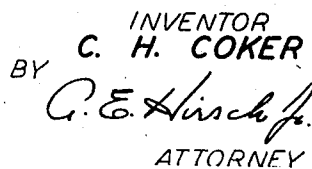


المجلس الأعلى للمعوقين
بمبنى وزارة الشؤون الاجتماعية
الرياض - ١١٥٦١

3,439,122

Sheet 1 of 2



April 15, 1969

C. H. COKER

3,439,122

SPEECH ANALYSIS SYSTEM

Filed June 15, 1966

Sheet 2 of 2

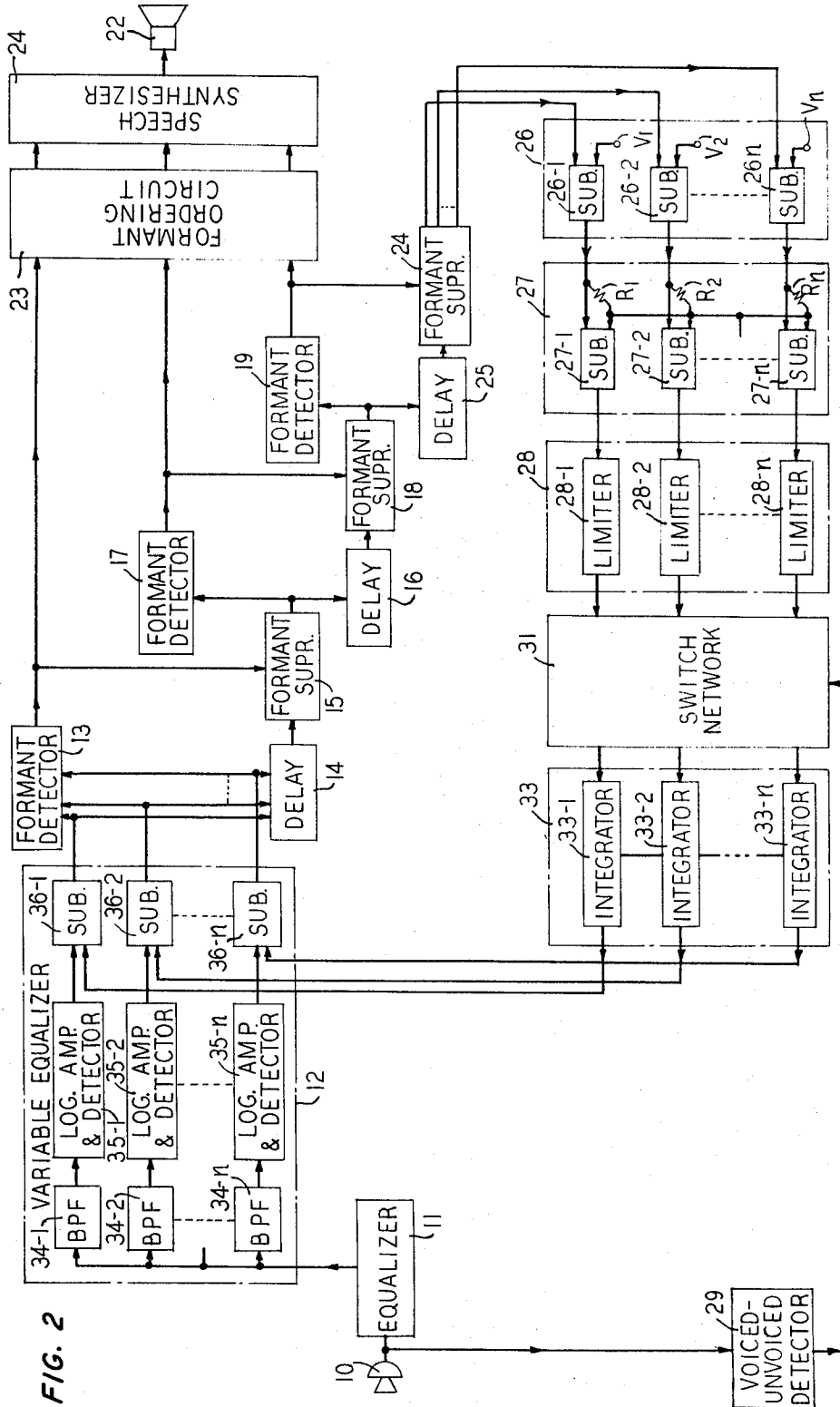


FIG. 2

1

3,439,122

SPEECH ANALYSIS SYSTEM

Cecil H. Coker, Berkeley Heights, N.J., assignor to
Bell Telephone Laboratories, Incorporated, Berkeley
Heights, N.J., a corporation of New York
Filed June 15, 1966, Ser. No. 557,756
Int. Cl. H04m 11/06; G01r 23/16

U.S. Cl. 179—1

6 Claims

ABSTRACT OF THE DISCLOSURE

In a formant detector wherein formants are sequentially detected and then suppressed, a variable equalizer is used to alter the frequency spectrum of the incoming wave in response to an error signal which is derived by averaging the difference between the spectrum of the actual incoming signal and that of a desired optimum spectrum.

This invention pertains to the analysis of speech waves and, more particularly, to the analysis of speech waves in bandwidth compression systems.

In order to make more economical 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 facsimile 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," specific forms of which are described in J. L. Flanagan Patent 2,891,111, issued June 16, 1959, and H. L. Barney Patent 2,819,341, issued Jan. 7, 1958. In a resonance vocoder, the distinctive information-bearing characteristics represented by the control signals and reconstructed at the receiving terminal are the frequency locations of selected peaks or maxima in the speech amplitude spectrum. These selected maxima, formants, correspond to vocal tract resonances, that is, they correspond to frequency regions of relatively effective transmission through a talker's vocal tract. Generally, it is the maxima corresponding to the three principal vocal tract resonances which are selected.

In a typical resonance vocoder analyzer, for example, an analyzer of the type described in the above-mentioned Flanagan patent, the spectrum of an incoming speech wave is divided into three fixed frequency subbands, and each subband embraces a frequency range within which a particular formant normally occurs. From the speech frequency components lying within a subband there is derived a narrow band control signal representative of the frequency at which a formant peak occurs in that frequency subband of the spectrum. However, as pointed out in M. R. Schroeder Patent 2,857,465, issued Oct. 21, 1958, it is an empirical fact that there is substantial overlapping between the frequency ranges within which formants normally occur. From time to time, therefore, a particular fixed frequency subband will contain more than one formant peak, a condition that will result in an erroneous indication of formant frequency location by the narrow band control signal. For example, the narrow band signal derived by the previously-mentioned Flanagan system represents the frequency location of the largest amplitude speech component within the subband, which generally

