

Jan. 11, 1955

M. J. DI TORO ET AL

2,699,464

FUNDAMENTAL PITCH DETECTOR SYSTEM

Filed May 22, 1952

3 Sheets-Sheet 1

Fig. 1

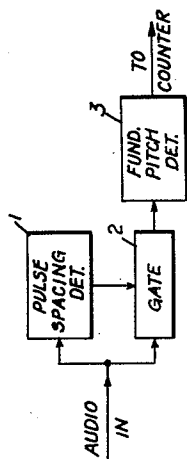
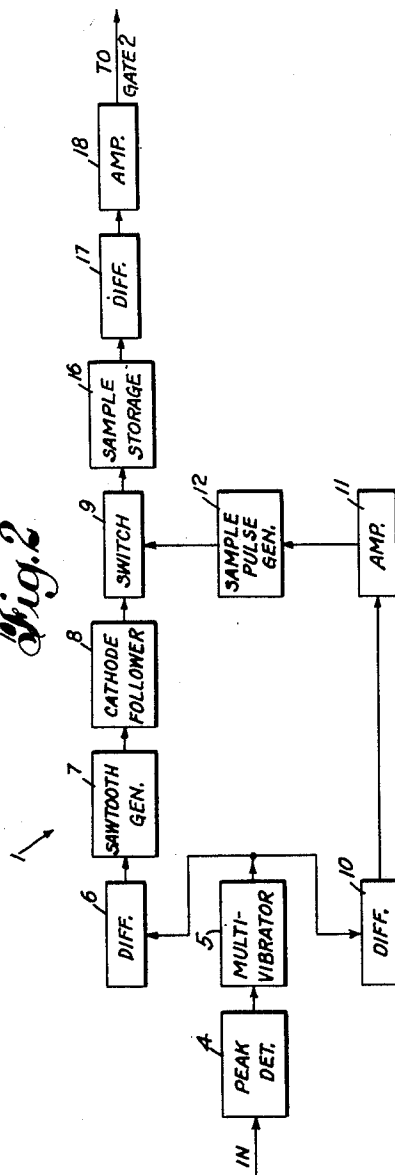


