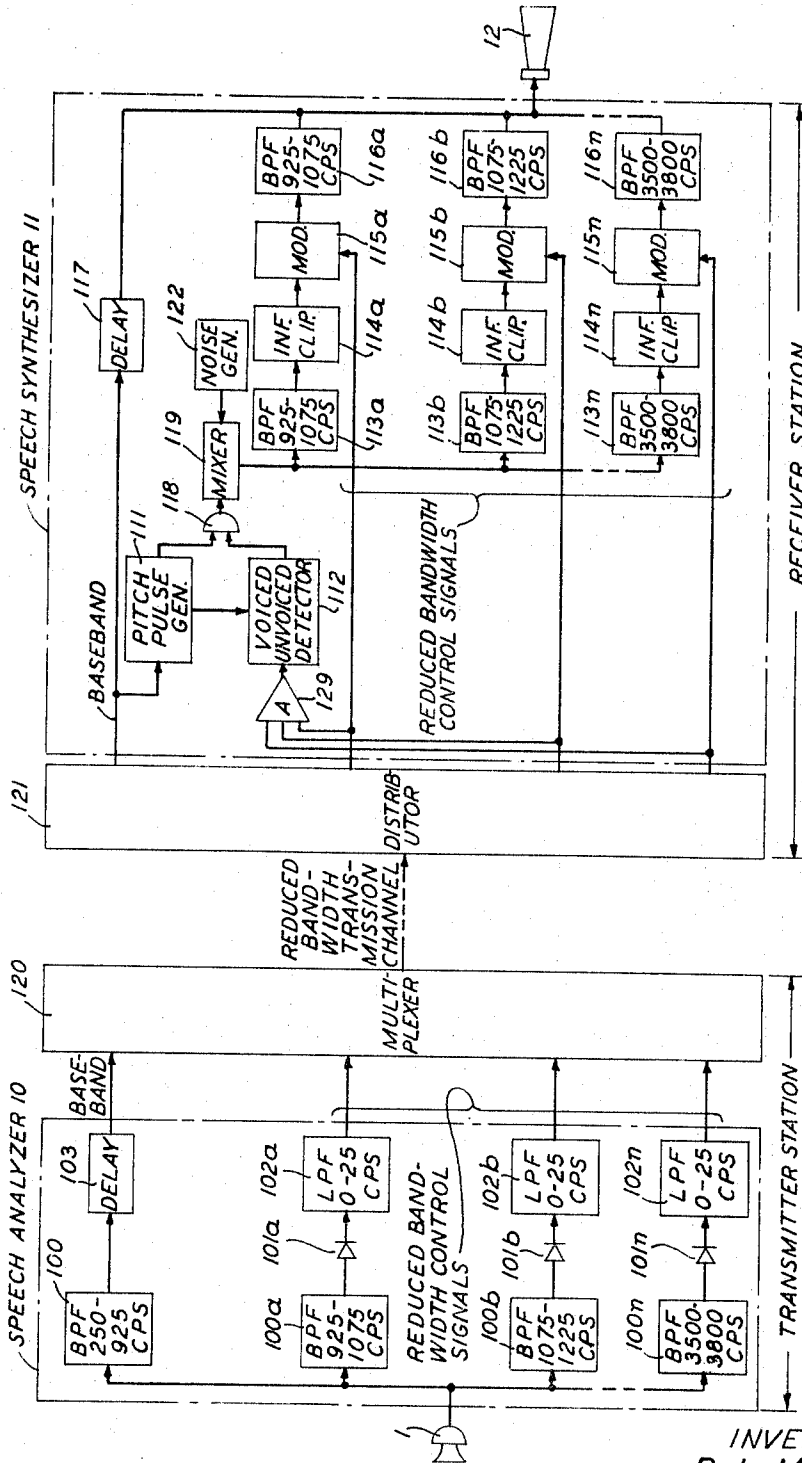


March 4, 1969

R. L. MILLER

3,431,362

VOICE-EXCITED, BANDWIDTH REDUCTION SYSTEM EMPLOYING PITCH  
FREQUENCY PULSES GENERATED BY UNENCODED BASEBAND SIGNAL  
Filed April 22, 1966



