

Oct. 9, 1962

J. F. KAISER

3,057,960

NORMALIZED SOUND CONTROL SYSTEM

Filed March 13, 1961

3 Sheets-Sheet 1

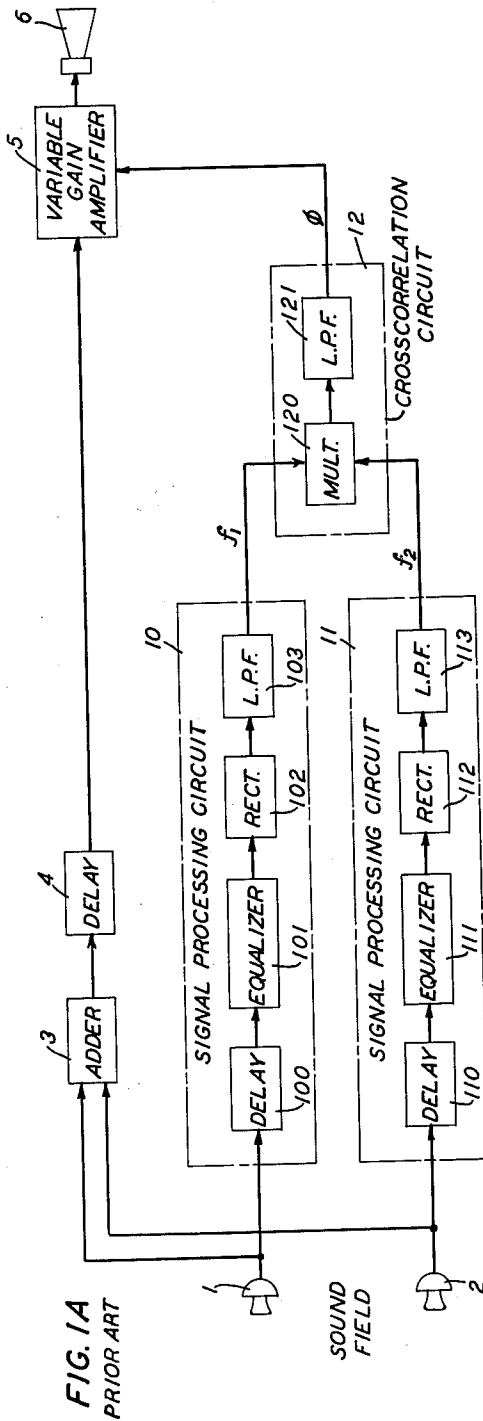
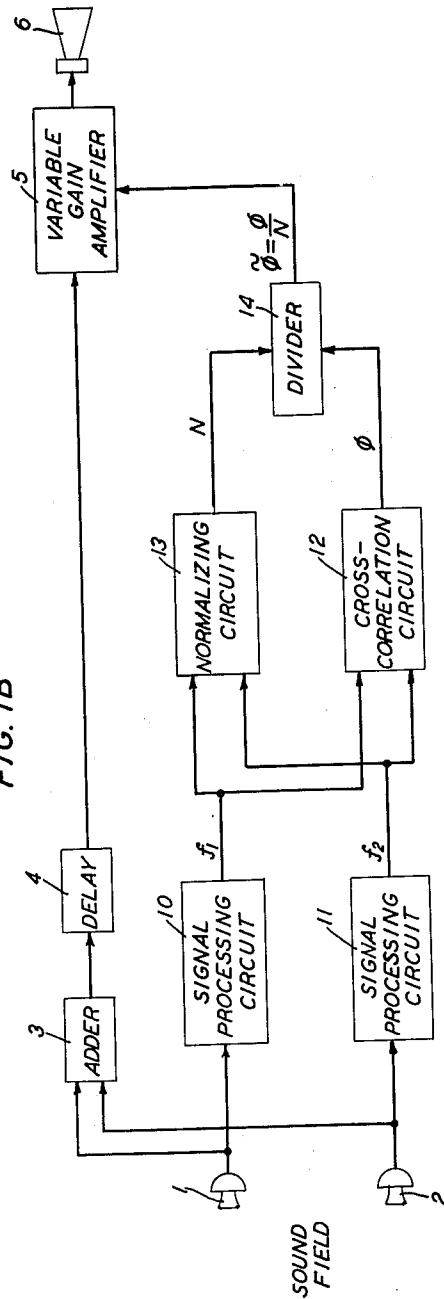


