(19) Europäisches Patentamt
European Patent Office
Office européen des brevets

(11) Publication number: **0 107 945**
B1

(12) EUROPEAN PATENT SPECIFICATION

(45) Date of publication of patent specification: **18.03.87**

(51) Int. Cl.⁴: **G 10 L 5/04**

(21) Application number: **83306228.4**

(22) Date of filing: **14.10.83**

(54) Speech synthesizing apparatus.

(30) Priority: **19.10.82 JP 183410/82**

(43) Date of publication of application:
09.05.84 Bulletin 84/19

(45) Publication of the grant of the patent:
18.03.87 Bulletin 87/12

(84) Designated Contracting States:
DE FR GB NL

(56) References cited:
EP-A-0 058 130
DE-A-2 531 006
US-A-3 975 587

ICC '79 CONFERENCE RECORD (INTERNATIONAL CONFERENCE ON COMMUNICATIONS), vol. 3, 10th-14th June 1979, Boston, pages 39.4.1-39.4.5, New York, USA, E. VIVALDA et al.: "Unlimited vocabulary voice response system for Italian"

(73) Proprietor: **Kabushiki Kaisha Toshiba**
72, Horikawa-cho Saiwai-ku
Kawasaki-shi Kanagawa-ken 210 (JP)

(72) Inventor: **Nitta, Tsuneo**
207 Seya-Heights 2446 Seya-cho
Seya-ku Yokohama-shi (JP)
Inventor: **Nomura, Norimasa**
6 Takenomaru
Naka-ku Yokohama-shi (JP)
Inventor: **Sumita, Kazuo**
2-7-505, Kumugidai-Danchi 1404 Kawashima-cho
Hodogaya-ku Yokohama-shi (JP)

(74) Representative: **Freed, Arthur Woolf et al**
MARKS & CLERK 57-60 Lincoln's Inn Fields
London WC2A 3LS (GB)

EP 0 107 945 B1

Courier Press, Leamington Spa, England.

(58) References cited:
ICASSP 81 (IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING), 30th,31st March, 1st April 1981, Atlanta, vol. 1, pages 110-113, IEEE, New York, USA, H. DETTWEILER: "An approach to demisyllable speech synthesis of german words"

1978 IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH & SIGNAL PROCESSING, 10th-12th April 1978, Tulsa, Oklahoma, pages 577-580, IEEE, New York, USA, C.H. SHADLE et al.: "Speech synthesis by linear interpolation of spectral parameters between dyad boundaries"

ASSP-24, No. 6, Dec. 76, p. 446

Description

This invention relates to a speech synthesizing apparatus for synthesizing speech in accordance with input character strings.

Recently, various speech synthesizing apparatuses for synthesizing speech on the basis of the sentence data to be applied as character strings have become known. For example, in an apparatus for synthesizing speech by rule, various speech segments of predetermined units are preliminarily registered as a format of acoustic parameter in a speech segment file, and the corresponding acoustic parameter data is selectively read out from this speech segment file in accordance with the input phoneme data string. The speech data is synthesized on the basis of this acoustic parameter data read out in accordance with a predetermined synthesizing rule. As described above, in this speech synthesizing apparatus, a desired sentence can be generated at a desired speaking speed since the speech is synthesized in accordance with a predetermined synthesizing rule.

This apparatus for synthesizing speech by rule is mainly divided, for example, into a V—C—V synthesizing apparatus using a chain consisting of vowel, consonant and vowel as a speech segment of one unit, and a C—V synthesizing apparatus using a monosyllable consisting of consonant and vowel as a speech segment of one unit in dependence upon the format of the speech segment to be registered in the speech segment file. Reference characters V and C used herein represent a vowel segment and a consonant segment, respectively.

Fig. 1 is a schematic block diagram of a conventional speech synthesizing apparatus. This speech synthesizing apparatus includes a phoneme converting circuit 2 for converting input character code string into phoneme data string including accent information in accordance with predetermined phoneme conversion rule and accent rule, a speech segment file 4 in which a plurality of speech segments in the form of monosyllable have been stored, an interpolating circuit 6 which sequentially reads out the speech characteristic parameter data of the corresponding speech segment from the speech segment file 4 in accordance with the phoneme data string from the phoneme converting circuit 2 and then interpolates these speech characteristic parameter data, and a speech synthesizer circuit 8 for generating speech data by filter-processing the parameter data from this interpolating circuit 6.

In the apparatus for synthesizing speech data by rules of this kind, phonemes must of course be converted with high accuracy to obtain more natural speech with high quality, but it is also required to obtain speech characteristic parameters which represent, with a high fidelity, the characteristics of the speech generated by a human being. For example, when speech is continuously generated, there may be a case where a certain monosyllable in this speech is coarticulated by monosyllables before and after the above-mentioned monosyllable. When a monosyllable formed of consonant-vowel (C1—V1) syllable is independently generated, the acoustic energy pattern (speech characteristic parameter) of the speech segment of this monosyllable exhibits the inherent characteristics of the consonant C1 and vowel V1 with high fidelity as schematically shown in Fig. 2. However, in the case where this monosyllable is successively generated together with other monosyllables, the acoustic energy pattern (speech characteristic parameter) of the speech segment of the C1—V1 monosyllable will be changed as shown in Figs. 3A and 3B in dependence upon, for example, whether the subsequent monosyllable is a C2—V2 syllable or a C3—V3 syllable. In other words, this monosyllable is coarticulated by the subsequent C2—V2 monosyllable and is changed to a C11—V11 monosyllable, or it is coarticulated by the subsequent C3—V3 monosyllable and is changed to a C12—V12 monosyllable. Therefore, in order to generate the speech which is more natural and has high quality and is as similar as possible to the speech that is actually generated by a human being, it is required to generate the speech in consideration of the coarticulation between the successive speech segments. However, with a conventional speech synthesizing apparatus, only unnatural speech is obtained because it generates speech by simply coupling the phonemes regardless of the influence due to the coarticulation.

EP—A—58130 discloses that discreet sound elements corresponding to consonant portions, steady-state vowel portions and transition elements. However, in this prior art, transition elements are composed of a combination of a consonant portion and a coarticulated vowel and it is thus necessary to prepare a large number of such transition elements in order to synthesize natural speech.

It is an object of the present invention to provide a speech synthesizing apparatus for synthesizing clear and natural speech.

According to the invention, there is provided a speech synthesizing apparatus comprising a data generation circuit for generating phoneme string data; memory means in which consonant and vowel characteristic parameter data representative of consonant and vowel segments are stored and which has a consonant segment file in which a plurality of consonant characteristic parameter data representative of a plurality of consonant segments, each of which has a consonant portion and a transient segment to a vowel segment, are stored, and a vowel segment file in which a plurality of vowel characteristic parameter date representative of a plurality of steady-state vowel segments are stored; control means for allowing the corresponding consonant and vowel characteristic parameter data to be generated from said memory means in accordance with said phoneme string data; and synthesizing means for synthesizing a speech signal on the basis of said consonant and vowel characteristic

parameter data from said memory means; and including a parameter data series generation circuit for generating a series of consonant and vowel characteristic parameter data on the basis of the consonant and vowel characteristic parameter data from said consonant and vowel characteristic paramteter data from said consonant and vowel segment files, and a synthesis circuit for synthesizing the speech signal on the basis of the parameter data series from said parameter data series generation circuit, characterized in that said vowel segment file further stores a plurality of vowel characteristic parameter data representative of a plurality of coarticulated vowel segments, each of said steady-state and coarticulated vowel segments being formed of one frame parameter data, said control means generates time length data indicative of a vowel duration length in accordance with the phoneme string data from said data generation circuit, and said parameter data series generation circuit includes a repetition circuit which derives the vowel characteristic parameter data from said vowel segment file the number of times corresponding to said time length data.

