

July 11, 1967

J. L. FLANAGAN

3,330,910

FORMANT ANALYSIS AND SPEECH RECONSTRUCTION

Filed May 8, 1964

4 Sheets-Sheet 1

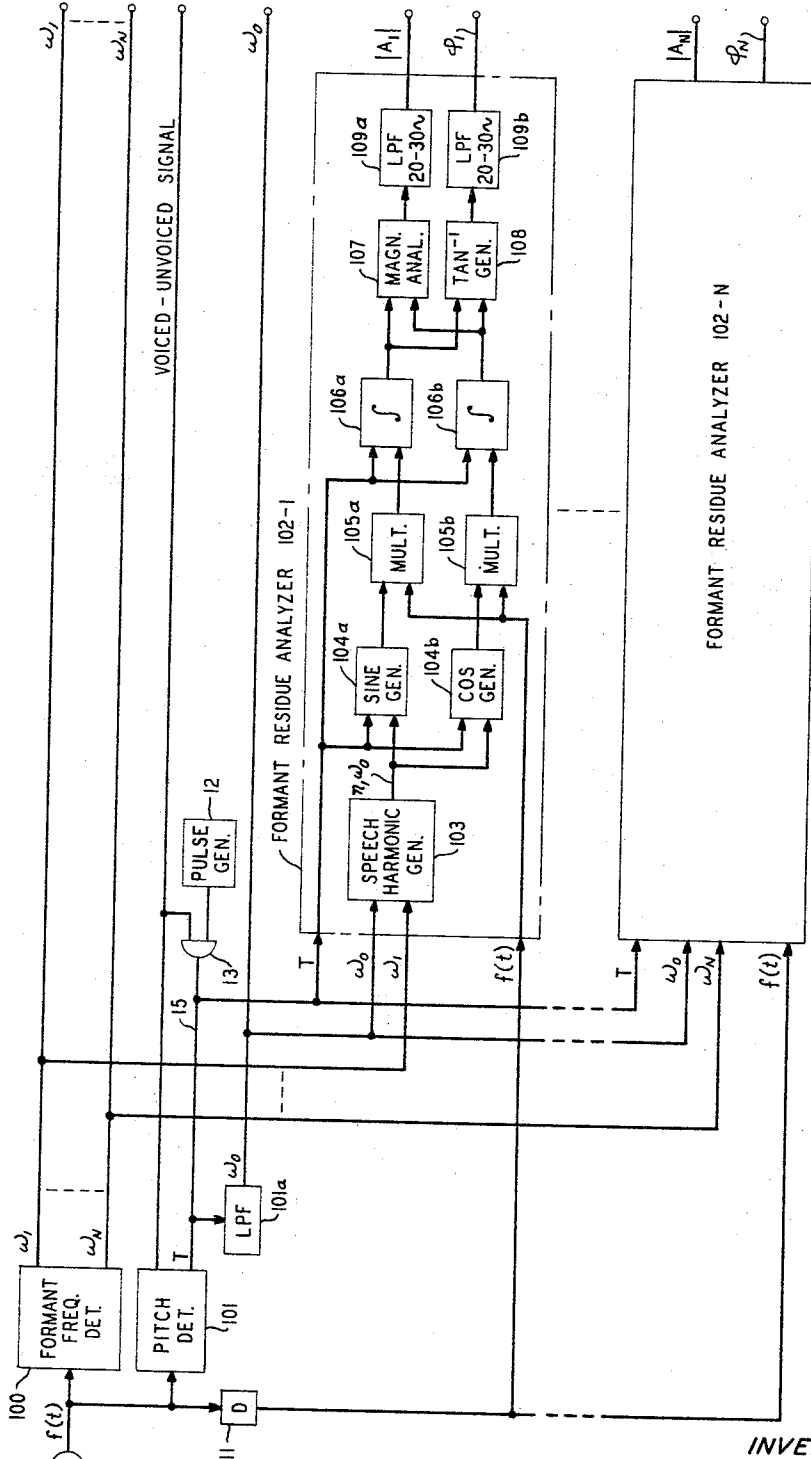


FIG. 1

INVENTOR  
J. L. FLANAGAN  
BY *Jules P. Kirsch*  
ATTORNEY

July 11, 1967

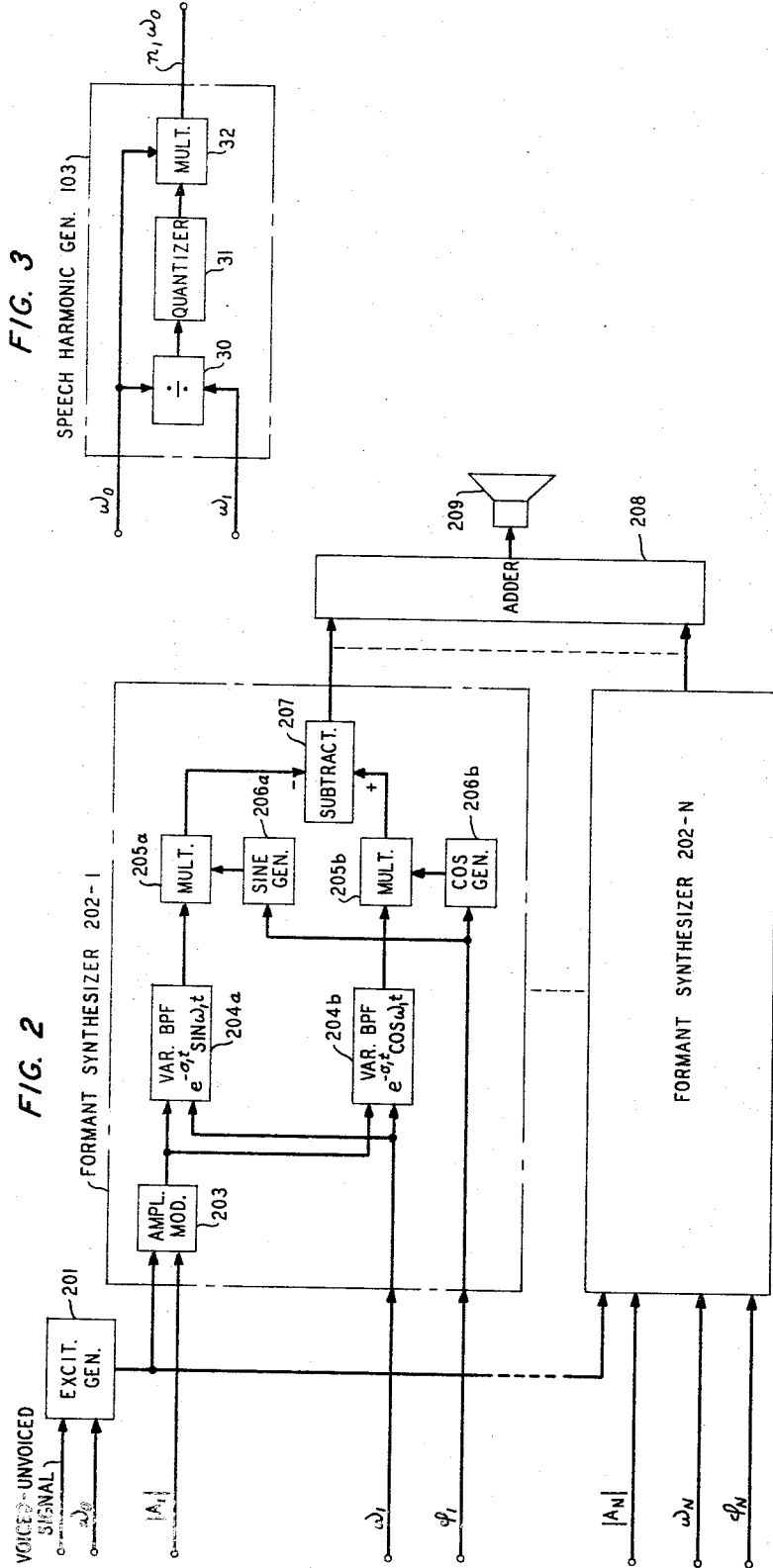
J. L. FLANAGAN

3,330,910