Fig. 2



INVENTORS
 MICHAEL J. DI TORO
 WALTON GRAHAM
 BERNARD M. DWORK
 BY *Ray P. Lantry*
 ATTORNEY

Jan. 11, 1955

M. J. DI TORO ET AL

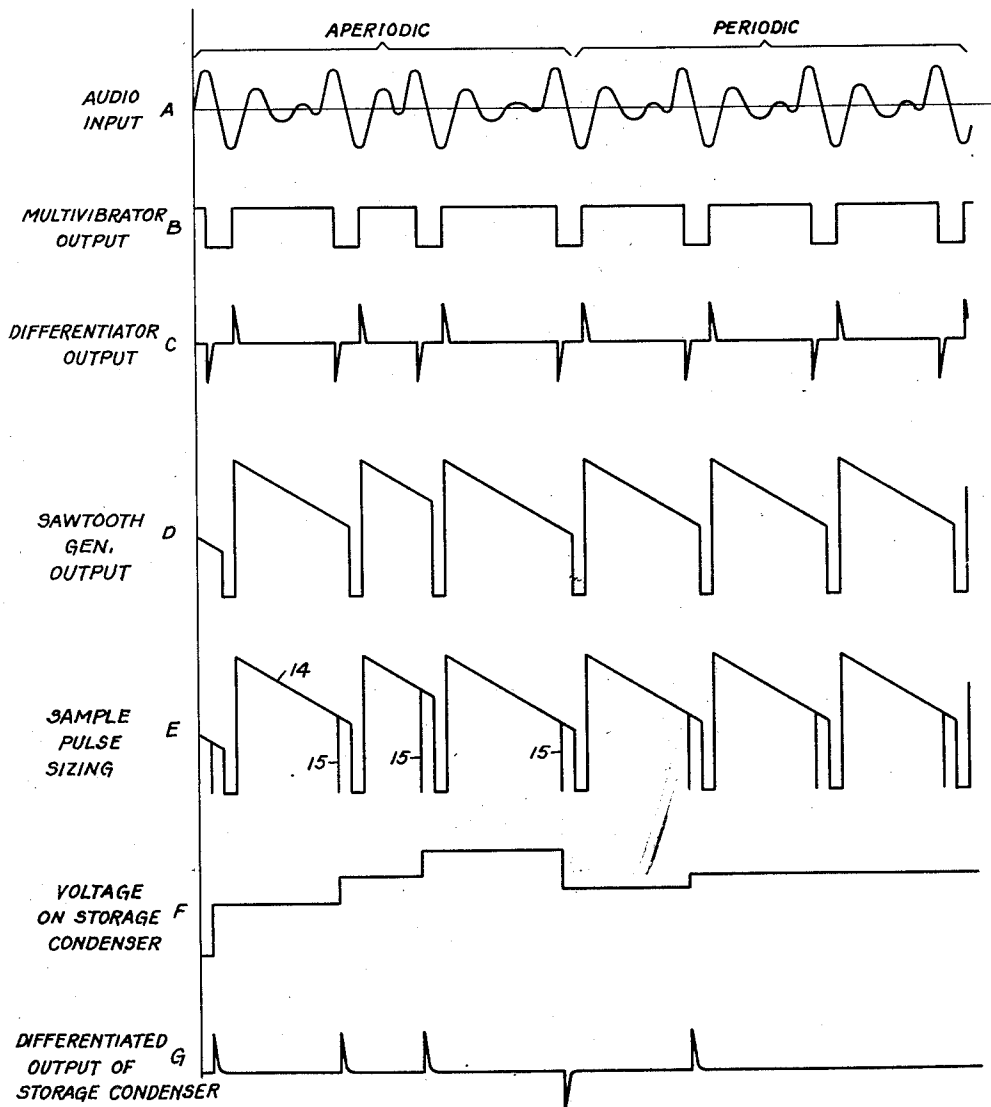
2,699,464

FUNDAMENTAL PITCH DETECTOR SYSTEM

Filed May 22, 1952

3 Sheets-Sheet 2

Fig. 3



INVENTORS
MICHAEL J. DI TORO
WALTON GRAHAM
BERNARD M. BWORK
BY *Russ P. Landry*
ATTORNEY

Jan. 11, 1955

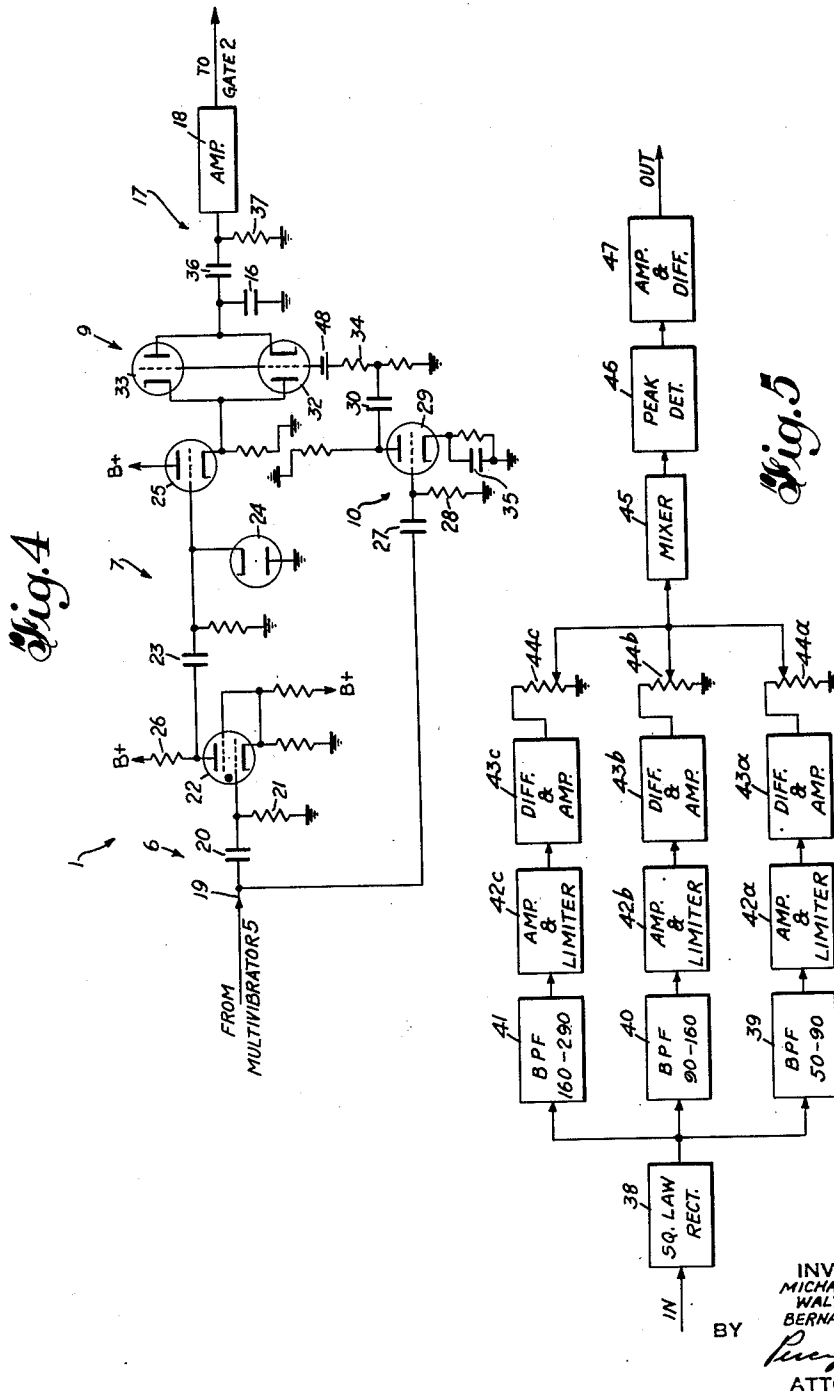
M. J. DI TORO ET AL

2,699,464

FUNDAMENTAL PITCH DETECTOR SYSTEM

Filed May 22, 1952

3 Sheets-Sheet 3



INVENTORS
MICHAEL J. DI TORO
WALTON GRAHAM
BERNARD M. DWORK
BY
Rory P. Lantry
ATTORNEY

1

2,699,464

FUNDAMENTAL PITCH DETECTOR SYSTEM

Michael J. Di Toro, Bloomfield, N. J., and Walton Graham, New York, and Bernard M. Dwork, Bronx, N. Y., assignors to International Telephone and Telegraph Corporation, a corporation of Maryland

Application May 22, 1952, Serial No. 289,345

11 Claims. (Cl. 179—1)

This invention relates to fundamental pitch detector systems and more particularly to electrical means by which the fundamental pitch of the voiced sounds of a person's speech may be instantaneously and continually derived.

Human speech is composed of time sequence voiced (quasi-periodic) sounds and of unvoiced (aperiodic) sounds. Periodic or voiced sounds are those made by using the vocal cords. In normal speech the voiced sounds have an envelope waveform which is of a recurrent nature. The major distinguishing characteristic between periodic and aperiodic sounds is the recurrent nature of the former and the non-recurrent nature of the latter.

