June 27, 1967 J. L. KELLY, JR 3,328,525

SPEECH SYNTHESIZER

Filed Dec. 30, 1963 2 Sheets-Sheet 1

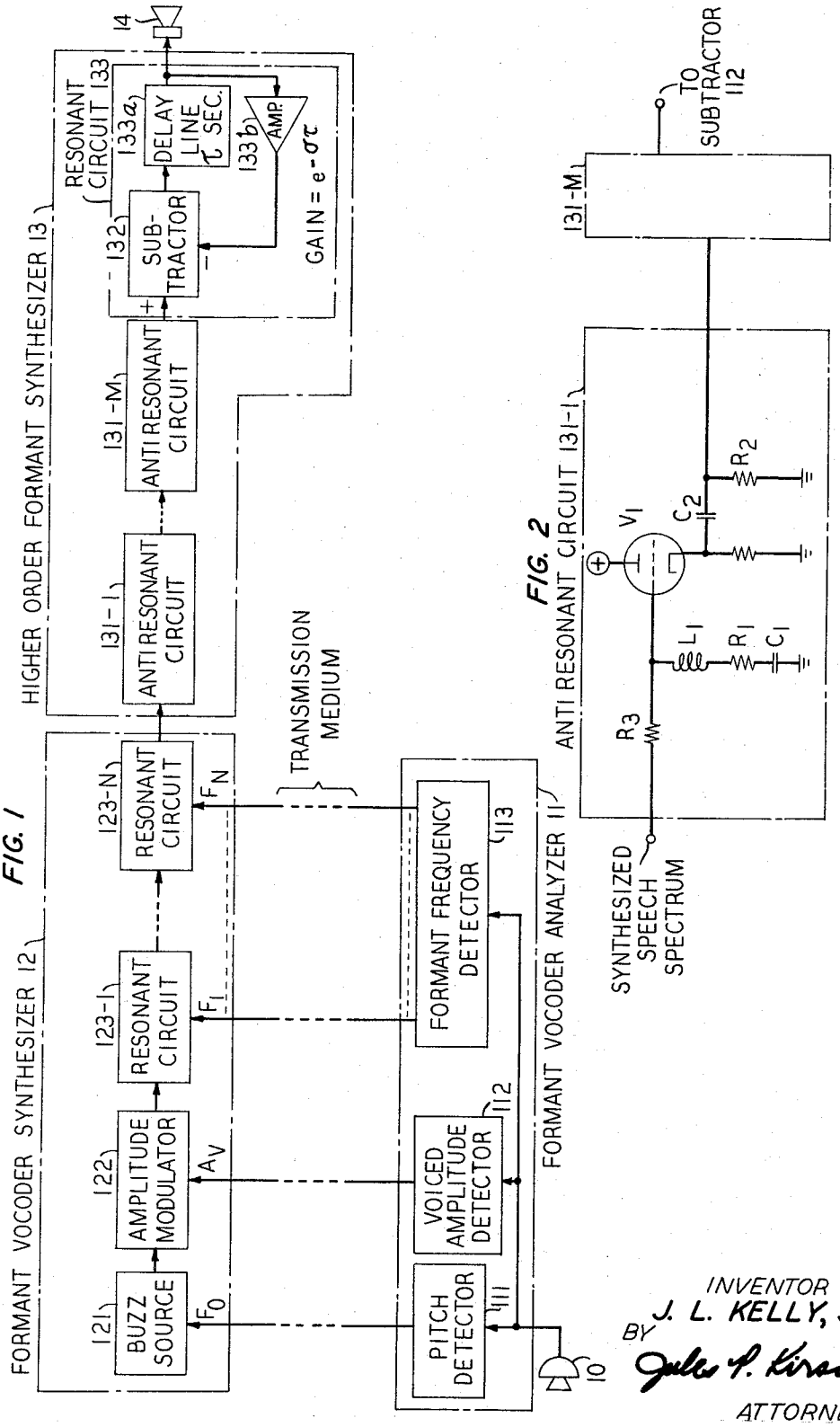


FIG. 1

FIG. 2

INVENTOR
J. L. KELLY, JR
BY Jules A. Kirsch
ATTORNEY

June 27, 1967 J. L. KELLY, JR 3,328,525

SPEECH SYNTHESIZER

Filed Dec. 30, 1963 2 Sheets-Sheet 2

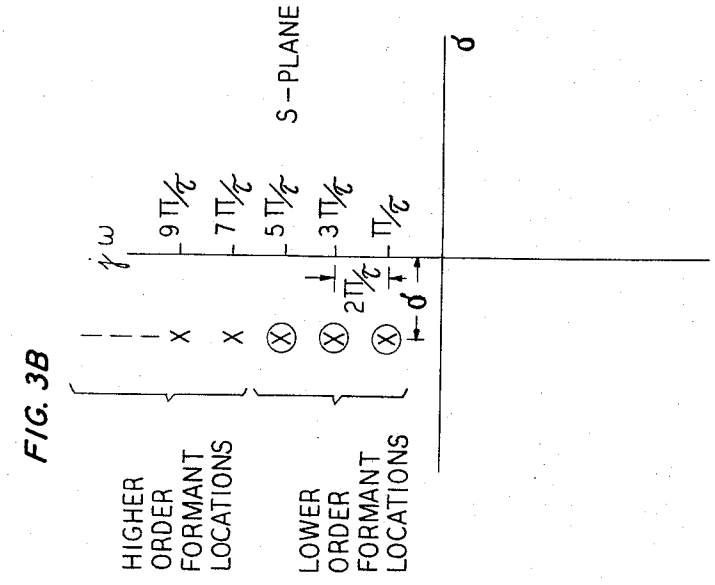


FIG. 3B

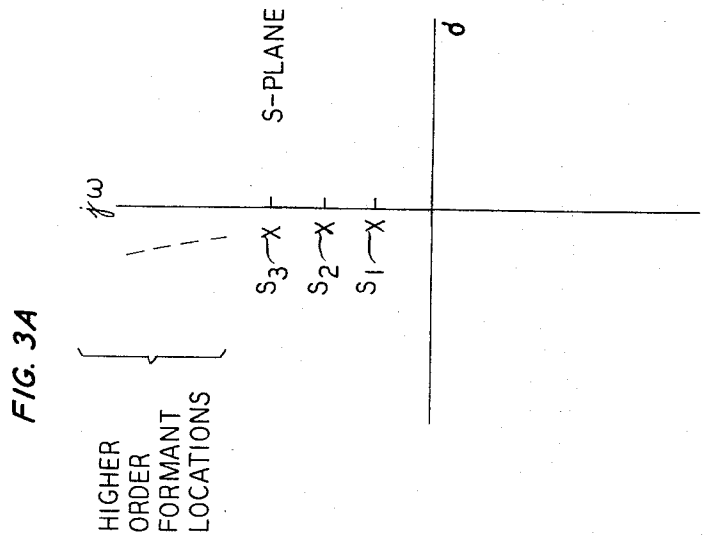


FIG. 3A

United States Patent Office 3,328,525 Patented June 27, 1967

3,328,525

SPEECH SYNTHESIZER

John L. Kelly, Jr., Berkeley Heights, N.J., assignor to Bell Telephone Laboratories, Incorporated, New York, N.Y., a corporation of New York

Filed Dec. 30, 1963, Ser. No. 334,354
4 Claims. (Cl. 179—1)

This invention relates to the synthesis of speech, and in particular to the synthesis of natural sounding speech in bandwidth compression systems.

Conventional speech communication systems, for example, commercial telephone system, typically convey human speech by transmitting an electrical facsimile of the acoustic waveform produced by a human talker. Because of the redundancy of human speech, however, facsimile transmission is a relatively inefficient way to transmit speech information, and it is well known that the information contained in a typical speech sound may be transmitted over a channel of substantially narrower bandwidth than that required for facsimile transmission of the speech waveform.