FORMANT ANALYSIS AND SPEECH RECONSTRUCTION

Filed May 6, 1964

4 Sheets-Sheet 2



July 11, 1967

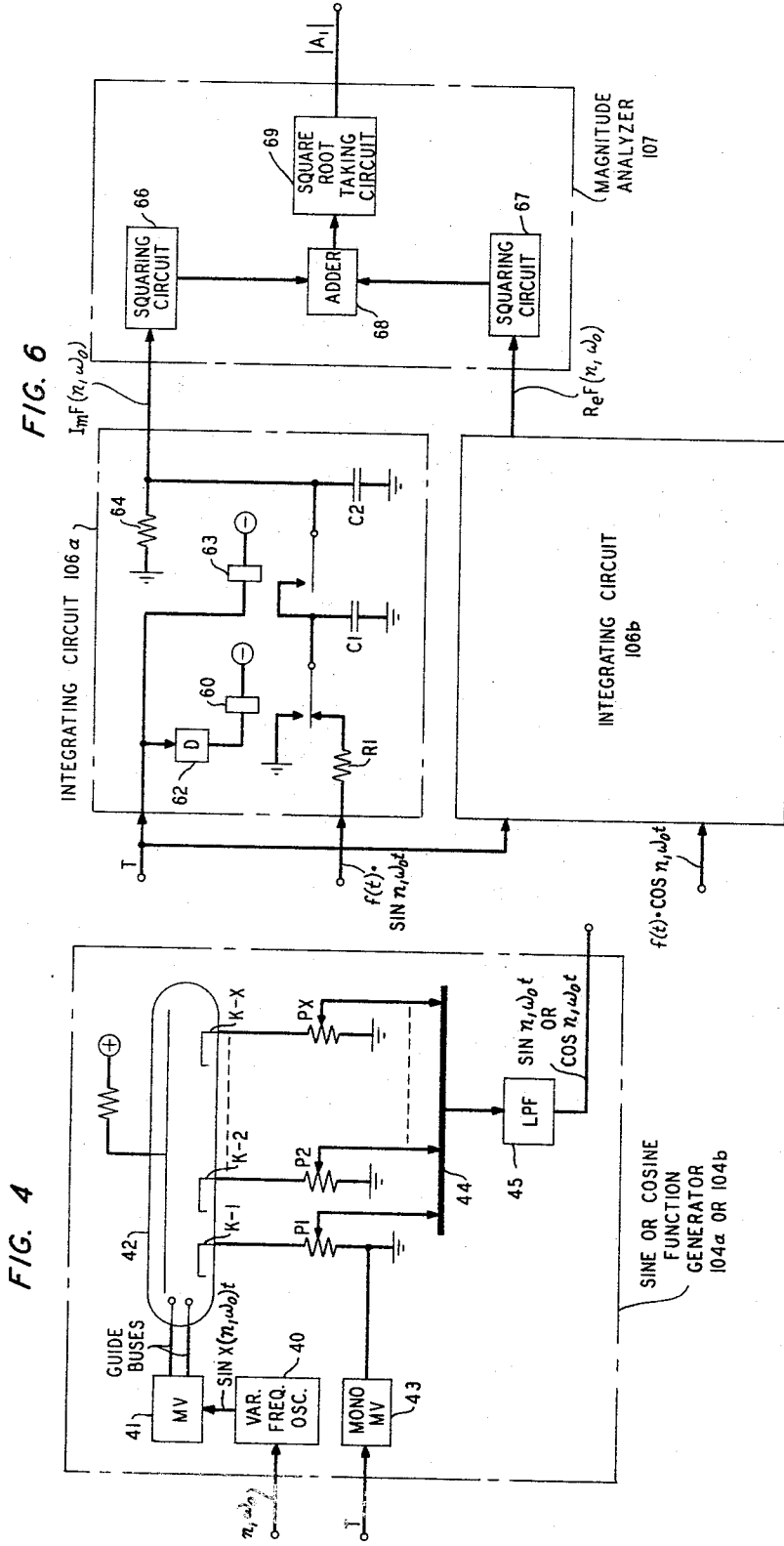
J. L. FLANAGAN

3,330,910

FORMANT ANALYSIS AND SPEECH RECONSTRUCTION

Filed May 6, 1964

4 Sheets-Sheet 3



July 11, 1967

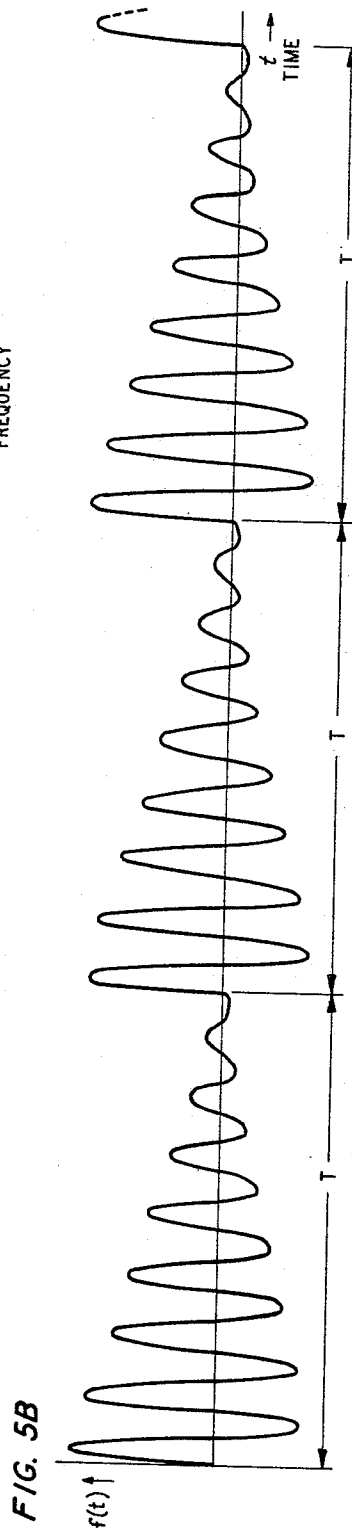
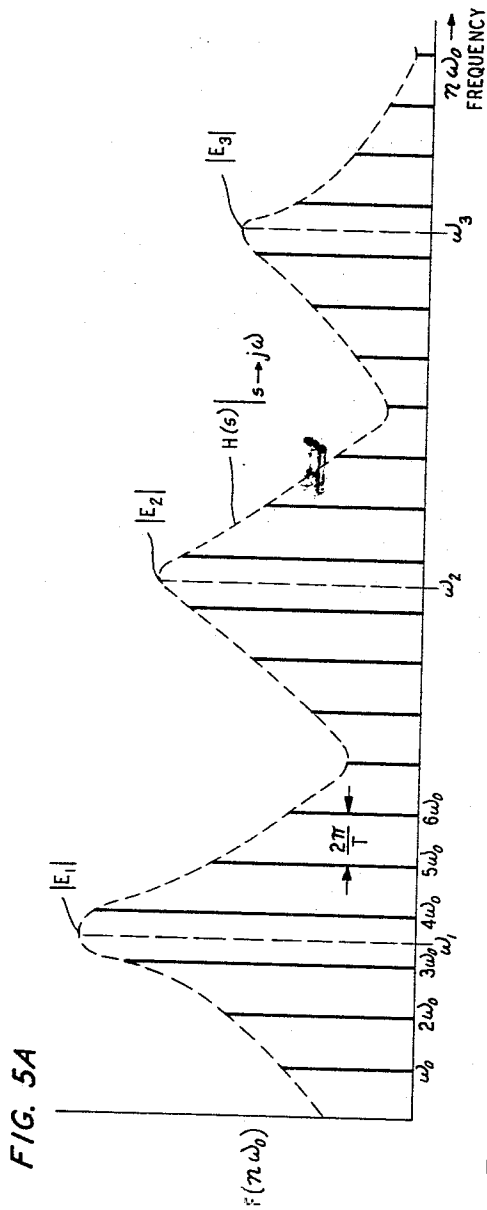
J. L. FLANAGAN

