

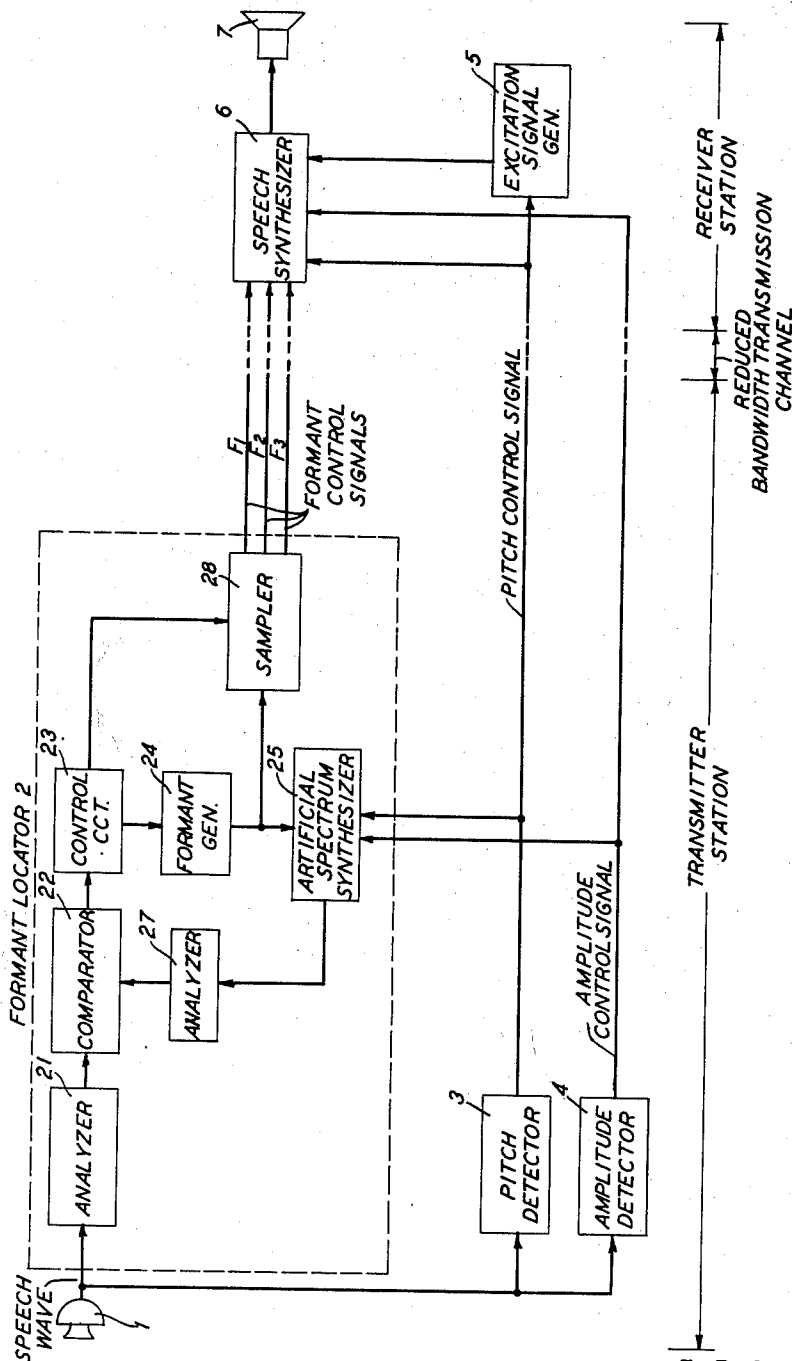
March 31, 1964

E. E. DAVID, JR., ETAL
AUTOMATIC FORMANT LOCATOR

3,127,477

Filed June 27, 1962

4 Sheets-Sheet 1



INVENTORS E. E. DAVID, JR.
M. V. MATHEWS
J. E. MILLER
BY G. E. Hirsch Jr.
ATTORNEY

March 31, 1964

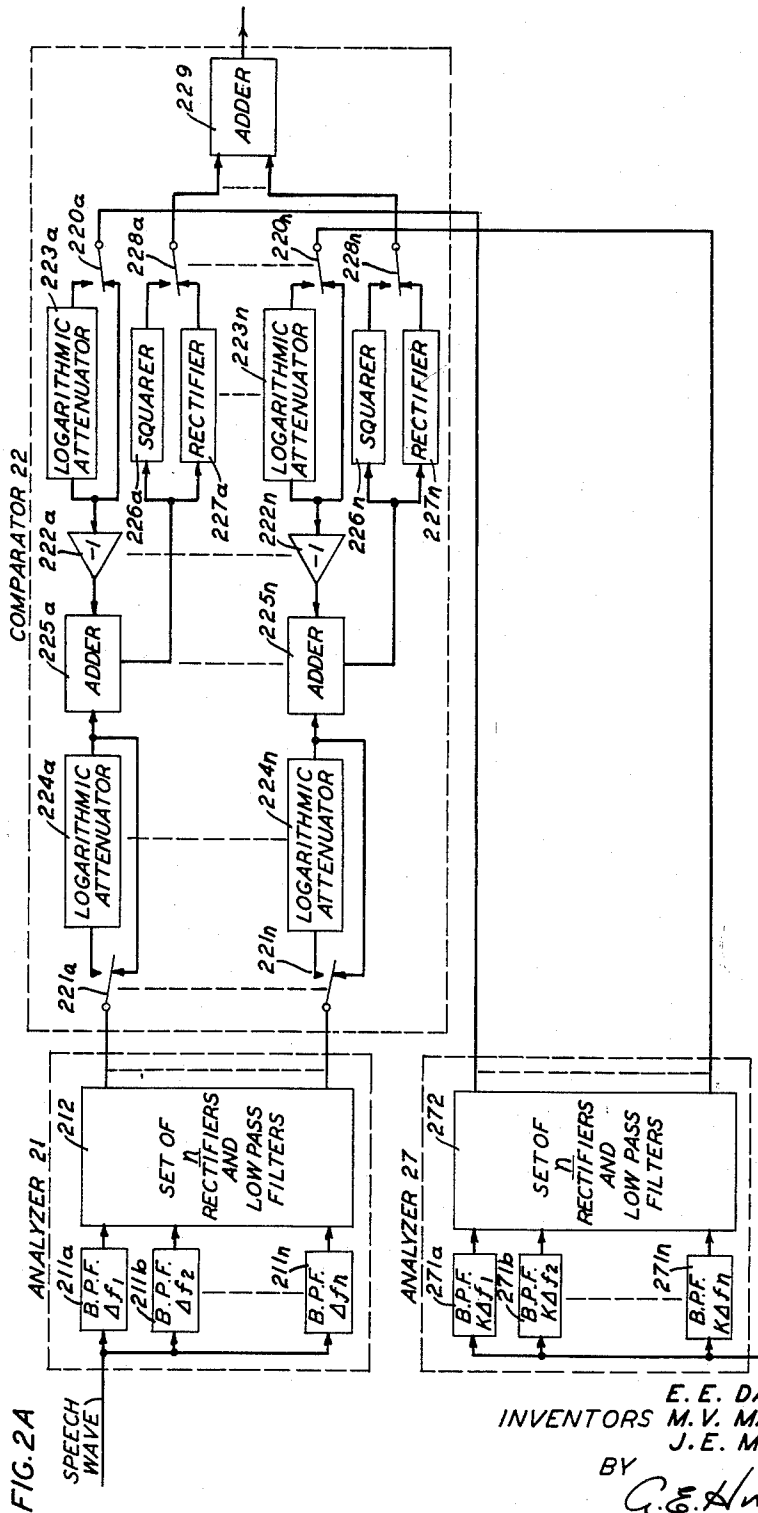
E. E. DAVID, JR., ETAL

3,127,477

AUTOMATIC FORMANT LOCATOR

Filed June 27, 1962

4 Sheets-Sheet 2



E. E. DAVID, JR.
INVENTORS M. V. MATHEWS
J. E. MILLER

BY *G. E. Hirsch Jr.*
ATTORNEY

March 31, 1964

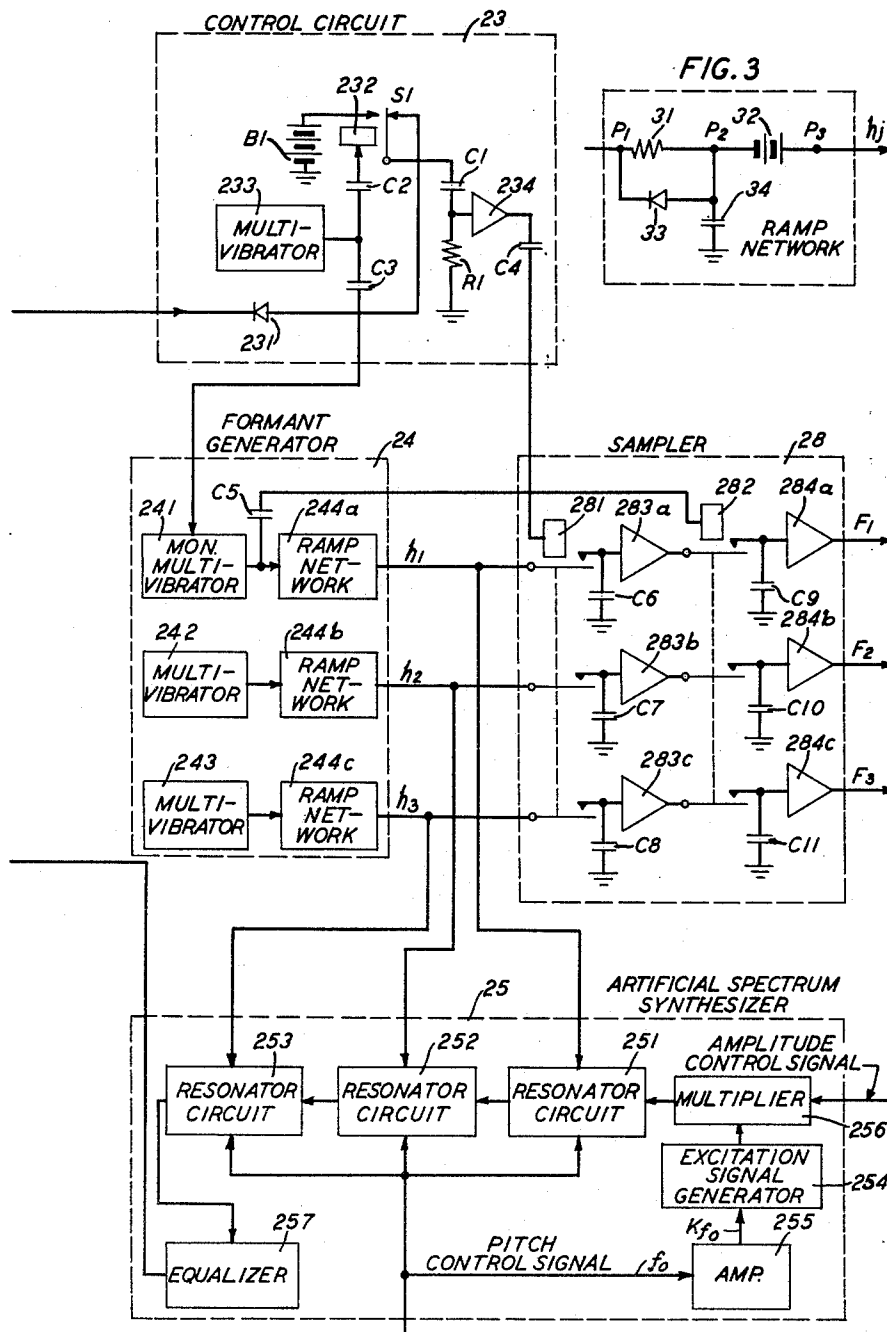
E. E. DAVID, JR., ETAL
AUTOMATIC FORMANT LOCATOR

3,127,477

Filed June 27, 1962

4 Sheets-Sheet 3

FIG. 2B



E. E. DAVID, JR.
INVENTORS
M. V. MATHEWS
J. E. MILLER

BY *G. E. Hirsch*
ATTORNEY

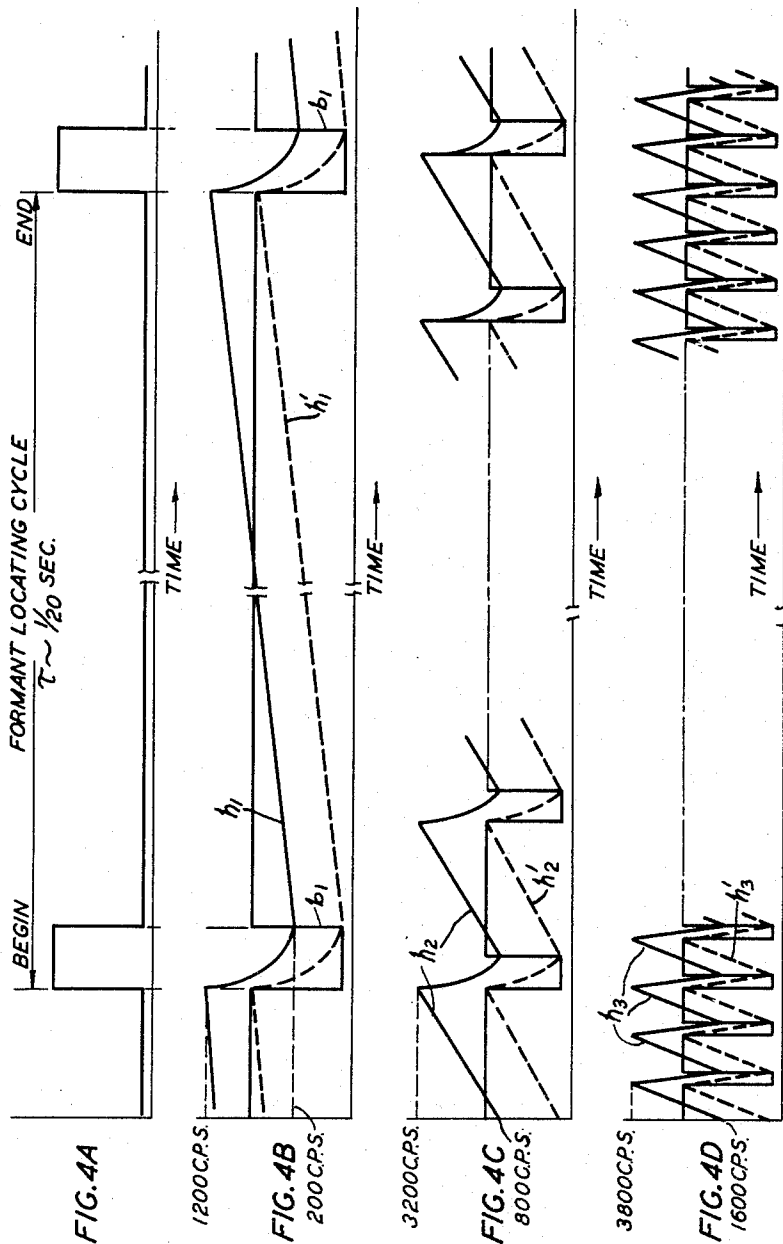
March 31, 1964

E. E. DAVID, JR., ETAL
AUTOMATIC FORMANT LOCATOR

3,127,477

Filed June 27, 1962

4 Sheets-Sheet 4



E. E. DAVID, JR.
INVENTORS M. V. MATHEWS
J. E. MILLER

G. E. Hirsch Jr.
ATTORNEY

1

2

3,127,477

AUTOMATIC FORMANT LOCATOR

Edward E. David, Jr., Berkeley Heights, Max V. Mathews, New Providence, and Joan E. Miller, Plainfield, N.J., assignors to Bell Telephone Laboratories, Incorporated, New York, N.Y., a corporation of New York

Filed June 27, 1962, Ser. No. 205,663
15 Claims. (Cl. 179-15.55)

This invention relates to speech systems in which the information content of a wide-band speech wave is represented by a group of narrow-band signals.

Of the various information-bearing characteristics of human speech, the locations of formants, or peaks in the envelope of the speech amplitude spectrum, are known to be among the most significant. Several speech communications systems have utilized formant locations to conserve transmission channel bandwidth, since formant locations may be specified by a small number of narrow-band control signals whose total bandwidth is substantially smaller than that of a typical speech wave, for example, a speech wave of telephone quality. One of the best known of these systems is the so-called resonance vocoder, several varieties of which are described by E. E. David, Jr., in *Signal Theory in Speech Transmission*, vol. CT-3, I.R.E. Transactions on Circuit Theory, page 239 (1956).

