

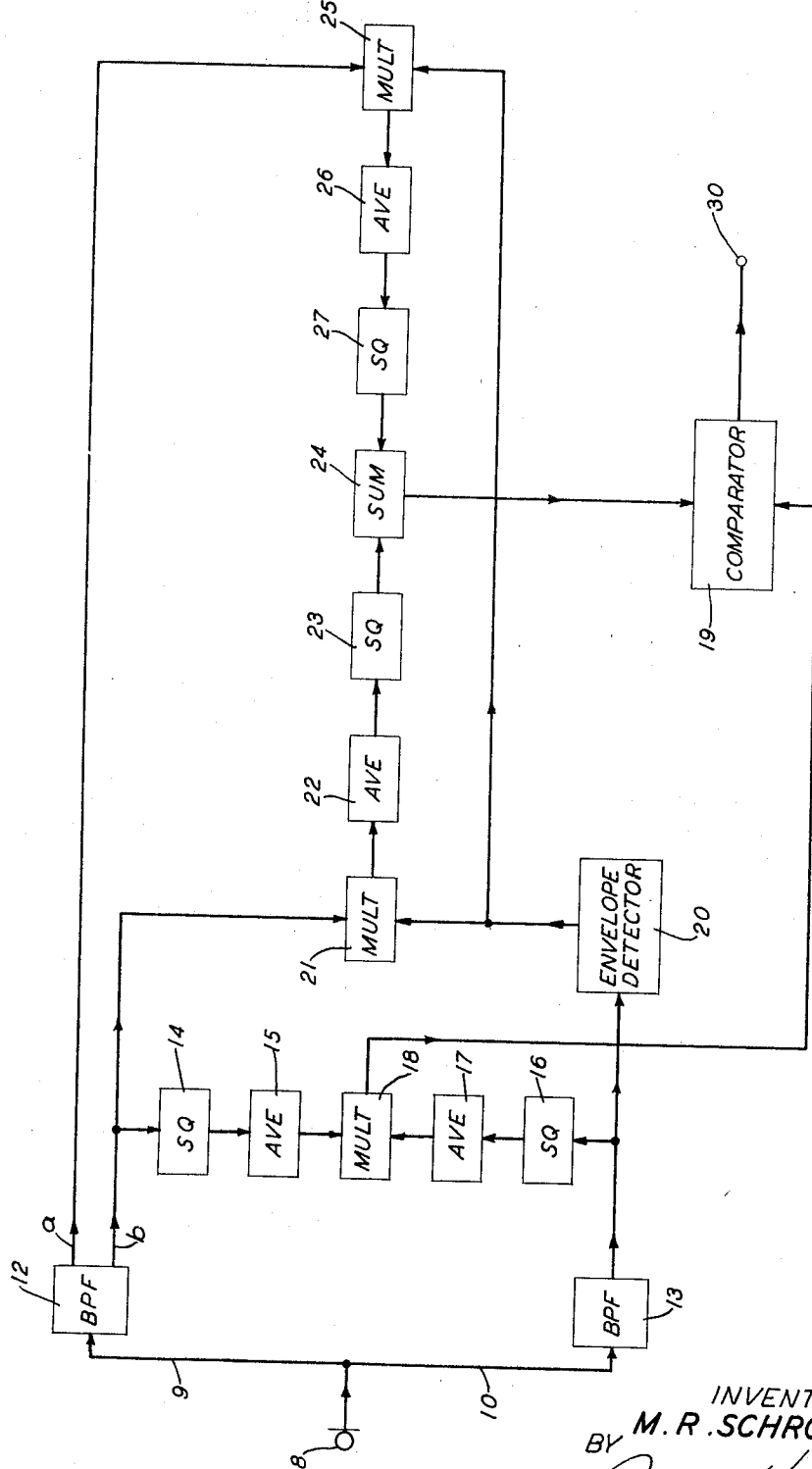
April 21, 1970

M. R. SCHROEDER

3,507,999

SPEECH-NOISE DISCRIMINATOR

Filed Dec. 20, 1967



INVENTOR
BY M. R. SCHROEDER
Arnell C. L. Smith, Jr.
ATTORNEY

1

3,507,999

SPEECH-NOISE DISCRIMINATOR

Manfred R. Schroeder, Gillette, N.J., assignor to Bell Telephone Laboratories, Incorporated, Murray Hill, N.J., a corporation of New York