3,330,910

FORMANT ANALYSIS AND SPEECH RECONSTRUCTION

Filed May 6, 1964

4 Sheets-Sheet 4



1

2

3,330,910

**FORMANT ANALYSIS AND SPEECH RECONSTRUCTION**

James L. Flanagan, Warren Township, Somerset County, N.J., assignor to Bell Telephone Laboratories, Incorporated, New York, N.Y., a corporation of New York  
 Filed May 6, 1964, Ser. No. 365,353  
 5 Claims. (Cl. 179—15.55)

This invention relates to bandwidth compression systems, and in particular to the transmission of speech information over narrow bandwidth media in coded form.

Conventional speech communications systems, for example, telephone systems, convey human speech by transmitting an electrical facsimile of the acoustic waveform produced by human talkers. It has been recognized, however, that facsimile transmission of the speech waveform is a relatively inefficient way to transmit speech information, because of the redundancy of human speech. That is, the amount of information contained in a typical speech wave may be transmitted over a communication channel of substantially narrower bandwidth than that required for facsimile transmission of the speech waveform. In order to make more economic use of the frequency bandwidth of transmission media, a number of bandwidth compression arrangements have been devised for transmitting the information content of a speech wave over a channel whose bandwidth is substantially narrower than that required for transmission of the speech waveform. Bandwidth compression systems typically include at a transmitter terminal an analyzer for deriving from an incoming speech wave a group of relatively narrow-band control signals representative of selected information-bearing characteristics of the speech wave and at a receiver terminal, a synthesizer for reconstructing from the control signals a replica of the original speech wave.

One of the best known bandwidth compression arrangements is the so-called resonance or formant vocoder, in which the distinctive speech characteristics represented by the control signals are the frequencies and amplitudes of the so-called formants or normal modes of vibration of the vocal tract. Descriptions of formant vocoders may be found in H. L. Barney Patent 2,819,341, issued Jan. 7, 1958 and in an article by J. L. Flanagan entitled Note on the Design of "Terminal-Analog" Speech Synthesizers, volume 29, Journal of the Acoustical Society of America, page 306 (1957). Although the conventional formant vocoder is relatively efficient from the standpoint of transmitting information, the replica speech wave reconstructed from the transmitted control signals does not sound as natural as speech transmitted by facsimile waveform systems.

The bandwidth compression system of the present invention may also be characterized as a formant vocoder but it attempts to improve the quality of the replica speech wave by encoding a set of variables which more accurately describe the speech formants. It is well known that the transmission properties of the human vocal tract may be characterized as a superposition of damped sinusoids, each sinusoid corresponding to one of the formants of the vocal tract. There are four variables that specify the characteristics of each damped sinusoid, so that in principle each formant is exactly specified by a set of four parameters; namely, the amplitude and phase angle of a quantity generally referred to as the formant residue, the formant frequency, and the formant damping or bandwidth. The present invention recognizes that the sinusoid characterization of speech formants provides a precise characterization of the transmission properties of the human vocal tract, but in order to achieve an improvement in quality without reducing bandwidth economy by too large a factor, each sinusoid is approximated in the present inven-

tion by assuming a predetermined constant value for the formant damping variable and by measuring the three other variables. Thus the information content of a speech wave is encoded in this invention in terms of a group of  $3 \cdot N$  control signals, where  $N$  is an integer representing the number of formants or damped sinusoids to be employed in describing the transmission properties of the vocal tract.

The three parameters selected in the present invention to represent each sinusoid are the formant frequency and the magnitude and phase angle of the formant residue, a different residue being associated with each formant. At the transmitter terminal of this invention, the frequencies of a selected group of formants are located, after which a measure of the magnitude and phase angle of the residue associated with each formant is obtained from the magnitude and phase angle of the harmonic component of the speech amplitude spectrum which is determined to be closest in frequency to each formant being represented. Each of these three parameters is represented by a relatively narrow bandwidth control signal, and together with so-called voiced-unvoiced and pitch control signals representative of the properties of the excitation applied to the vocal tract, these control signals represent in coded form the information content of the incoming speech wave. The control signals are transmitted over a reduced bandwidth channel to a receiver station, where there is generated from the control signals a group of damped sinusoids corresponding to the selected group of formants whose frequencies are located at the transmitter terminal. This group of sinusoids is then combined to form a natural sounding replica of the original speech wave.

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

FIG. 1 is a block diagram showing the transmitter terminal of this invention;

FIG. 2 is a block diagram illustrating the receiver terminal of this invention;

FIG. 3 is a block diagram illustrating in detail certain components of the system shown in FIG. 1;

FIG. 4 is a diagram illustrating in detail certain other components of the transmitter terminal shown in FIG. 1;

FIGS. 5A and 5B are graphic illustrations of assistance in explaining the operation of the present invention; and

FIG. 6 is a diagram showing in detail still other components of the transmitter terminal in FIG. 1.

*Theoretical considerations*

The transmission properties of the human vocal tract may be described by taking the ratio of the Laplace transform,  $U_2(s)$ , of the volume velocity of air through the lips,  $u_2(t)$ , to the Laplace transform,  $U_1(s)$ , of the volume velocity of air through the vocal cords or glottis,  $u_1(t)$ , this ratio being commonly known as the transfer characteristic or transfer function of the vocal tract. As shown by J. L. Flanagan and A. S. House in Development and Testing of a Formant-Coding Speech Compression System, vol. 28, Journal of the Acoustical Society of America, page 1099 (1956), the vocal tract transfer function for vowel sounds may be expressed as

$$H_v(s) = \frac{U_2(s)}{U_1(s)} = \frac{\pi}{k} \frac{k s_k s_k^*}{(s - s_k)(s - s_k^*)} \quad (1)$$

where

$\frac{\pi}{k}$

denotes the product of  $k$  terms,  $s = (\sigma + j\omega)$  denotes the complex frequency variable;  $s_k = (-\sigma_k + j\omega_k)$  is a com-

plex number representing a formant or normal mode of vibration of the vocal tract; and

$$s_k^* = (-\sigma_k - j\omega_k)$$

is the complex conjugate of  $s_k$ . For nonvowel sounds, the vocal tract transfer function becomes

