

June 20, 1967

C. H. COKER

3,327,058

SPEECH WAVE ANALYZER

Filed Nov. 8, 1963

3 Sheets-Sheet 1

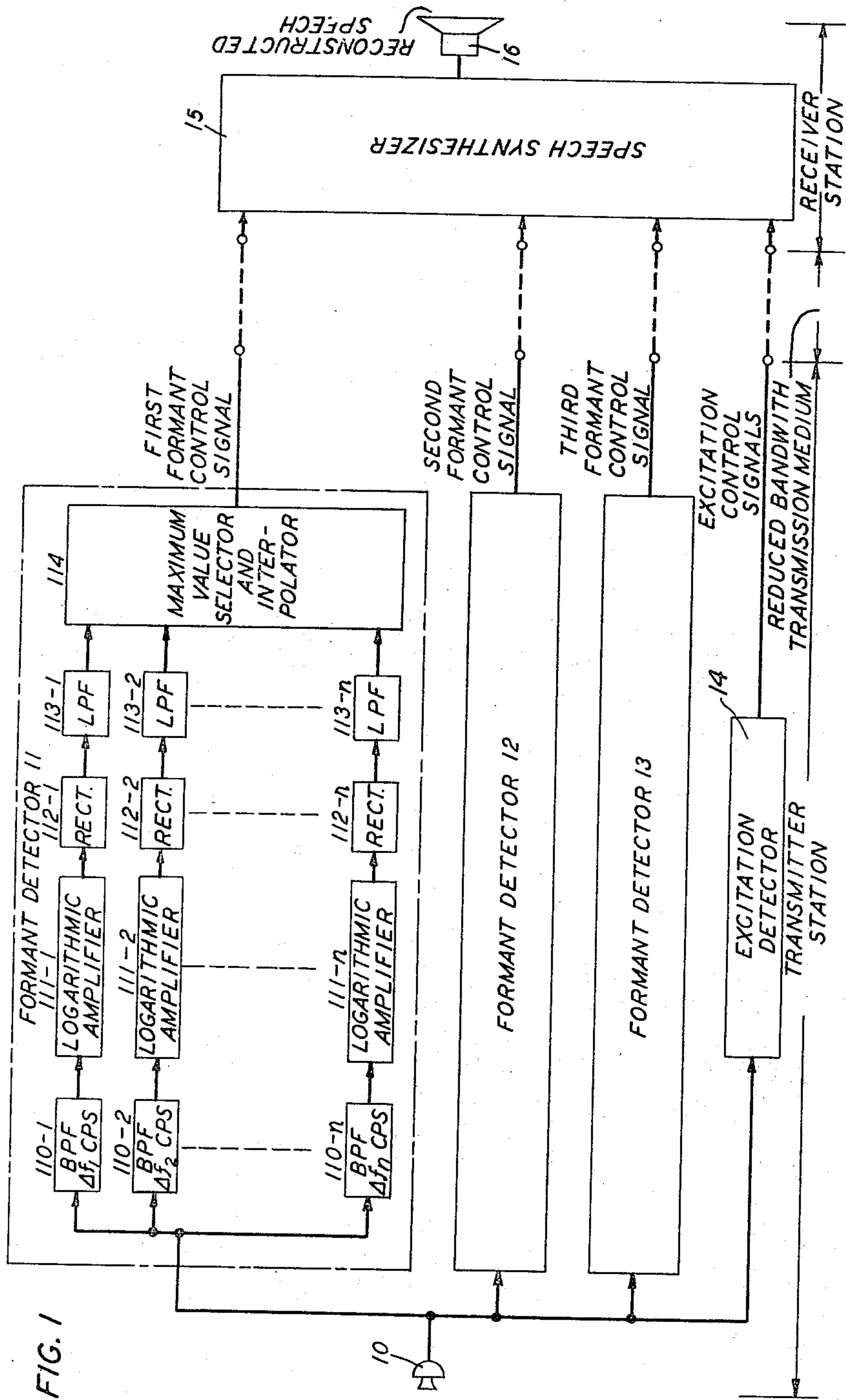


FIG. 1

INVENTOR
C. H. COKER
BY *John P. Kirsch*
ATTORNEY

June 20, 1967