Filed Dec. 20, 1967, Ser. No. 692,230

Int. Cl. G101 1/02

U.S. Cl. 179—1

9 Claims

ABSTRACT OF THE DISCLOSURE

The presence of speech is detected in a speech-noise signal by examining signal components in a plurality of contiguous spectral regions for common periodicity. By comparing a signal representative of the degree of common periodicity between such spectral components with a reference signal proportional to the power of the received signal difficulties caused by relative variations in the noise level and in the frequency of the speech or noise signals are eliminated.

This invention relates to wave analyzing apparatus, and more specifically to apparatus for identifying the presence of a speech wave in an ambient noise environment.

BACKGROUND OF THE INVENTION

Communication signals in general and speech signals in particular are frequently mixed with extraneous noise which may mask the desired signal. Such noise may originate in the communication system itself or may be generated by unwanted sources at the transmission point. In either situation, the noise is likely to vary in both amplitude and frequency and may occupy the same frequency spectrum as the desired voice signal thereby making the identification of a speech signal particularly difficult.

There are several important features which distinguish a speech signal from noise. As is well-known, normal speech signals are composed of voiced and unvoiced speech elements. Voiced speech elements have an amplitude spectrum composed of a number of individual frequency components of various amplitudes which occur at harmonics of the fundamental frequency of the sound. These voiced speech signal components are periodic with a common fundamental period in all parts of the frequency spectrum. Noise and unvoiced speech sounds normally have a more continuous, nonharmonic frequency structure. Even when noise does contain discrete frequency constituents, these constituents rarely have a single fundamental period.

Field of the invention

The ability to distinguish between voiced speech and noise or unvoiced speech is important in many communication systems. In voice operated equipment and in automatic microphone directing systems, the ability to distinguish speech from noise quickly and automatically is particularly critical. Similarly, the coding schemes employed in many vocoder speech communication systems require that the transmitting station distinguish between voiced and unvoiced speech portions before coding.

DESCRIPTION OF THE PRIOR ART

In vocoder systems, the separation of voiced and unvoiced speech portions is often accomplished by dividing a speech signal into two spectral regions and subtracting the signal in the high frequency region from the low frequency signal. Since a voiced sound has its predominant energy components centered in the low frequency portion of the spectrum and an unvoiced sound is composed of predominantly higher frequency energy, this difference is generally positive when voiced energy is present and

2

negative otherwise. However, such processing is not adequate for distinguishing speech in a high level variable ambient noise environment since it is sensitive to relative amplitude differences in the spectral components. Further, such systems do not indicate the relative strength of speech energy present. Other methods of voiced speech identification including time domain analysis by autocorrelation have been employed but have not proved entirely satisfactory.

SUMMARY OF THE INVENTION

Consequently, it is an object of the present invention to detect the presence and relative intensity of a speech signal in a high level variable ambient noise environment.

In accordance with the present invention, two signals are derived from the signal to be analyzed. The first is a reference signal and is representative of the power of the received signal. The reference signal is derived by dividing the signal to be analyzed into two contiguous frequency subbands, extracting signals proportional to the power in each subband, and multiplying these signals together.

The second is a comparison signal and is proportional to the degree of common periodicity between the high and low frequency components of the received signal. The comparison signal is derived by processing the signal in either one of the subband channels by a periodicity-preserving nonlinear process to produce a signal with frequency components which overlap the components of the signal in the other subband channel. The two overlapping subband signals are then multiplied together in such a way as to compensate for phase differences between them. The resulting product, averaged over a period of time, is the comparison signal.

The reference and comparison signals are compared in an appropriate comparator or threshold detector to determine the presence or absence of speech or the proportion of speech in the input.

BRIEF DESCRIPTION OF THE DRAWING

The invention will be fully apprehended from the following description of an illustrative embodiment thereof, taken in conjunction with the appended drawing which is a block schematic diagram of apparatus for detecting the presence of speech in a variable ambient noise environment.

