May 9, 1967 J. L. DE CLERK ETAL 3,319,002

ELECTRONIC FORMANT SPEECH SYNTHESIZER

Filed May 24, 1963 5 Sheets-Sheet 1

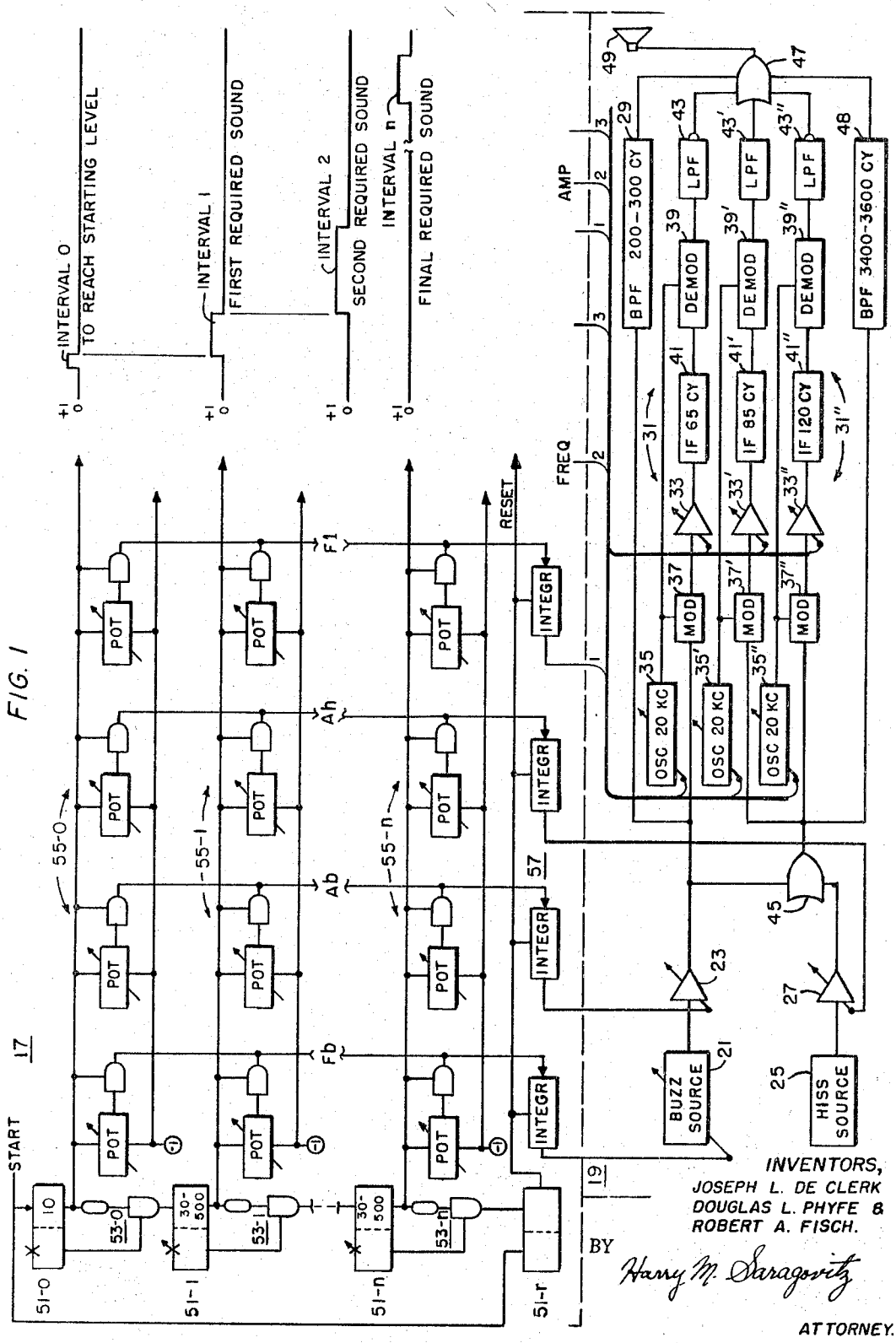


FIG. 1

INVENTORS,
JOSEPH L. DE CLERK
DOUGLAS L. PHYFE &
ROBERT A. FISCH.

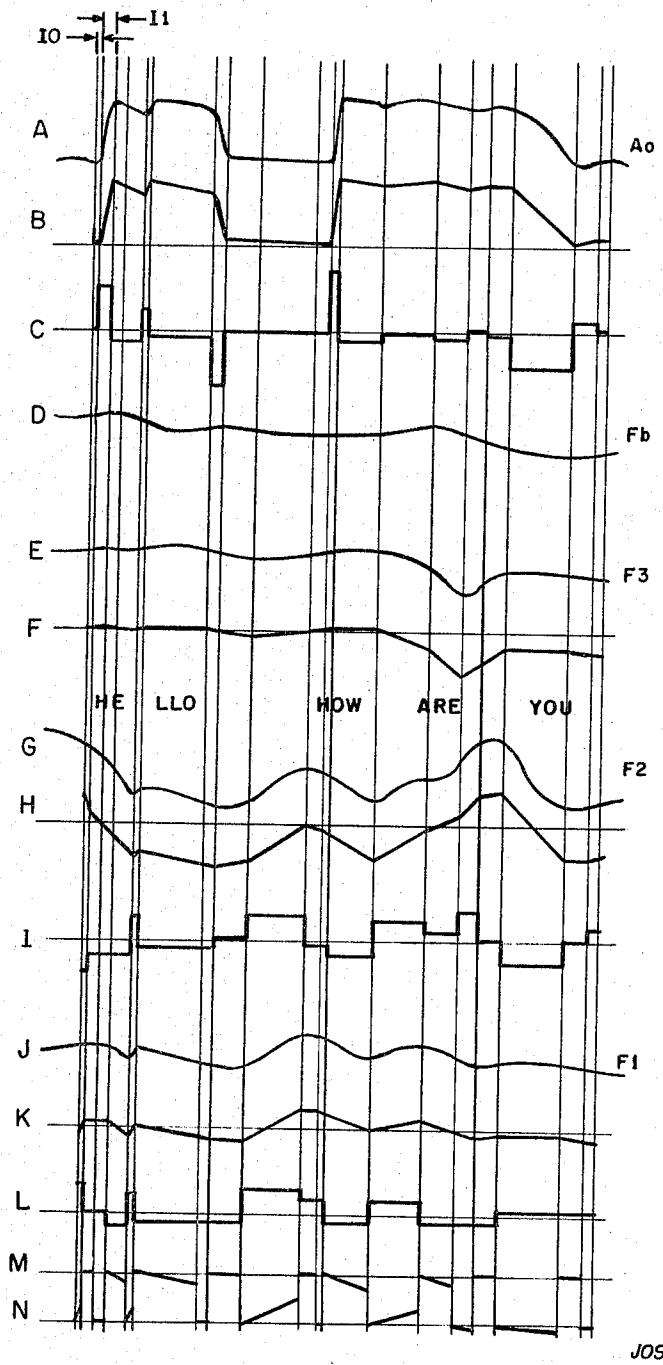
BY Harry M. Saragovitz

ATTORNEY.

ELECTRONIC FORMANT SPEECH SYNTHESIZER

Filed May 24, 1963 5 Sheets-Sheet 2

FIG. 2



INVENTORS,
JOSEPH L. DE CLERK
DOUGLAS L. PHYFE &
ROBERT A. FISCH.

