

[54] **SPEECH SYNTHESIZER RESPONSIVE TO A DIGITAL COMMAND INPUT**

[75] Inventor: **Richard T. Gagnon**, Birmingham, Mich.

[73] Assignee: **Scitronix Corporation**, Birmingham, Mich.

[22] Filed: **July 21, 1972**

[21] Appl. No.: **274,029**

Related U.S. Application Data

[63] Continuation-in-part of Ser. No. 119,473, March 1, 1971, abandoned.

[52] U.S. Cl. **179/1 SA**
[51] Int. Cl. **G101 1/10**
[58] Field of Search **179/1 SA, 15.55 R, 15.55 T**

References Cited

UNITED STATES PATENTS

3,102,165	8/1963	Clapper	179/1 SA
3,158,685	11/1964	Gerstman	179/1 SA
3,268,660	8/1966	Flanagan	179/1 SA
3,319,002	5/1967	DeClerk	179/1 SA
3,328,525	6/1967	Kelly	179/1 SA
3,573,374	4/1971	Focht	179/1 SA
3,624,302	11/1971	Atal	179/1 SA

OTHER PUBLICATIONS

Rabiner, Digital Formant Synthesizer for Speech Syn-

thesis Studies, JASA, 1967, p. 822-828.

Flanagan, Synthetic Voices for Computers, IEEE Spectrum, 10/70, pp. 22-45.

Rabiner, A Model for Synthesizing Speech by Rule, IEEE Trans. on Audio 3/69, p. 7-13.

Primary Examiner—Kathleen H. Claffy

Assistant Examiner—Jon Bradford Leaheey

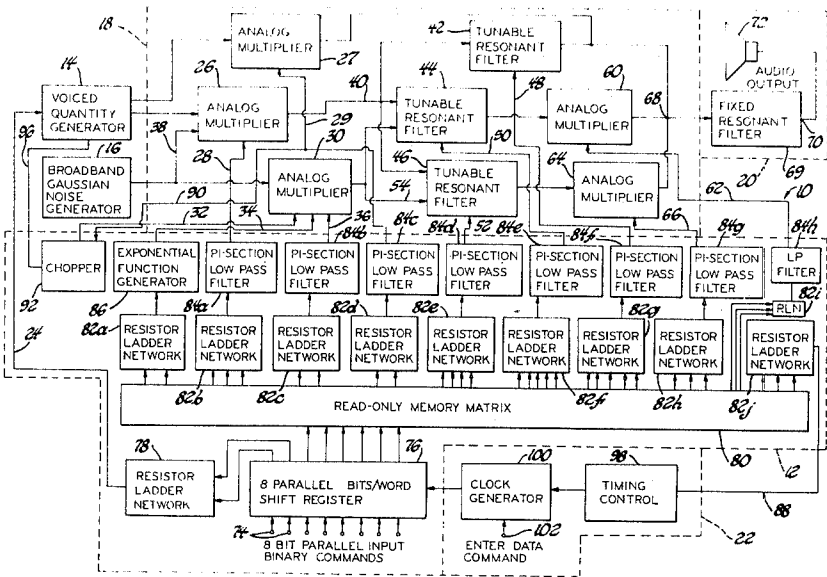
Attorney, Agent, or Firm—Fisher, Krass, Young & Gerhardt

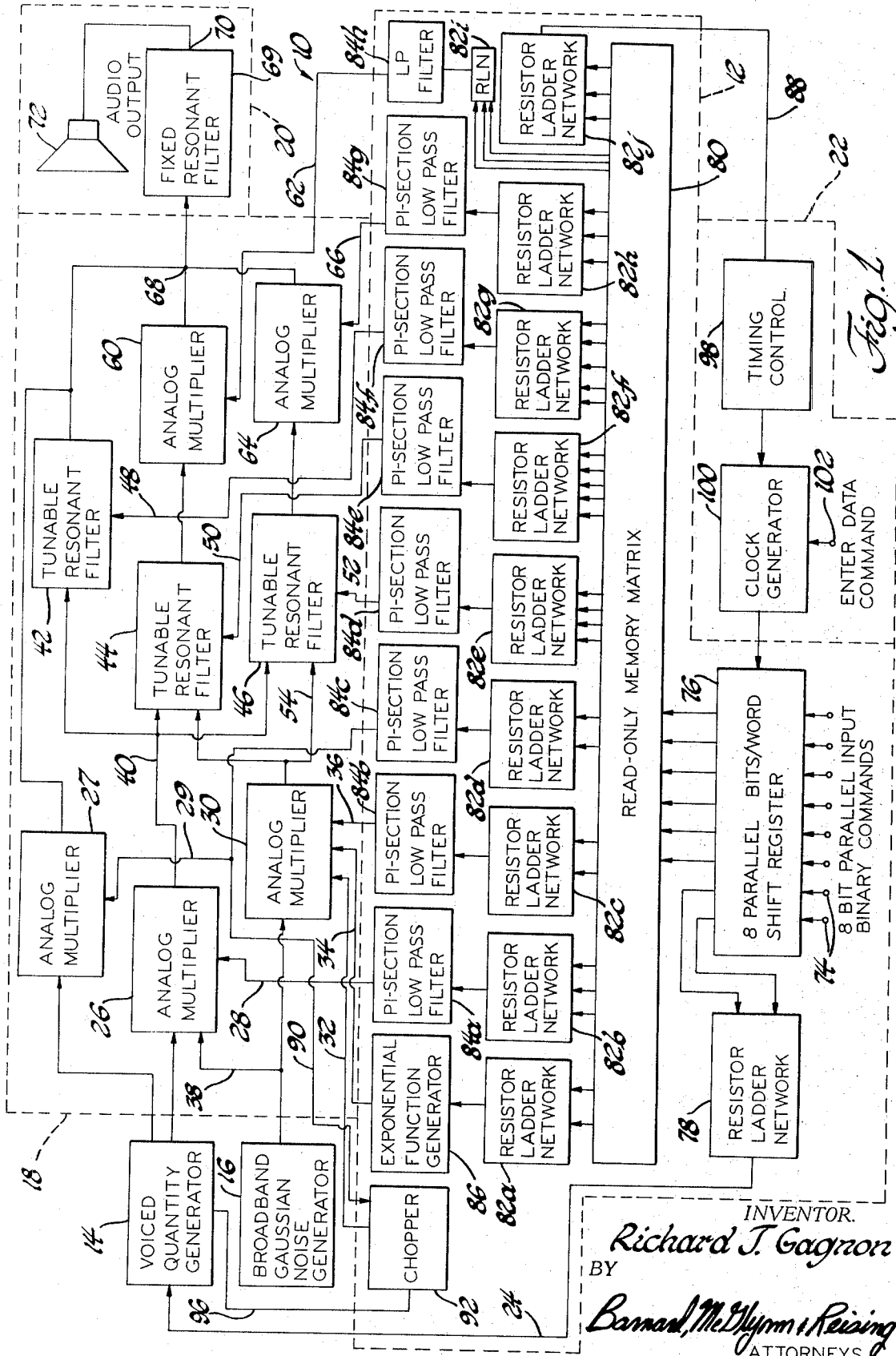
[57]

ABSTRACT

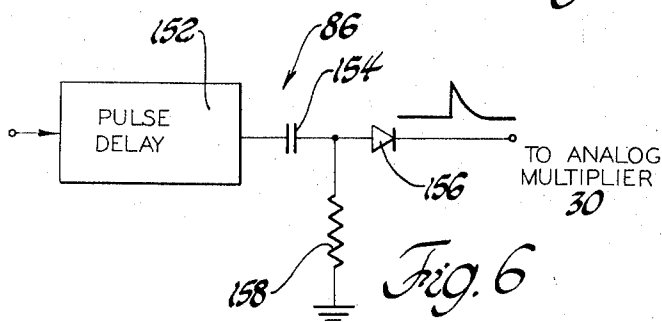
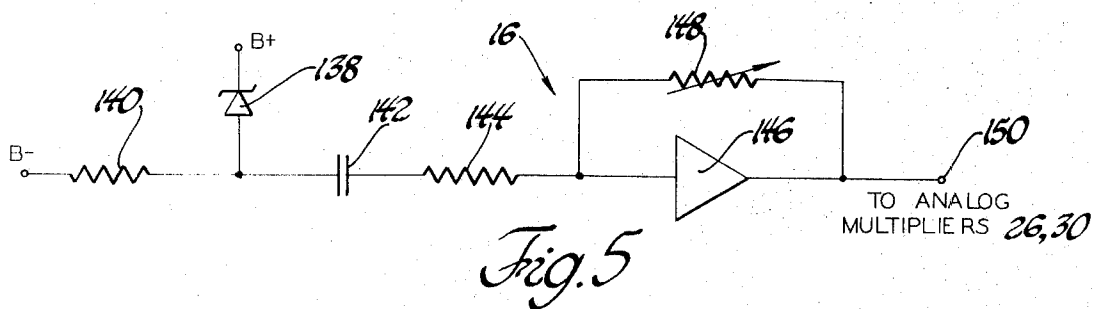
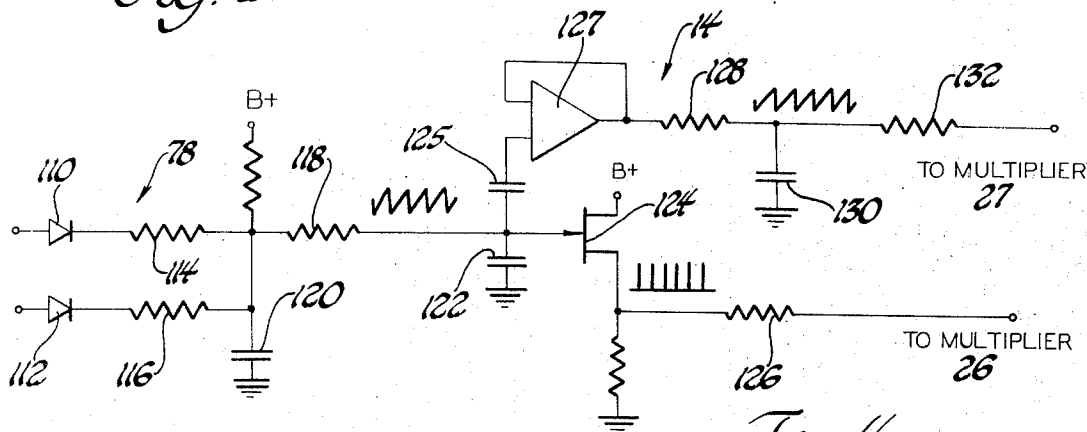
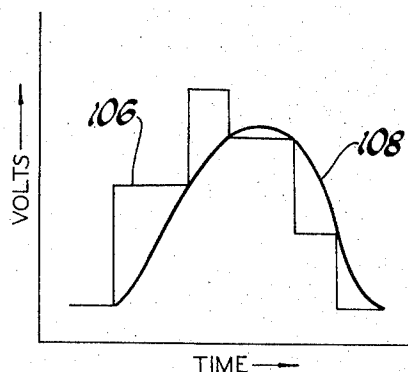
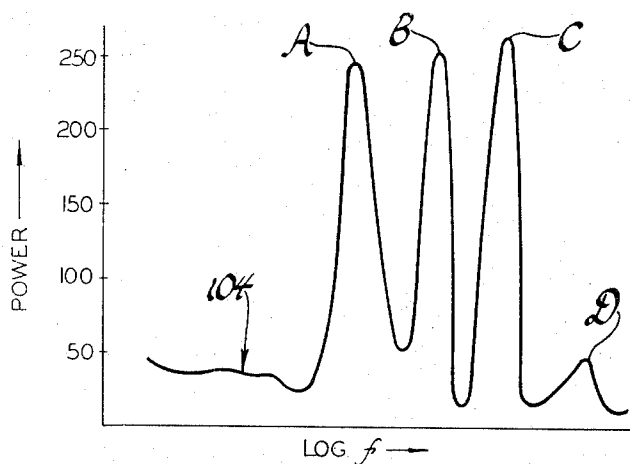
A voice synthesizer including a voiced quantity generator, an unvoiced quantity generator, a control circuit including amplitude and resonance control devices, and an input system for receiving eight-bit digital signals and for producing step function analog control signals to the various phoneme forming control devices. Slow acting filters are connected between the step function generators and the control devices to produce smooth transitions between the analog levels, thus, to realistically simulate the smoothly-changing dynamic quality of human speech. Circuit details are disclosed.