A number of arrangements for compressing or reducing the amount of bandwidth employed in the transmission of speech information have been proposed, one of the best known of these arrangements being the so-called "resonance vocoder." A specific version of the resonance vocoder is described in J. C. Steinberg Patent 2,635,146, issued Apr. 14, 1953.

The distinctive feature of resonance vocoder systems is the transmission of speech information in terms of narrow bandwidth control signals representative of the frequency locations of selected peaks or maxima in the speech amplitude spectrum which correspond to the principal formants or resonances of the human vocal tract. A typical resonance vocoder system includes at a transmitter station an analyzer for deriving from an incoming speech wave a group of narrow bandwidth control signals including formant contral signals representative of the frequencies of selected formant peaks in the speech spectrum. After transmission to a receiver station, the control signals are applied to a synthesizer that is provided with controllable resonant circuits for shaping an artificial spectrum to have peaks at frequencies specified by the formant control signals, thereby reconstructing a replica of the spectrum of the original speech wave.

From the standpoint of efficiency, it is of course desirable to transmit the smallest number of control signals consistent with a desired level of intelligibility and naturalness in the reconstructed speech wave. It is well known that the first three formants in order of frequency contribute most to the intelligibility of speech; accordingly, it is common practice to transmit three formant control signals representative of the locations of these three principal formant locations. From the standpoint of speech quality, however, it has been found that higher frequency formants contribute significantly to the naturalness of reconstructed speech, but of course the transmission of additional formant control signals requires a greater transmission channel bandwidth, thereby decreasing the bandwith efficiency of a resonance vocoder system.

One arrangement that improves the quality of reconstructed speech without decreasing bandwidth efficiency is described in an article by J. L. Flanagan and A. S. House, Development and Testing of a Formant—Coding Speech Compression System, volume 28, Journal of the Acoustical Society of America, page 1099, (1957). In the Flanagan-House system, the three principal lower order formants are represented by control signals that are transmitted from an analyzer to a synthesizer, as in a conventional resonance vocoder, but in addition to shaping an artificial

spectrum to have three peaks at frequencies specified by the three transmitted formant control signals, the synthesizer is provided with a separate fixed frequency resonant circuit that shapes the artificial spectrum to have a fourth peak at a fixed frequency corresponding to the average location of a fourth, high frequency formant. However, the human vocal tract is characterized by an infinite number of resonances or formants, and therefore the Flanagan-House arrangement does not specify completely the higher order speech formants.

The present invention improves the quality of speech reconstructed in a resonance vocoder without decreasing bandwidth efficiency by providing at a resonance vocoder receiver station a novel arrangement for shaping an artificial spectrum to have an infinite number of peaks at selected fixed frequencies corresponding to the locations of higher order speech formants. In the apparatus of this invention there is provided a novel resonant circuit having an infinite number of resonances at selected fixed frequencies, the higher frequency resonances of this resonant circuit corresponding to the frequencies of higher order speech formants. One or more of the lower frequency resonances of this circuit may lie within the frequency range of the formants represented by transmitted control signals, hence these lower frequency resonances are canceled or removed by separate antiresonant circuits having antiresonances at fixed frequencies corresponding to the unwanted lower frequency resonances of the resonant circuit. In the present invention an artificial spectrum is first shaped in conventional fashion by adjustable resonant circuits to have lower order formant peaks at frequencies specified by transmitted formant control signals, following which the artificial spectrum is further shaped by the apparatus of the present invention to have higher order formant peaks at fixed frequencies specified by the uncanceled resonances of the resonant circuit provided in this invention.

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

FIG. 1 is a block diagram showing a complete resonance vocoder system embodying the principles of this invention;

FIG. 2 is a circuit diagram showing in detail a specific embodiment of an antiresonant circuit employed in this invention; and

FIGS. 3A and 3B are diagrams of assistance in explaining the features of this invention.