BY Harry M. Saragovitz

ATTORNEY

May 9, 1967 J. L. DE CLERK ETAL 3,319,002

ELECTRONIC FORMANT SPEECH SYNTHESIZER

Filed May 24, 1963 5 Sheets-Sheet 3

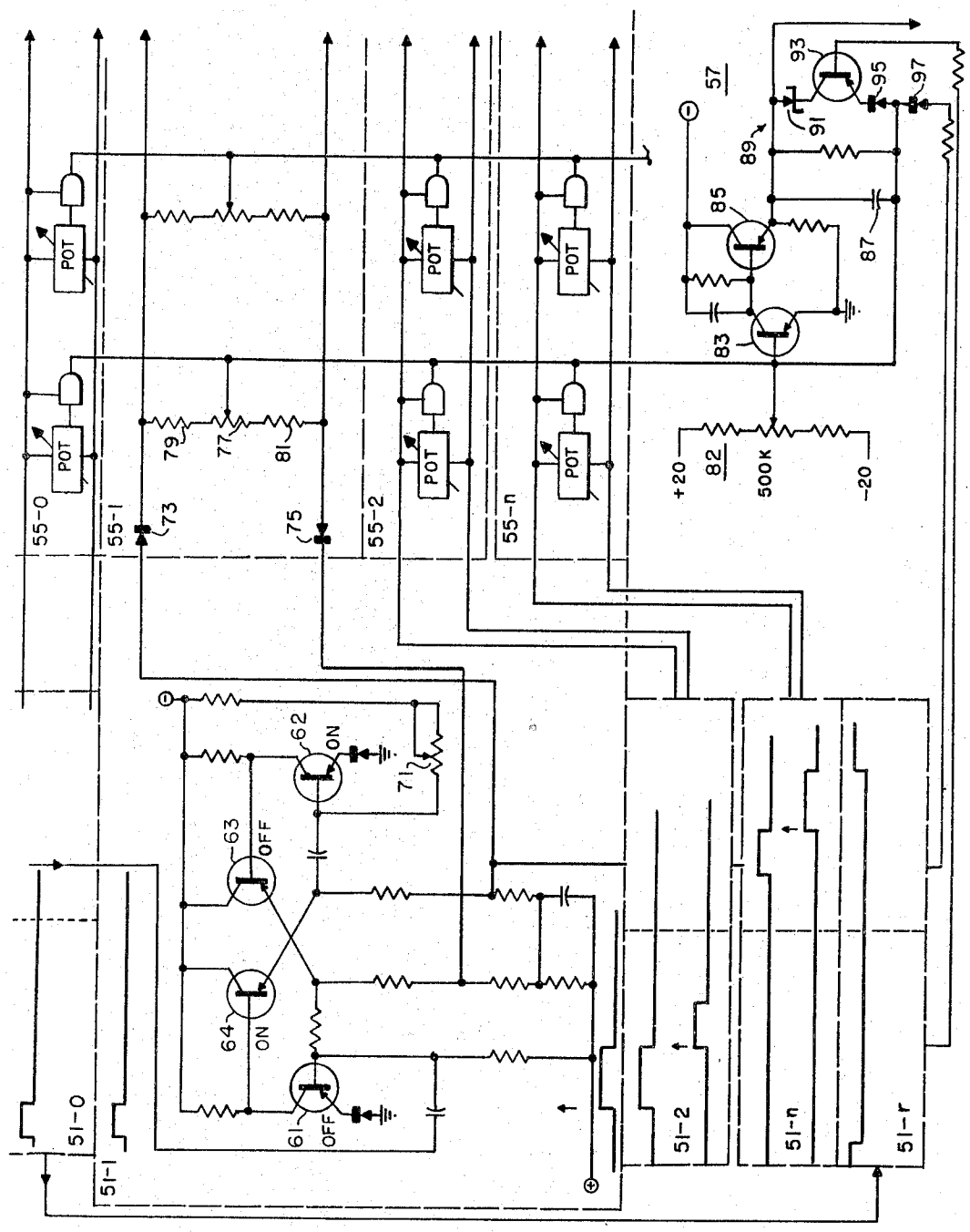


FIG. 3

INVENTORS,
JOSEPH L. DE CLERK
DOUGLAS L. PHYFE &
ROBERT A. FISCH.

BY Harry M. Saragovitz

ATTORNEY.

May 9, 1967 J. L. DE CLERK ETAL 3,319,002

ELECTRONIC FORMANT SPEECH SYNTHESIZER

Filed May 24, 1963 5 Sheets-Sheet 4

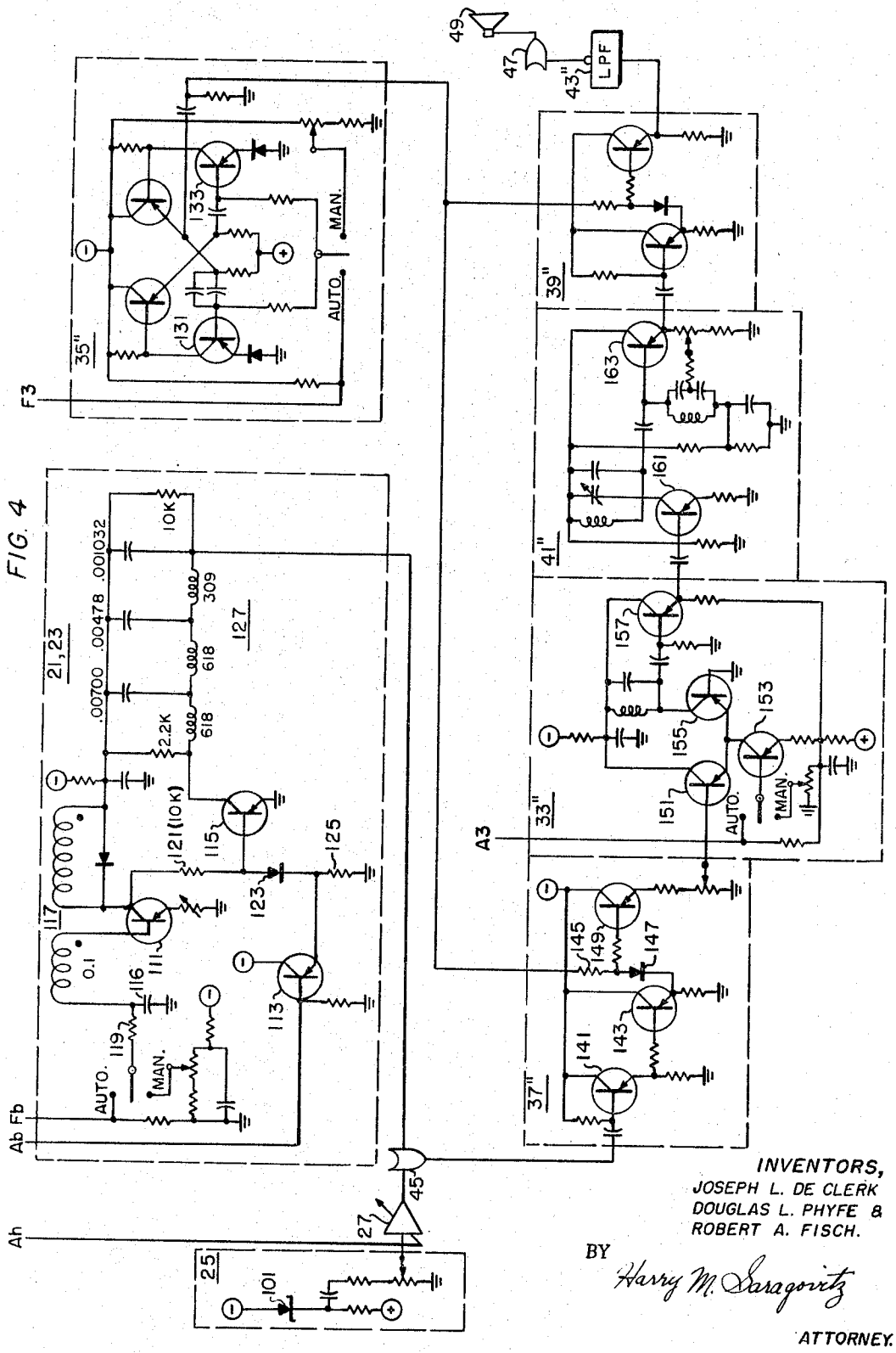


FIG. 4

INVENTORS,
JOSEPH L. DE CLERK
DOUGLAS L. PHYFE &
ROBERT A. FISCH.