35 Claims, 12 Drawing Figures

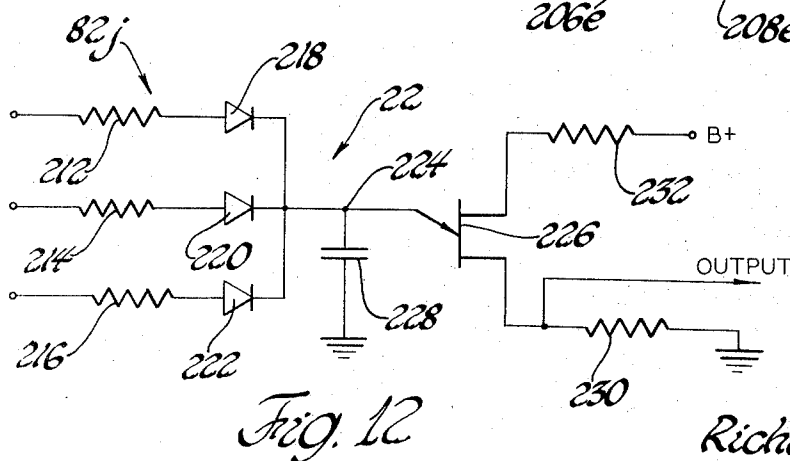
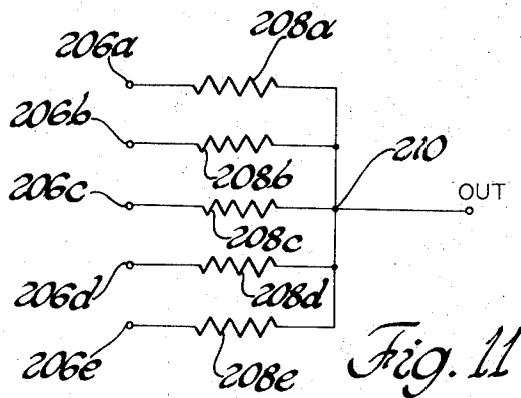
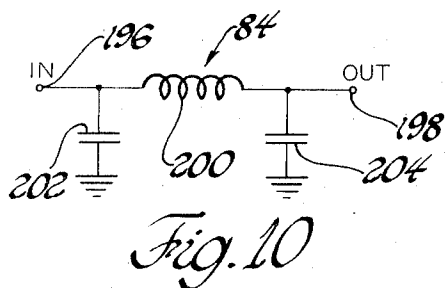
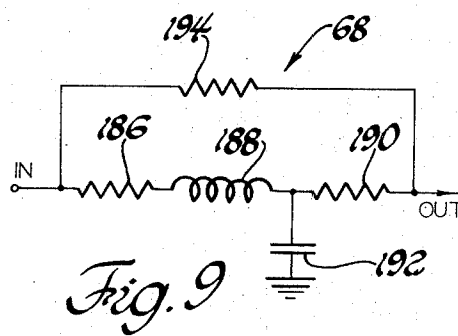
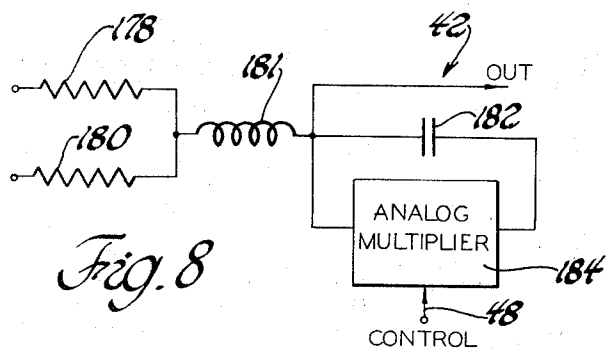
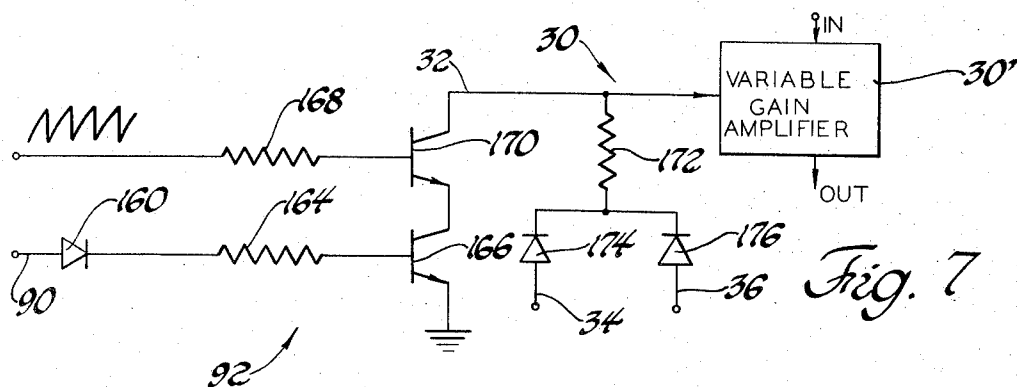




INVENTOR.
Richard J. Gagnon
 BY
Barnard, McElroy & Reising
 ATTORNEYS



INVENTOR.
Richard J. Gagnon
 BY
Barnard, McElvyn & Reising
 ATTORNEYS



INVENTOR.
 Richard T. Gagnon
 BY
 Bernard, McEllynn & Rising
 AT TORNEYS

SPEECH SYNTHESIZER RESPONSIVE TO A DIGITAL COMMAND INPUT

This is a Continuation-In-Part application of U.S. Ser. No. 119,473, now abandoned filed Mar. 1, 1971, entitled "Voice Synthesizer."

INTRODUCTION

This invention relates to voice and speech synthesis apparatus and particularly to speech synthesis apparatus employing phoneme generation devices.

BACKGROUND OF THE INVENTION

The prior art contains numerous examples of apparatus for producing an output which resembles human speech. The prior art would appear to divide into two categories. The first category includes devices for recording and storing words and phrases in such a fashion as to be retrievable to construct sentences and paragraphs. These devices may be thought of as encyclopedias rather than synthesizers and are of generally limited applicability due to the storage and retrieval requirements of a large vocabulary. The other category of such devices include those which synthesize speech elements, usually on a phonetic basis, such phonetic elements being assembled into words in response to input command signals. Since there are far fewer phonemes than words in practically any language, the phoneme generation approach is preferable from the standpoint of minimizing data storage and retrieval requirements while maintaining a widely variable speech capability. The phoneme generators which are shown in the prior art are nevertheless generally quite complex, apparently due to the prevailing notion that speech is best synthesized by a full electronic analogy of the human vocal tract. This notion results in a synthesizer having a large number of controlled elements, this in turn placing a complexity requirement on the programming means for controlling the elements in response to the input commands.

GENERAL STATEMENT OF INVENTION

The present invention is a speech synthesizer of the phoneme generating type and has for its primary objective the vast reduction of complexity in speech synthesizers along with an improvement in the realism and understandability of the synthesized speech. In accordance with the present invention, human speech is synthesized in a system wherein the input addressing or selection function is preferably accomplished digitally while actual phoneme formation is accomplished in an analog fashion using analog control devices. Moreover, the control signals which are generated and applied to the control devices occur in coded combinations thereby to control analog amplitude, resonance, and timing devices to form acoustic effects which realistically simulate the acoustic effects which define human speech.

In accordance with the invention, a realistic simulation of human speech is accomplished by producing actual speech effects through simple electronic devices. According to one feature of the preferred embodiment hereinafter described, such realistic speech effects are accomplished by generating basic voiced and unvoiced phoneme qualities and controlling the modulation and resonance of combinations of these quantities using variable amplitude analog control functions of a smooth and relatively slowly changing waveform. This

waveform is generated by passing a step function representing an analog version of a digital command through a simple network having a finite response time. Thus, the analog control signals exhibit smooth rather than abrupt transitions between levels introducing similar smooth and nonabrupt transitions between successive phonemes. Thus, the dynamic sluggishness of the human vocal tract is simulated as far as the effect on output speech is concerned.

According to another feature of one embodiment of the invention as hereinafter described, speech quality is improved by synthesizing the unvoiced component relating to air escape or breathing effect of unvoiced phonemes. This is accomplished by combining a portion of the unmodulated unvoiced phoneme quantity with the inflection controlled voiced phoneme quantity such that all succeeding amplitude and resonant control of the voiced quantity inherently affects the unvoiced component as well.

