

[54] **BANDWIDTH COMPRESSION SYSTEM
IN PHONETIC SOUND SPECTRUM**

[72] Inventor: Meguer V. Kalfaian, 962 Hyperion Ave.,
Los Angeles, Calif. 90029

[22] Filed: Dec. 14, 1970

[21] Appl. No.: 97,893

[52] U.S. Cl. 179/1 SA

[51] Int. Cl. G101 1/00

[58] Field of Search 179/1 SA, 15.55 R; 324/77

[56] **References Cited**

UNITED STATES PATENTS

3,431,356 3/1969 Copel.....179/1 SA

3,296,374 1/1967 Clapper 179/1 SA

Primary Examiner—Kathleen H. Claffy

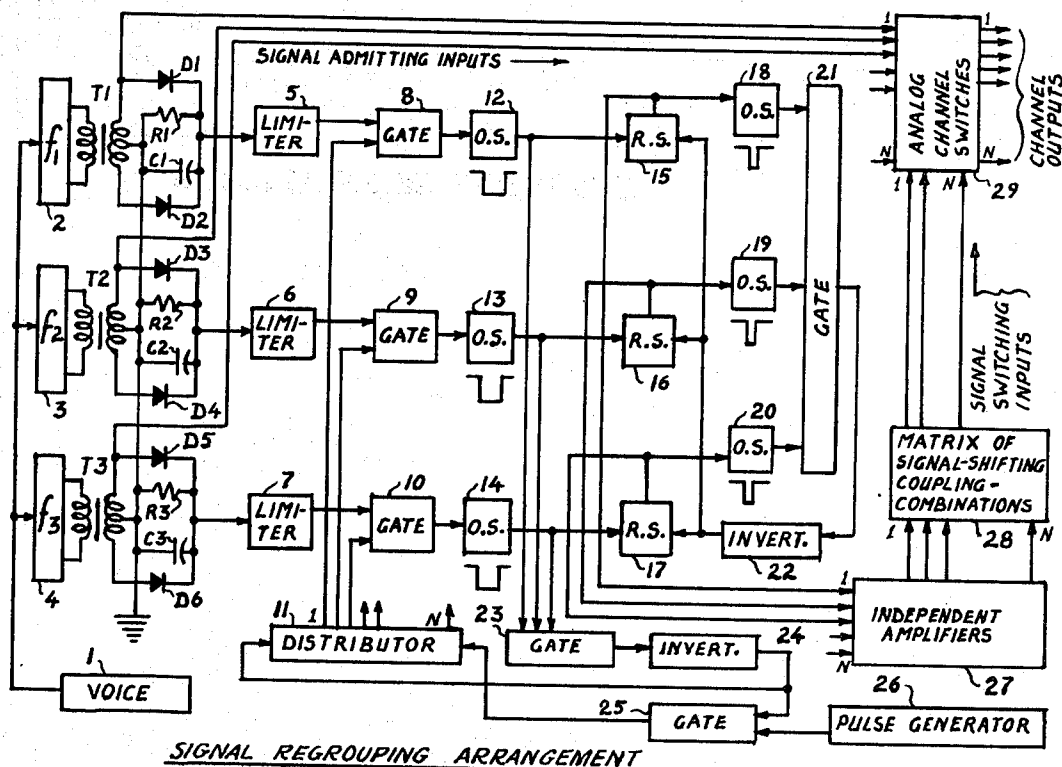
Assistant Examiner—Jon Bradford Leaheey

[57]

ABSTRACT

In sub-band divisions of speech sound waves the detected signals derived from the sub-bands are continually regrouped at the outputs of a numerically arranged bank of channels in an order that, the signal derived from the lowest frequency (pitch) in the sub-bands is continually shifted to the output of the first channel, while the other signals are shifted to the numerically positioned channel outputs by the same factors of multiplication from the first channel as the other sub-bands differ from said lowest frequency.

