Oct. 14, 1969 V. D. HOGUE 3,472,964

VOCAL RESPONSE SYNTHESIZER

Filed Dec. 29, 1965 4 Sheets-Sheet 1

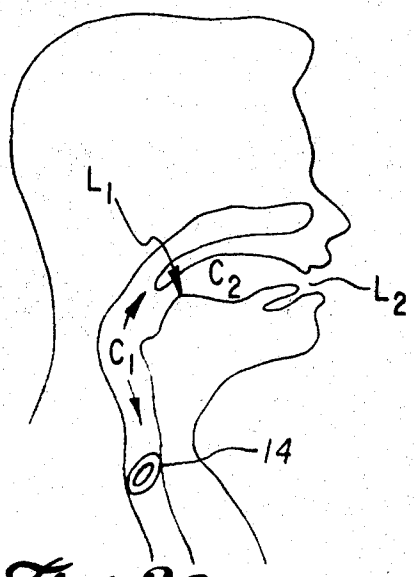


Fig. 2a

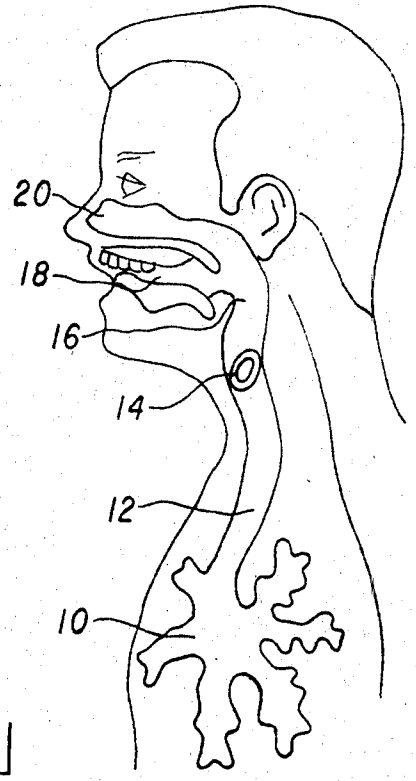


Fig. 1

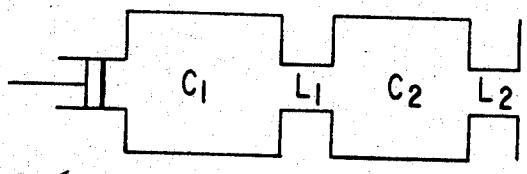


Fig. 2b

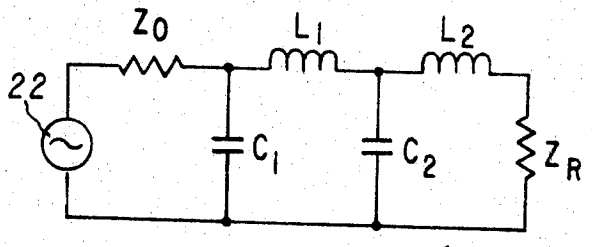


Fig. 2c.

