and an output point for delivering at said output point a signal proportional to the quotient of a dividend signal by a divisor signal of which the dividend signal is applied to said first input point while the divisor signal is applied to said second input point which comprises a first element having an exponential current-voltage characteristic connected to the first input point, a second element having a similar exponential current-voltage characteristic connected to the second input point, means for applying said dividend and divisor signals as currents to said first and second input points, respectively, thereby to produce a first voltage which is proportional to the logarithm of said first input signal and a second voltage which is proportional to the logarithm of said second input signal, means for subtracting said second voltage from said first voltage, and a third element having a similar exponential current-voltage characteristic for deriving from the difference of said first and second voltages a current which is exponentially related to said voltage difference and hence directly related to the quotient of said dividend signal by said divisor signal.

11. In combination with a source of a speech wave, apparatus for continuously determining the continuously varying fundamental period of said speech wave which comprises means for delaying said wave by each of a plurality of different time lags distributed over a range extending substantially from zero to the longest period of said speech wave, means for individually comparing each delayed wave singly with the original undelayed wave, means for identifying that one of said different time lags for which the delayed wave most closely resembles the undelayed wave as determined by said comparison, means for rejecting all others of said time lags, means for continually altering the time lag identified to preserve said closest resemblance as said fundamental period changes, and means for developing a signal continuously representative of said varying identified time lag.

12. Apparatus as defined in claim 11 wherein said comparing means comprises means for determining the autocorrelation of said speech wave for various delays.

13

13. Apparatus as defined in claim 12 wherein said identifying means comprises means for selecting that value of the time lag for which said autocorrelation as determined is a maximum.

14. Apparatus as defined in claim 11 wherein said comparing means comprises means for multiplying the delayed wave by the undelayed wave to form, for each value of the time lag, a product.

15. In combination with apparatus as defined in claim 14, means for averaging each of said products.

16. In combination with a source of a speech wave, apparatus for continuously determining the continuously varying fundamental period of said speech wave which comprises means for deriving from said wave a plurality of variously delayed replicas of said wave, the lengths of the individual delays characterizing said replicas covering a range extending substantially from zero to the longest speech wave period, means for comparing each of said replicas with the undelayed wave, means for identifying the replica which most closely matches the orig-

14

inal wave as determined by said comparison, means for rejecting all others of said replicas, whereby the delay characterizing the replica thus identified at each moment is equal to the length of said fundamental period at that moment, means for continually altering said identification to preserve said closest match as said fundamental period changes, and means for developing a signal continuously representative of the delay characterizing the replica momentarily identified.

References Cited in the file of this patent

UNITED STATES PATENTS

2,098,956	Dudley	Nov. 16, 1937
2,243,526	Dudley	May 27, 1941
2,401,405	Bedford	June 4, 1946
2,508,620	Peterson	May 23, 1950
2,580,421	Guanella	Jan. 1, 1952
2,705,742	Miller	Apr. 5, 1955