A problem common to systems employing a formant representation of speech information is the difficulty in locating quickly and accurately the frequencies at which formants occur. As explained in M. R. Schroeder Patent 2,857,465, issued October 21, 1958, a large part of this difficulty is attributable to the fact that formant locations are not fixed, but vary over wide frequency ranges. Further, adjacent formants vary within frequency ranges that overlap to a considerable extent, so that a formant may at times occur within the frequency range ordinarily occupied by one of its neighboring formants. As a result, prior systems that attempted to locate formants within fixed, preassigned frequency ranges tended to yield ambiguous results when two adjacent formants occurred within the frequency range assigned to only one of the formants. The present invention, however, undertakes to locate individual formants without ambiguity, even when adjacent formants lie within a frequency range ordinarily occupied by only one formant.

In this invention, formants are located automatically on the basis of a systematic comparison between the amplitude spectrum of an incoming speech wave and an artificial amplitude spectrum whose formants are continuously varied. The comparison is made by matching the frequency components of the speech spectrum against the corresponding frequency components of the artificial spectrum, with the formant locations of the best matching version of the artificial spectrum being selected to represent the true speech formant locations.

To make the spectrum matching process indicate only the degree to which the formant locations of the artificial spectrum approximate the unknown formant locations of the incoming speech spectrum, only the formant locations of the artificial spectrum are varied, and in every other respect the artificial spectrum is constructed to be substantially identical with the speech spectrum. This is accomplished by having the frequency components of the artificial spectrum occur at harmonics of the fundamental frequency of the incoming speech wave, and by causing the energy of the artificial spectrum to be equal to that of the incoming speech wave.