2

occurs in the vicinity of a formant peak. Hence, a frequency subband designed to embrace the normal frequency range of the second formant, for example, occasionally embraces both the first and second formant peaks, making it highly probable that on occasion the narrow band control signal will represent the first instead of the second formant peak.

The previously-mentioned Barney patent discloses a system that prevents the occurrence of two formants within a single frequency subband by determining formant locations sequentially, instead of simultaneously, by providing a tandem arrangement of high pass filters having variable low-frequency cutoff points. The first narrow band control signal derived in the Barney system represents the location of the first formant, and this signal is employed to adjust the low-frequency cutoff point of a high pass filter through which the incoming speech wave is passed prior to locating the second formant. By adjusting the low-frequency cutoff point of the high pass filter to block the passage of that portion of the speech spectrum which contains the preceding first formant, the frequency subband from which the second formant is determined does not contain the first formant. The same procedure is followed for determining the frequency location of the third formant.

It will be appreciated, however, that the success of the Barney system depends to a large extent upon the adjustment of the low-frequency cutoff point of each of the high pass filters. Since several relatively large speech components typically occur in the vicinity of a formant, the cutoff point must exclude not only the frequency component with the largest amplitude nearest the preceding formant peak but also the relatively large frequency components immediately following the largest amplitude component. For example, if the cutoff point is set at too low a frequency, one or more of the components immediately following the largest component may be included in the frequency subband that is supposed to contain the next higher order formant. Since the components in the vicinity of a lower order formant generally have larger amplitudes than the components in the vicinity of a higher order formant, the presence of components from the vicinity of a preceding formant, in the subband of a subsequent formant, will prevent accurate determination of the location of the subsequent formant.