INVENTOR  
R. L. MILLER  
BY *A. E. Hirsch*  
ATTORNEY

1

3,431,362

## VOICE-EXCITED, BANDWIDTH REDUCTION SYSTEM EMPLOYING PITCH FREQUENCY PULSES GENERATED BY UNENCODED BASEBAND SIGNAL

Ralph L. Miller, Chatham, N.J., assignor to Bell Telephone Laboratories, Incorporated, New York, N.Y., a corporation of New York

Filed Apr. 22, 1966, Ser. No. 544,498

U.S. Cl. 179-15.55

Int. Cl. H04b 1/66; H04m 1/00

6 Claims

### ABSTRACT OF THE DISCLOSURE

Bandwidth reduction of a speech signal is obtained by separating the signal into a relatively low frequency or baseband signal and a relatively high frequency signal. The high frequency signal is divided into contiguous subbands and a reduced bandwidth control signal is produced for each subband. The baseband signal and control signals are transmitted. At the receiver, the baseband signal is used to generate one pulse per pitch period of the original speech signal. A voiced-unvoiced detector selectively gates the pitch period pulses to a mixer where the pulses are mixed with random noise signals. The output of the mixer is combined with the received control signals to reconstruct the high frequency subbands of the original speech signal. The original speech signal is synthesized by combining the reconstructed high frequency subbands with the received baseband signal.

This invention pertains to the transmission of reduced bandwidth speech signals and, more particularly, to the reduced bandwidth transmission of telephone quality speech signals.

Attempts to improve the performance of channel vocoder systems have been plagued by the so-called "pitch problem," i.e., the problem of determining, with the accuracy demanded by the ear, whether a speech signal is voiced or unvoiced and, if voiced, the pitch. With the advent of the voice-excited vocoder (VEV) described, e.g., in Patent 3,030,450, issued to M. R. Schroeder on Apr. 17, 1962, it was thought that the need for determining the fundamental pitch of a speech signal had been obviated. In the VEV, the excitation signal for the vocoder synthesizer is obtained from the uncoded low-frequency components of the speech signal, i.e., the baseband, by a process called "spectrum flattening." This process is described in the above-cited Schroeder patent and in Patent 3,139,487, issued to B. F. Logan et al. on June 30, 1964. The success of this method of excitation was immediate. The excitation signal in a VEV has, inherently, the correct periodicity: for an aperiodic input, the output of the spectrum-flattener is also aperiodic, and, for a periodic input, the output will be periodic with the same periodicity.

With the development of the voice-excited vocoder, synthesized speech has lost much of its unnatural "buzzy" quality. However, with the modern demand for high-fidelity vocoders, the conventional VEV has not been

2

found suitable in those particular applications requiring exacting standards.

The "spectrum flattening" process when applied to the received baseband signal produces a series of pulses for each pitch period of the speech signal. This series of pulses, when considered in terms of the impulse response of the synthesizing filter system, tends to fill in the whole pitch period of the synthesized speech with a series of peaks; in contradistinction to the simple damped sinusoid which is characteristic of voiced speech waves. The subjective effect of this series of peaks is to produce speech signals having, for the want of a better term, a very "mushy" quality. Measurements have indicated that this distortion is primarily due to phase variations in the harmonic spectral components. In addition, the spectrum flattening process tends to produce a "buzz-like" quality for unvoiced sounds; the sensitivity of the ear is such that an attempt to simulate what is essentially white noise with a series of quasi-random peak signals is bound to be detected.

It is therefore an object of this invention to overcome these difficulties and thus improve the quality and naturalness of synthesized speech in a voice-excited, bandwidth reduction system.

The present invention improves the quality and naturalness of synthesized speech by incorporating two seemingly incompatible concepts, namely, pitch determination and voice excitation. The spectrum flattener of the conventional voice-excited vocoder is replaced by a pitch pulse generator and a voice-unvoiced detector. Responsive to the received uncoded baseband signal, a sequence of excitation pulses is generated having a repetition rate corresponding to the fundamental frequency of the speech signal. Thus, one pulse per pitch period of the original speech wave is generated. This excitation signal, after appropriate filtering, closely resembles the damped sinusoid of voiced speech waves. The sequence of excitation pulses is selectively combined with random noise signals and the resultant signal is applied as an excitation function to the vocoder synthesizer channel control modulators. The continuous presence of random noise signals has been found to improve the naturalness and quality of the synthesized speech. Thus, by the practice of this invention, synthesized speech is obtained with higher quality and greater naturalness than that available in voice-excited and conventional vocoders.