An important feature of this invention is its ability to provide a highly accurate, fully automatic indication of formant locations for a wide variety of voiced sounds and

for a wide variety of talkers. This feature is attained by constructing the artificial spectrum to have its formants occur at all possible combinations of locations, thereby obtaining at least one set of formant locations in the artificial spectrum which closely matches the unknown formant locations of the spectrum of the incoming speech wave. In the present invention, the artificial spectrum is constructed from a set of locally generated formant signals, each formant signal varying continuously over a preselected range of values corresponding to the range of all possible frequencies within which a particular formant may occur. To make the formants of the artificial spectrum occur at all possible combinations of locations, in each pair of adjacent formant signals the higher order formant signal varies over its entire range of values before the other, lower order formant signal has changed appreciably in value. Thus, in the case of an artificial spectrum constructed from a set of three locally generated formant signals, the second formant signal varies once over its entire range of values before the first formant signal has changed appreciably in value, while the third formant signal completes one entire variation in value before the second formant signal has appreciably changed in value. In this manner, by the time that the first formant signal has completed a single variation over its entire range of values, the second and third formant signals have each completed a large number of variations over their respective ranges of values, with the third formant signal completing approximately the same number of variations relative to the variations of the second formant signal as the second formant signal completes relative to the first formant signal. As a result, the formants of the artificial spectrum constructed according to the principles of this invention occur at substantially all possible combinations of locations during a single complete variation of the first formant signal.

