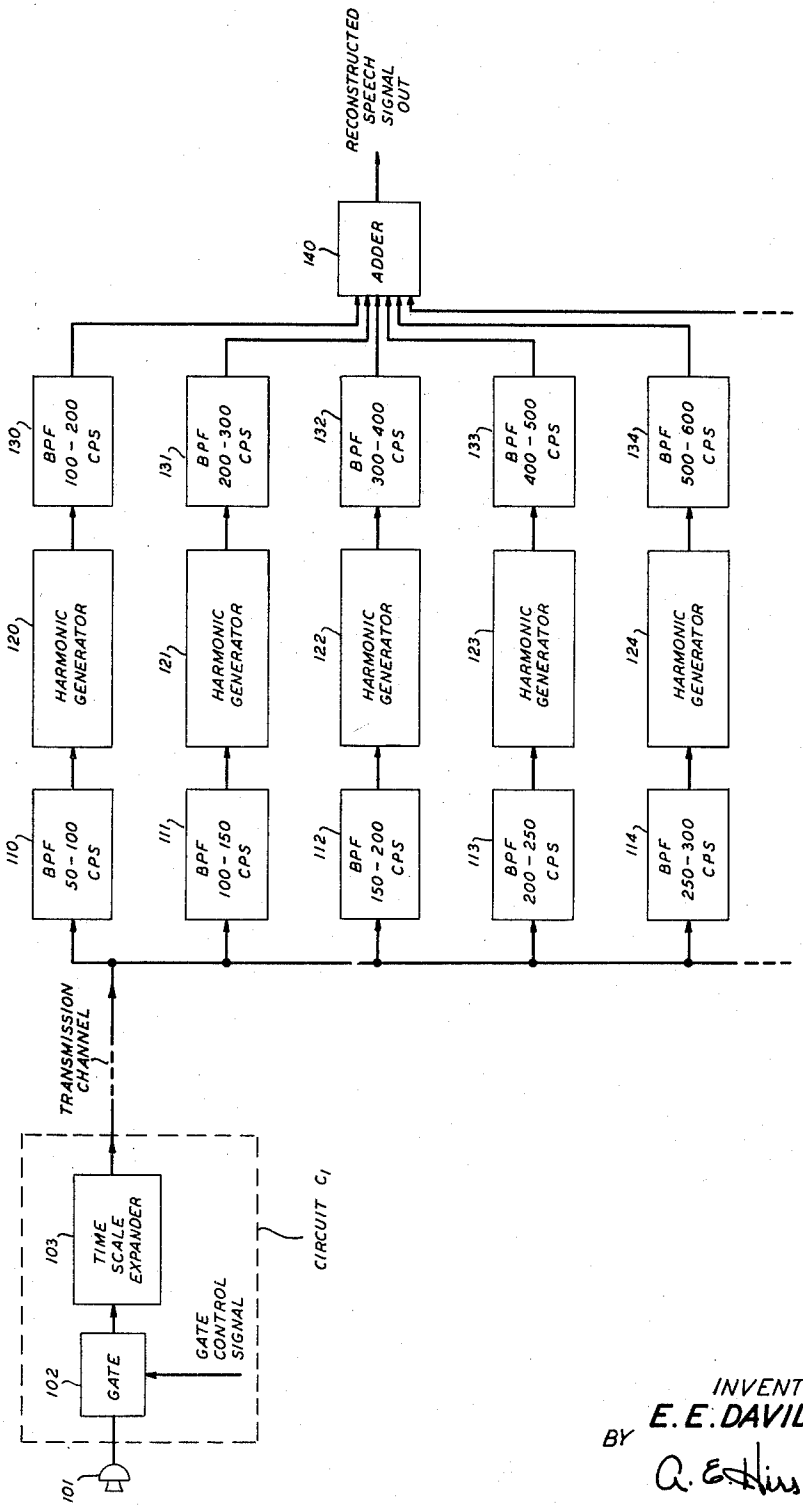
March 31, 1964 E. E. DAVID, JR 3,127,476