The invention may be more fully understood from the following detailed description of an illustrative embodiment thereof taken in connection with the appended drawing, in which:

The single figure is a schematic block diagram of apparatus for transmitting telephone quality speech signals over a reduced bandwidth channel with a negligible loss of intelligibility or naturalness.

In the drawing, there is shown a source of telephone quality speech signals, for example, a telephone transmitter 1, which may be of any conventional construction. A 3,600 cycles per second bandwidth speech signal originating in the telephone transmitter is applied in parallel to band-pass filter 100 and to band-pass filters 100a through 100n of speech analyzer 10. The pass bands of filter 100 and filters 100a through 100n are chosen to

divide the speech signal into a relatively narrow band of low-frequency components, and a relatively wide band of high-frequency components, respectively. The lower limit of the narrow band of low-frequency components, or baseband, is set at the highest low-frequency cut-off point of commercial telephone circuits, approximately 250 cycles per second. The upper limit is set at a frequency that will insure that the baseband contains accurate information regarding the fundamental pitch frequency of a wide range of typical human voices; for example, the upper limit is set at about 925 cycles per second. The relatively wide band of high-frequency components extends from 925 cycles per second to 3,800 cycles per second, and filters 100a through 100n subdivide this relatively wide band into a number of contiguous subbands whose bandwidths are sufficiently small to define with accuracy the individual high-frequency components of the speech signal; for example, the contiguous subbands from 925 cycles per second to 3,000 cycles per second may have bandwidths of 150 cycles per second; while the contiguous subbands from 3,000 cycles per second to 3,800 cycles per second may have somewhat broader bandwidths to produce a total of approximately  $n=15$  subbands of high-frequency components.

To the output terminal of each band-pass filter 100a through 100n there is connected a half-wave rectifier, 101a through 101n, followed by a low-pass filter 102a through 102n, where the cut-off frequency of each low-pass filter is about 25 cycles per second. The output signal of each low-pass filter is a reduced bandwidth control signal whose instantaneous magnitude is representative of the instantaneous amplitudes of the high-frequency components in its associated subband. The total bandwidth of the baseband and the reduced bandwidth control signals is on the order of 1,200 cycles per second, representing a one-third reduction of the 3,600 cycles per second bandwidth of the original signal from telephone transmitter 1.

By suitably multiplexing the baseband and the control signals in a conventional multiplexer 120, the speech signal originating in transmitter 1 may be transmitted in modified form over a reduced bandwidth transmission channel, as indicated in FIG. 1. Before multiplexing, however, the baseband output of filter 100 is passed through a conventional delay element, illustrated by element 103 of analyzer 10, to compensate, as required, for the delay introduced by low-pass filters 102a through 102n in deriving the group of reduced bandwidth control signals from the high-frequency components of the speech signal. Of course, multiplexing need not be used; any other method or mode of transmission may be suitable.

At the receiver station, the multiplexed signals are separated by a suitable distributor 121, and the baseband is applied in parallel to pitch pulse generator 111 and to delay element 117, of speech synthesizer 11, while the control signals are applied to the control terminals of modulators 115a through 115n of the synthesizer. The baseband serves two purposes: it indirectly furnishes a group of excitation signals for reconstruction of the high-frequency components of the original speech signal; and it is directly combined with the reconstructed high-frequency components to form a replica of the original speech signal.

To derive the group of excitation or fundamental pitch pulse signals, the baseband is applied to pitch pulse generator 111. Simultaneously, the energy content of the high-frequency component control signals is sampled by summing amplifier 129 and a signal proportional to the total energy thereof is applied to voiced-unvoiced detector 112. It is well known that during voiced portions of speech, energy tends to be concentrated in the low-frequency speech components, while during unvoiced portions of speech, energy tends to be concentrated in the high-frequency speech components. Detector 112 compares the energy content of the high-frequency components with a signal, developed by generator 111, which is proportional

to the energy content of the low-frequency or baseband components. A signal is therefore developed by detector 112 characterized by two amplitude levels: the first corresponding to a voiced signal and the second corresponding to an unvoiced signal. This binary signal is applied to logic gate circuit 118.

Pitch pulse generator 111, in addition to generating a signal proportional to the energy content of the baseband signal, i.e., for application to detector 112, also generates a train of pulse signals having a repetition rate corresponding to the fundamental pitch of the speech signal. This train of pulse signals serves as the excitation signal for the control channel modulators 115a through 115n, of speech synthesizer 11. By using one pulse per pitch period as the excitation signal, in accordance with the present invention, synthesized speech is obtained of a higher quality and greater naturalness.