C. H. COKER

3,327,058

SPEECH WAVE ANALYZER

Filed Nov. 8, 1963

3 Sheets-Sheet 2

FIG. 2

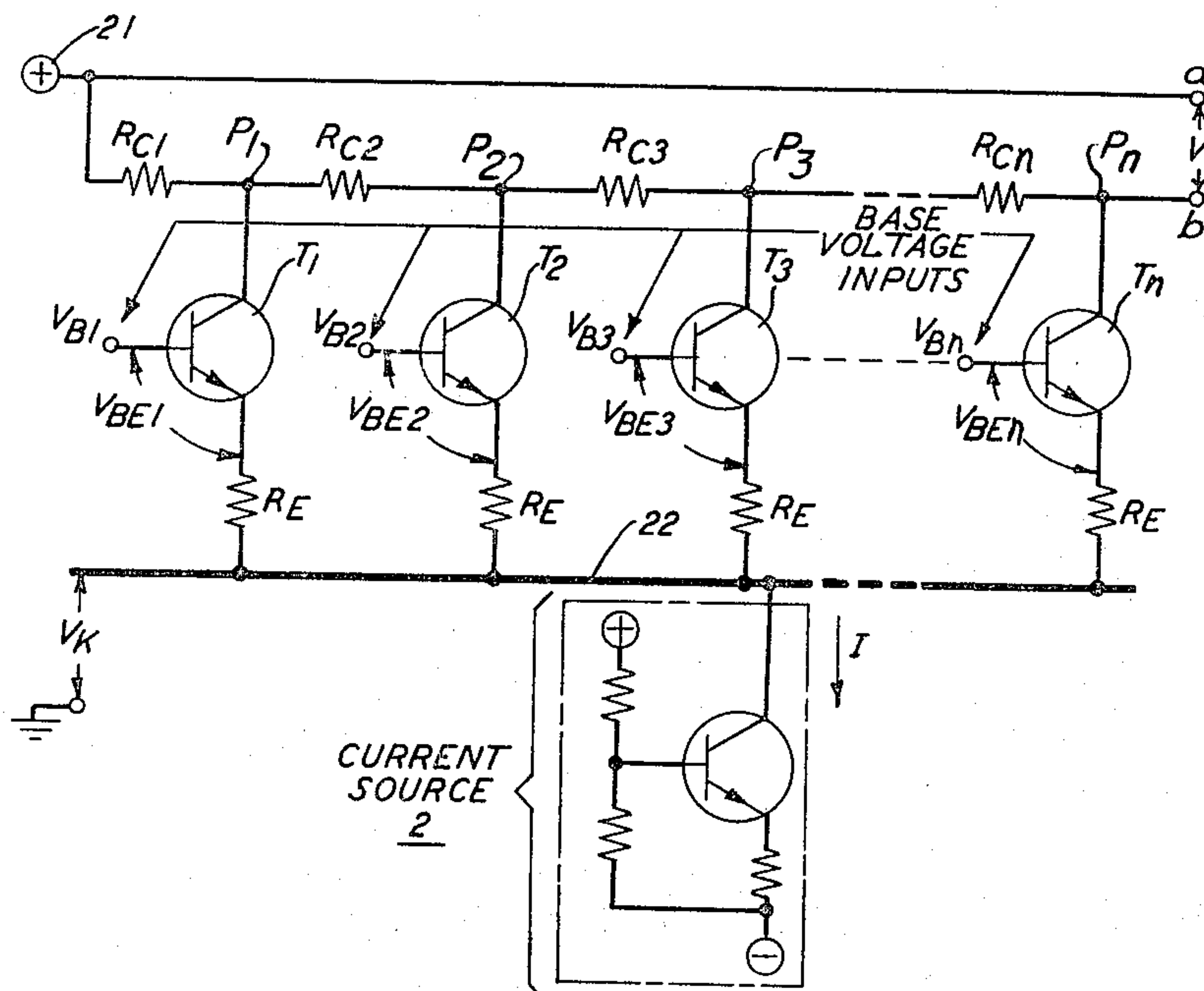


FIG. 3A

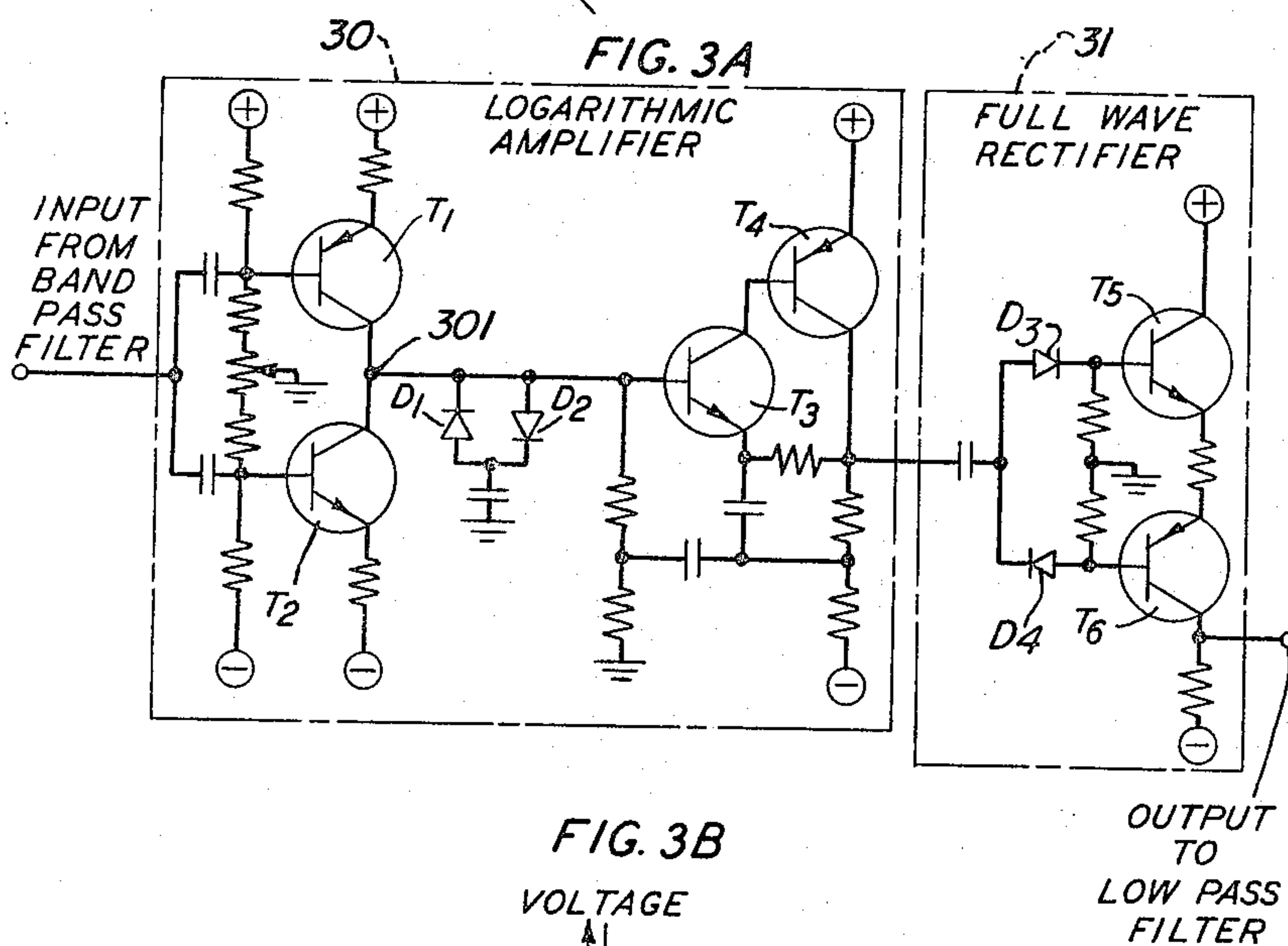
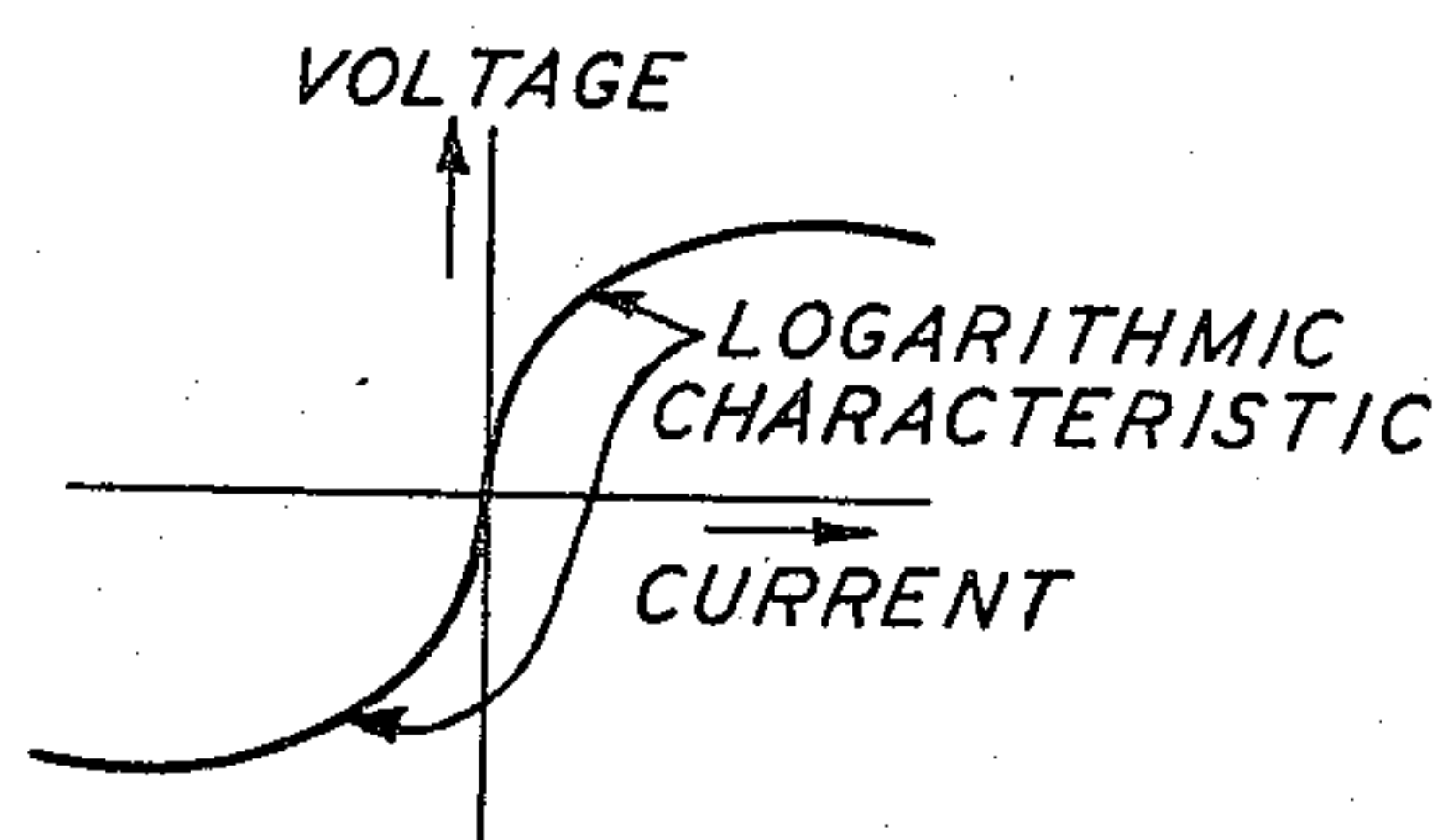