In some forms of speech frequency compression systems, it is necessary to determine the fundamental pitch of the voiced sounds. The problem of determining for speech the periodicity in time (or its reciprocal, the fundamental pitch frequency) requires a device which will determine if the sound is periodic or aperiodic and will then detect the periodicity of the voiced sounds.

In some forms of transmission systems for the transmission of voice currents, such as a telephone system utilizing a military type carbon microphone, the frequency response of the system in the lower portion of the voice spectrum is not sufficient to pass the fundamental frequency itself. When the fundamental frequency (together with some of the higher harmonics) is missing from a voiced sound, the wave shape of the sound envelope is still periodic and of a recurrent nature even though the waveform is changed. Thus, even after a sound passes through a transmission system which changes its spectrum profile and eliminates the fundamental frequency together with some higher harmonics, the essential characteristic that distinguishes a voiced sound from an unvoiced sound is still present in the received waveform. To determine the fundamental pitch of a voiced sound whose fundamental frequency is missing from the received voice currents, electronic means are required which distinguish between aperiodic and periodic waveforms, independent of the amplitude and wave shape, and determines the fundamental pitch of the voiced sounds independently of what the voiced sound is, or how loud it may be, or how it may be altered in the wave shape passing through mutilating transducers.

One of the objects of this invention, therefore, is to provide electrical means by which the fundamental pitch of voiced sounds of human speech may be instantly and continually derived.