Another important feature of the present invention is its ability to obtain very rapidly the formant locations of the best matching version of the artificial spectrum. This is accomplished by adjusting the rate of variation of the first formant control signal to be approximately equal to the rate at which speech formants typically change location in connected speech sounds, so that the first formant signal varies over its complete range of values in about the same interval of time that formants in the incoming speech wave typically change their locations. In addition, the spectrum matching process is repeated cyclically at a rate corresponding to the rate at which the first formant signal varies over its complete range of values. Thus, during each cycle, a separate determination of formant locations is made, and at the end of each cycle the values of the formant signals which produced the best matching version of the artificial spectrum during the cycle are converted into a group of narrow-band signals representing the speech formant locations for that cycle. Because the cycles recur at about the same rate as variations in speech formant locations, formants are located in this invention at a rate suitable for voice communications systems such as the telephone.

The invention will be fully understood from the following descriptions of illustrative embodiments thereof, taken in conjunction with the appended drawings, in which:

FIG. 1 is a block diagram showing the automatic formant locating apparatus of this invention in a resonance vocoder system;

FIGS. 2A and 2B are block diagrams showing in detail apparatus for locating formants in accordance with the principles of this invention;

FIG. 3 is a schematic circuit diagram of a ramp network useful in implementing the apparatus shown in FIG. 2; and

FIGS. 4A, 4B, 4C, and 4D are waveform diagrams

of assistance in explaining certain features of this invention.

Resonance Vocoder Application

Referring first to FIG. 1, the formant locating apparatus of this invention is shown in a resonance vocoder speech transmission system. At the transmitter station of this system, an incoming speech wave from source 1, for example, a conventional microphone, is applied simultaneously to formant locator 2 of the present invention, and to pitch detector 3 and amplitude detector 4. Formant locator 2, whose structure is illustrated in outline form in FIG. 1 and in detail in FIGS. 2A and 2B, produces a group of narrow-band control signals, denoted F_1 , F_2 , and F_3 , representative of the locations of the formants of the incoming speech wave. Pitch detector 3, which may be of any well-known variety, generates a pitch control signal indicative of whether the speech wave represents either a voiced or an unvoiced sound at a given instant, and if the sound is voiced, the pitch control signal also indicates the fundamental frequency of the sound. Amplitude detector 4, which may also be of conventional design, for example, a rectifier followed by a low-pass filter, produces an amplitude control signal indicative of the energy of the speech wave.

The formant, pitch, and amplitude control signals produced at the transmitter station specify the information-bearing characteristics of the incoming speech wave with a high degree of accuracy, but the total bandwidth of these signals is substantially smaller than that of the incoming speech wave. Hence these control signals may be transmitted to a distant receiver station over a reduced bandwidth transmission channel, a suitable channel being indicated in FIG. 1 by broken lines.

At the receiver station, an excitation signal comprising, for example, a train of uniform amplitude pulses, is generated from the pitch control signal by an excitation signal generator 5 of any desired sort, and the excitation signal, together with the formant control signals and the amplitude control signal, is applied to speech synthesizer 6, for instance, a conventional resonance vocoder synthesizer. Synthesizer 6 reconstructs from the incoming signals a replica of the original speech wave, and reproducer 7, which may be a conventional loudspeaker, converts the speech wave replica into audible sound.

General Principles

Before proceeding to an explanation of the details of formant locator 2 illustrated in FIGS. 2A and 2B, it will be convenient at this point to describe the general features of formant locator 2 which are outlined in FIG. 1. The formant locating process embodied by formant locator 2 comprises a rapid succession of repetitive cycles initiated by a repetitive cycle control signal generated by control circuit 23 and delivered to formant generator 24. In response to this control signal, formant generator 24 repetitively produces in each cycle a set of formant signals comprising, for example, three formant signals representing the three principal formants found in most voiced speech sounds. Each of these formant signals varies continuously over a range of values corresponding to the usual frequency range of a particular formant, and the combination of values represented by the set of formant signals as a whole is made to vary in such a manner that during each formant locating cycle, the formant signals collectively represent all possible combinations of formant locations. The variation of the set of formant signals is explained in detail below in connection with the description of FIGS. 2A and 2B.

The set of continuously variable formant signals from generator 24 is applied to artificial spectrum synthesizer 25, together with the pitch control signal from detector 3 and the amplitude control signal from detector 4. Synthesizer 25 constructs from the incoming signals an artificial amplitude spectrum whose formants continuously vary in location in synchronization with the continuous

variations in value of the applied formant signals. From synthesizer 25, the artificial spectrum is passed to analyzer 27, which separates the artificial spectrum into its individual frequency components. Analyzer 21 similarly separates the incoming speech spectrum from source 1 into its individual frequency components, and the two sets of frequency components from analyzers 27 and 21 are sent to comparator 22.

From the two sets of frequency components from analyzers 21 and 27, comparator 22 derives a so-called error signal whose magnitude is representative of the difference between the artificial spectrum and the speech spectrum during each formant locating cycle. The difference between the two spectra which is represented by the magnitude of the error signal is the difference between the formant locations of the speech spectrum and the formant locations of the artificial spectrum, because in every other respect the two spectra are identical. Thus as symbolized by the lines connecting detectors 3 and 4 with synthesizer 25, the artificial spectrum is constructed to have the same fundamental frequency and the same energy as the speech spectrum, whereas the formant signals from generator 24 cause the formants of the artificial spectrum to occur at substantially all possible combinations of locations during each formant locating cycle.

Theoretically, then, there will be at least one point in each formant locating cycle when the formant locations of the artificial spectrum will be identical with the formant locations of the speech spectrum, and, correspondingly, the magnitude of the error signal at this point will be zero. However, because of noise and other imperfections, the artificial spectrum will never be exactly identical with the speech spectrum, but there is a point in each cycle at which the difference between the two spectra is at a minimum, this minimum difference being indicated by a corresponding minimum in the magnitude of the error signal.

To determine the formant locations of the artificial spectrum at the point in each cycle at which the difference between the artificial spectrum and the speech spectrum is a minimum, the error signal from comparator 22 is passed to control circuit 23, where the error signal is continuously examined to determine its minimum magnitude. At the instant in each cycle that circuit 23 detects a minimum magnitude, it delivers a sample control signal to sampler 28, and in response, sampler 28 samples and stores the values of the formant signals at that instant. When the cycle ends, these values are converted into formant control signals F_1 , F_2 , and F_3 , to indicate the locations of the formants in the spectrum of the incoming speech wave.

The difference between the formant locations of the artificial spectrum and the formant locations of the speech spectrum may be measured in several ways. For example, the magnitude of the error signal may represent the sum of squared differences between the amplitudes of corresponding frequency components of the two spectra, the sum of absolute differences between the amplitudes of corresponding frequency components, or the sum of squared differences between the logarithms of the amplitudes of corresponding frequency components. Apparatus for deriving error signals representative of each of these measures of difference is illustrated in the drawings and described below. Other measures of difference may be utilized, if desired; for example, see the article by K. N. Stevens entitled *Toward a Model for Speech Recognition*, vol. 32, *Journal of the Acoustical Society of America*, page 50 (1960).