In the described embodiment, each consonant characteristic parameter data stored in the consonant segment file represents the consonant segment including a consonant portion and a transient segment to the vowel segment; therefore, it is possible to easily obtain the interpolated characteristic parameter data between this consonant characteristic parameter data and the succeeding vowel characteristic parameter data read out from the vowel segment file, thereby making it possible to clearly and naturally synthesize a speech even for a coarticulated monosyllable.

An embodiment of the invention will now be described, by way of example, with reference to the accompanying drawings, in which:

Fig. 1 is a schematic block diagram of a conventional speech synthesizing apparatus;

Fig. 2 shows the schematic acoustic energy pattern of a monosyllable independently generated;

Figs. 3A and 3B show the schematic acoustic energy pattern of coarticulated monosyllables;

Fig. 4 shows the schematic acoustic energy pattern of consonant and vowel segments registered in consonant and vowel segment files used in this invention;

Figs. 5A and 5B show waveforms of [a]-sound included in different speeches;

Figs. 6A and 6B show power spectra of selected frames in the [a]-sounds shown in Figs. 5A and 5B;

Figs. 7A to 7C show a speech signal, power spectra and power sequence of a monosyllable "go";

Fig. 7D shows similarity between the power spectrum having the maximum power in the power sequence of Fig. 7C and other power spectra;

Fig. 8 is a block diagram of a speech synthesizing apparatus according to one embodiment of this invention;

Fig. 9 shows power spectra obtained in the speech synthesizing apparatus of Fig. 8; and

Fig. 10 is a flowchart illustrating the operation of the speech synthesizing apparatus shown in Fig. 8.

As shown in Fig. 4, consonant segments each including a consonant portion and a transient segment which changes from this consonant portion to a vowel segment are registered as a consonant segment C in the consonant segment file, and vowel segments including steady-state and coarticulated vowel segments are registered as a vowel segment V in the vowel segment file.

Figs. 5A and 5B shows waveforms of a second [a]-sound of speech [hakata] and an [a]-sound of speech [kiai]. Fig. 6A shows a power spectrum in the frame A of [a]-sound shown in Fig. 5A. Fig. 6B shows a power spectrum in the frame B of [a]-sound shown in Fig. 5B. As is obvious from these Figs. 5A, 5B, 6A and 6B, the power spectrum of [a]-sound of [kiai] which is strongly affected due to the coarticulation is different from the power spectrum of the second [a]-sound of speech [hakata] which is not so affected due to the coarticulation. As described above, the speech characteristic parameters representative of the power spectra of different kinds of [a]-sounds are registered in the vowel segment file in dependence upon the degree of the influence due to the coarticulation.

Figs. 7A to 7C show a speech signal, power spectrum and power sequence of a monosyllable "go" when it was generated. Fig. 7D indicates similarity between the power spectrum having the maximum power in the power sequence shown in Fig. 7C and other power spectra. In Fig. 7D, time point t1 is determined as a boundary point between consonant and vowel, that is, in this example, the time point t1 is determined as a time point at which the similarity becomes smaller than a predetermined value when the similarity between the power spectrum having the maximum power and the power spectra which sequentially appear toward the direction in which a consonant was generated is sequentially calculated. The speech characteristic parameter data representing the power spectra generated during the period from the time when the consonant had been generated to the time point t1, in this example, the power spectra of three frames, is registered as a consonant segment data in the consonant segment file. In addition, the speech characteristic parameter data representing the power spectrum of one frame generated after a predetermined number of frames from the time point t1, preferably indicative of the power spectrum having the maximum power is registered as a vowel segment data in the vowel segment file.

The formats of the speech characteristic parameters to be registered in the consonant and vowel segment files are determined in accordance with the speech synthesizing apparatus to be used. For example, in the Formant synthesiz-

ing apparatus, the speech characteristic parameter is determined by the Formant frequency, its band width and voiced-unvoiced information. On the other hand, in the linear prediction synthesizing apparatus, the speech characteristic parameter is determined by the linear prediction coefficient and voiced-unvoiced information.

Fig. 8 shows a block diagram of a speech synthesizing apparatus for synthesizing speech by rule as one embodiment according to the present invention. This speech synthesizing apparatus includes a consonant segment file 10, a vowel segment file 12, a phoneme converting circuit 14, and a control circuit 16 for generating output data such as consonant segment address data, vowel segment address data, pitch data, etc. in response to the output data from the phoneme converting circuit 14. As already described with reference to Fig. 4, a plurality of speech characteristic parameter data respectively representing a plurality of consonant segments each of which has a consonant portion and a transient segment are stored in the consonant segment file 10. A plurality of speech characteristic parameter data respectively representing a plurality of steady-state vowel and coarticulated vowels are stored in the vowel segment file 12. The phoneme converting circuit 14 reads out the corresponding phoneme string data and accent data from a phoneme dictionary and an accent dictionary (not shown) on the basis of the character code string corresponding to word, clause or sentence, and then supplies to the control circuit 16. This phoneme converting circuit 14 is introduced in, for example, "Letter-to-Sound Rules for Automatic Translation of English Text to Phonetics" by Honey S. Elovitz et al. from Naval Research Lab. (ASSP—24, No. 6, Dec 76, p. 446).

The control circuit 16 serves to supply the consonant segment address data and vowel segment address data to the consonant segment file 10 and the vowel segment file 12, respectively, in accordance with the phoneme string data from the phoneme converting circuit 14. At the same time, the control circuit 16 writes the time data corresponding to the time duration of a vowel to be generated and the accent data from the phoneme converting circuit 14 into a random access memory (RAM) 16A. Where the control circuit 16 generates the consonant and vowel segment address data corresponding to the consonant and vowel which are included in a monosyllable supplied from the phoneme converting circuit 14, the segment address data are determined in accordance with not only the phoneme data indicative of the monosyllable, but also the phoneme data representing a succeeding monosyllable from the phoneme converting circuit 14, for example.

The speech characteristic parameter data from the consonant segment file 10 is supplied to a first input port of an interpolation circuit 18, while the speech characteristic parameter data from the vowel segment file 12 is supplied to a second input port of the interpolation circuit 18 and to a repetition circuit 20. The interpolation circuit 18 calculates a predetermined number of speech characteristic parameter data on the basis of the speech characteristic parameter data indicative of the consonant segment which is constituted by the power spectrum of three frames from the consonant segment file 10 and the speech characteristic parameter data indicative of the vowel segment of the power spectrum of one frame from the vowel segment file 12. The calculated speech parameter data respectively represent a corresponding number of vowel segments each having the spectrum of one frame and interpolated between the input consonant and vowel segments. The repetition circuit 20 repeatedly fetches from the vowel segment file 12 the speech characteristic parameter data by the number of frames corresponding to the vowel time duration data stored in the RAM 16A.

The speech characteristic parameter data from the interpolation circuit 18 and repetition circuit 20 are supplied through a switch 24 to a buffer register 22 in this order. The speech characteristic parameter data from this buffer register 22 is supplied to an interpolation circuit 26. This interpolation circuit 26 interpolates a predetermined number of speech characteristic parameter data between these two speech characteristic parameter data on the basis of the speech characteristic parameter data of the successive two frames from the buffer register 22. The speech characteristic parameter data from this interpolation circuit 26 are sequentially supplied to a speech synthesizer 28. This speech synthesizer 28 sequentially filter-processes the speech characteristic parameter data from the interpolation circuit 26 according to the pitch period data generated from a pitch generation circuit 30 in accordance with the accent data of the RAM 16A, and then generates a speech signal.

The operation of the speech synthesizing apparatus shown in Fig. 8 will be described with reference to a power spectrum shown in Fig. 9, and a flowchart shown in Fig. 10.

The phoneme converting circuit 14 supplies the phoneme string data and accent data to the control circuit 16 in accordance with the input character code series. This control circuit 16 writes the time length data representing the time duration of a vowel to be generated and the pitch data regarding a speech generating pitch in the RAM 16A on the basis of the phoneme data and accent data from the phoneme converting circuit 14, respectively. Furthermore, the control circuit 16 supplies the consonant segment address data and vowel segment address data corresponding to the phoneme string data from the phoneme converting circuit 14 to the consonant segment file 10 and the vowel segment file 12, respectively. In this case, the control circuit 16 simultaneously generates the switch control signal to set the switch 24 into the first switching position.

