

[72] Inventor Joseph F. Schanne  
Cheltenham, Pa.  
[21] Appl. No. 708,323  
[22] Filed Feb. 26, 1968  
[45] Patented Apr. 20, 1971  
[73] Assignee RCA Corporation

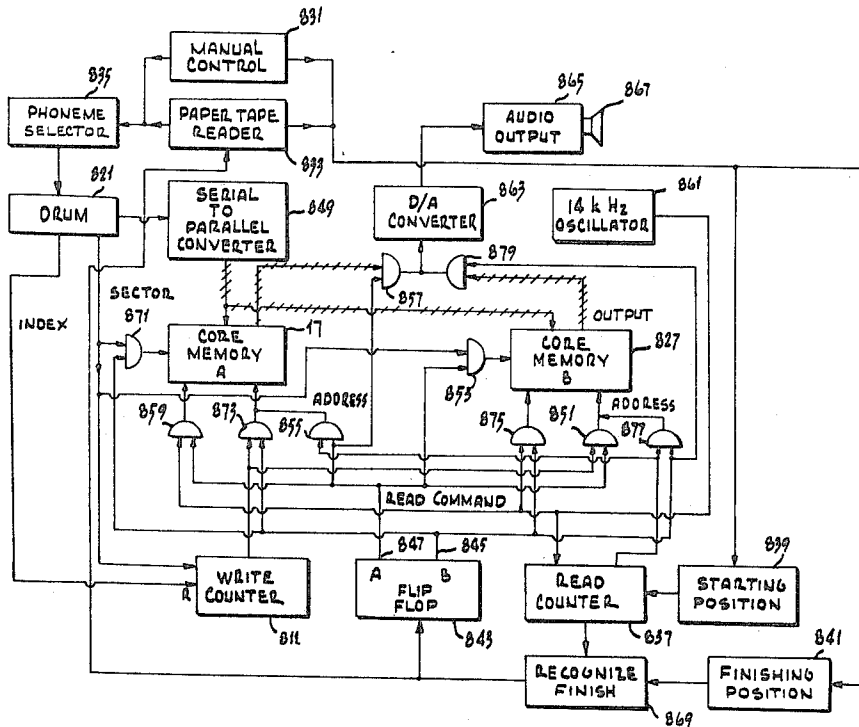
Primary Examiner—William C. Cooper  
Assistant Examiner—Jon Bradford Leahey  
Attorney—H. Christoffersen

[54] **SPEECH SYNTHESIZER PROVIDING SMOOTH TRANSITION BETWEEN ADJACENT PHONEMES**  
6 Claims, 8 Drawing Figs.

[52] U.S. Cl. .... 179/1  
[51] Int. Cl. .... G101 1/02  
[50] Field of Search ..... 179/1 (AS);  
340/15 (2)

[56] **References Cited**  
**UNITED STATES PATENTS**  
3,286,235 11/1966 Sinn ..... 340/152  
3,319,002 5/1967 Clerk et al. .... 179/1(AS)  
3,344,233 9/1967 Tufts ..... 179/1(AS)

**ABSTRACT:** An apparatus for synthesizing speech from phonemes is described. The smooth transition of one phoneme into the next is accomplished by a timewise truncation of the end of the leading phoneme and of the beginning of the following phoneme so that the formants are continuous over the junction. The apparatus described stores phonemes in digital fashion to permit their retrieval, starting and stopping the retrieval so that the desired truncation is achieved. A drum is used to store all the phonemes required and delays between phonemes are prevented by using core memories as temporary storage, transferring from the drum to one of the core memories while concurrently extracting the preceding phoneme from another core memory to be converted to sound.