According to still another feature of the illustrative embodiment to the invention hereinafter described, the inflection amplitude, and waveform modulated phoneme quantities are applied in combinations to a plurality of tuned resonant filters preferably of the single pole type, each filter being controllably turned to a frequency pole which is found in the frequency-power spectrum of an annunciated phoneme. These filters are controlled by separate analog control signals generated in response to digital phoneme commands. Such filters, as hereinafter specified, may be constructed using a voltage controlled capacitor circuit, this capacitor circuit being responsive to the analog control signals to vary the tuning of the resonant filter circuits.

According to still another feature of one embodiment of the invention, the resonant outputs of the tunable resonant filters are directed through additional individual control devices, such as variable gain amplifiers for amplitude control and, thence, through a fixed resonant single pole filter to simulate a nasal resonance quality. This final output is applicable to speaker apparatus and the like for reproducing speech of a high quality suitable for direct or indirect transmission to a human listener.

According to a feature of another embodiment of the invention, means are provided for the simulation of phoneme interactions; i.e., the timing effects which appear as a result of two or more consecutive phonemes which tend to produce modifications in the normal phoneme pronunciation and timing. One such interaction is involved between voiced phonemes, such as *e* and *a*, and phonemes involving nasal sounds such as *n* and *m*. This interaction is hereinafter called a "nasal closure." Another such interaction occurs between a voiced phoneme and an unvoiced fricative such as *t*, *k*, or *p*, and such interaction is hereinafter termed a "fricative stop." A third such interaction occurs between pure voiced phonemes and voiced fricatives such as *b*, and such interaction is hereinafter called a "vocal closure." Nasal closures, fricative stops, and vocal closures are all simulated by the generation of signals which result in a relatively abrupt amplitude reduction, followed by a relatively abrupt increase, in a phoneme interaction amplitude curve which would otherwise show a more smoothly occurring amplitude change. Various implementations for accomplishing this simulation are disclosed hereinafter with reference to the

second of two specific embodiments of the invention.

Still other features of the invention are described hereinafter including the alternate inversion of several resonant frequency components in any given phoneme to give increased resonant frequency definition, alternate means for accomplishing voiced fricative modulation, and many other features. The many useful application of the subject device along with the aforementioned and other features and advantages of the invention will become more apparent from a reading of the following specification which sets forth specific embodiments of the invention in detail. This specification is to be taken with the accompanying drawing.

BRIEF DESCRIPTION OF THE DRAWING

FIG. 1 is a block diagram of a first embodiment of the synthesizer system;

FIG. 2 is a power frequency spectrum of an exemplary phoneme;

FIG. 3 is an output signal waveform indicating the smooth transitional qualities of the analog control signal;

FIG. 4 is a schematic circuit diagram of a portion of the system of FIG. 1,

FIG. 5 is another schematic circuit diagram of another portion of the system of FIG. 1;

FIG. 6 is another schematic circuit diagram of another portion of the system of FIG. 1;

FIG. 7 is another schematic circuit diagram of another portion of the system of FIG. 1;

FIG. 8 is another schematic circuit diagram of another portion of the system of FIG. 1;

FIG. 9 is another schematic circuit diagram of another portion of the system of FIG. 1;

FIG. 10 is another schematic circuit diagram of another portion of the system of FIG. 1;

FIG. 11 is another schematic circuit diagram of another portion of the system of FIG. 1;

FIG. 12 is another schematic circuit diagram of another portion of the system of FIG. 1;

FIG. 13 is a block diagram of an alternative embodiment of the invention with portions in schematic detail;

FIG. 14 is a waveform diagram illustrating the timing of certain signals developed in the circuit of FIG. 13;

FIG. 15 is a schematic circuit diagram of a modification or optical feature for the circuit of FIG. 13; and,

FIG. 16 is a waveform diagram illustrating the operation of the circuit of FIG. 15.

BLOCK DIAGRAM DESCRIPTION OF FIRST EMBODIMENT

Referring to FIG. 1 there is shown a block diagram of a speech synthesizer system 10 which embodies the present invention. System 10 comprises an input section 12 which is responsive to eight-bit digital commands to generate analog control signals in 37 unique combinations, each representing the amplitude, resonance, and timing coordinates of the acoustic speech elements, hereinafter usually called phonemes, to be produced by the system 10. System 10 further includes a first generator 14 for producing a basic voiced phoneme quantity and a second generator 16 for producing a basic unvoiced phoneme quantity. The basic voiced

and basic unvoiced phoneme quantities from generators 14 and 16 are modulated and resonated in a control section 18 which operates under the control of the analog control signals which are supplied by the input section 12. The single output line of the control section 18 is connected to an output section 20 for suitable amplification and reproduction purposes. The phoneme generation rate and timing characteristics are established by a timing section 22 which receives control signals from the input section 12 as shown.

Describing the circuit of FIG. 1 in greater detail, the generator 14, which will be described in greater detail with reference to FIG. 4, produces periodic output signals at a basic frequency which approximates that of the typical human voice. Generator 14 produces both a rounded sawtooth waveform and a harmonic-rich impulse waveform, the two waveforms occurring at the same frequency. Frequency control for inflection and pitch control purposes may be accomplished by way of a control line 24 which carries an analog control signal to be described from the input section 12. The basic impulse-type periodic waveform representing the voiced quantity from generator 14 is connected to the input of an analog multiplier 26 which may be implemented as a variable gain amplifier. The gain of amplifier-multiplier 26 is set by a control signal appearing on control line 28. The sawtooth waveform is connected to the input of analog multiplier 27 which operates under the control of signals on line 29.

In a similar fashion the second generator 16 which comprises a broadband Gaussian noise generator described in greater detail with reference to FIG. 5 has the output thereof connected to an analog multiplier 30 which may also be implemented using a variable gain amplifier. The amplifier 30 is controlled in gain for consequent amplitude control of the signal from generator 16 by way of three separate control signals which appear on control lines 32, 34, and 36, respectively. The unvoiced phoneme quantity from the generator 16 is also applied by way of line 38 to the input of the analog multiplier 26 where it is combined with or superimposed on the impulse waveform of the basic voiced phoneme quantity which is produced by generator 14 prior to any modulation in the analog multiplier 26. Thus, each voiced phoneme which is produced by the system 10 carries a broadband Gaussian noise component to simulate the breath escape or unvoiced noise component which is inherently present in all voiced phoneme quantities of human speech.

The output of analog multiplier 26 appears on line 40 and is connected commonly to the inputs of single pole, tunable resonant filters 42, 44, and 46. The filters 42, 44, and 46 are each tunable by means of analog control signals applied thereto by way of control lines 48, 50, and 52, respectively, to resonate the applied waveforms at selected frequencies. In contrast with the possibly more well-known bandpass filter, the single pole tunable resonant filter used herein exhibits a significant transmissivity over nearly the entire range of signals applied thereto but further exhibits a resonance or high transmissivity at a given frequency value. For purposes to be hereinafter described, the filter 42 is tunable to resonance across a first low-range of frequencies extending from approximately 100 Hz to approximately 1,000 Hz. The second filter 44 is tunable to resonance across an intermediate band of frequencies extending from approximately 500 Hz to 3,000 Hz and third filter

46 is controllably tunable to resonance across a high band extending from approximately 1,000 Hz to 4,000 Hz. As shown in FIG. 1, the output of the analog multiplier 26 which carries the voiced phoneme quantity is applied to all three of the filters 42, 44, and 46 since voiced phoneme constituents tend to appear in all of the frequency ranges common to human speech. The output of analog multiplier 30 appearing on line 54, however, is applied only to the inputs of filters 44 and 46 since the unvoiced phoneme quantity tends to appear only in the intermediate and higher frequency bands of human speech.

Continuing with the description of the control section 18, the output of filter 42 is connected directly to an output terminal 68. The output of filter 44, however, is connected to an analog multiplier 60 for amplitude control of the compound output thereof in accordance with the magnitude of the signal on control line 62. Finally, the output of filter 46 is connected to an analog multiplier 64 for amplitude modulation of the compound signal according to the magnitude of the signal on control line 66. Thus, up to three resonant poles, each positionable in frequency and regulable in amplitude, may be generated by the control section 18 of system 10 to produce phonemes having both voices, unvoiced, and compound qualities. The outputs from the analog multipliers 60 and 64 are connected to the output terminal 68 which also represents the input to the output section 20.

The output section 20 comprises a single pole fixed resonance filter 69 which synthesizes the nasal resonance of the human vocal tract. The frequency pole of filter 69 occurs at approximately 4,000 Hz and may be fixed in frequency position. Accordingly, no control line from input section 12 need be provided. The output of the filter 69 appears at terminal 70 for application to an audio output stage such stage normally including a broad band amplifier and speaker 72.

Looking now to the input section 12, eight-bit parallel input binary commands are applied to the input terminals 74 of a shift register 76 which operates under the control of the timing section 22. The eight-bit input commands are readily generated by means of a computer or other business machine, such as one having a phonetic keyboard and a diode encoder matrix. Eight-bits are selected to provide sufficient information capacity for 60 or more phoneme commands although in normal practice only 37 elemental acoustic effects need be employed to generate a high-quality human speech. Two of the bits which are transferred in parallel from the shift register 76 at the time a phoneme is to be generated are applied to a resistor ladder network 78 having two or more preferably binary weighted resistors for the generation of a variable amplitude analog step function which controls the inflection or basic frequency output of the voiced phoneme quantity generator 14. The output of resistor ladder network 78 appears on the control line 24 which is connected to the generator 14 as previously described. Inflection control need not be elaborate, four basic levels of frequency being adequate under normal circumstances. The other six bits of each eight-bit word shifted out of shift register 76 in parallel form are applied to a read-only memory matrix 80 which produces unique combinations of signals on the 32-output lines emanating therefrom. Matrix 80 may be a diode matrix having connector pins at fixed locations to convert each six-bit input signal

into a unique output signal combination having up to 32 individual bits. Alternatively, a magnetic core matrix may be employed, the wiring of such a matrix being fixed so that no information writing function can be performed but exhibiting a fixed transformation response to preselected input signals, i.e., six bit input combinations produce 32 bit output combinations. This approach is believed preferable to an alterable memory; however, an alterable memory may be employed where computer capacity is available. Where such capacity is not available, the matrix 80 may be fabricated in a plug-in fashion so that it may be removed and exchanged for another matrix of slightly different makeup, thus, to enable the system 10 to represent different languages, genders, accents, and so forth.

