

Dec. 18, 1962

E. E. DAVID, JR

3,069,507

AUTOCORRELATION VOCODER

Filed Aug. 9, 1960

7 Sheets-Sheet 1

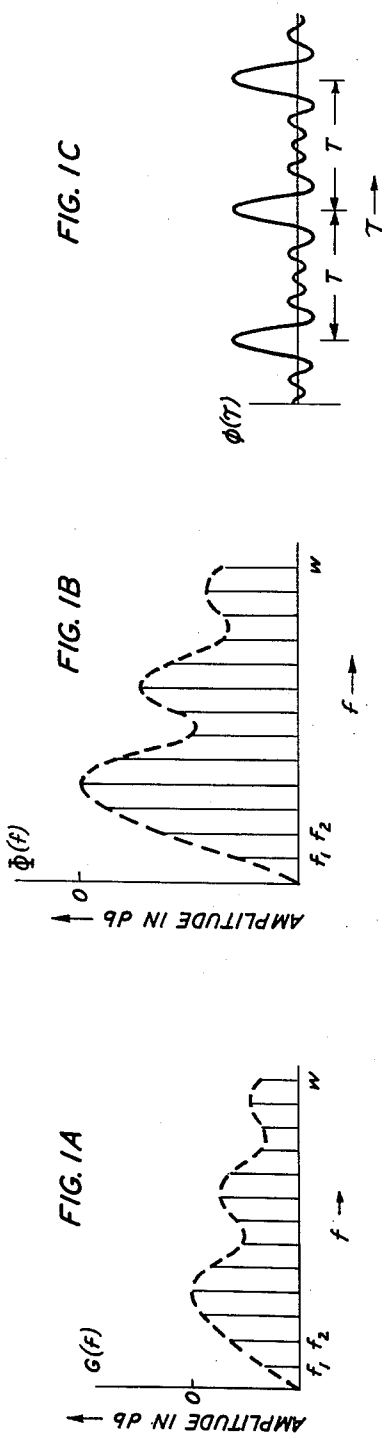
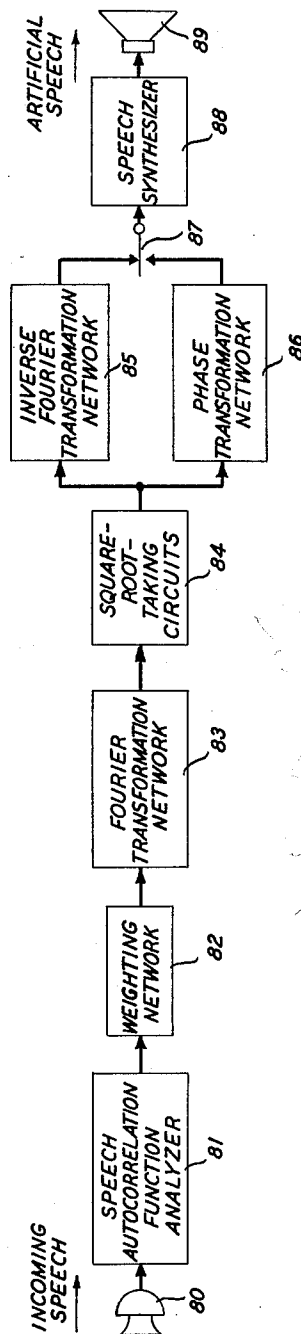


FIG. 1D



INVENTOR  
 E. E. DAVID, JR.  
 BY *A. E. Hirschfeld*  
 ATTORNEY

Dec. 18, 1962

E. E. DAVID, JR

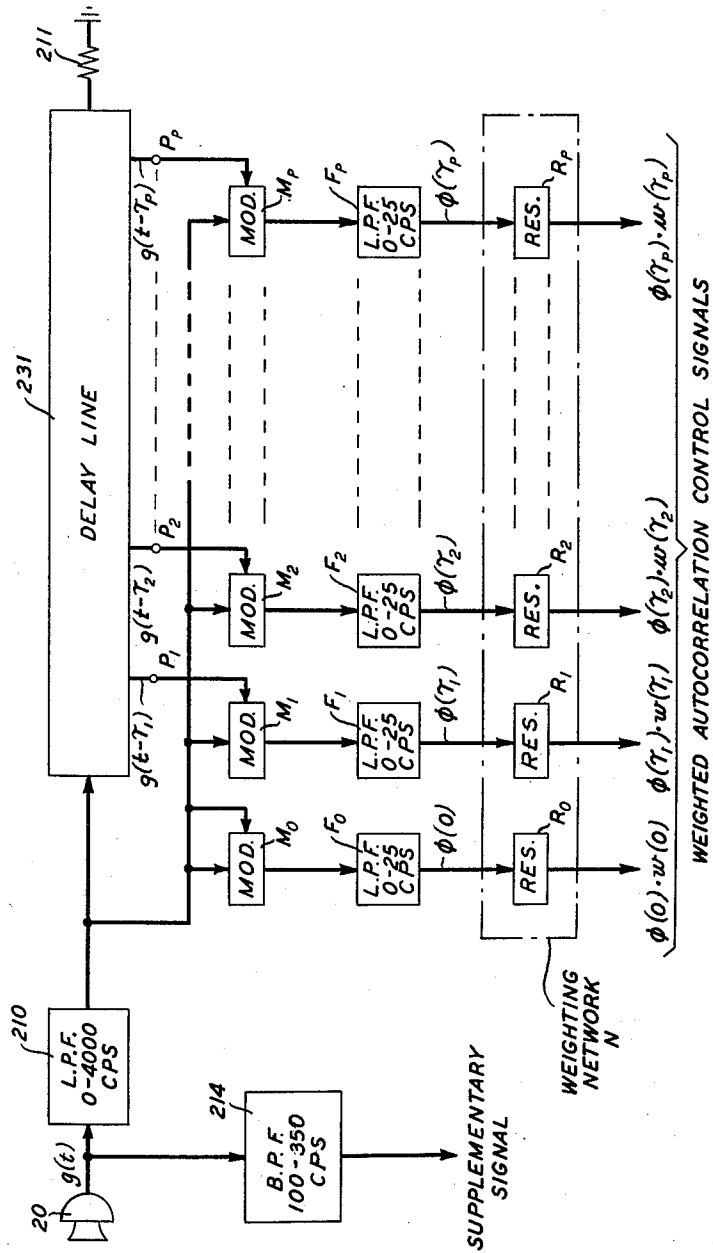
3,069,507

AUTOCORRELATION VOCODER

Filed Aug. 9, 1960

7 Sheets-Sheet 2

FIG. 2



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

Dec. 18, 1962

E. E. DAVID, JR  
AUTOCORRELATION VOCODER

3,069,507

Filed Aug. 9, 1960

7 Sheets-Sheet 3

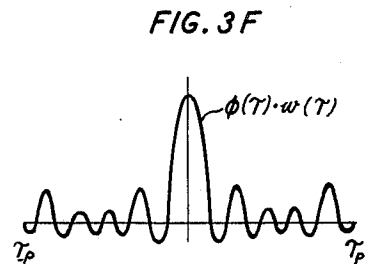
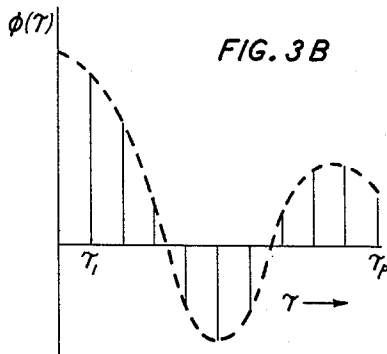
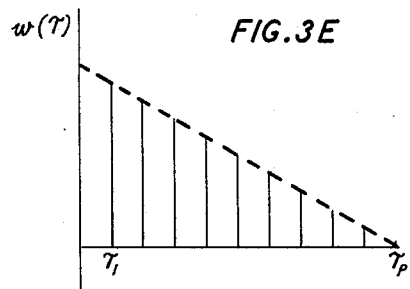
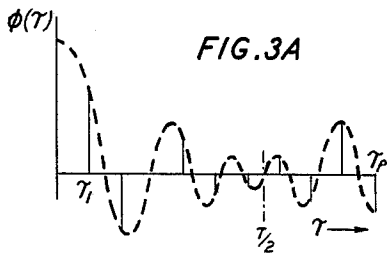