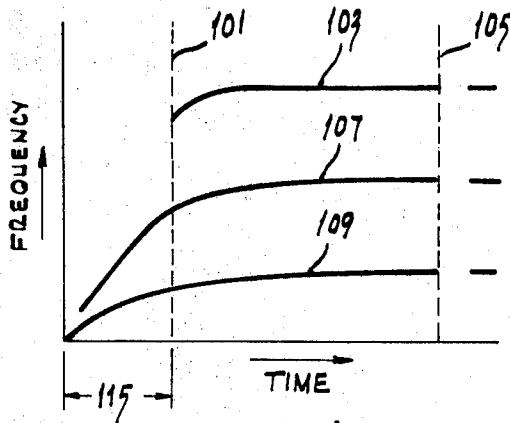


Fig. 1

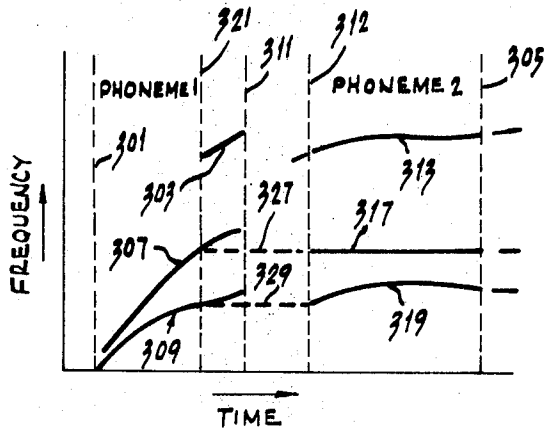


Fig. 3

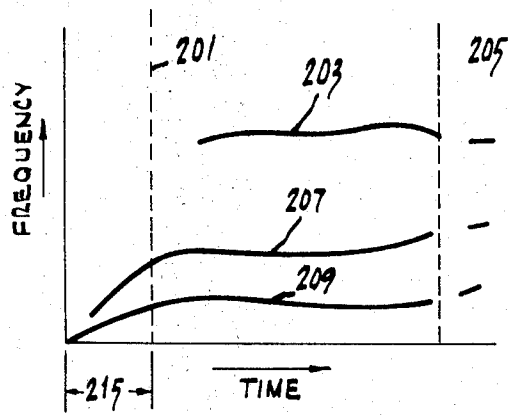


Fig. 2

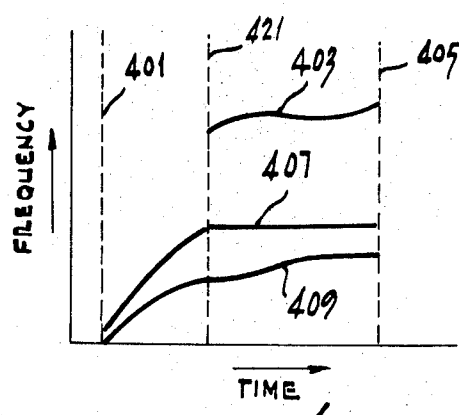


Fig. 4

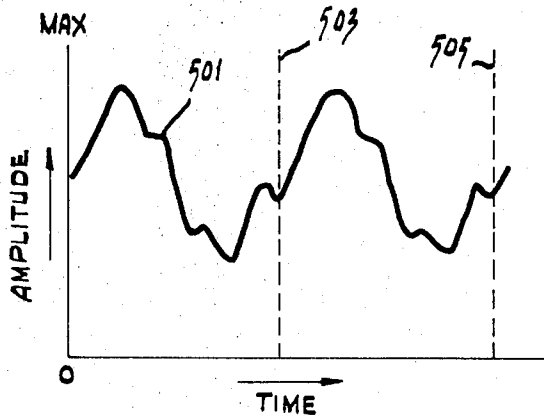


Fig. 5

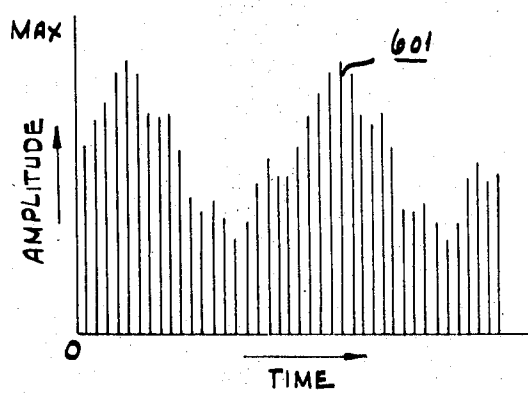


Fig. 6

INVENTOR  
JOSEPH F. SCHANNE  
BY *H. Christofferson*  
ATTORNEY

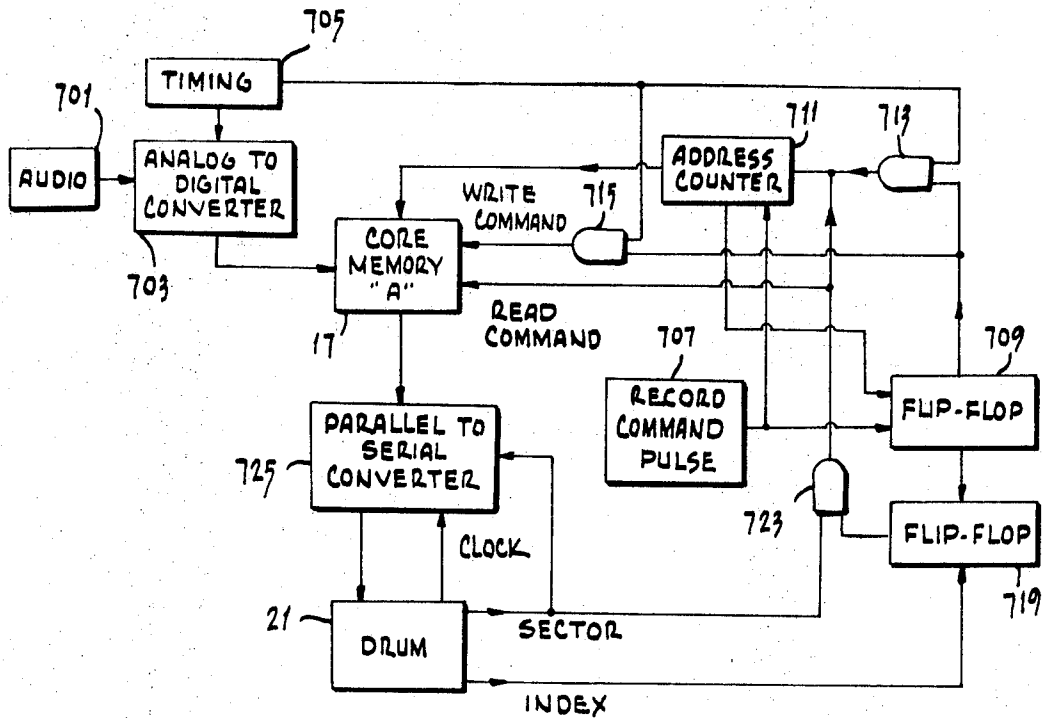


Fig. 7

INVENTOR  
**JOSEPH F. SCHANNE**  
BY *H. Ch...*  
ATTORNEY



### SPEECH SYNTHESIZER PROVIDING SMOOTH TRANSITION BETWEEN ADJACENT PHONEMES

The invention herein described was made in the course of or under a contract or subcontract thereunder with the Department of the Air Force.

The invention herein described was made in the course of or under a contract or subcontract thereunder with the Department of the Air Force.

#### CROSS-REFERENCES TO RELATED APPLICATIONS

A patent application, Ser. No. 708,389, titled **SPEECH SYNTHESIZER**, filed by Thomas B. Martin concurrently herewith and assigned to the assignee of this application contains related subject matter. FIGS. 1 through 6 are the same in both applications.

#### BACKGROUND OF THE INVENTION

Speech is a series of complex sounds generated and controlled by the larynx, tongue, oral and nasal cavities, and force of the breath. The abilities of persons to speak and to understand one another are acquired characteristics that tend to mask the implicit complications involved. The synthesis of speech by means other than human must take into account all the factors, however insignificant, that comprise understandable spoken words.

The recording of speech, as well as music, is usually done in an analog fashion. That is, the continuous changes in amplitude and frequency are maintained upon the storage medium. The reproduction of speech can be effected by reconvertng the recorded signals into audible sound.

In the synthesis of speech, more than mere reproduction is desired. The objective of synthesized speech is the conversion of abstract facts or stored concepts into understandable speech to communicate the said facts or concepts to persons who want to know what they are.

There are many methods of accomplishing this desired result. The most obvious is to record all the sentences possible within the framework of all the facts that the user might desire or require. For even a small number of facts, however, the storage requirements for all permutations and combinations of the facts involved becomes prohibitive.