BY Harry M. Saragovitz

ATTORNEY.

May 9, 1967 J. L. DE CLERK ETAL 3,319,002

ELECTRONIC FORMANT SPEECH SYNTHESIZER

Filed May 24, 1963 5 Sheets-Sheet 5

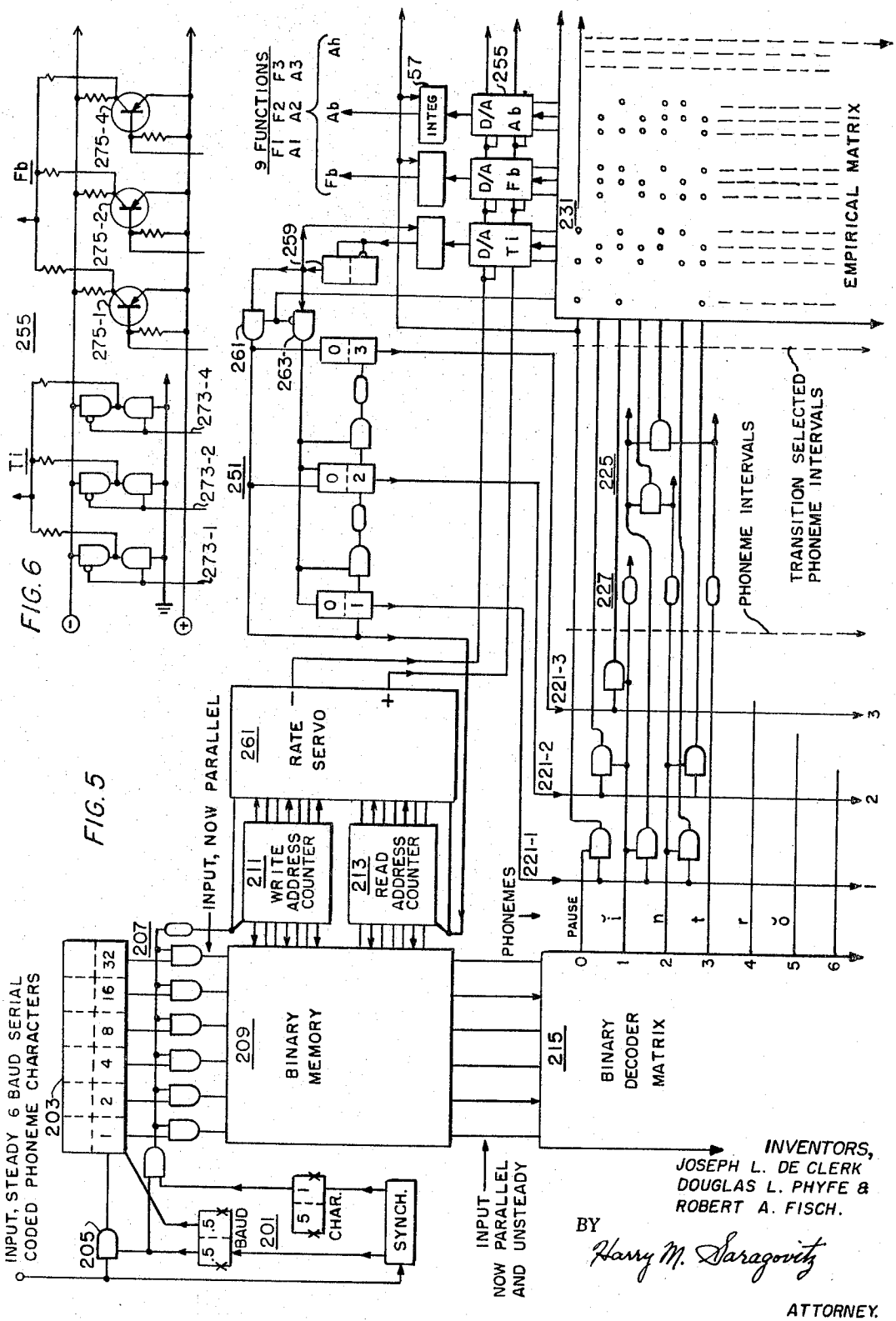


FIG. 6

FIG. 5

INVENTORS,
JOSEPH L. DE CLERK
DOUGLAS L. PHYFE &
ROBERT A. FISCH.

BY Harry M. Saragovitz

ATTORNEY.

United States Patent Office 3,319,002
Patented May 9, 1967

3,319,002
ELECTRONIC FORMANT SPEECH SYNTHESIZER

Joseph L. De Clerk, Red Bank, and Douglas L. Phyfe, Ridgewood, N.J., and Robert A. Fisch, Danbury, Conn., assignors to the United States of America as represented by the Secretary of the Army

Filed May 24, 1963, Ser. No. 283,118
3 Claims. (Cl. 179—1)

The invention described herein may be manufactured and used by or for the Government for governmental purposes, without the payment of any royalty thereon.

This invention relates to techniques used for synthesis of artificial speech sounds, the empirical function generators required for such synthesis, control circuits responsive to such functions, and typical systems and methods in which the means and techniques might be used. One of the major purposes for the invention is to provide for an increase in communication efficiency, for example, by reducing speech to a bit code generally resembling teletype, then synthesizing the sounds from the code. In this case each code group represents about 1 of 40 common sounds, phonemes, or shorthand characters rather than 1 of 26 ordinary letters, only very generally corresponding to certain sounds and probably requiring more letters than phonemes because of various silent letters, diphthongs, etc. The invention is directed particularly to synthesizing the sounds. Such synthesis also would be useful to provide for mutes to communicate by "voice" and to analyze speech defects for treatment. These sounds or combinations could easily be coded, providing they could be isolated from a voice input and recombined to provide a voice output. Thus the channel capacity required for transmitting voice information would be similar to that for the same information content by teletype. The present invention is concerned with the recombining to a voice output, particularly the synthesis of the actual sounds and the generation of complex voltage functions to control such synthesis.

In order to recognize the advantages in potential applications of the present invention it will be helpful to compare several common forms of communication. Assuming a normal speech rate of 100 words per minute, 6 letters per word and 5 bits per letter ($2^5=32$ possible combinations for 26 letters and 6 special operations such as a shift to numerals, etc. instead of letters) or 5 phonemes per word and 6 bits per phoneme ($2^6=64$ possible combinations for 40 phonemes), ordinary speech information might be transmitted at 3000 bits per minute or 50 bits per second corresponding to only 25 cycle bandwidth either in teletype or the present system. On the other hand, reasonable speech quality for ordinary telephone transmission requires 3000 cycle bandwidth, 120 times as great, and for radio broadcasting 10,000 cycle bandwidths, 400 times as great. For further comparison a (European) television picture also at 50 interlaced fields per second (25 complete frames per second) using about 250 lines per field and 500 bits per line or 125,000 bits per field, requires 125,000 times as much bandwidth as teletype or the present system and over 300 times as great as radio broadcasting.