FIG. 1B



INVENTOR
BY J. F. KAISER
G. E. Hirsch
ATTORNEY

Oct. 9, 1962

J. F. KAISER

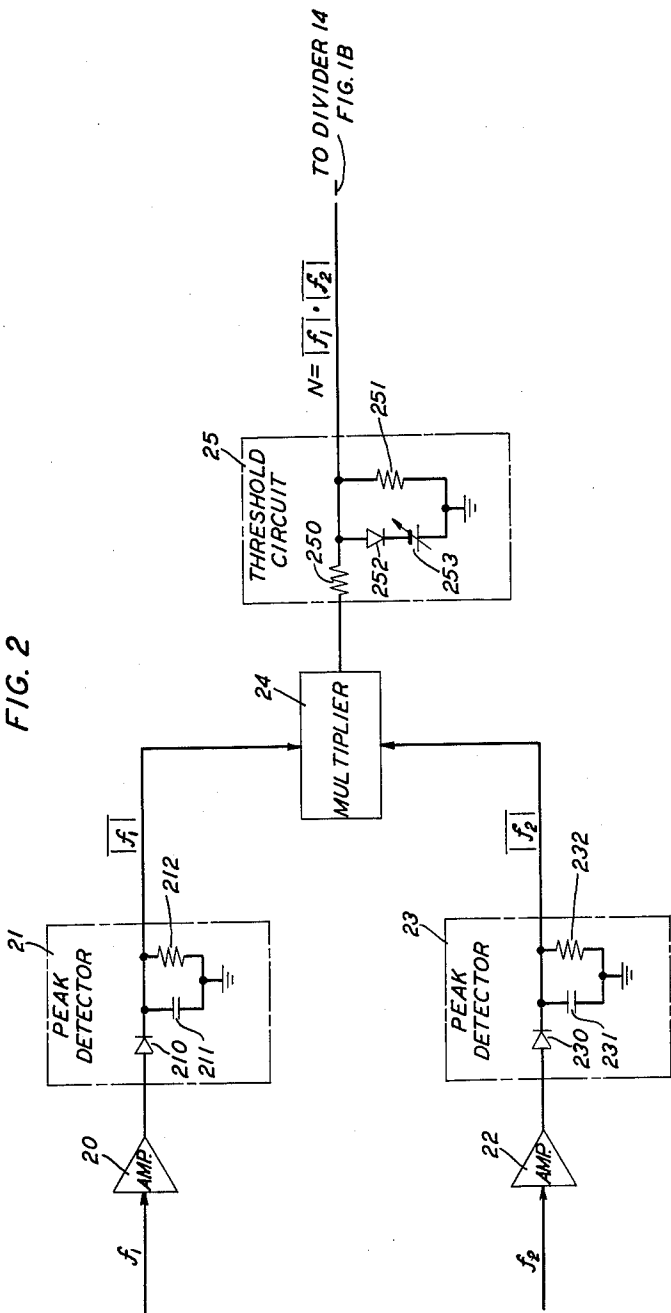
3,057,960

NORMALIZED SOUND CONTROL SYSTEM

Filed March 13, 1961

3 Sheets-Sheet 2

FIG. 2



INVENTOR
J. F. KAISER
BY
G. E. Hirsch
ATTORNEY

Oct. 9, 1962

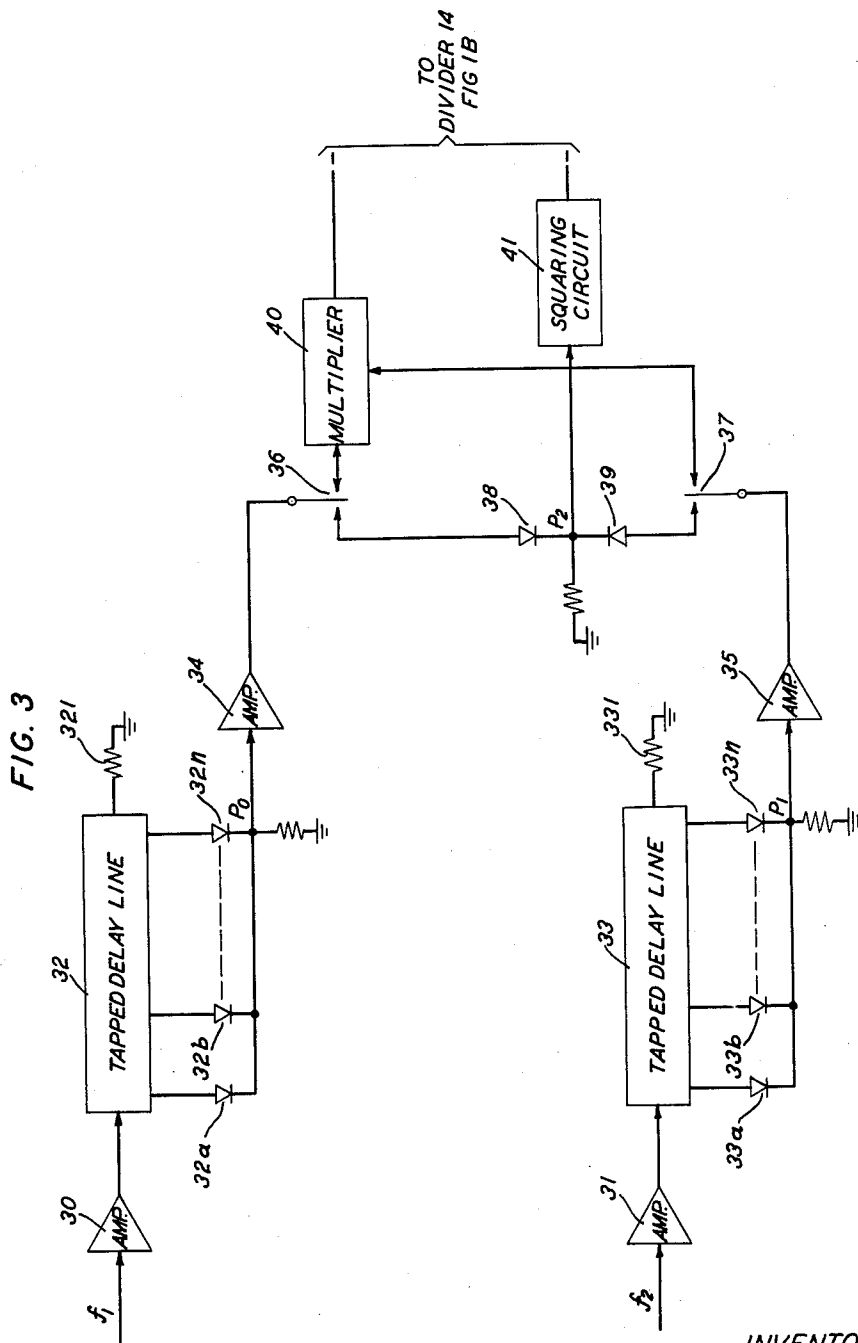
J. F. KAISER

3,057,960

NORMALIZED SOUND CONTROL SYSTEM

Filed March 13, 1961

3 Sheets-Sheet 3



INVENTOR
J. F. KAISER
BY
G. E. Hirsch
ATTORNEY

1

3,057,960

NORMALIZED SOUND CONTROL SYSTEM
James F. Kaiser, Summit, N.J., assignor to Bell Telephone Laboratories, Incorporated, New York, N.Y., a corporation of New York

Filed Mar. 13, 1961, Ser. No. 95,087

9 Claims. (Cl. 179—1)

This invention relates to sound control systems for improving the intelligibility of speech from a talker located in a noisy environment.

The sound control system described by E. E. David, Jr. in a copending application filed December 15, 1959, Serial No. 859,684, improves the intelligibility of speech from a talker located in a noisy environment by simulating in part of the so-called "cocktail party effect." The cocktail party effect is a psychoacoustic mechanism by which a human listener is able to concentrate on speech from a specific talker, despite a surrounding noise or speech babble whose sound intensity is approximately equal to that of the speech of the preferred talker. In the David system, two spatially separated microphones are placed in a sound field containing both speech from a preferred talker and background noise and speech of equal or higher intensity from other talkers. The output signals of the two microphones are combined to form a composite signal, and the amplitude of the composite signal is modulated or varied by a control signal. The properties of the control signal are selected to reinforce speech sounds from the preferred talker and to suppress noise and speech sounds from other talkers. This selective reinforcement and suppression in the David system improves the intelligibility of speech from the preferred talker and decreases the intelligibility of speech from other talkers, thereby simulating the ability of the human listener to concentrate on speech from a preferred talker in spite of background noise of equal or higher intensity.