FIG. 3C

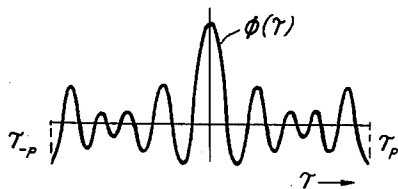


FIG. 3D

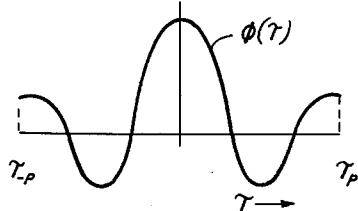
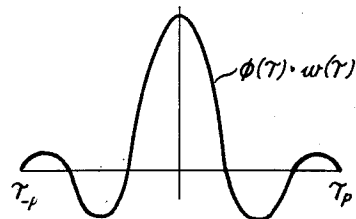


FIG. 3G



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

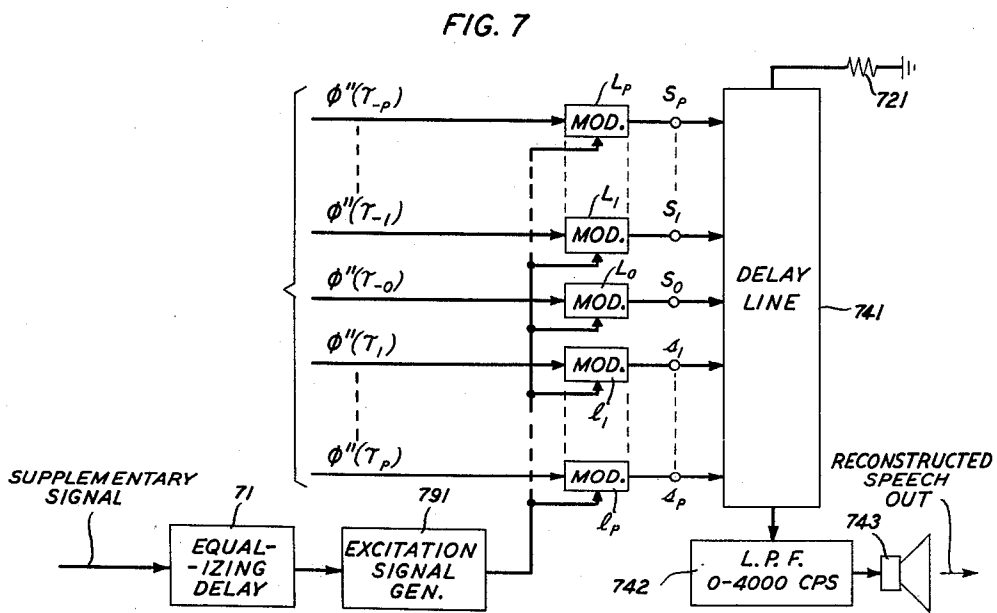
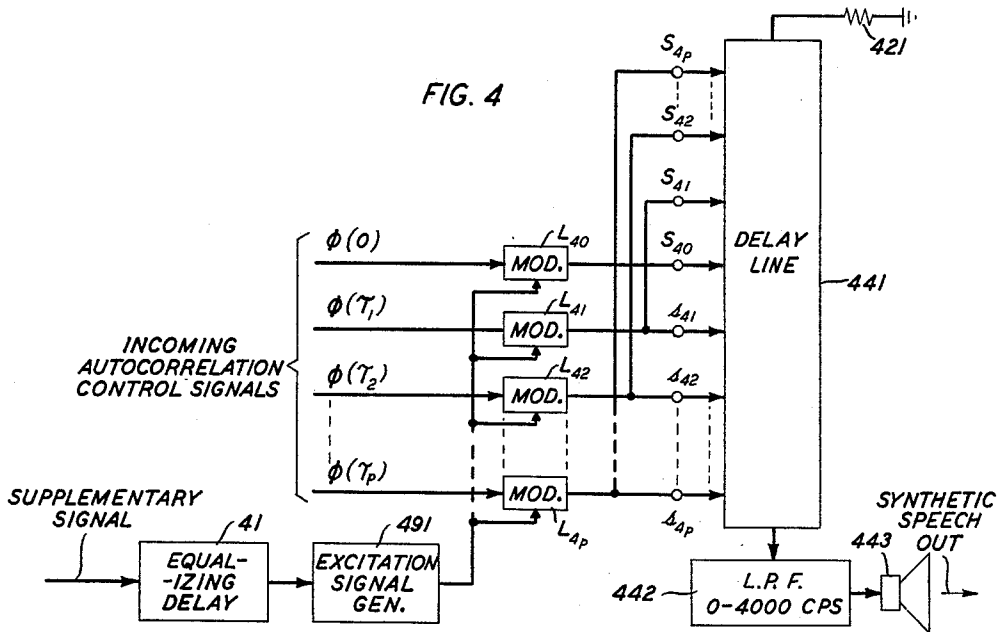
Dec. 18, 1962

