VOICENG DETECTION SYSTEM

John H. King, Jr., Endwell, N.Y., assignor to Interna-
tional Business Machines Corporation, Armonk, N.Y.,
a corporation of New York
Filed Sept. 29, 1966, Ser. No. 582,912

InCl. Cl 180/1/00 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 speech
spectrum.

The DC summation signal is generated by passing the
sound waveforms through a filter bank having a band pass
width ranging between 50–5000 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 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 funda-
mental frequency lying between 70–150 Hz, then rectify-
ing 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 sys-
tems for synthesizing speech sounds, and analyzing speech
sounds generated by individual speakers, and more recent-
ly 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 de-
tection utilizes a single filter having a band pass range
covering only the low end of the voice spectrum for ex-
tracting the voice energy. Another approach utilizes a sin-
gle 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 funda-
mental 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 frequency
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 conse-
quence, 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 over-
come 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 periodic-
ity 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 di-
rected to a voicing detection system which has a high
degree of reliability and is less costly than voicing de-
tection 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 wave-
form.

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 in-
vention, 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 ar-
angement of the principal means constituting the voicing
detection system. In FIG. 1 voice sounds are entered
into the system by way of a microphone 1 which trans-
lates sound energy into electrical energy which is am-
plified by means of an amplifier 1a 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
3–1a, 3–2a to 3–15a (individual lines between 3–2a and
3–15a not shown) to individual rectifiers in a rectifier
bank 4. The rectified outputs from the rectifier bank
are transmitted to a power measuring means 9 for
the glottal vibrator by way of lines 5a, 5b through 5e.
The same rectified outputs are also directed to a total spectral
energy detecting and measuring means 29 by way of the
lines 4–1a, 4–2a through 4–15a, a low pass filter bank
25 and lines 27a, 27b through 27o. By means of the
power measuring means 9 the periodicity property of
a speech waveform power spectrum is reduced to sub-
stantially 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 20d. By appropriate means to be described, a low output signal level is manifested by the rectified waveforms when the 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 3000 cycles.

The operation here is 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 microphones 1a, 1b which are converted to 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 bandwidth 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 lowest filter 3-15 has a center frequency of 3000 cycles per second. The filter bank 3, accordingly, provides a plurality of orthogonal signal channels controlled by the continuously 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 orators, 4-1, 4-2 through 4-15, by way of lines 3-1a, 3-2a through 3-15a. Rectified outputs from the rectifiers are transmitted by way of lines 4-1a, 4-2a through 4-15a 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 27a, 27b through 27c which meet at a common juncture 28. The junction 28 is connected by way of a line 28a to one input 30a of a DC summing amplifier 30 having a second input 30b connected to ground. Output 30c and the input 30a 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 a 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 39 which is responsive to the rectified waveforms which is an indication of the presence of voicing energy on output line 20d. By appropriate means to be described, a low output signal level is manifested by the rectified waveforms when the degree of periodicity of the power 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 3000 cycles.

The operation here is 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 microphones 1a, 1b which are converted to 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 bandwidth 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 lowest filter 3-15 has a center frequency of 3000 cycles per second. The filter bank 3, accordingly, provides a plurality of orthogonal signal channels controlled by the continuously 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 orators, 4-1, 4-2 through 4-15, by way of lines 3-1a, 3-2a through 3-15a. Rectified outputs from the rectifiers are transmitted by way of lines 4-1a, 4-2a through 4-15a 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 27a, 27b through 27c which meet at a common juncture 28. The junction 28 is connected by way of a line 28a to one input 30a of a DC summing amplifier 30 having a second input 30b connected to ground. Output 30c and the input 30a 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 a DC signal level representing the degree of periodicity of the power spectrum of the speech sounds entered into the system.
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,

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

4. A voicing detection system as in claim 3, including a plurality of low pass filters responsive to said rectified waveforms for issuing DC component voltages.

5. A voicing detection system as in claim 4, further including a plurality of low pass filters responsive to said rectified waveforms for issuing DC component voltages.

References Cited

UNITED STATES PATENTS

2,151,091 3/1939 Dudley
3,247,322 4/1966 Savage et al.

KATHLEEN H. CLAFFY, Primary Examiner
C. JIRAUCH, Assistant Examiner