ARTIFICIAL RECONSTRUCTION OF SPEECH

Filed Feb. 2, 1960 3 Sheets-Sheet 1

FIG. 1



INVENTOR
E. E. DAVID, JR.
BY A. E. Hirsch Jr.
ATTORNEY

March 31, 1964 E. E. DAVID, JR 3,127,476

ARTIFICIAL RECONSTRUCTION OF SPEECH

Filed Feb. 2, 1960 3 Sheets-Sheet 2

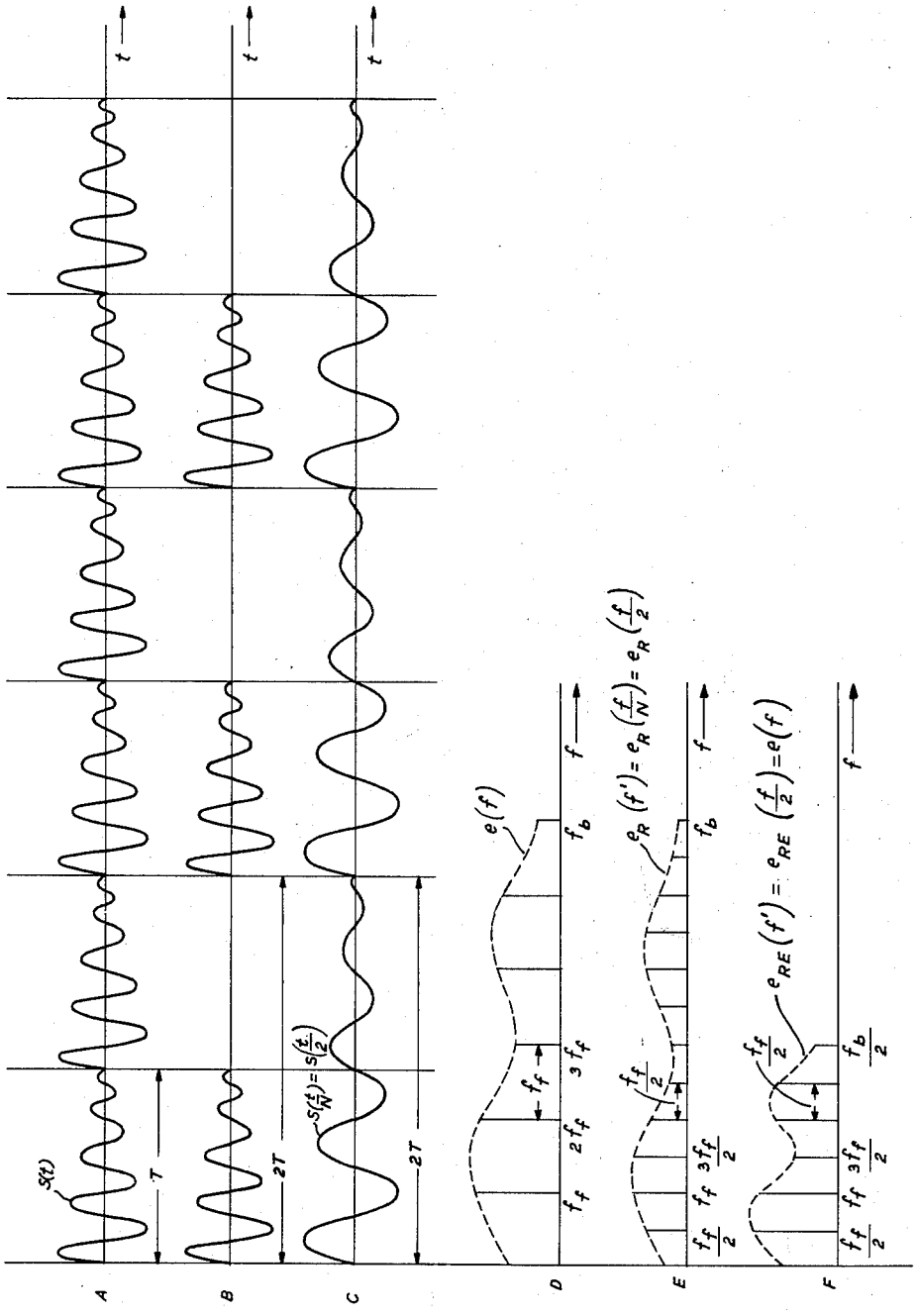


FIG. 2

INVENTOR
BY E. E. DAVID, JR.
C. E. Hirsch Jr.
ATTORNEY

March 31, 1964 E. E. DAVID, JR 3,127,476

ARTIFICIAL RECONSTRUCTION OF SPEECH

Filed Feb. 2, 1960 3 Sheets-Sheet 3

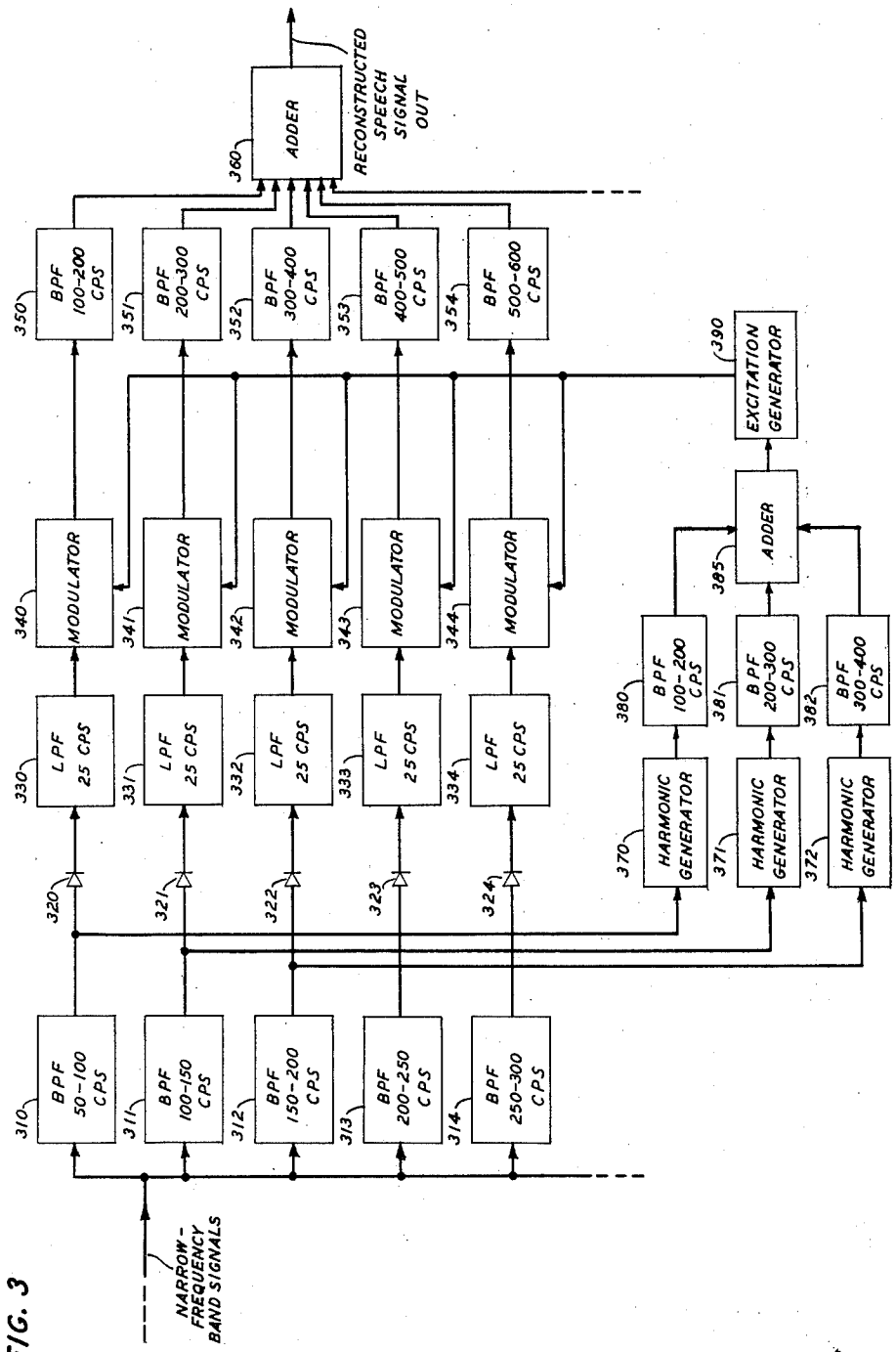


FIG. 3

INVENTOR
E. E. DAVID, JR.
BY A. E. Hirsch Jr.
ATTORNEY

United States Patent Office 3,127,476 Patented Mar. 31, 1964

3,127,476

ARTIFICIAL RECONSTRUCTION OF SPEECH

Edward E. David, Jr., Berkeley Heights, N.J., assignor to Bell Telephone Laboratories, Incorporated, New York, N.Y., a corporation of New York

Filed Feb. 2, 1960, Ser. No. 6,301
13 Claims. (Cl. 179—15.55)

This invention relates to the reconstruction of artificial speech from narrow-band transmitted signals, and has for its principal object the improvement of quality of such artificial speech.

