

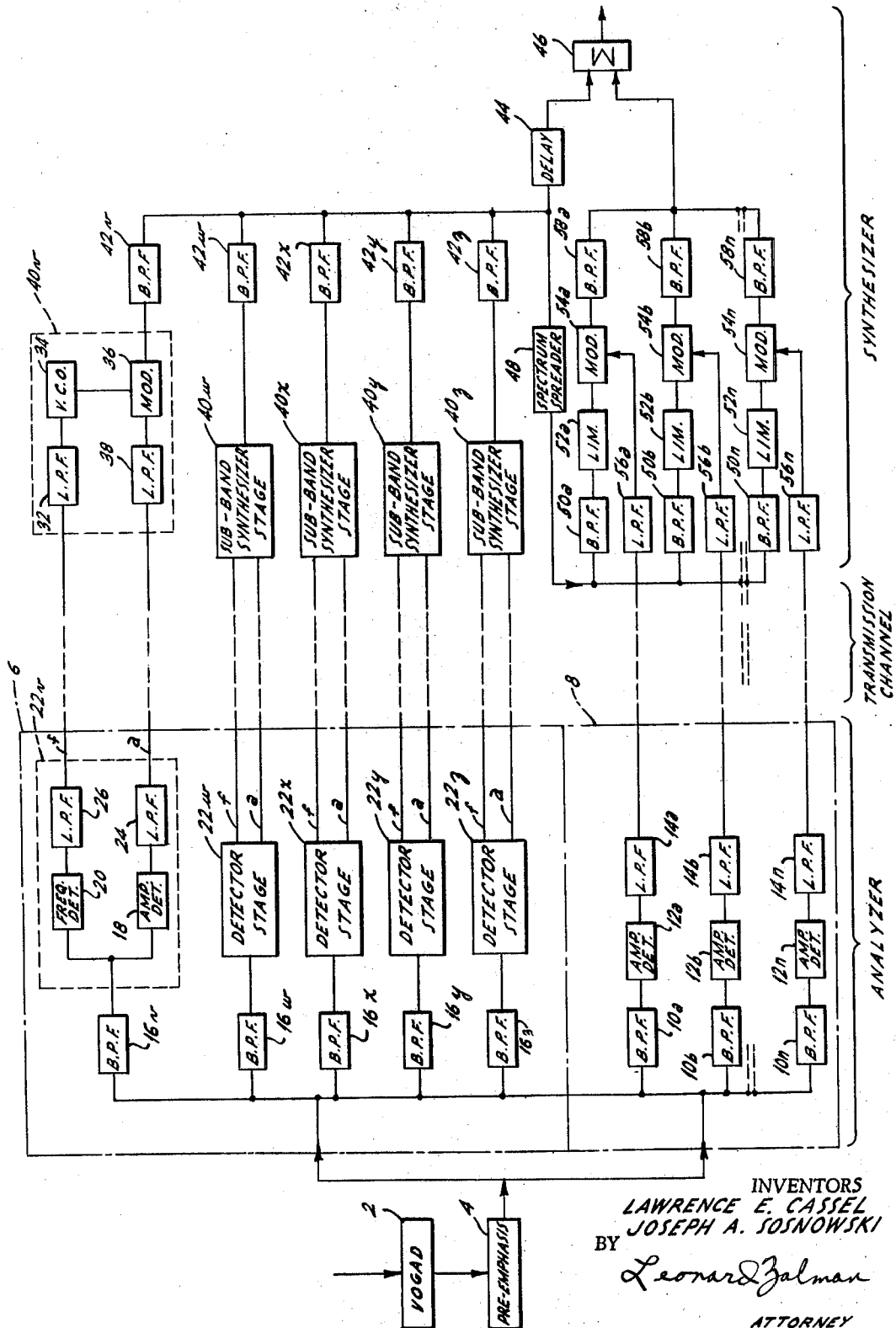
March 10, 1970

L. E. CASSEL ET AL

3,499,991

VOICE-EXCITED VOCODER

Filed Aug. 1, 1967



1

2

3,499,991

VOICE-EXCITED VOCODER

Lawrence E. Cassel, Norristown, and Joseph A. Sosnowski, Conshohocken, Pa., assignors to Philco-Ford Corporation, Philadelphia, Pa., a corporation of Delaware
Filed Aug. 1, 1967, Ser. No. 657,661