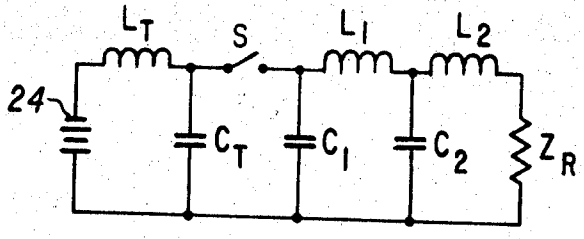


Fig. 2d

INVENTOR
VERNON D. HOGUE

BY René E. Grossman

ATTORNEY

VOCAL RESPONSE SYNTHESIZER

Filed Dec. 29, 1965 4 Sheets-Sheet 2

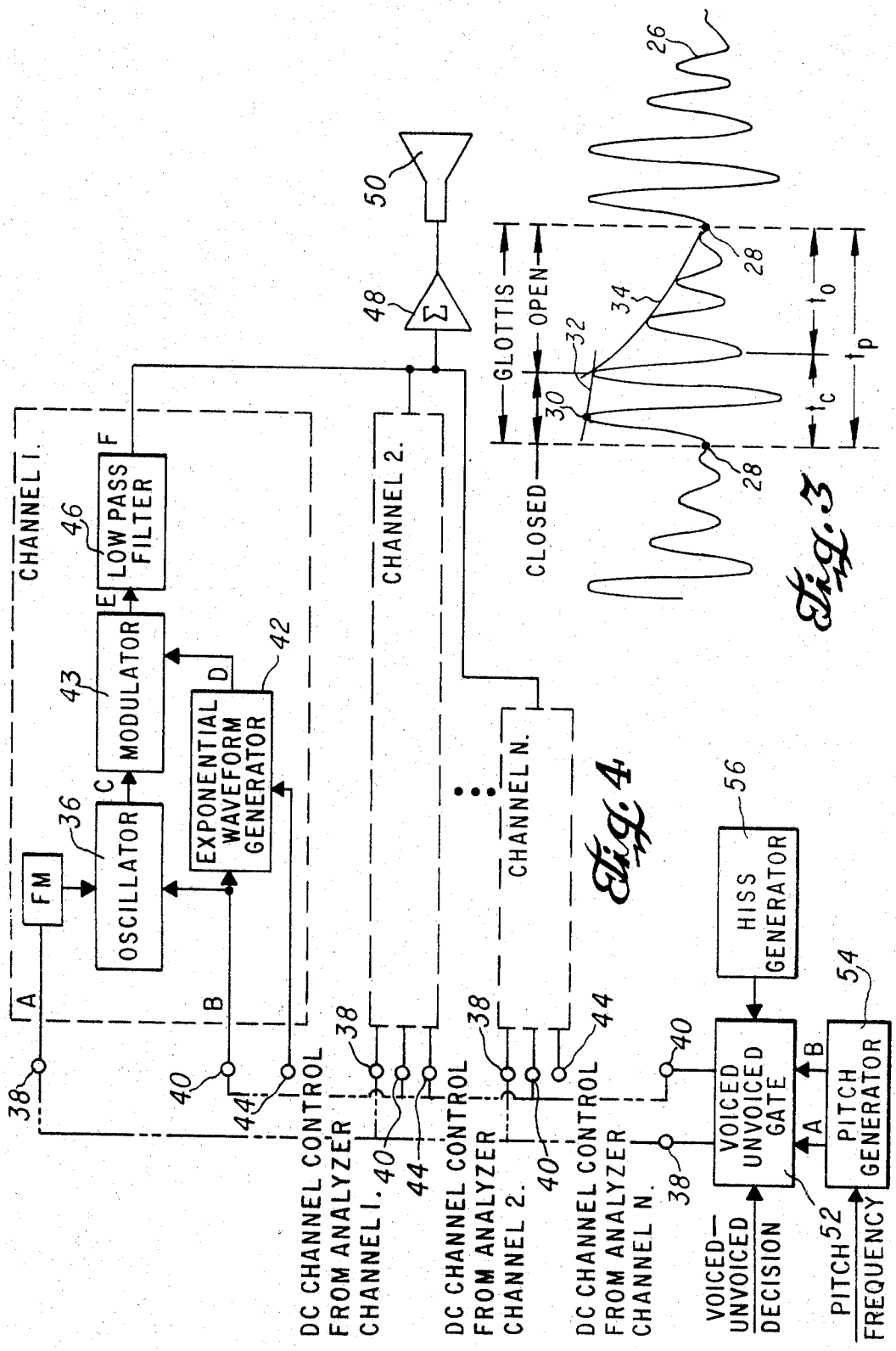


Fig. 4