The almost periodic nature of voiced speech has formed the basis of several systems that reduce the periodic redundancy of speech, and, by so doing, are able to transmit the information content of a speech signal over a smaller band of frequencies than is required for the transmission of the entire speech signal. Systems of this type, which shall be referred to hereinafter as pitch synchronous processing systems, are described in Patent 2,115,803, granted to H. W. Dudley on May 3, 1938, in Patent 2,860,187, granted November 11, 1958, to E. E. David, Jr., et al., and elsewhere. In such systems, an electrical speech signal is divided at the transmitter station into a sequence of sections in synchrony and in phase with its fundamental period. Each section contains N periods, and $N-1$ of these periods are eliminated in each section. The one retained period in each section is expanded in the time dimension to fill the gap left by the eliminated periods, thus reducing the width of the frequency band necessary for its transmission. After transmission over a narrow-frequency band channel, an artificial speech signal is reconstructed at the receiver station from the reduced bandwidth periods.

Both the Dudley and David et al. systems referred to above perform a period-by-period reconstruction of speech as a function of time at the receiver station: both systems restore the retained periods of the original speech signal by compressing the time scale of the transmitted periods, and both systems fill the blank intervals created by compression with $N-1$ artificial periods.

Filling the blank intervals, however, causes distortion in the artificial speech reconstructed by both the Dudley and the David et al. systems, the distortion arising primarily from discontinuities at the junctions of the compressed periods and the periods inserted in the intervals. This distortion is especially severe in transient portions of speech where either the voice pitch or the formant structure is changing, and it is evident subjectively as a low-frequency, speech-correlated "gargle" or "rumble."

It is a specific object of the present invention to reduce distortion in artificially reconstructed speech by eliminating the above discontinuities.

The objects of the present invention are achieved by reconstructing a frequency domain replica of the original speech signal, instead of a period-by-period, time domain replica. The apparatus of the invention operates upon the frequency characteristics of the transmitted periods, without compressing them in time and without creating blank intervals of any kind. Since there are no blank intervals to be filled, the discontinuities referred to above are eliminated, with a consequent improvement in the subjective quality of the reconstructed speech.

There are other systems that also operate in the frequency domain to reconstruct speech from narrow-band signals, for example, the channel vocoder system described in H. W. Dudley Patent 2,151,091, issued March 21, 1939, and a modification thereof described in patent application of M. R. Schroeder, Serial No. 774,173, filed November 17, 1958, now Patent No. 3,030,450. But in their present form these systems are unable to synthesize speech from the narrow-band signals produced by pitch synchronous processing.

It is a further object of the present invention to reconstruct artificial speech in the frequency domain solely from the narrow-band signals created by pitch synchronous processing.

The present invention analyzes the frequency characteristics of both the original signal and the reduced bandwidth signal produced by pitch synchronous processing in terms of the frequency components which completely specify the respective signals. The analysis reveals an important set of relationships between the components of the two signals: the components of the two signals are in a one-to-one correspondence such that the amplitudes of corresponding components are equal, and the frequencies of corresponding components differ only by a constant factor N.

Apparatus for reconstructing artificial speech based upon these relationships first obtains the frequency components of the reduced bandwidth signal at the receiver station. Harmonics of each component are then produced with amplitudes equal to the amplitude of the individual component, and therefore equal also to the amplitude of the corresponding component of the original signal. Finally, by selecting the Nth harmonic of each set of harmonics, a group of frequency components equal in both frequency and amplitude to the corresponding components of the original signal is obtained. Since a speech signal is completely specified by its frequency components, the group of Nth harmonics thus selected defines an accurate replica of the original speech signal.

The invention will be fully apprehended from the following detailed description of preferred embodiments thereof taken in connection with the appended drawings, in which:

FIG. 1 is a block schematic diagram showing a system embodying a preferred form of the invention;

FIG. 2 is a group of waveform diagrams of assistance in explaining the operation of FIG. 1; and

FIG. 3 is a block schematic diagram showing another embodiment of the invention.

Before describing the apparatus of the present invention, it is desirable to examine analytically the time domain and the frequency domain representations of voiced speech both before and after pitch synchronous processing, in order that the principles embodied in the invention will be more readily understood.

Voiced speech is represented in the time domain by a periodic function of time, $S(t)$, with periods of length T. A typical voiced portion of speech is shown in curve A of FIG. 2.