Int. Cl. H04m 1/19

U.S. Cl. 179—1

7 Claims

ABSTRACT OF THE DISCLOSURE

A "voice-excited" vocoder in which a substantial replica of the "base band" portion of the input speech wave is reconstructed at the synthesizer from a plurality of slowly varying signals representative of the frequency-vocal energy distribution of adjacent portions of the "base band" of the input speech wave.

Although human speech includes frequency components extending from approximately 70 c.p.s. to 9,000 c.p.s., it is possible to transmit intelligible human speech through a narrower passband. In ordinary telephone transmission, the passband allotted to one speech channel commonly extends from approximately 100 c.p.s., as its lower limit, to an upper limit of approximately 3200 c.p.s. This passband, which is generally sufficient to ensure intelligible speech reproduction, represents a compromise between the desirability of high fidelity of reproduction and the desirability of economy in the allocation of channel space. However, even this restricted passband presents problems where many messages are to be transmitted over a limited number of transmission channels.

Many systems have been proposed for transmitting speech over a fraction of the passband ordinarily allotted to a single speech channel in a telephone system. Most of these systems are based upon the channel vocoder of H. W. Dudley described in U.S. Patent No. 2,151,031, issued Mar. 21, 1939. In this system an input speech wave is supplied to a bank of contiguous bandpass filters each of which is followed by a detector and a low pass filter with a cutoff frequency of approximately 20 c.p.s. The output of each low pass filter is a slowly varying signal whose instantaneous amplitude represents the instantaneous amplitude of the vocal energy in the frequency band with which it is associated. In addition to these signals, the vocoder produces a signal representative of the fundamental or pitch frequency of the input speech wave and a signal which indicates whether a particular sound is voiced (periodic) or unvoiced (aperiodic). The signals are transmitted to a receiver station and are there utilized to synthesize an artificial speech wave having the characteristic pitch and frequency-vocal energy distribution of the input speech wave. For this purpose, the synthesizer includes a "buzz" source and a "hiss" source which produce voiced and unvoiced signals, respectively, in response to the incoming signals. The frequency of the "buzz" source is determined by the pitch signal.

The artificial speech wave synthesized by a channel vocoder is usually intelligible but it lacks quality and naturalness. These degradations are, in part, due to the difficulty of determining the fundamental frequency (pitch) of the input speech wave and of identifying voiced segments of the input speech wave, and, in part, to the requirement that the pitch and voicing of the artificial speech wave match that of the input speech wave.

To avoid these problems, the "voice-excited" vocoder was developed. In the "voice-excited" vocoder, the low frequency portion of the input speech wave (up to ap-

proximately 1,000 c.p.s.), called the "base band," is transmitted to the synthesizer without a reduction in bandwidth. The "base band" signal performs the function of both the "hiss" source and the "buzz" source of the conventional channel vocoder. Correlation between the pitch and the voicing of the input speech wave and the pitch and the voicing of the synthesized speech wave is achieved by spreading the frequency spectrum of the "base band" signal, by nonlinear distortion, to generate harmonic frequency components of the "base band" up to the highest frequency band to be synthesized. Since the frequency sub-bands of a "voice-excited" vocoder are derived directly from a portion of the input speech wave, the periodicity and voicing of the input speech wave is automatically reproduced without resort to "hiss" and "buzz" sources.