$$H_c(s) = \frac{U_2(s)}{U_1(s)} = K \frac{\prod_{j=1}^k (s - s_j)(s - s_j^*)}{\prod_{k=1}^k (s - s_k)(s - s_k^*)} \quad (2)$$

where  $K$  is a factor that includes the fixed radiation correction and  $j < k$ . Both Equation 1 and Equation 2 may also be written in terms of a partial fraction expansion that is formally the same for both vowel sounds and non-vowel sounds,

$$H(s) = \frac{U_2(s)}{U_1(s)} = \sum_k \left[ \frac{A_k}{(s - s_k)} + \frac{A_k^*}{(s - s_k^*)} \right] \quad (3)$$

where the quantity denoted  $A_k$  is called the complex residue in the pole  $s_k$  and is a function of all of the  $s_k$ 's. That is, for vowel sounds, the residue  $A_m$  in the  $m$ th pole,  $s_m$ , may be written

$$A_m = (s - s_m)H_v(s), \text{ for } s \rightarrow s_m \quad (4a)$$

$$= \frac{\prod_{k \neq m} k s_k \cdot s_k^*}{(s - s_m^*) \prod_{\pi \neq m} (s - s_k)(s - s_k^*)} \Big|_{s \rightarrow s_m} \quad (4b)$$

whereas for nonvowel sounds the residue  $A_m$  in the  $m$ th pole,  $s_m$ , may be written

$$A_m = (s - s_m)H_c(s) \text{ for } s \rightarrow s_m \quad (5a)$$

$$= \frac{\prod_{j \neq m} j (s - s_j)(s - s_j^*)}{(s - s_m^*) \prod_{\pi \neq m} (s - s_k)(s - s_k^*)} \quad (5b)$$

The vocal tract transfer function for both vowel and nonvowel sounds may also be expressed in terms of the inverse transform of the partial fraction expansion indicated in Equation 3, as indicated in the above-mentioned article

$$h(t) = \sum_k 2|A_k|e^{-\sigma_k t} \cos(\omega_k t + \varphi_k) \quad (6)$$

where  $h(t)$  is the inverse transform of  $H(s)$ ,  $|A_k|$  is the magnitude of the  $k$ th residue,  $A_k$ , as given by either Equation 4a or 5a, and  $\varphi_k$  is the phase angle of the  $k$ th residue, since

$$A_k = |A_k|e^{j\varphi_k} \quad (7)$$

It is observed in Equation 6 that the inverse transform of the transfer function of the vocal tract is defined by a sum of damped sinusoids, one for each pole or formant, so that each formant is completely specified by a set of four variables,  $A_k$ ,  $\sigma_k$ ,  $\omega_k$ , and  $\varphi_k$   $k=1, 2, \dots$ , describing the characteristics of each sinusoid. Both vowel and non-vowel sounds are therefore completely represented by  $(4k)$  quantities, where  $k$  is an integer denoting the number of formants or poles to be represented. For example, it is well known that from the standpoint of listener perception a large number of vowel sounds are accurately represented by three poles or formants, hence such sounds may be completely represented by  $4 \cdot 3 = 12$  parameters, supplemented, of course, by parameters describing the characteristics of the excitation applied to the vocal tract. However, formant bandwidth,  $\sigma_k$ , plays a relatively small role in influencing speech quality, hence only three of these four quantities are encoded in the present invention for each formant to be represented, namely, the magnitude and phase angle of each residue and the formant frequency. It is to be noted however that the bandwidth of each formant may be measured and encoded, if desired, with a consequent increase in the frequency bandwidth required for transmission of all of the coded signals. As described below in the detailed explanation of the

apparatus of this invention, the formant frequency may be determined by any one of a number of well-known techniques; for example, see H. L. Barney Patent 2,819,341, issued Jan. 7, 1958. However, in this invention the derivation of a pair of coded control signals representative of the magnitude and phase angle of each formant residue is based upon the following approximations.

As shown in the article by J. L. Flanagan entitled "Note on the Design of 'Terminal-Analog' Speech Synthesizers, Journal of the Acoustical Society of America, vol. 29, page 306 n. 12 (1957), the amplitude of a residue  $|A_k|$  is approximately equal to the maximum of the speech amplitude spectrum  $|E_k|$ , at the frequency  $\omega_k$  of the  $k$ th formant, that is,

$$|A_k| \cong \sigma_k |E_k| \quad (8)$$

where  $\sigma_k$  is the formant bandwidth which is assumed to be both small in comparison with the formant frequency  $\omega_k$  and relatively constant. FIG. 5A shows an idealized version of a voiced speech sound spectrum having an envelope with several maxima,  $|E_1|$ ,  $|E_2|$ ,  $|E_3|$  corresponding to speech formants at frequencies  $\omega_1$ ,  $\omega_2$ ,  $\omega_3$ , respectively. It is observed in FIG. 5A that the spectral maxima corresponding to speech formants do not generally coincide with the harmonically related speech frequency components occurring at integral multiples of the fundamental speech frequency,  $\omega_0$ . Since it is the speech frequency components and not the maxima in the spectral envelope which are most readily available for measurement, the present invention approximates the residue magnitudes and phase angles by determining the amplitudes and phase angles of the speech frequency components which are nearest in frequency to the spectral maxima corresponding to speech formants. Thus if  $|F(n\omega_0)|$  and  $\psi_n$  respectively denote the amplitude and phase angle of the speech component closest in frequency to the frequency location of the  $k$ th formant,  $\omega_k$ , then in the present invention the residue magnitude is approximated with the speech component amplitude

$$|A_k| \cong (\text{constant}) |F(n\omega_0)|$$

and the residue phase angle is approximated with the speech component phase angle

$$\varphi_k \cong \psi_n$$

The amplitude and phase angle of the  $n$ th harmonic frequency component of a speech wave are obtained in the following manner in this invention.

A single period  $T$  of a voiced portion of a speech wave  $f(t)$ , illustrated in FIG. 5B, may be expressed in terms of a Fourier series expansion

$$f(t) = \sum_{n=-\infty}^{\infty} F(n\omega_0) e^{jn\omega_0 t} \quad (9a)$$

where the characteristics of the individual frequency components of  $f(t)$  are represented by the complex Fourier coefficient,  $F(n\omega_0)$

$$F(n\omega_0) = \frac{1}{T} \int_0^T f(t) e^{-jn\omega_0 t} dt, \quad -\infty \leq n \leq +\infty \quad (9b)$$

and  $\omega_0$  is the fundamental radian frequency corresponding to the fundamental pitch period,  $T$

$$\omega_0 = \frac{2\pi}{T} \quad (9c)$$

Since the magnitudes of the Fourier coefficients for negative frequencies are equal to the magnitudes of the corresponding Fourier coefficients for positive frequencies, that is

$$|F(-n\omega_0)| = |F(n\omega_0)|$$

and since the phase angles for negative frequencies are equal to the corresponding phase angles for positive frequencies, that is

$$\varphi(-n\omega_0) = -\varphi(n\omega_0)$$

only positive values of  $n$  need to be considered in evaluating the magnitude and phase angle of a desired Fourier coefficient.