Voiced speech is represented in the frequency domain by a Fourier spectrum consisting of frequency components that are harmonics of the fundamental pitch frequency, f_f, and whose amplitudes are a function of frequency, represented by $e(f)$, $f=0, f_f, 2\cdot f_f, 3\cdot f_f \ldots$. Curve D of FIG. 2, which shows the Fourier spectrum of the periodic function $S(t)$ of curve A, consists of a number of regularly spaced lines of different amplitudes, the lines occurring at harmonics of the fundamental pitch frequency. Curve D also illustrates the band-limited case to be considered in the following discussion, in which the spectral amplitudes are defined as zero for frequencies greater than f_b, the cut-off frequency, that is

$$e(f)=0 \text{ for } f>f_b \tag{1}$$

The time domain and the frequency domain representations of voiced speech are related to each other in the following manner:

The fundamental pitch frequency is the reciprocal of the fundamental pitch period,

$$f_f=\frac{1}{T} \tag{2}$$

and the amplitude of each spectral component is given by the Fourier theorem,

$$e(f)=\frac{1}{T}\int_0^T S(t)e^{-2\pi ift}dt,\ f=0,\ f_f,\ 2\cdot f_f,\ 3\cdot f_f\ \ldots \quad (3)$$

Referring now to FIG. 1, an electrical speech signal, generated, for example, in microphone **101**, is applied to circuit C_1 at a transmitter station. The principal elements of circuit C_1 are gate **102** and time scale expander **103**, connected in tandem. The details of the circuit C_1 are well known, and an entirely suitable one is shown in FIG. 1 of the David et al. patent referred to above.

Circuit C_1 operates both to reduce the time domain redundancy of voiced speech, which appears as a repetition of identical or almost identical periods of the speech function, $S(t)$, and to reduce the frequency bandwidth necessary to transmit voiced speech. As the following analysis demonstrates, the operations of circuit C_1 also produce important relationships between the spectral components of the original signal and the spectral components of the reduced redundancy signal, which form the basis of the present invention.

Gate **102** of circuit C_1 reduces the redundancy of voiced speech by eliminating $N-1$ of every N periods of the signal from microphone **101**, in response to a gate control signal. Curve B of FIG. 2 shows the result of passing the speech signal shown in curve A through gate **102** for the case in which the gate control signal is selected to eliminate one or every two periods of $S(t)$, that is, N is selected to be equal to two.

The elimination of $N-1$ of every N periods by gate **102** lengthens the fundamental pitch period from T to $N\cdot T$ in the time domain, and, by Equation 2, decreases the fundamental pitch frequency from

$$\frac{1}{T}\text{ to }\frac{1}{N\cdot T}=\frac{f_f}{N}$$

in the frequency domain.

Since the spectral components of a voiced speech signal are harmonics of the fundamental frequency, the reduction in fundamental frequency by a factor N is accompanied by an N-fold increase in the number of spectral components. This is graphically illustrated by curve E of FIG. 2, in which N is equal to two.

The elemination of $N-1$ of every N periods by gate **102** also reduces the energy of the speech signal by a factor N, which is reflected in smaller amplitudes of the spectral components of the retained periods, shown in curve E of FIG. 2. The amplitudes of the spectral components of the retained periods are given by a new function of frequency e_R, in which

$$e_R(f')=\frac{1}{NT}\int_0^{NT} S(t)e^{-2\pi if't}dt,\ f'=0,\ \frac{f_f}{N},\ \frac{2\cdot f_f}{N},\ \frac{3\cdot f_f}{N}\ \ldots \quad (4)$$

where

$$f'=\frac{f}{N}$$

is the new frequency variable.

The retained periods passed by gate **102** are applied to time scale expander **103**, where they are expanded over the $N-1$ blank intervals left by the eliminated periods. Heretofore, expander **103** has been conceived of primarily as a device for reducing the bandwidth necessary for the transmission of the retained periods. As shown in mathematical terms below, expander **103** produces two other equally important effects: it increases the energy of the retained periods by a factor N, thereby exactly canceling the N-fold energy reduction brought about by gate **102**; and it reduces by a factor N the number of spectral components of the retained periods, thereby exactly canceling the N-fold increase in components brought about by gate **102**. The action of expander **103** therefore produces a one-to-one correspondence between the spectral components of the original signal and the reduced bandwidth signal, in which corresponding components are of equal amplitudes. Thus the net effect of the time domain operations of circuit C_1 upon the frequency characteristics of a periodic speech signal can be summarized as a change in frequency scale by a factor

$$\frac{1}{N}$$

By operating upon the reduced bandwidth signal to restore the proper frequency scale, the present invention synthesizes a replica of the original signal.