An approach to reducing the required amount of storage is to store phrases instead of sentences. The storage required is still very large for only a few facts. A further reduction is possible by storing words and combining them, under suitable control, into sentences. This has been done but results in a limited vocabulary. The same problems are encountered using syllables.

The most successful approach compatible with a large vocabulary without a prohibitively large storage requirement has been to use the basic speech unit, the phoneme.

A phoneme is a group of like or related sounds, varying under different phonetic conditions. Forty phonemes are involved in speaking English and they can be categorized into seven groups.

The first three groups comprise the vowel sounds. The first group consists of 10 simple vowels; the second, the six complex vowels; and the third, the four semivowels and liquids.

The fourth group is the six plosives, or explosive sounds.

The fifth consists of the three nasal consonants.

The sixth group is comprised of nine fricatives or spirants, characterized by frictional rustling of the breath against some part of the oral passage as it is emitted.

The seventh group consists of two affricatives. These are a stop or explosive sound followed by a slow separation of the articulating organs, so that the last part is a fricative, or spirant, with corresponding organic position.

Table I (below) lists the phonemes by group as described above. Each of the phonemes is illustrated in a simple comprehensive work indicating by the usual pronunciation the sound of the phoneme, which is underlined for identification.

### TABLE I

Elementary Sounds (Phonemes) of the English Language

TABLE I.—ELEMENTARY SOUNDS (PHONEMES) OF THE ENGLISH LANGUAGE

I.—Simple vowels:	IV.—Plosives:
(1) <u>fit</u>	(1) <u>bad</u>
(2) <u>feet</u>	(2) <u>drive</u>
(3) <u>let</u>	(3) <u>give</u>
(4) <u>bat</u>	(4) <u>pot</u>
(5) <u>but</u>	(5) <u>toy</u>
(6) <u>not</u>	(6) <u>cat</u>
(7) <u>law</u>	
(8) <u>book</u>	V.—Nasal consonants:
(9) <u>boot</u>	(1) <u>may</u>
(10) <u>bird</u>	(2) <u>now</u>
	(3) <u>sing</u>
II.—Complex vowels:	VI.—Fricatives:
(1) <u>pain</u>	(1) <u>zero</u>
(2) <u>go</u>	(2) <u>vision</u>
(3) <u>house</u>	(3) <u>very</u>
(4) <u>ice</u>	(4) <u>that</u>
(5) <u>boy</u>	(5) <u>hat</u>
(6) <u>few</u>	(6) <u>fat</u>
III.—Semi-vowels and liquids:	(7) <u>thing</u>
(1) <u>you</u>	(8) <u>shed</u>
(2) <u>we</u>	(9) <u>sat</u>
(3) <u>late</u>	VIII.—Affricatives:
(4) <u>rate</u>	(1) <u>church</u>
	(2) <u>judge</u>