The magnitudes of the Fourier coefficients  $F(n\omega_0)$  constitute the amplitude spectrum of  $f(t)$ , and may be obtained by rewriting Equation 8b in terms of sine and cosine functions

$$F(n\omega_0) = \frac{1}{T} \int_0^T f(t) \cos(n\omega_0 t) dt - j \frac{1}{T} \int_0^T f(t) \sin(n\omega_0 t) dt \quad (10a)$$

so that the magnitude or absolute value of  $F(n\omega_0)$  is given by

$$|F(n\omega_0)| = \frac{1}{T} \left\{ \left[ \int_0^T f(t) \cos(n\omega_0 t) dt \right]^2 + \left[ \int_0^T f(t) \sin(n\omega_0 t) dt \right]^2 \right\}^{1/2} \quad (10b)$$

Equations 10a and 10b therefore specify that in order to obtain the magnitude of the  $n$ th harmonic component of the speech amplitude spectrum, it is necessary to multiply the incoming speech wave  $f(t)$  by  $\cos(n\omega_0 t)$  and  $\sin(n\omega_0 t)$  respectively, integrate the products  $f(t) \cos(n\omega_0 t)$  and  $f(t) \sin(n\omega_0 t)$  over the period  $T$ , square each of the integrated products, and determine the square root of the sum of the squared integrated products. It is this square root which is employed in the present invention to represent the residue magnitude,  $|A_k|$ .

The phase angle of the  $n$ th harmonic component may be derived from the relationship in Equation 10a, which may be rewritten

$$F(n\omega_0) = a - jb$$

where  $a$  represents the real (Re) part of  $F(n\omega_0)$ ,

$$a = \text{Re} F(n\omega_0) = \frac{1}{T} \int_0^T f(t) \cos(n\omega_0 t) dt \quad (11a)$$

and  $b$  represents the imaginary (Im) part of  $F(n\omega_0)$ ,

$$b = \text{Im} F(n\omega_0) = \frac{1}{T} \int_0^T f(t) \sin(n\omega_0 t) dt \quad (11b)$$

so that the phase angle  $\psi_n$  is given by the expression

$$\psi_n = \tan^{-1} \frac{-b}{a} \quad (12a)$$

$$= \tan^{-1} \left[ \frac{-\text{Im} E(\omega_n)}{\text{Re} F(\omega_n)} \right] \quad (12b)$$

Hence the phase angle of the  $n$ th harmonic component is obtained by determining the inverse tangent of the quotient formed by dividing the negative imaginary part of  $F(n\omega_0)$  by the real part of  $F(n\omega_0)$ , and it is this quotient which is employed in the present invention to represent the residue phase angle,  $\varphi_n$ .

#### Apparatus—complete system

Referring first to FIG. 1, this drawing illustrates the transmitter terminal of a speech bandwidth compression system embodying the principles of this invention. An incoming speech signal denoted  $f(t)$  from source 10, which may be a conventional microphone of any desired construction, is applied simultaneously to formant frequency detector 100, pitch detector 101, and through delay element 11 to formant residue analyzers 102-1 through 102-N. Formant frequency detector 100 may be any one of a number of well-known arrangements for determining the frequencies of selected speech formants; for example, detector 100 may be of the type described in H. L. Barney Patent 2,819,341, issued Jan. 7, 1958, which derives  $N$  narrow band control signals representative of the frequency locations  $\omega_1$  through  $\omega_N$  of  $N$  selected speech formants,  $N=2, 3, \dots$ . A typical choice for  $N$  is 3 so that detector 100 develops three formant frequency control signals representative of the frequency locations of the first three speech formants, it being understood that addi-

tional control signals representing the frequencies of higher order formants may be obtained if desired.

Pitch detector 101, which may also be of well-known construction, derives from the incoming speech signal a voiced-unvoiced control signal which indicates whether the incoming speech signal represents a voiced or an unvoiced sound at a given instant, and a pitch period signal comprising a train of uniform amplitude pulses having a period  $T$  that follows the period of the speech signal. A suitable detector is disclosed in A. J. Prestigiaco Patent 3,020,344, issued Feb. 6, 1962.

Each of the formant frequency signals  $\omega_1$  through  $\omega_N$  is applied to a corresponding formant residue analyzer circuit 102-1 through 102-N, and the pitch period signal is applied directly to each analyzer 102-1 through 102-N, as well as through low-pass filter 101a. Filter 101a smooths or averages the uniform amplitude pulses to produce a pitch frequency signal which indicates the fundamental or pitch frequency  $\omega_0$ . Analyzers 102-1 through 102-N are similar in structure, and hence only analyzer 102-1 is shown in detail. Within analyzer 102-1, the pitch frequency signal  $\omega_0$  and the formant frequency control signal  $\omega_1$  are applied to speech harmonic generator 103, which is shown in detail in FIG. 3. As explained below, generator 103 derives from these two control signals an output signal, denoted  $n_1\omega_0$ , which represents the frequency of the harmonic component of the speech amplitude spectrum which is nearest in frequency to the formant frequency  $\omega_1$ . The output signal developed by generator 103 is delivered in parallel to sine and cosine function generators 104a and 104b, one of which is shown in detail in FIG. 4 and further described below. Generators 104a and 104b are also supplied with the pitch period signal,  $T$ , and generator 104a produces an output signal comprising  $n$  cycles of a sine wave  $\sin(n_1\omega_0 t)$  generated within a single speech period of duration  $T$  seconds, while generator 104b produces an output signal comprising  $n$  cycles of a cosine wave  $\cos(n_1\omega_0 t)$  generated within a single speech period of duration  $T$  seconds.

In accordance with Equations 10a and 10b, each of the output signals of generators 104a and 104b is multiplied by the incoming speech signal  $f(t)$  in multipliers 105a and 105b to obtain product signals respectively proportional to  $f(t) \sin(n_1\omega_0 t)$  and  $f(t) \cos(n_1\omega_0 t)$ . From multipliers 105a and 105b each of the product signals is delivered to a corresponding integrating circuit 106a and 106b respectively. As specified by Equations 11a and 11b and as shown in detail in FIG. 6, integrating circuits 106a and 106b respectively derive from each of the product signals an output signal representative of the imaginary and real parts of the complex Fourier coefficient  $F(n_1\omega_0)$  which is utilized in this invention to approximate the complex residue,  $A_1$ , of the pole or formant located at the frequency  $\omega_1$ .

The integrated output signals of circuits 106a and 106b are sent to magnitude analyzer 107 and inverse tangent function generator 108. Analyzer 107, which is also shown in detail in FIG. 6, obtains an output signal representative of the magnitude of the harmonic component  $|F(n_1\omega_0)|$  which approximates the magnitude of the residue,  $A_1$ , and circuit 108, which may be of the type shown in W. J. Karplus and W. W. Soroka, Analog Methods (2d ed.) page 87 (1959), derives in accordance with Equations 12a and 12b an output signal representative of the phase angle  $\varphi_{n_1}$  which approximates the phase angle  $\varphi_1$  of the residue  $A_1$ .

The residue magnitude signal, denoted  $|A_1|$ , from analyzer 107 and the residue phase angle signal, denoted  $\varphi_1$ , from function generator 108 are respectively applied to low-pass filters 109a and 109b in order to derive slowly varying control signals which may be transmitted together with corresponding control signals from analyzers 102-2 to 102-N over a reduced bandwidth medium to the receiver terminal of this invention, illustrated in FIG. 2. Filters 109a and 109b may each have a cutoff frequency

on the order of 20 to 30 cycles per second, since the vocal tract and therefore the formant residues changes relatively slowly. In addition to the control signals from analyzer 102-1 through 102-N, the formant frequency control signals from detector 100 and the voiced-unvoiced and pitch frequency control signals from pitch detector 101 and low-pass filter 101a, respectively, are transmitted over the reduced bandwidth medium to the receiver station of the present invention. At the receiver terminal of this invention the transmitted control signals are employed to reconstruct a natural sounding replica of the original speech signal  $f(t)$ .