FIG. 3B



June 20, 1967

C. H. COKER

3,327,058

SPEECH WAVE ANALYZER

Filed Nov. 8, 1963

3 Sheets-Sheet 3

FIG. 4

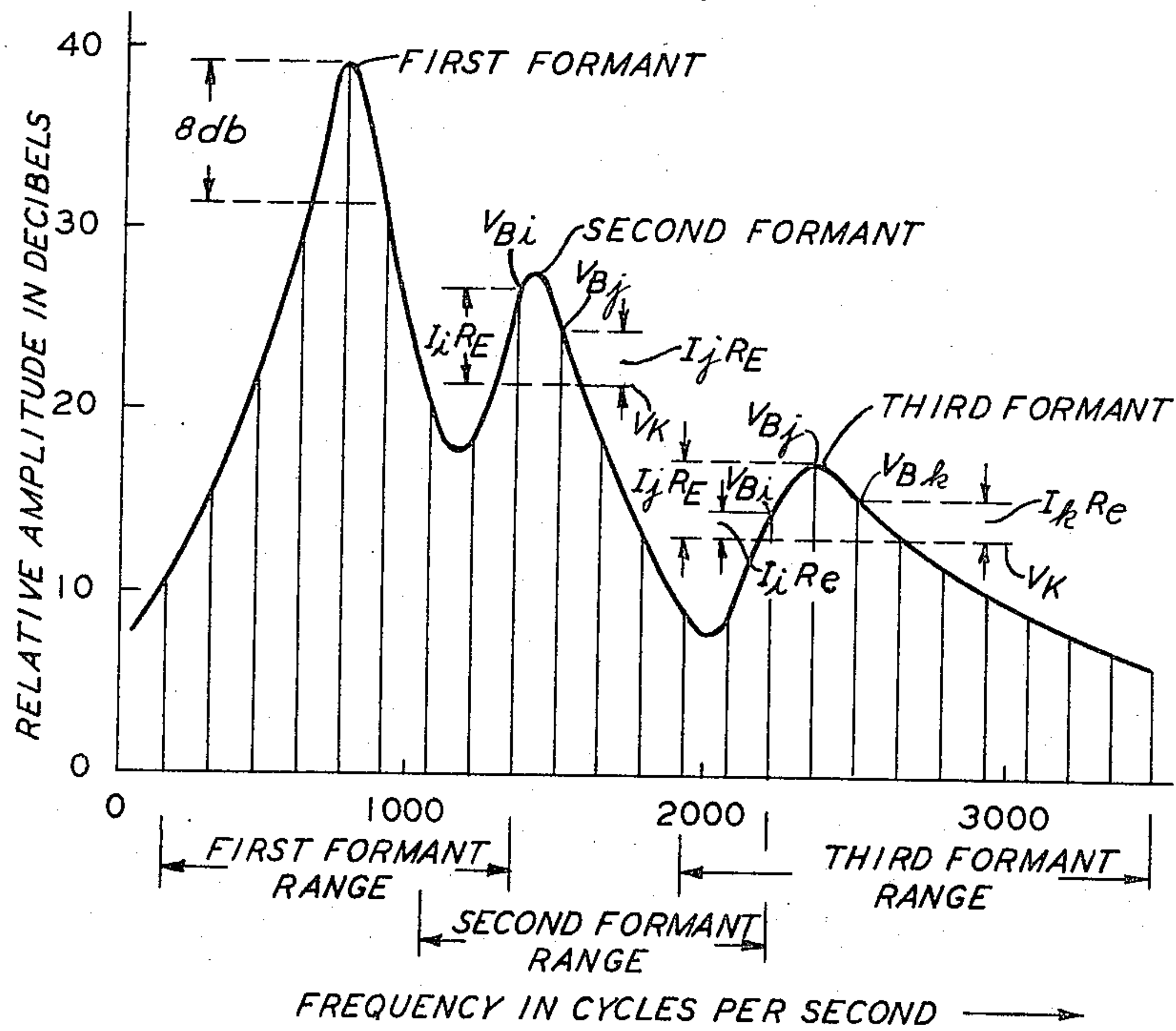
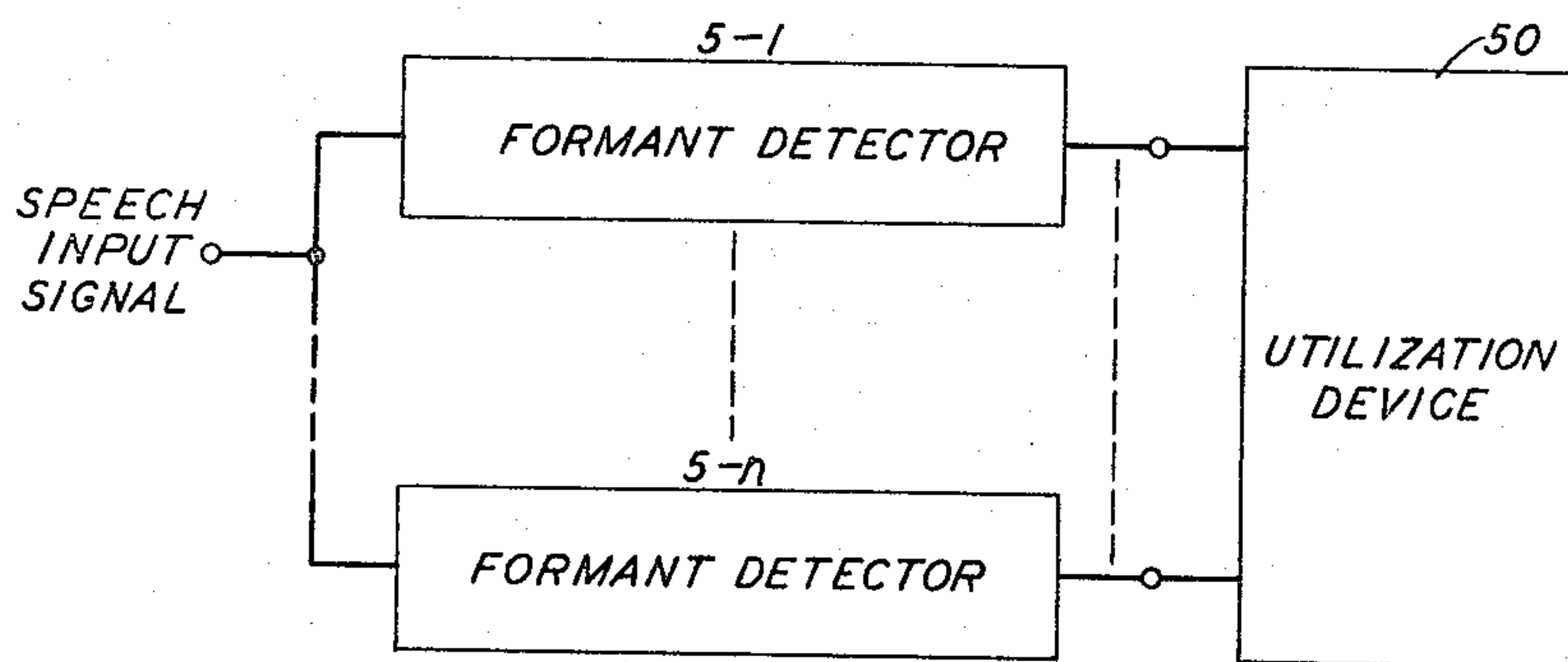