As a result of the automatic reproduction of the periodicity and voicing of the input speech wave, artificial speech waves produced by a "voice-excited" vocoder are superior to those produced by a channel vocoder. However, to accommodate the entire "base band," "voice-excited" vocoders require a larger transmission passband than channel vocoders.

It is therefore an object of the present invention to provide a new "voice-excited" vocoder.

It is a further object of the present invention to provide a "voice-excited" vocoder that requires less transmission bandwidth than prior art "voice-excited" vocoders.

According to the present invention, the aforementioned "voice-excited" vocoder problem, i.e. the requirement of a large transmission bandwidth for the "base band" of the input speech wave, is resolved by transmitting to a speech synthesizer a plurality of slowly varying signals representative of the frequency distribution and vocal energy distribution of adjacent portions of the "base band." These slowly varying signals, which supply information sufficient to synthesize a high quality speech wave, requires only a fraction of the transmission bandwidth required for the unmodified transmission of the "base band."

In a preferred embodiment of the present invention the bandwidth required for a "voice-excited" vocoder is reduced by supplying the "base band" portion of the input speech wave to means for dividing the "base band" into a plurality of contiguous frequency sub-band signals. Each frequency sub-band signal is supplied to means which produce slowly varying signals representative of the frequencies present in each sub-band and of the vocal energy corresponding to these frequencies.

These signals, together with signals representative of the vocal energy of the high frequency portion of the input speech wave, are transmitted to a synthesizer which comprises means responsive to said signals for reconstructing a signal representative of the frequency-vocal energy characteristics of each of said frequency sub-bands, means for combining said reconstructed sub-band signals to produce a control signal which is substantially a replica of the "base band" portion of the input speech wave, means utilizing said control signal for generating the high frequency portion of the speech wave to be synthesized, and means for combining said control signal and said high frequency portion of the speech wave to produce the synthesized speech wave.

For a better understanding of the present invention together with other and further objects thereof reference should now be had to the following detailed description which is to be read in conjunction with the accompanying drawing which is a block diagram showing a speech communication system in accordance with the present invention.

Referring to the drawing, there is shown the analyzer and the synthesizer of the "voice-excited" speech communication system of the present invention. At the analyzer an electrical representation of a speech wave, such as produced by a standard telephone carbon microphone, is supplied through a "Vogad" (voice operated gain adjusting device) 2 and a pre-emphasis network 4 to the "base band" processing portion 6 of the analyzer and to the high frequency processing portion 8 of the analyzer. The "Vogad" 2 maintains the input signal to network 4 at a relatively constant amplitude and the pre-emphasis network 4 compensates for the unequal frequency-vocal energy distribution characteristic of human speech.

The high frequency portion 8 of the analyzer comprises a plurality of bandpass filters 10a, 10b . . . 10n. In accordance with conventional vocoder techniques, each of the filters 10a, 10b . . . 10n is constructed to pass adjacent frequency portions of the input speech wave extending from approximately 1100 c.p.s. to approximately 3200 c.p.s. Bandpass filters 10a, 10b . . . 10n are connected to conventional amplitude detectors 12a, 12b . . . 12n, respectively, which are followed by low pass filters 14a, 14b . . . 14n, respectively. Amplitude detectors 12a, 12b . . . 12n can include amplifier and peak detector stages and each of the filters 14a, 14b . . . 14n can have a band pass of approximately 30 c.p.s. The outputs of filters 14a, 14b . . . 14n are slowly varying control signals whose instantaneous amplitude represents the instantaneous amplitude of the vocal energy in the frequency band with which it is associated.