E. E. DAVID, JR  
AUTOCORRELATION VOCODER

3,069,507

Filed Aug. 9, 1960

7 Sheets-Sheet 4



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

Dec. 18, 1962

E. E. DAVID, JR

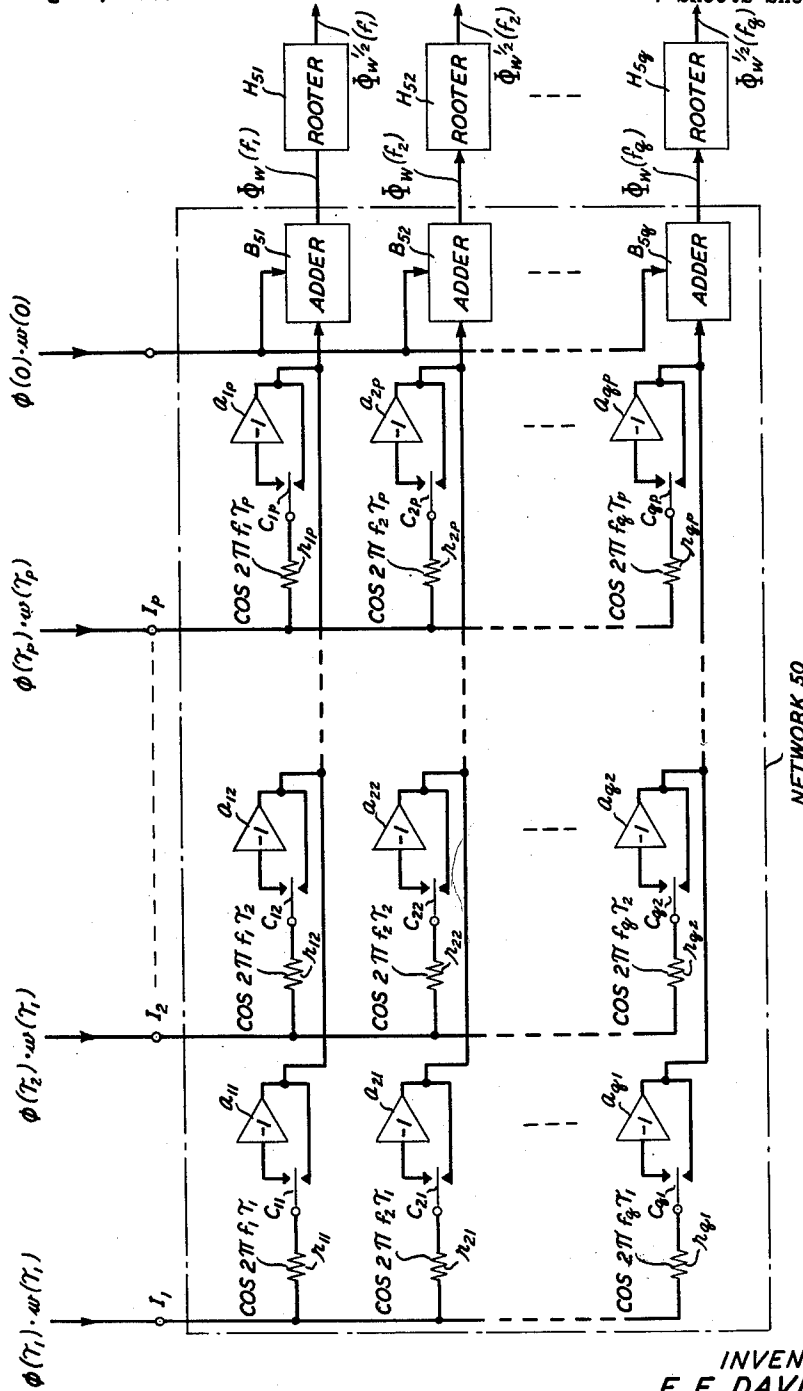
3,069,507

AUTOCORRELATION VOCODER

Filed Aug. 9, 1960

7 Sheets-Sheet 5

FIG. 5A



INVENTOR  
E. E. DAVID, JR.  
BY *G. E. Hirsch*  
ATTORNEY

Dec. 18, 1962

E. E. DAVID, JR

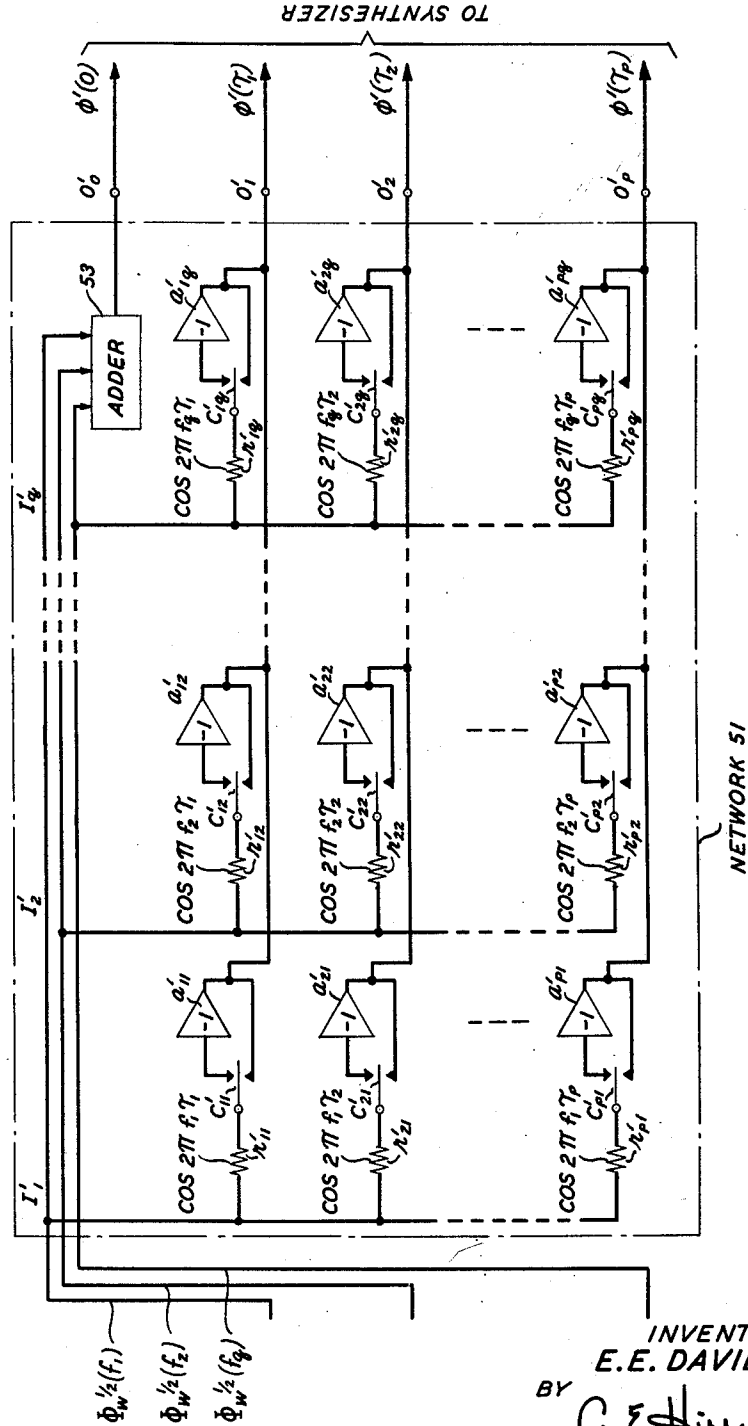
3,069,507

AUTOCORRELATION VOCODER

Filed Aug. 9, 1960

7 Sheets-Sheet 6

FIG. 5B

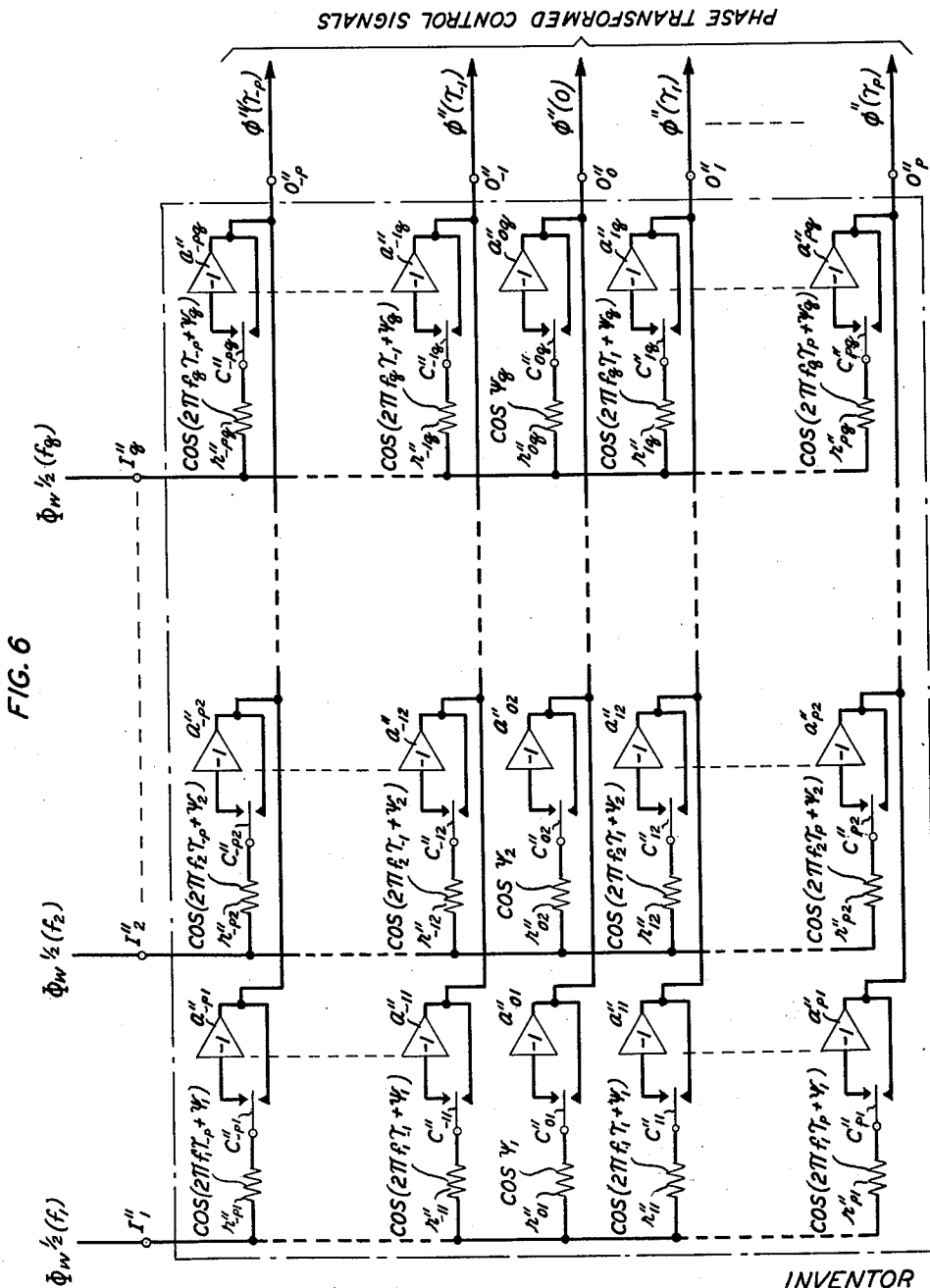


INVENTOR  
E. E. DAVID, JR.  
BY  
A. E. Hirsch, Jr.  
ATTORNEY

AUTOCORRELATION VOCODER

Filed Aug. 9, 1960

7 Sheets-Sheet 7



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

1

3,069,507

## AUTOCORRELATION VOCODER

Edward E. David, Jr., Berkeley Heights, N.J., assignor to Bell Telephone Laboratories, Incorporated, New York, N.Y., a corporation of New York  
Filed Aug. 9, 1960, Ser. No. 48,422  
10 Claims. (Cl. 179—15.55)

This invention relates to the transmission of speech over narrow-band channels, and particularly to the narrow-band transmission of speech in terms of autocorrelation functions.