FIG. 5



1

3,327,058

SPEECH WAVE ANALYZER

Cecil H. Coker, Berkeley Heights, N.J., assignor to Bell Telephone Laboratories, Incorporated, New York, N.Y., a corporation of New York

Filed Nov. 8, 1963, Ser. No. 322,389

5 Claims. (Cl. 179-1)

This invention relates to the analysis and transmission of speech, and in particular to the analysis and transmission of speech in bandwidth compression systems.

In order to make more economic use of the frequency bandwidth of speech transmission channels, a number of bandwidth compression arrangements have been devised for transmitting the information content of a speech wave over a channel whose bandwidth is substantially narrower than that required for transmission of the speech wave itself. Bandwidth compression systems typically include at a transmitting terminal an analyzer for deriving from an incoming speech wave a group of narrow bandwidth control signals representative of selected information bearing characteristics of the speech wave, and at a receiving terminal a synthesizer for reconstructing from the control signals a replica of the original speech wave.

One well-known bandwidth compression system is the so-called "resonance vocoder," a specific form of which is described in J. L. Flanagan Patent 2,891,111, issued June 16, 1959. In a resonance vocoder, the distinctive information bearing characteristics which are represented by the control signals and which are reconstructed at the receiving terminal are the frequency locations of selected peaks or maxima in the speech amplitude spectrum. These spectral maxima correspond to the principal vocal tract resonances or formants; that is, they correspond to frequency regions of relatively effective acoustic transmission through a talker's vocal tract.

In order to reconstruct natural sounding, intelligible speech in a resonance vocoder, it is necessary to determine with accuracy the frequencies at which formants or maxima occur in the speech spectrum, since human listeners are able to detect variations in formant frequencies which are on the order of 50 cycles per second. One system for determining formant locations is described in the above-mentioned J. L. Flanagan patent, in which the spectrum of an incoming speech signal is divided into three frequency subbands, each of which embraces the characteristic frequency range of a particular formant. Each frequency subband is separated into its individual frequency components, and from these components the Flanagan system selects the frequency of the component having the maximum amplitude within a subband to represent the location of a formant occurring within that subband.

It is evident that in the Flanagan system the determination of formant locations is predicated upon the assumption that within each selected spectral subband the frequency of a single speech component having an amplitude larger than the amplitude of any other component accurately represents a formant location. In general, however, formants do not coincide exactly with speech frequency components, so that the frequency of a component having a maximum amplitude within a frequency subband is generally an approximation to the true location of a formant.

The present invention provides a system for accurately determining formant locations by recognizing that in general a formant is most likely to occur at an intermediate frequency lying between several components in the vicinity of a maximum in the speech spectrum. Because the components that occur in the vicinity of a spectral maximum usually have larger amplitudes than any other components

2

within a particular frequency subband, the present invention interpolates a formant location among the frequencies of the components having the largest amplitudes within a subband. Further, since a formant does occasionally coincide with a single speech component having an amplitude substantially larger than the amplitude of any other component within the subband, the present invention also provides for this situation by selecting the frequency of such a component to represent the formant location.

The present invention determines whether a formant location is to be represented by the frequency of a single component or whether a formant location is to be interpolated among the frequencies of several speech components according to their relative amplitudes by first converting the absolute amplitudes of the speech components within a particular frequency subband into logarithmic values according to a predetermined scale. Then, from the logarithmic values thus obtained there is derived a variable threshold against which the logarithmic values of the speech component amplitudes are compared. If no more than a single logarithmic value exceeds the threshold, then a single predetermined frequency corresponding to the speech component having the largest amplitude is selected to represent the formant location. On the other hand, if more than one logarithmic value exceeds the threshold, then the formant location is represented by a frequency interpolated among the predetermined frequencies corresponding to the speech components having amplitude whose logarithms exceed the threshold, the exact value of the interpolated frequency depending upon the relative magnitudes of the differences between the variable threshold and the logarithmic values of the speech components that exceed the threshold. The variable threshold against which the logarithmic values are compared is derived from the logarithmic values by setting the threshold at a level such that the sum of the differences between the threshold and the logarithmic values that exceed the threshold is equal to a predetermined constant.

An important feature of the present invention is its ability to interpolate formant locations among speech component frequencies based on relative, rather than absolute speech component amplitudes, thereby insuring accurate determination of formant locations despite wide variations in absolute speech component amplitudes due to corresponding wide variations in overall speech energy. This feature of the present invention is attained through the conversion of absolute amplitudes into logarithmic values, and the interpolation of formant locations according to the relative magnitudes of logarithmic differences, for logarithmic differences correspond to ratios of absolute amplitudes, hence the present invention interpolates formant locations despite variations in overall speech energy.

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

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

FIG. 2 is a schematic diagram illustrating in detail the maximum value selector and interpolator of this invention;

FIG. 3A is a schematic diagram illustrating in detail certain components of the systems shown in FIG. 1;

FIG. 3B is a graph of assistance in explaining the operation of a portion of the circuit shown in FIG. 3A;

FIG. 4 is a graph of assistance in explaining certain principles of the present invention; and

FIG. 5 is a block schematic diagram illustrating that the principles of this invention may be utilized in systems other than bandwidth compression systems.