Referring first to FIG. 1, elements **11** and **12** respectively represent the analyzer and synthesizer portions of a typical formant vocoder system. Formant vocoder analyzer **11**, which is ordinarily located at a transmitter station, includes a pitch detector **111**, a voiced amplitude detector **112**, and a formant frequency detector **113**, which respectively derive from a speech wave from source **10**, for example, a conventional microphone, a group of narrow bandwidth control signals respectively representative of the fundamental glottal excitation frequency, F_0, the amplitude of the glottal excitation, A_V, and the frequencies $F_1, \ldots, F_N$, $N=2,3, \ldots$, of selected maxima in the spectrum of the incoming speech wave. The frequency locations of these maxima in the speech spectrum correspond to natural frequencies of the formants or normal modes of vibration of the human vocal tract, and as the shape of the vocal tract is deformed during the articulation of different speech sounds, the natural or formant frequencies and the corresponding locations of spectral maxima also change. It is generally accepted that the most important formants are the three having the lowest frequencies, and in most formant vocoders, the formant control signals derived by an analyzer represent the first

three formants; that is, three is a suitable value for N in the apparatus of FIG. 1.

The control signals derived by analyzer **11** are delivered by way of a suitable transmission medium, indicated by broken lines, to a synthesizer **12** located at a receiver station. In synthesizer **12** there is provided a buzz source **121** that generates a relatively flat artificial amplitude spectrum comprising a plurality of relatively uniform amplitude harmonics of the fundamental frequency F_O, and an amplitude modulator **122** that adjusts the uniform amplitude of the harmonics of the artificial spectrum from source **121** to represent the glottal excitation amplitude A_V. Synthesizer **12** is further provided with a cascade of uncoupled resonant circuits **123–1** through **123–N**, each of which has an adjustable resonance that is individually tuned by a corresponding one of the N formant control signals transmitted from analyzer **11**. By successively passing the artificial spectrum of uniform amplitude harmonics from modulator **122** through circuits **123–1** through **123–N**, the spectrum is shaped by the resonances of circuits **123–1** through **123–N** to resemble the spectrum of the original speech wave from source **10**. Thus the adjustable resonances of circuits **123–1** through **123–N** shape the spectrum from modulator **122** to have N maxima at frequencies F_1 through F_N corresponding to the locations of the N formant peaks in the original speech spectrum which are represented by the N transmitted formant control signals.

It is evident that the synthesized spectrum developed at the output terminal of synthesizer **12** is limited in its resemblance to the original speech spectrum in that the number of peaks in synthesized spectrum is determined by the number of transmitted formant control signals. Thus in the usual situation where three formant control signals representative of the three principal speech formants are transmitted, the synthesized spectrum developed by synthesizer **12** has only three maxima. A closer resemblance to original speech spectrum is obtained in the present invention by passing the synthesized spectrum from synthesizer **12** through higher order formant synthesizer **13** in order to shape further the synthesized spectrum to have additional maxima at frequencies corresponding to the locations of maxima in the original speech spectrum which are not represented by transmitted control signals.

Within synthesizer **13**, the synthesized spectrum from synthesizer **12** is passed through M series-connected antiresonant circuits **131–1** through **131–M** to resonant circuit **133**, which comprises a delay line **133***a* of length τ seconds in negative feedback relation through subtractor **132** with an amplifier **133***b* having a gain $\epsilon^{-\sigma\tau}$ less than unity. As described in detail below, resonant circuit **133** theoretically has an infinite number of resonances at fixed frequencies dependent upon the length τ of delay line **133***b*. The fixed resonances of feedback circuit **133** shape the incoming synthesized spectrum from synthesizer **12** to have maxima which correspond to desired higher order formant locations not specified by the transmitted formant control signals. Since feedback circuit **133** has fixed resonances at low frequencies as well as at high frequencies, and since the synthesized spectrum from synthesizer **12** has already been shaped to have N maxima corresponding to N low frequency formants, one or more of the low frequency resonances of feedback circuit **133** must be canceled. Cancellation of M selected low frequency resonances of feedback circuit **133** is accomplished by the M antiresonant circuits **131–1** through **131–M**, $M=1,2,\ldots$, detailed illustration of a suitable antiresonant circuit being shown in FIG. 2 and explained below. After the synthesized spectrum from synthesizer **12** has been further shaped by synthesizer **13**, the spectrum may be converted into intelligible speech sounds by a suitable transducer **14**, for example, a conventional loudspeaker.