It is now assumed, for example, that the input character code series including the character codes representative of two successive mono-

syllables of [goma] was supplied to the phoneme converting circuit 14. In this case, the control circuit 16 supplies the consonant and vowel segment address data coresponding to consonant segment [g] and vowel segment [o] to the consonant and vowel segment files 10 and 12, respectively, on the basis of the phoneme data corresponding to the two successive monosyllables of [goma] generated from the phoneme converting circuit 14. Due to this, the first to third speech characteristic parameter data corresponding to the power spectra of three frames indicative of consonant segment [g] in Fig. 9 are read out from the consonant segment file 10. The fourth speech characteristic parameter data corresponding to the power spectrum of one frame indicative of vowel [o] is read out from vowel segment file 12. The interpolation circuit 18 calculates the fifth to eighth speech characteristic parameter data indicative of the power spectrum of a predetermined number of frames, in this example, four frames between consonant segment [g] and vowel segment [o] shown in Fig. 9, on the basis of the third speech characteristic parameter data read out from the consonant segment file 10 and the fourth speech characteristic parameter data read out from the vowel segment file 12. Next, this interpolation circuit 18 supplies the 1st to 3rd speech characteristic parameter data from the consonant segment file 10, the 5th to 8th speech characteristic parameter data thus calculated, and the 4th speech characteristic parameter data from the vowel segment file 12 to the buffer register 22 through the switch 24 in this order in response to the interpolation control signal from the control circuit 16.

Thereafter, the switch 24 is set into the second switching position by the switching control signal from the control circuit 16. The control circuit 16 then supplies the control pulses of the number corresponding to the vowel time duration data stored in the RAM 16A to the repetition circuit 20 and through an OR gate 32 to the buffer register 22. Thus, the repetition circuit 20 fetches the speed characteristic parameter data from the vowel segment file 12 a corresponding number of times in response to the control pulse from the control circuit 16, and sequentially supplies to the buffer register 22. In this way, as shown in Fig. 9, the speech characteristic parameter data representing the power spectra similar to the power spectra shown in Fig. 7B is stored in the buffer register 22. In Fig. 9, the power spectra shown by the solid lines indicate the power spectra corresponding to the speech characteristic parameter data read out from the consonant and vowel segment files 10 and 12, and the power spectra shown by the broken lines represent the power spectra calculated by the interpolation circuit 18 and the power spectra generated from the repetition circuit 20.

Next, the control circuit 16 supplies the interpolation control signal through the OR gate 32 to the buffer register 22 and also supplies the interpolation control signal to the interpolation circuit 26, thereby allowing the speech characteristic parameter data in the buffer register 22 to be sequentially sent to the interpolation circuit 26. The interpolation circuit 26 then creates a predetermined number of interpolated speech characteristic parameter data on the basis of the speech characteristic parameter data of the successive two frames sent from the buffer register 22 and sequentially supplies to the speech synthesizer 28. In this case, the control circuit 16 simultaneously reads out the accent data stored in the RAM 16A and supplies to the pitch generation circuit 30, thereby allowing this pitch generation circuit 30 to generate the pitch period data. The speech synthesizer 28 synthesizes the speech signal including the pitch information in accordance with the speech characteristic parameter data from the interpolation circuit 26 and the pitch period data from the pitch generation circuit 30 and then generates the synthesized speech signal.

Although the present invention has been described above with respect to one embodiment, this invention is not limited to only this embodiment. For example, the repetition circuit 20 is constituted in such a manner that it fetches the vowel characteristic parameter data from the vowel segment file 12 in response to the control pulses from the control circuit 16. However, it may be possible to modify this repetition circuit 20 such that a high-level signal is generated from the control circuit 16 over the period of time corresponding to the time length data, and that the repetition circuit 20 fetches the vowel characteristic parameter data at a fixed interval from the vowel segment file 12 in response to this high-level signal. In addition, although a plurality of vowel characteristic parameter data each of which represents one frame power spectrum have been stored in the vowel segment file 12, the vowel characteristic parameter data each of which represents a plurality of power spectra can be stored in this vowel segment file.

Claims

1. A speech synthesizing apparatus comprising: a data generation circuit (14) for generating phoneme string data; memory means (10 and 12) in which consonant and vowel characteristic parameter data representative of consonant and vowel segments are stored and which has a consonant segment file (10) in which a plurality of consonant characteristic parameter data representative of a plurality of consonant segments, each of which has a consonant portion and a transient segment to a vowel segment, are stored, and a vowel segment file (12) in which a plurality of vowel characteristic parameter data representative of a plurality of steady-state vowel segments are stored; control means (16) for allowing the corresponding consonant and vowel characteristic parameter data to be generated from said memory means (10 and 12) in accordance with said phoneme string data; and synth-

esizing means (18, 20, 22, 24, 26, 28 and 30) for synthesizing a speech signal on the basis of said consonant and vowel characteristic parameter data from said memory means (10 and 12); and including a parameter data series generation circuit (18, 20 and 24) for generating a series of consonant and vowel characteristic parameter data on the basis of the consonant and vowel characteristic parameter data from said consonant and vowel characteristic parameter data from said consonant and vowel segment files (10 and 12), and a synthesis circuit (22, 26, 28 and 30) for synthesizing the speech signal on the basis of the parameter data series from said parameter data series generation circuit (18, 20 and 24), characterized in that said vowel segment file (12) further stores a plurality of vowel characteristic parameter data representative of a plurality of coarticulated vowel segments, each of said steady-state and coarticulated vowel segments being formed of one frame parameter data, said control means (16) generates time length data indicative of a vowel duration length in accordance with the phoneme string data from said data generation circuit (14), and said parameter data series generation circuit (18, 20, 24) includes a repetition circuit (20) which derives the vowel characteristic parameter data from said vowel segment file (12) the number of times corresponding to said time length data.

2. A speech synthesizing apparatus according to claim 1, characterized in that said parameter data series generation circuit further includes: an interpolation circuit (18) for calculating a predetermined number of interpolated characteristic parameter data on the basis of the consonant and vowel characteristic parameter data from said consonant and vowel segment files (10 and 12); and a data selection circuit (24) for sequentially and selectively supplying the characteristic parameter data from said interpolation circuit (18) and said repetition circuit (20) to said synthesis circuit (22, 26, 28 and 30).

3. A speech synthesizing apparatus according to claim 2, characterized in that said data selection circuit is a switching circuit (24) whose switching position is controlled in response to a switching control signal from said control means (16).

4. A speech synthesizing apparatus according to claim 2, characterized in that said data generation circuit (14) generates accent data together with said phoneme string data and said control means (16) generates pitch data in accordance with said accent data, and that said synthesis circuit (22, 26, 28 and 30) synthesizes the speech signal on the basis of the parameter data series from said parameter data series generation circuit (18, 20 and 24) and the pitch data from said control means (16).

5. A speech synthesizing apparatus according to claim 2, characterized in that said synthesis circuit comprises: an interpolator (26) which receives the parameter data series from said parameter data series generation circuit (18, 20 and 24) and calculates a predetermined number of interpolated parameter data on the basis of two successive parameter data; and a synthesizing unit (28) for synthesizing the speech signal on the basis of the parameter data from said interpolator (26).

6. A speech synthesizing apparatus according to claim 5, characterized in that said data generation circuit (14) generates accent data together with said phoneme string data and said control means (16) generates pitch data in accordance with said accent data, and that said synthesis circuit (22, 26, 28 and 30) synthesizes the speech signal on the basis of the parameter data series from said parameter data series generation circuit (18, 20 and 24) and the pitch data from said control means (16).