In one embodiment of the David invention, the control signal is derived by cross-correlating the two microphone signals; that is, the two microphone signals are multiplied together and the product of the two signals is averaged over an appropriate time interval. By suitably delaying one or the other of the two microphone signals before cross-correlating them, those portions of the two microphone signals which correspond to speech sounds from the preferred talker are brought into time coincidence, thereby causing the magnitude of the cross-correlation control signal to be relatively large when speech sounds from the preferred talker are present, and to be relatively small when speech sounds from the preferred talker are not present. The magnitude of the cross-correlation signal thus provides an accurate measure of the presence or absence of speech from the preferred talker, and modulation of the composite signal by the cross-correlation signal serves to reinforce the preferred talker's speech sounds, which are coherent with the cross-correlation signal, and to suppress the background noise and speech sounds, which are incoherent with the cross-correlation signal. Varying the amplitude of the composite signal in this fashion increases the intelligibility of speech from the preferred talker and decreases the intelligibility of noise and speech from background talkers.

Although the David system achieves a significant improvement in the intelligibility of speech from a particular talker, in the embodiment that utilizes the cross correlation of the microphone signals as a control signal the improvement in intelligibility is accompanied by an audible distortion in intensity. This distortion is particularly noticeable during changes in the intensity of speech sounds from the preferred talker.

2

Investigation into the source of distortion in the cross-correlation version of the David system revealed that the magnitude of the cross-correlation signal is a function of more than one independent variable. The most important independent variable is the degree of time coincidence between speech sounds arriving at the two microphones, since it is by adjusting this variable that speech sounds from the preferred talker are brought into time coincidence, thereby creating large magnitudes in the cross-correlation signal which serve to reinforce speech from the preferred talker. A second independent variable is the average intensity of the sound energy arriving at the two microphones, and this variable also increases the magnitude of the cross-correlation signal. It has been determined by applicant that this additional source of increased magnitude produces an excessive reinforcement of speech from the preferred talker which is audible as the distortion in intensity referred to above.

It is the principal object of the present invention to improve the intelligibility of speech sounds from a preferred talker located in a noisy environment without distorting the intensity of the preferred speech sounds.

It is a specific object of this invention to eliminate distortion from the cross-correlation version of the David system by removing the dependence of the magnitude of the cross-correlation control signal upon the average intensity of the sound energy reaching the two microphones.

The present invention eliminates distortion from the David system by normalizing the cross-correlation signal to render it insensitive to the average intensity of the sound energy reaching the two microphones, without disturbing the dependence of the cross-correlation signal upon the degree of time coincidence between speech sounds arriving at the two microphones. In this invention, normalizing comprises dividing the cross-correlation signal by a normalizing signal that is also a function of the average intensity of the sound energy arriving at the microphones, thereby effectively canceling the dependence of the cross-correlation signal upon average sound intensity. Modulation of the composite signal by the normalized cross-correlation signal of the present invention enhances the intelligibility of speech sounds from a preferred talker without distorting the intensity of the preferred speech sounds.

This invention provides two alternative systems for deriving a normalizing signal. In the first system, the normalizing signal is derived by obtaining the product of the average magnitudes of the two microphone signals. In the second system, the largest value of each microphone signal over several pitch periods is obtained, and the normalizing signal is derived either by multiplying together the largest values of the two microphone signals or by squaring the larger of the two largest values.

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

FIG. 1A is a schematic diagram of the cross-correlation sound control system described in the aforementioned David application;

FIG. 1B is a block diagram of a sound control system embodying the normalizing apparatus of this invention;

FIG. 2 is a schematic diagram of a preferred embodiment of the normalizing apparatus of this invention; and

FIG. 3 is a schematic diagram of normalizing apparatus alternative to that of FIG. 2.

Theoretical Considerations

Referring to FIG. 1A, there is shown a cross-correlation version of the sound control system described in the aforementioned David application. Spatially separated

microphones 1, 2 are placed in a sound field containing both speech sounds from a preferred source and background noise from other sources, and each microphone generates an electrical signal in response to the sound energy it receives. The microphone signals are combined in adder 3 to form a composite signal whose amplitude is modulated or adjusted in a variable gain amplifier 5. Amplifier 5 is controlled by a cross-correlation signal derived from the microphone signals by signal processing circuits 10, 11 and cross-correlation circuit 12. Delay element 4 synchronizes the composite signal with the control signal, and reproducer 6 converts the modulated composite signal into audible sound waves in which the intelligibility of speech sounds from the preferred source is substantially improved.