The low frequency portion 16 of the analyzer consists of a plurality of bandpass filters 16v, 16w, 16x, 16y and 16z. In accordance with the present invention the filters 16v, 16w, 16x, 16y and 16z are constructed to pass adjacent frequency portions of the input speech wave extending from approximately 100 c.p.s. to approximately 1100 c.p.s. For example, bandpass filters 16v, 16w, 16x, 16y and 16z can be proportioned to pass vocal energy in the sub-bands extending from 100 to 300 c.p.s., 300 to 500 c.p.s., 500 to 700 c.p.s., 700 to 900 c.p.s., and 900 to 1100 c.p.s., respectively. Additional sub-bands (not shown) can be employed, if desired, by providing additional parallel paths identical to those shown and proportioned to pass the desired frequency bands. Since the frequencies measured in any one channel are harmonics of the fundamental frequency, the centroid of each of the filters 16v, 16w, 16x, 16y and 16z should be multiples of the fundamental frequency. Thus, for a female voice with a fundamental frequency of approximately 200 c.p.s., the centroids of the filters 16v, 16w, 16x, 16y and 16z should be 200, 400, 600, 800 and 1000 c.p.s., respectively.

Filter 16v is coupled to a conventional amplitude detector 18 and to a frequency detector 20 of a detector stage 22v which also includes low pass filters 24 and 26 coupled to the detectors 18 and 20, respectively. The output of filter 24, which can have a band pass of approximately 50 c.p.s., is a slowly varying signal whose instantaneous amplitude represent the instantaneous amplitude of the vocal energy in the frequency band defined by filter 16v.

Detector 20, which may be a conventional ratio detector, generates a signal the instantaneous amplitude of which is substantially proportional to the instantaneous variation of the frequency of the input signal thereto from a mean value, e.g., the centroid frequency of filter 16v. Low pass filter 26 passes the amplitude variations of the signal generated by detector 20 up to a frequency of approximately 30 c.p.s.

Bandpass filters 16w, 16x, 16y and 16z are coupled to detector stages 22w, 22x, 22y and 22z, respectively. Each of the detector stages 22w, 22x, 22y and 22z, which can be similar to detector stage 22v, produces a pair of slowly varying signals; the instantaneous amplitude of one signal being representative of the instantaneous amplitude of the vocal energy of the input signal thereto and the instantaneous

amplitude of the other signal being proportional to the instantaneous variation of the frequency of the input signal thereto from a mean value, which, in each case, can be the centroid of the bandpass filter associated therewith. It has been discovered that the low frequency signals from detector stages 22v, 22w, 22x, 22y, and 22z supply information sufficient to control the synthesis of artificial speech.

The output signals of detector stages 22v, 22w, 22x, 22y and 22z and the output signals of filters 14a, 14b . . . 14n are combined and transmitted by conventional wire facilities or electromagnetic systems to the synthesizer. For example, the output signals can be time multiplexed together and transmitted in digital form. Alternatively, the output signals can be systematically arranged adjacent each other on a RF carrier by conventional heterodyning techniques.

At the synthesizer, the combined signal transmitted from the analyzer is separated into the corresponding components developed at the analyzer. The output of low pass filter 26 is supplied through a low pass filter 32 to a voltage controlled oscillator 34, which reconstructs a substantial replica of the frequency spectrum of the portion of the input speech wave defined by filter 16v. The output of oscillator 34 is then amplitude modulated in modulator 36 by the output of low pass filter 24, which has been processed through a low pass filter 38. System components 32, 34, 36 and 38 comprise a sub-band synthesizer stage 40v. The output of stage 40v is a signal representative of the frequency-vocal energy distribution of the portion of the input speech wave defined by filter 16v.

The output signals of detector stages 22w, 22x, 22y and 22z are supplied to sub-band synthesis stages 40w, 40x, 40y and 40z, respectively, which can be similar to stage 40v. Each stage 40w, 40x, 40y and 40z produces a signal representative of the frequency-vocal energy distribution of the portion of the input speech wave defined by filters 16w, 16x, 16y and 16z, respectively.

The output signals of stages 40v, 40w, 40x, 40y and 40z are processed through bandpass filters 42v, 42w, 42x, 42y and 42z, which are identical to those of the analyzer and which remove undesirable harmonic components introduced during the processing of the speech wave, and then combined additively to form a combined output signal occupying the frequency band between 100 and 1100 c.p.s. This signal is a substantial replica of the "base band" portion of the input speech wave supplied to the analyzer.