The present invention is an improvement of the speech analysis system, which overcomes these inaccuracies, disclosed in the copending application of C. H. Coker, Ser. No. 322,390, filed Nov. 8, 1963, now Patent No. 3,327,057. In the invention of that application, the accurate determination of the principal formant locations is facilitated by selectively removing an entire preceding formant peak before determining the location of a subsequent formant. In order to prevent one of the formant detectors from erroneously selecting the location of a preceding low-frequency region of relatively high energy instead of a formant peak, a fixed equalizer is utilized to alter the spectrum. The fixed equalizer is also used to account for certain so-called "invariant" characteristics or frequency distortions of the speech input. These characteristics are attributable to the glottal source and the radiation characteristics of the mouth, as described by G. Fant, Acoustic Theory of Speech Production (1959). In addition, the preprocessing or fixed equalizer also compensates for other distortion introduced by the microphone, amplifier and surrounding environment.

These characteristics are invariant only in the short term sense, that is, they can vary widely from one speaker to another or from one situation to another. A fixed equalizer must necessarily provide equalization which is a compromise or an average to accommodate most speakers and conditions. It is readily apparent, therefore, that a

fixed equalizer may be highly inaccurate in those situations where a variety of different speakers and conditions must be accommodated.

In the present invention the fixed equalizer is complemented by a variable equalizer which changes its processing characteristics in accordance with the nature of the applied signal, and not in accordance merely with some predetermined average value.

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

FIG. 1 is a block diagram illustrating a complete speech transmission system embodying the principles of this invention; and

FIG. 2 is a block diagram illustrating in detail a particular speech transmission system embodying the principles of the present invention.

In FIG. 1, an incoming speech wave from source 10, which may be a conventional transducer for converting speech sounds into a corresponding electrical wave, is applied to fixed equalizer 11. Equalizer 11, which may be of the type disclosed in the aforementioned copending application, Ser. No. 322,390, serves to adjust the amplitudes of the frequency components of the speech wave in order to approximately optimize the operation of the formant detecting apparatus. The equalized speech wave from equalizer 11 is simultaneously applied to formant detector 13, via variable equalizer 12, discussed later, and to delay element 14. Formant detector 13 may be of any well-known construction; a suitable formant detector is described in the copending application of C. H. Coker, Ser. No. 322,389, filed Nov. 8, 1963, now Patent No. 3,327,058. Detector 13 derives from a selected frequency subband of the equalized speech wave a first narrow band control signal representative of the location of the speech formant that normally occurs in that subband, for example, the first speech formant. Delay element 14 serves to delay the equalized speech wave from equalizer 11 by an amount sufficient to compensate for the delay, if any, introduced by formant detector 13 in the detection of a speech formant. The delayed equalized speech wave from delay element 14 is then applied to the input terminal of formant suppressor 15.

Formant suppressor 15, which may have one of the alternative forms shown in copending application, Ser. No. 322,390, is controlled by the narrow band control signal from detector 13. It suppresses a formant peak by suppressing all of the frequency components in the vicinity of the peak in the incoming speech wave, from delay element 14, which correspond to the formant represented by the narrow band control signal. By suppressing all of the frequency components in the vicinity of a formant peak, before detecting the location of the next formant peak, the frequency subband within which the next formant peak normally occurs will not contain any large amplitude components from the vicinity of the suppressed formant peak. Since the detection of formant locations is often based upon the frequency location of the largest amplitude components within a particular frequency subband, the suppression of frequency components in the vicinity of one formant peak prevents these components from being mistakenly recognized as indicating the location of another formant.