Fig. 3

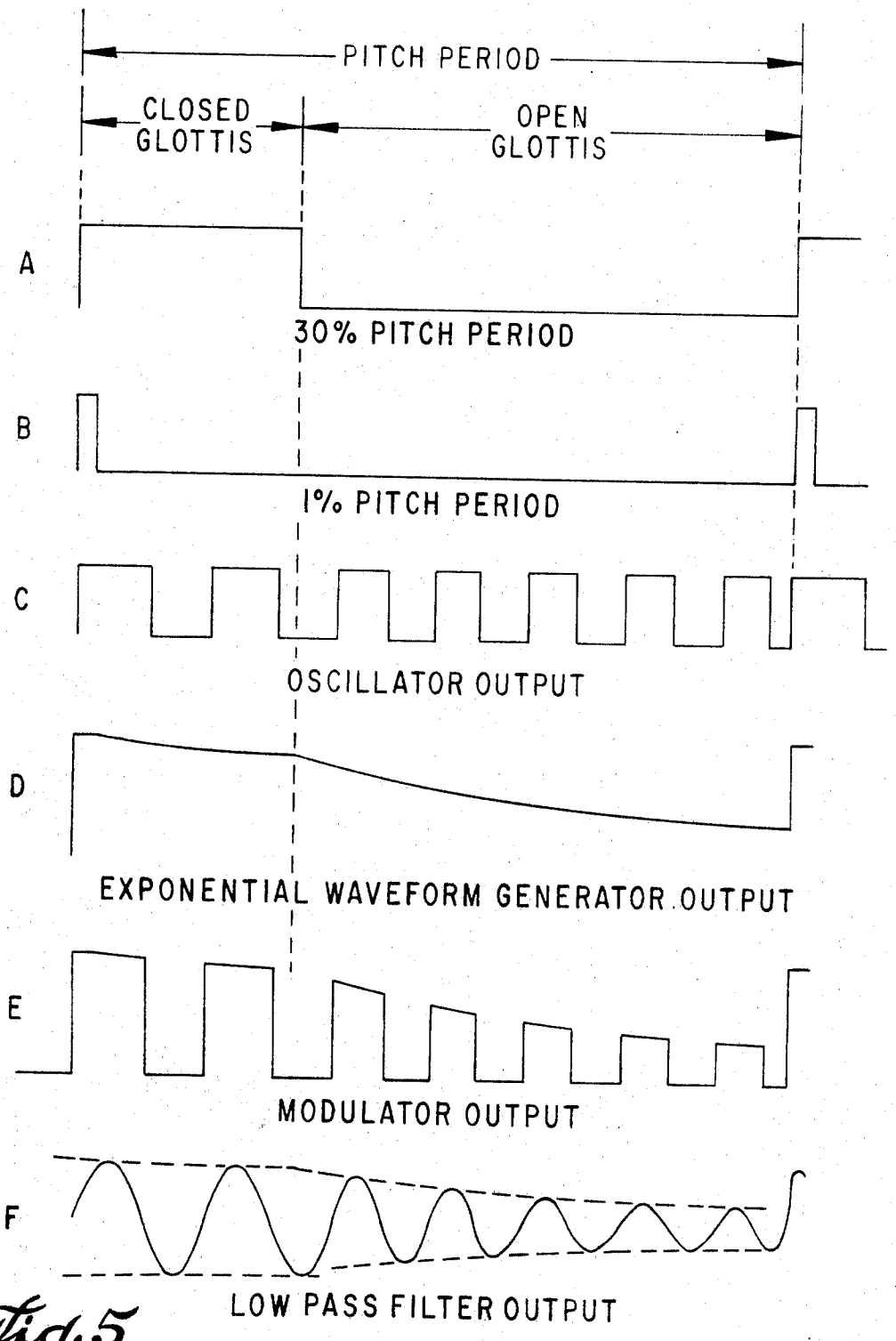

PITCH PERIOD
CLOSED GLOTTIS
OPEN GLOTTIS
A
30% PITCH PERIOD
B
1% PITCH PERIOD
C
OSCILLATOR OUTPUT
D
EXPONENTIAL WAVEFORM GENERATOR OUTPUT
E
MODULATOR OUTPUT
F
LOW PASS FILTER OUTPUT