One purpose of teletype is to avoid operator training as required in receiving Morse Code, etc. Recording and visual or other examination of the bits to determine the meaning would defeat this purpose even for anyone so trained and usually limit the practical speed of reception; therefore the teletype receiver performs the decoding, using mechanical type bars to print ordinary letters, or for very high speeds electrostatically operated monogram matrices of about 35 possible dots, readily recognizable as ordinary letters. One might suggest that transmitter tape or message sheet could be prepared in such dot form, sending corresponding bits to be easily recognized, but requiring 7 times as much channel capacity (175 cycle bandwidth), a prohibitive increase since the teletype coding art is already highly developed and the apparently simpler coding is not a major advantage.

It it also noted that the same coded message intended for voice reproduction according to this invention could be stored by printing using rather than usual letters a slightly expanded "alphabet" of phoneme characters easily read without special training, although a very general familiarity with stenography, stenotypy, etc. might be helpful. (Speech and Hearing, Fletcher, 1929, Van Nostrand, p. 84, also recently publicized Pittman phonetic spelling system being tested in England.)

At the transmitter end the code might be prepared automatically from a speech input or manually on a phoneme keyboard device, a compromise between a typewriter spelling fully in common letters and a stenotypy device using much oversimplified spelling, both phonetic only in a very crude sense.

For convenience in analysis of the operation of this invention it may be desirable to consider its analogy to the operation of the brain and the mouth. In some respects this may involve an over simplification since it is probable that the brain controls the operation of the mouth but also responds partially to the sounds received in the ear and also to basically mechanical return signals to the brain representing the actual operation of the vocal tract. In the present case there are no such return signals to the brain and it can merely provide the instruction to the mouth assuming that such instruction is actually followed. A precise line of demarcation as between the brain and mouth probably does not exist but for convenience the entire time domain function generator is considered as the brain. The time domain function generator internally includes a series of square wave portions which determine the slope of the corresponding portions of the output functions; one might consider such square wave pulses as the instruction to the mouth regarding the time variation required to form the desired sounds. On the other hand, an integrator is used whose output corresponds to the instantaneous position required of the vocal members. It is a matter of viewpoint whether this integration occurs in the brain as here assumed or in the mouth.

In analyzing the operation of the synthesizer or mouth portion it will be helpful to recognize that the vocal tract or airway involves a path from the lungs as the source of airflow, the larynx membrane vibrated by such airflow, and various cavities, the lower pharynx up to the soft palate (velum tongue) controlling two alternative or parallel paths, thru the nasal pharynx or the oral cavity modified by the tongue, teeth, and lips where the sounds can be further modified. In the synthesizer there are a plurality of formant portions corresponding to these cavities of the vocal tract.

There may also be questions as to how much more belongs in the "brain" portion. The human normally hears someone else, consciously thinks of an appropriate response in words of a particular language, subconsciously prescribes the necessary sound formation for such words by a memory-like process, then provides the complex time function instruction to the vocal system. As initially described only this last aspect of the brain operation will be considered; the previous aspects are considered already complete without any restriction as to the time required for settings of appropriate potentiometers, etc. In the case of a coded speech communication system the decoding would have to prescribe the

necessary sound formation very rapidly as in the case of the actual brain noted above.

Much work has been attempted with acoustical models of the vocal system, but the controls are extremely complex. The channel vocoder involves an attempt to select variable amplitude components from a substantial number of bands in the voice spectrum; again to get useful quality the controls are complex. In the present case relatively few sources are controlled, each in a slightly more complex manner, but with a relative very limited overall complexity considering the attainable quality.

The principal object of this invention is the synthesis of artificial speech having a substantially natural sound in as simple manner as possible. A further object is the provision of a complex time domain function generator used for control of such synthesis. A further object is the provision of a simple amplitude control varied by such a complex function. A still further object is the provision of a narrow band frequency control varied as to the particular frequency range by such a complex function. Various further objects of the invention will become apparent from the accompanying drawings and particularly the following description and claims.

In such drawings:

FIG. 1 represents a preliminary application of the invention for synthesizing various illustrative sound sequences.

FIG. 2 represents typical waveforms produced by the complex function generator for a particular sentence, illustrated as "Hello, how are you?"

FIG. 3 illustrates typical detailed circuits for the complex function generator or "brain" portion of FIG. 1, while FIG. 4 similarly illustrates the synthesizer or "mouth" portion of FIG. 1.

FIG. 5 represents a more general application of the invention for synthesizing actual information sound sequences as might be transmitted over wire or wireless communication channels, using binary information storage devices for the now even more complex function generation.

FIG. 6 represents a typical digital to analog converter used in FIG. 5 to convert the binary information to the digital form required for the complex function generation.

For simplicity in reducing the need for word legends, etc. the drawings use various symbols generally corresponding to the logic symbols of U.S. Army MIL-STD-806A, such as:

(a) the almost universal D shaped shield for AND gate,

(b) a generally triangular shield with one concave and two convex sides for usual OR gate (with an X thru it in the case of an EXclusive-OR group of gates as in half adders, mod-2 adders, etc.),

(c) a mere triangular shield for amplifiers (where this function is not merely implied as inherent in other components but actually significant to the operation as in case of inversion, variable gain, operational networks, etc.),

(d) a small circle (or half circle to distinguish from other uses of a circle) for NOT or INHibit inputs or outputs of gates, etc.,

(e) a rounded end narrow rectangle for a delay device,

(f) a short rectangle (suggestive of the two alternative "sides") for a two-state (binary) stable or quasi-stable circuit such as a flip-flop,

(g) a long rectangle (suggestive of the several binary stages with various alternative arrangements of gates, delays, etc., not usually shown in detail) for a shift register.

In the case of binaries, the two-state characteristic of such sides is often further emphasized by a dotted divider line with:

(a) Outputs from either or both sides,

(b) Ordinary inputs to either or both sides,

(c) Complement or count input at such divider line (implying the gating function of such an input, and for sequential inputs permitting use somewhat analogous to EXclusive-OR gates),

(d) An X, suggestive that an input is non-essential to the particular side (as in the stable side of monostable or in both sides of astable circuits, which may also have an ordinary input for such purposes as synchronizing),

(e) A common input, direct to one side and thru a NOT circuit to the other side, suggestive of Schmitt trigger operation (a binary output but not strictly binary input).

Arrangement and extra detail also aid drawing clarity for true logic, analog, or even schematic diagrams, such as:

(a) Never more than one analog input to AND gate symbols, always shown on flat side,

(b) Divider lines in shift registers and other block symbols,

(c) Greater detail for deviations from conventional features,

(d) Symmetry, analogy, and other orderly plan, usually reducing crossed leads and permitting signal flow from upper left to lower right, with arrows for exceptions or emphasis,

(e) In place of unnecessary word legends, numerical or other characters to identify "weight" significance of the sides of common binaries, the time in bauds, milliseconds (ms.), microseconds (μs), etc. of delays and unstable sides of binaries, the transition levels of Schmitt triggers in volts, amperes, etc. (or merely polarity), or mere reference to related components or signals. The weights usually provide a convenient orderly plan for identifying various code combinations whether ultimate use involves actual quantitative significance (sometimes non-linear) of an output amplitude or merely identification of letters or other elements having no apparent orderly relation whatever to the weights.

These symbols avoid language problems involved in word legends, save space, usually identify direction and avoid need to consider polarity, and are considerably simpler than corresponding schematics and sometimes clearer. Even an elementary circuit often is further clarified as a few simple symbols, such as a transition circuit (as used in sequence trains, counters, shift registers, etc.) shown in logic form as an AND gate with direct and delayed inputs from the two sides of the binary; the corresponding schematic of an elementary capacitor coupling would also be rather simple, per se, but would require details of the associated circuits to avoid ambiguity.