The output signal of suppressor 15 is applied to a delay element 16 in order to delay the output signal of suppressor 15 by an amount of time sufficient to compensate for the delay, if any, introduced by detector 17 in deriving a second narrow band control signal representative of another speech formant location. The second narrow band control signal from detector 17 is applied to the control terminal of formant suppressor 18, while the delayed output signal from suppressor 15 is applied to the input terminal of suppressor 18. Suppressor 18, which functions in a manner similar to suppressor 15, serves to suppress in the output signal of suppressor 15

the frequency components in the vicinity of that speech formant which corresponds to the formant represented by the narrow band control signal developed by detector 17. The output signal developed by suppressor 18 and delivered to formant detector 19 therefore has two fewer formants than are found in the original speech wave, so that in the situation where the apparatus of FIG. 1 is designed to locate the three principal speech formants, the output signal of suppressor 18 contains only one principal formant. The output signal of suppressor 18 is passed to formant detector 19, and detector 19 derives from this output signal a third narrow band control signal representative of still another speech formant, for example, the third principal speech formant.

The narrow band control signals developed at the output terminals of detectors 13, 17 and 19 may be utilized to reconstruct a replica of the original speech wave in a suitable synthesizer 21, where it is understood that synthesizer 21 is to be supplied with the usual additional control signals necessary to specify completely the speech characteristics. A suitable synthesizer is disclosed in the previously-mentioned Barney patent. Speech sounds may be reproduced from the reconstructed wave by a suitable transducer 22, for example, a loudspeaker of conventional design.

It is important to note at this point, however, that in order for synthesizer 21 to reconstruct a replica of the speech wave having formants that occur at the same frequency locations as the formants of the original speech wave, it is necessary that the narrow band control signals from detectors 13, 17 and 19 unambiguously identify particular formants of the original speech wave. If the three principal formants are ordered in terms of their relative locations on the frequency scale, then, for a given speech sound, the second formant occurs at higher frequencies than the first formant and lower frequencies than the third formant, and the third formant occurs at higher frequencies than the first and second formants. Although fixed equalizer 11 may be constructed so that the first, second, and third narrow band control signals developed by detectors 13, 17 and 19, respectively, represent the first, second, and third principal speech formants in that order, it is contemplated that other types of equalizers may be employed, in which case, the first, second, and third narrow band control signals respectively developed by the detectors may not necessarily represent the first, second, and third principal formant locations in that order. In the latter event, it is necessary to distinguish between the three narrow band control signals so that the proper narrow band signal may be applied to the proper input point of synthesizer 21, and this may be accomplished by passing the three narrow band control signals through formant ordering circuit 23. Formant ordering circuit 23, which may be of the type disclosed in the above-mentioned copending application Ser. No. 322,390, rearranges the narrow band control signals as necessary in order that the narrow band control signals representative of the formant locations appear in the desired sequence.

In accordance with the present invention, a third formant suppressor 24 is used to suppress the third principal speech formant. Accordingly, the output signal developed by suppressor 18 is delivered to a delay element 25 while the third narrow band control signal is applied to formant suppressor 24. In a manner analogous to that described above, the speech waves appearing at the output of suppressor 18 are delayed by element 25 and applied to suppressor 24. Thus, at the output of suppressor 24 there appears the original speech signal devoid of its first three principal formants. This suppressed speech wave is utilized to develop an error signal indicative of the distortion still present in the speech wave due to the failure of fixed equalizer 11 to properly compensate for the variable characteristics of the speech input, as discussed above. The error signal developed corresponds to the

difference between the spectrum of the signal appearing at the output of suppressor 24 and a predetermined optimum spectrum established by weighting network 26. Therefore, the suppressed speech wave emanating from suppressor 24 is applied to a weighting network 26. The weighting network may comprise a plurality of lumped elements which exhibit the predetermined optimum frequency response desired. Such networks are well known; see for example, Guillemin, *Synthesis of Passive Networks* (1957). Alternatively, network 26 may be of the type shown in FIG. 2 and discussed below.

The operation of weighting network 26 may be more fully understood by first assuming that fixed equalizer 11 is precisely optimum for the particular input signal applied to transducer 10, and that variable equalizer 12 is adjusted to have no effect on the applied speech signals. Then the spectrum of the signal at the output of equalizer 11 would be composed of the three principal speech formants and the residual equalized spectrum; i.e., remaining spectral components, of the speech signal. Since suppressors 15, 18, and 24 remove these three principal formants, the spectrum of the signal appearing at the output of suppressor 24 will correspond to the remaining equalized spectrum of the input speech signal. Weighting network 26 is chosen to have a frequency response which is the inverse of this optimum equalized residual spectrum. Thus, the spectrum of the signal appearing at the output of network 26 will be flat, that is, it will contain all frequency components of the speech wave in equal proportions.