Patentansprüche

1. Einrichtung zur Sprachsynthese, umfassend: eine Datenerzeugungsschaltung (14) zum Erzeugen von Phonemereihen- oder -kettendaten, eine Speichereinrichtung (10 und 12), in welcher für Konsonanten- und Vokalsegmente repräsentative Konsonanten- und Vokalcharakteristik-Parameterdaten gespeichert sind und die eine Konsonantensegmentdatei (10), in welcher eine Vielzahl von für eine Vielzahl von Konsonantensegmenten, die jeweils einen Konsonantenteil und ein Einschwing- oder Übergangssegment zu einem Vokalsegment aufweisen, repräsentativen Konsonantencharakteristik-Parameterdaten gespeichert sind, sowie eine Vokalsegmentdatei (12), in welcher eine Vielzahl von für eine Vielzahl von Einschwing- oder Dauerzustands-Vokalsegmenten repräsentativen Vokalcharakteristik-Parameterdaten gespeichert sind, aufweist, eine Steuereinheit (16), um die entsprechenden Konsonanten- und Vokalcharakteristik-Parameterdaten von der Speichereinrichtung (10 und 12) nach Maßgabe der Phonemekettendaten erzeugen zu lassen, und (eine) Zusammensetzeinrichtung(en) (18, 20, 22, 24, 26, 28 und 30) zum Synthetisieren bzw. Zusammensetzen eines Sprachsignals auf der Grundlage der Konsonanten- und Vokalcharakteristik-Parameterdaten von der Speichereinrichtung (10 und 12), sowie mit einer Parameterdatenreihen-Erzeugungsschaltung (18, 20 und 24) zum Erzeugen einer Reihe von Konsonanten- und Vokalcharakteristik-Parameterdaten auf der Grundlage der Konsonanten- und Vokalcharakteristik-Parameterdaten aus den Konsonanten- und Vokalcharakteristik-parameterdaten von den Konsonanten- und Vokalsegmentdateien (10 und 12) und einer Syntheseschaltung (22, 26, 28 und 30) zum Zusammensetzen des Sprachsignals auf der Grundlage der Parameterdatenreihen von der Parameterdatenreihen-Erzeugungsschaltung (18, 20 und 24), dadurch gekennzeichnet, daß die Vokalsegmentdatei (12) ferner eine Vielzahl von Vokalcharakteristik-Parameterdaten speichert, die für eine Vielzahl von mitartikulierten Vokalsegmenten repräsentativ sind, jedes der Dauerzustand- und mitartikulierten Vokalsegmente aus

Einfeld-Parameterdaten gebildet ist, die Steuereinheit (16) Zeitlängendaten, welche eine Vokaldauer angeben, nach Maßgabe der Phonemekettendaten von der Datenerzeugungsschaltung (14) erzeugt und die Parameterdatenreihen-Erzeugungsschaltung (18, 20, 24) einen Wiederholungskreis (20) aufweist die Vokalcharakteristik-Parameterdaten von der Vokalsegmentdatei (12) mit einer Häufigkeitszahl entsprechend den Zeitlängendaten ableitet.

2. Einrichtung zur Sprachsynthese nach Anspruch 1, dadurch gekennzeichnet, daß die Parameterdatenreihen-Erzeugungsschaltung weiterhin aufweist: einen Interpolationskreis (18) zum Berechnen einer vorbestimmten Zahl von interpolierten Charakteristikparameterdaten auf der Grundlage der Konsonanten- und Vokalcharakteristik-Parameterdaten von den Konsonanten- und Vokalsegmentdateien (10 und 12) sowie einen Datenwählkreis (24) zum sequentiellen und selektiven Liefern der Charakteristikparameterdaten vom Interpolationskreis (18) und vom Wiederholungskreis (20) zur Syntheseschaltung (22, 26, 28 und 30).

3. Einrichtung zur Sprachsynthese nach Anspruch 2, dadurch gekennzeichnet, daß der Datenwählkreis ein (Um-)Schaltkreis (24) ist, dessen Schaltstellung in Abhängigkeit von einem Schaltsteuersignal von der Steuereinheit (16) steuerbar ist.

4. Einrichtung zur Sprachsynthese nach Anspruch 2, dadurch gekennzeichnet, daß die Datenerzeugungsschaltung (14) Akzentdaten zusammen mit den Phonemekettendaten und die Steuereinheit (16) Tonlagendaten nach Maßgabe der Akzentdaten erzeugen und daß die Syntheseschaltung (22, 26, 28 und 30) das Sprachsignal auf der Grundlage der Parameterdatenreihe von der Parameterdatenreihen-Erzeugungsschaltung (18, 20 und 24) und der Tonlagendaten von der Steuereinheit (16) synthetisiert bzw. zusammensetzt.

5. Einrichtung zur Sprachsynthese nach Anspruch 2, dadurch gekennzeichnet, daß die Syntheseschaltung umfaßt: einen Interpolator (26), der die Parameterdatenreihe(n) von der Parameterdatenreihen-Erzeugungsschaltung (18, 20 und 24) abnimmt und eine vorbestimmte Zahl von interpolierten Parameterdaten auf der Grundlage zweier aufeinanderfolgender Parameterdaten berechnet, und eine Zusammensetzeinheit (28) zum Zusammensetzen des Sprachsignals auf der Grundlage der Parameterdaten vom Interpolator (26).

6. Einrichtung zur Sprachsynthese nach Anspruch 5, dadurch gekennzeichnet, daß die Datenerzeugungsschaltung (14) Akzentdaten zusammen mit den Phonemekettendaten und die Steuereinheit (16) Tonlagendaten nach Maßgabe der Akzentdaten erzeugen und daß die Syntheseschaltung (22, 26, 28 und 30) das Sprachsignal auf der Grundlage der Parameterdatenreihe von der Parameterdatenreihen-Erzeugungsschaltung (18, 20 und 24) und der Tonlagendaten von der Steuereinheit (16) synthetisiert bzw. zusammensetzt.

Revendications

1. Appareil de synthèse de la parole, comprenant: un circuit générateur de données (14) servant à produire des données de chaînes de phonèmes; un moyen de mémoire (10 et 12), dans lequel des données de paramètres caractéristiques de consonnes et de voyelles représentatives de segments de consonnes et de voyelles sont emmagasinées et qui possède un fichier de segments de consonnes (10) dans lequel plusieurs données de paramètres caractéristiques de consonnes représentatives de plusieurs segments de consonnes, chacun desquels possède une partie consonne et un segment transitoire vers un segment de voyelles, sont emmagasinées, et un fichier de segments de voyelles (12) dans lequel plusieurs données de paramètres caractéristiques de voyelles représentatives de plusieurs segments de voyelles stationnaires sont emmagasinées; un moyen de commande (16) servant à permettre que les données de paramètres caractéristiques de consonnes et de voyelles correspondantes soient produites par ledit moyen de mémoire (10 et 12) en fonction desdites données de chaînes de phonèmes; et un moyen de synthèse (18, 20, 22, 24, 26, 28 et 30) servant à synthétiser un signal de parole sur la base desdites données de paramètres caractéristiques de consonnes et de voyelles venant dudit moyen mémoire (10 et 12); et comportant un circuit générateur de série de données de paramètres (18, 20 et 24) servant à produire une série de données de paramètres caractéristiques de consonnes et de voyelles sur la base des données de paramètres caractéristiques de consonnes et de voyelles obtenues à partir desdites données de paramètres caractéristiques de consonnes et de voyelles venant desdits fichiers de segments de consonnes et de voyelles (10 et 12), et un circuit de synthèse (22, 26, 28 et 30) servant à synthétiser le signal de parole sur la base de la série de données de paramètres venant dudit circuit générateur de série de données de paramètres (18, 20 et 24), caractérisé en ce que ledit fichier de segments de voyelles (12) comprend en outre plusieurs données de paramètres caractéristiques de voyelles représentatives de plusieurs segments de voyelles coarticulés, chacun desdits segments de voyelles stationnaires et coarticulés étant formés de données de paramètres d'un bloc unique, ledit moyen de commande (16) produit des données de longueur de temps indicatives de la durée des voyelles en fonction des données de chaînes de phonèmes venant dudit circuit générateur de données (14), et ledit circuit générateur de série de données de paramètres (18, 20, 24) comporte un circuit de répétition (20) qui extrait dudit fichier de segments de voyelles (12) les données de paramètres caractéristiques de voyelles un nombre de fois qui correspond à ladite donnée de longueur de temps.