It is not enough, however, merely to reproduce a sequence of recorded phonemes to produce artificial or synthesized speech. Three conditions must be met in the production of natural sounding synthetic speech from phonemes, viz:

1. there must be continuity in the speech waveform at the junction of phonemes;

2. there must be continuity in the pitch periods across the phoneme boundaries; and

3. there must be continuity of the constituent frequency components between phonemes.

The constituent frequencies of a phoneme can be considered as the dominant frequencies called formants. It is well known that any complex periodic waveform can be synthesized by a combination of sine waves of proper frequencies, amplitudes, and phase relations. The characteristic sound of a phoneme can be reproduced recognizably by the combination of no more than three formants, each of which may or may not vary with respect to time.

The synthesis of speech from phonemes requires, therefore, selecting the proper phoneme sequence and merging the formants of each at their junction points so that there are no discontinuities in the resulting speech.

Some of the approaches to providing smooth transition between phonemes have been described by Dudley et al. in Pat. No. 2,771,509; by David et al. in 2,860,187; and by Gerstman et al. in 3,158,685.

In this prior art, transitions between phonemes are provided by special circuits that provide the required continuities across the junction or store several forms of each phoneme so that the proper one could be selected to provide the continuity at the junction.

An object of this invention is to provide artificial, or synthesized, speech of improved quality and requiring as little storage of sound as possible for an unrestricted vocabulary.

Another object of this invention is to produce synthetic speech in response to control signals that determine the information to be transmitted.

A further object of this invention is the transmission of speech by means of pulses to reduce the bandwidth requirements.

Another object of this invention is to provide means for converting the output data of electronic computers or other control devices into understandable speech.

#### BRIEF SUMMARY OF THE INVENTION

A speech synthesizer embodying the invention includes a first storage for storing phoneme signals required to produce speech, transferring and selecting means to extract a predetermined sequence of phonemes from the first storage to one of more than one second storages, means for extracting the phonemes from the second storages in order, and for converting them to audible sound. The second storages are provided with means whereby the first and last locations of the phoneme signals extracted can be varied in response to control signals. Furthermore, provision is made to load the first storage with phoneme signals.

#### BRIEF DESCRIPTION OF THE DRAWING

FIG. 1 represents the approximate variations of the formants in the spoken word "WED;"

FIG. 2 represents the approximate variations of the formants in the spoken word "WADE;"

FIG. 3 illustrates how a junction to provide continuity of formants between phonemes is determined;

FIG. 4 shows the resultant formants of two phonemes from FIG. 3 joined as illustrated;

FIG. 5 represents two periods of a typical periodic complex waveform involved in speech;

FIG. 6 shows the pulses resulting when the waveform of FIG. 5 is sampled at periodic intervals;

FIG. 7 is a block diagram of an embodiment of the present invention for loading a drum with phoneme signals; and

FIG. 8 is a block diagram of an embodiment of the present invention.

#### DETAILED DESCRIPTION

FIG. 1 shows the formants 103, 107, and 109 for the spoken word "WED" as they might appear on a spectrogram with solid lines depicting the midpoint of the bands of frequencies present. For instance, the lowest frequency formants 107 and 109 between the origin and the ordinate 101 constitute the /W/ phoneme which, in the word "WED," is shown to be made up of two frequencies, both of which increase with time during the time frame 115. The vowel sound of the /EH/ phoneme is composed of three formants 103, 107, and 109 between the ordinates represented by the dashed lines 101 and 105. The final consonant /D/ occurs after a short pause at the end of the vowel sound. For smooth, intelligible speech, the formants of each phoneme must be continuous with those of the following and preceding phonemes across their junctions. The ordinate 101 in FIG. 1 represents one such junction between the /W/ and /EH/ phonemes; the formants 107 and 109 blend smoothly together and are continuous across the junction 101.

FIG. 2 is a similar representation of the spectrogram of the spoken word "WADE." The /W/ phoneme consists of the lower two formants 207 and 209 in the time frame 215 which is delineated by the origin and the ordinate represented by the dashed line 201. The /AY/ phoneme consists of the three formants 203, 207 and 209 between the dashed lines 201 and 205. The formants of the /W/ and /AY/ phonemes blend smoothly at the junction therebetween represented by the ordinate 201.