Now suppose that the characteristics of the input speech signal change; then the remaining spectrum after removal of the three speech formants will no longer be optimally equalized, since fixed equalizer 11 is adjusted for a different input speech signal. Thus, portions of the residual spectrum appearing at the output of network 26 will have greater or lesser amplitudes than others. The differences in amplitudes between these spectral portions or components and the average of the entire residual spectrum corresponds to the error deviation of equalizer 11. A signal proportional to this error deviation, after additional processing as discussed below, is utilized to control variable equalizer 12 to compensate for the deviation. Accordingly, the output signal of weighting network 26 is applied to error detector 27 which develops a signal corresponding to the difference between the optimum spectral distribution and the spectral distribution actually present.

The error signal developed by detector 27 is occasionally subject to random perturbations which are a function of the speech signal and apparatus utilized. To prevent these random perturbations, which generally have greater amplitudes than an average error signal, from affecting the operation of the error control loop, limiter 28 clips any signals which have amplitudes above a predetermined level. Signals developed by error detector 27 are meaningless during periods when the input signal to transducer 10 either is not present, or is unvoiced. Accordingly, a conventional voiced-unvoiced detector 29 develops a signal during these intervals, which activates switch 31 via relay 32 and thus opens the error control loop.

Assuming that the input speech signal is voiced, the error signal after passing through limiter 28 is applied to integrator 33, which develops an accumulated control signal, over a period of time, which is utilized to selectively alter the response of variable equalizer 12. Variable equalizer 12 may be of any well-known type, active or passive, which responsive to an applied control signal alters its frequency response. A preferable equalizer is shown in FIG. 2 and its operation is discussed below.

Thus, in accordance with the present invention, the applied input speech signal is optimally equalized regardless of changes in the spectral distribution of the applied signal.

The speech transmission system of FIG. 2 illustrates in detail the operative principles of the present invention. Advantageously, the system has been found to operate optimally when the formant suppressor disclosed in the aforementioned application, Ser. No. 322,390, and shown in FIG. 4 thereof, is utilized. In a manner analogous to that discussed above, an incoming speech wave from source 10 is applied to fixed equalizer 11, which alters the spectrum of the applied speech wave in such a fashion to approximate an optimum spectrum. The speech wave after passage through equalizer 11 is applied to variable equalizer 12. Variable equalizer 12 comprises a plurality of contiguous or overlapping bandpass filters 34-1 through 34-n. The pass bands of filters 34-1 through 34-n span the entire frequency range of the incoming approximately equalized speech signal so that there is developed at the output terminals of these filters a group of alternating signals representative of the frequency components of the incoming speech wave. Each bandpass filter 34-1 through 34-n is followed by a conventional logarithmic amplifier and detector 35-1 through 35-n, respectively. Elements 35-1 through 35-n develop from the group of alternating signals a corresponding group of unidirectional voltages proportional to the logarithms of the amplitudes of the components of the incoming speech wave. The group of unidirectional voltages is combined in subtractors 36-1 through 36-n with the error signals emanating from integrators 33. As discussed above, the error signals correspond to the difference between the actual spectrum present and an optimum spectrum established by weighting network 26. Since unidirectional voltages applied to the subtractors are proportional to the logarithms of the amplitudes of the speech wave components, subtraction or addition corresponds to division or multiplication, respectively, of a nonlogarithmic signal. Thus, in each channel a unidirectional signal is modified in amplitude to correspond to the idealized spectral distribution desired. Speech wave components, after passage through subtractors 36-1 through 36-n, are applied in parallel to formant detector 13 and delay element 14. It is to be noted that although a plurality of inputs are shown for delay element 14, only one output conductor is illustrated. It is to be understood, of course, that this output conductor has a corresponding plurality of leads. However, in order to reduce the complexity of the diagram, a single conductor has been shown connecting all formant suppressors and the delay elements and formant detectors that they activate. As previously mentioned, the formant suppressors are advantageously of the type discussed and shown in the copending application of C. H. Coker, Ser. No. 322,390.