Among speech coding systems for the conservation of transmission channel bandwidth, one of the best known is the channel vocoder described in H. W. Dudley Patent 2,151,091, issued March 21, 1939. At the transmitter terminal of the Dudley vocoder, the speech amplitude spectrum is divided into frequency bands by a bank of band-pass filters, and the energy contained within each band is represented by a narrow-band control signal. After transmission over a reduced bandwidth channel to a receiver station, the control signals adjust the energy of selected frequency bands of an excitation spectrum generated at the synthesizer. The energy-adjusted frequency bands are then combined to produce synthetic speech.

Synthetic speech produced by the Dudley vocoder is distorted by the inherent limitations of the band-pass filters that it employs. For ideal reproduction of speech, the speech amplitude spectrum should be represented by a group of points; as practiced by the Dudley vocoder, however, the finite widths of the band-pass filters produce "points" that are in fact averages of many points within the frequency bands passed by the filters.

It is a specific object of this invention to reduce distortion and to eliminate the need for band-pass filters by transmitting speech in terms of nearly ideal points on its autocorrelation function.

In this invention, an incoming speech wave is correlated with itself at an analyzer terminal to obtain a number of samples of each period of the speech autocorrelation function. The bandwidth required to transmit these samples is much smaller than that required to transmit the original speech wave, since the speech autocorrelation function changes very little from one period to the next. The samples are transmitted over a narrow-band channel to a synthesizer, where they are used in the reconstruction of artificial speech by serving as control signals to adjust the amplitude of an excitation signal generated at the synthesizer.

The intelligibility of artificial speech reconstructed from autocorrelation control signals is impaired by the squaring of the speech amplitude spectrum inherent in the autocorrelation function representation of speech. Squaring the speech amplitude spectrum changes its shape, thereby altering the characteristics of most sounds.

It is a specific object of the present invention to improve the intelligibility of speech reconstructed from autocorrelation control signals by performing a square-root-taking operation upon the amplitude spectrum of the speech autocorrelation function samples.

The amplitude spectrum of the autocorrelation samples is obtained by passing the samples through a first network of resistance elements, which performs a Fourier transformation upon the autocorrelation samples. The spectrum signals produced by this first network are then rooted by a group of square-root-taking circuits, whose rooted spectrum output signals represent an amplitude spectrum of the same shape as the amplitude spectrum of the original speech wave. Artificial speech reproduced from the rooted spectrum signals is therefore free of the

2

distortion due to amplitude spectrum squaring. In order to reproduce artificial speech from the rooted spectrum signals, they are converted into autocorrelation samples by a second Fourier transformation network also composed of resistance elements.

One of the properties of autocorrelation functions generally is an amplitude spectrum whose phase angles are all zero; hence speech reconstructed from autocorrelation signals will also have an amplitude spectrum with zero phase angles. Phase angles other than zero, however, may be desired in the amplitude spectrum of high quality artificial speech.

Accordingly, it is a specific object of this invention to reconstruct artificial speech having any desired phase spectrum.

In order to produce artificial speech with a given phase spectrum, the second Fourier transformation network in the square-root-taking apparatus referred to above is replaced by a phase transformation network that converts the rooted spectrum signals into correlation signals having a predetermined phase spectrum. The operation of the phase transformation network is based upon a Fourier transformation in which the desired phase angles have been inserted.

The symmetry of the speech autocorrelation function about the center of each period permits the function to be reproduced from samples of half of each period, thereby achieving a further reduction in transmission channel bandwidth. Variations in the period of the autocorrelation function, however, require variations in the number of half-period samples in order to reproduce the speech autocorrelation function exactly. To vary the number of half-period samples in synchrony with variations in the period of the autocorrelation function requires complex apparatus; for example, see the copending patent application of E. E. David, Jr., and J. R. Pierce filed this date, Serial No. 48,423.

Accordingly, it is a specific object of this invention to produce good quality artificial speech by reconstructing a symmetrical speech wave from a fixed number of autocorrelation samples, regardless of variations in the period of the speech autocorrelation function.

In this invention, a fixed number of samples of the speech autocorrelation function is obtained at the analyzer terminal, the number of samples being kept small, consistent with good quality speech, in order to conserve transmission channel bandwidth. The samples transmitted to the synthesizer terminal represent a portion of each autocorrelation period, and by deriving from each sample two equal amplitude, symmetrically located points on each reconstructed period, an artificial speech wave whose symmetry approximates the symmetry of the original speech autocorrelation function is reconstructed.

Because of variations in the period of the autocorrelation function, it is not possible to represent accurately each half period of the autocorrelation function with a fixed number of samples. A period that is either too long or too short with respect to the interval spanned by the fixed number of samples is truncated at some point other than the end of the period. The fixed number of samples thus represents truncated portions of each period of the autocorrelation function, and the points of truncation appear as abrupt discontinuities in artificial periods reconstructed from the samples, causing distortion in the artificial speech.

It is a further object of the present invention to minimize distortion in speech reconstructed from samples of truncated autocorrelation periods.

By passing the autocorrelation samples through a weighting network, the magnitudes of the individual samples are reduced by various predetermined amounts,

3

thereby reducing the discontinuities and the associated distortion in the artificial speech. The amount by which the magnitude of each sample is reduced depends upon the particular weighting function that is chosen. An optimum weighting function minimizes distortion with a negligible loss in vocal characteristics. Since the previously mentioned amplitude unquaring process is to be performed upon the samples after weighting, the weighting function must possess a real and nonnegative amplitude spectrum in order to assure physically realizable results from the square-root-taking process.

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

FIGS. 1A, 1B, and 1C are a group of waveform diagrams of assistance in explaining this invention;

FIG. 1D is a schematic block diagram showing a complete speech transmission system based upon the principles of this invention;

FIG. 2 is a block schematic diagram showing apparatus for coding speech in terms of a fixed number of autocorrelation control signals;

FIGS. 3A, 3B, 3C, 3D, 3E, 3F, and 3G are a group of waveforms of assistance in explaining the operation of the apparatus of FIG. 2;

FIG. 4 is a block schematic diagram showing apparatus for reconstructing artificial speech from autocorrelation control signals;

FIG. 5A is a schematic block diagram showing apparatus for unquaring the amplitude spectrum of the autocorrelation control signals;

FIG. 5B is a schematic block diagram showing apparatus for converting rooted amplitude spectrum signals into their autocorrelation function counterparts;

FIG. 6 is a schematic block diagram showing apparatus for generating autocorrelation control signals having any desired phase spectrum from rooted amplitude spectrum signals; and

FIG. 7 is a schematic block diagram showing apparatus for reconstructing artificial speech having any desired phase spectrum from phase-transformed autocorrelation control signals.

### Mathematical Foundations

A speech wave  $g(t)$  with period  $T$  may be expanded in a Fourier series

$$g(t) = \sum_n G(f_n) \cos(2\pi f_n t + \psi_n) \quad (1)$$

where the coefficients  $G(f_n)$  constitute the amplitude spectrum of  $g(t)$  and the phase angles  $\psi_n$  constitute the phase spectrum.

The autocorrelation function of  $g(t)$  is defined as the average

$$\varphi(\tau) \equiv \frac{1}{T} \int_0^T g(t) \cdot g(t-\tau) dt \quad (2)$$

where  $\varphi(\tau)$  has the same period as  $g(t)$  and  $\tau$  represents the amount of time by which  $g(t)$  is delayed before being multiplied together with the undelayed speech wave.

From Wiener's theorem, the autocorrelation function may also be expanded in a Fourier series,

$$\varphi(\tau) = \sum_n \Phi(f_n) \cos(2\pi f_n \tau) \quad (3)$$

where

$$\Phi(f_n) = G^2(f_n) \quad (4)$$

that is, the amplitude spectrum of  $\varphi(\tau)$  is the square of the amplitude spectrum of  $g(t)$ . Further, it is noted in a comparison of Equations 1 and 3 that all of the phase angles of the original speech wave have become zero in the speech autocorrelation function.