Fig. 5

Oct. 14, 1969 V. D. HOGUE 3,472,964

VOCAL RESPONSE SYNTHESIZER

Filed Dec. 29, 1965 4 Sheets-Sheet 4

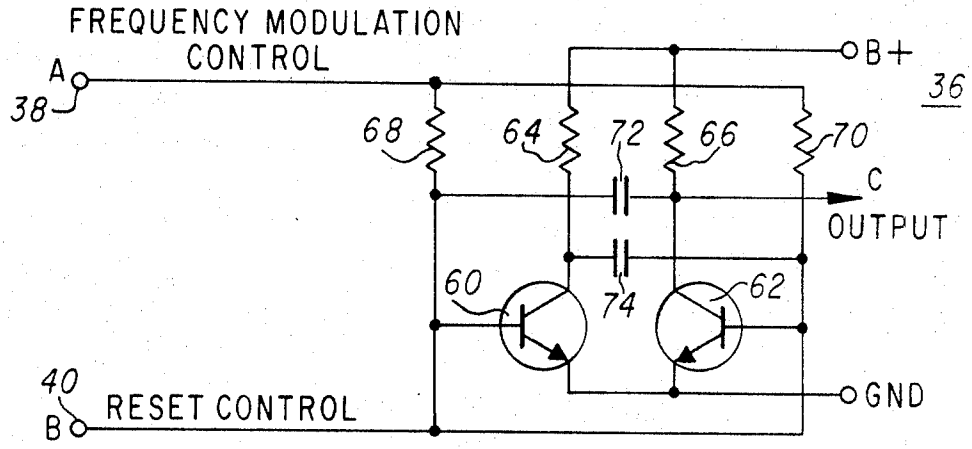


Fig. 6

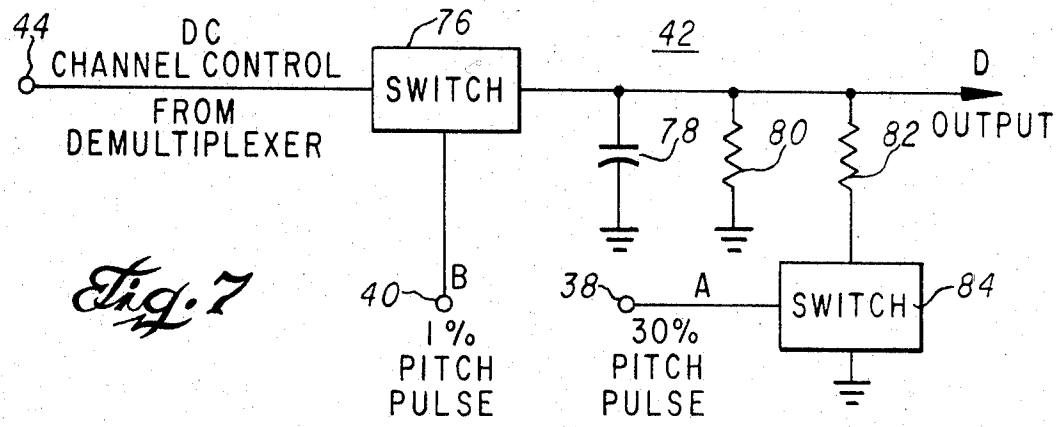


Fig. 7

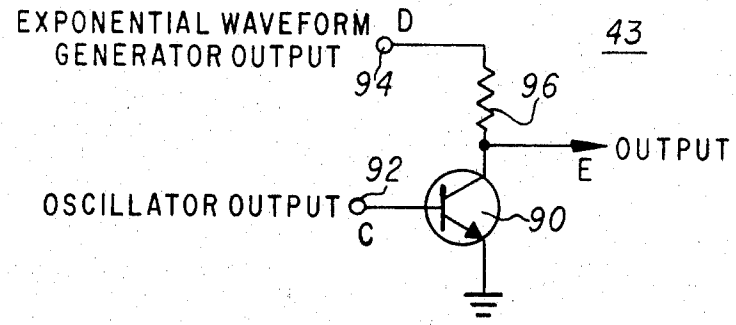


Fig. 8

United States Patent Office 3,472,964
Patented Oct. 14, 1969

3,472,964

VOCAL RESPONSE SYNTHESIZER

Vernon D. Hogue, Dallas, Tex., assignor to Texas Instruments Incorporated, Dallas, Tex., a corporation of Delaware

Filed Dec. 29, 1965, Ser. No. 517,231
Int. Cl. H03b 25/00
U.S. Cl. 179—1 8 Claims

ABSTRACT OF THE DISCLOSURE

Disclosed is an apparatus for synthesizing the sound produced by the changing resonant conditions in the cavities of the human vocal track during a pitch period, said apparatus including a free running oscillator whose output signal is frequency modulated at one frequency during the closed glottis condition and a second frequency during the open glottis condition, an exponential waveform generator for producing first and second exponential decays during the closed and open glottis conditions, and means for modulating the frequency modulated oscillator output signal with the output from the exponential waveform generator.

This invention relates to the synthesis of speech and more particularly to the apparatus for more closely approximating the human vocal speech waveforms.

A typical vocoder system comprises both an analyzer and a synthesizer. The analyzer forms no part of this invention but for the purpose of clarity, a brief description of a typical analyzer will afford a better understanding of the workings of the novel synthesizer portion of the present invention. An analyzer (not shown) is a device which takes an original sound and analyzes it for it characteristics of fundamental pitch and fundamental and harmonic energy distribution. This analysis is performed by the use of a plurality of channels, each including a bandpass filter, a full-wave rectifier and a low pass filter. Each bandpass filter will pass a different band of harmonic frequencies having a certain intensity, the full wave rectifier will rectify the AC signal obtained and the low pass filter will then filter out the high frequencies in the waveform and give a DC level which is proportional to the intensity of the frequencies occurring in the bandpass filter of that particular channel. This DC level information from the low pass filters is then fed into an analyzer multiplexer, thereby producing a serial train of DC levels representative of the intensity of the frequency spectrum in the various bandpass filters.

Further information which is derived from the original sound in the analyzer portion is the fundamental pitch frequency and whether the sounds are voiced or unvoiced speech sounds. Voiced sounds are those sounds in which the acoustic energy is derived from the vibration of the vocal cords and the frequency spectrum is characterized by discrete frequency components. Unvoiced or fricative sounds are not derived from the vibration of the vocal cords but from the passage of an air stream through a restricted aperture at some point in the mouth or throat; these unvoiced sounds produce a relatively high pitched noise or "hiss." The fundamental pitch frequency information and the information from an unvoiced-voiced detector which describes the type of speech sound present are also sent to the multiplexer for subsequent use in the synthesizer and may be put in digital form and transmitted to the synthesizer which then reconstructs as nearly as possible the original speech sound.

In a conventional system for the artificial production of vocal or other sounds a bank of bandpass filters is needed in the synthesizer that is similar in bandwidth and center frequency to the corresponding analyzer bandpass filters.

The synthesizer filters are excited with a pulse whose repetition rate is that of the pitch fundamental for a voiced sound or a random repetition rate for an unvoiced sound. The amplitude of these pulses, at the input of each synthesizer bandpass filter, is controlled by a slowly varying DC level from the respective analyzer channel. The output of each synthesizer bandpass filter is summed to form the synthesized speech output. Thus, this type of synthesizer requires many bandpass filters, the quality of reproduction improving with the increase in number of channel filters.

The result of exciting a bandpass filter with a pulse, as in the above-mentioned conventional synthesizer, is a slow buildup and a non-exponential decaying speech waveform. This accomplishes the opposite response from the actual human vocal cavities being pulsed at a pitch frequency rate, the actual response being a rapid buildup and exponential decay (bandwidth modulation) of the speech waveform. This characteristic of the vocal tract has not been given consideration in prior art synthesizers.

Another characteristic that has been completely disregarded heretofore in the synthesis of speech is the frequency modulation of the speech waveform due to the loading of the vocal tract during that portion of the pitch period when the glottis or vocal cords are open. The frequency modulation is produced because the length of the vocal tract is being modulated. During that fraction of the pitch period when the glottis is open, the sub-glottal region becomes an extension of the vocal tract, thus increasing the volume of the back cavity causing a frequency variation.