Signal processing circuits 10, 11 are of similar construction, comprising, respectively, delay elements 100, 110, equalizers 101, 111, rectifiers 102, 112, and low-pass filters 103, 113. The delay elements bring into time coincidence the portions of each microphone signal attributable to speech sounds from the preferred talker; the equalizers pre-emphasize selected frequencies; and the rectifiers, which are preferably of the half-wave variety, together with the low-pass filters serve to make the polarity of the microphone signals unidirectional and to pass only the prominent features of the unidirectional signals. The processed microphone signals from circuits 10, 11 are applied to cross-correlation circuit 12, where the two signals are multiplied together in multiplier 120 and the product of the two signals is averaged by low-pass filter 121 to form a cross-correlation control signal.

The cross-correlation operation performed by circuit 12 may be expressed as follows:

$$\begin{aligned}\varphi(t) &= \overline{f_1(x) \cdot f_2(x+t)} \\ &= \overline{f_1(x+t) \cdot f_2(x)}\end{aligned}\quad (1)$$

where $f_1(x)$, $f_2(x)$ denote the microphone signals, the independent variable t represents the amount of time by which one of the microphone signals is delayed with respect to the other microphone signal in order to create time coincidence, the superposed bar denotes the time averaging performed by low-pass filter 121 upon the product of the delayed microphone signals, and $\varphi(t)$ is the cross-correlation signal derived by circuit 12.

It is well known that for two functions that are in some manner related to each other, the magnitude of the cross correlation of the two functions, as expressed in Equation 1, depends primarily upon the independent variable t , that is, the cross-correlation function is a measure of the time coherence between related portions of the two functions; for example, see Y. W. Lee, T. P. Cheatham, Jr., J. B. Wiesner, "Application of Correlation Analysis to the Detection of Periodic Signals in Noise," volume 38, Proceedings of the I.R.E., page 1165 (1950). Thus the magnitude of the cross-correlation function increases as the time coherence of the two functions increases, starting from a minimum value when the related portions of the two functions are most incoherent, and reaching a maximum value when the related portions are exactly time coincident.

In the case of the cross-correlation control signal derived from the two microphone signals by the apparatus of FIG. 1A, the independent variable t is adjusted via delay elements 100, 110 to bring into time coincidence the portions of both microphone signals which are attributable to speech sounds from the preferred source. This adjustment of the variable t causes the magnitude of the cross-correlation signal to be relatively large when speech sounds from the preferred source are present, and to be relatively small, with rare exceptions caused by chance coincidences, at all other times. Varying the amplitude of the composite signal from adder 3 in amplifier 5 in response to the cross-correlation signal from circuit 12 improves the intelligibility of speech from the pre-

ferred talker by selectively reinforcing those portions of the composite signal due to preferred speech sounds and suppressing those portions due to sounds from other sources.

Analysis of Equation 1 reveals, however, that the magnitude of the cross correlation of the two microphone signals is also dependent upon the average intensity of the sounds arriving at the microphones, as well as upon the degree of time coherence between the sounds. This is demonstrated, for example, by increasing the average sound intensity at each of the microphones 1, 2 by constant factors K_1 , K_2 , respectively; then from Equation 1, the magnitude of the cross-correlation function is increased by a factor $K_1 \cdot K_2$, that is,

$$\begin{aligned}\varphi(t) &= \overline{K_1 f_1(x) \cdot K_2 f_2(x+t)} \\ &= K_1 \cdot K_2 \overline{f_1(x) \cdot f_2(x+t)}\end{aligned}\quad (2)$$

It has been determined that the dependence of the cross-correlation signal upon the product of the sound intensities, an expression in Equation 2, produces in amplifier 5 a reinforcement of the composite signal in excess of the reinforcement required for improvement of intelligibility, the excess reinforcement causing the audible distortion of intensity previously discussed. The present invention eliminates this distortion at its source: the cross-correlation signal is normalized before applying it to the control terminal of amplifier 5 in order to remove its dependence upon the product of the average sound intensities at the two microphones. The dependence of the normalized cross-correlation signal upon the degree of time coherence between the microphone signals is preserved, however, and modulation of the composite signal in response to the normalized cross-correlation signal improves the intelligibility of speech sounds from a preferred source without distorting the intensity of the preferred speech sounds.