4

Referring now to FIG. 1A, there is shown the amplitude spectrum of a typical voiced sound. FIG. 1B, on the other hand, shows the amplitude spectrum of the autocorrelation function corresponding to the speech wave of FIG. 1A. It is observed in a comparison of the curves that squaring the amplitude spectrum, as given in Equation 3, doubles the differences between peaks of the amplitude spectrum, thereby suppressing the relatively small peaks and changing the characteristics of the sound.

FIG. 1C shows several periods of a typical speech autocorrelation function, where it is noted that the autocorrelation function is symmetrical about the center of each period.

The amplitude spectrum of the autocorrelation function may also be expanded in a Fourier series

$$\Phi(f_n) = \sum_N \varphi(\tau_N) \cos(2\pi f_n \tau_N) \quad (5)$$

where, from the symmetry of the autocorrelation function about the center of each period, the amplitude spectrum may be expressed in terms of autocorrelation function samples over either of the half periods

$$-\frac{T}{2} \leq \tau - N \leq 0 \text{ or } 0 \leq \tau_N \leq \frac{T}{2}$$

Hence Equation 5 may be rewritten

$$\Phi(f_n) = \sum_{N \geq 0} \varphi(\tau_N) \cos(2\pi f_n \tau_N) \quad (6)$$

### Complete Speech Transmission System

Referring first to FIG. 1D, an incoming speech wave from source 80 is applied to speech autocorrelation function analyzer 81, which derives samples of the speech autocorrelation function from the speech wave. The details of source 80, analyzer 81, and the other elements of FIG. 1D are described below. The samples are passed through weighting network 82 in order to reduce the magnitudes of discontinuities in the sampled autocorrelation function.

The amplitude spectrum of the weighted samples from network 82 is unquarred by Fourier transformation network 83 and square-root-taking circuits 84. The rooted spectrum signals from circuits 84 are converted into autocorrelation function samples with zero phase angles by Fourier transformation network 85, and into correlation function samples having predetermined nonzero phase angles by phase transformation network 86. Switch 87 may be manually set to pass the output signals of either network 85 or network 86 to a speech synthesizer 88, depending upon which set of signals is desired for a particular application of this invention. A speech wave is reconstructed from the output signals of either network 85 or network 86 by speech synthesizer 88, and artificial speech is reproduced from the reconstructed speech wave by reproducer 89.

Analyzer 81 is located at a transmitter terminal, while speech synthesizer 88 and phase transformation network 86 are located at a receiver terminal. Elements 82, 83, 84, and 85, however, may be located at either the transmitter terminal or the receiver terminal.

### Analyzer

Referring now to the analyzer apparatus of FIG. 2, an incoming speech wave  $g(t)$  from source 20, for example, a transducer of any suitable variety, is passed through low-pass filter 210. Filter 210 is proportioned to pass only those frequencies in the band from 0 to  $W$  cycles per second, where  $W$  may be chosen to be 4,000.

The band-limited speech wave output of filter 210 is applied simultaneously to a tapped delay line 231, for example, a tapped acoustic delay line, and to a bank of multipliers, for example, modulators  $M_1, M_2, \dots, M_p$ , each of which is provided with two input terminals and one output terminal. Delay line 231, which is terminated

5

in a matched impedance 211 to prevent reflection, is provided with taps  $P_1, P_2 \dots P_p$ , at which appear signals proportional to the speech wave at various delay times,  $g(t-\tau_1), g(t-\tau_2) \dots g(t-\tau_p)$ , where  $\tau_1, \tau_2 \dots \tau_p$  are the various delay times corresponding to taps  $P_1, P_2 \dots P_p$ , respectively.

Modulators  $M_1, M_2 \dots M_p$ , in addition to receiving the undelayed speech wave at one of their input terminals, have their second input terminals connected to delay line taps  $P_1, P_2 \dots P_p$ , respectively, to receive the variously delayed speech wave as a second input signal. The modulators develop at their output terminals signals proportional to the products  $g(t) \cdot g(t-\tau_1), g(t) \cdot g(t-\tau_2) \dots g(t) \cdot g(t-\tau_p)$ , which are passed to a bank of averaging devices, for example, low-pass filters  $F_1, F_2 \dots F_p$ , each having a cutoff of 25 cycles per second. Filters  $F_1, F_2 \dots F_p$  develop at their output points signals proportional to averages of the product signals received from modulators  $M_1, M_2 \dots M_p$ . From Equation 2, these signals are proportional to samples of the autocorrelation function at specific delay times,  $\varphi(\tau_1), \varphi(\tau_2) \dots \varphi(\tau_p)$ .

A signal proportional to  $\varphi(0)$  is obtained by applying the undelayed speech wave to both input terminals of modulator  $M_0$ , and by connecting the output terminal of  $M_0$  to low-pass filter  $F_0$ .

#### Weighting Network

The number of taps with which delay line 231 is provided, and the number of associated modulators and filters, determine the number of autocorrelation samples appearing at the output points of filters  $F_1, F_2 \dots F_p$ . From the previously noted symmetry of the speech autocorrelation function, transmission channel bandwidth is conserved by sampling half periods of the autocorrelation function, the symmetry of each period being restored at the synthesizer from the transmitted samples; for example, a 3 millisecond period need only be sampled over the delay interval 0 to  $1\frac{1}{2}$  milliseconds.

Variations in the period of the autocorrelation function, however, prevent accurate sampling of each half period with a fixed number of taps, modulators, and filters. A period that is either too short, FIG. 3A, or too long, FIG. 3B, with respect to the sampling interval  $(\tau_p - \tau_1)$ , is truncated at the last sampling point  $\tau_p$ , and symmetrical periods synthesized from samples of portions of truncated periods contain abrupt discontinuities at the points of truncation, as shown in FIGS. 3C and 3D, respectively. These abrupt discontinuities in the synthesized periods produce distortion in the artificial speech.

In order to reduce distortion due to discontinuities, the samples obtained by the apparatus of FIG. 2 are passed through a weighting network N that reduces the magnitude of each sample, thereby reducing the discontinuities and the associated distortion in the artificial speech.

As shown in FIG. 2, the weighting network N is located at the analyzer terminal, but, if desired, it may be located at the synthesizer terminal. Weighting network N consists of a group of resistors  $R_0, R_1, R_2 \dots R_p$ , one for each autocorrelation sample, connected to the output terminals of filters  $F_0, F_1, F_2 \dots F_p$ , respectively. The resistance values of the elements of circuit N are determined by the particular weighting function selected on the basis of the following criteria: (a) discontinuities in periods reconstructed from the weighted samples must be very small, consistent with the preservation of important speech characteristics; and (b) the amplitude spectrum of the weighted samples must be real and nonnegative. Many suitable weighting functions satisfying these criteria are available, for example, the class of decreasing autocorrelation functions, the decreasing property satisfying (a), and the autocorrelation property satisfying (b). One such function is shown graphically in FIG. 3E.

The requirement that the weighting function have a real and nonnegative amplitude spectrum deserves further ex-

6

planation. As will be explained below in connection with the description of the apparatus of FIG. 5, a square-root-taking operation is performed upon the amplitude spectrum of the weighted autocorrelation samples, thus requiring the spectral values to be real and nonnegative in order to obtain physically realizable results. Passing the autocorrelation samples through the weighting network is equivalent to multiplication in the time domain and convolution in the frequency domain. For example, if  $w(\tau)$  is the weighting function and  $W(f_n)$  is its amplitude spectrum, then the weighted autocorrelation function is  $\varphi(\tau) \cdot w(\tau)$ , and the amplitude spectrum of the weighted autocorrelation function is

$$\Phi_w(f_n) = \Phi(f_n) * W(f_n) = G^2(f_n) * W(f_n) \quad (6a)$$

where \* denotes convolution. From Equation 4,  $\Phi(f_n)$  is real and nonnegative, hence for  $\Phi_w(f_n)$  to be real and nonnegative, it is sufficient that  $W(f_n)$  be real and nonnegative.