Before describing the theoretical considerations underlying the construction of resonant circuit **133**, it will be helpful at this point to describe the properties of the human vocal tract during vowel production in terms of Laplace transform notation. 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 glottis, $U_2(t)$, this ratio being commonly known as the transfer characteristic of the vocal tract, can be approximated by a rational transfer function having the following form:

$$\frac{U_2(s)}{U_1(s)}=\frac{k^{\Pi}s_k-s_k^*}{k^{\Pi}(s-s_k)(s-s_k^*)} \tag{1}$$

where $s=(\sigma+jw)$ denotes the complex frequency variable; $s_k=(\sigma_k+jw_k)$ is a complex number representing a formant or normal mode of vibration of the vocal tract; and s_k^* is the complex conjugate of s_k. Equation 1 indicates that the glottis-to-mouth transfer functions for vowel sounds has only poles, denoted s_k, and no zeros, and the the poles coincide with the normal modes of vibration. The locations of these poles are illustrated graphically in the pole diagram of FIG. 3A, where the X's indicate the locations in the s-plane of the complex numbers s_1, s_2, s_3, representing the first three poles or formants, and the dashed lines indicate higher frequency poles.

Since the vocal tract is a distributed acoustic system, it has in theory an infinite number of natural frequencies which change in value with time as the vocal tract is deformed during the articulation of different speech sounds, and correspondingly, the poles of the transfer characteristic in Equation 1 also change with time. However, only at relatively low frequencies, usually including no more than the first three natural frequencies, do the natural frequencies change substantially in value with deformations of the vocal tract, whereas at high frequencies, the natural frequencies asymptotically approach a uniform spacing in frequency.

Synthesizer **13** of the present invention is provided with an infinite number of natural resonant frequencies having a uniform spacing in frequency corresponding to that of the higher order natural frequencies of the vocal tract by constructing resonant circuit **133** and antiresonant circuits **131–1** through **131–N** in the following manner. As shown in FIG. 1, resonant circuit **133** comprises delay line **133***a* of length τ in feedback relation through subtractor **132** with amplifier **133***b* having gain $\epsilon^{-\sigma\tau}$, where a suitable value for $\epsilon^{-\sigma\tau}$ may be on the order of 2/3. In Laplace transform notation, the transfer characteristic of delay line **133***a* is $\epsilon^{-s\tau}$, hence the combined transfer characteristic of delay line **133***a* and amplifier **133***b* is the product.

$$\epsilon^{-s\tau}\cdot\epsilon^{-\sigma\tau}=\epsilon^{-\tau(s+\sigma)} \tag{2}$$

Since the elements **133***a* and **133***b* are in negative feedback relation with each other through subtractor **132**, the product in Equation 2 corresponds to the familiar β in the transfer characteristic relating the incoming signal F_1 applied to the minuend terminal of subtractor **132** to the outgoing signal F_2 developed at the output terminal of circuit **133**,

$$\frac{F_2}{F_1}=\frac{1}{1+\beta}=\frac{1}{1+\epsilon^{-\tau(s+\sigma)}} \tag{3}$$

