April 28, 1970 J. H. KING, JR 3,509,281

VOICING DETECTION SYSTEM

Filed Sept. 29, 1966

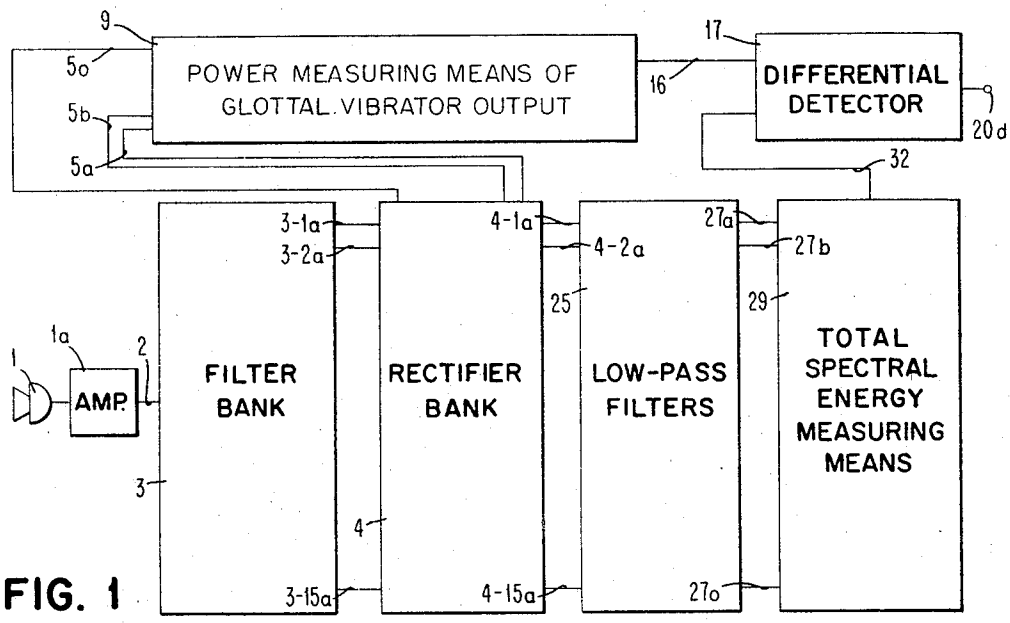


FIG. 1

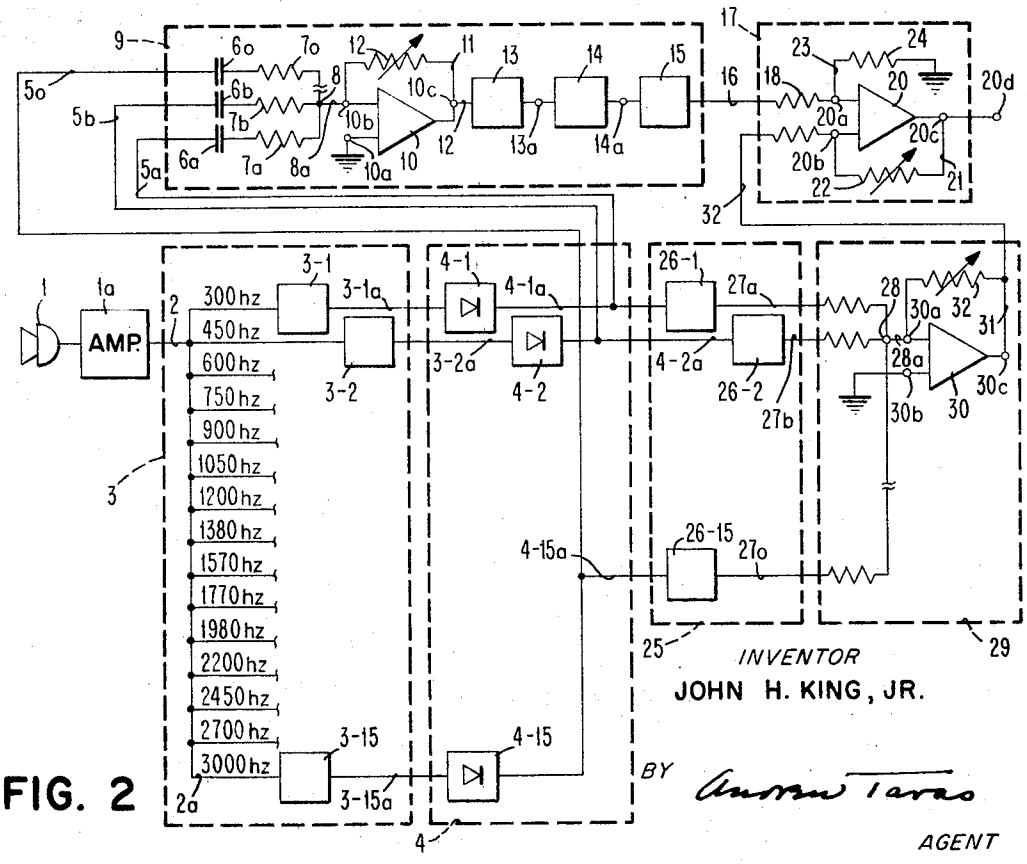


FIG. 2

INVENTOR
JOHN H. KING, JR.

BY

AGENT

United States Patent Office 3,509,281 Patented Apr. 28, 1970

3,509,281

VOICING DETECTION SYSTEM

John H. King, Jr., Endwell, N.Y., assignor to International Business Machines Corporation, Armonk, N.Y., a corporation of New York

Filed Sept. 29, 1966, Ser. No. 582,912
Int. Cl. G10l 1/00
U.S. Cl. 179—1 5 Claims

ABSTRACT OF THE DISCLOSURE

The relative magnitude of voiced sounds in a speech spectrum is determined by comparing a quasi DC signal representing the voiced sounds against a DC summation signal representing the total spectral energy in the sound spectrum.

The DC summation signal is generated by passing the sound waveforms through a filter bank having a band pass width ranging between 300–3,000 Hz. The outputs of the latter are rectified and these rectified outputs are fed through a low pass filter designed to pass signals between 0 and 15 Hz. The rectified outputs representing total sound energy are finally summed by a DC amplifier to provide the DC summation signal.

The quasi DC signal is developed by summing the same rectified outputs representing total sound energy, by means of an AC amplifier and passing the AC summed signal through a filter which is tuned to pass the voice fundamental frequency lying between 70–150 Hz., then rectifying the output of the voice fundamental and finally passing the rectified voice fundamental output through a low pass filter having a band pass width of 15 Hz. and below to provide the quasi DC signal.

The present invention relates to waveform analysis and more particularly to voicing detection predicated upon the periodicity of the power spectrum in a voiced speech sound spectrum. The prior art is replete with a variety of systems for synthesizing speech sounds, and analyzing speech sounds generated by individual speakers, and more recently analyzing the speech waveforms of different speakers by means of a single system.