Circuit Details

Referring now to FIGS. 2A and 2B, these drawings illustrate in detail the structure of the formant locating apparatus of this invention. In control circuit 23 of FIG. 2B, multivibrator 233, which may be a conventional free-running type, is constructed to oscillate at a fre-

quency equal to the rate at which it is desired to perform the formant locating process. For example, if it is desired to perform a single cycle of the formant locating process in one-twentieth of a second, then multivibrator 233 is constructed to oscillate at a frequency of 20 cycles per second. A few of the output pulses generated by multivibrator 233 are shown in somewhat idealized form in FIG. 4A, where for purposes of illustration a single period, τ , of the output pulses is shown to be approximately equal to one-twentieth of a second.

The output pulses of multivibrator 233 comprise the cycle control signal previously referred to, and are employed to initiate each cycle of the formant locating process by coupling them through capacitor C3 to monostable multivibrator 241 of formant generator 24. By this arrangement, the positive-going portions of these pulses successively trigger multivibrator 241 to its unstable state, causing a corresponding succession of positive-going pulses denoted b_1 to appear at the output terminal of multivibrator 241. The relationship of the output pulses of multivibrator 241 to the output pulses of multivibrator 233 is illustrated graphically by a comparison of the waveform denoted b_1 in FIG. 4B with the waveform in FIG. 4A.

The positive-going output pulses of multivibrator 241 are converted by ramp network 244a, which is illustrated in detail in FIG. 3, into a succession of ramp-shaped signals that constitute the first formant signal, h_1 , of this invention. As shown in FIG. 4B, each period of the h_1 signal increases continuously over a range of values corresponding to the range of frequencies within which the first speech formant is defined to occur. Thus, if the first speech formant is defined as that peak in the envelope of the speech amplitude spectrum which occurs in the frequency range from 200 to 1,200 cycles, then as shown in FIG. 4B, the range of values over which h_1 varies in a single period is made to correspond to this frequency range. However, as indicated by the dashed waveform denoted h_1' , other suitable frequency ranges may be defined for the first formant, if desired, with a corresponding shift in values for the first formant signal.

The ramp network illustrated in FIG. 3 comprises a resistor 31 in series with a capacitor 34, and a diode 33 connected in parallel with resistor 31. This arrangement converts pulses appearing at point P1 into a ramp-shaped voltage at point P2, the diode 33 serving to allow capacitor 34 to discharge rapidly at the end of each formant locating cycle. Energy source 32 connected between points P2 and P3 is chosen to make the ramp-shaped voltage, h_j , $j=1, 2, 3$, produced at point P3 start at a value corresponding to the lowest expected frequency at which the j th formant may occur. By proper selection of values for the elements shown in FIG. 3, the ramp network is adapted to convert the various output pulses of multivibrators 241, 242, and 243 of generator 24 into formant signals.

Formant signals representing the locations of the two other principal speech formants are generated by multivibrators 242 and 243 followed by ramp networks 244b and 244c, respectively. However, additional multivibrators and associated circuitry may be provided if it is desired to locate higher order formants in addition to the three principal ones. Unlike multivibrator 241, multivibrators 242 and 243 may be of the free-running variety, and as shown in FIGS. 4C and 4D, the output pulses of these multivibrators need not be synchronized with the formant locating cycle. FIGS. 4C and 4D also illustrate that the values of formant signals h_2 and h_3 derived from the positive-going output pulses of multivibrators 242 and 243 correspond to overlapping frequency ranges, and waveforms h_2' and h_3' indicate that other frequency ranges may be selected for the formant signals, if desired.

Because of the variable nature of speech formant locations, the formant locations of the artificial spectrum

constructed from the formant signals in synthesizer 25 must occur at all possible combinations of locations in order for the formant locating process to yield accurate results for a wide variety of sounds and a wide variety of talkers. All possible combinations of formant locations are obtained by generating the formant signals in the following fashion: As illustrated in a comparison of FIGS. 4B and 4C, the frequency of oscillation of multivibrator 242 is made sufficiently high to enable a single period of the formant signal h_2 produced at the output terminal of ramp network 244b to vary over its complete range of values while the first formant signal h_1 is changing by a relatively small amount. Similarly, as shown by a comparison of FIGS. 4C and 4D, the frequency of oscillation of multivibrator 243 is made sufficiently high with respect to that of multivibrator 242 to enable a single period of the formant signal h_3 produced at the output terminal of ramp network 244c to vary over its complete range of values while the second formant signal h_2 is changing by a relatively small amount. Suitable frequencies of oscillation for multivibrators 242 and 243 may be on the order of 200 cycles per second and 2,000 cycles per second, respectively. In this way, the set of formant signals generated in each formant locating cycle represents substantially all possible combinations of formant locations, and therefore the formants of the artificial spectrum constructed from these signals vary through substantially all possible combinations of locations during a single formant locating cycle.

From formant generator 24, formant signals h_1 , h_2 , and h_3 are passed to artificial spectrum synthesizer 25, where they are individually applied to series-connected resonator circuits 251, 252, and 253, respectively, which may be of similar construction to those described in E. S. Weibel Patent 2,817,707, issued December 24, 1957; however, it is to be understood that parallel-connected resonator circuits or other resonance vocoder synthesizer circuits are also suitable for constructing an artificial spectrum. Resonator circuits 251 through 253 are also supplied with the pitch control signal to distinguish between voiced and unvoiced portions of the incoming speech wave, as shown in the Weibel patent, since the resonator circuits must be altered to produce an artificial spectrum whose characteristics reflect both voiced and unvoiced portions of the incoming speech wave.

As previously explained, the artificial spectrum is made to resemble the speech spectrum so that the only variable feature of the artificial spectrum is its formant locations. One way in which this is accomplished is by using the pitch control signal derived from the incoming speech wave to generate an excitation signal for the construction of the artificial spectrum. In synthesizer 25, therefore, the pitch control signal is applied (through amplifier 255) to a suitable excitation signal generator 254, comprising for example, conventional buzz and hiss sources, which for voiced sounds produces a train of uniform amplitude pulses whose fundamental frequency is determined by the magnitude of the pitch control signal. However, because of the rapid variation in value of the formant signals, particularly the second and third, h_2 and h_3 , it is necessary to increase by a factor k the fundamental frequency, f_0 , represented by the pitch control signal. This may be achieved by passing the pitch signal through a conventional voltage amplifier 255 having a gain constant equal to k before applying the signal to generator 254.