Referring first to FIG. 1, this drawing illustrates a complete speech communication system embodying the principles of the present invention. At the transmitter station, an incoming speech wave from source 10, which may be a conventional transducer for converting speech sounds into a corresponding electrical wave, is delivered simultaneously to formant detectors 11, 12, and 13 and to excitation detector 14. Formant detectors 11, 12, and 13 are similar in construction, but as described in detail below, each of these detectors is designed to obtain a relatively narrow band control signal indicative of the frequency location of a particular formant of the sound represented by the incoming speech wave. As shown in FIG. 1, detectors 11, 12, and 13 may be designed to obtain narrow band control signals respectively indicative of the first, second and third formants. Examples of peaks in the spectrum of a typical voiced speech sound which correspond to these formants are illustrated in FIG. 4, in which the vertical lines represent the frequency components of the speech sound.

Excitation detector 14, which may be of any conventional design, derives from the incoming speech wave a group of narrow band excitation control signals indicative of the characteristics of the excitation that produced this speech wave; for example, detector 14 may be a conventional pitch detector that derives from the speech wave a pitch control signal representing the fundamental frequency of voice portions of the speech wave and a voiced-unvoiced control signal indicating whether the speech wave at a given instant represents a voiced speech sound or an unvoiced speech sound. Alternatively, if desired, excitation detector 14 may comprise the pitch detector and voice amplitude detector, the unvoiced amplitude detector, and the unvoiced control signal generator shown and described in the copending application of E. E. David, Jr., and J. L. Flanagan, Ser. No. 235,703, filed Nov. 6, 1962, now issued as Patent 3,190,963 on June 22, 1965.

The narrow band control signals derived by detectors 11, 12, 13, and 14 represented selected characteristics of the original speech wave which convey sufficient speech information for the reconstruction of an intelligible replica of the original speech wave, but the collective bandwidth of these control signals is substantially smaller than the bandwidth of the original speech wave. Hence these control signals may be transmitted to a distant receiver station over a reduced bandwidth transmission medium, the medium being indicated in FIG. 1 by broken lines.

At the receiver station the transmitted narrow band control signals are utilized in a speech synthesizer 15 to reconstruct an artificial speech wave that is a highly intelligible, natural sounding replica of the original speech wave. Speech synthesizer 15 may be any conventional formant vocoder synthesizer; for example, synthesizer 15 may be of the variety shown in the above-mentioned David, Jr.-Flanagan Patent 3,190,963. Synthesizer 15 produces a reconstructed speech wave having a spectrum whose peaks are adjusted to follow closely the locations of the formants of the original speech wave as indicated by the transmitted formant control signals. Speech sounds may be obtained from the original wave by applying the reconstructed wave to a suitable transducer, for example, a conventional loudspeaker 16.

Referring back to formant detectors 11, 12, and 13 at the transmitter station, only detector 11 is shown and described in detail, since the detectors are structurally similar, except for certain design characteristics which depend upon the particular formant to be located, as described below. Within detector 11 the incoming speech wave is applied in parallel to a bank of n contiguous or overlapping bandpass filters 110-1 through 110- n , where the pass bands of the filters are respectively denoted Δf_1 through Δf_n in cycles per second. The pass bands of filters 110-1 through 110- n collectively span a frequency subband of the speech spectrum which embraces the fre-

quency range within which the first speech formant normally occurs, and similarly the pass bands of the corresponding banks of bandpass filters in detectors 12 and 13 collectively span frequency subbands of the speech spectrum which embrace the respective frequency ranges within which the second and third formants normally occur. Since the frequency ranges of the various formants overlap to some extent, it is evident that one or more of the bandpass filters in each of the detectors 11, 12, and 13 will have pass bands embracing the same frequencies, as indicated by the division of the frequency scale in FIG. 4. Further, although FIG. 1 illustrates only three formant detectors for determining the first three formants, which are the most important formants for speech intelligibility, it is to be understood that additional formant detectors may be provided if it is desired to determine the frequency locations of higher order formants.

Depending upon the frequency content of the incoming speech wave, there will be developed at the output terminals of a number of the filters 110-1 through 110- n alternating waves whose frequencies and amplitudes represent the frequencies and amplitudes of corresponding frequency components of the incoming speech wave within the frequency subband covered by the pass bands of filters 110-1 through 110- n . The output terminal of each filter 110-1 through 110- n is connected to the input terminal of a corresponding logarithmic amplifier 111-1 through 111- n , each of which is followed by a corresponding rectifier 112-1 through 112- n , and a corresponding low pass filter 113-1 through 113- n , all connected in series. A detailed schematic of a suitable logarithmic amplifier and rectifier is shown in FIG. 3A and described below, but conventional logarithmic amplifiers and rectifiers may be employed if desired, while low pass filters 113-1 through 113- n may be of any well-known design. The logarithmic amplifier, rectifier, and low pass filter following each bandpass filter serve to derive from the alternating output signal of a preceding bandpass filter a unidirectional or narrow band signal having a magnitude that represents the logarithm of the instantaneous amplitude of the corresponding speech wave component indicated by a preceding sinusoidal output signal.

In this invention the voltages of the unidirectional output signals of low pass filters 113-1 through 113- n , which represent the amplitudes of the corresponding speech frequency components are described in terms of decibels but other logarithmic scales are equally suitable. A suitable scale factor for these unidirectional signals is 0.1 volt per decibel, where, as shown in FIG. 4, the decibel represents amplitude on a logarithmic scale according to the usual definition. By making the unidirectional signals represent the amplitudes of speech frequency components on a logarithmic or decibel scale, the maximum value selector and interpolator 114 in formant detector 11 is able to determine formant locations with a high degree of accuracy, as will be explained in the following detailed description of maximum value selector 114.

Turning now to FIG. 2, this drawing illustrates in detail the structure of the maximum value selector and interpolator employed in each of the formant detectors 11, 12, and 13. The output terminal of each of the n low pass filters preceding the maximum value selector and interpolator is connected to the base of a corresponding transistor T_1 through T_n , where the emitters of transistors T_1 through T_n are each connected through an identical resistor R_E to a common emitter bus 22. Transistors T_1 through T_n may be NPN transistors of conventional design. The collectors of transistors T_1 through T_n are each connected through a corresponding tap P_1 through P_n of a plurality of series connected resistors R_{c1} through R_{cn} to a source of positive potential 21.

A constant current I from conventional current source 2 is delivered to the common emitter bus 22, and depending upon the relative magnitudes of the incoming unidirectional signals which produce base voltages V_{B1}

5

through V_{Bn} at the bases of the corresponding transistors, the current I is directed to flow through one or more of the transistors and the associated resistors R_{c1} through R_{cn} to vary the magnitude of the output voltage V between points a and b . The output voltage V may be calibrated to assume a predetermined value V_j corresponding to a formant centered on the pass band Δf_j of bandpass filter 110- j by choosing R_{ci} such that

$$\left[\sum_{i=1}^j R_{ci} \right] I = V_j \quad (1)$$

where it is to be understood that calibration of V is made in the order proceeding from R_{c1} through R_{cn} . In this manner, the output voltage may be calibrated to assume n predetermined different values V_1 through V_n corresponding to the n possibilities that a formant location coincides with a particular speech component whose frequency corresponds to the center frequency of the pass band of a particular bandpass filter. However, as explained in detail below, the output voltage produced by the maximum value selector and interpolator also assumes intermediate values between adjacent voltages V_j and V_{j+1} to represent formant locations that occur at intermediate points between filter center frequencies.