In a manner analogous to the operation of the apparatus shown in FIG. 1 and described above, formant suppressors 15, 18 and 24 delete, from the incoming speech wave spectrum, the three principal speech formants. Also, formant detectors 13, 17 and 19 develop control signals indicative of the formants present in the applied speech signal which, after processing by formant ordering circuit 23 and processing by a speech synthesizer 21, are reproduced as a replica of the speech signal by transducer 22. Thus, at the output of suppressor 24 there appears the original speech signal devoid of its first three principal formants. This suppressed speech wave is applied to weighting network 26 which advantageously comprises a plurality of subtractors 26-1 through 26-n. Each subtractor has an additional input, V_1 through V_n , which corresponds to a source of fixed potential of a predetermined value and polarity. Since the signals applied to weighting network 26 corresponds to the logarithms of the speech component amplitudes, subtraction or addition corresponds to division or multiplication, respectively, of a nonlogarithmic signal. Thus each speech component may be altered in a predetermined fashion to optimally correspond to a desired spectrum.

After processing by weighting network 26 the altered speech components are applied to error detector 27. Detector 27 develops a signal corresponding to the difference between the optimum spectral distribution established by the fixed potential sources of weighting network 26 and the spectral distribution actually present. This is accomplished by the use of subtractors 27-1 through 27-n. Each speech frequency component, after alteration by weighting network 26, is applied to one of these subtractors. In addition, all of the speech frequency components are averaged together by the summing network comprising resistors R₁ through R_n. A signal proportional to the average of the amplitudes of all the speech components is applied to each subtractor. Portions of the residual speech spectra which have not been properly equalized by fixed equalizer 11 will have greater or lesser amplitudes than the other remaining portions of the spectrum. The difference in amplitude between these spectral portions and the average of the entire residual spectrum corresponds to the error deviation of equalizer 11. Accordingly, signals proportional to this error deviation are developed by subtractors 27-1 through 27-n. Each error signal is applied, respectively, to a bank of limiters 28-1 through 28-n which remove random perturbations occasionally present in the applied speech signal. Signals developed by limiters 28 are meaningless during periods when the input signal to transducer 10 either is not present or is unvoiced. Accordingly, a conventional voiced-unvoiced detector 29 responsive to the output of transducer 10 develops a signal during these intervals which activates switch network 31 and thus opens the error control loop. Switch network 31 may conveniently comprise a plurality of switches, one in each speech component channel, operated by a relay responsive to signals developed by detector 29. Assuming that the input speech signal is voiced, the error signal after passing through switch network 31 is applied to element 33 which comprises a plurality of individual integrators 33-1 through 33-n. Each integrator develops an accumulated error control signal, over a period of time, which is utilized to selectively alter the response of variable equalizer 12 by altering the amplitude of the applied speech component signals through the medium of subtractors 36-1 through 36-n.

Thus an applied input speech signal is optimally equalized regardless of changes in the spectral distribution of the applied signal or the environment in which the apparatus is located.

Although this invention has been described in terms of speech transmission systems, it is to be understood that the principles of this invention are applicable to such related fields as automatic speech recognition, speech processing and automatic message recording and reproduction. In addition, it is to be understood that the above-described embodiments are merely illustrative of the numerous arrangements which may be devised for the principles of this invention by those skilled in the art without departing from the spirit and scope of this invention.