Equation 3 indicates that the transfer characteristic F_2/F_1 has poles at the points where

$$\epsilon^{-\tau(s+\sigma)}=-1 \tag{4}$$

and since $\epsilon^{-s}=1$ is periodic, the positive frequency poles of F_2/F_1 occur at odd integral multiples of $j\pi$,

$$-\tau(s_n+\sigma)=-j(n\pi)$$

$$s_n=\sigma+j\frac{n}{\tau}\pi,\ (n=1,3,\ldots) \tag{5a}$$

From Equation 5*a* it is evident that F_2/F_1 has an infinite number of poles, the radian frequency of the nth pole being

$$\omega_n=\frac{n}{\tau}\pi,\ (n=1,\ 3,\ \ldots) \tag{5b}$$

or, since $\omega_n=2\pi f_n$ denotes frequency in cycles per second,

$$f_n=\frac{n}{2\tau},\ (n=1,\ 3,\ \ldots) \tag{5c}$$

The spacing in frequency between poles is uniform, being $2\pi/\tau$ in radius and $1/\tau$ in cycles per second. Further, each pole has the same constant real part, $-\sigma$, which represents the so-called formant damping, and which is manifested in the speech spectrum by the bandwidth of the formant peaks.

FIG. 3B illustrates graphically the locations of the poles of F_2/F_1, in which it is noted that a particular uniform spacing in frequency may be obtained by suitably choosing the length τ of delay line **133***a* and the corresponding factor τ in the gain of amplifier **133***b*. A suitable value for the length of delay line **133***a* τ may be on the order of one millisecond, which corresponds to the round trip delay of the human vocal tract, thereby placing the poles of feedback circuit **133** at frequencies of approximately 500, 1500, 2500 . . . , cycles per second. Similarly, a suitable value for the formant bandwidth factor σ in the gain of amplifier **133***b* may be on the order of 400 nepers per second, corresponding to a formant bandwidth of about 130 cycles per second.

Depending upon the value selected for the length of delay line **133***a*, one or more of the fixed poles or resonances of resonant circuit **133** may occur within the frequency range of formants represented by the control signals transmitted from analyzer **11** to synthesizer **12**. For example, by selecting the length of delay line **133***a* to be one millisecond, the first three fixed poles of circuit **133** occur at 500, 1500, and 2500 cycles per second, which respectively lie within the frequency ranges of the first three speech formants typically represented by transmitted control signals. In this situation it is therefore necessary to remove or cancel the three lower order poles of resonant circuit **133** which lie within the frequency range of the formants represented by transmitted control signals in order to prevent interference with the maxima previously synthesized in the artificial spectrum from synthesizer **12**. FIG. 3B illusrates the situation in which the first three poles of circuit **133** are to be canceled, as indicated by the three "X's" enclosed in circles.

Antiresonant circuits **131–1** through **131–M**, which precede feedback circuit **133**, are designed to cancel M=1,2, . . . , unwanted lower order poles of feedback circuit **133** in the following manner. In order for a particular antiresonant circuit, say **131–1**, to cancel a corresponding one of the poles of resonant circuit **133**, say the pole denoted $s_1=(-\sigma+j\omega_1)$, it is necessary for the transfer characteristic of circuit **131–1** to have a single zero at $s_1=(-\sigma+j\omega)$, and no other poles or zeros. That is, if Z_1 denotes the transfer characteristic of circuit **131–1**, then Z_1 must be proportional to $(s-s_1)$ $(s-s_1^*)$, or

$$Z_1=A[(s-s_1)(s-s_1^*)] \tag{6}$$

where A is a constant. A suitable realization of a circuit having a transfer characteristic of the type specified by Equation 6 is shown in detail in FIG. 2, it being understood that other antiresonant circuits having a suitable transfer characteristic may be employed, if desired. Further, it is understood that the transfer characteristic required for the cancellation of other unwanted poles by antiresonant circuits, **131–2**, . . . , **131–M**, may be obtained by substituting other poles s_2, s_2^*; . . . ; s_M, s_M^* for the quantities s_1, s_1^* in Equation 6.