The maximum value selector and interpolator circuit shown in FIG. 2 operates in the following manner. When one of the unidirectional voltages applied to the base of its corresponding transistor is substantially more positive than any of the other unidirectional voltages, all of the current I passes through the corresponding transistor and the resistors R_c that lie between the collector of the conducting transistor and the positive power supply 21 to produce an output voltage V whose magnitude has been calibrated to indicate a predetermined frequency location in accordance with Equation 1. This mode of operation may be understood from the following relationship between the common emitter bus voltage, denoted V_k , and the voltage across the i th transistor T_i , in the event that T_i is conducting the current I_i ,

$$V_k = V_{B1} - I_i R_E \quad (2)$$

where I_i is no greater than the total current I . It is seen from FIG. 2 that so long as V_k has a positive impedance to ground, and if all of the other base voltages remain constant but the base voltage V_{B1} increases, then the current I_i must increase, thereby increasing the common emitter bus voltage V_k . Hence as V_{B1} increases, eventually a point is reached at which V_k will be more positive than the base voltages of the transistors other than T_i , thereby making the other transistors nonconductive and allowing all of the current I to flow through T_i , as stated above.

A graphic illustration of the operation of the maximum value selector and interpolator circuit in the situation where all of the current I flows through a single transistor is shown by the first formant in FIG. 4. It is observed in FIG. 4 that the first formant coincides with the fifth harmonic of the speech wave and it is further observed that the fifth harmonic is about 8 decibels larger than the next largest harmonic, the sixth harmonic. In this situation there will be obtained a unidirectional voltage at the output terminal of one of the low pass filters having an amplitude which is correspondingly larger than the next largest unidirectional voltage, the magnitude of the difference between the largest and next largest unidirectional voltage being determined by the scale factor that relates voltage to decibels. For example, with a scale factor of $\frac{1}{10}$ of a volt per decibel, the unidirectional voltage representing the fifth harmonic will be $\frac{8}{10}$ of a volt larger than the unidirectional voltage representing the sixth harmonic. Assuming that the fifth harmonic coincides with the center frequency of the pass band of the fifth bandpass filter, 110-5, there will appear at the base of corresponding transistor T_5 a base voltage that is at least $\frac{8}{10}$ of a volt larger than the base voltage appearing at the base of any other transistor. By choosing the resistance of each of the

6

emitter resistors R_E and the value of the current I so that the product $I \cdot R_E$ is less than $\frac{8}{10}$ of a volt, for example, by choosing resistance and current values so that $I \cdot R_E = 0.6$ volt, it will be seen from Equation 2 that the common emitter bus voltage V_k is more positive than any of the other base voltages, thereby making all of the other transistors nonconductive. As a result, the output voltage V will be equal to the value calibrated for filter 5, that is,

$$V = (R_{c1} + R_{c2} + R_{c3} + R_{c4} + R_{c5}) \cdot I$$

It is to be noted at this point that in the above example the output voltage V would have the same value even if the fifth harmonic did not coincide exactly with the center frequency of the pass band of the fifth bandpass filter. That is, so long as the fifth harmonic lay somewhere within the pass band of the fifth bandpass filter and the resulting unidirectional voltage exceeded the product $I \cdot R_E$, T_5 would conduct all of the current and the magnitude of the output voltage would be determined by Equation 1. However, it has been found that the error between the frequency represented by V and the frequency of the component passed by the corresponding bandpass filter is a tolerable one.

In the example given above it was assumed for purposes of illustration that a formant was characterized by a relatively sharp peak in the speech spectrum so that the formant coincided with a single frequency component having an amplitude substantially larger than the amplitude of any other component within the frequency range of a particular formant. The determination of a formant location by the maximum value selector and interpolator in this situation required only identification of the frequency component having the largest amplitude within the particular formant frequency range. In general, however, these conditions do not prevail since a formant is not necessarily centered upon a single, relatively large speech frequency component; rather, it is more usual for a formant to occur in the vicinity of two or more speech frequency components having relatively large amplitudes, as shown in FIG. 4 by the second and third formants. Accurate determination of a formant location in this situation requires interpolation between the frequencies of two or more speech components having the largest amplitudes within the frequency range of a particular formant. It is a particular feature of the maximum value selector and interpolator shown in FIG. 2 that the output voltage V indicates the occurrence of a formant between the frequencies of two or more speech components having the largest relative amplitudes within a particular formant frequency range by interpolating between the two or more largest components according to their relative amplitudes.

Referring back to FIG. 2, it is evident from Equation 2 that when two transistors, T_i and T_j conduct respective current I_i and I_j , the common emitter bus voltage V_k is given by

$$V_k = V_{B1} - I_i R_E = V_{Bj} - I_j R_E \quad (3)$$

When the magnitudes of V_{B1} and V_{Bj} become sufficiently large to make V_k more positive than the base voltages applied to any other transistors, the other transistors are made nonconductive, and the sum of the respective currents I_i and I_j conducted by T_i and T_j is equal to the total current I , that is,

$$I = I_i + I_j \quad (4)$$

In this situation, Equation 3 may be rewritten

$$V_k = \frac{1}{2} \left[\sum_{n=1}^j V_{Bn} - I R_E \right] \quad (5)$$

that is, the magnitude of V_k is dependent upon both the magnitudes of the base voltages V_{B1} and V_{Bj} and the magnitude of the predetermined constant $I R_E$. The common emitter bus voltage V_k thus serves as a variable threshold against which the magnitudes of all the base voltages are compared, the magnitudes of the two base voltages V_{B1} and V_{Bj} exceeding the magnitude of V_k by

individual amounts $I_i R_E$ and $I_j R_E$ whose sum is equal to the predetermined constant $I R_E$. This is illustrated graphically in FIG. 4 where it is seen that the sum of the differences between V_k and V_{Bi} and V_{Bj} is equal to

$$I_i R_E + I_j R_E = I R_E \quad (6)$$

From Equation 3 the currents I_i and I_j may be expressed in terms of the difference between the base voltages applied to T_i and T_j ,

$$I_i = \frac{1}{2} \left[\frac{(V_{Bi} - V_{Bj})}{R_E} + I \right] \quad (7a)$$

$$I_j = \frac{1}{2} \left[\frac{(V_{Bj} - V_{Bi})}{R_E} + I \right] \quad (7b)$$

Equations 7a and 7b reveal that I_i and I_j vary linearly with the difference in base voltages, with the relative magnitudes of I_i and I_j depending upon which of the two base voltages is larger. For example, if the two base voltages are equal, that is, if $V_{Bi} = V_{Bj}$, then the current I divides equally between T_i and T_j so that $I_i = I_j = \frac{1}{2} I$.