During unvoiced sounds, no pitch pulses are produced by detector 101, hence in order to detain control signals for the operation of analyzers 102-1 through 102-N the voiced-unvoiced signal from pitch detector 101 is used to control a gate 13 to which there is connected a conventional pulse generator 12. Pulse generator 12 produces a train of uniform amplitude pulses at a conveniently low rate, for example 50 pulses per second. Gate 13 passes the pulses from generator 12 to line 15 and thence to analyzers 102-1 through 102-N, as well as to low-pass filter 101a to derive an arrow band signal representative of the frequency of the pulses from generator 12. Thus during unvoiced sounds, each of the analyzers 102-1 through 102-N derives from the pulse train from generator 12 and from the low frequency signal from filter 101a a pair of output signals representative of the magnitude and phase angle of the short time unvoiced speech spectrum evaluated at the formant frequencies specified by the control signals  $\omega_1$  through  $\omega_N$  from detector 100.

#### Receiver terminal

Turning now to FIG. 2, this drawing illustrates the receiver terminal of this invention which constructs from the incoming narrow band control signals derived at the transmitter terminal a sum of N damped sinusoidal waves representing the inverse transform of the vocal tract transfer function specified by Equation 6, each sinusoidal wave corresponding to one of the N formants analyzed at the transmitter terminal. It is this sum of damped sinusoidal waves which constitutes the replica of the original speech wave reconstructed in the present invention. At the receiver terminal, the voiced-unvoiced and pitch frequency control signals are applied to an excitation generator 201, which may comprise conventional buzz and hiss sources. Thus when the voiced-unvoiced control signal indicates the presence of a voiced sound in the original speech, signal generator 201 produces a train of uniform amplitude pulses have a period T specified by the pitch control signal, whereas when the voiced-unvoiced control signal indicates the presence of an unvoiced sound generator 201 produces a train of randomly occurring uniform amplitude pulses. The excitation signal from generator 201 is applied in parallel to formant synthesizers 202-1 through 202-N, and in addition each synthesizer is supplied with the three control signals representing the frequency, residue magnitude and residue phase angle associated with the sinusoid characterization of a particular formant. Since synthesizers 202-1 through 202-N are similar in structure and operation, only synthesizer 202-1 is shown in detail.

Each synthesizer 202 develops one of the damped cosine waves specified by Equation 6. Thus Equation 6 may be written

$$h(t) = 2|A_1|e^{-\sigma_1 t} \cos(\omega_1 t + \varphi_1) + 2|A_2|e^{-\sigma_2 t} \cos(\omega_2 t + \varphi_2) + \dots + 2|A_N|e^{-\sigma_N t} \cos(\omega_N t + \varphi_N) \quad (13a)$$

and the  $j$ th term of the sum on the right-hand side of Equation 13a may be expanded by means of a trigonometric identity so that

$$2|A_j|e^{-\sigma_j t} \cos(\omega_j t + \varphi_j) = 2|A_j|e^{-\sigma_j t} [\cos(\omega_j t) \cos(\varphi_j) - \sin(\omega_j t) \sin(\varphi_j)] \quad (13b)$$

The  $j$ th synthesizer 202- $j$  of the present invention develops from the excitation signal produced by generation

201 and from the transmitted control signals  $|A_j|$ ,  $\omega_j$ ,  $\varphi_j$  a damped cosine wave according to Equation 13b.

Within synthesizer 202-1, the excitation signal from generator 201 is applied to amplitude modulator 203 together with the residue magnitude signal,  $|A_1|$ , associated with the first formant, so that the amplitude of the excitation signal is adjusted to be proportional to the instantaneous value represented by the residue magnitude signal. The amplitude adjusted excitation signal from modulator 203 is delivered simultaneously to variable bandpass filters 204a and 204b, which may be similar to the circuit shown in FIG. 5, of H. L. Barney Patent 2,819,341 issued Jan. 7, 1958, where filter 204a is suitably connected to produce a damped sine impulse response specified by the expression  $e^{-\sigma_1 t} \sin \omega_1 t$ , and filter 204b is suitably connected to produce a damped cosine impulse response specified by the expression  $e^{-\sigma_1 t} \cos \omega_1 t$ . The quantity  $\sigma_1$  is a constant selected to correspond to a desired bandwidth for the first formant and  $\omega_1$  is the variable center frequency of the filter pass band whose location on the frequency scale is controlled by the formant frequency control signal  $\omega_1$ . The output signal of filter 204a is therefore a damped sine wave of the form  $|A_1|e^{-\sigma_1 t} \sin \omega_1 t$ , and the output signal of filter 204b is therefore a damped cosine wave of the form

$$|A_1|e^{-\sigma_1 t} \cos \omega_1 t$$

The damped sine wave from filter 204a is applied to multiplier 205a, and the damped cosine wave from filter 204b is applied to multiplier 205b. Also connected to multiplier 205a is the output terminal of sine function generator 206a, which is responsive to the incoming narrow band  $\varphi_1$  signal and develops an output signal representative of  $\sin \varphi_1$ . Any desired sine function generator may be employed; for example, see pages 81-87 of the previously cited Karplus and Soroka text. Multiplier 205a therefore develops an output signal indicative of the product

$$|A_1|e^{-\sigma_1 t} \sin \omega_1 t \cdot \sin \varphi_1$$

Similarly, cosine function generator 206b is responsive to the incoming narrow-band  $\varphi_1$  control signal and develops an output signal proportional to  $\cos \varphi_1$ , so that multiplier 205b generates an output signal indicative of the product

$$|A_1|e^{-\sigma_1 t} \cos \omega_1 t \cdot \cos \varphi_1$$

The output signal of multiplier 205a is applied to the subtrahend or negative input terminal of subtractor 207, which may be of any well-known design, while the output signal of multiplier 205b is applied to the minuend or positive input terminal of subtractor 207. The difference signal produced at the output terminal of subtractor 207 is therefore proportional to the first term in Equation 13a.

Correspondingly, synthesizers 202-2 through 202-N each synthesize an output signal proportional to the second through Nth terms of Equation 13a. The output signals of synthesizers 202-1 through 202-N are additively combined in a conventional adder 208 to form an electrical wave that closely follows the specifications provided by Equations 6 and 13a. Reproducer 209 converts the electrical wave from adder 208 into a sound wave which is a natural sounding replica of the original speech sound, where reproducer 209 may be a loudspeaker of any desired variety.

#### Speech harmonic generator

Referring now to FIG. 3, this drawing illustrates in detail speech harmonic generator 103 shown in FIG. 1. The incoming fundamental pitch signal,  $\omega_0$ , from filter 101a is applied to the divisor terminal of a conventional divider circuit 30, and the incoming formant frequency signal  $\omega_1$  is applied to the dividend terminal of divider 30. Since in general the formant frequency  $\omega_1$  is not an integral multiple of the fundamental pitch frequency  $\omega_0$ , the quotient signal, denoted  $q$ , represents a number lying between two integers,  $j$  and  $k$ ,  $j < q < k$ . Therefore, in order



to obtain the frequency of the speech harmonic component nearest in frequency to  $\omega_1$ , the present invention passes the quotient signal to a quantizer 31; for example, a quantizer of the type described in B. M. Oliver Patent 2,773,980, issued Dec. 11, 1956.