Another object of this invention is to provide electrical means by which the missing fundamental pitch of voiced sounds may be derived from the received voice currents of an electrical transmission system.

A further object of this invention is to provide means to obtain the fundamental pitch of the voiced sounds by detecting the output of a mixer fed by a plurality of frequency adjacent filters.

Still a further object of this invention is to provide means for separating periodic from aperiodic wave shapes by detecting a change in spacing between recurrent pulses each representative of a relatively large amplitude peak in an input waveform.

A feature of this invention is the use of a pulse spacing detector of novel design whose output is dependent upon the change in spacing between adjacent peaks of the input audio signal. Voiced sounds, having a periodic characteristic, of the input audio signal are fed to a fundamental pitch detector by means of a gating circuit which is responsive to the output of the pulse spacing detector. The input

2

signal to the fundamental pitch detector is passed through a plurality of frequency adjacent filters each approximately one octave wide and the fundamental frequency is coupled to a mixer. The output of the fundamental pitch detector comprises a series of pulses at the fundamental frequency and may be fed to a usual counter circuit or utilized in other ways as desired.

The above-mentioned and other features and objects of this invention will become more apparent by reference to the following description taken in conjunction with the accompanying drawings, in which:

Fig. 1 shows a block diagram of a system for determining the fundamental pitch of the voiced sounds present in an audio input wave form;

Fig. 2 shows a block diagram of a pulse spacing detector for use in the system of Fig. 1;

Fig. 3 shows a graphic illustration of a set of curves helpful in the illustration of a pulse spacing detector of Fig. 2;

Fig. 4 is a schematic circuit diagram of the pulse spacing detector of Fig. 2; and

Fig. 5 shows a block diagram of a fundamental pitch detector for use with the system of Fig. 1.

Referring to Fig. 1, a system for deriving the fundamental pitch of voiced sounds present in a person's speech, in accordance with the principles of this invention, is shown, comprising a pulse spacing detector 1, which determines from the periodic or aperiodic nature of the instantaneous audio input whether the sound is voiced or unvoiced and if the sound is voiced, the output of the pulse spacing detector 1 controls gating circuit 2 in such a manner that the voiced sound of the audio signal is coupled to the fundamental pitch detector 3 which determines the fundamental pitch of the input voiced sound and feeds its output, consisting of a series of pulses at the fundamental frequency rate, to a counter circuit or other utilizing equipment.

Referring to Figs. 2 and 3, the block diagram of a pulse spacing detector, and some waveforms created therein, are shown for use in explanation of the system of Fig. 1 for determining whether an input audio signal, which may have its fundamental frequency missing due to the characteristics of the transmission system, is voiced or unvoiced. Assume for purposes of explanation that the instantaneous input audio signal is a voiced or periodic sound as shown in Fig. 3, curve A. The input audio signal is coupled through a peak detector 4 to a multivibrator 5. The output of the multivibrator 5, as shown in curve B, comprises a series of pulses of constant width but of variable spacing, responsive to the output of the peak detector 4, which is indicative of the large peak amplitude portions of the original audio input waveform. A series of pips indicative of the duration of the pulses from multivibrator 5, as shown in curve C, are formed in differentiating network 6. The output of the differentiating network 6 is fed to a sawtooth generator 7 whose sawtooth shaped output is responsive to the series of pips from differentiator network 6, as shown in curve D. The sawtooth wave from generator 7 is fed through a cathode follower circuit 8 to a switch 9.

The output of multivibrator 5 is also fed to a second differentiator network 10, the output of which is coupled to an amplifier 11. Amplifier 11 is so biased that only the negative pips, corresponding in time to the leading edges of the pulses from multivibrator 5 are amplified and coupled to a sample pulse generator 12.

Thus, as shown in curve E, the signal waveforms present in the sample pulse generator 12 are the sawtooth wave 14 from generator 7 and pulses 15 from amplifier 11.