The magnitude of the output voltage V reflects the flow of the currents I_i and I_j through transistors T_i and T_j , respectively, because

$$V = \left(\sum_{n=1}^i R_{cn} \right) I_i + \left(\sum_{n=1}^j R_{cn} \right) I_j \quad (8a)$$

and since $I_i = I - I_j$,

$$V = \left(\sum_{n=1}^i R_{cn} \right) (I - I_j) + \left(\sum_{n=1}^j R_{cn} \right) I_j$$

which may be rewritten

$$V = \left(\sum_{n=1}^i R_{cn} \right) I + \left[\left(\sum_{n=1}^j R_{cn} \right) - \left(\sum_{n=1}^i R_{cn} \right) \right] I_j \quad (8b)$$

Similarly, by substituting $I_j = I - I_i$ in Equation 8a, the output voltage V may also be expressed as

$$V = \left(\sum_{n=1}^j R_{cn} \right) I - \left(\sum_{n=1}^j R_{cn} - \sum_{n=1}^i R_{cn} \right) I_i \quad (8c)$$

In Equations 8b and 8c it is seen that the output voltage V interpolates between the condition in which T_i conducts all of the current and the condition in which T_j conducts all of the current. Thus the first term on the right-hand side of Equations 8b and 8c respectively represents V_i and V_j , the magnitudes of V when all of the current I flows only through T_i or only through T_j , and the second term on the right-hand side of Equation 8b represents the additional voltage due to the flow of current I_j through T_j , while the second term on the right-hand side of Equation 8c represents the amount by which V_j is reduced due to the flow of I_i through T_i . The exact formant location represented by V is therefore seen to be dependent upon the relative amplitudes of the two largest speech components as reflected in the relative magnitudes of the corresponding base voltages V_{Bi} , V_{Bj} , that exceed the threshold V_k , and produce the resulting currents I_i , I_j , according to the equations above.

It is believed to be evident from the foregoing explanation that the maximum value selector and interpolator also interpolates among more than two relatively large frequency components. That is, the value of the output voltage V may be shown to be a function of the respective currents I_i , I_j , I_k , ..., I_N flowing through a number of transistors T_i , T_j , T_k , ..., T_N in FIG. 2, in an extension of the result shown in Equation 8a, that is

$$V = \left(\sum_{n=1}^i R_{cn} \right) I_i + \left(\sum_{n=1}^j R_{cn} \right) I_j + \left(\sum_{n=1}^k R_{cn} \right) I_k + \dots + \left(\sum_{n=1}^N R_{cn} \right) I_N \quad (9)$$

Thus in the situation where base voltages V_{Bi} , V_{Bj} , V_{Bk} , ..., V_{BN} are respectively applied to transistors T_i , T_j , T_k , ..., T_N to cause respective currents I_i , I_j , I_k , ..., I_N to flow, the value of V_k is given by

$$\begin{aligned} V_k &= V_{Bi} - I_i R_E \\ &= V_{Bj} - I_j R_E \\ &= V_{Bk} - I_k R_E \\ &= \dots \\ &= V_{BN} - I_N R_E \end{aligned} \quad (10a)$$

However, the relationship between V_k and the base voltages is the same in the case of several conducting transistors as it is in the cases of one or two conducting transistors: when the magnitudes of V_{Bi} , V_{Bj} , V_{Bk} , ..., V_{BN} become sufficiently large to make V_k more positive than the base voltages applied to any of the other transistors, the other transistors are made nonconductive, and the sum of the respective currents I_i , I_j , I_k , ..., I_N conducted by T_i , T_j , T_k , ..., T_N is equal to the total current I .

$$I = I_i + I_j + I_k + \dots + I_N \quad (10b)$$

In this situation, Equation 10a may be rewritten

$$V_k = \frac{1}{N} \left[\sum_{n=1}^{i+N} V_{Bn} - I R_E \right] \quad (11)$$

The specific case of three base voltages V_{Bi} , V_{Bj} , V_{Bk} greater than the variable threshold V_k is illustrated graphically in FIG. 4, where it is seen that in the vicinity of the third formant that the sum of the differences between V_k and V_{Bi} , V_{Bj} , V_{Bk} is equal to

$$I_i R_E + I_j R_E + I_k R_E = I R_E \quad (12)$$

Equation 11 therefore specifies the conditions under which the maximum value selector and interpolator apparatus of this invention interpolates a formant frequency location among the frequencies of any number of relatively large speech components within a frequency subband which have magnitudes greater than the variable threshold V_k , where the sum of the differences in magnitude between the amplitudes of the two or more relatively large components and the variable threshold is equal to a predetermined constant, and where the variable threshold exceeds the amplitude of any other component.

It is important to observe that the accuracy achieved by the maximum value selector of this invention in determining formant locations is due in large part to the interpolation capability of the maximum value selector and interpolator since most formants occur at an intermediate point between several speech components. By converting absolute speech component amplitudes into logarithmic values, this accuracy is maintained despite wide variations in overall speech energy which produce corresponding wide variations in absolute speech component amplitudes, since the differences in logarithmic values which determine whether interpolation is performed according to Equation 3 correspond to ratios of absolute values.

The discussion above has distinguished between the cases in which the output voltage of the maximum value selector and interpolator may represent either the frequency of a single speech component having an amplitude that is larger than the amplitude of any other component by more than a predetermined amount or a frequency interpolated among several speech components having amplitudes that differ by less than the predetermined amount. However, it may be desirable to interpolate between two different indications of the same frequency component, which may be obtained by constructing bandpass filters 110 to have partially overlapping pass bands with gentle cutoffs so that the same frequency component is represented by two alternating voltages of different amplitudes. In this way the output voltage of the maximum value selector and interpolator of this invention will indicate that a speech component does not coincide exactly with the center frequency of a particular bandpass filter since the output voltage V in this case will represent an interpolation between two different indications of the same frequency component.

Turning now to FIG. 3A, this drawing illustrates a detailed schematic of a suitable logarithmic amplifier and a suitable rectifier which may be employed in the formant detector of this invention shown in FIG. 1. An incoming alternating voltage representative of a speech frequency component lying within the pass band of a preceding bandpass filter 110 is applied to the input point of the logarithmic amplifier 30, from which point complementary symmetry transistors T_1 and T_2 acting together as a high impedance current driver reproduce the applied voltage as an alternating controlled current at the common collector output point 301. The alternating controlled current is passed to ground through the two diodes D_1 and D_2 , which have been selected for logarithmic forward current-voltage characteristics. By connecting the two diodes to present a forward diode characteristic as a load to current in either direction, there is obtained the combined logarithmic current-voltage characteristic shown in FIG. 3B. The overall effect of the current driver and its logarithmic nonlinear load impedance is to generate across the diodes D_1 , D_2 an output voltage whose instantaneous amplitude is proportional to the logarithm of the sinusoidal input voltage amplitude. Transistors T_3 and T_4 form a high input impedance feedback amplifier that increases the amplitude of the logarithmic voltage developed across diodes D_1 , D_2 sufficiently to overcome nonlinearities in diodes D_3 , D_4 of the following full wave rectifier 31. From element 30 the amplified logarithmic voltage is delivered to element 31, in which the two diodes D_3 and D_4 and the two transistors T_5 and T_6 together form a transformerless full wave rectifier. Diodes D_3 and D_4 each work as a half wave rectifier, one for each half of the wave, while transistors T_5 and T_6 act as a difference amplifier amplifying on one-half cycle the rectified output signal of D_3 and on the other half cycle the rectified output signal of D_4 . However, since the output signals of D_3 and D_4 are of opposite polarity, and since these diodes are connected to opposite polarity input terminals of the difference amplifier, each half cycle of the alternating input signal has the same polarity in the output signal of the difference amplifier. The output signal thus developed at the output terminal of rectifier 31 is a unidirectional voltage having an amplitude proportional to the logarithm of the amplitude of a speech frequency component represented by the alternating voltage applied to logarithmic amplifier 30. The unidirectional output signal of rectifier 31 is then passed to a low pass filter 113, as shown in FIG. 1, in order to remove unwanted high frequency components from the unidirectional signal.