Complete System

Referring now to FIG. 1B, the normalizing apparatus of this invention is shown incorporated in the sound control system of FIG. 1A. The processed microphone signals from circuits 10, 11 are applied to normalizing circuit 13 as well as to cross-correlation circuit 12. Normalizing circuit 13, which is described in detail below, derives from the microphone signals a normalizing signal, N , which is a nonlinear function of the average magnitudes of the microphone signals. By applying the normalizing signal to the divisor terminal of divider 14, which may be of any suitable construction, and by applying the cross-correlation signal to the dividend terminal of divider 14, there is obtained at the output terminal of divider 14 a normalized cross-correlation signal that is primarily dependent upon the degree of time coherence between related portions of the microphone signals and that is relatively insensitive to the product of the average sound intensities at the microphones.

Preferred Normalizing Circuit

FIG. 2 illustrates a preferred embodiment of the principles of this invention, in which the microphone signals from signal processing circuits 10, 11 are applied to amplifiers 20, 22, respectively, which are constructed in a conventional manner to provide low impedance driving sources for peak detectors 21, 23. Peak detectors 21, 23 comprise, respectively, diodes 210, 230, and RC circuits consisting of capacitors 211, 231, and resistors 212, 232. The peak detectors have a charging time constant on the order of 100 microseconds and a discharge or holding time constant of about 100 milliseconds, in order to derive from each microphone signal a unidirectional or single polarity signal whose peaks coincide with the commencement of the periods of voiced sounds, and whose average amplitude is approximately equal to the average magnitude of each of the microphone signals, denoted

$$|f_1|, |f_2|$$

respectively, as shown in FIG. 2. Since the average magnitude of each of the microphone signals is proportional to the average sound intensity at each microphone, by multiplying together the output signals of peak detectors 21, 23 in multiplier 24, there is developed at the output terminal of multiplier 24 a normalizing signal, N, approximately proportional to the product of the average sound intensity at each microphone,

$$N = |f_1| \cdot |f_2| \quad (3)$$

By dividing the cross-correlation signal ϕ , by the normalizing signal N, as shown in divider 14 of FIG. 1B, there is obtained a normalized cross-correlation signal $\tilde{\phi}$ where

$$\tilde{\phi} = \frac{\phi}{N} \quad (4)$$

It is observed in Equation 4 that an increase in sound intensity at each microphone by factors K_1, K_2 , does not affect the magnitude of the normalized cross-correlation signal, since the factor $K_1 \cdot K_2$ appears in both the numerator and the denominator of Equation 4.

In order to prevent the denominator of Equation 4 from passing to zero at some time and thereby producing an erroneous normalized cross-correlation signal at the output terminal of divider 14, the normalizing signal is passed through threshold circuit 25. Threshold circuit 25 prevents the normalizing signal from falling below a predetermined nonzero value, and comprises a voltage divider network consisting of resistors 250, 251, with a diode 252 biased by an adjustable energy source 253 connected in parallel with resistor 251. Diode 252 and source 253 are shown in FIG. 2 connected for a negative polarity normalizing signal; however, for a positive polarity normalizing signal, the polarities of diode 252 and source 253 are reversed.

Alternative Normalizing Circuit

FIG. 3 illustrates alternative apparatus for deriving a normalizing signal, which is somewhat faster-acting than the apparatus shown in FIG. 2. The processed microphone signals from circuits 10, 11 of FIG. 1B are applied to amplifiers 30, 31 and then passed to tapped delay lines 32, 33, which are terminated in matching impedances 321, 331 to prevent reflection. The taps of each delay line 32, 33 are connected to a common output point P_0 , P_1 , through diodes 32a through 32n and 33a through 33n, respectively, and the signal formed at each of the common output points is the largest of all the signals appearing simultaneously at the taps of each delay line. By making the total delay of the delay lines equal to several pitch periods, that is, about 10 milliseconds delay with $n=15$ equally spaced taps, the largest value of each microphone signal over several pitch periods is obtained at each of the common output points.