The above two characteristics or second-order modulations occurring in the human vocal system in a single fundamental pitch period have been largely ignored in the synthesis of speech.

It is therefore an object of this invention to provide an apparatus which superimposes the second-order modulations on the synthesized speech waveforms during a single fundamental pitch period.

Another object of this invention is to provide a synthesizer which eliminates the need for bandpass filters.

A further object of the invention is to provide a new and improved speech synthesizer which lends itself to microminiaturization because of the absence of inductors necessary for the channel filters, said filters comprising the bulk of the size and weight of a conventional synthesizer.

A still further object of the present invention is to provide a synthesizer which is economical in cost, size and weight.

For a more complete understanding of the present invention and for further objects and advantages thereof, reference may now be had to the following description taken in conjunction with the appended claims in which:

FIGURE 1 is a cross section of the human vocal system.

FIGURE 2*a* is a cross section of the human vocal tract with its representative acoustical impedances.

FIGURE 2*b* is the acoustic analog of the human vocal tract.

FIGURE 2*c* is the conventional electrical analog of the human vocal tract.

FIGURE 2*d* is a modified electrical analog of the human vocal tract.

FIGURE 3 is the steady state vowel waveform for one band of harmonics to be reconstructed in one channel of the synthesizer.

FIGURE 4 is the block diagram of the synthesizer system of the present invention.

FIGURE 5 is the waveform outputs at various points in the system of FIGURE 4.

FIGURE 6 shows in schematic form the basic oscillator circuit.

FIGURE 7 is the circuit for the envelope generator.

FIGURE 8 is the modulator circuit of the present invention.

Referring now to FIGURE 1, there is illustrated a cross section of the human vocal system with the lungs **10**, trachea **12**, glottis or vocal cords **14**, pharynx **16**, oral cavities **18** and nasal cavities **20** shown in their respective positions. The lungs are the primary energy source for most speech sound, providing a stream of air which flows through the trachea, glottis, pharynx, mouth and nose to the outside world. To produce audible sounds, the air stream is modulated. For voiced sound, this is accomplished by the glottis **14** which opens and closes in a periodic manner and produces the fundamental pitch frequency or period. The glottis vibrates in other ways to generate a harmonic spectrum. The pitch frequency will vary for voiced sounds depending upon the tone changes of the individual speaker. The spectrum produced by the glottis is modified before reaching the outside world by the resonances of the pharynx **16** and oral and nasal cavities **18** and **20**, respectively. These resonant cavities reinforce harmonics of the glottis spectrum in certain frequency regions thereby producing formants of speech.

FIGURE 2*a* illustrates the cross section of the human vocal tract with representative acoustic impedances included therein. The acoustic analog of FIGURE 2*b* illustrates the two resonant LC circuits (C_1L_1 and C_2L_2) which represent the pharynx and oral and nasal cavities as shown in FIGURE 2*a*.

FIGURE 2*c* illustrates the conventional electrical analog of the human vocal tract and comprises a voltage generator **22** which represents the driving function of the glottis **14** (of FIGURE 2*a*) and Z_o characterizes the impedance of the glottis. Connected to one end of the characteristic impedance Z_o and the voltage generator **22** are the electrical analogs of the resonant cavities C_1, L_1, and C_1, L_2, which terminate into load impedance Z_R.

The electrical analog of FIGURE 2*c* does not take into account the second-order modulations before-mentioned which occur in the human vocal tract. FIGURE 2*d* illustrates the modified electrical analog required to take into effect these modulations. The electrical analog of FIGURE 2*c* does not take into account the change in the characteristic impedance Z_o as the glottis opens and closes during a pitch period. L_T and C_T of FIGURE 2*d* represent the impedances looking from the glottis back into the trachea. The switch S connected between the common terminal of L_T and C_T and the tuned circuits represents the opening and closing of the glottis while the DC source **24** represents the lungs **10** (of FIGURE 1). In the electrical analog of FIGURE 2*d*, if the switch remains closed, the circuit reaches a steady state current condition; if the switch is opened, the current falls to zero. If the switch is opened and closed periodically, a harmonically-related AC signal is superimposed upon the DC level which is analogous to the action that takes place in the human vocal tract.

It is important to note that the second-order modulations occur because the resonances are slightly detuned by the impedances looking back into the trachea when the switch S is closed. This causes a frequency and Q shift in the tuned circuits each time switch S (that is, the glottis) opens and closes. This, in turn, creates a frequency modulation and a bandwidth modulation of the speech waveform during a pitch period.

FIGURE 3 illustrates the steady state vowel waveform **26** for one band of harmonics to be reconstructed in one channel of the synthesizer. The fundamental pitch period, t_p, is the time required for the glottis to close, open and then close again. The period that the glottis is closed is represented on waveform **26** by t_c while the open period is represented by t_o. Certain features which are present in waveform **26** which are important in the synthesis of speech should be noted. Firstly, there is a phase discontinuity at the beginning of the closed glottis event which means that the waveform at point **28** will always start having the same phase. Therefore in the synthesis of this waveform at the beginning of each pitch period, it will be necessary to reset some type of frequency generator so that the proper phase relationship is repeated for each cycle.