1 Claim, 2 Drawing Figures



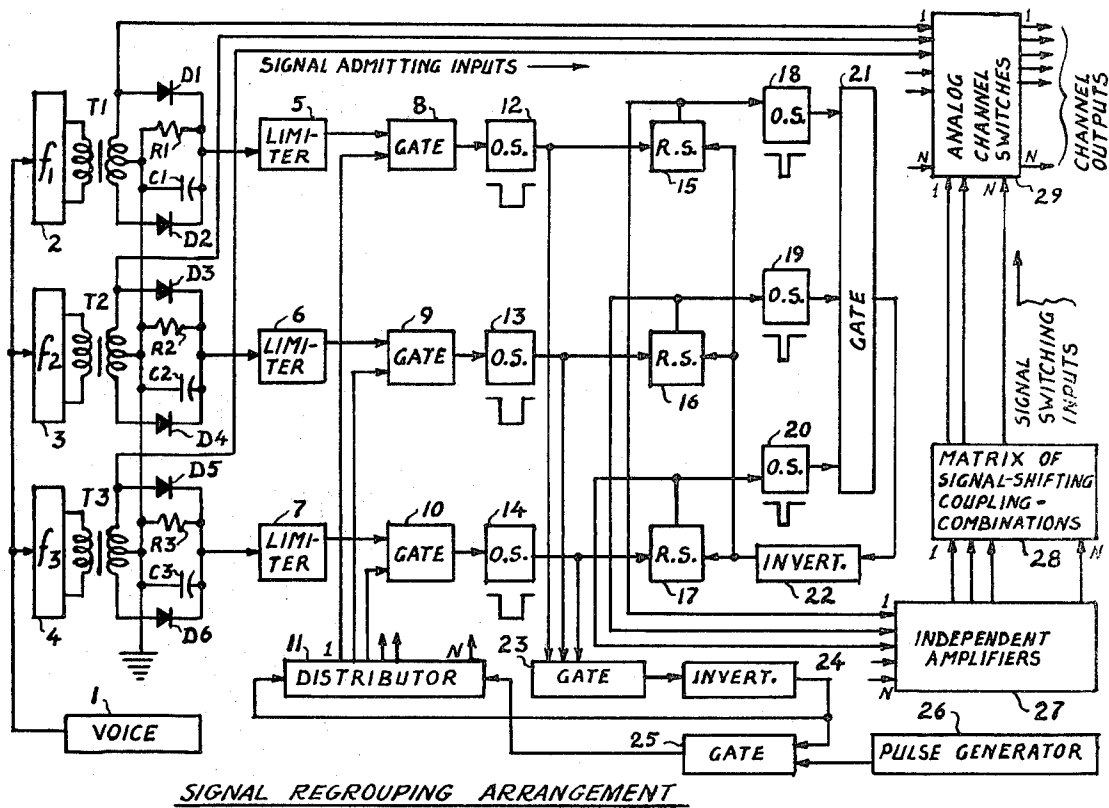
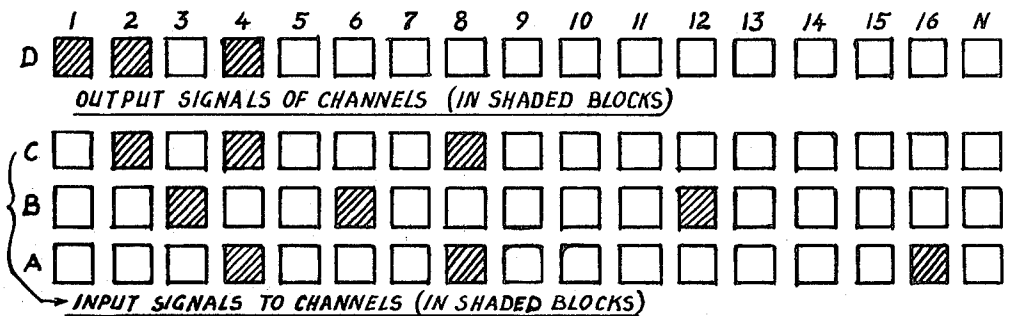


Fig. 2



THE ILLUSTRATION SHOWS HOW A GROUP OF INPUT SIGNALS IN VARYING NUMERAL REGIONS ALONG A BANK OF NUMERICALLY ARRANGED CHANNELS AT A-B-C ARE SHIFTED AND COMPRESSED TO A REFERENCE REGION AT D, AT THE OUTPUTS OF THE CHANNELS

INVENTOR

Fig. 1

Alvin V. Kalfay

BANDWIDTH COMPRESSION SYSTEM IN PHONETIC SOUND SPECTRUM

This invention relates to speech sound wave analysis, and more particularly to an arrangement for normalizing the spectrum variations that occur in the sound wave, so as to facilitate analytical simplicity and accuracy of extracting the different informations that the speech sound wave contains.