2. Appareil de synthèse de la parole selon la revendication 1, caractérisé en ce que ledit circuit générateur de série de données de paramètres comporte en outre: un circuit d'interpolation (18)

servant à calculer un nombre prédéterminé de données de paramètres caractéristiques interpolées sur la base des données de paramètres caractéristiques de consonnes et de voyelles venant desdits fichiers de segments de consonnes et de voyelles (10 et 20); et un circuit de sélection de données (24) servant à délivrer séquentiellement et sélectivement les données de paramètres caractéristiques venant dudit circuit d'interpolation (18) et dudit circuit de répétition (20) audit circuit de synthèse (22, 26, 28 et 30).

3. Appareil de synthèse de la parole selon la revendication 2, caractérisé en ce que ledit circuit de sélection de données est un circuit de commutation (24) dont la position de commutation est commandée en fonction d'un signal de commande de commutation venant dudit moyen de commande (16).

4. Appareil de synthèse de la parole selon la revendication 2, caractérisé en ce que ledit circuit générateur de données (14) produit des données d'accent en même temps que lesdites données de chaînes de phonèmes et ledit moyen de commande (16) produit des données de hauteur de son en fonction desdites données d'accent, et en ce que ledit circuit de synthèse (22, 26, 28 et 30) synthétise le signal de parole sur la base de la série de données de paramètres venant dudit circuit générateur de série de données de paramètres (18, 20 et 24) et des données de hauteur de son venant dudit moyen de commande (16).

5. Appareil de synthèse de la parole selon la revendication 2, caractérisé en ce que ledit circuit de synthèse comprend: un interpolateur (26) qui reçoit la série de données de paramètres de la part du circuit générateur de série de données de paramètres (18, 20 et 24) et calcule un nombre prédéterminé de données de paramètres interpolées sur la base de deux données de paramètres successives; et une unité de synthèse (28) servant à synthétiser le signal de parole sur la base des données de paramètres venant dudit interpolateur (26).

6. Appareil de synthèse de la parole selon la revendication 5, caractérisé en ce que ledit circuit générateur de données (14) produit des données d'accent en même temps que lesdites données de chaînes de phonèmes et ledit moyen de commande (16) produit des données de hauteur de son en fonction desdites données d'accent, et en ce que ledit circuit de synthèse (22, 26, 28 et 30) synthèse le signal de parole sur la base de la série de données de paramètres venant dudit circuit générateur de série de données de paramètres (18, 20 et 24) et des données de hauteur de son venant dudit moyen de commande (16).

FIG. 1

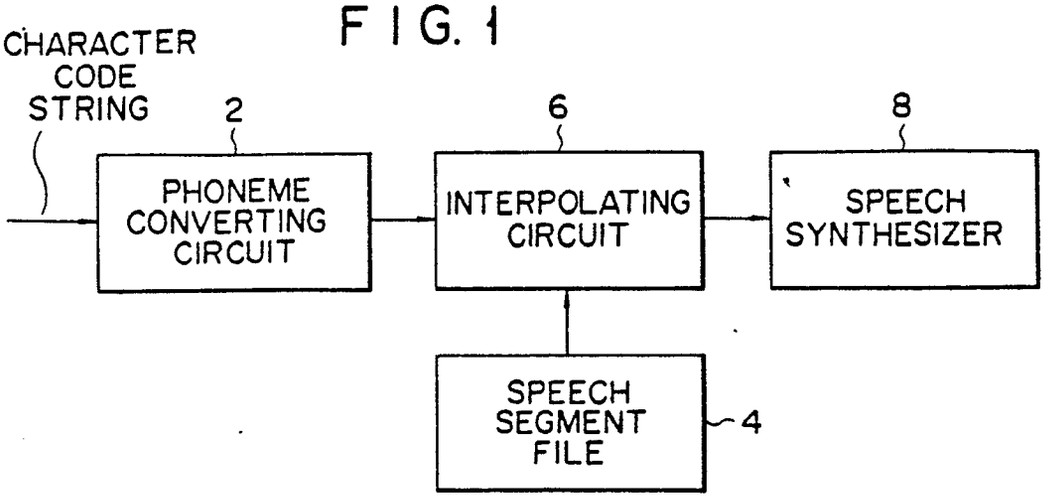


FIG. 2

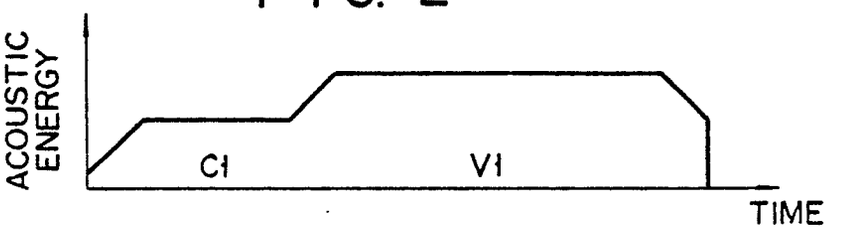


FIG. 3A

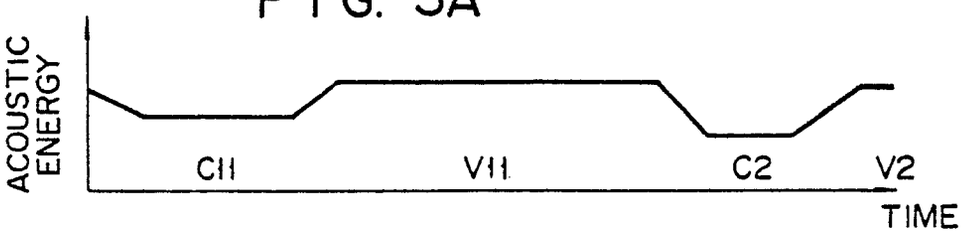


FIG. 3B

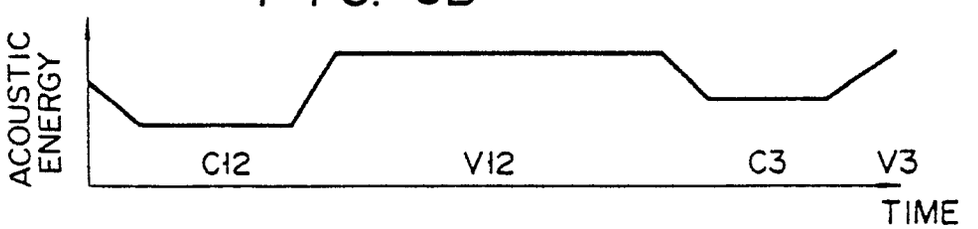
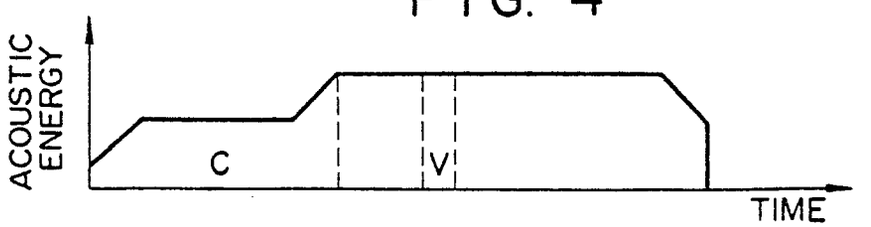