Comparing FIGS. 1 and 2, the /W/ phoneme in the word "WED" occupies a time frame 115 that is longer in duration than that 215 of the /W/ phoneme in the word "WADE." Also, in FIG. 1 the lowest two formants 107 and 109 of the phoneme /EH/ are lower respectively than the lowest two formants 207 and 209 of the phoneme /AY/ in FIG. 2. The

/W/ phoneme formants in the word "WADE" in FIG. 2 are similar to the /W/ phoneme formants in the word "WED" in FIG. 1 over the same period of time. The outstanding difference between the two /W/ phonemes is that of FIG. 2 is truncated at an earlier point in time than the /W/ phoneme in FIG. 1.

FIG. 3 shows two phonemes not joined, but rather separated by some time interval. The sound depicted in FIG. 3 would be two complete phonemes spoken separately and distinctively.

If the two phonemes depicted in FIG. 3 are to be joined as part of speech synthesis, it is obvious that moving the terminating point in time 311 of the first phoneme into coincidence with the beginning point in time 312 of the second phoneme would result in discontinuities at the junction line so formed. The formants 303 of the first phoneme would end abruptly and the formants 313 of the second phoneme would immediately begin at different frequencies from those of the first. These abrupt changes in frequencies would result in distortions that would destroy the intelligibility of the speech being synthesized.

By extending lines 327 and 329 from the formant beginnings of the second phoneme, it can be noted in FIG. 3 that such lines will intersect the formants 307 and 309 respectively of the first phoneme. The point in time 321 determined by the intersection is a point at which the first phoneme can be truncated for a smooth transition of the formants from the first phoneme into those of the second phoneme. If the aforedescribed intersections do not occur at each formant of the first phoneme at the same time, the beginning of the second phoneme is changed so that they do. A slight amount of discontinuity is permissible so that the point in time at which the intersections occur need not be exactly the same.

If the first phoneme in FIG. 3 is truncated at the point in time 321 depicted by the described intersection and this point in time is made to coincide with the beginning point in time 312 of the second phoneme, a junction is formed across which the formants are continuous. This is shown in FIG. 4 wherein the formants 407 and 409 are continuous across the junction 421, the combined phonemes starting at the same point in time 401 as that 301 of the first phoneme in FIG. 3 and the end of the combined phonemes occurring at a point in time 405 earlier than that 305 of the second phoneme in FIG. 3.

Another method of implementing this technique is to store individual phonemes in a manner that permits selected phonemes to be retrieved, truncated, and reproduced in a sequence previously determined, as, for example, by a control device such as a computer, to synthesize a desired speech pattern. In the described embodiment of this invention, the phonemes are stored digitally by taking periodic samples of the amplitude of the wave shape of each phoneme and converting the magnitudes into binary numbers. The binary numbers obtained are stored in sequence for each phoneme.

FIG. 5 shows two periods of a typical waveform. The line 501 representing the amplitude of the wave as a function of time traces a complex path from the origin to the end of the first period 503. The line 501 then traces a similar path to the end of the second period 505. If the amplitude of such a wave is measured at periodically occurring points in time that occur many times during the period of the wave being sampled, a series of numbers will result that will permit a close approximation of the original wave to be produced by generating individual amplitudes, as determined by the series, at intervals of time that are the same as those at which the measurements were taken. The more samples that are taken during a wave period, the more accurate the reproduction will be.

FIG. 6 is an example of a sample that could be taken of the waveform depicted in FIG. 5. Each of the amplitude plots 601 represents the instantaneous value of the continuously varying amplitude of the line 501 of FIG. 5 at a corresponding point in time.

To demonstrate one method, by way of example, employing such a technique, refer to FIGS. 7 and 8. The loading mode will be described first.

In FIG. 7, the audio signal 701 provides one of the inputs to an analog-to-digital (A/D) converter 703. The other input to the A/D converter 703 is derived from the timing signal source 705 so that the output of the A/D converter represents the instantaneous amplitude of the audio signal at the time of this input. By way of example, the amplitudes of the pulses can be divided into 128 divisions. Each magnitude can then be represented by a binary number of seven bits from the minimum value (0000000) to the maximum value (1111111). The AC zero level is at 64 (1000000). (The reference level is actually taken as approximately 5 percent off center. The direction depends on the number of inversions through the amplifier. The reason for this offset is that the amplitude of the sound waves caused by the expulsion of breath is greater than that caused by the actions of the muscles in the larynx.)

On receipt of record command pulse 707, a flip-flop 709 is triggered to enable the signals from the timing source 705 (1) to advance a triggerable address counter 711 through an enabling gate 713, and (2) to permit write command signals to a core memory 17 through an enabling gate 715. Each pulse from the timing source 705 will advance the address counter 711 one location address, permit the output of the A/D converter 703 to be sent to core memory 17, and produce a write command pulse via gate 715 that causes the output of the A/D converter 703 to be transferred into the core memory 17 at a location specified by the address counter 711.