A typical system is symbolically illustrated in block diagram form in FIG. 1. The top part of the figure relates to the function generator 17 ("brain") which provides suitable approximations to the complex functions for controlling the several parameters of the desired sounds. This part of the figure is abridged (1) to show elements for these functions controlling buzz frequency, buzz and hiss amplitudes, and one of the formant frequencies only, omitting those for the other two formant frequencies and for the amplitudes of all three, and also (2) for each such function to show elements only for the initial or zero control interval, used merely to attain an appropriate starting level for the actual sounds desired, and the first and last sound intervals, in each of which substantially straight line segments approximate corresponding intervals of the desired complex functions. The lower part of the figure relates to the sound synthesizer 19 ("mouth"), including circuits controlled by the function generators, and will be described first.

In the synthesizer, the buzz source 21 supplies the generally low frequency "voiced" components, characteristic of vowels, controllable both in frequency and amplitude as indicated by the arrows thru the buzz source 21 and thru its output amplifier 23; the hiss source 25 supplies the generally higher frequency noise-like "unvoiced" components, characteristic of consonants, controllable only in amplitude as indicated by the arrow thru its output ampli-

fier 27. The buzz source output is supplied to a simple bandpass filter 29 having a pass-band from about 200 to 300 cycles, to provide a typical background tone to the voice, affected only moderately by the words to be formed.

The buzz source output is also supplied thru the first formant controlled network 31, comprising modulator 37, variable amplifier 33, narrow band IF filter 41, demodulator 39, and low pass filter 43. In this network both amplitude and dominant frequency components are significantly controlled to produce the desired characteristics of the words to be formed, as indicated by the arrows thru the amplifier 33 and the carrier oscillator 35 typically around 20 kc. This oscillator output is applied to modulator 37 and demodulator 39 resulting in a temporary frequency translation up to and back from a range near 20 kc. This range is designated as intermediate frequency (IF) by analogy to prevailing heterodyne broadcast receiving techniques. Both amplitude and frequency are most conveniently controlled in this IF range, particularly the frequency, which is controlled by means of the selectivity of a narrow band IF filter 41 (approximately 65 cycles bandwidth in the case of the first formant). As would be apparent from the frequency of the carrier oscillator, the filter operates near 20 kc., but the legend indicates only the bandwidth. The low pass filter 43 removes the higher frequency modulation products, retaining only the audio frequency buzz source components as variably limited in the IF stages.

Assuming a buzz source output from 65 cycles to 4 kc. and an IF filter pass-band from 22.0 to 22.065 kc., a 20 kc. oscillator output would convert the buzz source output to a range from 20.065 and 24.0 kc. (and 16.0 to 19.935 kc.). When the band limited filter output is converted back to the audio range the outputs would be from 2.0 to 2.065 kc., excluding 42.0 to 42.065 kc. by filter 43. However, by modifying the oscillator output to 19 kc. the IF ranges would be 19.065 to 23.0 kc. (and 15.0 to 18.935 kc.) and the audio output would be altered to a range from 3.0 to 3.065 kc. Similarly by a 21.5 kc. oscillator output the audio output would be altered to a range from 500 to 565 cycles. The complementary relation of the two conversions at one location assures that the now limited audio output corresponds to that portion of the buzz source output, and only one set of sidebands need by considered. Unlike most ordinary IF systems in this case any IF amplitude gain is secondary, the primary purpose is to use the high selectivity at a particular (IF) frequency range for providing a similar selectivity at a very different and readily variable (audio) frequency range, thus simulating the audio filtering characteristics of the vocal tract.

The buzz source output is also combined with the hiss source output in buffer 45, shown as an OR gate, and is supplied thru second and third formant controlled networks 31′ and 31″ analogous to that already described except that bandwidths are 85 cycle and 120 cycle respectively for these formants. This combined signal is supplied also thru bandpass filter 48, having a range of about 3400 to 3600 cycles, somewhat analogous to filter 29 in that it involves mainly background sound, but in this case high frequency components. The five modified outputs originating from the buzz source, and in some cases the hiss source, are supplied to an output device or speaker 49, after being combined in buffer 47, shown as an OR gate. The outputs of filters 43 and 43″ are phase reversed as indicated by inverter or NOT symbols where the corresponding leads are connected. In operation the first or second formant alone provides some resemblance to normal speech, but not always intelligible, while the first and second formants together provide an output which is readily intelligible, and the five components combine to provide an output which is reasonably natural in sound.

The necessary frequency and amplitude controls can readily be made voltage sensitive by the use of combinations of existing circuitry as described below. Thus, appropriate complex time-varying functions are required for control of the particular sounds to be produced. These functions normally involve widely variable time intervals within which increasing, constant, or decreasing voltage values are required. Such functions can be most conveniently generated within reasonable limits of accuracy by establishing appropriate levels of current, and integrating to obtain a reasonably smooth voltage function, as indicated by the function generator 17 in the upper part of FIG. 1. While nine functions are required for control of the synthesizer as shown, the rates in many cases change simultaneously in the several functions; thus a single timing chain is used to control the generation of many functions. This timing chain involves a plurality of monostable circuits 51–0 to 51–*n* and a similar bistable circuit 51–*r*, connected by transition circuits 53–0 to 53–*n*. The number of monostable circuits depends very generally on the greatest length of sounds to be synthesized. A start input to the unstable side of circuit 51—0 shifts it to the unstable state causing a temporary output from such side. This circuit returns to its stable state after a typical interval of about 10 milliseconds (ms.). At this time the direct and delayed inputs to the AND gate of transition circuit 53–0 actuates the next monostable circuit 51–1.

In this and further binaries the time is widely variable from around 30 ms. for very short sound intervals of consonants such as, b, d, k, p, t, etc. and the very short vowel sounds of a, e, i, o, u, to around 500 ms. which is more than adequate for very long sound intervals of consonants such as m, r, z and very long vowel sounds including oo, ou, etc. The sequence may continue thru any number of stages, and the transition output 53–*n* of the last stage to be used may actuate a reset bistable circuit 51–*r* for reasons to appear later. Since this is bistable it also requires a START input at the side corresponding to the stable side of circuits 51–0 to 51–*n*. The chain may even become a ring to repeat the same sounds or to allow re-setting of some stages while others are in use.

Besides providing timing the outputs of binaries 51–0 to 51–*n* are also used in various slope control circuit groups 55–0 to 55–*n*, one circuit of each group being used for control of frequency of buzz source F_b, amplitude of buzz source A_b, etc. Output voltages between the sides of each binary stage typically would differ by 10 volts, but for convenience in analysis as binary signals the levels may be designated merely as the usual 0 and 1 units. Each slope control circuit group includes a potentiometer to set a current value, during the unstable interval of its corresponding monostable circuit, suited to use in its corresponding integrator 57 F_b, A_b, etc. Ordinarily a potentiometer is assumed to control voltage, but in this case the integrator operates at very low impedance, thus permitting a linear integration over a substantial interval and also avoiding feedback among the various inputs; therefore the current is the significant parameter. Symbolically the restriction to a particular interval is illustrated by an AND gate in the potentiometer output, controlled by the monostable circuit output to be effective only during the particular interval. Actually the AND gate function would be accomplished by diodes at the potentiometer inputs connected to both outputs of the monostable circuits as will be illustrated below.

Since the potentiometer outputs may be set from +1 unit to −1 unit, when applied to the integrators during appropriate times they will determine the slope of the integrator outputs, either rising or falling. During the preliminary interval, the potentiometer outputs controlled by monostable 51–0 determine the slope and preliminary level reached by the integrators to start the desired sounds; thereafter in each successive interval, the corresponding potentiometers determine the slopes of the linear increments for approximating the complex func-

ions needed to complete the sounds. The functions controlling amplitude normally would be very low at the beginning of any sounds. Those controlling frequency might be of almost any value depending on the particular sounds.