In the speech recognition art, speech characteristics have been derived by various systems to the exclusion of voicing energies. Only recently has it been considered to utilize the process of voicing energy in the determination of a speech characteristic as part of a total recognition system. One approach considered by the prior art in voicing detection utilizes a single filter having a band pass range covering only the low end of the voice spectrum for extracting the voice energy. Another approach utilizes a single broad band filter having a range covering the entire speech spectrum. Both of these approaches have been highly unreliable from the standpoint that in the former approach a greater portion of the harmonic content was undetected whereas in the latter approach rectification of the unfiltered waveform as a means for extracting the fundamental is highly unreliable due to the wide variations in the amplitudes of the various harmonics.

The present invention is directed to a voice detection system which employs a sound spectrum analyzer having a plurality of individual band pass filters each having a minimum band pass width greater than the highest fundamental frequency of the voice spectrum, each filter passing a minimum of two harmonics. With this consideration in mind, the present invention employs a spectrum analyzer of 15 filters which is compatible with a fundamental range of 150 cycles. By virtue of the present invention, the periodic aspect of the power spectrum for all voiced sounds issues as a modulated waveform having an envelope with a

periodicity equal to the voice fundamental. As a consequence, the periodic property of the power spectrum for all voiced sounds may be measured with a high degree of accuracy and reliability since the outputs corresponding to the voice sounds are highly correlated. On the other hand, random noise, background noises and speech sounds other than the voiced sounds provide highly complex waveforms which are highly uncorrelated.