The 32 digital signal output lines of matrix 80 are connected in various combinations to a plurality of resistor ladder networks 82a through 82j which convert the digital (binary) input signals into analog step functions of corresponding value. The number of digital inputs to each of the resistor ladder networks 82a through 82j indicates the exponential relation to the number of analog amplitude variations available in the output waveform. A suitable resistor ladder network is illustrated in FIG. 11.

Proceeding with the description of the input section 12 of system 10, the analog step function outputs of resistor ladder networks 82b through 82i are connected to individual lowpass filters 84a through 84h, respectively, each filter having a finite response time on the order of 70 milliseconds thereby to smooth out the abrupt amplitude variations between the step function levels and present relatively smooth transitions and slowly varying amplitude levels to the various devices which are controlled by the analog control signals. The output of filter 84a appears on line 28 and is applied to the analog multiplier 26 for amplitude control of the voiced phoneme quantity as mixed with a portion of the unvoiced phoneme quantity. The output of filter 84b appears on line 36 and is applied as one of the three control signals to the analog multiplier 30 for amplitude control of the unvoiced phoneme quantity. The output of filter 84c appears on line 29 and is a slowly varying analog function (voltage) applied to analog multiplier 27 for amplitude control of the voiced phoneme quantity applied to terminal 68. The output of filter 84d appears on line 52 and is applied to the third tunable resonant filter 46 to establish the location of the frequency pole in the high-frequency range. The output of filter 84e is applied to the intermediate range tunable resonant filter 44 by way of line 50 to determine the location of the frequency pole in the intermediate range. The output of filter 84f appears on line 48 and is applied to the control terminal of tunable resonant filter 42 to determine the frequency pole of the voiced phoneme quantity in the low-frequency range. The output of filter 84g is applied to the control input of analog multiplier 64 by way of control line 66 to determine the amplitude modulation of the compound signal from filter 46. The output of filter 84h is applied by way of control line 62 to the analog multiplier 60 for controlling the amplitude of the compound signal in the intermediate resonant range.

The output of resistor ladder network 82a is applied not to a low-pass filter but to an exponential function generator 86 the details of which are illustrated in FIG.

6 for the generation of plosive phonemes. The output of exponential function generator 86 appears on control line 34 and is applied as the second of the three inputs to the analog multiplier 30 for the amplitude modulation of unvoiced phoneme quantities. The output of resistor ladder network 82j is applied directly to the timing section 22 by way of control line 88.

A final control function of the read only memory matrix 80 is the control of a chopper circuit 92 which provides the third input to multiplier 30. The control signal for chopper 92 appears on line 90 is taken from the output of filter 84c and is, thus, present any time a voiced phoneme is generated. The signal turns on the chopper circuit 92 to modulate the unvoiced phoneme quantity applied to the analog multiplier 30. The modulation of control signal is applied to the multiplier 30 by way of line 32. The amplitude modulation is effectively a square waveform variation between full amplitude and zero amplitude and occurs at a rate which is low relative to the center frequency of the unvoiced signal component. The frequency control on chopper 92 is derived from the generator 14 which provides the voiced phoneme quantity by way of line 96. The result is that the unvoiced component of compound phonemes is modulated at the fundamental voiced pitch frequency. The chopper 92 is, thus, activated during the phoneme interval which periodically drops the output of the analog multiplier 30 to zero during each voiced cycle. Compound phonemes for which chopper 92 is effective include *th* as in "then," *s* as in "leisure," *v*, *j*, and *z*. Phonemes having only unvoiced components and, thus, those in which the chopper 92 is not actuated are *s*, *ch*, *sh*, *f*, *th* as in "thin," and *h*. The chopper 92 is, of course, on during pure voiced phonemes, but to no effect since the control from filter 84b overrides the others and maintains the multiplier 30 cut off.

Looking now to the timing section 22, the control line 88 from resistor ladder network 82j is connected to the timing control circuit 98 which in turn is connected to the clock generator 100. The clock generator 100 is connected to the shift input of the shift register 76 to control the intervals between the transfer of eight-bit words from the register 76 to the read-only memory 80. The clock intervals for the various phonemes vary between thirty and 150 milliseconds, the interval or duration of each phoneme being established by the input word which is transferred from the register 76 to the memory 80. In other words, each phoneme times itself through the timing control circuit 98. The timing control circuit may be any of a number of well known devices for effecting the timed generation of signal pulses, a preferred form being hereinafter described. An "enter data" command terminal 102 is provided on the clock generator 100 for enabling the clock while data is being entered into the shift register 76.

OPERATION OF FIRST EMBODIMENT

Reviewing now the operation of the system 10 of FIG. 1, the phonemes required to synthesize a word or group of words are loaded into the shift register 76 in eight-bit words, and shifted toward the eight-bit parallel transfer position in the register 76 by means of the clock generator 100. The phonetic construction of words is common knowledge and will not be repeated in its entirety herein. The eight bits which define each phoneme include two inflection bits which are trans-

ferred to the resistor ladder at work 78 in digital form. The ladder network 78 converts the digital signals to an analog step function which is applied by way of line 24 to the generator 14. Accordingly, the generator 14 is actuated at all times to supply the voiced phoneme quantity to analog multiplier 36 even though the quantity is only actually employed to produce acoustic effects having voiced components. The other six bits determine the digital command signals which are to be applied to the resistor ladder networks 82a through 82j and, thus, the analog step function outputs which are applied to the exponential function generator 86, the low-pass filters 84a through 84j, and the timing control circuit 98. The unvoiced phonemes *s*, *sh*, *f*, *th* as in "thin", and *h* are indicated by an absence of signals from either chopper 92 or exponential function generator 86. In general, the aforesaid phonemes are controlled by the signals appearing on line 36 as applied to analog multiplier 30. In addition, the tunable resonant filters 42, 44, and 46 are actuated with suitable signals during unvoiced phonemes to simulate the resonant quality of the vocal tract during the formation of such phonemes. The aforementioned phonemes *th* as in "then", *s* as in *leisure*, *v*, *j*, and *z* all include both voiced and unvoiced phonemes, the unvoiced phonemes being essentially a noise component which in the human vocal tract is formed by air turbulently passing through a constriction in the vocal tract. This, of course, occurs simultaneously with a pressure wave from the vocal cords, this being simulated in the system 10 by suitable modulation of the basic voiced quantity from generator 14. For all of these phonemes the chopper 92 is actuated by means of a suitable signal in the form of a voltage on line 90. The signal is applied by way of line 32 to the analog multiplier 30 to amplitude modulate the unvoiced component at the voiced component frequency.

For plosive phonemes the exponential function generator 86 is activated by one of four analog signal levels from the resistor ladder network 82a. These plosive phonemes include voiced phonemes such as *b*, *d*, and *g* as well as unvoiced phonemes *k*, *ch*, *p*, and *t*. All of these phonemes when produced by the human vocal tract involve the build up of a pressure behind a constriction in the vocal tract followed immediately by a release of pressure to produce an exponentially decaying Gaussian noise function. This function is carried out in the system 10 by means of the exponential function generator 86 which initially delays the transmission of the phoneme through the analog multiplier 30 for the period during which a capacitor charge is built up and then immediately discharges the capacitor through a resistor to produce a decaying analog amplitude modulation by way of multiplier 30. This modulation, of course, is worked upon the unvoiced Gaussian noise component produced by generator 16.

Reference to FIG. 2 shows a frequency power spectrum analysis for the phoneme *u* and indicates the operation of the three tunable resonant filters 42, 44, and 46 during phoneme formation. It can be seen that the waveform 104 representing the frequency power spectrum of the enunciated phoneme as a first pole *A* or resonant peak at about 650 Hz this pole being established by and within the range of the resonant filter 42. A second pole *B* occurs at approximately 1,120 Hz and is established by the resonant filter 44. A third pole *C* occurs at approximately 2,100 Hz and is established by

the resonant filter 46. A fourth much lower power resonance D is indicated at approximately 4,000 Hz and represents the resonant contribution of the single pole fixed resonant filter 69, that is, the nasal resonant synthesis. It is to be understood that each phoneme exhibits a frequency-power spectrum which can be displayed in graphic form such as the phoneme power-frequency spectrum 104 of FIG. 2. Not all phonemes, of course, exhibit the three major poles of the waveform 104 but rather each phoneme exhibits a variation of pole frequency location and relative amplitude thereby making each phoneme unique in the specification of signals to be applied by way of control lines 48, 50, and 52 to the filters 42, 44, and 46, respectively. The tunable bands for the filters 42, 44, and 46 necessarily overlap inasmuch as some phonemes exhibit poles of which two may occur within the range of one filter.

Referring now to FIG. 3, there is shown a typical analog step function waveform 106 of the type which may be produced by the resistor ladder networks 82a through 82j of FIG. 1. The superimposed smooth waveform 108 represents the smoothed and delayed version of the analog step function which results from passing the step function 106 through the low-pass filters 84. It will be noted that since the response time of the filters 84 is on the order of 70 milliseconds, whereas some phonemes have a duration of only 30 milliseconds, there are phonemes for which the exactly prescribed response is never fully achieved in the system 10 of FIG. 1. Again, this tends to increase and enhance the realistic quality of the operation of system 10 by smoothly blending successive phonemes into one another in the same fashion that the human vocal tract normally operates. It will, thus, be perceived by those skilled in the art that the phoneme intervals should be somewhat centered around the filter response times. Accordingly, if the speech rate is varied, say by proportional shortening or lengthening of all phoneme intervals, a corresponding variation in filters 84 may be required to preserve intelligibility.

SCHEMATIC CIRCUIT DIAGRAMS

Referring now to FIG. 4, schematic details of an illustrative resistor ladder network 78 and a voiced phoneme generator 14 are shown. The resistor ladder 78 comprises diode 110 and 112 which direct current from the read-only memory matrix 80 through summing resistors 114 and 116 to an input resistor 118. A capacitor 120 smooths out current transitions. Current through resistor 118 charges a capacitor 122 until the threshold voltage of unijunction transistor 124 is reached. At this time the transistor 124 is conductive through the path from the dc supply voltage $B+$ through the resistor 126 to ground discharging capacitor 122 and causing an impulse voltage across resistor 126. This cycle repeats periodically and the impulse output is applied to multiplier 26. A saw-tooth voltage is developed across capacitor 122 which is sampled by an input path comprising capacitor 125 to the high impedance amplifier 127. The output of amplifier 127 is applied to a resistor 128 and a capacitor 130 which removes all higher frequency components. Output resistor 132 applies the somewhat rounded saw-tooth waveform shown to analog multiplier 27. The impulse component of the periodic waveform which is produced across resistor 126 in the circuit of FIG. 4 includes a broad spectrum of energy which can be resonated in

the tunable filters 42, 44, and 46 of FIG. 1. The low-frequency component produced by the amplifier 127 and the following circuitry passes through the analog multiplier 27 and is recombined with the impulse component at terminal 68 adding realism to the speech output. The current from the read-only matrix 80 applied by way of diodes 110 and 112 controls the frequency of the unijunction oscillator-transistor 124 and, thus, the pitch or inflection of the voiced phoneme quantity which is produced by the generator 14 illustrated in FIG. 4.