The value of the gain constant k of amplifier 255 is determined by the following considerations. It is well known that the width, Δf , of a formant is substantially larger than the rate at which formant locations change; that is, Δf may be on the order of 100 cycles per second, while the rate of change of formant locations is on the order of 20 cycles per second. In the construction of artificial speech by resonator circuits of the type described in the above-mentioned Weibel patent, this relationship

is preserved by adjusting the time constant, RC, of each resonator circuit, according to the inverse relationship between Δf and RC given by Equation 2 of the Weibel patent.

In the present invention, the rate of change of formant locations in the artificial spectrum constructed by synthesizer 25 is substantially higher than 20 cycles per second, because formant signal h_3 , for example, makes about 2000 variations over its entire range of values in a second. Therefore, the rate of change of formant locations in the artificial spectrum produced by synthesizer 25 is in excess of 2000 cycles per second. In order to preserve the proper relationship between formant widths and rate of formant change, formant widths in the artificial spectrum must exceed 10,000 cycles per second and the time constants of the resonator circuits must be adjusted accordingly. In addition, this increase on the frequency scale of formant widths must be attended by a corresponding shift of formant locations on the frequency scale in order to maintain a natural relationship between formant frequencies and formant widths. For example, if formant widths of the artificial spectrum are larger than the formant widths of the speech spectrum by a factor of 100 or more, then formant frequencies must similarly be increased by a factor of 100 or more. This shift on the frequency scale is accomplished by constructing the artificial spectrum in synthesizer 25 from an excitation signal whose frequency components are correspondingly scaled up in frequency relative to the frequency components of the incoming speech wave. In the present invention this scaling up in frequency is achieved by passing the pitch control signal through amplifier 255, whose gain constant, k , is made equal to a value in the interval between 100 and 1000, as required, thereby increasing by a factor k both the fundamental frequency of the excitation signal and the formant frequencies of the artificial spectrum.

The second way in which the artificial spectrum is made to resemble the spectrum of the incoming speech wave is to make the total energy of the artificial spectrum equal to that of the speech spectrum. This is accomplished by adjusting the amplitude of the excitation signal from generator 254 in multiplier 256 under the control of an amplitude control signal, derived, for example, in the fashion shown in FIG. 1. From multiplier 256, the excitation signal is applied to the input terminal of resonator circuit 251 for synthesis of the artificial spectrum.

In the spectrum constructed by resonator circuits 251 through 253, there are three variable formant locations representing substantially all possible combinations of the three principal formant locations, but it is well known that additional, higher order formants do occur in voiced speech sounds. To compensate for this omission in the artificial spectrum appearing at the output terminal of resonator circuit 253, the artificial spectrum is passed through equalizer 257, which adjusts the amplitudes of the higher frequency components of the artificial spectrum to bring them into closer correspondence with similar components of the speech spectrum. From equalizer 257 the artificial spectrum is sent to analyzer 27 in FIG. 2A.

In analyzer 27 of FIG. 2A, the artificial spectrum is applied in parallel to band-pass filters 271a through 271n, which are constructed to separate the artificial spectrum into its individual frequency spectrum. At the same time, in analyzer 21, the spectrum of the incoming speech wave is similarly being separated into its individual frequency components by band-pass filters 211a through 211n. It is noted that where the pass bands of filters 211a through 211n in analyzer 21 are denoted Δf_1 through Δf_n , respectively, the pass bands of filters 271a through 271n in analyzer 27 are denoted $k\Delta f_1$ through $k\Delta f_n$, in order to take into account the shift on the frequency scale of the artificial spectrum by a factor k introduced in synthesizer 25.

Following the band-pass filters in analyzers 21 and 27, a set of n rectifiers followed by a set of n low-pass filters is provided, shown as elements 212 and 272, respectively, to derive in well-known fashion from the individual frequency components of each of the two spectra a group of direct-current signals representative of the amplitudes of the frequency components. The two groups of direct-current signals from analyzers 21 and 27 are then delivered to comparator 22, in which there is obtained an error signal representative of a selected measure of the difference between the formant locations of the artificial spectrum and the formant locations of the speech spectrum during each formant locating cycle. As illustrated in FIG. 2A, the error signal obtained by comparator 22 may be made to represent any one of several measures of the difference in formant locations between the artificial spectrum and the speech spectrum, depending upon the settings of switches 220a through 220n, 221a through 221n, and 228a through 228n. One of these measures is the sum of squared differences between the amplitudes of corresponding components, which is obtained by setting switches 220a through 220n to connect the output terminals of analyzer 27 to polarity inverters 222a through 222n, for example, minus one amplifiers, and by setting switches 221a through 221n to connect the output terminals of analyzer 21 to one of the input terminals of conventional adder circuits 225a through 225n. The output terminal of each polarity inverter is connected to the other input terminal of one of the adders and the signals developed at the output terminals of the adders represent differences in amplitude between corresponding frequency components of the two spectra. Each of these difference signals is applied in parallel to a squarer and a rectifier, denoted as elements 226a through 226n and 227a through 227n, respectively. At the output terminal of each of the squarers, which may be of well-known construction, there is produced a signal whose amplitude is proportional to the square of the amplitude of the input signal, and by setting switches 228a through 228n to connect all of the squarers to adder 229, there is formed at the output terminal of adder 229 an error signal representative of the sum of squared differences between the amplitudes of corresponding frequency components of the two spectra. The quantity represented by the error signal may be expressed by the following equation

$$E_s = \sum_i (A_i - B_i)^2 \quad (1)$$

where A_i represents the amplitude of the i th component of the speech spectrum, B_i represents the amplitude of the i th component of the artificial spectrum, and E_s denotes the sum of squared differences.

Another measure of the difference in formant locations between the artificial spectrum and the speech spectrum is the sum of absolute differences between amplitudes of corresponding frequency components, where if E_A denotes this sum, then

$$E_A = \sum_i |A_i - B_i| \quad (2)$$

An error signal representative of E_A is obtained at the output terminal of adder 229 by changing the switching arrangement in comparator 22 so that switches 228a through 228n connect rectifiers 227a through 227n to adder 229, with switches 220a through 220n and 221a through 221n remaining in the same position used for obtaining E_s .

Still another measure of the difference in formant locations is the sum of squared differences between logarithms of the amplitudes of corresponding frequency components of the two spectra. If this sum is denoted E_L , then

$$E_L = \sum_i (\log A_i - \log B_i)^2 \quad (3)$$

The error signal appearing at the output terminal of adder 229 may be made to represent E_L by setting switches 220a through 220n to connect the output terminals of analyzer 27 to the input terminals of conventional logarithmic at-

tenuators 223a through 223n, and by setting switches 221a through 221n to connect the output terminals of analyzer 21 to logarithmic attenuators 224a through 224n. The construction of logarithmic attenuators is well known; for example, each of the attenuators may comprise a passive network whose output signal has an amplitude that is a linear function of the logarithm of the input signal amplitude. By setting switches 228a through 228n to connect squarers 226a through 226n to adder 229, the output signal of adder 229 is an error signal representative of E_L .