The output of sample pulse generator 12 comprises a series of pulses corresponding in time to the leading edges of the pulses from the output of multivibrator 5 and of an amplitude dependent upon the amplitude of the sawtooth wave from generator 7 at that instant. This series of pulses is fed to switch 9 which as heretofore explained is also coupled to the output of the sawtooth generator 7. The sawtooth wave from generator 7 is gated by switch 9 responsive to the pulse output of generator 12 and the output from switch 9 is coupled to a sample storage condenser 16. The voltage stored on condenser 16, as shown in curve F, is of an amplitude dependent upon the gating

3

of the sawtooth wave in switch 9. The voltage change across condenser 16 is differentiated by network 17 whose output, as shown in curve G, comprises a series of pips having an amplitude and polarity dependent upon the time interval between changes of voltage stored on condenser 16 and the amplitude change between successive voltages stored on condenser 16.

When the input audio signal comprises voiced (periodic) sounds, there is no difference in the time spacing of the pulses from sample pulse generator 12 and thus no change in voltage stored on the sample storage condenser 16, but for an unvoiced (aperiodic) sound input there exists a continual change in adjacent sample pulse spacings resulting in a large output from differentiating network 17, which is amplified in circuit 18 to control gate 2 of Fig. 1 to prevent unvoiced sounds from being coupled to the fundamental pitch detector 3.

Referring to Fig. 4, the circuit diagram in schematic form of a portion of the pulse spacing detector of Fig. 2 is shown. The output of the multivibrator 5 is coupled via line 19 to the differentiating network 6 comprising a condenser 20 and resistor 21. Differentiated output of the multivibrator 5 is fed to one grid of a gas-filled discharge device 22 which comprises part of the sawtooth generator 7.

Gas tube 22 is generally non-conducting and is not affected by the negative pips, curve C, produced by the differentiation of the leading edges of the pulses, curve B, from the output of multivibrator 5. However, the positive pips from the output of the differentiating network, produced from the trailing edges of the output pulses from multivibrator 5, fire the gas tube 22 and thereby remove any charge present on condenser 23 coupling the plate of tube 22 to the rest of the circuit. Diode 24, connected between ground and the far side of condenser 23, enables rapid discharge of condenser 23 when tube 22 fires. A short time after gas tube 22 fires it is extinguished due to resistor 26, connecting the plate of tube 22 to a positive voltage source. Resistor 26 is large and produces a large voltage drop which lowers the plate voltage below the firing level of the gas tube 22. When the gas tube 22 is extinguished current begins to build up a charge on condenser 23 altering the bias present on the grid of tube 25. This operating cycle produces the sawtooth wave shown in Fig. 3, curve D.

The pulses from multivibrator 5 are also fed to a differentiating network 10, comprising a condenser 27 and resistor 28. The output of this network is coupled to the grid of an amplifier tube 29 which is operated near its saturation point so that only the negative pips from the differentiating network, corresponding to the leading edges of the pulses from multivibrator 5, are amplified. The amplified negative pips from the plate of tube 29 are coupled through condenser 30 to the switch 9 along with the sawtooth wave from tube 25. Switch 9 comprises two triodes 32 and 33 connected in parallel, back to back. The output of tube 29 is coupled to the grids of both triodes 32 and 33 through a high resistance 34. The plate of tube 33 and the cathode of tube 32 are coupled to a sample storage condenser 16. If condenser 16 is uncharged the pulse from tube 29 causes tube 32 to conduct and charge the storage condenser 16 to the value of the sawtooth wave from tube 25 at the instant of sampling. When the sampling pulse from tube 29 stops, tube 32 ceases to conduct due to bias from battery 48. During this time tube 33 remains inoperative because it is of opposite polarity from tube 32 and the pulse from tube 29 so biases tube 33 that it is prevented from conducting. Thus, when tube 32 ceases to conduct, the storage condenser 35 is open-ended and maintains the charge impressed upon it when tube 32 was conducting.

If the input audio wave to peak detector 4 is periodic then the output of multivibrator 5 will be periodic and the amplitude of the sawtooth wave from generator 7 will be equal in value each time a sampling pulse from tube 29 causes tube 32 to conduct. Under such conditions, the next time tube 32 conducts it will couple the instantaneous voltage of the sawtooth wave from tube 25 to an equal voltage stored on condenser 16 causing no change in the voltage across the condenser 16. However, if, due to an aperiodic characteristic of the input audio signal, the next sampling pulse should come after a smaller interval of time than the preceding interval then the voltage of the sawtooth wave will be greater than the voltage stored on condenser 16. When tube 32 conducts it will cause con-

4

denser 16 to be charged to the new voltage level causing a change in voltage across condenser 16. If the next sampling pulse occurs after a longer interval of time than the preceding one then the instantaneous voltage of the sawtooth wave at the time of sampling will be less than the voltage stored on condenser 16 due to the preceding sampling pulse. Since the voltage stored on condenser 16 will be greater than the voltage of the sawtooth wave at the instant of sampling, tube 33 will conduct, discharging condenser 16 until it is of equal potential with the instantaneous value of the sawtooth wave from tube 25.