The second feature of note is the rapid buildup and exponential decay of waveform **26**. As was mentioned previously, in conventional speech synthesizers which used the pulsing of a bandpass filter, a response was obtained opposite to the one actually occurring as illustrated in FIGURE 3, that being the rapid buildup and exponential decay. The apparatus of this invention incorporates means to generate this type of function.

The third feature observed from the waveform **26** is the change of decay rate (bandwidth) with the change in glottal conditions. The Q of a resonant cavity is defined as the ratio of the resonant frequency of the formant to the bandwidth of the formant of speech. With the glottis (of FIGURE 2*a*) closed, the resonant cavities of the vocal tract form a high Q circuit (that is, narrow bandwidth) which means that there will be a slow rate of decay of the steady state vowel waveform during the period that the glottis is closed. This is illustrated by exponential decaying line **32** which connects the peaks **30** of the speech waveform during the closed glottis condition and illustrates a slow rate of decay. As was explained earlier, when the glottis opens, additional loading is created by the sub-glottal region which becomes an extension of the resonant cavities, thus lowering the Q of the resonant circuits. Accordingly, for the open glottis condition, a low Q resonant cavity is present in the tract which, in turn, means that there is a larger bandwidth present and the speech waveform will decay at a faster rate than in the closed glottis condition. This can be seen from line **34** which connects the peaks of the waveform **26** with the envelope defined by line **34** describing a greater exponential decay rate than the envelope defined by line **32**. Therefore in going from the closed to the open glottis conditions, the bandwidth of the resonant circuits in the vocal tract is changed or modulated as shown by the two exponential decay rates defined by lines **32** and **34**.

The fourth feature which should be noted in waveform **26** is the change of resonant frequency from the closed glottis to the open glottis condition with the lower frequency occurring in the closed glottis state. The frequency difference between these two states varies between 30 and 200 c.p.s. Thus, in going from the closed to the open glottis condition there is a frequency modulation of waveform **26** with the higher frequency occurring in the open glottis state.

FIGURE 4 illustrates the system block diagram of the synthesizer of the present invention while FIGURE 5 represents the output waveform occurring at various points in the system of FIGURE 4. The synthesizer of FIGURE 4 will comprise a demultiplexer (not shown) which converts the serial DC information from the analyzer multiplexer into parallel form for subsequent use in the DC channel control portion of the synthesizer. The synthesizer has N channels, the number of channels corresponding to the number of channels in the analyzer portion of the vocoder. The portion of FIGURE 4 labeled channel **1** discloses the details of a typical channel. It comprises a free-running oscillator **36** oscillating at a first and second frequency below and above the center frequency to which its corresponding analyzer bandpass filter is tuned (the center frequency of the bandpass filter in analyzer channel **1**, in this case) and having input terminals **38** and **40**. A pulse is provided at input terminal **38** (waveform A of FIGURE 5, the 30% pitch period pulse) which varies or modulates the frequency of oscillator **36** at a first rate during the closed glottis condition and a second rate during the open glottis condition. The ratio of the closed-to-total pitch period of the glottis varies over a wide range depending upon the speaker. This ratio for male speakers has been observed to be a

mean of 37%, but for purposes of illustration, waveform A assumes a 30% ratio. Waveform B of FIGURE 5, the 1% pitch period pulse, is provided at input terminal **40** and resets the oscillator at the beginning of each pitch period. Accordingly, the output from the oscillator (waveform C of FIGURE 5) is a frequency modulated wave which is reset at the initiation of the pitch period.

Exponential waveform generator **42** provides the bandwidth modulation required for channel **1** of the synthesizer. This generator provides an output (waveform D of FIGURE 5) having two varying exponential decay rates, the greater exponential decay rate occurring during the open glottis condition while the slower decay rate occurs during the closed glottis condition. Waveform B (the 1% pitch pulse) which resets oscillator **36** also resets exponential waveform generator **42**, enabling pitch synchronism to be obtained. The initial amplitude of the exponential waveform generator output (waveform D) is determined by the DC control level from analyzer channel **1**. As was explained above, the output from the low pass filter of analyzer channel **1** represents a DC level which is proportional to the intensity of the frequencies occurring in the bandpass filter for analyzer channel **1**. This DC level is obtained by the synthesizer from the demultiplexer (not shown) and provided to terminal **44**. Accordingly, the output waveform of generator **42** describeds a waveform having an initial amplitude equal to the DC level of the corresponding analyzer channel with a first exponentially decreasing waveform occurring while the glottis is closed and a second exponentially decreasing waveform (having a greater decay rate than the first decaying waveform) continuing until the initiation of the next pitch period.