FIG. 4



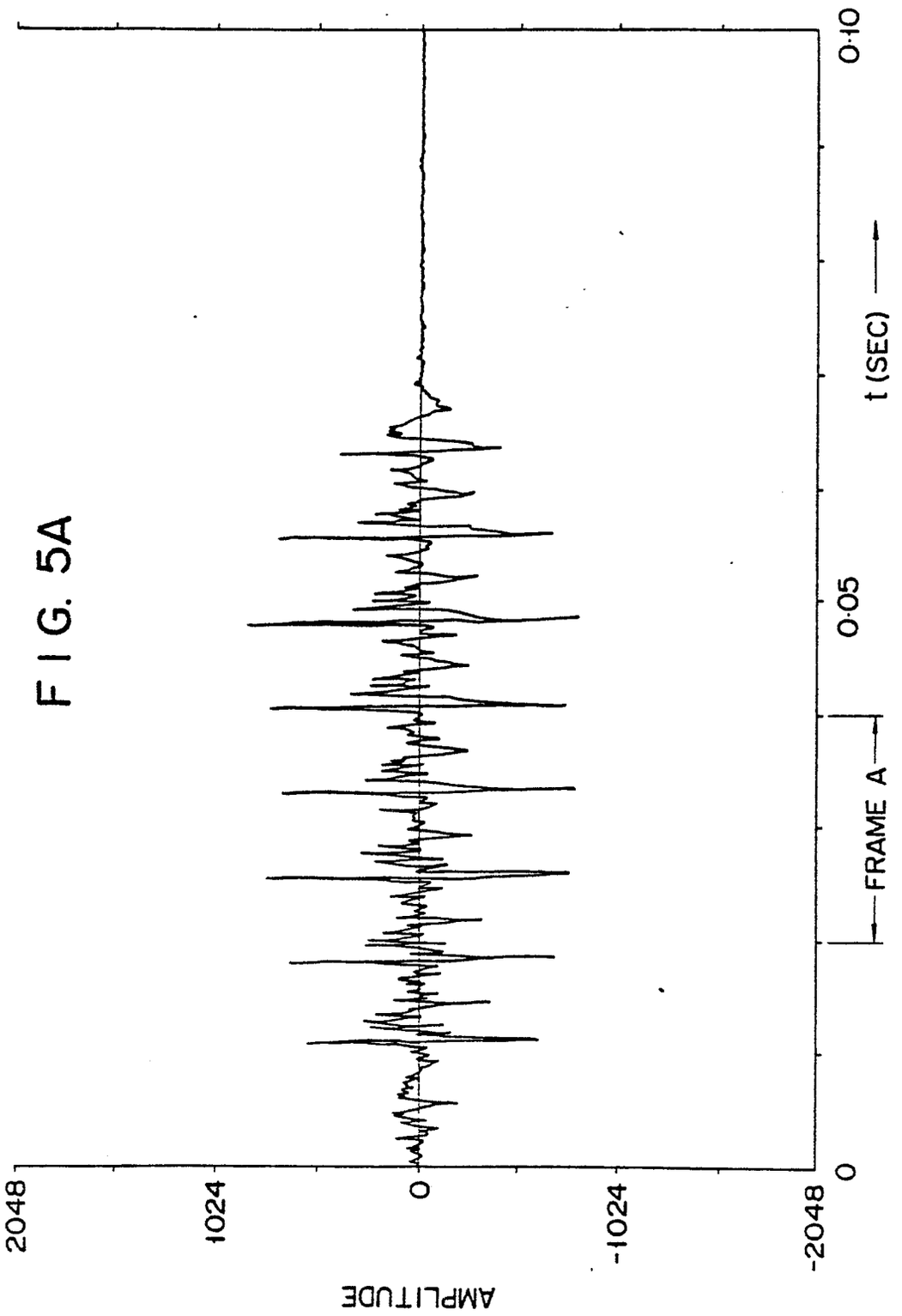


FIG. 5A

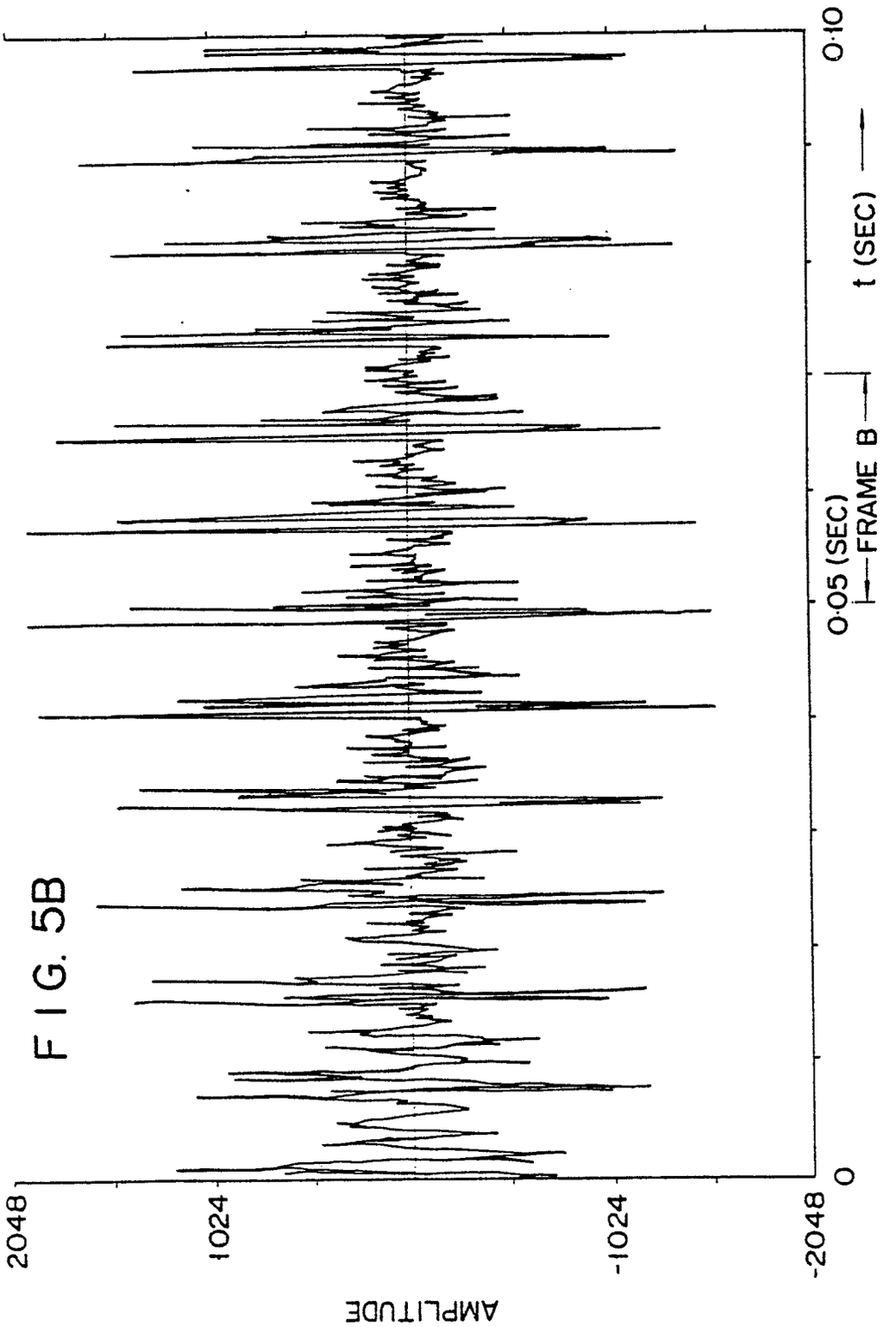


FIG. 5B

FIG. 6A

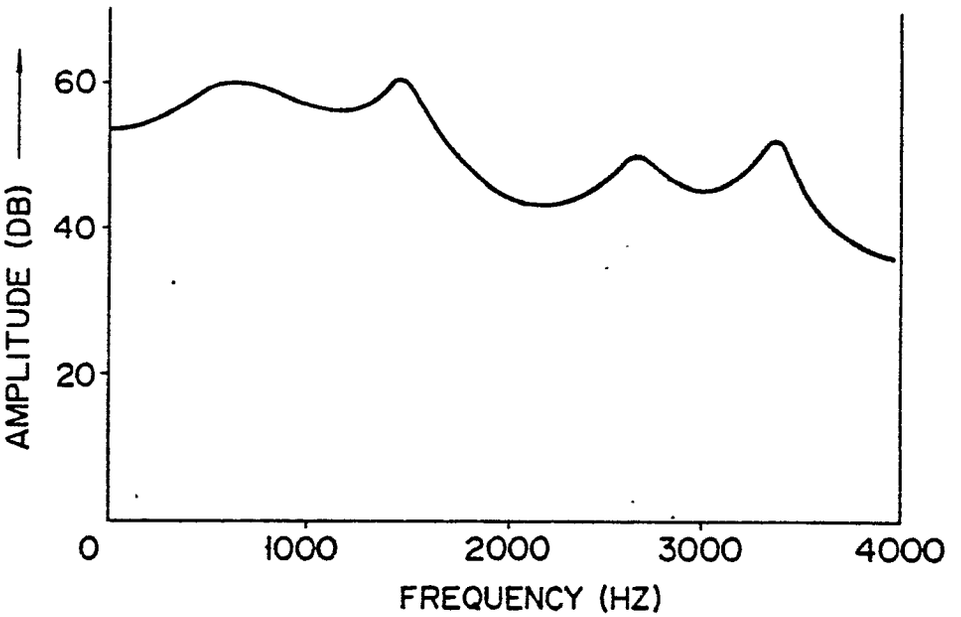

AMPLITUDE (DB)
60
40
20
0
1000
2000
3000
4000
FREQUENCY (HZ)

FIG. 6B