The combined signal is supplied through a delay network 44 to an adder network 46. Delay network 44 compensates for delays produced during the synthesis of the high frequency portion of the artificial speech wave. The combined signal is also supplied to a frequency spectrum spreader or harmonic generator 48 which, by means of nonlinear distortion of the input signal thereto, produces a wide-band excitation signal containing harmonics of the fundamental frequencies present in the "base band" of the input speech wave over a range of frequencies between 1100 c.p.s. and 3200 c.p.s. A suitable frequency spectrum spreader for the system of the present invention is described in U.S. Patent No. 3,030,450, issued to M. R. Schroeder on Apr. 17, 1962, entitled "Band Compression System." Since the excitation signal generated by spectrum spreader 48 is derived directly from the "base band" signal, it has inherently the correct periodicity; for an aperiodic input, its output is also aperiodic and, for a periodic input its output is periodic with the same periodicity.

The wide-band excitation signal is supplied to band pass filters 50a, 50b . . . 50n, which are identical to those in the high frequency portion of the analyzer. The output signals of band pass filters 50a, 50b . . . 50n are supplied through limiters 52a, 52b . . . 52n, respectively, to modulators 54a, 54b . . . 54n, respectively, where they are modulated by the output signals of low pass filters 14a, 14b . . . 14n, respectively, which have been processed

5

through low pass filters 56a, 56b . . . 56n. The resulting modulated waves are then bandpass filtered by filters 58a, 58b . . . 58n, which are also identical to filters 10a, 10b . . . 10n, respectively, to eliminate undesirable harmonics created in the modulation process and then linearly summed and added to the delayed "base band" signal in summer 46 to produce an output signal which, when properly de-emphasized, has a frequency-vocal energy characteristic which is a substantial replica of the input speech wave.

As previously described, the system of the present invention requires the transmission of only ten slowly varying control signals to produce a signal which is a substantial replica of the "base band" portion of the input speech wave. When 30 c.p.s. control signals are employed, a total transmission bandwidth of 300 c.p.s. (10×30 c.p.s.) is required. Thus, the "voice-excited" vocoder system of the present invention requires a "base band" transmission bandwidth which is less than one-third that required by prior art "voice-excited" vocoders.

While the invention has been described with reference to a particular embodiment thereof, it will be apparent that various modifications and other embodiments thereof will occur to those skilled in the art within the scope of the invention. For example, since spectrum spreader 48 generates an output signal which is dependent only on the frequency components of the input signal thereto, the output signals of the oscillators of stages 40v, 40w, 40x, 40y and 40z can be combined and supplied directly to spectrum spreader 48. In this modified system, the "base band" portion of the synthesized speech wave can be generated in the manner previously described. Accordingly, we desire the scope of our invention to be limited only by the appended claims.

We claim:

1. A "vocoder" system for transmitting an electrical representation of a speech wave from a transmitter station to a receiver station which comprises, at said transmitter, first means for producing a first plurality of slowly varying signals representative of the vocal energy distribution of the low frequency portion of said speech wave and a second plurality of slowly varying signals representative of the frequency distribution of the low frequency portion of said speech wave, second means for producing a third plurality of signals representative of the vocal energy distribution of the high frequency portion of said speech wave, means for transmitting said signals to said receiver station, and at said receiver, third means responsive to said first and second plurality of signals for generating a signal representative of the low frequency portion of said speech wave, fourth means responsive to both said third plurality of signals and said signal representative of the low frequency portion of said speech wave for generating a signal representative of said high frequency portion of said speech wave, and fifth means for combining said signals representative of said high and low frequency portions of said speech wave.

2. The system of claim 1 in which said first means comprises a plurality of contiguous bandpass filters for dividing said low frequency portion of said speech wave into a plurality of sub-bands, an amplitude detector and a frequency detector coupled to the output of each of said filters, and low pass filters coupled to said detectors.