The present invention is, accordingly, directed to overcome the disadvantage of high cost and the inability of the prior art systems to be accurately responsive to a variety of speech spectrums. The capability of the present invention is generally achieved by means of a system which obtains a measure of the total voice spectral energy and compares it with a measure of the degree of periodicity of the voice power spectrum to provide an output which is indicative of the presence of voicing energy.

The primary object of the present invention is thus directed to a voicing detection system which has a high degree of reliability and is less costly than voicing detection systems of the prior art.

Another object resides in the capabilities of the present invention to provide more meaningful data at lower costs than the prior art systems.

Yet another object resides in the provision of a highly sophisticated system which derives meaningful voicing data predicated upon detecting and measuring the degree of periodicity of the power spectrum of the speech waveform.

Still another object resides in the provision of a voicing detection system which has high discrimination for the periodicity property of the power spectrum of voiced speech sounds.

The foregoing and other objects, features and advantages of the invention will be apparent from the following more particular description of preferred embodiments of the invention, as illustrated in the accompanying drawings.

In the drawings:

FIG. 1 is a schematic showing the arrangement of the principal means forming the voicing detection system.

FIG. 2 is a detail drawing of the voicing detection system.

A general understanding of the present invention may now be had from FIG. 1 which shows a schematic arrangement of the principal means constituting the voicing detection system. In FIG. 1 speech sounds are entered into the system by way of a microphone **1** which translates sound energy into electrical energy which is amplified by means of an amplifier **1***a* and entered by way of line **2** into a spectrum analyzer **3** which is essentially a filter bank constituted of a plurality of individual filters whose output waveforms are passed by way of lines **3–1***a*, **3–2***a* to **3–15***a* (individual lines between **3–2***a* and **3–15***a* not shown) to individual rectifiers in a rectifier bank **4**. The rectified outputs from the rectifier bank **4** are transmitted to a power measuring means **9** for the glottal vibrator by way of lines **5***a*, **5***b* through **5***o*. These same rectified outputs are also directed to a total spectral energy detecting and measuring means **29** by way of the lines **4–1***a*, **4–2***a* through **4–15***a*, a low pass filter bank **25** and lines **27***a*, **27***b* through **27***o*. By means of the power measuring means **9** the periodicity measure of a speech waveform power spectrum is reduced to substantially a DC signal level which is passed on by way of line **16** to a differential detector **17** for final comparison with a similar DC signal level on line **32** which is indicative of the total spectral energy. Detection of the total voice spectral energy is accomplished by means **29** which translates the energy to substantially a DC signal level that is also transmitted to the differential detector **17** (a bistable device) by way of a line **32**. The differential

detector **17** provides an indication of the presence of voicing energy on output line **20***d*. By appropriate means to be described, a low output signal level is manifested by the detector **17** when meaningful degree of periodicity of the speech spectrum is predominant, and a high signal level when the speech spectrum is essentially constituted of fricative and noise sounds with little or no meaningful degree of periodicity of the power spectrum. The meaningful periodicity of the power spectrum of voiced speech is the result of voiced energy present in the speech spectrum. This voiced energy is present during intervals of speech where excitation of the vocal tract is provided by the glottal vibrator. During these intervals the voiced sounds are predominantly rich in harmonics that are integer multiples of the fundamental frequency which for the male voice extends from about 70 cycles to 150 cycles per second in normal speech, the meaningful spectrum of which extends from 300 cycles to somewhat beyond 3,000 cycles.