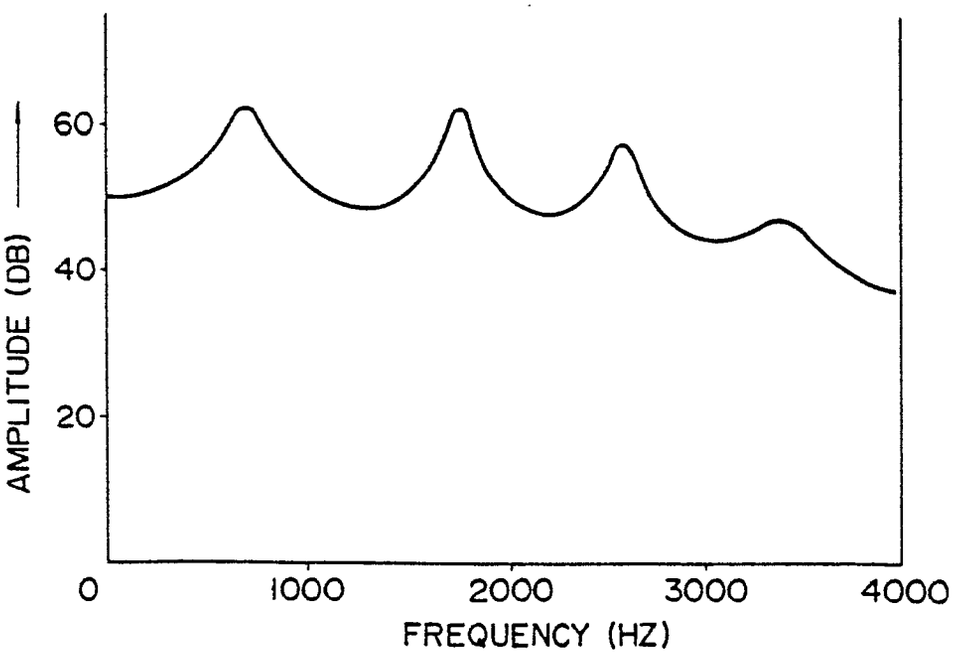

AMPLITUDE (DB)
60
40
20
0
1000
2000
3000
4000
FREQUENCY (HZ)

FIG. 7A

FIG. 7B

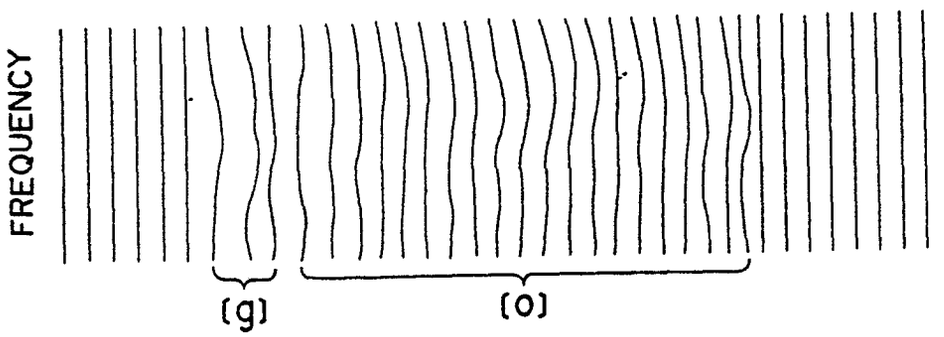


FIG. 7C

FIG. 7D

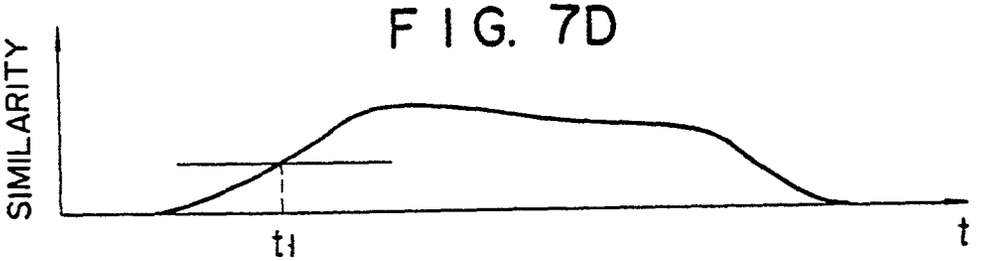
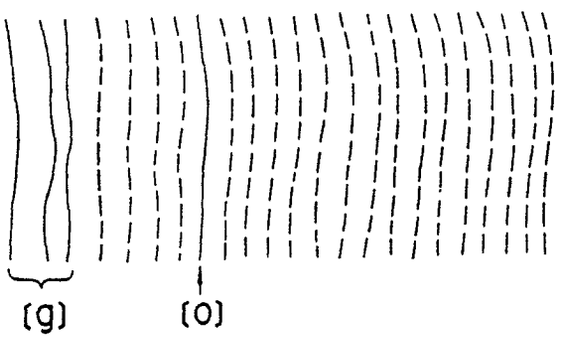