To prevent erroneous excitation signals from being applied to the channel modulators during unvoiced, i.e., aperiodic portions of the speech wave, the binary voiced-unvoiced signal, previously referred to, is employed to gate the pitch period pulses applied from generator 111 to logic circuit 118, thereby blocking spurious pulses generated during unvoiced portions of the speech wave.

Any of the multitudinous pitch generators and voiced-unvoiced detectors developed in the past for use with conventional channel vocoders, and well known to those skilled in the art, may be used in the apparatus of the present invention. For example, the apparatus for deriving pitch information from a speech wave disclosed in Patent 3,020,344, issued to A. J. Prestigiacomo on Feb. 6, 1962, is suitable for use in the present invention.

Thus, a train of pulses, having a predetermined relationship to the baseband signal, is applied via gate circuit 118 to mixer 119, e.g., a hybrid network. An output noise signal of generator 122, which may be of any well-known type, is continuously applied to mixer 119. In the absence of a voiced signal, gate 118 is inhibited by the signals emanating from detector 112, and the noise signal output of generator 122 is applied to the channel modulators to provide an aperiodic excitation function. When the speech signal is periodic, i.e., voiced, the pitch pulse train is mixed with the output of noise generator 122 and applied to modulator 115a through 115n. The amplitude of the noise signal is adjusted so as to be insignificant in comparison with the amplitude of the pitch pulses. However, in the absence of the pitch pulses the noise signal, of course, is predominant. The continuous presence of a noise signal, though at a reduced level, has been found to improve the naturalness and quality of the synthesized speech.

The derived excitation signal is applied in parallel to a bank of band-pass filters 113a through 113n, whose pass bands are identical with the pass bands of filters 100a through 100n of analyzer 10. Filters 113a through 113n divide the spectrum of the excitation signal into subbands identical in frequency with the subbands into which the band of high-frequency components is subdivided at the transmitter station. Each subband of signals from filters 113a through 113n is passed to an infinite clipper 114a through 114n, respectively, and each infinite clipper, which may be of any desired sort, generates from the components of each subband an excitation function for the reconstruction of one of the subbands of high-frequency components of the original speech signal.

Each excitation signal is applied to the input terminal of one of the conventional modulators 115a through 115n, and each of the reduced bandwidth control signals from distributor 121 is applied to the control terminal of one of the modulators. The control signals adjust the amplitudes of the excitation signals and the amplitude-adjusted excitation signals are filtered by band-pass filters 116a through 116n, which have pass bands identical with filters 100a through 100n of analyzer 10, to reconstruct

the high-frequency subbands of the original speech signal.

A replica of the original speech signal is synthesized by combining the reconstructed high-frequency subbands with the delayed baseband output of element 117, where the delay of the baseband is made equal to the delay introduced in the reconstruction of the high-frequency subbands. The synthesized speech is converted into highly intelligible, natural sounding speech by reproducer 12, for example, a conventional telephone receiver.

It is to be understood that the above-described arrangements are merely illustrative of applications of the principles of the invention. Numerous other arrangements may be devised by those skilled in the art without departing from the spirit and scope of the invention. For example, variable resonators, similar to those used in the well-known resonance vocoder, may be incorporated with the pitch pulse generator to reduce the peak factor of the excitation signal.

What is claimed is:

1. Apparatus for the reduced bandwidth transmission of speech signals which comprises:
  - a source of a speech signal,
  - means for dividing said speech signal into a relatively narrow band of selected low-frequency components and a relatively wide band of selected high-frequency components,
  - means for deriving from said band of high-frequency components a plurality of reduced bandwidth control signals representative of a corresponding plurality of selected subbands of said band of high-frequency components,
  - means for transmitting said band of low-frequency components of said speech signal and said control signals to a receiver station,
  - means at said receiver station for generating from said band of low-frequency components of said speech signal a sequence of pulse signals separated by intervals corresponding to the fundamental period of said speech signal,
  - means responsive to said sequence of pulse signals and said control signals for reconstructing said high-frequency components of said speech signal,
  - and means for combining said low-frequency components of said speech signal with said reconstructed high-frequency components to synthesize a replica of said speech signal.
2. Apparatus for transmitting a telephone quality speech signal over a reduced bandwidth channel with a negligible loss of intelligibility or naturalness which comprises:
  - a source of a telephone quality speech signal,
  - means for dividing said speech signal into a narrow band of selected low-frequency components and a wide band of selected high-frequency components,
  - means for deriving from said band of high-frequency components a plurality of reduced bandwidth control signals representative of the amplitudes of a corresponding plurality of selected subbands of said band of high-frequency components,
  - means for transmitting said band of low-frequency components of said speech signal and said plurality of control signals to a receiver station,
  - means at said receiver station for generating from said band of low-frequency components of said speech signal an excitation signal characterized by one pulse for each pitch period of said speech signal,
  - means responsive to said transmitted low-frequency components and said plurality of control signals for developing a signal representative of the voice-unvoiced nature of said speech signal,
  - a source of noise signals,
  - means jointly responsive to said excitation signal and said representative signal for selectively combining said excitation signal and said noise signals,
  - means jointly responsive to said control signals and said

combined signal for developing replicas of said subbands of high-frequency components, and means for combining said band of low-frequency components with said replicas to generate a synthesized speech signal.

3. Voice-excited vocoder synthesizer apparatus jointly responsive to applied baseband components of a speech signal and control signals representative of the energy content of selected spectrum bands of said speech signal comprising:

means responsive to said baseband components for generating pulse signals separated by an interval corresponding to the fundamental period of said speech signal,

means responsive to said baseband components and said control signals for developing a signal indicative of the voiced-unvoiced nature of said speech signal,

means responsive to said indicative signal for selectively transmitting said pulse signals,

means for generating noise signals,

means for combining said transmitted pulse signals and said noise signals,

means for selectively filtering said combined signals,

means jointly responsive to said filtered signals and said control signals for reproducing said selected bands of said speech signal,

and means responsive to said baseband components and said reproduced bands of speech for developing a synthesized speech signal.

4. Speech synthesis apparatus comprising:

means responsive to applied baseband components of a speech signal for generating an excitation function characterized by pulse signals having a periodicity corresponding to the fundamental pitch of said speech signal,

means responsive to said applied baseband components and applied signals representative of the energy content of higher subbands of said speech signal for generating an indication of the form of said speech signal,

means responsive to said generated indication for selectively transmitting said pulse signals,

means for mixing said transmitted pulse signals with random noise signals,

means responsive to said mixed signals and said representative higher subband signals for reconstructing the higher subbands of said speech signal,

and means jointly responsive to said reconstructed higher subbands and said baseband components for developing a synthesized replica of said speech signal.

5. In a speech synthesizing system the combination that comprises:

a source of a narrow band of low-frequency components of a speech signal,

a source of reduced bandwidth control signals representative of the amplitudes of subbands of high-frequency components of said speech signal,

means for generating from said low-frequency components a pulse signal having a repetition rate corresponding to the fundamental pitch of said speech signal,

means under the influence of said control signals and said pulse signal for constructing replicas of said high-frequency subbands of said speech signal,

and means for combining said low-frequency components of said speech signal with said replicas of said high-frequency subbands of said speech signal.

6. Speech synthesizing apparatus comprising:

a source of a narrow band of low-frequency components of a speech signal,

a source of reduced bandwidth control signals representative of the amplitudes of subbands of high-frequency components of said speech signal,

means for generating from said low-frequency components a periodic signal having a repetition rate corre-

7

sponding to the fundamental pitch of said speech  
signal,  
a source of random noise signals,  
means for selectively combining said periodic signal  
and said random noise signals, 5  
means under the influence of said control signals and  
said combined signals for constructing replicas of  
said high-frequency subbands of said speech signal,  
and means for combining said low-frequency compo- 10  
nents of said speech signal with said replicas of said  
high-frequency subbands of said speech signal.

8

References Cited

UNITED STATES PATENTS

|           |        |              |             |
|-----------|--------|--------------|-------------|
| 3,030,450 | 4/1962 | Schroeder    | 179—15.55   |
| 3,139,487 | 6/1964 | Logan et al. | 179—15.55   |
| 3,325,596 | 6/1967 | Manley       | 179—15.55 X |

ROBERT L. GRIFFIN, *Primary Examiner.*

W. S. FROMMER, *Assistant Examiner.*

U.S. Cl. X.R.

179—1