Turning now to FIG. 2, the synthesized speech spectrum from synthesizer **12** is applied through a sufficiently high resistance R_3 to pass a constant current through the series connected inductance, resistance and capacitance elements respectively denoted L_1, R_1, and C_1, and to apply a constant voltage to cathode folower V_1. The output voltage of cathode follower V_1 is differentiated by capacitance element C_2 and resistance element R_2, and the differentiated output signal is passed to the next antiresonant circuit.

The impedance of elements L_1, R_1, C_1, in Laplace transform notation, may be written

$$L_1s+R_1+\frac{C_1}{s}$$

and differentiating elements R_2 and C_2 change this impedance by a multiplicative factor s, so that the transfer characteristic of circuit **131–1** may be written

$$Z_1=s\left(L_1s+R_1+\frac{C_1}{s}\right)$$

$$=s^2+\frac{R_1}{L_1}+\frac{C_1}{L_1} \tag{7}$$

Equating the right-hand side of Equation 6, without the constant factor A, to the right-hand side of Equation 7 and expanding the product $(s-s_1)(s-s_1^*)$,

$$s^2+\frac{R_1}{L_1}s+\frac{C_1}{L_1}=s^2+2\sigma s+(\sigma^2+\omega_1^2) \tag{8}$$

From Equation 8 it is evident that the values of the inductance, resistance, and capacitance elements L_1, R_1, and C_1 may be determined from the predetermined values of σ and ω_1 according to the following relations:

$$\frac{R_1}{L_1}=2\sigma \tag{9a}$$

$$\frac{C_1}{L_1}=\sigma^2+\omega_1^2 \tag{9b}$$

$$L_1=1 \tag{9c}$$

or

$$R_1=2\sigma \tag{9d}$$

$$C_1=\sigma^2+\omega_1^2 \tag{9e}$$

Although this invention has been described in terms of a resonance vocoder system of the type shown in FIG. 1, it is to be understood that applications of the principles of this invention are not limited to this particular system, but include other resonance vocoder systems as well as various kinds of speech processing equipment in which speech formants are synthesized. In addition, it is to be understood that the above-described embodiments are merely illustrative of the numerous arrangements that may be devised for 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. In a resonance vocoder synthesizer,

a source of a plurality of control signals including a pitch control signal, an amplitude control signal, and a group of formant control signals representative of the frequencies of selected low frequency formant peaks in the spectrum of an original speech wave,

first synthesizing means responsive to said plurality of control signals for developing an artificial speech spectrum having a first group of peaks at frequencies represented by said formant control signals, and

second synthesizing means for shaping said artificial speech spectrum to have a second group of peaks at selected fixed frequencies representative of high frequency speech formants, said second synthesizing means including

a resonant circuit having a transfer characteristic with no zeroes and an infinite number of poles at equally spaced predetermined frequencies, wherein the higher frequency poles of said transfer characteristic corre-

spond in frequency to high frequency speech formants, and

a plurality of series-connected antiresonant circuits preceding said resonant circuit, each of said antiresonant circuits having a transfer characteristic with no poles and a zero at a predetermined frequency corresponding to an unwanted pole in said transfer characteristic of said resonant circuit, and

means for applying said artificial speech spectrum to said second synthesizing means.

2. In combination with a resonance vocoder synthesizer that generates an artificial speech spectrum having peaks at selected low frequencies corresponding to selected low frequency formant peaks in the spectrum of an original speech wave, apparatus for introducing additional peaks into said artificial speech spectrum at selected frequencies corresponding to selected high frequency formants which comprises

a resonant circuit having a transfer characteristic with no zeros and an infinite number of poles at equally spaced fixed frequencies, wherein the higher frequency poles of said transfer characteristic correspond in frequency to selected high frequency speech formants, and

a plurality of series-connected antiresonant circuits in preceding circuit relation with said resonant circuit for cancelling a corresponding plurality of unwanted poles in the transfer characteristic of said resonant circuit.