Looking now to FIG. 5, the broad band Gaussian noise generator 16 for the production of the unvoiced phoneme quantity is illustrated in schematic detail, again it being understood that the circuits of FIGS. 4 through 12 are merely illustrative of the preferred form of implementation. In FIG. 5 a semiconductor diode 138 is reversed biased beyond its breakdown voltage by positive and negative voltage supplies $B+$ and $B-$ connected through the diode 138 by way of resistor 140 which limits current flow. A very strong noise component results, this component being applied through the capacitor 142 and a resistor 144 to the input of operational amplifier 146 having a variable resistive feedback path 148. The output terminal 150 is, of course, connected to the input of the analog multipliers 26 and 30 shown in FIG. 1.

Looking now to FIG. 6 the exponential function generator 86 of the system 10 of FIG. 1 is shown in schematic form. The exponential function generator 86 comprises an input which is received from a resistor ladder network 82a and a pulse delay circuit 152 such as a standard one-shot multivibrator. The output of pulse delay circuit 152 is a digital pulse which is coupled through the capacitor 154 and the isolating diode 156. The RC exponential circuit is established by means of the resistor 158 which is connected between the capacitor 154 and ground. The output of the circuit 86 of FIG. 6 is connected to control the analog multiplier 30 modulating its Gaussian noise output with an exponentially decaying envelope. As previously mentioned, this signal is used in forming all plosive phonemes.

Looking now to FIG. 7, the details of the interconnection between the control devices which produce signals on lines 32, 34, and 36 and the analog multiplier 30 which is controlled thereby are shown. As previously mentioned, the analog multiplier 30 is mainly a variable gain amplifier 30' the input from the broad band Gaussian noise generator 16 being connected into the top thereof and the output to filters 44 and 46 being taken from the bottom thereof as seen in FIG. 7. The control signals on input lines 32 and 34 are abrupt and well defined in character whereas the control signal on line 36 coming from low-pass filter 84b tends to be smooth and slowly varying in character. The chopper 92 includes the diode 160, and series resistor 164 connected to the base electrode of NPN transistor 166. The saw-tooth waveform appearing at the terminal forming the common junction between capacitors 122 and 125 and resistor 118 of FIG. 4 is applied across resistor 168 to the base of a second NPN transistor 170. The collector-emitter circuits of transistors 166 and 170 are connected in series between the input of analog multiplier 30 and ground. Thus, the transistors serve as digital switches with their collector circuits in series. When transistor 166 is turned on by the read-only

memory matrix 80 through diode 160, transistor 170 chops the control voltage across resistor 172, thus, chopping the output of the multiplier 30. This simulates amplitude modulation of unvoiced components in voiced phonemes. The exponential function generator is also connected to the analog multiplier 30 through control resistor 172 by way of diode 174. The low-pass filter 84b is similarly connected to the analog multiplier through a diode 176 and the control resistor 172. The control lines 34 and 36 to the diodes 174 and 176 are indicated in conformity with the numbering of the circuits of FIG. 1.

Looking now to FIG. 8, a typical single-pole, tunable resonant filter 42 is shown. The filter shown in FIG. 8 may be that of any of the filters 42, 44, and 46 in the system 10 of FIG. 1. Representative summing resistors 178 and 180 are shown as combining the voiced and unvoiced phoneme quantities to the resonant circuit comprising inductor 181, capacitor 182, and an analog multiplier 184. The gain of the analog multiplier 184 is controlled by means of the control signal applied thereto by way of line 48, for example. The control signal applied to the multiplier 184 alters the gain of the analog multiplier 184, changing the apparent size of the capacitor 182, thus, changing the resonant frequency of the filter 42 in the circuit of FIG. 8. Resistors 178 and 180 limit the Q of the filter 42 to match the Q of the mechanical resonant filters in the vocal tract.

Looking to FIG. 9, the fixed resonant filter 68 is shown to comprise the series combination of resistor 186, inductor 188, and resistor 190, a shunt capacitor 192 being connected to ground from the point between the inductor 188 and the resistor 190. A resistor 194 is connected in parallel relationship to the entire series combination of resistor 186, conductor 188, and resistor 190 to add a portion of the input to the output. As previously mentioned, the filter 68 adds a fourth fixed resonance to the audio output waveform.

Looking now to FIG. 10, a typical low-pass filter 84 for smoothing the transitions between the levels of the analog step functions generated by the resistor ladder networks is shown. Filter 84 is a pi-section filter comprising input and output terminals 196 and 198 joined by a series inductor 200. The opposite sides of the inductor 200 are shunted to ground by capacitors 202 and 204. The capacitors 202 and 204 and the inductor 200 are chosen to produce a smooth, 70 millisecond transition at the output upon application of a step function at the input.

FIG. 11 is a representative five-input resistor ladder

network. As shown in FIG. 11, there are five input terminals 206a through 206e for applying digital signal quantities to binary weighted resistors 208a through 208e, respectively. All the resistors are combined at a summing terminal 210 which represents the analog output point for the resistor ladder network. Thus, each resistor ladder network represents a digital-to-analog converter wherein equal voltage digital signals are converted to an analog step function, the amplitude value of the step function being determined by the number of inputs energized and the weighing values of the resistors 208. As previously mentioned, the resistors are preferably weighted in a binary order, e.g., 200 ohms, 100 ohms, 50 ohms, 25 ohms, and 12.5 ohms, to produce a plurality of selectable amplitude steps. It is to be understood that weighing sequences of other than the binary type may be employed. In addition, it is to be understood that other types of ladder networks may also be employed, using more and fewer resistors.

Looking now to FIG. 12, the details of the timing section 22 of FIG. 1 are shown. The resistor ladder network 82j comprises resistors 212, 214, and 216 connected in series with respective diodes 218, 220, and 222. The cathode terminals of the diodes are connected to a common summing point 224 which is connected to the input of unijunction transistor 226. A shunt capacitor 228 is charged by the voltage appearing at summing point 224 and, upon conduction of transistor 226, discharges through resistor 230. The positive supply source B+ is connected to the opposite primary terminal of transistor 226 through a resistor 232. Thus, various charging rates for capacitor 228 may be selected in accordance with the values of the resistors 212, 214, and 216. When the charge across the capacitor 228 reaches the threshold of the unijunction transistor 226, the capacitor 228 discharges through the resistor 230 producing an output to the clock generator 100 which steps the shift register 76. Therefore, each digital command includes a timing bit or group of bits which determines the time interval during which the six-bit digital word is presented to the read-only memory matrix 80.

The following values are given by way of example to indicate the degree of amplitude modulation and resonant frequency positioning for the production of frequency power spectrums typically taken to correspond with the indicated phonemes some of which are connectives. These values are given purely by way of example and are not intended to limit the circuit and system diagrams shown herein to these specific values.

PHONEME PARAMETERS

PHO- NEME	FILTER (42)	AMPL- ITUDE (26)	FILTER (44)	AMPL- ITUDE (60)	FILTER (46)	AMPL- ITUDE (64)	AMPL- ITUDE (27)	TIME Milli- seconds
E	350	2.0	2200	2.0	2700	5.0	3.0	170
R	480	4.0	1300	2.0	1580	3.0	3.0	120
U	630	3.0	1160	1.5	2700	1.5	3.0	140
A	520	3.0	2200	2.0	2700	5.0	3.0	140
L	480	4.0	1000	0.6	2800	0.5	3.0	140
Pause	480	0.0	1700	0.0	2550	0.0	0.0	30
N	420	0.8	1950	0.3	2700	0.5	1.5	100
O	520	3.0	900	1.0	2550	1.0	3.0	120
M	350	0.6	900	0.1	2550	0.5	1.5	120
O	800	4.0	1160	2.5	2700	1.5	3.0	170
E	630	2.5	1950	1.3	2800	3.0	3.0	170
U	420	3.0	960	1.5	2700	0.6	3.0	170
A	730	2.5	1950	1.3	2700	2.5	3.0	170
I	480	4.0	1950	1.5	2700	3.0	3.0	170
B	200	0.0	730	0.0	2200	0.0	1.5	100

PHONEME PARAMETERS - Continued

PHO- NEME	FILTER (42)	AMPL- ITUDE (26)	FILTER (44)	AMPL- ITUDE (60)	FILTER (46)	AMPL- ITUDE (64)	AMPL- ITUDE (27)	TIME Milli- seconds
AW	730	3.0	960	5.0	2700	2.5	3.0	140
W	350	3.5	730	3.0	2700	1.0	1.5	100
Y	350	1.5	2200	1.5	2700	4.0	3.0	140
D	200	0.6	1950	0.0	3300	0.0	1.5	50
G	260	0.8	1950	0.0	2550	0.0	1.5	50
OO	480	6.0	960	4.0	2700	1.5	3.0	170
NG	380	1.0	2200	0.3	2200	0.4	1.5	170
S	420	0.0	1700	0.0	3700	0.7*	0.0	100
H	550	0.0	1700	0.3*	2550	0.2*	0.0	75
SH	380	0.0	1950	0.5*	2550	1.5*	0.0	120
F	480	0.0	1160	0.0	2700	0.2*	0.0	140
TH	420	0.0	1950	0.2*	3300	0.4*	0.0	100
V	480	0.0	1160	0.0	2700	0.6*	1.5	75
Z	420	0.0	1700	0.0	3700	0.4*	1.5	100
Space	480	0.0	1700	0.0	2550	0.0	0.0	60
J	380	0.0	1950	0.3*	2700	1.5*	1.5	75
TH	420	0.0	1950	0.2*	2800	0.6*	1.5	75
T	480	0.0	1950	0.1*	3300	0.4*	0.0	140
K	350	0.0	1950	0.0	2700	0.4*	0.0	140
P	420	0.0	1300	0.1*	2700	0.2*	0.0	140
CH	480	0.0	2200	0.3*	2700	1.5*	0.0	140
I	520	3.0	1800	1.5	2700	3.0	0.0	50

* = Unvoiced Component

In summary, the system of FIG. 1 synthesizes human speech by simulating the acoustic effects of human speech; that is, it produces sequences of elemental acoustic effects by defining the parameters of those effects; that is, frequency of resonance, amplitude, time interval, and waveshape. The system, because of its intentionally sluggish analog response characteristics, does not always realize the full values of the elemental parameters given above except, of course, for time duration, but makes only approaches to such values within the specified time intervals, thus, to simulate the slowly and smoothly changing dynamics of human speech. The commands may, thus, be thought of as coordinates of elemental acoustic effects, which coordinates are seldom realized precisely but which in many instances, are merely approached within the specified phoneme time interval.