Quantizer 31 is provided with a suitable plurality of preadjusted control level voltages corresponding to an anticipated range of values for the quotient signal from divider 30, so that the output condition of quantizer 31 signifies the value of the integer closest to the quotient signal value. For example, for the quantizer in harmonic generator 103 of the first formant residue analyzer 102-1, the first formant of a typical adult male voice may occur within the frequency range from about 300 to 1,000 cycles per second, and the fundamental pitch frequency may vary from about 50 to 150 cycles per second, hence the speech harmonic nearest in frequency to the first formant frequency may be from 2 to 20 times the fundamental pitch frequency. Accordingly, quantizer 31 is provided with 19 control level voltages each corresponding to one of the integers from 2 to 20, inclusive. Similarly, the quantizer in each of the corresponding speech harmonic generators of analyzers 102-2 through 102-N is provided with a plurality of control levels corresponding to the anticipated range of integers for the associated formant. Quantizer 31 is connected to one of the input terminals of multiplier 32, and the incoming fundamental pitch frequency signal is connected to the other input terminal of multiplier 32, thereby developing at the output terminal of multiplier 32 a signal representative of the frequency  $n_1\omega_0$ , of the speech harmonic closest the formant frequency  $\omega_1$ , where  $n_1$  represents the integer closest to  $q$ , the number represented by the quotient signal of divider 30. The output signal of multiplier 32 is then utilized in sine and cosine function generators 104a and 104b in the fashion previously described.

#### Sine/cosine function generator

Referring now to FIG. 4, the signal representative of the harmonic closest in frequency to the first speech formant,  $n_1\omega_0$ , is applied to variable frequency oscillator 40, and oscillator 40 develops at its output terminal a sine wave  $\sin n_1\omega_0 t$ . This wave is then sent to multivibrator 41 which produces a square wave at radian frequency  $X(n_1\omega_0)$ , where the value of X need only be greater than that required by the Nyquist interval for the sine or cosine function to be generated; for example, a convenient and completely adequate number would be 10. This square wave causes the glow discharge of multicathode gas transfer tube 42 to step progressively from cathode K-1 to cathode K-X at a rate equal to  $X(n_1\omega_0)$  divided by  $2\pi$  so that during each period of the sine wave X sample voltages are developed across output potentiometers P-1 through P-X. Tube 42 may be of the type described in A. H. Lytel, Industrial Electronics, pages 214 through 218 (1962). To insure synchronism of the cosine and sine functions with the pitch pulses, the glow discharge spot is returned to the first cathode K-1 at the end of every pitch period by a brief negative pulse from multivibrator 43, where multivibrator 43 may be of the conventional monostable or one-shot variety having a low output impedance. Multivibrator 43 is triggered by the incoming pitch period pulses so that the glow discharge spot is returned to cathode K-1 at the beginning of each speech period. It is therefore evident that in each speech period the glow discharge in tube 42 will step through  $n_1$  complete cycles. Each cathode is connected to a common output bus 44 through a potentiometer P-1 through P-X, where each potentiometer is set at a resistance value which corresponds to a particular sine or cosine sample value within a single sine or cosine period. On bus 44 there is developed a sequence of sine or cosine samples during each speech period, and the output terminal of bus 44 is connected to a low-pass filter 45 which smooths the sequence of  $n_1-X$  samples during each period T to develop

a sine or cosine wave of the form  $\sin(n_1\omega_0 t)$  or  $\cos(n_1\omega_0 t)$ .

#### Integrating circuit and magnitude analyzer

Referring now to FIG. 6, this drawing illustrates in detail an integrating circuit for performing the integration required of circuits 106a and 106b in FIG. 1. During a pitch period the incoming product signal  $f(t)\sin n_1\omega_0 t$  is applied through resistance R1 to capacitor C1, the time constant of resistor R1 and capacitor C1 being long compared to the pitch period T. Therefore the voltage accumulated on capacitor C1 is proportional to the integral of the incoming product signal over the pitch period. At the end of a given pitch period an incoming pitch period pulse is applied to relay 60 through delay element 62 and to relay 63, hence relay 63 is briefly energized before relay 60 so that relay 63 transfers the voltage on capacitor C1 to capacitor C2, where the capacitance of capacitor C2 is small in comparison with the capacitance of capacitor C1. Subsequent to the brief energizing of relay 63, relay 60 is briefly energized, thereby discharging capacitor C1 prior to the start of the next pitch period. The voltage developed across high-impedance load resistor 64 and appearing at the output terminal of integrating circuit 106a is therefore proportional to the imaginary part of the speech amplitude spectrum at the harmonic frequency  $n_1\omega_0$ , as specified by Equation 11b. Hence, relay 63 acts to sample and hold each value of  $ReF(n_1\omega_0)$  calculated for the pitch period. Correspondingly, integrating circuit 106b, which is identical with integrating circuit 106a, develops from the incoming product signal  $f(t)\cos n_1\omega_0 t$  an output signal proportional to the real part of the speech amplitude spectrum at the harmonic frequency  $n_1\omega_0$ , in accordance with Equation 11a.

From integrating circuits 106a and 106b the signals representative of the real and imaginary parts of the amplitude spectrum of the speech wave are respectively delivered to squaring circuits 66 and 67 of magnitude analyzer 107. Circuits 66 and 67 may be of any desired variety for obtaining an output signal proportional to the square or second power of an input signal; for example, see pages 78 through 81 of the previously mentioned Karplus and Soroka text. The output signals of circuits 66 and 67 are combined by adder circuit 68 and the sum signal developed at the output terminal of adder 68 is passed to square root taking circuit 69, which may also be of the type described on pages 78 through 81 of the Karplus and Soroka text. The output signal of circuit 69 is therefore proportional to the square root of the sum of the squares of the integrated product signals from circuits 106a and 106b, hence according to Equation 10b the output signal of circuit 69 is proportional to the magnitude of the speech amplitude spectrum at the harmonic frequency  $n_1\omega_0$ , and in this invention this signal is employed as an approximation of the formant residue magnitude ( $A_1$ ) in the manner described above.

Although this invention has been described in terms of a speech communication system of the type shown in FIGS. 1 and 2, it is to be understood that applications of the principles of this invention are not limited to such systems, but include the fields of automatic speech recognition, speech processing, and automatic message recording and reproduction. In addition, it is to be understood that the above-described embodiments are merely illustrative of the numerous arrangements that may be devised from the principles of this invention by those skilled in the art without departing from the spirit and scope of the invention.

What is claimed is:

1. A bandwidth compression system that comprises a transmitter terminal including a source of an incoming speech wave, means for deriving from said speech wave a plurality of formant frequency control signals representative