In accordance with a preferred embodiment of the invention, and referring to the drawing, an acoustic wave is converted to an electrical analog signal by transducer 8. To determine whether or not speech is present in this analog signal, the signal is applied to bandpass filtering networks 12 and 13 by means of signal channels 9 and 10.

Bandpass filter 12 is adjusted to pass all signal components below a selected frequency and block all others and is constructed to provide two outputs in quadrature phase, at terminals 12a and 12b. The design of such filtering networks is well-known. Bandpass filtering network 13 is adjusted to pass all signal components above the selected frequency. This separation of the input signal into two contiguous frequency subbands is a first step in the comparison of the low and high frequency energy components of the input signal. When the comparison of these components reveals that they are periodic with a common fundamental period, the presence of speech is indicated. A reference for such comparison is derived as follows:

A first output of network 12, at 12b in the drawing, is delivered to the input of squaring network 14. Network 14 is of well-known design and produces a signal at its output terminal proportional to the square of the signal at its input. The output of squaring network 14 is applied to averaging network 15 which produces an output signal proportional to the input signal averaged over a time

period. This squaring and averaging process produces a signal proportional to the signal power in the output of network 12.

The output of BPF network 13 is similarly applied to a squaring network, 16, and then applied to averaging network 17. Output signals from averaging networks 17 and 15 are applied to multiplying network 18. Network 18 produces a "reference" signal proportional to the product of the power in the signal passed by filter 12 and the power in the signal passed by filter 13. This reference signal is applied to comparator network 19 for comparison with a "comparison" signal.

To derive a "comparison" signal which is representative of the degree of common periodicity between the signal in filters 13 and 14, the output of BPF network 13 is applied to envelope detecting network 20 which produces a signal proportional to the envelope of the applied signal. Frequency components contained in this envelope signal overlap the components of the signal passed by filter 12. The product of these signals will accentuate common periodicities between them. The output of envelope detecting network 20 is applied to multiplying network 21 together with the first output of BPF network 12. Multiplying network 21 produces a first product signal proportional to the product of the applied signals. This first product signal is applied to averaging network 22 which provides an output proportional to the first product signal averaged over a time period. By applying this averaged signal to squaring network 23 a signal proportional to the square of the averaged signal is produced. This averaged and squared first product signal is applied to summing network 24 for addition to a second product signal.

The second product signal, which is required to compensate for phase differences between the low and high frequency components of the received signal, is formed as follows. The second output 12a of bandpass filtering network 12, which carries a signal 90 degrees out of phase with the signal on 12b, and the output of envelope detecting network 20 are applied to product network 25 which produces a signal proportional to the product of these two signals. This signal is applied to averaging network 26 and then to squaring network 27, which have the same effect as networks 22 and 23. The resulting averaged and squared signal is applied to summing network 24, the output of which is proportional to the sum of the two signals applied to it. This output is directed to comparator circuit 19 for comparison with the reference signal.

Comparator 19 may, in one illustrative embodiment, be designed to provide a binary output. In this case, when the comparison signal is a specified increment larger than the reference signal or when the ratio of the two signals exceeds a given ratio, a voiced signal in the input is indicated, for example, by a binary "one" signal in output channel 30. When the comparison signal differs from the reference signal by less than the specified increment or the ratio falls below the given ratio, the absence of voiced signal components in the input signal is indicated, for example, by a binary "zero" signal in output channel 30. In an alternative embodiment, the comparator may provide a signal in channel 30 proportional to the ratio of the comparison signal to the reference signal. Such a signal represents the quantity of speech energy present at the transducer 8. Comparator circuits of either construction are well-known in the art.

It should be apparent from the foregoing discussion that the comparator's operation is independent of relative changes in the amplitudes of the high and low frequency components of the speech-noise signal. For example, if the amplitude of the high frequency component of the input signal increases by a factor of 2 and the amplitude of the low frequency signal component increases by a factor of 3 the reference signal increases by a factor of 36. Under these circumstances, the comparison signal similarly increases by a factor of 36 thereby facilitating

the detection of speech in the presence of a variable noise signal.