What is claimed is:

1. Apparatus for determining the frequency locations of selected formants of speech sounds which comprises: means for receiving an incoming speech wave, fixed equalizer means supplied with said speech wave for enhancing the relative amplitudes of the high frequency components of said speech wave by predetermined relative amounts and for eliminating selected frequency components of said speech wave, thereby to develop an equalized speech wave, variable equalizer means supplied with said equalized speech wave for compensating for the distortion inherent in said equalized speech wave, first detector means in circuit relation with said variable equalizer means for deriving from said compensated speech wave a first control signal representative of the frequency location of a formant of said equalized speech wave,

first suppressor means under the control of said first control signal and responsive to said variable equalizer means for individually suppressing in said compensated speech wave each frequency component in the vicinity of the formant represented by said first control signal to obtain a first suppressed formant speech wave,

second detector means in circuit relation with said first suppressor means for deriving from said first suppressed formant speech wave a second control signal representative of the frequency location of a formant of said first suppressed formant speech wave,

second suppressor means under the control of said second control signal and supplied with said first suppressed formant speech wave for individually suppressing in said suppressed formant speech wave each frequency component in the vicinity of the formant represented by said second control signal to obtain a second suppressed formant speech wave,

third detector means in circuit relation with said second suppressor means for deriving from said second suppressed formant speech wave a third control signal representative of the frequency location of a formant of said second suppressed formant speech wave,

third suppressor means under the control of said third control signal and supplied with said second suppressed formant speech wave for individually suppressing in said suppressed formant speech wave each frequency component in the vicinity of the formant represented by said third control signal to obtain a third suppressed formant speech wave,

means responsive to said third suppressed formant speech wave for developing an error signal which corresponds to the difference between the spectrum of said third suppressed speech wave and a predetermined optimum spectrum,

means responsive to said error signal for selectively altering the response of said variable equalizer.

and speech synthesizing means responsive to said first, second and third control signals for reconstructing a replica of said incoming speech wave.

2. Apparatus as defined in claim 1 wherein said variable equalizer means comprises:

means for developing a plurality of signals representative of the spectral components of said equalized speech wave,

means responsive to said representative signals for developing a plurality of signals each individually proportional to the logarithms of the amplitudes of the spectral components of said equalized speech wave, and means responsive to said proportional logarithmic signals and said error signal for altering the amplitude of each spectral component to establish a predetermined optimum spectrum.

3. Apparatus as defined in claim 1 wherein said means for developing an error signal comprises:

means for developing a plurality of signals proportional to the difference between the amplitudes of each of the spectral components of said third suppressed formant speech wave and a predetermined optimum value for each spectral component,

means for developing a signal proportional to the average of said proportional difference signals,

and means for developing a plurality of signals proportional to the difference between each of said proportional difference signals and the average of said proportional difference signals.

4. Speech analyzer apparatus for developing signals representative of the locations of selected formants of a speech wave comprising:

means for receiving an incoming speech wave, variable equalizer means supplied with said speech wave for compensating said speech wave,

means responsive to said compensated speech wave for developing a plurality of signals representative of the

9

frequency locations of the formants of said speech wave,

means for simultaneously suppressing in said speech wave the formants represented by said plurality of signals,

means responsive to said suppressed speech wave for developing an error signal which corresponds to the difference between the spectrum of said suppressed speech wave and a predetermined optimum spectrum, and means responsive to said error signal for selectively altering the frequency response of said variable equalizer.

5. Apparatus as defined in claim 4 wherein said variable equalizer means comprises:

means for developing a plurality of signals representative of the spectral components of said speech wave, means responsive to said representative signals for developing a plurality of signals each individually proportional to the logarithms of the amplitudes of the spectral components of said speech wave, and means responsive to said proportional logarithmic signals and said error signal for altering the amplitude of each spectral component to establish a predetermined optimum spectrum.

6. Apparatus as defined in claim 4 wherein said means for developing an error signal comprises:

10

means for developing a plurality of signals proportional to the difference between the amplitudes of each of the spectral components of said suppressed speech wave and a predetermined optimum value for each spectral component,

means for developing a signal proportional to the average of said proportional difference signals,

and means for developing a plurality of signals proportional to the difference between each of said proportional difference signals and the average of said proportional difference signals.

References Cited

UNITED STATES PATENTS

3,327,058	6/1967	Coker.	
3,327,057	6/1967	Coker.	
2,891,111	6/1959	Flanagan.	
2,866,001	12/1958	Smith	179—15.55

20 KATHLEEN H. CLAFFY, *Primary Examiner.*

R. P. TAYLOR, *Assistant Examiner.*

U.S. Cl. X.R.

25 179—15.55; 324—77