To detect the minimum magnitude of the error signal in each formant locating cycle, the minimum magnitude being indicative of the best matching version of the artificial spectrum in each cycle, the output point of comparator 22 is connected to the input point of control circuit 23 in FIG. 2B, where the magnitude of the error signal is continuously examined during each cycle. Returning to FIG. 2B, each positive-going output pulse generated at the beginning of each cycle by multivibrator 233 is applied through capacitor C2 to actuate relay 232 which draws the armature S1 into momentary contact with energy source B1, thereby charging capacitor C1. Energy source B1 is constructed to charge capacitor C1 to a value greater than the largest expected magnitude of the error signal from comparator 22. The error signal obtained during each cycle by comparator 22 is delivered through diode 231 and the normally closed back contact of relay 232 to capacitor C1, so that whenever the magnitude of the error signal is less than the voltage on C1, current flows through diode 231 to make the voltage on C1 equal to the magnitude of the error signal. This current also produces a pulse of voltage in resistor R1, which is amplified by amplifier 234 and passed through capacitor C4 to operate relay 281 of sampler 28 momentarily. The momentary operation of relay 281 samples the values of formant signals h_1 , h_2 , and h_3 at this instant and transfers the sampled values to capacitors C6, C7, and C8, respectively. Since current flows through diode 231 only when the magnitude of the error signal falls below the voltage on C1, each flow of current during a formant locating cycle indicates a smaller magnitude of the error signal than that which caused the previous current flow. Correspondingly, each time that the armatures of relay 281 connect the output terminals of ramp networks 244a through 244c to capacitors C6 through C8, respectively, the charges on capacitors C6 through C8 are changed to correspond to the values of the formant signals at that instant. At the end of each formant locating cycle, diode 231 will have last conducted at the minimum magnitude of the error signal for that cycle, and the charges on capacitors C6 through C8 will represent the values of the formant signals which produced this minimum.

The charges on capacitors C6 through C8 are transferred via amplifiers 283a through 283c to capacitors C9 through C11, respectively, by the operation of relay 282 after the end of each formant locating cycle. The operation of relay 282 is controlled by the output pulses of multivibrator 241 acting through capacitor C5, these output pulses occurring after the end of one cycle and at the beginning of the next cycle as shown in FIG. 4B. The charges transferred to capacitors C9 through C11 are available at the output terminals of amplifiers 284a through 284c, respectively, to serve as formant control signals, for example, in the fashion described above in connection with FIG. 1, while capacitors C6 through C8 are employed in locating the next set of formant control signals.

Although this invention has been described in terms of speech communications systems such as the resonance vocoder of FIG. 1, it is to be understood that applications of this invention are not limited to such systems, but include the fields of automatic speech recognition systems, speech processing systems, and other signalling systems.

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 system for automatically locating speech formants which comprises a source of a first control signal representative of the fundamental frequency of an incoming speech wave, a source of a second control signal representative of the energy of said speech wave, means for generating a plurality of formant signals whose values collectively vary to represent substantially all possible combinations of formant locations during each of a succession of selected time periods, means for synthesizing an artificial spectrum from said formant signals and said first and second control signals, first analyzing means for deriving from said artificial spectrum a first group of signals representative of the amplitudes of the individual frequency components of said artificial spectrum, second analyzing means for deriving from the spectrum of said speech wave a second group of signals representative of the amplitudes of the individual frequency components of said speech spectrum, comparing means in circuit relation with said first and second analyzing means for obtaining from said first and second groups of amplitude signals an error signal representative of a selected measure of the difference in formant location between said artificial spectrum and said speech spectrum, means supplied with said error signal for sampling the values of said formant signals which correspond to the minimum magnitude of said error signal during each of said selected time periods.

2. Apparatus as defined in claim 1 wherein said comparing means comprises means for obtaining from said first and second groups of amplitude signals an error signal representative of the sum of squared differences between the amplitudes of corresponding frequency components of the artificial spectrum and the speech spectrum.

3. Apparatus as defined in claim 1 wherein said comparing means comprises means for obtaining from said first and second groups of amplitude signals an error signal representative of the sum of absolute differences between the amplitudes of corresponding frequency components of the artificial spectrum and the speech spectrum.

4. Apparatus as defined in claim 1 wherein said comparing means comprises means for obtaining an error signal representative of the sum of squared differences between logarithms of the amplitudes of corresponding frequency components of the artificial spectrum and the speech spectrum.

5. Apparatus for automatically locating peaks in the envelope of the spectrum of a speech wave which comprises a source of an excitation signal whose fundamental frequency is proportional to the fundamental frequency of an incoming speech wave and whose amplitude is proportional to the energy of said speech wave, means for repetitively synthesizing from said excitation signal an artificial spectrum having a plurality of variable formants that occur at all possible combinations of locations during each of a succession of selected intervals of time, a source of said incoming speech wave connected to a comparing means supplied with said artificial spectrum, wherein said comparing means obtains during each of said selected intervals of time an error signal that measures the difference in formant locations between said artificial spectrum and the spectrum of said speech wave, and means responsive to said error signal for selecting the formant locations of said artificial spectrum which correspond to the minimum magnitude of said error signal to represent the formants of said speech spectrum in each of said selected intervals of time.

6. A system for automatically locating formants which comprises a source of an incoming speech spectrum, means for constructing an artificial spectrum whose energy is

proportional to the energy of said speech spectrum, whose frequency components occur at harmonics of the fundamental frequency of said speech spectrum, and whose formants occur at substantially all possible combinations of locations during each of a succession of predetermined intervals of time, comparing means supplied with said speech spectrum and said artificial spectrum for obtaining an error signal whose magnitude is indicative of the difference in formant locations between said speech spectrum and said artificial spectrum during each of said time intervals, and means responsive to said error signal for identifying the formants of said artificial spectrum which correspond to the minimum magnitude of said difference signal in each of said predetermined intervals of time.