The voltage change across condenser 16 is differentiated at 17 by condenser 36 and resistor 37. The differentiator network output comprises a series of pips whose polarity is plus, minus, or zero depending upon whether the time interval between successive pulses is shorter, longer, or equal to the preceding interval and the amplitude of the pips is proportional to the change in spacing between successive time intervals. This output, which is actually a measure of the interval of time between successive pulses, is amplified and controls gate 2 of Fig. 1 in such a manner that only voiced (periodic) sounds present in the original speech input are coupled to the fundamental pitch detector 3.

Referring to Fig. 5, a block diagram of a fundamental pitch detector for use with the circuit of Fig. 1 is illustrated wherein only the voiced sounds present in the original speech input as determined by the pulse spacing detector heretofore explained are coupled to the input of the fundamental pitch detector. The input signal from gate 2 is coupled to a square law rectifier 38 (such as a crystal diode) which distorts the input waveform and assures the presence of a fundamental frequency component. The output of the square law network 38 is coupled to three frequency adjacent band pass filters 39, 40, and 41 each being slightly less than one octave wide. Filter 39 passes frequencies between 50 and 90 C. P. S., filter 40 passes frequencies between 90 and 160 C. P. S., and filter 41 passes between 160 and 290 C. P. S. The band width of each filter is such that the fundamental frequency of the input signal, restored by square law network 38 is passed by only one filter. It is possible that higher harmonics of a fundamental frequency of the input signal passed by filter 39 will appear in one or both of the filters 40 and 41 having higher cutoff frequencies. The output of each band pass filter is coupled to one of three similar channels, each channel comprising an amplifier and limiter circuit 42a, 42b, and 42c and differentiating network and amplifier circuit 43a, 43b, and 43c. The output of each channel comprises a narrow pulse each time the output of the associate filter goes to zero. The frequency output of each filter is equal to the frequency of the pulse output from each channel. The amplitude of the pulse output of each channel is adjusted by rheostats 44a, 44b, and 44c in such a manner that the output of the lowest band pass filter 39, middle band pass filter 40, and highest band pass filter 41 are in the ratio of 4:2:1. The amplitude adjusted output is such that the fundamental signal of the input energy will appear strongest in the combined output of the three channels. The amplitude adjusted outputs of the three channels are coupled to a mixer 45 whose output comprises a series of pulses such that those with the greatest amplitude are spaced by the fundamental period regardless of which band pass filter transmits the fundamental frequency. The series of pulses from mixer 45 are then coupled to a peak detector 46, which responds only to the pulses of the greatest amplitude and produces a sawtooth shaped wave whose fundamental frequency is that of the output of the lowest frequency channel which has an output. Hence the sawtooth wave from peak detector 46 is at the fundamental frequency. This sawtooth wave output from peak detector 46 is amplified, then differentiated, then amplified again in circuit 47 in such a manner that the final output from circuit 47 comprises a series of pips at the fundamental frequency, which, as shown in Fig. 1, is coupled to a counter-circuit or other utilizing means which are responsive to the fundamental frequency of the voiced sounds of the original speech input.

While we have described our invention in connection with certain specific examples, it is to be clearly understood that such examples are not to be construed as a limitation on the scope of our invention as set forth in the objects thereof or in the accompanying claims.

We claim:

1. A system for deriving a series of pulses having a frequency equal substantially to the fundamental pitch of voiced sounds, comprising means to detect voiced sounds in a signal source, a plurality of frequency adjacent band pass filters, means to adjust the output of each of said band pass filters so that energy of the lowest frequency passed will have the greatest relative amplitude, means to generate a series of pulses responsive to the frequency energy having the largest relative amplitude, and switching means to couple said signal source to said plurality of filters in accordance with detection of voiced sounds in the signals of said source.