The invention is contemplated to be particularly useful in the vocoder systems for first narrowing the widely varying sub-bands of the speech sound waves into a fixed number of numerically reduced channels, so that the narrow band control signals, which are usually derived from the sub-bands, as means for frequency compressed transmission, can be derived more simply and accurately than by the systems heretofore employed. The arrangement may also be used for selection of the pitch frequency from the speech sound wave, wherein the widely varying regions of the sub-bands of the sound may continually be shifted to a fixed narrow region, in an order that the lowest sub-band, which represents the pitch frequency, is always shifted to a fixed channel, as means for automatically selecting the pitch frequency from that fixed channel. In another practice that the arrangement may be used advantageously is the voice-print systems, wherein it may be desired that the voice pattern is narrowed to a fixed visual location (without destroying the original characteristics of the voice information), for facilitating greater ease and accuracy in pointing out to a specific speaker that the voice may be assumed to have belonged to. Thus the particular arrangement that will be described in the following specification is contemplated to be useful in a scope that embrace various other forms of systems and arrangements which the invention is capable of assuming in practice.

The art of making speech recognition automata has for some time been growing to enter into a complex structure of practices, which may gradually assume as an integral part of the present day computer systems. Devising an automata for speech recognition, however, has at present been a most difficult task for the scientist, because of the enormously wide variations that occur in the sound waves. Thus in order to facilitate these and other attempts into a practical realization, I contemplate to provide an arrangement, as disclosed herein, which is capable of narrowing down those variables that are most difficult to control.

Other desirable aspects of the invention will also become apparent in the following specification, in conjunction with the accompanying drawings, wherein:

FIG. 1 is a graph showing how the widely varying frequency regions of the speech sound waves are shifted to a fixed narrow region at the outputs of a numerically arranged bank of channels. The graph shows at A, B and C three groups of signals having the same mutually related distribution ratio along the sound spectrum, but distributed along different regions in the sound spectrum. When any one of these groups of signals are applied to the inputs of the bank of channels, the outputs will have the same narrow band region, as shown at D. And FIG. 2 shows a switching arrangement for continually hunting, by way of a signal-distributing generator, and connecting the signals derived from various frequency components of the sound to different channels in prearranged coded combinations, so as to obtain the bandwidth narrowing condition as shown in FIG. 1.

Frequency normalization has been proposed previously under various titles, such as, signal conversion, frequency conversion, spectrum normalization, and frequency standardization. The problem of frequency normalization has been a critical one, been favored however, and for that reason the practice has not been favored in experimental systems for speech recognition automata. In carrying out an exemplary method of spectrum normalization, the group of resonances in the arriving wave are regrouped in a reference frequency region in the sound spectrum in an order that, the lowest frequency (pitch) in the group is converted to a reference pitch frequency, and the other frequencies are converted to frequencies by the same factors of multiplication from the reference frequency as

they differ from the original pitch frequency. Since in transmitting techniques of information, as known and practiced, one set of parameters may be substituted for another without loss of definition as long as the independent parameters remain unchanged, it is readily seen that the converted frequencies are relocated in fixed (standard) positions in the sound spectrum, and therefore, easily adapted to any type of analytical processing that may be desired. For practical purposes, however, the method just mentioned involves critical adjustments. For example, assuming that the various resonances in the sound wave are harmonically related to the lowest (pitch) frequency, the values may be shown, as $F_1 \div f_0$, $F_2 \div 2f_0$, $F_3 \div 3f_0 \dots F_n \div nf_0$, where (F_0) is the varying components and ($F_1 \div f_0$) is the fixed reference pitch frequency component. Obviously, the artificial generation of the (F) components $2f_0$, $F_3 \div 3f_0 \dots F_n \div nf_0$, where (f_0) is the varying components; and ($F_1 \div f_0$) is the fixed shown product values would involve critical and undesired control systems. But as stated above, one method of signal conversion may be substituted by another without changing the specific information to be analyzed. Thus, instead of changing the varying frequencies in the sound wave into fixed frequencies, we may first derive detected signals from the pass-band filters (for subdividing the sound into sub-bands), and regroup them in an arrangement of numerical order, such as, 1, 2, 3, ... n , which by simulation may assume the values as, $1 = F_1 \div f_0$, $2 = F_2 \div 2f_0$, $3 = F_3 \div 3f_0$, and $n = F_n \div nf_0$, where (1) represents the fixed reference numeral (fixed reference pitch frequency). By such numerical conversion (or digital conversion, as far as frequency components are concerned, without changing the amplitude components), we may now deal with on-and-off conditions which can be established in the highest order of control and stability with the present day digital techniques. Accordingly, the novel switching system used herein may be briefly described, as in the following:

There are used a bank of channels arranged in numerical order, 0 such as, 1, 2, 3, ... n , as described above, each one of which is provided with a plurality of signal-admitting inputs, and a plurality of signal-switching inputs, respectively. The detected signals derived from various resonances (sub-band divisions in the sound) are applied to the plurality of signal-admitting inputs, so that anyone of the signals can be admitted to the output of anyone of the channels by the operation of a respective signal-switching input. Thus in order to obtain normalization of the frequency variations, a plurality of prearranged combinations of groups of switching signals are applied to the plurality of signal-switching inputs for regrouping the detected signals admitted to the outputs of the bank of channels sequentially until a specific group of the detected signals are regrouped along a reference standard region of the numerically arranged channel outputs. This may be explained graphically in FIG. 1, as in the following:

In FIG. 1 there are shown three different groups of signals at the inputs of the signal-regrouping channels at A, B, and C, which are regrouped at the reference outputs of the channels, as at D. For example, and for the purpose of deriving only the mutually related frequency ratios of the group of resonances in a complex sound wave, as passed through a plurality of sub-band filters, we may first derive detected signals from the outputs of these filters, and instead of analyzing the original frequency ratios, we may now analyze the numerical ratios of the numerical locations of the filters (sequenced in a numerical order) in which the signals are derived from. Thus assuming that a group of detected signals are applied to the A inputs (shaded blocks) of the numerically arranged channels, it is seen that the second signal is in a numerical location (along the numerically sequenced order) as the second numerical harmonic of the first, and the third signal is in a numerical location as the fourth numerical harmonic of the first signal. Assuming now that the same group of signals are distributed along the channel inputs, at B, the same mutually related numerical ratios are observed—this also relating to the signals shown at C. by such examples, all that is required to stan-

standardize these numerical ratios is to switch these input signals to the numerically arranged outputs of the channels, as at D, wherein, the first signal is switched to the channel-1, the second input signal is switched to the channel-2 output, and the third signal is switched to the channel-4 output—the numerical ratios still being as, second and fourth harmonics of the first numeral, for simplicity of selection and analysis. At this point, however, the problem remains as to how to determine what combination of switching that is required for each of the group of signals at A, B or C, in order to obtain the reference signal regrouping at D. This is done simply by a prearranged matrix of a plurality of switching combinations, which are applied to the bank of channels sequentially until one of the combinations establishes the required switching at D.

In reversing the process just mentioned, we could also record numerically distinguishable signals derived from those regrouped switching combinations that were responsible for narrowing down the original group of signals, to regroup again into the original wide-band signals by reverse switching combinations. This is particularly useful in vocoder systems, wherein the signals in shaded blocks at D could be converted first into narrow bandwidth control signals for narrow bandwidth transmission to a receiving end, including auxiliary control signals representing the numerical positions of the switching combinations, and regroup the received group of signals in reverse switching combinations under control by said auxiliary signals, in order to translate the narrow band transmitted signals into the original wide band signals for natural speech reproduction. Needless to say, of course, that the grouping or regrouping of the signals may be in any secret coding form that may be desired. Since the regrouped signals at D represent any one of the groups of signals at A, B and C, the process of signal regrouping will also be useful in voice-print systems for simpler and more accurate analysis of the voice.