FIG. 2 illustrates some typical functions involved in synthesizing a simple sentence, "Hello, how are you?" The first curve A is representative of the overall amplitude or intensity A*o*; this would not actually be produced by the function generator but is typical of the several separate intensity curves, and therefore useful for illustration. According to the invention such a curve is to be approximated by a series of straight line sections as illustrated in B. The differential of this curve B, analogous to the desired input to corresponding integrators **57**, is shown in C. Curve D represents the actual function required to control the buzz source frequency F*b*; for this sentence the variations in this curve although important are rather small and not particularly helpful to analysis. Curve E represents the function required to control the frequency of the third formant F3; this curve also involves rather small variations and only the straight line approximation has been illustrated by curve F. Curve G represents the function required for controlling the frequency of the second formant F2 and curve J the function required for controlling the frequency of the first formant F1, the two most significant formants for this sentence. In each case the straight line approximation and the derivative are represented by the curves H and I and K and L respectively. The entire preliminary change in the function is shown concentrated in the very brief (10 ms.) preliminary interval $\pm o$ for convenience; the actual sound becomes significant only in the first interval I1. This approximation is noticeable only in formants F1 and F2 for this particular sentence. Since the various straight line sections of each curve are assumed to be generated in separate circuits, each section in full detail would correspond to a separate curve; however to save space separate sections of only curve K are illustrated, and alternated on each of two lines M and N. It will be recognized that the interval lengths would be set at the monostable circuits **51**–1 to **51**–*n* of FIG. 1, while the amplitudes of the differential curves would be set at the various potentiometers in groups **55**–**0** to **55**–*n* according to the particular sentence synthesized. The most important practical application of the invention would require that it be suited to synthesizing sound corresponding to a wide variety of incoming information. In this case both the intervals and amplitudes of the derivatives should be electronically controlled as the information is received as will be illustrated below in connection with FIGS. 5 and 6.

FIGS. 3 and 4 show suitable circuits which may be used in the function generator **17** and synthesizer **19** respectively of FIG. 1. Referring first to the function generator of FIG. 3 the monostable circuit **51**–1 is a substantially conventional binary or trigger circuit having usual common emitter control transistors **61** and **62** and also emitter follower cross coupling and output transistors **63** and **64**. To cause instability in one state the cross coupling from the emitter terminal of the output transistor **64** to the base terminal of the normally "ON" control transistor **62** is capacitive rather than resistive. To adjust the duration of the unstable state this base is forward biased thru a potentiometer **71**, thus allowing change in the RC time constant.

Outputs are taken from the emitter circuits of both coupling transistors thru diodes **73** and **75** having a polarity to conduct only during the unstable state, thus energizing the various potentiometers **77** only during the corresponding intervals. The potentiometers have current limiting resistors **79** and **81** at each terminal, and their outputs are combined in appropriate integrators **57**, shown as operational amplifiers. This type of circuit has very low input impedance, with input voltage substantially constant in spite of current changes; therefore, the several potentiometers connected to the integrator input load the particular potentiometer energized but do not cause any feedback among the several outputs. An additional potentiometer circuit **82** helps to stabilize the level of the integrator when no other signals are applied. The integrator circuit **57** includes a typical common emitter amplifier transistor **83** with emitter follower transistor **85** to provide a low impedance output, together producing an output signal opposite in phase to the input. A large capacitor **87** from output to input provides a negative feedback to avoid substantial input voltage change and to integrate the input current.

Such a circuit is subject to considerable D.C. drift. To assure starting from the same charge on the capacitor for each sound sequence, a reset circuit **89** in parallel with the capacitor includes a Zener diode **91** in the collector circuit of a control transistor **93**. Such a diode has a very stable breakdown voltage under reversed bias. An ordinary diode **95** of very high reverse resistance in the emitter circuit of the control transistor avoids minority carrier current during integration, which would otherwise include this current, in the same manner as the intended signals. The base is back biased during the successive integrations by one output of the bistable reset circuit **51**–*r* then forward biased to reset the integrator. At the same time, a current is supplied to the input terminal from the other output of reset **51**–*r* thru ordinary diode **97**, to restrict operation to the reset period yet assure sufficient current so that the Zener diode operates in the range where its breakdown voltage is stable.

In the case of the reset trigger circuit **51**–*r* the operation is made bistable, but the elements can still be identified by analogy to the circuit **51**–1. The output from the emitter circuit of the cross coupling and output transistor corresponding to the output transistor **63** is positive during operation of **51**–**0** to **51**–*n*, then negative providing forward base bias in transistor **93**. The output from the other emitter circuit thru diode **97** assures ample Zener diode current to hold condenser **87** discharged until a new start pulse actuates **51**–**0** and **51**–*r* to corresponding states. A function generator of this type can readily provide a linear piece-wise approximation to any time domain function with any desired degree of precision, as used, for example, in programming analog computers, control of output display devices in radar sets, etc.

Referring now to FIG. 4 the hiss source **25** includes a Zener diode **101** as the actual source of noise energy, operating at an unstable point on its reverse biased characteristic—the "knee of the curve." This Zener diode is supplied with a suitable source of current and is coupled to the output thru a variable amplifier **27**, closely similar to the amplifier **33**" to be described below.

The variable frequency buzz source **21** and variable amplifier **23** are closely interrelated and therefore are shown in a single group including an astable blocking oscillator transistor **111**, a variable amplitude clipper controlled by an emitter follower transistor **113**, and a further common emitter transistor **115**, driving a 4 kc. 6 pole low pass Butterworth filter. When the exponentially changing voltage on the base capacitor **116** leaves the cutoff region of the transistor **111** the blocking oscillator action is triggered thru the base feedback coil of its transformer **117**. This capacitor is biased thru a resistor **119** from a source of variable potential depending on the desired blocking oscillator frequency but is returned to the cutoff region by the feedback action. The voltage and time constant are arranged to provide pulse repetition frequencies from 60 to 330 cycles per second, overlapping the maximum useful range. The switching from manual to automatic operation in this and other circuits is merely to simplify testing and adjustment. A diode across the

collector coil of transformer **117** protects the transistor from excessive voltages due to transformer back swing. A resistor in the emitter circuit of transistor **111** manually variable from 0 to about 150 ohms serves to vary the pulse width and thus the relative amplitude of the harmonics of the pulse repetition frequency.

The output from the collector of the blocking oscillator is also connected in series thru load resistor **121**, diode **123** and further resistor **125** to ground, with the output of emitter follower transistor **113** connected between the diode and resistor **125**, and the base input of a common emitter transistor **115** connected between the diode and resistor **121**. The emitter follower transistor **113** is controlled by the output of corresponding integrator **57** A*b* to control the amplitude of the output provided from the blocking oscillator thru the clipper action of the diode **123**. The diode clipper works as follows. With no voltage from integrator **57** A*b* the base of transistor **113** and thus the emitter are both held near ground. Since the peak of the output pulse is near ground, diode **123** should not conduct at any time. As the voltage at the base of transistor **113** is made more negative the voltage at its emitter will also decrease and the diode will conduct, clipping the pulse peak, whenever the pulse peak is more positive than the emitter of transistor **113**. Thus with a range of voltage at the buzz amplitude connector of from zero to minus 20, the pulse amplitude will vary from 20 to zero. The resistor **121** is necessary to reduce loading of the blocking oscillator with a resultant decrease in pulse width. The output of common emitter transistor **115** has its collector terminal properly biased and also connected thrus a low pass Butterworth filter **127** to the output terminal. The proper resistance, inductance and capacitance values are shown on the filter elements for operation up to the desired frequency of 4 kc.

The outputs of the hiss source **25** and the buzz source **21** with their amplifiers **27** and **23** are combined in a buffer **45** shown as an OR gate and supplied to the modulator **37''** as well as other circuits.