To appreciate more fully the manner in which the voicing system detects the periodic structure of the speech waveform power spectrum as well as the total spectral energy in the speech spectrum, reference is invited to FIG. 2 which shows in detail the preferred embodiment. In FIG. 2, sound waves enter the system by way of the microphone **1** and are converted by means of the amplifier **1***a* into electrical waveform signals which enter the filter bank **3** by way of line **2**. The filter bank **3** comprises fifteen individual filters, three of which are shown and referenced as **3–1**, **3–2**, and **3–15**. The filters employed here are commonly known as a twin T-type, each having a band width of approximately 300 cycles and each is tuned to a desired band width and center frequency. The frequencies indicated in the filter bank **3** denote the center frequencies to which each of the 15 filters is tuned. For example, the topmost filter **3–1** has a center frequency of 300 cycles per second, and the lowermost filter **3–15** has a center frequency of 3,000 cycles per second. The filter bank **3**, accordingly, provides a plurality of orthogonal signal channels controlled by the contiguously tuned filters each providing, during a voiced speech interval, a modulated waveform the envelope of which contains the fundamental voice frequency. Thus, modulation results from the combination of waveforms which constitute the harmonic components of the fundamental frequency. In this particular embodiment a minimum of two harmonic components provide a modulated waveform. These modulated output waveforms are passed through full wave rectifiers, or detectors, **4–1**, **4–2** through **4–15**, by way of lines **3–1***a*, **3–2***a* through **3–15***a*. Rectified outputs from the rectifiers are transmitted by way of lines **4–1***a*, **4–2***a* through **4–15***a* through a low pass filter bank **25** containing fifteen individual low pass filters of which only three are shown, namely **26–1**, **26–2** and **26–15**. These low pass filters are designed to extract what may be considered a DC component level from the rectified outputs of the rectifiers. The low pass filters are tuned to accept frequencies below 15 cycles per second to extract the DC components, which components are transmitted to the total voice spectral energy detecting and measuring means **29** by way of resistive paths **27***a*, **27***b* through **27***o* which meet at a common juncture **28**. The junction **28** is connected by way of a line **28***a* to one input **30***a* of a DC summing amplifier **30** having a second input **30***b* connected to ground. Output **30***c* and the input **30***a* are interconnected by way of a negative feedback path **31** containing a potentiometer **32**. The summing amplifier together with the resistive input network constitutes a summing network which in response to incoming rectified waveforms of low frequency issue a DC component representing a measure of the total spectral energy of the speech sounds presented to the system. This total energy DC signal level is passed on to the differential detector **17** where it will be compared against another DC signal level representing the degree of periodicity of the power spectrum of the speech sounds entered into the system.

The detection of the periodic structure of the speech waveform power spectrum is achieved by the power measuring means **9** which is responsive to the rectified waveforms appearing on lines **4–1***a*, **4–2***a* through **4–15***a*. The latter are coupled by way of capacitors **6***a*, **6***b* through **6***o* to the resistive network constituted of resistive paths **7***a*, **7***b* through **7***o* terminating at junction **8** at which junction the rectified waveforms are presented to an input terminal **10***b* of an operational summing amplifier **10**. A second input **10***a* of the amplifier **10** is connected to ground. An output **10***c* is interconnected with the input terminal **10***b* by way of a negative feedback path **11** containing potentiometer **12**. The amplified output from the amplifier **10** appears on output line **12** connected to a band pass filter **13** having a band width of 70 cycles to 150 cycles corresponding to the frequency range of the voice fundamental frequency. The output appearing at **13***a* of the filter **13** reflects an amplified fundamental frequency which is rectified after passing through full wave rectifier **14**. The rectified output appearing on line **14***a* is passed through a low pass filter **15**, tuned to permit passage of frequencies of 15 cycles per second and below. The output from the latter reflects a quasi DC signal which is a measure of the degree of periodicity of the power spectrum of the speech sound. This quasi DC signal passes through line **16** connected by way of a resistor **18** to input terminal **20***a* of an amplifier **20** forming part of the differential detector **17**. The input terminal **20***a* is further connected to ground by way of a path **23** containing a resistor **24**. The amplifier **20** has its input and output terminals interconnected by means of a positive feedback path **21** containing potentiometer **22**.

The DC quasi signal and the summation DC signal voltages representing respectively the measures of total voiced energy and total spectral energy are weighted appropriately by adjusting the gains of the summing amplifiers **10** and **20** respectively so that for voiced speech sounds the DC level on line **16** exceeds the DC level on line **32** thus causing the differential detector **17** to switch to one of its stable states so that during unvoiced speech sounds or if no speech is occurring the DC level on line **16** will be less than the DC level on line **32**, thus causing the differential detector **17** to switch to its other stable state. The indication of the presence or absence of voiced speech is given by the voltage level appearing at the output terminal **20***d* of the differential detector **17**.