DESCRIPTION OF SECOND EMBODIMENT

Referring now to FIG. 13, a second embodiment of the voice synthesizer system is disclosed. The circuit 300 shown in FIG. 13 is very similar in overall organization to the circuit of FIG. 1 and employs the same input and timing components found in FIG. 1. In particular, circuit 300 is designed to be responsive to the analog signals which are obtained from the plurality of filters 302 found at 84 in FIG. 1, as well as the resistor ladder networks 82, read-only memory matrix 80, and other input and timing control elements as are illustrated in FIG. 1. Accordingly, these elements have been omitted in the illustration of FIG. 13 to avoid duplication.

The overall organization of FIG. 13 involves the use of a vocal oscillator 304 which is the equivalent of the voiced quantity generator 14 in the circuit of FIG. 1, and a fricative or noise source 306 which is the equivalent of the Gaussian noise generator 16 in the circuit of FIG. 1. In addition, circuit 300 of FIG. 13 employs individual amplitude control units for the vocal oscillator and fricative sources as well as a plurality of tunable resonant filters, individual amplitude control units for those filters, and a summation unit to provide an audio output signal suitable for driving a loud speaker, such as that illustrated at 72 in FIG. 1. However, the specific circuit configuration illustrated in FIG. 13 differs somewhat from that of the circuit of FIG. 1 and, in addition, certain additional features including the aforementioned phoneme interaction means are employed in the

circuit of FIG. 13.

More specifically, the output of vocal oscillator 304 is applied to an amplitude control unit 308, preferably in the form of an analog multiplier, and thence to a summing amplifier 310. Similarly, the output of the fricative source 306 is applied to an amplitude control unit 312 and thence to the summing amplifier 310. Each of the amplitude control units 308 and 312 is subject to individual control via control lines 350 340, respectively, from the filter networks 302 in exactly the same manner as the analog multipliers 26 and 30 are controlled in the circuit of FIG. 1. The output of the summing amplifier 310 is connected commonly to the inputs of parallel-connected, tunable resonant filters 313, 314, 316, and 318. Again each of the tunable resonant filters is subject to individual control by an analog control signal generated in the appropriate filter networks 302 just as the tunable resonant filters 42, 44, and 46 are controlled in the circuit of FIG. 1. Amplitude control on each of the individual resonant constituents is provided by amplitude control units 320, 322, 324, and 326 which are connected to the filters 313, 314, 316, and 318, respectively. Again, each of the amplitude control units preferably takes the form of an analog multiplier and is subject to individual control from the filter networks 302. The outputs of the amplitude control units representing the four resonant frequency components in a given phoneme are separately connected to inputs of a summing amplifier 328, the output of which represents the audio signal for suitable amplification and broadcast. It will be observed that circuit 300 provides for control over all four phoneme frequency constituents whereas the circuit of FIG. 1 provides for control of only three.

The output of tunable resonant filter 313 is passed through an inverter 330 before being applied to the amplitude control unit 320. Similarly, the output of tunable resonant filter 316 is passed through an inverter 332 before it is applied to the amplitude control unit 324. On the other hand, the outputs of the tunable resonant filters 314 and 318 are noninverted, giving rise to a pattern of alternate inversion and noninversion in the four characteristic frequency components which go to make up any given phoneme. The inverters 330 and 332 will be understood by those skilled in the art to produce a 180° phase shift in the variable electrical signals passed therethrough. It has been found that this

gives rise to a deepening of the valleys between the resonant peaks A, B, C, and D of the phoneme spectrum 104, as illustrated in FIG. 2. This deepening of the valleys is accompanied by greater definition in the individual frequency components and greater realism in the speech product.

The circuit 300 of FIG. 13 employs means to produce a chopping effect on the fricative component during a voiced fricative in much the same manner as the chopper 92 operates in the circuit of FIG. 1. However, in FIG. 13 the chopping effect is accomplished by means of the connection via line 334 on the output of the amplitude control unit 308 to a low-pass filter 336 and thence to the base of an NPN transistor 338 connected to shunt to ground the analog control signal which appears at any time on the control line 340 for the amplitude control unit 312. In short, the analog output of amplitude control unit 308, which occurs at the voiced oscillator frequency, operates to modulate the fricative amplitude control signal which is being applied to the fricative amplitude control unit 312. The low-pass filter 336 operates to filter out any high frequency components in the analog signal from control unit 308 and periodically renders the transistor 338 conductive so as to ground out the analog control signal on line 340. Since this occurs at the vocal or voiced component frequency rate, the combined output applied to summer 310 is a proper voiced fricative.

It can be seen that the chopping effect of the transistor 338 is effective to modulate the fricative or noise component only at such times as a voiced component in the generated phoneme exists.

It will be observed that the modulating signal on line 334 is obtained from the analog multiplier-amplitude control unit 308 and, hence, is a more smoothly varying signal than might be obtained directly from the output of the vocal oscillator 304. This smoothly varying or analog characteristic of the modulating signal on line 334 is preferred to that of a sharply varying or digital signal as might be obtained directly from the vocal oscillator 304 and tends to enhance the realism of the overall speech product.

The first of the means for effecting phoneme interaction simulation is disclosed in FIG. 13 with reference to the formation of a so-called fricative stop. This phoneme interaction is exemplified by the abrupt amplitude modulation which occurs during normal vocal pronunciation of the word "mast." Cose analysis indicates that the *s* and *t* sounds are quite clearly separated from one another by a short and abrupt amplitude reduction.

In accordance with the embodiment of the invention illustrated in FIG. 13, this is accomplished by monitoring the phoneme commands as they are shifted through the input shift register and generating in the read-only memory matrix 80 a signal indicating the need for a fricative stop command as a result of the occurrence of a fricative phoneme such as *t* or *p*. When such a fricative occurs, a signal is generated on filter network output line 341 to actuate a fricative stop timer circuit 342 for a predetermined period which is less than that of a full phoneme time.

Looking to FIG. 14, the top line indicates a phoneme time during which a fricative such as *t* is to be generated. The middle line of FIG. 14 indicates the normal phoneme command for that fricative phoneme, this command being applied to amplitude control unit 312

by way of line 340. The bottom line of FIG. 14 indicates the generation of a fricative stop signal of short duration, this signal being generated by the fricative stop timer unit 342. This signal is applied to the base of a transistor 344, the emitter of which is connected to ground through a resistor 346 and the collector of which is connected to a junction in line 340. Accordingly, during the generation of the fricative stop signal, the normal fricative command voltage on line 340 is shunted to ground through transistor 344 so as to produce an abrupt amplitude reduction between the fricative phoneme and the preceding phoneme but substantially during the fricative phoneme time interval.

Another of the phoneme interactions is involved in the formation or simulation of a vocal closure, such as the phoneme *b* following and/or preceding a voiced phoneme such as *e*. In the circuit of FIG. 13 this is accomplished by means of a similar phoneme sensor arrangement in the input shift register and a similar responsive signal generation combination in the read-only memory matrix 80 to actuate a control line 351 by way of a suitable filter and resistor ladder network. Control line 351 is connected to the base electrode of the transistor 348, the emitter of which is connected to ground through resistor 352 and the collector of which is connected to a junction on amplitude control line 350. The amplitude control line 350 carries the normal phoneme command which is applied to the amplitude control unit 308 for the vocal oscillator or voiced quantity generator 304. Accordingly, the application of a signal to line 351 shunts the phoneme command to ground for a short period equal to less than the full phoneme interval by rendering transistor 348 momentarily conductive. In this fashion, an abrupt amplitude modulation for the voiced component is accomplished in much the same way as the fricative stop is accomplished.

The emitter electrodes of both transistors 344 and 348 are preferably connected to ground through suitable resistors such as 346 and 352 so as to prevent a total reduction to ground potential in the respective amplitude control signals and to regulate as desired the degree of amplitude modulation which is effected.

The third phoneme interaction means in the circuit 300 of FIG. 13 is provided for the simulation of a nasal closure such as the voiced quantity *n* or *m* or *ng*. This phoneme interaction also involves amplitude modulation during the phoneme interval for the nasal closure, but is carried out by selective modulation of the major frequency components rather than direct amplitude modulation of either the voiced or unvoiced phoneme quantities as is the case with respect to fricative stops and vocal closures.

In FIG. 13, the nasal closure amplitude modulation is accomplished by means of a transistor 366 which is connected to shunt to ground the amplitude control signals normally applied to control lines 354, 356, and 357 which are associated with amplitude control units 326, 324, and 322, respectively. As was explained with reference to the circuit of FIG. 1, the control lines 354, 356, and 357 carry suitable amplitude control signals for the analog multiplier amplitude control units and are subject to short term grounding for amplitude control units and are subject in the same fashion as the control lines 340 and 350, as previously described. To accomplish this, the control currents or control voltages are shunted substantially to ground through respective di-

odes 360, 362, and 364 associated with control lines 354, 356, and 357, respectively. The cathodes of each of the diodes are connected to the collector electrode of transistor 366 and thence to ground through the emitter electrode of the transistor and through a resistor 368. The transistor 366 is rendered conductive by means of a signal on control line 370 which is applied through a resistor 372 to the base or control electrode of the transistor 366. The diodes 360, 362, and 364 are provided to prevent reverse current flow from one control line to another and, thus, to permit the simultaneous control of all three control lines by way of a single transistor.