The outputs from oscillator **36** and generator **42** are fed to modulator **43**. This modulator modulates the oscillator output with the exponential waveform generator output to cause the amplitude of the oscillator output to vary in accordance with the amplitude of the output of waveform generator **42** as shown by waveform E of FIGURE 5. The output of modulator **43** is fed into low pass filter **46** which filters out the high frequencies in the output of modulator **43** which results in an output as shown in waveform F of FIGURE 5. The output from filter **46** and all of the other outputs from the other channels in the synthesizer are combined and fed into summing means **48** which sums together all of the outputs from the respective low pass filters of the various channels and reconstructs the original speech waveform which is heard from speaker **50**.

The voiced-unvoiced decision information obtained in the analyzer portion of the vocoder is transmitted from the demultiplexer (not shown) to the voiced-unvoiced gate **52**. During the time that a voiced event occurs, gate **52** allows pitch generator **54** to transmit pulses A and B via terminals **38** and **40** to the synthesizer channels which are in time synchronism with the pitch frequency determined by the analyzer (that is, with the opening and closing of the glottis). This means that during the time that a voiced decision is occurring, oscillator **36** in the various channels will be reset in synchronism with the pitch frequency and a predetermined frequency modulation of each oscillator output will occur.

When an unvoiced sound occurs, gate **52** blocks the pitch generator **54** and allows the hiss generator output to be applied to terminals **38** and **40**. The output of hiss generator **56** comprises an analog output of randomly occurring frequencies. It will be remembered that when an unvoiced sound is made the glottis is not in use, and accordingly the oscillators are allowed to be reset and frequency modulated at a random rate.

FIGURE 6 illustrates the schematic diagram of oscillator **36** shown in FIGURE 5. Oscillator **36** is a free-running multivibrator which may be reset at terminal **40** and frequency modulated at terminal **38**. The reset terminal **40** is electrically connected to the base of transistors **60** and **62**, the emitters of said transistors each being connected to ground. The collectors of transistors **60** and **62** are connected to the B+ supply through collector resistors **64** and **66**, respectively. The frequency modulation control terminal **38** is electrically connected to the base of transistors **60** and **62** through resistors **68** and **70**, respectively. Capacitor **72** is connected from the base of transistor **60** to the collector of transistor **62**, while capacitor **74** is electrically connected from the base of transistor **62** to the collector of transistor **60**. The output from the oscillator circuit is taken from the collector of transistor **62**.

In operation, the frequency of the oscillator is determined by the time constant of resistors **68** and **70**, capacitors **72** and **74** and the voltage at terminal **38**. it should be noted that in the usual configuration of a multivibrator circuit, terminal **38** is connected to the B+ supply, thereby allowing the oscillator to run at a constant frequency. In the embodiment shown in FIGURE 6, terminal **38** is connected to a variable voltage, the waveform of said voltage being shown in FIGURE 5, waveform A for a voiced sound. During the period that the glottis is closed (which for purposes of illustration is assumed to be 30% of the total pitch period), the voltage level of waveform A is greater than the period when the glottis is open. This greater voltage level causes the output waveform to decrease in frequency from the center frequency of the respective analyzer channel corresponding to the period when the glottis is closed. When the lower voltage level (corresponding to the open glottis condition shown in waveform A of FIGURE 5) is imposed at terminal **38** of oscillator **36**, the frequency of the oscillator exceeds the center frequency of the corresponding analyzer channel. In this manner, frequency modulation of the oscillator is obtained below and above the center frequency of the corresponding analyzer channel.

While waveform A of FIGURE 5 is used to frequency modulate the oscillator, waveform B of FIGURE 5 in phase with waveform A is used for resetting the oscillator. Oscillator **36** must be reset so that all channel oscillators start out in phase with the occurrence of each pitch pulse during a voiced sound. Not only are the oscillators started in phase with the occurrence of the pitch pulse, but each oscillator must start oscillating at its normal free-running frequency with the first cycle after being reset. This is accomplished by setting the proper voltage levels on terminal **40** of each of the two frequency determining capacitors **72** and **74** at the initiation of the resetting event. Accordingly, a frequency modulated pitch-synchronous output (waveform C of FIGURE 5) is obtained from oscillator **36**.

FIGURE 7 is a partial schematic diagram of the exponential waveform generator **42**. This generator comprises switch **76** (which may be a transistor switch, for example), said switch having two input terminals **40** and **44**. The output of switch **76** is connected to capacitor **78** with the other side of capacitor **78** being grounded. In parallel with said capacitor are resistors **80** and **82**, each being connected to ground with the latter resistor **82** being connected through switch **84** to ground. The pulse applied to terminal **38** of switch **84** determines the length of time that said switch is open.