From the largest signal obtained at each of the common output points there is derived a normalizing signal, either by multiplying together the two signals from the common output points or by squaring the larger of the two signals. The two largest signals formed at each common output point are passed through amplifiers 34, 35, and depending upon the positions in which switches 36, 37 are set, the two largest signals are applied either to the input terminals of diodes 38, 39 or to the input terminals of multiplier 40. Diodes 38, 39 are connected to a common output point P_2 and the signal developed at this point is the larger of the two largest signals from delay lines 32, 33. The magnitude of the larger of the two signals passed by diodes 38, 39 is squared in squaring circuit 41 to form a normalizing signal. When switches 36, 37 connect the two common output points P_0, P_1 of delay lines 32, 33 to the input terminals of multiplier 40, however, the product signal developed at the output terminal of multiplier 40 from the largest signal

at each common output point is the normalizing signal.

It is to be understood that the above-described arrangements are merely illustrative of applications of the principles of the invention. Numerous 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. In a system for the control of sound, the combination that comprises a source of a pair of information bearing signals derived from the output signals of a pair of spatially separated sound detecting devices, means for cross-correlating said pair of information bearing signals, means for normalizing the output signal of said cross-correlating means to eliminate its dependence upon the average amplitudes of said information bearing signals, and means under the control of said normalized cross-correlation signal for adjusting the amplitude of the sum of the output signals of said sound detecting devices.

2. In a sound control system, the combination that comprises a pair of spatially separated microphones, means for additively combining the output signals of said microphones, means for individually processing the output signals of said microphones to emphasize selected features of said signals, means for averaging the product of said processed microphone signals to develop a cross-correlation signal, means for deriving a normalizing signal from said processed microphone signals, means for dividing said cross-correlation signal by said normalizing signal to obtain a control signal, and means under the influence of said control signal for varying the amplitude of said additively combined microphone signals.

3. Apparatus as defined in claim 2 wherein said means for deriving a normalizing signal comprises two parallel subpaths each of which is provided with an input terminal, an output terminal, and an amplifier and a peak detector connected in series between said input and output terminals, means for applying each of said processed microphone signals to the input terminal of one of said subpaths, a multiplier provided with two input terminals and an output terminal, means for connecting the output terminal of each of said subpaths to one of the input terminals of said multiplier, and threshold means connected between the output terminal of said multiplier and said dividing means.

4. Apparatus as defined in claim 3 wherein said peak detector comprises a diode and an RC circuit.

5. Apparatus as defined in claim 3 wherein said threshold means comprises a voltage divider including a first resistor and a second resistor, a diode connected in series with an adjustable energy source, and means for connecting said series connected diode and energy source in parallel with said second resistor.

6. In combination with a cross-correlation sound control system, apparatus for reducing distortion which comprises a source of a pair of information bearing signals representative of the sound energy at two spatially separated points in a sound field, a source of a cross-correlation signal derived from said pair of information bearing signals whose magnitude is a function of the time coherence between said signals and of the average amplitudes of said signals, means for deriving from each of said information bearing signals a signal proportional to the largest value over several pitch periods of each of said information bearing signals, means for combining said largest value signals to form a normalizing signal, and means for dividing said cross-correlation signal by said normalizing signal.

7. Apparatus as defined in claim 6 wherein said means for deriving from each of said information bearing signals a signal proportional to the largest value over several pitch periods of each of said information bearing signals comprises two parallel signal paths, one for each of said information bearing signals, wherein each of said signal paths includes a first amplifying means provided with an input terminal and an output terminal, a delay

7

line provided with an input terminal and a plurality of taps, means for connecting the output terminal of said first amplifying means to the input terminal of said delay line, a plurality of diodes in one-to-one correspondence with the taps of said delay line, wherein each of said diodes is provided with an input terminal and an output terminal, means for connecting each tap of said delay line to the input terminal of one of said diodes, means for connecting the output terminals of said diodes to a common output point, a second amplifying means provided with an input terminal and an output terminal, means for connecting said common output point to the input terminal of said second amplifying means, and means for connecting the output terminal of said second amplifying means to said combining means.

8. Apparatus as defined in claim 6 wherein said means for combining said average magnitude signals to form a

8

normalizing signal comprises means for multiplying together said average magnitude signals.

9. Apparatus as defined in claim 6 wherein said means for combining said average magnitude signals to form a normalizing signal comprises two diodes, one for each of said average magnitude signals, wherein each of said diodes is provided with an input terminal and an output terminal, means for applying each average magnitude signal to the input terminal of one of said diodes, means for connecting the output terminals of said diodes to a common output point, squaring means provided with an input terminal and an output terminal, means for connecting said common output point to the input terminal of said squaring means, and means for connecting the output terminal of said squaring means to said dividing means.

No references cited.