FIG. 3E illustrates an example of a suitable weighting function  $w(\tau)$ , which decreases in value with increasing delay time. By choosing the values of resistors  $R_0, R_1, R_2 \dots R_p$  to correspond to the values of the weighting function at the same delay times as those of the incoming autocorrelation samples, the signals developed at the output terminals of the resistors are weighted samples

$$\varphi(0) \cdot w(0), \varphi(\tau_1) \cdot w(\tau_1), \varphi(\tau_2) \cdot w(\tau_2) \dots \varphi(\tau_p) \cdot w(\tau_p)$$

Artificial symmetrical periods reconstructed from a fixed number of samples after weighting are shown in FIGS. 3F and 3G, corresponding to the original periods shown in FIGS. 3A and 3B, respectively. The smoothness with which these periods begin and end is to be compared with the abrupt discontinuities of artificial periods reconstructed from the same samples without weighting, as illustrated in FIGS. 3C and 3D.

The weighted autocorrelation samples appearing at the output terminals of the weighting network N in FIG. 2 constitute a set of control signals from which artificial speech may be synthesized. As shown in FIG. 1C, the speech autocorrelation function changes very little from period to period, hence the variation of the control signals is very small. As a result, the control signals individually occupy relatively narrow-frequency bands, on the order of 25 cycles per second, and the entire group of control signals may be transmitted over a much narrower frequency band than is required for transmission of the original speech wave. The control signals may be transmitted from the analyzer terminal to the synthesizer terminal by means of any well-known transmission medium to meet the requirements of the particular application of this invention.

#### Supplementary Signal

The weighted autocorrelation control signals must be supplemented with a signal whose characteristics indicate whether the instantaneous speech sound is voiced or unvoiced, and if voiced, its fundamental pitch frequency. At the synthesizer an excitation signal is derived from the supplementary signal, the excitation signal characteristics being closely correlated with the characteristics of the supplementary signal. Artificial speech reconstructed from the excitation signal under the control of the autocorrelation samples thus preserves faithfully the characteristics of the original speech.

As shown in FIG. 2, the supplementary signal is derived by passing the speech wave output of source 20 through band-pass filter 214. Filter 214 is proportioned to pass the frequency band from 100 to 350 cycles per second, spanning the range of fundamental pitch frequencies of typical human talkers. This subband of frequencies is then transmitted as a supplementary signal to the synthesizer. The supplementary signal may also be a conventional voiced-unvoiced pitch signal, if desired.

## 7 Synthesizer

Artificial speech is reproduced from the transmitted autocorrelation control signals and supplementary signal at a synthesizer, a preferred embodiment of which is shown in FIG. 4. The synthesizer reconstructs an artificial speech wave with symmetrical periods by using the supplementary signal to generate an excitation signal and by using the autocorrelation samples as control signals to form from the excitation signal symmetrically located samples of an artificial period. The symmetry of the artificial periods reconstructed in this fashion approximates the symmetry of the original autocorrelation periods.

Referring now to the apparatus of FIG. 4, each of the incoming control signals is applied to the control terminal of a modulator whose output terminal is connected to delay line 441. Delay line 441 is terminated in a matched impedance 421 to prevent reflection and is provided with a number of taps disposed in symmetrically located pairs disposed about center tap  $S_{40}$ . The output terminal of modulator  $L_{40}$  is connected to center tap  $S_{40}$ , and the output terminal of each of the other modulators  $L_{41}, L_{42}, \dots, L_{4p}$  is connected to two taps  $S_{41}, S_{41}, S_{42}, S_{42}, \dots, S_{4p}, S_{4p}$ , respectively, disposed at equal intervals about center tap  $S_{40}$ .

The control signals  $\varphi(0), \varphi(\tau_1), \varphi(\tau_2), \dots, \varphi(\tau_p)$  are applied to the control terminals of modulators  $L_{40}, L_{41}, L_{42}, \dots, L_{4p}$ , respectively, and adjust the amplitude of an excitation signal supplied in parallel to the input terminals of the modulators from excitation signal generator 491. Generator 491 derives the excitation signal from the incoming supplementary signal, which is first passed through an equalizing delay 41 to synchronize the supplementary signal with the control signals. Excitation signal generator 491, which is fully described in a patent application of M. R. Schroeder, Serial No. 812,028, filed May 8, 1959, operates to provide the modulators with an excitation signal that is closely correlated with the voiced-unvoiced and fundamental frequency characteristics of the original speech wave as conveyed by the supplementary signal. If a conventional voiced-unvoiced pitch signal is employed as a supplementary signal, excitation signal generator 491 may be the usual buzz-hiss source.

At the modulators, the incoming control signals adjust the amplitude of the excitation signal from generator 491, and the amplitude-adjusted excitation signals derived by the modulators are passed to delay line 441 via the various tap connections. The output signals of the modulators reappear at the output terminal of delay line 441 as samples of symmetrical periods of an artificial wave. These samples are smoothed to form an artificial wave by filter 442, proportioned to eliminate all frequencies greater than 4,000 cycles per second. The electrical wave formed by filter 442 is converted into audible and intelligible speech by conventional reproducer 443 connected to the output terminal of filter 442.

### Amplitude Spectrum Unsquaring Network

The intelligibility of artificial speech synthesized directly from the weighted autocorrelation control signals produced by the analyzer of FIG. 2 is impaired by the amplitude spectrum squaring inherent in the autocorrelation function representation, as shown in Equation 6a and in FIGS. 1A and 1B. To improve quality and intelligibility, the shape of the amplitude spectrum of the artificial speech is made to resemble closely the shape of the amplitude spectrum of the original speech wave by subjecting the amplitude spectrum of the weighted autocorrelation control signals to a square-root-taking operation prior to synthesizing the artificial speech. From Equation 6a, taking the square root of the amplitude spectrum of the weighted autocorrelation samples produces

$$\Phi_{\#}^{\#}(f_n) = G(f_n) * W^{\#}(f_n) \quad (7)$$

## 8

from which it is seen that both

$$\Phi_{\#}^{\#}(f_n)$$

and  $G(f_n)$  have approximately the same shape.

Apparatus for unsquaring the amplitude spectrum of the weighted autocorrelation control signals is shown in FIG. 5A. The apparatus of FIG. 5A may be located in its entirety at either the analyzer or the synthesizer, or, if desired, the component parts may be conveniently divided between the two stations.

Network 50 of FIG. 5A converts the incoming control signals into signals representing specific values of the corresponding amplitude spectrum, in accordance with the Fourier transformation of Equation 5. The amplitude spectrum signals developed at the output terminals of network 50 are then passed through rooting circuits  $H_{51}, H_{52}, \dots, H_{5p}$ , which perform a square-root-taking operation upon the spectrum signals. The rooted spectrum signals derived by the rooting circuits represent an amplitude spectrum of approximately the same shape as the amplitude spectrum of the original speech wave, as given by Equation 7. The autocorrelation signals corresponding to the rooted spectrum signals are obtained by passing the latter through network 51 of FIG. 5B, which performs an inverse Fourier transformation in accordance with Equation 3. The autocorrelation signals appearing at the output terminals of network 51 constitute a group of control signals from which artificial speech may be reconstructed by applying them, together with a supplementary signal, to a synthesizer such that shown in FIG. 4.

Recalling the Fourier transformation of Equation 5, the amplitude spectrum of weighted autocorrelation control signals is represented by the following group of series:

$$\begin{aligned} \Phi_W(f_1) &= \varphi(0) \cdot w(0) + \varphi(\tau_1)w(\tau_1) \cos 2\pi f_1 \tau_1 \\ &\quad + \varphi(\tau_2)w(\tau_2) \cos 2\pi f_1 \tau_2 \\ &\quad + \dots + \varphi(\tau_p)w(\tau_p) \cos 2\pi f_1 \tau_p \\ \Phi_W(f_2) &= \varphi(0) \cdot w(0) + \varphi(\tau_1)w(\tau_1) \cos 2\pi f_2 \tau_1 \\ &\quad + \varphi(\tau_2)w(\tau_2) \cos 2\pi f_2 \tau_2 \\ &\quad + \dots + \varphi(\tau_p)w(\tau_p) \cos 2\pi f_2 \tau_p \\ &\quad \cdot \\ &\quad \cdot \\ &\quad \cdot \\ \Phi_W(f_q) &= \varphi(0) \cdot w(0) + \varphi(\tau_1)w(\tau_1) \cos 2\pi f_q \tau_1 \\ &\quad + \varphi(\tau_2)w(\tau_2) \cos 2\pi f_q \tau_2 \\ &\quad + \dots + \varphi(\tau_p)w(\tau_p) \cos 2\pi f_q \tau_p \end{aligned} \quad (8)$$

where  $f_q \leq W$ .