3. The system of claim 2 in which said third means of claim 1 comprises a plurality of stages, each of said stages producing a signal representative of the frequency-vocal energy distribution of a different one of said sub-bands in response to one of said first plurality of signals and a corresponding one of said second plurality of signals, and means coupled to each of said stages for producing a combined output signal therefrom having the desired frequency range.

4. The system of claim 3 in which said last mentioned

6

means of claim 3 includes a plurality of contiguous bandpass filters identical to said plurality of filters of said first means.

5. The system of claim 4 in which said fourth means of claim 1 comprises means responsive to said combined output signal of said stages for spreading the frequency spectrum of said signal to produce a wide-band excitation signal whose frequency spectrum encompasses the frequency spectrum of the high frequency portion of said speech wave, means controlled by said third plurality of signals and said wide-band excitation signal for generating a fourth plurality of signals representative of the frequency-vocal energy distribution of contiguous portions of the high frequency portion of said speech wave, and means for combining said fourth plurality of signals to produce said signal representative of said high frequency portion of said speech wave.

6. The system of claim 5 in which said fifth means of claim 1 includes time delay means.

7. The system of claim 1 in which said first means comprises a first plurality of contiguous bandpass filters for dividing said low frequency portion of said speech wave into a plurality of sub-bands, first plurality of series circuits each comprising an amplitude detector and a low pass filter, a different one of said first plurality of series circuits being connected to each filter of said plurality of filters, a second plurality of series circuits each comprising a frequency detector and a low pass filter, a different one of said second plurality of series circuits being connected to each filter of said plurality of filters; said second means comprises a second plurality of contiguous bandpass filters for dividing said high frequency portion of said speech wave into a plurality of sub-bands, a third plurality of series circuits each comprising an amplitude detector and a low pass filter, a different one of said third plurality of series circuits being connected to each filter of said second plurality of filters; said third means comprises a plurality of stages, each of said stages including a voltage controlled oscillator coupled to a modulator, each of said oscillators being coupled to said transmission means so that it receives the output signal of one of said second plurality of series circuits and each of said modulators being coupled to said transmission means so that it receives the output signal of a corresponding one of said first plurality of series circuits, a third plurality of contiguous bandpass filters identical to said first plurality of filters of said first means, each filter of said third plurality of filters being coupled to a different one of said modulators; said fourth means comprises a harmonic generator coupled to each of said modulators, a fourth plurality of bandpass filters identical to said second plurality of filters of said second means coupled to said harmonic generator, a second plurality of modulators coupled to said fourth plurality of filters, each of said second plurality of modulators being coupled to said transmission means so that it receives the output signal of one of said third plurality of series circuits; and said fifth means includes time delay means.

References Cited

UNITED STATES PATENTS

3,102,928	9/1963	Schroeder	179—1
3,431,362	3/1969	Miller	179—15.55

OTHER REFERENCES

IRE Transactions on Audio, May-June 1960, pp. 95-102; A Resonance Vocoder and Baseband Complement: A Hybrid System for Speech Transmission, James L. Flanagan.

KATHLEEN H. CLAFFY, Primary Examiner

J. B. LEAHEEY, Assistant Examiner

UNITED STATES PATENT OFFICE
CERTIFICATE OF CORRECTION

Patent No. 3,499,991 Dated March 10, 1970

Inventor(s) L. E. Cassel and J. A. Sosnowski

It is certified that error appears in the above-identified patent and that said Letters Patent are hereby corrected as shown below:

Claim 2, line 62, after "the" insert -- same --.

Claim 7, line 23, after the comma (,) insert -- a --.

SIGNED AND
SEALED
AUG 11 1970

(SEAL)

Attest:

Edward M. Fletcher, Jr.

Attesting Officer

WILLIAM E. SCHUYLER, JR.
Commissioner of Patents