With the above description on the usefulness and purposes of the present invention, the actual arrangement will now be described, as in the following:

In FIG. 2 the voice sound wave in block 1 is applied to a number of pass-band filters, of which only three filters are shown in blocks 2, 3 and 4. The number of pass-band filters for dividing the voice sound wave into sub-bands is a matter of choice, but for the purpose of speech sound wave analysis, reference may be made to my related disclosures in my patent applications Ser. No. 828,067 filed Apr. 29, 1969, now U.S. Pat. No. 3,622,706, Nov. 23, 1971, and Ser. No. 26,623 Filed Apr. 8, 1970, wherein I have shown complete numerical charts on how different combinations of channel switchings may be arranged. The outputs of blocks 2, 3 and 4 are applied to transformers T1, T2 and T3, respectively, and the voltages developed across the secondaries of these transformers are detected by the full-wave rectifying diodes D1 through D6; resistors R1 through R3; and wave smoothing capacitors C1 through C3, respectively. The outputs of these detected waves are applied to amplitude-limiter circuits in blocks 5 through 7, respectively, and the outputs of these limiters are applied in "1" level to one of the inputs of gates in blocks 8 through 10, respectively. Voltage-amplitude limiting circuits (signal-amplitude-threshold level) are known and often used in the art of electronics, and the purpose of using these limiters in blocks 5 through 7 is to make sure that all sub-band waves above a threshold amplitude level of the original sound wave are converted into constant amplitude-detected signals and applied to the inputs of gates 8 through 10 in "1" levels. The other inputs of gates 8 through 10 are sequentially driven into "1" levels (starting gate 8) by the distributor in block 11, so that any one of the gates 8 through 10 receiving a distributor (hunting) signal at "1" level, has also received an amplitude-limited signal at "1" level, will produce at its output a signal at "0" level and apply to an associate one-shot circuit. For example, if the speaker has the lowest pitched voice, and the pass-band filter in block 2 is tuned to the low pitch frequency of that vo-

ice, the input of gate 8 will have received a "1" level signal from the limiter in block 5 coincidentally with the distribution signal from block 11 starting from the input of gate 8. Thus the output of gate 8 operates the one-shot in block 12. Assuming however, that the lowest pitch frequency appears in the sub-band of block 3, then the input of gate in block 9 would have received a "1" level signal from the limiter in block 6. In this case, when the distributor in block 11 starts signal distribution (hunting) from the input of gate 8, the gate 8 does not operate because its other input is at "0" level. However, as the distribution continues to the input of gate 9, the gate 9 operates, which in turn operates the one-shot in block 13. This operation continues in similar fashion to any one of the gates 8 through 10 (and the *n*th gate not shown in the drawing), depending on which one of the gates operates first in the sequential signal distribution from block 11.

The output pulses of one-shots in blocks 12 through 14 are applied to the set inputs of the set-reset flip-flops in blocks 15 through 17, respectively. The set outputs of the flip-flops 15 through 17 are applied to the one-shots in blocks 18 through 20, respectively, for operation of these one-shots. The output pulses of one-shots 18 through 20 are in turn applied in "0" levels to the multi-inputs of the gate in block 21. Finally, the output of gate 21 is phase inverted in block 22 and fed back in parallel to the reset inputs of the flip-flops 15 through 17. Thus assume that the one-shot 12 is operated and it further operates the set-reset flip-flop in block 15 into set operation. This set operation of flip-flop 15 operates the one-shot 18, which in turn operates the gate in block 21, and finally the output of gate 21 operates all of the flip-flops (after passing through the phase inverter in block 22) into reset states. As will be seen by the pulse waveforms drawn under the one-shot blocks 12 through 14, and under the one-shot blocks 18 through 20, the output pulses of the former one-shots are preadjusted to be longer than the output pulses of the latter mentioned one-shots. Thus, after all of the set-reset flip-flops in blocks 15 through 17 are driven into reset operating states by feed-back from the gate 21, the output pulse of one-shot 12 remains long enough to hold the flip-flop in block 15 in set operating state. By such self-locking operation, it is seen that each time a distributor signal (pulse) is applied to one of the inputs of any one of the gates 12 through 14 coincidentally with a signal from its associated limiter in blocks 5 through 7, the flip-flops 15 through 17 are driven into reset operating states, except the one that is associated with the gate (gates 8 to 10) last operated.

In order to provide continual hunting by the distributor in block 11, the outputs of one-shots in blocks 12 through 14 are applied to the multi-input gate in block 23, the output of which is further phase inverted in block 24, and applied to the distributor in block 11 for reset operation. The phase inverted output of block 24 is also applied in "0" level to one of the inputs of gate in block 25 for preventing the output pulses of pulse generator in block 26 passing to the set operating input of the distributor in block 11. Thus the distributor in block 11 sets into operation for hunting the lowest existing frequency in the sub-bands, and re-starts immediately again for continual hunting.