By expanding the retained periods over periods of length $N\cdot T$, time scale expander **103** produces a new periodic function,

$$S\left(\frac{t}{N}\right)$$

with periods of length $N\cdot T$, shown in curve C of FIG. 2. By substitution in Equation 4, the amplitudes of the spectral components of the expanded periodic function

$$S\left(\frac{t}{N}\right)$$

are given by $e_{RE}(f')$, where

$$e_{RE}(f')=\frac{1}{NT}\int_0^{NT} S\left(\frac{t}{N}\right)e^{-2\pi if't}dt,\ f'=0,\ \frac{f_f}{N},\ \frac{2\cdot f_f}{N},\ \frac{3\cdot f_f}{N}\ \ldots \quad (5)$$

By the substitution

$$\frac{t}{N}=\tau$$

$dt=N\cdot d\tau$, and the relation

$$f_f=\frac{f}{N}$$

Equation 5 can be reduced to the simpler form

$$e_{RE}\left(\frac{f}{N}\right)=\frac{1}{T}\int_0^T S(\tau)e^{-2\pi if\tau}d\tau=e(f),\ f=0,\ f_f,\ 2f_f,\ 3f_f,\ \ldots \quad (6)$$

Hence Equation 6 reveals the important relationships between the spectral components of the reduced bandwidth signal and the original signal: (*a*) one-to-one correspondence; (*b*) equality of amplitude between corresponding components; and (*c*) the frequency scale factor

$$\frac{1}{N}$$

In a preferred form of the invention shown in FIG. 1, the reduced bandwidth signal produced by circuit C_1 is transmitted over a narrow-band channel to a receiver station, where it is applied to a number of relatively narrow-band filters, only a few of which are shown in the figure. In the embodiment shown in FIG. 1, N is assumed to be equal to two, but the invention is suitable for use with speech signals reduced in bandwidth by factors other than two. Each filter **110**, **111** . . . has a suitably narrow passband in order to pass a single frequency sine wave corresponding in frequency and in amplitude to a single spectral component of the reduced bandwidth signal. For example, a passband of 50 cycles per second is satisfactory. Since 20 or more spectral components are necessary for good quality speech, 20 or more filters of this bandwidth are required. It is to be understood, however, that sufficient spectral components for good quality speech may be obtained by employing, if desired, either fewer filters with passbands broader than those shown in FIG. 1, or a greater number of filters with passbands narrower than those shown in FIG. 1.

The sine wave outputs of filters **110**, **111** . . . are applied to linear harmonic generators **120**, **121** . . . respectively. Each generator, which may take any structural form well known in the art, produces harmonics of

the frequency of the sine wave applied to it, with amplitudes equal to the amplitude of the input sine wave. Filters **130, 131** . . . connected to the output terminals of generators **120, 121** . . . respectively, select the Nth harmonic from the harmonics produced by each generator to form a set of components whose frequencies are harmonics of the fundamental frequency of the original signal. For example, the frequency of the sine wave from filter **110** is

$$\frac{f_t}{N}=\frac{f_t}{2}$$

the harmonics produced by generator **120** are

$$1\cdot\left(\frac{f_t}{2}\right),\ 2\cdot\left(\frac{f_t}{2}\right),\ 3\cdot\left(\frac{f_t}{2}\right)\cdots$$

and filter **130** selects the Nth or second harmonic f_t. Similarly, the frequency of the sine wave from filter **111** is

$$2\cdot\left(\frac{f_t}{N}\right)=2\cdot\left(\frac{f_t}{2}\right)=f_t$$

the harmonics produced by generator **121** are

$$1\cdot f_t, 2\cdot f_t, 3\cdot f_t\ \ldots$$

and filter **131** selects the Nth or second harmonic, $2\cdot f_t$. In like fashion, filter **132** selects the Nth or second harmonic, $3\cdot f_t$, from the harmonics produced by generator **122**. Since f_t is the fundamental frequency of the original speech signal, the Nth harmonics selected by filters **130, 131** . . . are harmonics of f_t.

In addition, since the amplitudes of the Nth harmonics thus selected are equal to the amplitudes of the spectral components of the reduced bandwidth signal, they are also equal to the amplitudes of the spectral components of the original signal. Hence the combination of the output signals of filters **130, 131** . . . obtained, for example, by adder **140**, is a set of spectral components that approximates the Fourier spectrum of the original speech signal and therefore defines a replica of the original speech signal.