Successive digitized signals that compose the phoneme being stored are thereby stored in the core memory 17 starting at the lowest address of the memory 17. When the address counter 711 has been advanced to the highest address of the memory, indicating that the memory capacity has been reached, the flip-flop 709 is triggered again. The gate 713 is thereby disabled, inhibiting further advancement of the address counter 711. The gate 715 is also disabled, inhibiting write commands to the core memory 17. The state of the flip-flop 709 after the second triggering described enables the triggering circuit of a second flip-flop 719. The second flip-flop 719 is triggered after being enabled by the first index pulse received from a storage drum 21. The triggering of the second flip-flop 719 enables a gate 723, the other input of which is provided by a sector timing signal from the drum 21. The sector timing signal from the drum 21 occurs once for each of the digitized signals to be stored thereon. Each of the digitized signals stored in the core memory 17 consists of seven binary digits in the embodiment being described. The seven binary bits of each signal are transferred into and out of the core memory 17 in parallel, i.e., simultaneously. Storage on the drum 21 of the seven binary bits of each signal is performed serially, i.e., in consecutive order. The sector timing signal from the drum 21 causes the address counter 711 to advance one memory location and provides a control signal (read command) to cause the core memory 17 to transfer a digitized signal to a parallel-to-serial converter 725. These two functions of the sector timing signal are accomplished only when the gate 723 has been enabled by the second flip-flop 719. Another function of the sector timing pulse is to gate the read output of the core memory 17 into the parallel-to-serial converter 725.

The parallel-to-serial converter 725 is merely a seven-stage shift register into which the output of the core memory 17 is gated in parallel and the output of which is the result of shifting each successive stage into the last stage from where the output is taken. A clock timing pulse from the drum 21 occurs once for each bit to be transferred into the drum 21 from the parallel-to-serial converter 725. For each digitized signal extracted from the core memory 17, seven clock timing pulses from the drum 21 are required to store the seven binary bits in serial fashion on the drum 21. Furthermore, for every seven clock timing pulses that occur, one sector timing pulse occurs.

The address counter 711 used in the embodiment of the invention being described is based on modulo 4096. That is, when the contents of the address counter 711 is advanced to 4095 (in binary digits, 11111111111), the next advance resets the counter to zero (in binary digits, 00000000000). The loading of the core memory 17 is complete when the address counter 711 contents are advanced to 4095. The first sector timing pulse from the gate 723 causes the address counter 711 to be advanced to zero so that the extraction of the successive digitized signals begins at the first address of the core memory 17. The number of digitized signals transferred from the core memory 17 to the drum 21 via the parallel-to-serial converter 725 may be less than 4096. It is therefore necessary to reset the address counter 711 by the record command pulse 707 prior to loading the core memory 17.

The transfer from the core memory 17 to the drum 21 continues until another index pulse from the drum 21, signifying the drum has completed a revolution, triggers the second flip-flop 719. The gate 723 is disabled, inhibiting advancement of the address counter and preventing further command signals to the core memory 17.

The digitized signals comprised of seven binary bits each that constitute a phoneme are therefore recorded serially on a track of the drum 21 during one revolution. Additional phonemes are recorded on other tracks of the drum 21 by using other heads distributed axially along the drum surface. In the present embodiment, there are 128 such tracks for data. The index, sector, and clock pulses are each recorded on a separate track. There are seven clock pulses between sector pulses, and approximately 4,000 sector pulses between index pulses, the latter occurring once per revolution. Each track consists of an individual phoneme. During the loading mode, the tracks may be selected manually, as by means of switches. By selecting one of the data heads, the particular phoneme associated therewith can be recorded and later retrieved.

After all the phonemes to be used have been recorded, a specified sequence of phonemes can be extracted from the drum and transferred to one of two core memories alternately during a speech synthesis operation. The phonemes will then be extracted from the core memories in the same order, truncated to produce continuity of speech sounds, and converted to audible sounds. Transfer from the drum to one core memory and extraction from the other core memory for conversion to audible sounds will be accomplished concurrently. The details of how this is accomplished in the present embodiment will now be explained by reference to FIG. 8.

Three numbers are provided for each phoneme to be reproduced. The first designates which data track is to be read from the drum 21, thereby selecting the phoneme. The second number indicates a starting position and the third, a finishing position. The second and third numbers supplied indicating the starting and finishing positions for the phoneme being selected are delayed until the selected phoneme has been extracted from the drum as described below. These numbers are shown in the present embodiment to be supplied manually 831 or by a paper tape reader 833. It is apparent that such numbers could be provided by a complex control device such as a computer. The starting and finishing positions are predetermined so as to truncate each phoneme to blend properly with the preceding and succeeding phonemes respectively. The beginning and ending addresses are selected so that:

1. The value of the binary number is within 5 percent of 64 (1000000);

2. The formants in the phoneme are at frequencies which are contiguous with those in the end of the preceding phoneme with regard to the starting address and with those of the beginning of the following phoneme with regard to the ending address.