For final switching of the channels for signal regrouping, such as shown in FIG. 1, the outputs of set-reset flip-flops in blocks 18 through 20 are independently amplified in block 27, and applied to a coding matrix in block 28 which comprises a plurality of prearranged combinations of direct couplings to a plurality of analog switches, representing the plurality of channels in block 29. These are the bank of channels, at the outputs of which the sub-band signals are shifted, as described by way of the illustration in FIG. 1. The arrangement of the analog switches, connecting to a bank of channels, is shown in my related patent applications, as mentioned in the foregoing, and the first one of which is not patent issued. In these applications I have also shown a sub-band frequency chart using 30 and 32 sub-band filters for dividing the phonetic sound spectrum, but for the instant disclosure it will be preferable to use

38 sub-band filters to cover the lowest pitch frequency of human voice. The analog switching matrix in the above mentioned patent applications had been shown with capacitive couplings for the purpose of reducing the number of commercially available integrated analog switches. But with the presently available multi-element analog switches, with reduced prices, it is more economical to eliminate the capacitive couplings and use direct couplings, which will also provide simplicity of manufacturing the device shown herein. Similarly, the block 27 comprising independent amplifiers is included with the assumption that available integrated circuits use 5 volts supply voltage, whereas, the analog switches of the MOS field effect transistor type use -10 volts to be applied upon their gate electrodes for on-and-off operating states.

Thus while specific embodiments of the invention have been selected to describe the invention, it is obvious that various modifications, adaptations, and substitutions of parts may be made without departing from the true spirit and scope of the invention. For example, it may be desired that the outputs of block 29 are coupled to a utilization apparatus for analysis of the regrouped signals at the channel outputs. Then by such analysis, a pulse signal may be produced to control the distributor in block 11 into reset operation; instead of receiving such control from the output of block 24. In such operation, the distributor 11 would have to be allowed to continue its distribution until said utilization apparatus (not shown) produces said control pulse. In another example, the inputs of block 29 may be received from the detected terminals of the signals in T1 through T3, instead of the undetected terminals, as shown. In furtherance, the outputs of the flip-flops 15 through 17 may be used as control signals in the vocoder.

What I claim, is:

1. Signal conversion system of complex combinations of groups of signals distributed along a plurality of channels comprising a plurality of coupling means from the signals of said channels to a matrix of signal-regrouping combinations, each combination having been prearranged for a specific signal

regrouping under control of a control signal derived from a signal appearing at a specific location along said channels; a plurality of coupling means from the signals of said channels to a plurality of signal-amplitude-threshold sensing means, for deriving sensing signals; a plurality of AND gates; coupling means from said sensing signals to the first inputs of said AND gates; a sequentially operated pulse distributor having plurality of outputs, and a reset control input; coupling means from the distributor outputs to the second inputs of said AND gates, whereby only those AND gates which have received sensing signals at their first inputs operate by the distributor pulses; coupling means from the AND gates to a first plurality of one-shots; a plurality of set-reset flip-flops having first and second inputs for set and reset operating states, and said set operating states representing the control signals of said signals appearing along the channels; coupling means from said first plurality of one-shots to the set inputs of said flip-flops; a second plurality of one-shots operating in shorter pulse periods than said first plurality of one-shots; coupling means from the set operating states of said flip-flops to the second plurality of one-shots, coupling means from the set operating states of said plurality of flip-flops to said matrix for operation of specific signal-regrouping combinations; first mixer means for mixing the pulses of said first plurality of one-shots, and means for applying these mixed pulses to the reset control input of said distributor, whereby to effect reset of the distributor each time signal regrouping is established; second mixer means for mixing the pulses of the second plurality of one-shots, and means for applying these mixed pulses in parallel to the reset inputs of said flip-flops for reset operation except the flip-flop that has last operated in set state due to reason of its longer pulse duration than the pulse duration of the last operated one-shot of the second plurality of one-shots, and thereby holding said signal regrouping undisturbed until the following operation of signal regrouping is recycled.

* * * * *

40

45

50

55

60

65

70

75