The filtering action of filters **130, 131** . . . is improved by choosing the characteristics of harmonic generators **120, 121** . . . to correspond to the odd or even value of N. For example, if N is odd, a peak sampler that samples both positive and negative peaks of the input sine wave is appropriate as a harmonic generator, since it produces odd harmonics of the input frequency; that is, if the input frequency is f, the harmonics generated are f, $3f$, $5f$ . . . If N is even, a full-wave rectifier is appropriate as a harmonic generator, since it produces even harmonics, $2f$, $4f$, $6f$ . . . of the input frequency, f.

Although the apparatus of FIG. 1 has been described in connection with voiced speech signals, it is also suitable for use with unvoiced speech signals, if desired. Unvoiced speech signals are processed by circuit C_1, the processed signals are transmitted to the receiver station, and artificial unvoiced speech signals are reconstructed from the transmitted signals at the receiver station.

Another form of the invention is illustrated in the apparatus of FIG. 3. From the narrow-frequency band signals produced by pitch synchronous processing, there is derived an excitation signal and a set of control signals. The control signals are utilized to adjust the energy of the excitation signal to produce a replica of the Fourier spectrum of the original speech signal.

Pitch synchronously processed speech is applied to a number of relatively narrow-band filters, a few of which are represented in FIG. 3 by filters **310, 311** . . . . The passbands of these filters are selected in exactly the same fashion as filters **110, 111** . . . of FIG. 1, and N is again assumed to be equal to two; hence a value of 50 cycles per second is sufficient to pass a single-frequency sine wave corresponding in frequency and in amplitude to a single spectral component of the input signal. The operation of the embodiment shown in FIG. 3 is not limited, however, to this value of N, nor is it limited to the passbands of the filters shown. Equally suitable filter passbands may be selected by those skilled in the art.

Control signals are derived from the output signals of filters **310, 311** . . . by rectifiers **320, 321** . . . connected in series with low-pass filters **330, 331** . . . . The values of the cut-off frequencies of the low-pass filters are of the order of 25 cycles per second. The instantaneous magnitudes of the control signals represent the instantaneous amplitudes of the spectral components of the reduced bandwidth signal; hence by Equation 6, they also represent the instantaneous amplitudes of the corresponding components of the original signal. The control signals are applied to modulators **340, 341** . . . where they adjust the energy of an excitation signal supplied to the modulators from excitation generator **390**.

The excitation signal must be closely correlated with the original speech signal in order to reconstruct an accurate replica of the Fourier spectrum of the original speech signal. Close correlation is achieved by employing an excitation generator **390**, of the type described, for example, in the Schroeder application referred to above, which generates a signal having a broad-band Fourier spectrum that is a function of the fundamental frequency of the input signal applied to it. An input signal that is a function of the fundamental frequency of the original speech signal is obtained from the output signals of several of the band-pass filters lowest on the frequency scale, for example, filters **310, 311, 312**. The output terminals of these filters are connected to a set of linear harmonic generators **370, 371, 372**, connected in tandem with a set of band-pass filters **380, 381, 382**, which elements serve to obtain harmonics of f_t from harmonics of

$$\frac{f_t}{N}$$

in exactly the same fashion as described in connection with the apparatus of FIG. 1. These harmonics of f_t, suitably combined, for example, in adder **385**, form an appropriate input signal for excitation generator **390**. The excitation output of generator **390** is supplied simultaneously to modulators **340, 341** . . . where the amplitudes of the spectral components of the excitation signal are adjusted by the control signals from low-pass filters **330, 331** . . . to values equal to the amplitudes of the components of the original signal. The product signals of modulators **340, 341** . . . are passed through band-pass filters **350, 351** . . . which are scaled in conventional channel vocoder fashion, and the spectral components thus obtained are appropriately combined, for instance, in adder **360**, to form a replica of the Fourier spectrum of the original speech signal.

As explained in connection with the apparatus of FIG. 1, the filtering action of filters **380, 381, 382** may be improved, if desired, by adapting the characteristics of the harmonic generators to the odd or even value of N. In addition, the apparatus of FIG. 3, like the apparatus of FIG. 1, may be used, if desired, to reconstruct artificial unvoiced signals as well as artificial voiced signals.

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.

What is claimed is:

1. In a system for the synthesis of speech, a source of narrow-frequency band signals derived from a speech signal by a pitch synchronous processing system and characterized by a plurality of spectral components, means for generating harmonics of the fundamental frequency of said speech signal from said spectral components of said narrow-frequency band signals, and means for combin-

ing said harmonics to synthesize a replica of said speech signal.