3. Apparatus for synthesizing a plurality of peaks of predetermined width at selected fixed frequencies in an incoming amplitude spectrum which comprises

a plurality of series-connected antiresonant circuits each provided with a transfer characteristic having a single zero at

$$s_n=\left(-\delta+j\frac{n}{\tau}\pi\right),\ (n=1,\ 3,\ \ldots)$$

for preventing the occurrence of peaks at unwanted complex frequencies, said plurality of antiresonant circuits being provided with an input point and an output point,

a resonant circuit including a delay element having a delay of τ seconds in negative feedback relation with an amplifier having a gain characteristic $\epsilon^{-\sigma\tau}$, where σ represents said predetermined bandwidth for said peaks, so that said resonant circuit has a transfer characteristic with an infinite number of poles, s_n where

$$s_n=-\sigma+\frac{n}{\tau}\pi,\ (n=1,\ 3,\ \ldots)$$

means for applying said incoming amplitude spectrum to said input point of said plurality of antiresonant circuits, and

means for conencting the output point of said plurality of antiresonant circuits to said resonant circuit.

4. Apparatus for synthesizing an artificial spectrum having a plurality of peaks at selected high frequency locations so that said peaks in said artificial spectrum closely resemble the formant peaks at high frequency locations in the spectrum of an original speech wave, which comprises

means for developing an incoming artificial spectrum having peaks at selected low frequency locations corresponding to formant peaks at selected low frequency locations in said spectrum of said original speech wave,

a plurality $M(M=1, 2, \ldots)$ of series-connected antiresonant circuits for preventing the occurrence of peaks at a corresponding plurality of unwanted high

frequency locations, wherein the nth of said antiresonant circuits, $n=1,2, \ldots, M$, comprises

an input terminal,

a cathode follower provided with an input point and an output point,

a first resistance element connected between said input terminal and said input point of said cathode follower,

an inductance element L, a second resistance element R_n, and a first capacitance element C_n connected in series between said input point of said cathode follower and ground,

an output terminal,

a second capacitance element connected between said output point of said cathode follower and said output terminal, and

a third resistance element connected between said output terminal and ground,

wherein the values of said inductance element L, said second resistance element R_n, and said first capacitance element C_n are determined by the bandwidth σ and the frequency ω_n of said unwanted peak at said high frequency location according to the following relationships:

$$L=1$$
$$R=2\sigma$$
$$C=\sigma^2+\omega_n^2$$

means for aplying said incoming artificial spectrum to said input terminal of the first of said antiresonant circuits, $n=1$,

a resonant circuit including

subtracting means provided with a minuend terminal, a subtrahend terminal and an output terminal,

a delay element having a delay time of τ seconds and provided with an input terminal and an output terminal,

an amplifier having a gain $\epsilon^{-\sigma}\tau$ so that said resonant circuit has an infinite number of resonances at frequencies

$$\omega_i=\frac{(2i-1)}{\tau}\pi,\ i=1,\ 2,\ \ldots$$

said amplifier being provided with an input terminal and an output terminal,

means for connecting said output terminal of the last of said antiresonant circuits, $n=M$, to said minuend terminal of said subtracting means,

means for connecting said output terminal of said subtracting means to said input terminal of said delay element,

means for connecting said output terminal of said delay element to said input terminal of said amplifier, and

means for connecting said output terminal of said amplifier to said subtrahend terminal of said subtracting means,

thereby developing from said incoming artificial spectrum an outgoing artificial spectrum at said output terminal of said delay element, wherein said outgoing artificial spectrum has peaks at selected low frequency locations corresponding to said peaks of said incoming artificial spectrum and peaks at selected high frequency locations corresponding to resonances of said resonant circuit at frequencies ω_i, for i greater than M.

No references cited.

KATHLEEN H. CLAFFY, *Primary Examiner.*

R. MURRAY, *Assistant Examiner.*