7. A resonance vocoder system that comprises a source of an incoming speech wave connected in parallel to an automatic formant locator, a pitch detector, and an amplitude detector, wherein said pitch detector derives from said speech wave a pitch control signal indicative of the fundamental frequency of voiced portions of said speech wave, said amplitude detector derives from said speech wave an amplitude control signal indicative of the amplitude of said speech wave, and said formant locator, which is supplied with said pitch control signal and said voiced amplitude control signal, includes control means for generating a cycle control signal comprising a train of pulses, each of said pulses indicating the beginning of a formant locating cycle, means responsive to said cycle control signal for producing a plurality of continuously variable formant signals whose values during each formant locating cycle collectively represent all possible combinations of speech formant locations, synthesizing means for constructing an artificial spectrum from said formant signals, said pitch control signal, and said voiced amplitude control signal, first analyzing means for deriving from said artificial spectrum a first group of signals representative of the amplitudes of individual frequency components of said artificial spectrum, a second analyzing means for deriving from the spectrum of said speech wave a second group of signals representative of the amplitudes of individual frequency components of said speech spectrum, comparing means connected to said first and second analyzing means for deriving an error signal representative of a selected measure of the difference in formant locations between said artificial spectrum and said speech spectrum, means under the influence of said error signal for sampling the values of said formant signals when the magnitude of said error signal passes through its smallest value in each of said cycles, and means for converting the sampled values of said formant signals in each cycle into a succession of narrow-band formant control signals, means for transmitting said pitch control signal, said amplitude control signal, and said formant control signals to a receiver station, and at said receiver station, means for synthesizing a replica of said incoming speech wave from said transmitted control signals.

8. Apparatus for constructing an artificial amplitude spectrum whose formants occur at substantially all possible combinations of locations within each of a succession of predetermined intervals of time which comprises means for generating a train of first formant signals in one-to-one correspondence with a succession of predetermined intervals of time, each of said first formant signals varying over a range of values corresponding to all possible frequency locations of the first speech formant, means for generating a train of second formant signals, each of said second formant signals varying over a range of values corresponding to all possible frequency locations of the second speech formant at a rate sufficiently fast for each second formant signal to vary over its complete range of values before any of said first formant signals has changed appreciably in value, means for generating a train of third formant signals, each of said third formant signals varying over a range of values corresponding to all possible frequency locations of the

third speech formant at a rate sufficiently fast for each third formant signal to vary over its complete range of values before any of said second formant signals has changed appreciably in value, a source of a pitch control signal whose magnitude is indicative of the fundamental frequency f_0 of voiced portions of an incoming speech wave, means for increasing the magnitude of said pitch control signal by a factor k to preserve the natural relationship between formant frequencies and formant widths, means for deriving from said increased magnitude pitch control signal an excitation signal comprising a train of uniform amplitude pulses whose fundamental frequency is kf_0 , a source of an amplitude control signal representative of the energy of said speech wave, means responsive to said amplitude control signal for adjusting the amplitudes of said excitation signal pulses, and synthesizing means supplied with said first, second, and third formant signals, said pitch control signal, and said amplitude adjusted excitation signal for obtaining an artificial spectrum.

9. Apparatus for repetitively generating during each of a succession of uniform time intervals of predetermined length a set of formant signals representative of substantially all possible combinations of locations of the three principal formants of voiced speech sounds which comprises a source of a repetitive control signal having a uniform, predetermined period, means responsive to said control signal for generating a first series of pulses having the same period as said control signal, first converting means for deriving from said first series of pulses a corresponding train of first formant signals each of which varies over a range of values proportional to the frequency range of the first formant of voiced speech sounds, a source of a second series of pulses whose period is short relative to the period of said first series of pulses, second converting means for deriving from said second series of pulses a corresponding train of second formant signals each of which varies over a range of values proportional to the frequency range of the second formant of voiced speech sounds at a uniform rate such that a second formant signal completes its variation in value before a time coincident first formant signal has changed appreciably in value, a source of a third series of pulses whose period is short relative to the period of said second series of pulses, and third converting means for deriving from said third series of pulses a corresponding train of third formant signals each of which varies over a range of values proportional to the frequency range of the third formant of voiced speech sounds at a uniform rate such that a third formant signal completes its variation in value before a time coincident second formant signal has changed appreciably in value.

10. Apparatus as defined in claim 9 where said source of a repetitive control signal comprises a free running multivibrator whose period of operation is on the order of one-twentieth ($\frac{1}{20}$) of a second.

11. Apparatus as defined in claim 9 wherein said means for generating a first series of pulses comprises a monostable multivibrator.

12. Apparatus as defined in claim 9 wherein said source of a second series of pulses comprises a free running multivibrator whose period of operation is on the order of one two hundredth ($\frac{1}{200}$) of a second.

13. Apparatus as defined in claim 9 wherein said source of a third series of pulses comprises a free running multivibrator whose period of operation is on the order of one two thousandth ($\frac{1}{2000}$) of a second.

14. Apparatus as defined in claim 9 wherein each of said first, second, and third converting means comprises an input terminal and an output terminal, a resistor and a capacitor connected in series between said input terminal and ground, a diode connected in parallel with said resistor, and an energy source connected between said resistor and said output terminal.

15. Apparatus for constructing an artificial amplitude spectrum which comprises an amplifier provided with an input terminal and an output terminal and having a gain constant k , where the value of k lies between 100 and 1000, means for applying to the input terminal of said amplifier a pitch control signal whose magnitude is proportional to the fundamental frequency, f_0 , of an incoming speech wave, an excitation signal generator provided with an input terminal and an output terminal for producing a train of uniform amplitude pulses whose fundamental frequency is determined by the magnitude of the signal applied to the input terminal of said generator, means for connecting the output terminal of said amplifier to the input terminal of said generator, a multiplier provided with a control terminal, an input terminal, and an output terminal for adjusting the amplitude of an input signal applied to its input terminal in response to a control signal applied to its control terminal, means for applying to the control terminal of said multiplier an amplitude control signal representative of the energy of said speech wave, means for connecting the output terminal of said generator to the input terminal of said multiplier, three series-connected resonator circuits, each of which is provided with a first control terminal, a second control terminal, an input terminal and an output terminal, means for connecting the output terminal of said multiplier to the input terminal of the first of said resonator circuits, means for applying to the first control terminal of each of said three resonator circuits a first, second,

and third formant signal, respectively, wherein the first formant signal varies over a range of values corresponding to the frequency locations of the first formant of speech sounds, the second formant signal varies over a range of values corresponding to the frequency locations of the second formant of speech sounds at about ten times the rate of variation of the first formant signal, and the third formant signal varies over a range of values corresponding to the frequency locations of the third formant of speech sounds at about ten times the rate of variation of the second formant signal, means for applying said pitch control signal to the second control terminal of each of said resonator circuits, an equalizer circuit provided with an input terminal and an output terminal for adjusting the amplitudes of the higher frequency components of a signal applied to its input terminal, and means for connecting the output terminal of the third resonator circuit to the input terminal of said equalizer circuit, whereby the signal appearing at the output terminal of said equalizer circuit has an amplitude spectrum whose frequency components occur at harmonics of k times the fundamental frequency of said speech wave, whose energy is equal to that of said speech wave, and whose formants occur at substantially all possible combinations of locations during the time that it takes said first formant signal to vary over its complete range of values.

No references cited.