The carrier frequency oscillator **35''** is shown as an ordinary astable circuit generally similar to the monostable circuit **51–1** except that both cross couplings are capacitive. The frequency of operation of this astable circuit is controlled by using the output of integrator **57** F3 as the base bias of both control transistors **131** and **133**. This oscillator will generate square waves containing a very wide band of frequencies, but separate filtering is not required at this point since the later circuits will make use of only one component, assumed as the fundamental.

The modulator **37''** includes 2 stage emitter follower connected transistors **141** and **143** having an input from the buffer **45**. In the actual modulator stage, the output of the oscillator **35''** thru a resistor **145**, and the output of the prior emitter follower stage thru diode **147**, are combined and applied thru a coupling resistor to the base of a further emitter follower transistor **149**, whose output is coupled to a variable amplifier **33''**. The diode non-linearity provides the usual sum and difference modulation frequencies from the audio and the various carrier components, those from the harmonic components being superfluous.

Variable amplifier **33''** comprises a group of transistors **151**, **153** **155**, and their associated circuits including a further emitter follower output transistor **157**. The input transistor **151** is operated as an emitter follower with a "constant current" type of emitter load. The output at its emitter is directly coupled to the emitter of a common base amplifier transistor **155**. The emitter currents of both transistors are supplied from the collector of a further transistor **153** having a resistor in its emitter circuit and controlled by a variable potential at its base to allow a total current substantially proportional to such base potential. The signal voltage at the base of transistor **151** is coupled to the emitter input of transistor **155** with very little loss in amplitude, due to the high impedance emitter load on transistor **151**. At the same time, the emitter of transistor **155** is driven from an emitter follower stage having a low impedance characteristic. The current source approximates the ideal case of a large emitter supply voltage coupled thru a large resistance without requiring a high voltage power supply, and the use of the emitter follower **151** eliminates the necessity of by-passing the current source with the resulting loss in response time.

The input voltage from the integrator **57** A3 is applied to the base of transistor **153** which is a common emitter type of circuit. Because of the properties of the forward biased emitter base circuit the emitter voltage and current will closely track the base voltage. This property is relatively independent of such factors as transistor type and supply voltage levels. Since the voltage gains of a transistor vary with its emitter current, variation of the voltage at the base of transistor **153** will vary the overall currents and also the amplitude of the output signal on the collector of transistor **155**. The main purpose of this circuit is to provide control of amplitude by the application of a control signal to the base of transistor **153**. However, the collector load on transistor **155** is shown as a tuned circuit in the IF range and will therefore eliminate the effect of harmonics from the carrier oscillator **35''**; further band limiting occurs in the narrow band amplifier **41''**. This same variable gain amplifier is used in element **27** previously referred to, except that the IF tuned circuit load in the collector of transistor **155** is replaced by a mere resistance load since no particular frequency band is emphasized.

The narrow band tuned amplifier **41''** includes a first common emitter stage transistor **161** having a tuned collector circuit coupled to the base of an emitter follower transistor **163** providing the output. The emitter resistor of this transistor is arranged to provide an adjustable feedback to its tuned base circuit which can be adjusted to provide bandwidths of 65 to 120 cycles as previously indicated. Such a circuit is analyzed in some detail in "Highly Selective Bandpass Filters Using Negative Resistances," by M. Kawakami, P. Yanagisawa, H. Shibayama, Active Networks and Feedback System, Polytechnic Press, Polytechnic Institute of Brookly, p. 369. The output is applied to a demodulator circuit **39''** equivalent to the modulator circuit **37''** except that there is only one emitter follower stage in the input section. The losses of coils cause many difficulties in the construction of high selectively bandpass filters, because a given filter's characteristic is determined by the sum of the loss factors of each tuned circuit contained in the filter, regardless of its configuration. These difficulties are removed by the use of negative resistance. That is, it is possible to make some resonant circuit decrements of the filter negative by combining passive elements and negative resistances, so that the remaining circuit decrements can be selected to any realizable values using conventional passive elements. Therefore, highly selective bandpass filters may be realized from usual finite value Q elements by the addition of active elements. It is not unusual to need unrealizable high-Q elements to obtain selectivity and narrow bandwidth. The use of an active element such as the transistor has been proposed to multiply the Q of the resonant circuit by means of regenerative action. This means that a negative resistance realized by the active elements would be used to compensate the loss of the resonant circuit. The present circuit further develops the above ideas. The main purpose of the narrow band tuned amplifier is to provide selectivity and narrow bandwidth.

In FIGS. 1 to 4 a leisurely manual setting of the function generator potentiometers for the time intervals and the nine slopes in each interval has been assumed to synthesize a particular sentence. This corresponds to

the brain portion for providing the complex time domain function instruction to the vocal tract as noted above. For greatest usefulness the system must extend partly into the brain portion providing memory, permitting a simply coded alphabet character corresponding to a written letter or speech phoneme to be converted to the sound form expected in ordinary speech. It will be observed that this involves a rapidly variable substitute for the manual settings of the various potentiometers. This can be accomplished by digital equipment which can take a binary coded symbol similar to a teletype character and from this prescribe the number and length of successive time interval settings and the 9 slope settings in each interval, all corresponding to the particular character. The memory device for each of the 40 common phonemes would have to provide roughly from 4 to 16 levels of slope for each of the 10 adjustments to be made. This can readily be accomplished at very high speeds by fixed matrix memories of a general type known in the art.

FIG. 5 illustrates such a system for serial bit coded phoneme character signals somewhat analogous to teletype but having 6 bits per character which is more than enough for the phoneme alphabet. The incoming serially coded signal actuates timing circuits **201** shown as a synch control for a bit or baud clock circuit and a character clock circuit, both shown as synchronized astable circuits, and is stored in shift register **203**. The baud clock controls the input gate **205** for the register at one part of the cycle and the shift operation at another part of the cycle. When the register is full of serial bits from a single character (and therefore no bits from other characters) the character clock controls output gates **207** to read the register in parallel mode. Actually the information could be processed in serial mode but is more easily explained in parallel mode. Usual efficient teletype operation involves a very uniform character period, which should equal the average phoneme period, but individual phoneme periods would vary over a wide range; therefore it will be convenient to immediately store the parallel shift register outputs in the successive addresses of a suitable memory device **209**, in which the addresses are selected by a simple binary write address counter **211** actuated at the character period. A similar read address counter **213** is used in extracting the phoneme code identification but this must be stepped at an irregular period as will be indicated later. A random access memory has been assumed but the simple successive addressing requirement would permit use of less versatile memories such as a slack tape recorder if desired.

The parallel character data from the memory can be converted by an ordinary binary matrix **215** to identify the several phonemes represented. Except for the memory and the difference in number of bits per character much of the foregoing description of the left side of FIG. 5 corresponds to an ordinary teletype system. In both cases the binary "weight" has negligible significance as to the character represented, but in the present case merely for orderly arrangement the smaller weights are assumed to represent the commonest characters. The memory device is not essential to elementary teletype operation but is often used in coupling separately synchronized systems.

From this point on, the system for a time becomes less methodical and more empirical in nature depending on the desired degree of perfection in synthesizing the actual speech output. The phoneme alphabet of about 40 characters leaves the other 24 combinations ($2^6=64$) to be used for supplemental information if desired, such as personal characteristics of voices, stenography-like word signs for further economy over separate phonemes, etc.