The fact that there exists a unique adjustment of the gains of amplifiers **20** and **30** such that the relative voltage levels on lines **16** and **32** will occur in opposite fashion according to whether voiced or unvoiced speech is present can be best demonstrated by the following:

The ratio of the signal output from filter **13**, for an input signal to the filter bank **3** having a spectrum consisting of a fundamental frequency of say 100 Hz. and a series of integer harmonic frequencies all of equal amplitude to the output at filter **13** due to a random input signal having a flat noise spectrum with a normal amplitude distribution and the same average power as the harmonic signal will be about $15/\sqrt{15}$ according to the well-known way in which the correlated and uncorrelated components of a composite signal contribute to the amplitude of the composite signal which is a linear combination of the component parts. With actual speech signals the maximum discrimination between voiced and unvoiced speech sounds will occur for the open vowels such as *e, ae, ai, i,* etc. because these sounds have more harmonic energy in the high frequency portion of the spectrum than say the closed or semi-vowels such as *w, v, u, j, r,* etc. For any given total spectrum energy a purely fricative sound such as *f,* θ or *s* will give the least output from filter **15** at **16** whereas certain voiced fricatives such as *z* and *v* will give intermediate levels of output on line **16**. Thus, it is seen that the least reliable detection will occur for the voiced fricatives and the

most reliable discrimination will occur for the open vowels.

For specific applications the invention may be limited to the means for measuring the power output of the glottal vibrator to provide an output representing the composite waveform from which the DC component may be extracted and stored as separate speech characteristics representing the specific voicing power present in a speech spectrum.

While the invention has been particularly shown and described with reference to preferred embodiments thereof, it will be understood by those skilled in the art that the foregoing and other changes in form and details may be made therein without departing from the spirit and scope of the invention.

What is claimed is:

1. A voicing detection system comprising

a source of rectified waveforms representing sounds of a voice spectrum,

power measuring means comprising an operational summing amplifier, a voice fundamental filter, rectifying means and a low pass filter,

said operational amplifier responsive to said rectified waveforms for summing the present correlated voiced waveforms and providing a composite voltage output representing the sum of said voiced waveforms,

said filter responsive to said composite voltage for issuing a voice fundamental frequency signal, said rectifier responsive to said fundamental frequency signal for issuing a rectified voice fundamental signal, and

said low pass filter responsive to the rectified voice fundamental signal for providing a quasi DC signal voltage indicative of the degree of periodicity of said voiced waveforms in said voice spectrum.

2. A voicing detection system as in claim 1, further including a plurality of low pass filters responsive to said rectified waveforms for issuing DC component voltages,

total spectral energy detecting means responsive to said DC component voltages for providing a summation DC voltage output representing the total spectral energy in said voice spectrum, and

a differential detector responsive jointly to said quasi DC signal voltage and the total spectral energy summation DC signal voltage to provide an output indicative of the relative magnitude of said voiced waveforms.

3. A voicing detection system as in claim 2, in which said differential detector is a differential amplifier having positive feedback to provide an output indicative of the presence of voiced waveform signals regardless of the over-all speech sound intensity.

4. A voicing detection system as in claim 3, in which said voice fundamental filter is tuned to the voice fundamental frequency range of 70 cycles per second to 150 cycles per second, and said low pass filter is tuned to accept frequencies at and below 15 cycles per second.

5. A voicing detection system as in claim 4, further including a filter bank constituted of a plurality of channels each responsive to a select band of sound frequencies in the voice spectrum ranging from approximately 300 cycles per second to 3,000 cycles per second, to provide waveform outputs, constituted of at least two harmonics for the voiced sounds, including means responsive to said waveform outputs for issuing said source of rectified waveforms representing voiced sounds of the voice spectrum.

References Cited

UNITED STATES PATENTS

2,151,091	3/1939	Dudley	179—1.5
3,129,287	4/1964	Bakis.	
3,247,322	4/1966	Savage et al.	

KATHLEEN H. CLAFFY, Primary Examiner

C. JIRAUCH, Assistant Examiner