Looking now to FIG. 15 a further phoneme interaction system is provided for the accomplishment of amplitude control during a brief interval between a fricative and a pure voiced phoneme. In FIG. 15, differential adders 370 and 372 are shown to be connected into the control lines 350 and 340 for the voiced component amplitude control unit 308 and the fricative amplitude control unit 312, respectively. The differential adders 370 and 372 are connected in such a way as to cause a reduction in the amplitude control signal applied to the appropriate analog multiplier-amplitude control unit by a factor equal to seventy percent of the opposite phoneme component signal; i.e., the fricative amplitude control signal is reduced by seventy percent of the voiced component amplitude control signal and the voiced amplitude control signal is reduced by seventy percent of the fricative amplitude control signal.

Looking more specifically to FIG. 15, it can be seen that the voiced phoneme amplitude control signal on line 350 is connected by way of resistor 374 to the negative input of differential adder 370 and also by way of resistor 388 to the positive input of differential adder 372. Conversely, the fricative amplitude control signal on line 340 is connected by way of resistor 376 to the positive input of adder 370 and by way of resistor 386 to the negative input of adder 372. Adder 370 has a resistive feedback connection 378 between the output and the negative input thereof and similarly, adder 372 has a resistive feedback connection 390 between the output and the negative input thereof. The output of adder 370 is caused by proper selection of resistive values to be equal to one-hundred percent of the fricative amplitude control signal on line 340 less seventy percent of the voiced component amplitude control signal on line 350. This signal is applied by way of resistor 380 to the fricative amplitude control unit 312, as shown. A voltage clamp device in the form of an operational amplifier 382 having a unidirectionally conducting diode 384 in the feed-back path is employed to prevent negative voltages from flowing to the amplitude control unit 312.

The output of the differential adder 372 is caused by proper selection of resistive values to be equal to one-hundred percent of the voiced component amplitude control signal less seventy percent of the fricative amplitude control signal. This output is applied by way of resistor 392 to the voiced amplitude control unit 308, as shown. A clamp device in the form of an operational amplifier 394 having a diode feedback 396 is employed to prevent negative voltages from flowing to the amplitude control unit 308.

The overall effect of the circuit of FIG. 15 is shown in FIG. 16. In FIG. 16 the example selected involves successive phonemes *s* and *e*, the normal curves of

which tend to overlap and run into one another as indicated by the dotted or dashed lines in FIG. 16. However, because of the subtraction factor as between the two phonemes as carried out by the circuit of FIG. 15, a deep amplitude valley in the control signal amplitudes tends to occur between the two phonemes, thus, giving rise to another form of phoneme interaction or vocal stop.

In implementing the circuit 300 of FIG. 13, it is preferred to employ a vocal oscillator 304 having a saw-tooth output waveform. The type of waveform has the advantage of producing a distribution of frequency components with time and produces a more natural sounding speech than an impulse function of the type used in FIG. 1. The impulse function does not exhibit the time distribution of frequency components inherent in the saw-tooth, but rather drives all of the tunable filters with all frequencies at substantially the same time. This has a tendency to produce a rasp or buzz in the basic voiced quantity.

It is to be understood that the circuits of FIGS. 1 and 13 are representative of two fundamentally similar but specifically distinct embodiments and that each of the embodiments embraces specific features and concepts which are adaptable to the other embodiment. Accordingly, it is possible, as will be apparent to those skilled in the electronics art, to intermingle the various specific features of one circuit with the various specific features of the other circuit so that practically an infinite number of specific implementations is possible in accordance with the overall teachings of the present invention.

The embodiments of the invention in which an exclusive property or privilege is claimed are defined as follows:

1. A speech synthesizer system comprising: input means responsive to digital commands to produce analog control signal combinations representing the steady state characteristics of phonemes to be generated; first generator means for producing a first audio signal quantity representing a voiced phoneme constituent, second generator means for producing a second audio signal quantity representing an unvoiced phoneme constituent; first control means for modulating the amplitude of the voice phoneme quantity according to a first of said analog control signals; second control means for modulating the amplitude of the unvoiced phoneme quantity according to second of said analog control signals; a plurality of tunable resonant filters each being responsive to respective third analog control signals to assume selected resonant conditions according to the predetermined resonant frequency poles of the phonemes being generated, means connecting the outputs of the first and second control means to at least some of said filters, output means for cumulatively receiving the quantities passed by said filters, timing means responsive to fifth of said control signals to define the time intervals of said phonemes, and relatively slow acting filter circuit means connecting said analog control signals to said control means and tunable resonant filters and having transfer characteristics which, in sufficient time, duplicate the end parameters of the control signals but which, within the time intervals set by said timing means, substantially prevent said steady state characteristics from being reached during successive phoneme generation for synthesized speech.

2. A speech synthesizer system as defined in claim 1 wherein the input means comprises storage means for receiving the digital commands, decoder means for deriving from said digital commands a plurality of digital signal outputs, and digital-to-analog converter means for converting said outputs to said control signal combinations.

3. A speech synthesizer system as defined in claim 2 wherein the storage means operates under the control of the timing means to transfer the digital commands to the decoder means.

4. A speech synthesizer system as defined in claim 2 wherein the decoder means is a read-only memory matrix.

5. A speech synthesizer system as defined in claim 2 wherein the converter means includes a plurality of networks for converting each of said digital signal outputs to corresponding relatively abruptly variable analog step functions.

6. A speech synthesizer system as defined in claim 5 wherein the converter means further includes a plurality of filter circuits connected to the outputs of corresponding networks and having a finite response time to thereby smooth the relatively abruptly variable analog functions into relatively smoothly variable analog functions.

7. A speech synthesizer system as defined in claim 6 wherein each of the filter circuits is a low-pass filter including an inductive element to pass low frequency components of said functions and a capacitor to shunt high frequency components of said functions.

8. A speech synthesizer system as defined in claim 6 wherein the filter circuits are connected in a predetermined pattern to the first and second control means and to each of the plurality of tunable filter means to produce relatively smooth, dynamic transitions between phonemes formed sequentially by the system.

9. A speech synthesizer system as defined in claim 1 wherein the first generator means includes an amplifier for producing a selectively variable frequency, periodic waveform.

10. A speech synthesizer system as defined in claim 9 including means interconnecting the amplifier and the input means for varying the frequency of said periodic waveform for inflection control.

11. A speech synthesizer system as defined in claim 1 wherein the second generator means is a broadband Gaussian noise generator.

12. A speech synthesizer system as defined in claim 1 including control means connected operatively between the input means and the second control means for amplitude modulating the unvoiced phoneme quantity at a rate which is slow relative to the center frequency of said quantity.

13. A speech synthesizer system as defined in claim 12 wherein the control means is a chopper.

14. A speech synthesizer system as defined in claim 1 wherein the output of the second generator means is combined with the output of the first generator means at the first control means to add an unmodulated portion of the unvoiced phoneme quantity to each voiced phoneme quantity.

15. A speech synthesizer system as defined in claim 1 wherein said plurality of tunable resonant filters are at least three in number, each of said filters being tunable under the control of said control signals to resonate at frequencies within a selected band whereby

each of the phonemes produced may exhibit a resonant envelope having at least three selected amplitude poles.

16. A speech synthesizer system as defined in claim 15 wherein each of the tunable filters is connected to receive a different control signal from said input means.

17. A speech synthesizer system as defined in claim 16 including networks in said input means for producing analog control signal variations for the control of said tunable filters.

18. A speech synthesizer system as defined in claim 17 further including filter circuits in said input means and individually connected between the networks and the tunable filters for smoothing relatively abrupt amplitude transitions in the control signals as applied to the tunable filters thereby to smooth transitions between successive phonemes.

19. A speech synthesizer system as defined in claim 1 wherein the output means includes a fixed resonance filter to add a synthesis of nasal resonance to the phonemes received and produced thereby.

20. A speech synthesizer system as defined in claim 1 wherein the first and second control means are variable gain amplifiers.

21. A speech synthesizer system as defined in claim 20 including an exponential function generator connected operatively between the input means and the second control means for synthesizing plosive phonemes.

22. A speech synthesizer system as defined in claim 1 comprising means for modulating at least one of said analog control signals to effect an abrupt amplitude change in at least a constituent of a given phoneme in accordance with the interaction of said given phoneme and certain predetermined other adjacent phonemes.

23. A speech synthesizer as defined in claim 22 wherein said means for modulating includes a selectively operated shunt circuit connected to said first control means for modulating said first analog control signal during at least a portion of a given phoneme interval.

24. A speech synthesizer as defined in claim 22 wherein said means for modulating includes a selectively operated shunt circuit connected to said second control means for modulating said second analog control signal during at least a portion of a given phoneme interval.

25. A speech synthesizer system as defined in claim 22 wherein said means for modulating includes selectively operable shunt circuit means connected to at least one of said individual control means for modulating at least one of said still additional analog control signals thereby to effect an amplitude modulation of at least one of the resonant conditions during at least a portion of a given phoneme interval.

26. A speech synthesizer system as defined in claim 25 wherein said individual control means each includes an analog signal responsive multiplier and a control signal conductor connected between said input means and said control means, said shunt circuit means comprising a transistor switch having a control electrode and a pair of primary electrodes, said primary electrodes being connected to said control signal conductor in shunt relation to reduce the control signal thereon when said switch is conducting, and additional control

signal conductor means connected between said input means and said control electrode.

27. A speech synthesizer system as defined in claim 22 wherein said means for modulating comprises circuit means connected between said input means and said first and second control means for subtracting at least a portion of said first analog control signal from said second analog control signal and applying the result to said second control means.

28. A speech synthesizer system as defined in claim 27 wherein said means for modulating further comprises circuit means for subtracting at least a portion of said second analog control signal from said first analog control signal and applying the result to said first control means.

29. A speech synthesizer system as defined in claim 28 including signal clamping means connected to said circuit means for establishing minimum control signal levels for said first and second control means.

30. A speech synthesizer system as defined in claim 1 wherein the tunable filters are three in number, the first tunable filter being connected to receive a mixture of an unmodulated unvoiced phoneme quantity and the voiced phoneme quantity and being controllably tunable to resonate at frequencies within a first relatively low frequency band, the second tunable filter being connected to receive both the modulated voiced and modulated unvoiced phoneme quantities and being controllably tunable to resonate at frequencies within a second relatively intermediate frequency band, and the third tunable filter being connected to receive both the modulated voiced and modulated unvoiced phoneme quantities and controllably tunable to resonate at frequencies within a third relatively high frequency band.