2. A system for deriving a series of pulses having a repetition rate equal substantially to the fundamental pitch of voiced sounds of human speech, comprising means to detect a change in time interval between adjacent peaks in said human speech, a plurality of frequency adjacent band pass filters each less than one octave wide, means to adjust the relative amplitude of the output of each of said band pass filters so that the energy output of the band pass filter passing the lowest frequency will have the greatest relative amplitude, means to generate a series of pulses responsive to the frequency energy having the largest relative amplitude, and switching means controlled by the output of said detector means for coupling to said band pass filters only input speech components having substantially the same time interval between adjacent amplitude peaks.

3. A system according to claim 2, wherein said means to detect a change in time interval between adjacent amplitude peaks in said speech comprises means to form a series of pulses each of equal width responsive to the larger peaks in said speech, means to generate a sawtooth wave timed according to the trailing edges of said pulses, means to store the instantaneous voltage of said sawtooth wave coincident with the leading edges of said pulses, and means to detect changes in said stored voltage.

4. A system according to claim 2, wherein said means to adjust the relative amplitude of the output of said band pass filters includes means to attenuate the outputs of said filters proportional to the portion of the frequency range passed by each filter.

5. A system according to claim 2, wherein said means to generate a series of pulses responsive to the output of the band pass filter having the largest relative amplitude includes means to detect said largest relative amplitude peaks from the output of said band pass filters and means to differentiate said detected peaks.

6. A system for determining the fundamental pitch of voiced sounds in human speech when said speech has its fundamental missing, comprising means to detect a change in time interval between relatively large adjacent peaks in the signals of a speech source, means to distort said speech signals to reproduce said missing fundamental pitch, a plurality of frequency adjacent band pass filters each less than one octave wide, means to adjust the relative amplitude of the output of each of said band pass filters so that energy of the lowest frequency passed will have the greatest relative amplitude, means to generate a

series of pulses responsive to the frequency energy having the largest relative amplitude and switching means controlled by the output of said detector means for coupling to said distortion means only those speech signals having substantially the same interval between relatively large adjacent amplitude peaks.

7. A system according to claim 6, wherein said means to reproduce said fundamental pitch includes a square law distortion type network.

8. A system to detect a change in time interval between adjacent pulses of constant width, comprising means to generate a sawtooth wave timed according to the trailing edges of said pulses, means to store the instantaneous voltage of said sawtooth wave coincident with the leading edges of said pulses, and means to detect changes in voltage present on said storage means.

9. A system according to claim 8, wherein said means to store the instantaneous voltage of said sawtooth wave coincident with said leading edges includes a storage condenser, two electron discharge devices each including a plate, grid, and cathode, the plate of one and the cathode of the other of said devices being coupled to said storage condenser, the cathode of said one device and the plate of said other device being coupled to said sawtooth generating means, and means to apply a sampling pulse to said grids in accordance with the occurrence of said leading edges, whereby when the sampling pulse is impressed on said grids the instantaneous voltage of said sawtooth wave will be stored on said storage condenser.

10. A system for determining the fundamental pitch of voiced sounds in a signal source comprising a plurality of frequency adjacent band pass filters each less than one octave wide, means to apply signals from said source to said filters, means to adjust the relative amplitude of the outputs of said band pass filters so that the energy output of the lower frequency band pass filter will have the largest relative amplitude, and means to generate a series of pulses responsive to the frequency energy having the largest relative amplitude.

11. A system for determining the fundamental pitch of voiced sound having its fundamental pitch missing, comprising means to distort said voiced sounds, a plurality of frequency adjacent band pass filters each less than one octave wide, means to apply the distorted sounds to said filters, means to adjust the relative amplitude of the outputs of said band pass filters so that the energy output of the lower frequency band pass filter will have the largest relative amplitude, and means to generate a series of pulses responsive to the frequency energy having the largest relative amplitude.

References Cited in the file of this patent

UNITED STATES PATENTS

2,403,984	Koenig, Jr., et al. -----	July 16, 1946
2,406,825	French -----	Sept. 3, 1946
2,593,694	Peterson -----	Apr. 22, 1952
2,627,541	Miller -----	Feb. 3, 1953
2,640,880	Aigrain et al. -----	June 2, 1953