From the above equation, it is seen that in order to obtain signals proportional to specific values of the amplitude spectrum of the weighted control signals, it is necessary to multiply the various control signals by selected values of the cosine function and to add the resulting products.

Network 50 of FIG. 5A performs the required multiplication and addition by means of an array of  $q \cdot p$  resistors  $r_{ij}$  arranged in  $q$  rows and  $p$  columns, the input terminals of the resistors in the  $j$ th column being connected to a common input point,  $I_j$ , and the output terminals of the resistors in the  $i$ th row being connected to the input terminal of an adder,  $B_{5i}$ , of any well-known construction, through manually adjusted switches  $c_{ij}$ ,  $i=1, 2, \dots, q$ ,  $j=1, 2, \dots, p$ . Each row of resistors corresponds to a particular series in Equation 8, and the resistance values of the individual resistors in each row are proportional to the absolute values of the cosine factors appearing in the corresponding series. The position in which the switch connecting the output terminal of a given resistor to the input terminal of its adder is placed depends upon whether the corresponding cosine factor is positive, negative or zero. For a positive cosine value, the switch connects the output terminal of the resistor directly to its

common output point; for a negative cosine value, the switch connects the corresponding resistor to its common conductor via a polarity inverter  $a_{ij}$ ,  $i=1,2 \dots q$ ,  $j=1,2 \dots p$ , for example, a conventional minus one amplifier; and for a zero cosine value, the switch is placed in the open position.

The exact configuration of network 50 depends upon two factors: the frequency resolution or number of amplitude spectrum values desired in the reconstructed speech wave; and the number of autocorrelation samples obtained at the analyzer. As given by Equation 8, the frequency resolution determines the number of rows of resistors,  $q$ , and the number of autocorrelation samples determines the number of resistors in each row,  $p$ . Once these two factors have been established, the resistance values and the positions of the switches are fixed by Equation 8. For a particular application of this invention, the switches may be replaced by appropriate permanent connections, and only those polarity inverters required by the particular negative cosine values of Equation 8 need be employed.

The incoming weighted autocorrelation samples  $\varphi(\tau_1) \cdot w(\tau_1)$ ,  $\varphi(\tau_2) \cdot w(\tau_2) \dots \varphi(\tau_p) \cdot w(\tau_p)$  are applied to the input points  $I_1, I_2 \dots I_p$ , respectively, of network 50, and the sample  $\varphi(0) \cdot w(0)$  is applied in parallel to the second input terminals of adders  $B_{51}, B_{52} \dots B_{5q}$ . The signal developed at the output terminal of each adder is a linear combination of the weighted autocorrelation samples, which, in accordance with Equation 8, is proportional to a particular amplitude spectrum value.

The output signals  $\Phi_w(f_1), \Phi_w(f_2) \dots \Phi_w(f_q)$  of network 50 are applied to a bank of rooters  $H_{51}, H_{52} \dots H_{5q}$ , of any suitable variety. Each rooter develops at its output point a signal whose magnitude is proportional to the square root of the magnitude of the signal applied to its input point, that is,

$$\Phi_w^{\frac{1}{2}}(f_1), \Phi_w^{\frac{1}{2}}(f_2) \dots \Phi_w^{\frac{1}{2}}(f_q)$$

From Equation 7, these signals represent an amplitude spectrum of approximately the same shape as the amplitude spectrum of the original speech wave; hence, artificial speech reconstructed from either the rooted spectrum signals or their autocorrelation counterparts is substantially free of the distortion caused by amplitude spectrum squaring.

In order to reconstruct speech in a synthesizer of the type shown in FIG. 4 of this application, the rooted spectrum signals must be converted into their autocorrelation function counterparts. Apparatus for performing this conversion is based upon Equation 3, and a preferred embodiment is shown in FIG. 5B.

Referring now to the inverse Fourier transformation of Equation 3, the autocorrelation function corresponding to the rooted amplitude spectrum values produced by the rooters of FIG. 5A is given by the following group of series:

$$\begin{aligned} \varphi'(0) &= \Phi_w^{\frac{1}{2}}(f_1) + \Phi_w^{\frac{1}{2}}(f_2) + \dots + \Phi_w^{\frac{1}{2}}(f_q) \\ \varphi'(\tau_1) &= \Phi_w^{\frac{1}{2}}(f_1) \cos 2\pi f_1 \tau_1 + \Phi_w^{\frac{1}{2}}(f_2) \cos 2\pi f_2 \tau_1 + \dots \\ &\quad + \Phi_w^{\frac{1}{2}}(f_q) \cos 2\pi f_q \tau_1 \\ \varphi'(\tau_2) &= \Phi_w^{\frac{1}{2}}(f_1) \cos 2\pi f_1 \tau_2 + \Phi_w^{\frac{1}{2}}(f_2) \cos 2\pi f_2 \tau_2 + \dots \\ &\quad + \Phi_w^{\frac{1}{2}}(f_q) \cos 2\pi f_q \tau_2 \\ &\vdots \\ \varphi'(\tau_p) &= \Phi_w^{\frac{1}{2}}(f_1) \cos 2\pi f_1 \tau_p + \Phi_w^{\frac{1}{2}}(f_2) \cos 2\pi f_2 \tau_p + \dots \\ &\quad + \Phi_w^{\frac{1}{2}}(f_q) \cos 2\pi f_q \tau_p \end{aligned} \quad (9)$$

The operation of network 51 shown in FIG. 5B is based upon the above equation, and is similar in structure to network 50 of FIG. 5A. Network 51 consists of an array of  $p \cdot q$  resistors  $r'_{ij}$ ,  $i=1,2 \dots p$ ,  $j=1,2 \dots q$ , arranged in  $p$  rows and  $q$  columns the input terminals of

the resistors in each column being connected to a common input point  $I'_j$  and the output terminals of the resistors in each row being connected to a common output point  $O'_i$  through manually adjusted switches  $c'_{ij}$ . Each row of resistors corresponds to a particular series in Equation 9, and the resistance values of the individual resistors in each row are proportional to the absolute values of the cosine factors appearing in the corresponding series.

The position in which the switch connecting the output terminal of each resistor to its common output point is placed depends upon whether the corresponding cosine factor is positive, negative or zero. For a positive cosine value, the switch connects the corresponding resistor directly to its common output point; for a negative value, the switch connects the corresponding resistor to its common output point through polarity inverter  $a'_{ij}$ ,  $i=1,2 \dots p$ ,  $j=1,2 \dots q$ ; and for a zero value, the switch is placed in the open position.

The rooted spectrum signals from the rooters of FIG. 5A are applied to the input terminals of network 51, thereby developing at each common output point a linear combination of signals proportional to a particular series of Equation 9.

A signal proportional to  $\varphi'(0)$  is obtained by applying each of the rooted spectrum signals to an input terminal of adder 53; in accordance with Equation 9, the linear combination of rooted spectrum signals formed at the output terminal  $O'_q$  of adder 53 is proportional to  $\varphi'(0)$ .

The configuration of network 51 is determined by two quantities: the number of amplitude spectrum values produced by network 50; and the number of autocorrelation samples to be supplied to the synthesizer. As given by Equation 9, the number of rows of resistors,  $p$ , is determined by the number of autocorrelation samples to be supplied to the synthesizer, and the number of resistors in each row is determined by the number of rooted spectrum signals from the rooters. The establishment of these quantities fixes both the resistance values and the positions of the switches in network 51. For a particular application in which these quantities have been established, the switches may be replaced by appropriate permanent connections, and network 51 need contain only those polarity inverters actually required by negative cosine values in Equation 9.

The autocorrelation function signals  $\varphi'(0), \varphi'(\tau_1), \varphi'(\tau_2) \dots \varphi'(\tau_p)$  produced by network 51 constitute a set of control signals from which artificial speech may be reconstructed by a synthesizer of the type shown in FIG. 4 of this application. Artificial speech reconstructed from these autocorrelation control signals has an amplitude spectrum of approximately the same shape as that of the original speech wave, thus faithfully reproducing the original speech sounds.