The first number is transferred from the manual control 831 or the paper tape reader 833 to a phoneme selector register 835. The second number is transferred to a read counter 837

via an intermediate register 839. The third number is transferred to a holding register 841.

The phoneme associated with the starting and finishing position in the register will have been extracted from the drum 21 and stored in one of the core memories 17 or 827.

Two operations will be performed concurrently. The first transfers a phoneme from the drum 21 to one of the core memories 17 or 827, and the second extracts the phoneme in the other core memory and converts it to audible sound. A flip-flop 843 is provided to designate which core memory is involved in the first operation and which is involved in the second. For purpose of illustration, it will be assumed that the A-output 847 of the flip-flop 843 is true and that the B-output 845 is false. It is immaterial to the operations to be described which output is assumed to be true first.

The first operation of transferring from the drum 21 to one of the core memories will be described. The track to be read from the drum 21 is selected by the phoneme selector register 835. An index pulse from the drum 21 resets a write counter 811 to zero. The binary digits of the phoneme signal from the head selected on the drum 21 are gated by the clock pulse track into a serial-to-parallel converter 849. The parallel output of the converter 849 consists of seven lines to each of the core memories 17 and 827. The true A-output 847 of the flip-flop 843 will cause the binary digits from the converter 849 to be written into core memory B 827 by enabling the gates 851 and 853, associated with writing into core memory B 827. The address at which each character of seven bits is to be written is transmitted to core memory B 827 from the write counter 811 via the enabled gate 851. Every seven clock pulses from the drum 21 will be accompanied in time by a sector pulse, which is transmitted to core memory B 827 by the enabled gate 853 to cause the storage of the seven bits from the converter 849 to occur, and which also advances the write counter 811 by one count. Thus, each successive seven binary digit phoneme signal is transferred from the drum 21 through the converter 849 into the core memory B 827. The write counter 811 "wraps around" at a count of 4095 so that a maximum of 4096 characters can be transferred. When the drum 21 has made a complete revolution, all the characters comprising a complete phoneme will have been transferred to core memory B 827. An index pulse will reset the write counter 811 to zero and, with no change in the flip-flop 843, the same sequence of characters will be transferred again without altering the contents of the core memory B 827.

The second operation of extracting the signals from the other core memory and converting them to audible sound occurs concurrently with the first operation just described, and will now be described.

The true A-output 847 of the flip-flop 843 also enables the gates 855, 857 and 859 associated with reading from the core memory A 17. The address from which the characters are extracted from core memory A 17 is provided by the read counter 837 via the enabled gate 855. The original setting of the read counter 837 was supplied externally from either the manual control 831 or the paper tape reader 833. The timing pulse to cause the read-out (extraction) from the core memory A 17 is supplied by a 14kHz. oscillator 861 via the enabled gate 859. The timing pulse also advances the read counter 837 one location. The output of the core memory A 17 is a seven bit binary character which furnishes the input to a digital-to-analog converter 863 via the enabled gate 857. The output of the digital-to-analog (D/A) converter 863 is a continuously varying electrical signal, the amplitude of which is determined by the digital input. The continually varying output of the D/A converter is amplified by a suitable amplifier 865 and converted to audible sounds by a suitable transducer such as a speaker 867. Successive phonemes are read out from the core memory A 17 until the number in the read counter 837 is equal to the number in the finishing position register 841. The equality is detected by a comparator 869, the output of which triggers the flip-flop 843 and signals the paper tape reader 833 to furnish the second

and third numbers associated with the phoneme just transferred from the drum 21 to the core memory B 827 and to furnish the first number of the next phoneme to be so transferred.

5 The triggering of the flip-flop 843 causes the B-output 845, which was previously false, to become true, and the A-output 847, which was previously true, to become false.

The track on the drum 21 selected by the phoneme selector 835 is read out to the serial-to-parallel converter 849, timed as described above, and the output of the converter 849 is sent to both core memories. The true B-output 845 of the flip-flop 843 enables the gates 871 and 873 associated with writing into core memory A 17. Specifically, the sector timing pulse is supplied via the enabled gate 871 and the address via the enabled gate 873 from the write counter 811. The corresponding gates 853 and 851 of core memory B 827 are now disabled because the A-output 847 of the flip-flop 843 is false. The first operation is therefore performed using the alternate core memory.

20 The second operation is also performed using the other core memory because the true B-output 845 of the flip-flop 843 enables the gates 875, 877 and 879 associated with reading from the core memory B 827. The timing pulse from the oscillator 861 is supplied to the core memory B 827 via the enabled gate 875; the address from the read counter 837, via the enabled gate 877; and the output from the core memory B 827 is transmitted to the input of the digital-to-analog converter 863 via the enabled gate 879. The corresponding gates 855, 859 and 857 associated with the core memory A 17 are disabled because the A-output 847 of the flip-flop 843 is false.