FIG. 9



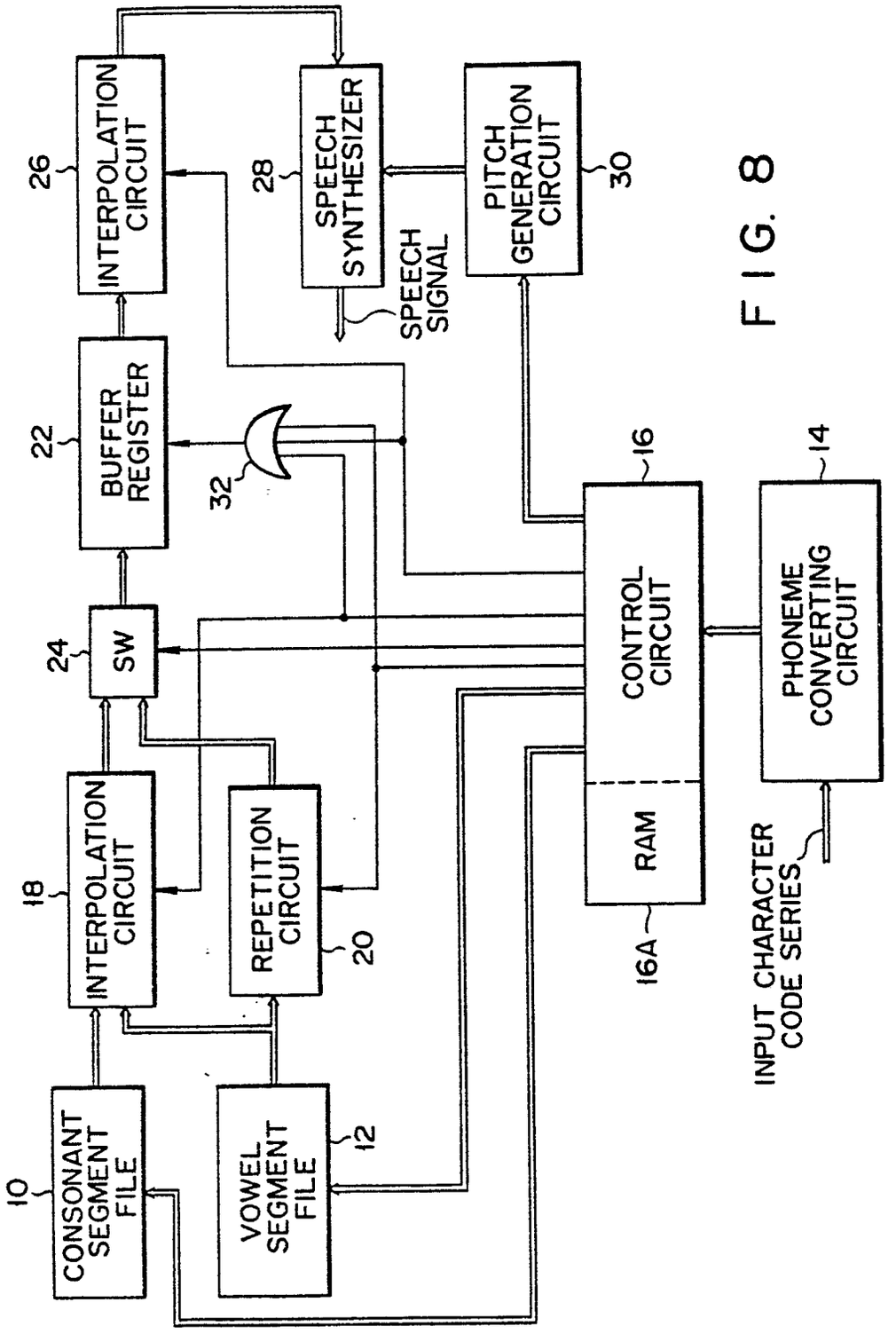
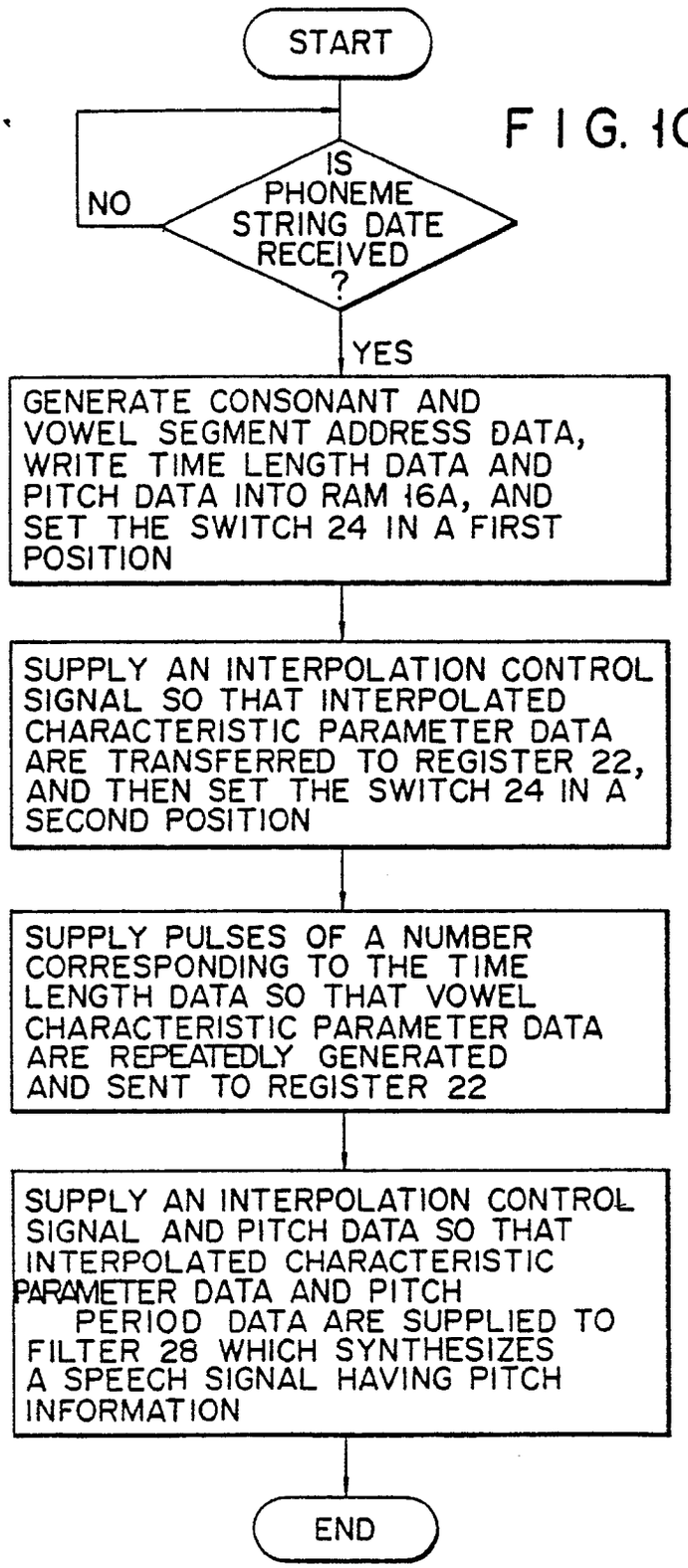

CONSONANT SEGMENT FILE
10
VOWEL SEGMENT FILE
12
INTERPOLATION CIRCUIT
18
REPETITION CIRCUIT
20
SW
24
BUFFER REGISTER
22
INTERPOLATION CIRCUIT
26
32
SPEECH SYNTHESIZER
28
SPEECH SIGNAL
PITCH GENERATION CIRCUIT
30
CONTROL CIRCUIT
16
RAM
16A
PHONEME CONVERTING CIRCUIT
14
INPUT CHARACTER CODE SERIES

FIG. 8


START
IS PHONEME STRING DATE RECEIVED ?
NO
YES
GENERATE CONSONANT AND VOWEL SEGMENT ADDRESS DATA, WRITE TIME LENGTH DATA AND PITCH DATA INTO RAM 16A, AND SET THE SWITCH 24 IN A FIRST POSITION
SUPPLY AN INTERPOLATION CONTROL SIGNAL SO THAT INTERPOLATED CHARACTERISTIC PARAMETER DATA ARE TRANSFERRED TO REGISTER 22, AND THEN SET THE SWITCH 24 IN A SECOND POSITION
SUPPLY PULSES OF A NUMBER CORRESPONDING TO THE TIME LENGTH DATA SO THAT VOWEL CHARACTERISTIC PARAMETER DATA ARE REPEATEDLY GENERATED AND SENT TO REGISTER 22
SUPPLY AN INTERPOLATION CONTROL SIGNAL AND PITCH DATA SO THAT INTERPOLATED CHARACTERISTIC PARAMETER DATA AND PITCH PERIOD DATA ARE SUPPLIED TO FILTER 28 WHICH SYNTHESIZES A SPEECH SIGNAL HAVING PITCH INFORMATION
END

F I G. 10