#### Phase Transformation Network

It is observed from Equation 3 that the output signals of network 51 represent an autocorrelation function with a zero phase spectrum. If desired, a function with a phase configuration other than zero may be obtained by the following transformation of the rooted spectrum values given by Equation 7:

$$\begin{aligned} \varphi''(\tau_p) &= \Phi_w^{\frac{1}{2}}(f_1) \cos (2\pi f_1 \tau_p + \psi_1) \\ &\quad + \Phi_w^{\frac{1}{2}}(f_2) \cos (2\pi f_2 \tau_p + \psi_2) + \dots \\ &\quad + \Phi_w^{\frac{1}{2}}(f_q) \cos (2\pi f_q \tau_p + \psi_q) \\ &\vdots \\ \varphi''(\tau_{-1}) &= \Phi_w^{\frac{1}{2}}(f_1) \cos (2\pi f_1 \tau_{-1} + \psi_1) \\ &\quad + \Phi_w^{\frac{1}{2}}(f_2) \cos (2\pi f_2 \tau_{-1} + \psi_2) + \dots \\ &\quad + \Phi_w^{\frac{1}{2}}(f_q) \cos (2\pi f_q \tau_{-1} + \psi_q) \\ \varphi''(0) &= \Phi_w^{\frac{1}{2}}(f_1) \cos \psi_1 + \Phi_w^{\frac{1}{2}}(f_2) \cos \psi_2 + \dots + \Phi_w^{\frac{1}{2}}(f_q) \cos \psi_q \end{aligned}$$

11

$$\begin{aligned} \varphi''(\tau_1) &= \Phi_{\omega}^{\omega}(f_1) \cos(2\pi f_1 \tau_1 + \psi_1) \\ &\quad + \Phi_{\omega}^{\omega}(f_2) \cos(2\pi f_2 \tau_1 + \psi_2) + \dots \\ &\quad + \Phi_{\omega}^{\omega}(f_q) \cos(2\pi f_q \tau_1 + \psi_q) \\ &\vdots \\ \varphi''(\tau_p) &= \Phi_{\omega}^{\omega}(f_1) \cos(2\pi f_1 \tau_p + \psi_1) \\ &\quad + \Phi_{\omega}^{\omega}(f_2) \cos(2\pi f_2 \tau_p + \psi_2) + \dots \\ &\quad + \Phi_{\omega}^{\omega}(f_q) \cos(2\pi f_q \tau_p + \psi_q) \end{aligned} \quad (10)$$

where  $\psi_i$ ,  $i=1,2,\dots,q$  are the desired phase angles.

It is noted in Equation 10 that samples of both halves of each period of the phase-transformed autocorrelation function must be computed, since for arbitrary phase angles other than  $0^\circ$  or  $180^\circ$ , the phase-transformed function is not symmetrical about the center of each period. Computing samples of the phase-transformed function is therefore performed at the synthesizer after transmission, in order to maintain the saving in transmission channel bandwidth effected by sampling portions of half periods of the original speech autocorrelation function at the analyzer terminal.

Referring now to phase transformation network 61 of FIG. 6, there is shown apparatus for deriving from the rooted spectrum signals produced, for example, by rooters  $H_{51}, H_{52}, \dots, H_{5q}$ , of FIG. 5A, a group of control signals having any desired phase spectrum in accordance with Equation 10. Network 61 consists of an array of resistors  $r_{ij}''$ ,  $i=-p \dots -2, -1, 0, 1, 2 \dots p$ ,  $j=1, 2 \dots q$ , arranged in  $2p+1$  rows and  $q$  columns, the input terminals of the resistors in the  $j$ th column being connected to a common input point  $I_j''$ , and the output terminals of the resistors in the  $i$ th row being connected to a common output point  $O_i''$  through switches  $c_{ij}''$ . Each row of resistors corresponds to a particular series in Equation 10, and the individual resistance values of the resistors in each row are proportional to the absolute values of the individual cosine factors in the corresponding series. The position in which the switch connecting a given resistor to its common conductor is placed depends upon whether the corresponding cosine factor is positive, negative or zero. For a positive cosine value, the switch connects the particular resistor directly to its common output point; for a negative cosine value, the switch connects the resistor to its common output point through a polarity inverter,  $a_{ij}''$  of any conventional design; and for a zero cosine value, the switch is placed in the open position.

The rooted spectrum signals from rooters  $H_{51}, H_{52}, \dots, H_{5q}$ , of FIG. 5A are applied to input points  $I_1'', I_2'', \dots, I_q''$ , respectively, of network 61. The signal formed at each common output point of network 61 is a linear combination of the rooted spectrum signals passed through the resistors in each row, and each linear combination is proportional to a particular series of Equation 10.

The output signals  $\varphi''(\tau_{-p}) \dots \varphi''(0), \varphi''(\tau_1) \dots \varphi''(\tau_p)$  of network 61 constitute a set of control signals from which speech may be reconstructed by a synthesizer of the type shown in FIG. 7 of this invention. Artificial speech reconstructed from these signals has an amplitude spectrum of approximately the same shape as that of the original speech wave and a phase spectrum of any desired configuration.

Referring now to FIG. 7, speech is reproduced from the output signals of network 61 of FIG. 6 by applying the control signals  $\varphi''(\tau_{-p}) \dots \varphi''(\tau_{-1}), \varphi''(0), \varphi''(\tau_1) \dots \varphi''(\tau_p)$ , to the control terminals of a bank of modulators  $L_p \dots L_1, L_0, l_1 \dots l_p$ , respectively. The incoming control signals adjust the amplitude of an excitation signal supplied to the modulators from excitation generator 791. The excitation signal is derived by generator 791 from the incoming supplementary signal,

12

which is first delayed by equalizing delay 71 to synchronize the supplementary signal with the control signals. Generator 791 is identical in construction and operation to generator 491 of FIG. 4, and supplies the modulators with an excitation signal that is closely correlated with the voiced-unvoiced and fundamental frequency characteristics of the original speech wave.

The amplitude-adjusted output signals of the modulators are passed to tapped delay line 741, which is terminated in a matched impedance 721 and is similar in construction to delay line 441 of FIG. 4, via taps  $S_p \dots S_1, S_0, s_1 \dots s_p$ . The reconstructed signals appearing at the output terminal of delay line 741 represent samples of the periods of a nonsymmetrical correlation function having a nonzero phase spectrum. These samples are converted into a continuous wave by filter 742, which passes only those frequencies between 0 and 4,000 cycles per second. Audible speech is obtained from the output wave of filter 742 by conventional reproducer 743.

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 narrow-band transmission of speech, the combination that comprises a source of a speech wave, means for correlating said speech wave with itself to obtain a constant number of control signals representative of portions of the autocorrelation function of said speech wave, means for reducing discontinuities in the speech autocorrelation function represented by said control signals by selectively reducing the magnitude of each of said control signals, means for transmitting said reduced magnitude control signals to a receiver station, and, at said receiver station, means for reconstructing an artificial speech wave from said transmitted control signals.

2. Apparatus as defined in claim 1 wherein said means for selectively reducing the magnitude of each of said control signals comprises a plurality of resistance elements in one-to-one correspondence with said control signals whose resistance values are proportional to a decreasing function with a nonnegative amplitude spectrum.

3. In a system for the narrow-band transmission of speech, the combination that comprises a source of a speech wave, means for correlating said speech wave with itself to obtain a fixed number of signals representative of portions of the autocorrelation function of said speech wave, means for selectively reducing the magnitude of each of said autocorrelation signals, means for unsquaring the amplitude spectrum of said reduced magnitude autocorrelation signals to produce a set of unrounded control signals, means for transmitting said unrounded control signals to a receiver station, and, at said receiver station, means for reconstructing an artificial speech wave from said unrounded control signals.

4. Narrow-band speech transmission apparatus that comprises a source of a speech wave, means for correlating said speech wave with itself to obtain a fixed number of control signals representative of portions of the speech autocorrelation function, means for reducing discontinuities in the speech autocorrelation function represented by the control signals by reducing the magnitude of each of said control signals by a predetermined amount, means for transmitting said reduced magnitude control signals to a receiver station, and, at said receiver station, means for unrounding the amplitude spectrum of said control signals, and means for reconstructing an artificial speech wave from said unrounded control signals.

5. Apparatus as defined in claim 4 wherein said means for unrounding the amplitude spectrum of said control signals comprises a first array of resistors arranged in