2. In a system for the transmission of speech, the combination that comprises a source of a speech signal, means for pitch synchronously processing said speech signal to form a narrow-frequency band signal, means for transmitting said narrow-frequency band signal to a receiver station, and means at said receiver station for developing from the spectral components of said narrow-frequency band signal a group of spectral components whose frequencies are harmonics of the fundamental frequency of the original speech signal, thereby to reconstruct a replica of the original speech signal.

3. In a system for the transmission of speech, the combination that comprises a source of a speech signal, means for dividing said speech signal into a sequence of time period sections, means for eliminating portions of said sections to leave retained portions and blank intervals, means for expanding said retained portions over said blank intervals, means for transmitting to a receiver station the expanded portions, means at said receiver station for synthesizing from the spectral components of said expanded portions artificial spectral components that closely approximate the spectral components of said original speech signal, and means for reconstructing a replica of the original speech signal from said artificial spectral components.

4. Speech transmission apparatus that comprises a source of a speech signal consisting of a sequence of periods, means for eliminating $N-1$ of each group of N successive periods to leave one retained period in each group, means for expanding each retained period over the N periods in its group, means for transmitting to a receiver station said expanded periods, means at said receiver station for obtaining the Nth harmonics of the spectral components of said expanded periods, and means for combining said Nth harmonics to synthesize a replica of said speech signal.

5. Apparatus for transmitting a speech signal that occupies a relatively wide-frequency range to a receiver station over a transmission channel that has a relatively narrow-frequency range, which comprises means for pitch synchronously processing the original speech signal to obtain narrow-frequency band signals, means for transmitting said narrow-frequency band signals to a receiver station, and, at said receiver station, means for obtaining the spectral components of said narrow-frequency band signals, means for synthesizing from said spectral components harmonics of the fundamental frequency of said original speech signal, and means for reconstructing a facsimile of the original speech signal from said harmonic components.

6. Apparatus as defined in claim 5 wherein said means for obtaining the spectral components of said narrow-frequency band signals comprises a set of selected narrow band-pass filters.

7. Apparatus as defined in claim 5 wherein said means for synthesizing harmonics of the fundamental frequency of said original speech signal comprises a set of harmonic generators connected in series with a set of selected narrow band-pass filters.

8. Apparatus as defined in claim 5 wherein said means for reconstructing a facsimile of the original speech signal from said harmonic components comprises a resistive adder for combining said harmonics.

9. Speech transmission apparatus which comprises a source of a speech signal, means for pitch synchronously processing said speech signal to form a reduced bandwidth signal whose fundamental frequency is

$$\frac{1}{N}$$

times the fundamental frequency of said speech signal, means for transmitting said reduced bandwidth signal to a receiver station, and, at said receiver station, means for obtaining the spectral components of said reduced bandwidth signal, means for generating harmonics of said spectral components, means for selecting the Nth harmonics of said generated harmonics, and means for reconstructing a replica of the original speech signal from said Nth harmonics.

10. Apparatus for the synthesis of speech which comprises a source of reduced bandwidth signals produced by a pitch synchronous processing system, means for deriving the spectral components of said reduced bandwidth signals, means for obtaining control signals representative of amplitude variations in said spectral components, means for generating an excitation signal from selected spectral components of said reduced bandwidth signal, and means responsive to said control signals for synthesizing an artificial speech signal from said excitation signal.

11. Apparatus for reconstructing an artificial speech signal from signals reduced in bandwidth by a factor N by a pitch synchronous processing system, which comprises means for obtaining selected spectral components of said reduced bandwidth signals, means for deriving control signals representative of amplitude variations of said selected spectral components, means connected in parallel with said control signal deriving means for generating harmonics from several of said selected spectral components, means for selecting the Nth harmonics of said generated harmonics, means for obtaining a speech excitation signal from said Nth harmonics, and means responsive to said control signals for reconstructing an artificial speech signal from said excitation signal.

12. Apparatus as defined in claim 11 wherein said means for deriving control signals comprises a set of linear rectifiers connected in tandem with a set of low-pass filters.

13. Apparatus as defined in claim 11 wherein said means for reconstructing an artificial speech signal from said excitation signal comprises a set of modulators connected in series with a set of selected band-pass filters.

References Cited in the file of this patent

UNITED STATES PATENTS

2,117,739	Miller	May 17, 1938
2,710,892	Dahlbom et al.	June 14, 1955
2,906,955	Edson et al.	Sept. 29, 1959
2,928,901	Bogert	Mar. 15, 1960