In a variation of the above embodiment of the invention, the input signal to be analyzed may be divided into more than two contiguous frequency subbands. In this case, each subband is paired with another and each subband pair is processed as described above. The criterion for determining the presence of speech will depend on the application in which the invention is employed. An example of such criterion is as follows: If any preselected number of subband pairs indicates the presence of speech, then speech is considered to be present in the input signal.

It is to be understood that the above-described arrangements are merely illustrative of application of the principles of the invention. Other arrangements may be devised by those skilled in the art without departing from the spirit and scope of the invention.

What is claimed is:

1. A speech discriminator comprising, means for separating an input signal into a plurality of contiguous frequency subband signals, a plurality of signal channels each supplied with one of said subband signals, means for deriving a reference signal representative of the signals in a first and a second of said channels, means for deriving a comparison signal representative of the degree of common periodicity between the signals in said first and said second signal channels, means for comparing said comparison signal with said reference signal.

2. A speech discriminator as defined in claim 1 wherein the signal in said first signal channel is lower in frequency than the signal in said second signal channel, and wherein said means for deriving said comparison signal includes means for deriving an envelope signal proportional to the envelope of the signal in said second signal channel, and means for combining said envelope signal with the signal in said first signal channel.

3. A speech discriminator as defined in claim 1 wherein said means for deriving a comparison signal includes nonlinear, periodicity preserving means for processing said signal in a selected one of said signal channels.

4. A speech discriminator as defined in claim 3 wherein said means for deriving a comparison signal further includes means for deriving a first signal proportional to the product of said processed signal in said selected signal channel and said signal in a second selected signal channel, means for deriving a phase shifted signal proportional to the signal in said second selected signal channel shifted in phase, means for deriving a second signal proportional to the squared and time averaged product of said phase shifted signal and said signal in said selected signal channel, and means for deriving a sum signal proportional to the sum of said first and said second signals.

5. Apparatus for identifying the presence of speech in an analog signal comprising, means for deriving a first signal proportional to the low frequency components of an input signal, means for deriving a second signal proportional to the envelope of the high frequency components of an input signal, means for deriving a reference signal indicative of the power in said input signal, means for compensating for phase differences between said first and said second signal, means for deriving a product signal proportional to the product of said first signal and said second signal, means for deriving a threshold measure from said reference signal, means for comparing said reference signal with said threshold quantity.

6. A speech discriminator comprising, means for separating an input signal into two contiguous frequency subband signals, first and second signal channels each provided with one of said subband signals, means for deriving a signal proportional to the signal in said first channel shifted in phase by 90 degrees, means for deriving an envelope signal proportional to the envelope of

5

the signal in said second signal channel, means for deriving a first product signal proportional to the product of said first subband signal and said envelope signal, means for deriving a second product signal proportional to the product of said envelope signal and said phase shifted signal, means for deriving a third signal proportional to the time average of the square of said first product signal, means for deriving a fourth signal proportional to the time average of the square of said second product signal, adder means for deriving a sum signal proportional to the sum of said third signal and said fourth signal, means for deriving a reference signal related to said first and second subband signals, means for comparing said reference signal with said sum signal, means for indicating the result of said comparison. 15

7. A speech discriminator as defined in claim 6 wherein said means for deriving a reference signal comprises, means for deriving a first power signal proportional to the power of said first subband signal, means for deriving a second power signal proportional to the power of said 20

6

second subband signal, and multiplier means for multiplying said first power signal with said second power signal.

8. A speech discriminator as defined in claim 6 wherein, said means for comparing said reference signal with said sum signal comprises a comparator which provides output signals proportional to the ratio of said sum signal and said comparison signal.

9. A speech discriminator as defined in claim 6 wherein said means for comparing said reference signal with said sum signal comprises a comparator which provides a binary output.

References Cited

UNITED STATES PATENTS

3,238,303 3/1966 Dersch ----- 179—1

WILLIAM C. COOPER, Primary Examiner

D. W. OLMS, Assistant Examiner