In operation, the exponential waveform generator **42** works in time synchronism with the resetting of oscillator **36**. That is, at the same time that the 1% pitch pulse (waveform B) resets oscillator **36**, this same pulse applied to terminal **40** closes switch **76** and allows capacitor **78** to charge during that 1% pitch pulse period to a value equal to the voltage at terminal **44**, that being the DC control level from the respective analyzer channel. Switch **84** is opened at the same time that switch **76** closes and remains open for 30% of the pitch period (waveform A of FIGURE 5), thus corresponding to the length of time that the glottis is closed. With capaci-

tor **78** charged to the value of the DC level of the respective analyzer channel, the capacitor then discharges at an exponential rate establishing the desired decay rate for the closed glottis condition. The decay rate during the closed glottis condition is determined by capacitor **78** discharging into resistor **80**. At the termination of the 30% pitch pulse period (which corresponds to the opening of the glottis), switch **84** closes such that the decay rate of capacitor **78** is determined by the parallel combination of resistors **80** and **82**, resulting in a faster exponential decay rate. Accordingly waveform D of FIGURE 5 is obtained at the output of exponential waveform generator **42**; this waveform has an initial amplitude determined by the DC channel control level of the analyzer channel, then decays exponentially at a rate determined by capacitor **78** discharging into resistor **80**, and then when the portion of waveform A representative of the open glottis condition occurs, switch **84** closes and the decay rate is increased as capacitor **78** is discharging into the parallel combination of resistors **80** and **82** creating a smaller time constant.

FIGURE 8 illustrates the modulator **43** of FIGURE 4 and comprises a transistor **90**, the emitter of which is connected to ground, the base terminal **92** is connected to the output of oscillator **36** and the collector is connected by terminal **94** to the output of exponential waveform generator **42** through collector resistor **96**. The output of the modulator is taken at the collector of transistor **90**.

In operation, the output of oscillator **36** applied to terminal **92** turns the transistor **90** on and off. In the on condition, the collector of transistor **90** is approximately at ground. In the off condition, the collector of transistor **90** is at the same level as the collector load input, that being the voltage of the exponential waveform generator output. The resulting waveform E (of FIGURE 5) illustrates the oscillator output modulated by the waveform generated by the exponential waveform generator **42**. As was described above, this modulator output is then filtered and summed with the output of the other channels to form the synthesized speech sound.

While the invention has been described with reference to a specific embodiment, it is to be understood that this description is not to be construed in a limiting sense. Various modifications of the disclosed embodiments, as well as other embodiments of the invention, will become apparent to persons skilled in the art without departing from the spirit and scope of the appended claims.

What is claimed is:

1. An apparatus for synthesizing the sound produced by the changing resonant conditions in the cavities of the human vocal tract during a pitch period defined by the closing, opening and re-closing of the glottis, comprising:

(a) free-running oscillator means having an output signal,

(b) means for frequency modulating said signal at a first frequency during the closed glottis condition and a second frequency during the open glottis condition,

(c) waveform generating means producing a first exponential decay during the closed glottis condition and a second exponential decay during the open glottis condition, and

(d) modulating means which modulates the frequency modulated oscillator output signal with the output from said waveform generating means.

2. An apparatus according to claim **1** including means connected to said waveform generating means for setting the initial amplitude of said first exponential decay.

3. An apparatus according to claim **1** including a low-pass filtering means connected to the output of said modulating means.

4. An apparatus according to claim **1** including resetting means connected to said oscillator means and said waveform generating means to reset each of said means at the initiation of each pitch period.

5. An apparatus according to claim **4** including means for randomly resetting said oscillator and waveform generating means.

6. An apparatus according to claim **1** further comprising:

(e) at least one additional free-running oscillator means having an output signal,

(f) at least one additional means for frequency modulating said output signal at a third frequency during the closed glottis condition and a fourth frequency during the open glottis condition,

(g) at least one additional waveform generating means producing a third exponential decay during the closed glottis condition and a fourth exponential decay during the open glottis condition,

(h) at least one additional modulating means which modulates the at least one additional modulated oscillator output signal with the output from said at least one additional waveform generating means, and

(i) means connected to each of the modulating means to sum their outputs.

7. An apparatus according to claim **6** including low-pass filtering means connected between the output of the modulating means and the summing means.

8. An apparatus for synthesizing the sound produced by the changing resonant conditions in the cavities of the human vocal track during a pitch period defined by the closing, opening and re-closing of the glottis, comprising:

(a) a free-running multivibrator having an output signal and including

(1) means for resetting said modulator at the initiation of each pitch period, and

(2) means for frequency modulating said signal at a first frequency during the closed glottis condition and a second frequency during the open glottis condition,

(b) an exponential waveform generator for producing a first exponential decay during the closed glottis condition and a second exponential decay during the open glottis condition, said generator including

(1) a variable resistive and capacitive network for producing variable decay rates and

(2) switching means interconnected to said network for changing the network time constant at substantially the same time that the glottis condition is changed,

(c) resetting means connected to said generator to reset said generator at the initiation of each pitch period, and

(d) means for modulating the frequency modulated output signal with the output from said exponential waveform generator.

References Cited

UNITED STATES PATENTS

3,330,910	7/1967	Flanagan	179—15.55
3,268,660	8/1966	Flanagan.	
3,083,266	3/1963	Mathews et al.	
2,817,707	12/1957	Wiebel.	

KATHLEEN H. CLAFFY, Primary Examiner

ARTHUR A. McGILL, Assistant Examiner