The alternate reading and writing to each core memory continues until all the desired phonemes selected have been converted to audible sound. In the described embodiment, the drum 21 revolves at an angular velocity of 1800 revolutions per minute. On revolution of the drum is necessary to transfer an entire phoneme. It requires therefor approximately 34 milliseconds to transfer a phoneme from the drum to a core memory. In addition, there is a latency period, i.e., waiting time for the index pulse, of almost 34 milliseconds. The read-out frequency from the other core memory is 14kHz., so that a character is retrieved every 71½ microseconds. Therefore, to transfer the phoneme into one core memory requires the amount of time needed to read out approximately 950 characters from the other core memory. The maximum capacity of each core memory is 4096 characters but this amount is never extracted because of truncation at the beginning and end of the phoneme. However, the number of characters extracted will always exceed the minimum required to provide the time to load the other core memory from the drum. The time required to set a new starting position is short enough that no discontinuity of the sound produced is detectable. The output of the digital-to-analog converter 863 is sustained long enough to prevent any minor discontinuity that may otherwise tend to occur.

50 The numbers supplied by the paper tape reader 833 or equivalent control device, such as a computer, are chosen to select the phonemes required in the proper sequence to produce the speech sounds desired and to truncate such phonemes to provide the maximum intelligibility.

Another possible embodiment of this invention employing only one core memory provides for extracting from the drum only that portion of each phoneme that is to be reproduced and storing all extracted phoneme portions serially in a large core memory. Extraction from the drum and transfer to the core memory would begin at the first, or lowest, core memory address. When the last, or highest, core memory address is reached, the transfer begins again at the first address. Extraction from the core memory of the digital signals to be converted to analog signals commences from the first address and proceeds sequentially to the last, at which time extraction would begin at the first address again. Suitable means for a specified number of drum revolutions provides proper timing.



I claim:

1. Apparatus for synthesizing speech comprising:

first storage means for storing a plurality of digitally-coded phonemes;

second and third storage means, each capable of storing at least one digitally-coded phoneme;

control means for selecting a predetermined sequence of phonemes from said first storage means;

transfer means for loading the successive digitally coded phonemes selected by said control means from the first storage means alternately to said second and third storage means;

read-out means operative concurrently with said transfer means for converting the digitally coded phoneme stored in said second storage means into audio signals during the time said third storage means is being loaded with the following phoneme by said transfer means and for converting the digitally coded phoneme stored by said third storage means into audio signals during the time said second storage means is being loaded with the following phoneme by said transfer means; and

addressing means coupled to said read-out means and said control means for causing the read-out means to finish the retrieval of a phoneme from one storage means and to start the retrieval of a phoneme from the other storage means at respective addresses at which the stored digital values represent contiguous audio frequencies.

2. The invention as claimed in claim 1 wherein the addressing means comprises:

address register means for storing and incrementing the address of the storage means whose contents are being retrieved;

start register means coupled to said control means for placing a starting address in said address register means; finish register means coupled to said control means for receiving therefrom the last address from which the contents of the addressed storage means is to be retrieved; and

recognition means responsive to the contents of the address register means and the finish register means for providing a signal to said control means when the last address has been specified.

3. The invention as claimed in claim 2 wherein the memory select means include a triggerable bistable multivibrator which changes state in response to the signal from the recognition means in said addressing means.

4. The invention as claimed in claim 3 wherein the control means includes means for specifying digitally, in sequence, groups of three numbers, each group comprising a first number designating the phoneme to be retrieved, a second number designating the start address, and a third number designating the finish address.

5. The invention as claimed in claim 4 wherein said first storage means stores a plurality of digitally coded phonemes serially; and further including

serial-to-parallel converting means coupled between said first storage means and said transfer means for converting the serially stored digitally-coded phonemes into groups of signals for parallel transfer to either the second or third storage means.

6. The invention as claimed in claim 1 wherein said first storage means comprises a serial memory and said second and third storage means comprises static memories.

35

40

45

50

55

60

65

70

75

UNITED STATES PATENT OFFICE  
CERTIFICATE OF CORRECTION

Patent No. 3,575,555 Dated April 20, 1971

Inventor(s) Joseph F. Schanne

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

Col. 1, lines 6-8	Delete the entire lines.
Col. 1, line 74	"work" should be ---word---
Col. 2, line 1	Delete the entire line.
Col. 2, line 30	"VIII" should be ---VII---
Col. 5, line 51	"l.e." should be ---i.e.---
Col. 8, line 74	after "means for" insert ---inhibi retrieval of a phoneme from the dr for---

Signed and sealed this 9th day of May 1972.

(SEAL)  
Attest:

EDWARD M. FLETCHER, JR.  
Attesting Officer

ROBERT GOTTSCHALK  
Commissioner of Patents