31. A speech synthesizer system as defined in claim 30 wherein the low, intermediate and high ranges are approximately 100 Hz to 1,000 Hz, 500 Hz to 3,000

Hz, and 1,000 Hz to 4,000 Hz, respectively.

32. A speech synthesizer system as defined in claim 30 wherein each of said tunable filters includes a tuning capacitor, and means for varying the voltage across the capacitor to tune said filter.

33. A speech synthesizer system as defined in claim 6 wherein at least some of the phonemes addressable via the input means are of a duration which is shorter than the response time of the filter circuits.

34. A speech synthesizer circuit comprising means for producing a basic voiced phoneme quantity; means for producing a basic unvoiced phoneme quantity; selectively addressable control means for variably modulating the amplitudes and resonance envelopes of combinations of the phoneme quantities, input means for generating a sequence of relatively abruptly varying analog control signals at predetermined occurrence intervals and representing nominal steady state values of phoneme amplitude and frequency content, and relatively slow response filter circuits connected to apply relatively smoothly varying counterparts of said analog control signals to the control means thereby to produce a smooth dynamic transitions between successive phonemes, the filter means being of such electrical transfer characteristics as to be capable of reconstructing at the outputs thereof the end parameters of said analog control signals in a time period which is substantially longer than the predetermined time intervals between the signals in said sequence whereby said steady state values typically cannot be reached within said phoneme time intervals.

35. A speech synthesizer circuit as defined in claim 34 and further comprising means for selectively effecting an abrupt amplitude change in at least a constituent of a phoneme in accordance with the predetermined interaction of said phoneme and at least one other phoneme which is adjacent in time.

* * * * *

40

45

50

55

60

65

UNITED STATES PATENT AND TRADEMARK OFFICE
CERTIFICATE OF CORRECTION

PATENT NO. : 3,836,717

DATED : September 17, 1974

INVENTOR(S) : Richard T. Gagnon

It is certified that error appears in the above—identified patent and that said Letters Patent is hereby corrected as shown below:

Insert the sheets of drawings containing Figures 13 through 16. (See Attachment).

On the Title Page, "12 Drawing Figures" should read

-- 16 Drawing Figures --.

Signed and Sealed this

Twenty-seventh **Day of** *May 1980*

[SEAL]

Attest:

SIDNEY A. DIAMOND

Attesting Officer

Commissioner of Patents and Trademarks

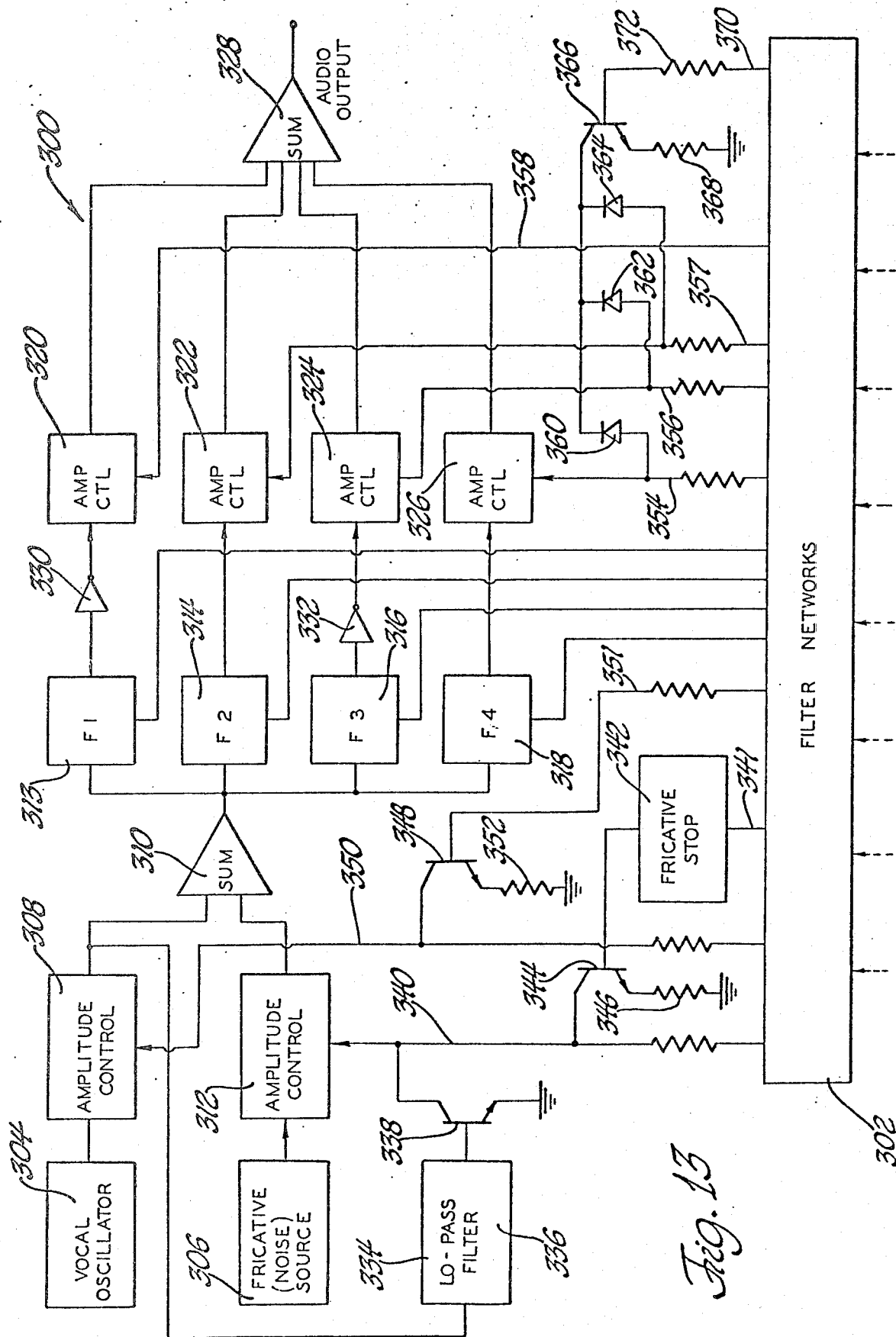


Fig. 13

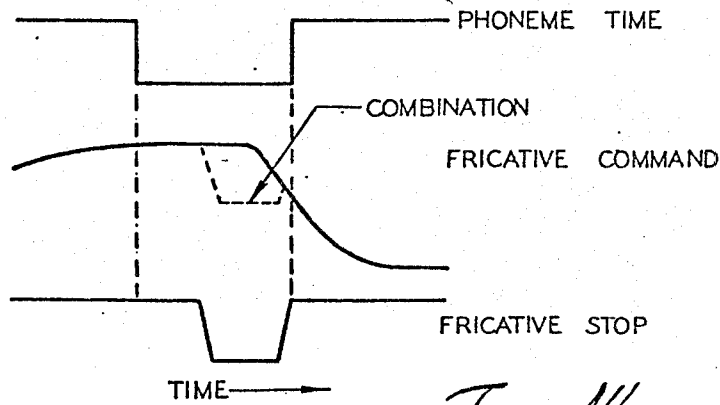


Fig. 14

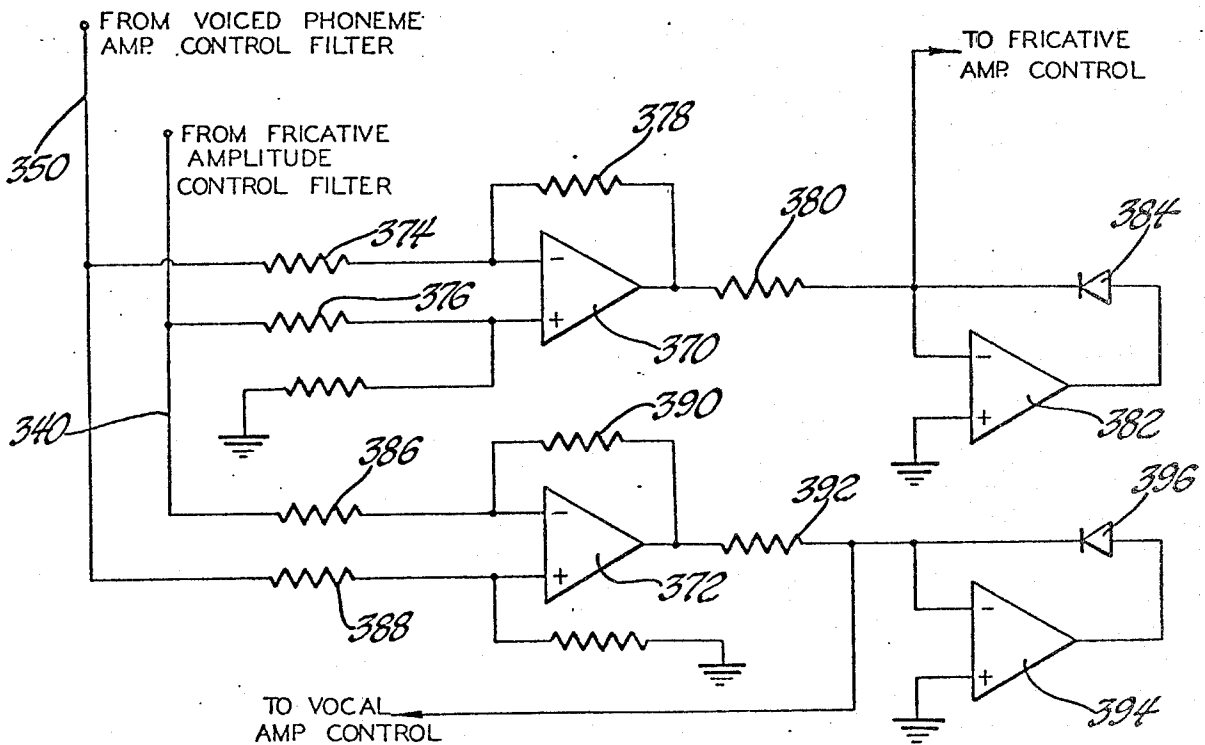


Fig. 15

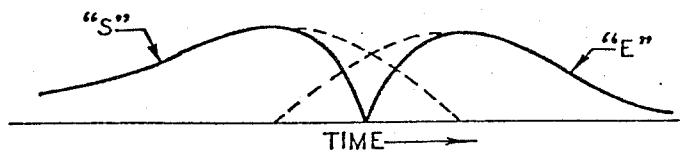


Fig. 16