- 11  
of the frequencies of selected maxima in the amplitude spectrum of said speech wave,  
means for obtaining from said speech wave a voiced-unvoiced control signal indicative of the presence of  
voiced and unvoiced portions of said speech wave and  
a pitch control signal indicative of the periodicity of  
voiced portions of said speech wave, and  
a plurality of analyzer means in one-to-one correspondence with said plurality of formant frequency control signals, each of said analyzer means being  
supplied with said speech wave, said pitch control signal, and said corresponding formant frequency control signal,  
wherein each of said analyzer means derives from said control signals a pair of magnitude and phase angle control signals respectively representative of the magnitude and phase angle of the harmonic component of said speech wave which is closest in frequency to the frequency of said selected maximum represented by said formant frequency control signal,  
means for transmitting said control signals to a receiver terminal, and at said receiver terminal,  
means for reconstructing a replica of said speech wave from said control signals.
2. A bandwidth compression system that comprises a transmitter terminal including  
a source of an incoming speech wave,  
means for deriving from said speech wave a plurality of formant frequency control signals representative of the frequencies of selected maxima in the amplitude spectrum of said speech wave,  
means for obtaining from said speech wave a voiced-unvoiced control signal indicative of the presence of voiced and unvoiced portions of said speech wave and a pitch control signal indicative of the periodicity of voiced portions of said speech wave, and  
a plurality of analyzer means in one-to-one correspondence with said plurality of formant frequency control signals, each of said analyzer means being supplied with said speech wave, said pitch control signal, and said corresponding formant frequency control signal,  
wherein each of said analyzer means derives from said control signals a pair of magnitude and phase angle control signals respectively representative of the magnitude and phase angle of the harmonic component of said speech wave which is closest in frequency to the frequency of said selected maximum represented by said formant frequency control signal,  
means for transmitting said control signals to a receiver terminal, and at said receiver terminal,  
means for generating an excitation signal from said voiced-unvoiced and pitch control signals,  
a plurality of synthesizer means in one-to-one correspondence with said plurality of analyzer means,  
wherein each of said synthesizer means generates a damped sinusoidal wave representing one of the formants of said speech wave from said excitation signal, a corresponding one of said formant frequency control signals, and a corresponding one of said pairs of magnitude and phase angle signals, and  
means for combining said damped sinusoidal waves generated by said plurality of synthesizer means to form a replica of said incoming speech wave.
3. Apparatus for synthesizing speech which comprises a source of a voiced-unvoiced control signal and a pitch control signal respectively indicative of the presence of voiced and unvoiced portions of an original speech wave and of the periodicity of voiced portions of said speech wave,  
a source of a plurality of groups of damped sinusoid control signals, each group of damped sinusoid control signals including a formant frequency signal representative of the frequency of a selected formant peak in the amplitude spectrum of said speech wave, an amplitude control signal representative of the amplitude of the harmonic component of said speech wave which is closest in frequency to said formant peak, and a phase control signal representative of the phase angle of said harmonic component of said speech wave which is closest in frequency to said formant peak,  
means for generating an excitation signal from said voiced-unvoiced and pitch control signals,  
a plurality of synthesizer means in one-to-one correspondence with said plurality of groups of damped sinusoid control signals,  
wherein each of said synthesizer means derives from said excitation signal and said corresponding group of damped sinusoid control signals a damped sinusoid having an amplitude, frequency, and phase angle controlled by said group of damped sinusoid control signals and a predetermined damping, and  
Means for combining said plurality of damped sinusoids derived by said plurality of synthesizer means.
4. Apparatus for determining the magnitude and phase angle of the harmonic frequency component of an incoming speech wave which lies closest in frequency to a selected formant of said speech wave which comprises a source of an incoming speech wave,  
first detector means for deriving from said speech wave a formant frequency signal indicative of the frequency location of a selected formant peak in the amplitude spectrum of said speech wave,  
second detector means for deriving from said speech wave a train of pitch marker pulses having a period indicative of the fundamental period of voiced portions of said speech wave and a pitch frequency signal indicative of the fundamental frequency of voiced portions of said speech wave, and  
analyzer means supplied with said speech wave, said formant frequency signal, said train of pitch marker pulses, and said pitch frequency signal for deriving a pair of control signals respectively indicative of the magnitude and phase angle of the harmonic component of said speech wave which is closest in frequency to said formant peak, said analyzer means comprising  
a speech harmonic generator for deriving from said formant frequency signal and said pitch frequency signal an output signal proportional to the frequency of the harmonic component of said amplitude spectrum which is closest in frequency to the frequency location of said selected formant peak,  
means for generating from said harmonic frequency signal a first sine wave and a first cosine wave each having a frequency proportional to the frequency of said harmonic component which is closest in frequency to the frequency location of said selected formant peak,  
means for multiplying said first sine wave and said first cosine wave by said speech wave to develop a first sine product signal and a first cosine product signal,  
first integrating means supplied with said sine product signal and said train of pitch marker pulses for obtaining at the end of each pitch period an integrated sine product signal proportional to the integral of said sine product signal over each pitch period represented by said pitch marker pulses,  
second integrating means in parallel with said first integrating means and supplied with said cosine product signal and said train of pitch marker pulses for obtaining at the end of each pitch period an integrated cosine product signal proportional to the integral of said cosine product signal over each pitch period represented by said pitch marker pulses,  
magnitude analyzer means supplied with said integrated sine product signal and said integrated cosine product

13

signal for obtaining a magnitude signal proportional to the square root of the sum of the second power of said integrated sine product signal and the second power of said integrated cosine product signal, and inverse tangent analyzer means in parallel with said magnitude analyzer means and supplied with said integrated sine product signal and said integrated cosine product signal for obtaining phase angle signal proportional to the inverse tangent of the quotient of the negative of said integrated sine product signal divided by said integrated cosine product signal, whereby said magnitude signal represents the magnitude of said harmonic component which is closest in frequency to said selected formant peak and said phase angle signal represents the phase angle of said harmonic component.

5. Apparatus for synthesizing a signal representative of the characteristics of a selected formant of a speech wave which comprises

- a source of an excitation signal comprising a train of uniform amplitude pulses having a period indicative of the presence of voiced and unvoiced portions of and original speech wave,
- a source of an incoming magnitude control signal indicative of the instantaneous amplitude of the harmonic component of said speech wave which is closest in frequency to said selected formant of said speech wave,
- amplitude modulating means responsive to said magnitude control signal and supplied with said excitation signal for adjusting the amplitude of said train of pulses to follow the amplitude represented by said magnitude control signal,
- a source of a formant frequency control signal indicative of the frequency location of said selected formant,
- a first variable bandpass filter responsive to said formant frequency control signal and supplied with said amplitude adjusted train of pulses, wherein said first filter has a pass band with a variable center frequency that follows the formant frequency represented by said formant frequency control signal and wherein said first filter has a predetermined damping factor so that said first filter produces a damped sinusoidal output wave with a predetermined damping and at a frequency determined by said formant frequency control signal,

14

- a second variable bandpass filter in parallel with said first variable bandpass filter and responsive to said formant frequency control signal as well as supplied with said amplitude adjusted train of pulses, wherein said second filter has a pass band with a variable center frequency that follows the formant frequency represented by said formant frequency control signal and wherein said second filter has a predetermined damping factor so that said second filter produces a damped cosinusoidal output wave with a predetermined damping and at a frequency determined by said formant frequency control signal,
- a source of an incoming phase control signal indicative of the instantaneous phase angle of said harmonic component which is closest in frequency to said selected formant of said speech wave,
- a sine function generator supplied with said phase control signal for generating a sine wave having a frequency specified by the phase angle represented by said phase angle signal,
- a cosine function generator in parallel with said sine function generator and supplied with said phase control signal for generating a cosine wave having a frequency specified by the phase angle represented by said phase angle signal,
- first multiplying means supplied with said damped sinusoidal output wave and said sine wave for developing a first product signal proportional to the product of said damped sinusoidal output wave and said sine wave,
- second multiplying means supplied with said damped cosinusoidal output wave and said cosine wave for developing a second product signal proportional to the product of said damped cosinusoidal output wave and said cosine wave, and
- means for subtracting said second product signal from said first product signal to obtain a difference signal indicative of the characteristics of said selected formant of said original speech wave.

No references cited.

JOHN W. CALDWELL, *Acting Primary Examiner.*

W. S. FROMMER, *Assistant Examiner.*