Since the various phonemes are made up of variable numbers of intervals, within which the functions have reasonably uniform rates, the phoneme outputs of the matrix are each transmitted thru AND gates of one or more gate groups **221**–1, –2, –3 controlled in succession by a modified shift register circuit **251**. For example, the "zero" binary output assumed to correspond to a *pause* between words would require only a single interval, while the "one" binary output assumed to correspond to short i (the most common phoneme sound in English) is assumed to be made up of three intervals *i*1, *i*2 and *i*3 and therefore has three such gates controlled by the register **251**. Furthermore, the transition from one phoneme to another may require that the first interval be arranged to provide different parameters depending on the next previous phoneme. This is represented by further gates **225** in the circuit of output *i*1, controlled thru delay networks **227** to provide a different operation depending on the prior phoneme. At this point it will be seen that the original coded phonemes have been identified and also separated into various intervals.

The necessary parameters of these many possible intervals can best be provided through a suitable empirical matrix system **231** to permit storing in binary form the somewhat empirical information to any desired precision. The single first column of the matrix represents merely the final interval of the particular character as explained below. In the case of the pause only the duration of the single interval is of interest and may be provided by the matrix. In all or most of the other intervals both duration and various other parameters relating to the frequency and amplitude are significant and would be set up in advance on the matrix. As each interval line into the matrix is energized the corresponding stored data is supplied to the various digital to analog converters **255**, which provide output levels analogous to the potentiometers **55** of FIG. 1 controlling integrating circuits **57** also as in FIG. 1. The converters **255** represent a return to more methodical modes of operation after the empirical modes involved in gate groups **221** and **225** and matrix **231**.

The integrator **57**–T*i* determines the time interval by actuating a level responsive circuit shown as a Schmitt Trigger **259**, which resets the same integrator ready for the next time interval and also advances the shift register **251** in a manner determined by the AND and INHibit gates **261** and **263** depending on the "last interval" reading in the first column of matrix **231**. If the advance pulse passes gate **261** it sets the register to state **1** and advances the read address counter **213**, while if it passes gate **263** it provides merely a normal shift operation in the register. This register is generally conventional but slightly simplified since it needs to provide only one output at a time and to reset to output **1** for each new character, otherwise progressing through outputs **1**, **2**, **3** as each interval is completed.

The other integrators need be reset only between words, that is, only at a pause, as indicated by the connection from the "pause" input lead of matrix **231** to the reset inputs of integrators **57** F_b, A_b, etc.

In order to utilize the benefits of the memory **209** without overrunning its capacity, the actual content at any time may be used to modify the usual timing as controlled by the matrix **231** and converter **255**–T*i* thru integrator **57**–T*i*. Since the other function generator outputs controlled by the matrix are also based on integration this change in timing would tend to alter the integrator outputs. Therefore the difference between the setting of the write and read address counters **211** and **213** and even the counter rates are combined in rate servo **261** to produce a variable D.C. control voltage which is supplied to all the digital to analog converters to correct the output values accordingly. In most converters the output would be either positive or negative to provide rise or fall in the function generator outputs; for this purpose the servo varies both positive and negative supply voltages for the converters. However, the converter **255**–T*i* need only provide timing of each interval, and therefore requires only one input polarity.

The matrix **231** would ordinarily be made up of a great many diodes in an empirical pattern to suit the

desired sound synthesis. It is illustrated with 3 bits per function although 2 or 4 bits might be desirable for some functions. FIG. 6 shows suitable digital to analog converters in various degrees of detail. In the first group 255–T*i*, with one variable D.C. input from the rate servo, the gate symbols indicate which of the resistors 271–1, –2, or –4 is merely grounded and which supplies a current to the corresponding integrator 57–T*i* depending on inputs on leads 273–1, –2, or –4 from matrix 231. Actually the gates are not essential and resistors could be connected directly to the matrix. In the second group 255–F*b* the gate functions are provided by the transistors 275–1, –2, –4 for both polarities needed in this group, but resistor functions are the same; it would not be economical to expand the already complex matrix and eliminate the very simple gates. In both these groups the weight designations have actual significance to evaluate the converter outputs and the corresponding slopes in the functions produced by the integrators. The 3 bits per function leads to eight ($2^3=8$) possible levels, the sum of all possible combinations of the weights 1, 2, and 4 indicated by the converter inputs; the difference in appearance of the output weighting resistors is to emphasize the relative currents to be combined in integrators 57.

It will be recognized that the circuits of FIG. 5 including the slight details in FIG. 6 provide an alternative function generator portion to replace the upper part 17 of FIG. 1, as partially detailed in FIG. 3, but in the case of FIG. 5 the system is arranged for general use rather than merely to generate a particular sentence. The code type of transmission permits further advantages of noise rejection, encryption, etc. The non-uniform write and read periods are also involved in a Variable Length Code Method and System of Leo H. Wagner application Ser. No. 208,134, now Patent No. 3,156,768, Nov. 10, 1964, for further channel economy; both techniques could be used in a single system with further storage economy since the same storage could serve both purposes. The sound synthesizer portion to be used with FIG. 5 corresponds to the lower portion of FIG. 1 as detailed in FIG. 4.

The invention has been illustrated in its elementary form and also applied to a typical system. Many other uses of the circuits and further variations in the system of the invention will be apparent to those skilled in the art.

What is claimed is:

1. A speech synthesizer comprising a first source of voiced sound energy, producing various harmonics and controllable both in amplitude and fundamental frequency, and a second source of unvoiced sound energy producing a wide range of substantially random sound frequency components, applied to the base of a first emitter follower connected transistor, a second transistor having an emitter load and an amplitude control signal applied to its base, and a third common base connected transistor providing a collector output circuit, the collector of said second transistor being connected only to the emitters of both said other transistors as a substantially constant current emitter load for said first transistor, controllable to vary the emitter current of said third transistor and its gain, and a filter system comprising a small number of channels, some supplied from one and some from both said sources, each channel having a very limited audio frequency bandwidth, controllable both in amplitude and in the particular audi frequency range, having a fixed band filter in anoth frequency range, a modulator input to and demodulat output from said filter, and a common carrier oscillat for both said modulator and demodulator, controllab thru a range to convert the desired audio-pass-band t and from the fixed band of said filter, a multiple tim domain function generator system having means to pr duce a series of variable length intervals, means to pr duce in each said interval variable parameters corr sponding to the slopes of the desired multiple functio during said intervals, means to combine said parameter for corresponding functions in sequence to produce cor tinuous functions of substantially rectangular steps, an integrating means to produce from each said rectangula stepped function a substantial approximation to sai desired functions, said desired functions controlling sai source and channel amplitudes and frequencies in a man ner corresponding to the principal components of nat ural speech, and means to provide a common combine output of the several channels as synthesized speech cor responding to said variable length and other parameters

2. A speech synthesizer comprising

a first source of voiced sound energy, producing vari ous harmonics and controllable both in amplitud and fundamental frequency,

a second source of unvoiced sound energy producin a wide range of substantially random sound fre quency components and controllable in amplitude

a filter system comprising a small number of chan nels, each of very limited audio frequency band width, some supplied from one and some from bot said sources, and controllable both in amplitude an in the particular audio frequency range,

a multiple time domain function generator system pro viding a succession of intervals, each of variable length, and providing variable parameters within each said interval for each of a plurality of func tions,

said functions controlling said source and channel amplitudes and frequencies in a manner corresponding to the principal components of natural speech,

and means to provide a common combined output of the several channels as synthesized speech corresponding to said variable length and other parameters.

3. A time domain function generator for producing a substantial approximation to a desired function in a series of linear sections, comprising

means to produce a series of variable length intervals corresponding to said linear sections,

means to produce in each interval a variable parameter corresponding to the slope of said linear sections,

means to combine said parameters in sequence to produce a continuous function of substantially rectangular steps,

and integrating means to produce from said rectangular stepped function a substantial approximation to said desired function.

No references cited.

KATHLEEN H. CLAFFY, *Primary Examiner*,

R. MURRAY, *Assistant Examiner*.