Although this invention has been described in terms of speech communication systems of the type shown in FIG. 1, it is to be understood that applications of the principles of this invention are not limited to these systems, but include such related fields as automatic speech recognition, speech processing, and automatic message recording and reproduction. Other applications of this invention are indicated in FIG. 5, in which formant detectors 5-1 through 5-n determine the locations of n selected formants of an incoming speech signal for use in a suitable utilization device 50. In addition, it is to be understood that the above-described embodiments are merely illustrative of the numerous arrangements which may be devised for the principles of this invention by those skilled in the art without departing from the spirit and scope of the invention.

What is claimed is:

1. A speech transmission system that comprises a transmitter station including:

- a source of an incoming speech wave,
- a plurality of formant detectors for deriving from said speech wave a corresponding plurality of formant control signals representative of the frequency locations of a corresponding selected plurality of formants of said speech wave, each of said formant detectors comprising

a plurality of means for separating a selected subband of the amplitude spectrum of said speech wave into its individual speech frequency components to obtain a group of alternating voltages representative of the amplitudes and frequencies of said speech frequency components,

means for obtaining from said group of alternating voltages a corresponding group of unidirectional voltages having magnitudes proportional to the logarithms of the amplitudes of corresponding speech components,

means under the control of said group of unidirectional voltages for generating a control signal representative of the frequency location of a selected formant of said speech wave according to the relative magnitudes of said unidirectional voltages,

said means for generating a control signal including means for deriving from said unidirectional voltages a threshold having a magnitude proportional to the magnitudes of the largest of said unidirectional voltages so that the sum of the differences between said threshold and the magnitudes of said unidirectional voltages which exceed said threshold is equal to a predetermined constant, and

means for deriving from the magnitudes of said unidirectional voltages which exceed said threshold a control signal having a magnitude proportional to the relative magnitudes of said unidirectional voltages which exceed said threshold, and

an excitation detector for deriving from said speech wave a group of excitation control signals indicative of selected characteristics of the excitation characteristics of said speech wave,

a transmission medium for delivering said formant control signals and said excitation control signals to a receiver station, and

at said receiver station,

means for reconstructing from said formant control signals and said excitation signals a replica of said incoming speech wave at said transmitter station.

2. Apparatus for determining the frequency location of a selected formant of a speech wave from the relative amplitudes of the frequency components of said speech wave which lie within a frequency subband that embraces the frequency range in which said formant may appear which comprises

a source of a speech wave,

means supplied with said speech wave for obtaining from said speech wave a group of unidirectional signals with magnitudes proportional to the logarithms of the amplitudes of the frequency components of said speech wave within said frequency subband, and

means under the control of said group of unidirectional signals for comparing the magnitudes of said unidirectional signals against a threshold derived from said magnitudes to generate a formant control signal with a magnitude proportional to the relative magnitudes of each of said unidirectional signals that exceeds said threshold, wherein the sum of the differences in magnitude between said threshold and each of said unidirectional signals is equal to a predetermined constant.

3. Apparatus as defined in claim 2 wherein said means for obtaining a group of unidirectional signals comprises a plurality of signal paths each including

an input terminal connected to said source of a speech wave, an output terminal connected to said means under the control of said group of unidirectional signals, and

a bandpass filter, a logarithmic amplifier, a rectifier, and a low pass filter connected in series between said input terminal and said output terminal.

4. Apparatus for determining the frequency location of a selected formant of a speech wave as defined in claim 2

wherein said means under the control of said group of unidirectional signals comprises

- a source of current of constant magnitude,
 - a first plurality of resistors in one-to-one correspondence with said group of unidirectional signals, wherein each of said first plurality of resistors has an input terminal connected in parallel to said source of current, an output terminal, and a uniform selected resistance, 5
 - a plurality of transistors in one-to-one correspondence with said group of unidirectional signals and said first plurality of resistors, wherein each of said transistors is provided with an emitter terminal connected to the output terminal of the corresponding resistor in said first plurality of resistors, a base terminal, and 10 a collector terminal,
 - a second plurality of resistors in one-to-one correspondence with said group of unidirectional signals and said plurality of transistors, wherein each of said second plurality of resistors is provided with a predetermined resistance, an input terminal, and an 20 output terminal,
 - a source of positive potential having an input point, means for connecting said second plurality of resistors in series with each other to form a resistance path having an output point connected to the input point of said source of positive potential and an input point, 25
 - means for connecting the collector terminal of each of said transistors to the input terminal of the corresponding one of said second plurality of resistors, and 30
 - means for applying each of said unidirectional signals to the base terminal of the corresponding one of said plurality of transistors to develop an output voltage between the input point of said source of positive potential and the input point of said resistance path. 35
5. Apparatus for determining the frequency locations of selected formants of a speech wave from the relative amplitudes of the frequency components of said speech wave which lie within frequency subbands that embrace the respective frequency ranges in which said formants may appear, which comprises 40
- a source of a speech wave,
 - a plurality of formant detectors connected in parallel to said source of a speech wave for deriving from said speech wave a corresponding plurality of output signals representative of the frequency location of a selected formant of said speech wave, each of said formant detectors including 45

- a plurality of contiguous bandpass filters provided with contiguous pass bands that span a frequency subband embracing the frequency range within which a pre-assigned formant normally occurs,
 - means for applying said speech wave in parallel to said bandpass filters to develop a group of alternating voltages representative of the amplitudes of the frequency components of said speech wave which lie within said subband,
 - a plurality of logarithmic amplifiers in one-to-one correspondence with said plurality of bandpass filters for deriving from said group of alternating voltages a corresponding group of logarithmic signals representative of the logarithms of the amplitudes of said alternating voltages,
 - a plurality of rectifiers in one-to-one correspondence with said plurality of logarithmic amplifiers for obtaining from said group of logarithmic signals a corresponding group of unidirectional voltages,
 - a plurality of low pass filters in one-to-one correspondence with said plurality of rectifiers for removing unwanted high frequency components from said group of unidirectional voltages to develop a corresponding group of narrow band control signals having magnitudes representative of the logarithms of the amplitudes of the frequency components of said speech wave, and
 - a maximum value selector and interpolator responsive to the magnitudes of said group of narrow band control signals for generating an output signal indicative of the frequency location of said preassigned formant, wherein 50
- said maximum value selector and interpolator includes means for deriving from said narrow band control signals a threshold having a magnitude proportional to the magnitudes of the largest of said narrow band control signals so that the sum of the differences in magnitude between said threshold and said narrow band control signals which exceed said threshold is equal to a predetermined constant, and
- means for deriving from the magnitudes of said narrow band control signals which exceed said threshold an output signal having a magnitude representative of the relative magnitudes of said narrow band control signals which exceed said threshold. 55

No references cited.

KATHLEEN H. CLAFFY, *Primary Examiner*.
R. MURRAY, *Assistant Examiner*.