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

(11) Publication number: **0 071 716**
B1

(12) # EUROPEAN PATENT SPECIFICATION

(45) Date of publication of patent specification: **26.08.87**

(51) Int. Cl.⁴: **G 10 L 9/18**

(21) Application number: **82105168.7**

(22) Date of filing: **14.06.82**

(54) **Allophone vocoder.**

(30) Priority: **03.08.81 US 289604**
03.08.81 US 289603
03.08.81 US 289690

(43) Date of publication of application:
16.02.83 Bulletin 83/07

(45) Publication of the grant of the patent:
26.08.87 Bulletin 87/35

(84) Designated Contracting States:
DE FR GB NL

(56) References cited:

IEEE INTERNATIONAL CONFERENCE ON ACOUSTIC, Speech and signal processing, April 10-12, 1978, Tulsa NEW YORK (US), V.N. GUPTA et al.: "Speaker-independent vowel identification in continuous speech", pages 546-548

(73) Proprietor: **TEXAS INSTRUMENTS INCORPORATED**
13500 North Central Expressway
Dallas Texas 75265 (US)

(72) Inventor: **Ott, Granville E.**
3408 57th Street
Lubbock Texas 79413 (US)
Inventor: **Lin, Kun-Shan**
8601 Utica
Lubbock Texas 79413 (US)

(74) Representative: **Leiser, Gottfried, Dipl.-Ing. et al**
Patentanwälte Prinz, Leiser, Bunke & Partner
Manzingerweg 7
D-8000 München 60 (DE)

(56) References cited:

IEEE TRANSACTIONS ON ACOUSTICS, Speech and signal processing, vol. ASSP-24, no. 2, April 1976, NEW YORK (US), R.B. THOSAR et al.: "An approach towards a synthesis-based speech recognition system", pages 194-196

Note: Within nine months from the publication of the mention of the grant of the European patent, any person may give notice to the European Patent Office of opposition to the European patent granted. Notice of opposition shall be filed in a written reasoned statement. It shall not be deemed to have been filed until the opposition fee has been paid. (Art. 99(1) European patent convention).

Courier Press, Leamington Spa, England.

EP 0 071 716 B1

(56) References cited:
RADIO FERNSEHEN ELEKTRONIK, vol. 29, no. 2, 1980, BERLIN (DE), W. TSCHESCHNER et al.: "Technische Systeme der automatischen Sprachverarbeitung", pages 83-87

THE JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, vol. 42, no. 2, August 1967, NEW YORK (US), D.R. REDDY: "Computer recognition of connected speech", pages 329-347

IBM TECHNICAL DISCLOSURE BULLETIN, vol. 22, no. 4, September 1979, NEW YORK (US), H.R. GRUB: "Dictation system", pages 1327,1328

Description

This invention relates to a vocoder system as defined in the precharacterizing part of claim 1. Such a system is known from IEEE INTERNATIONAL CONFERENCE ON ACOUSTIC, Speech and Signal Processing, April 10—12, 1978, V.N. GUPTA et al.: "Speaker-independent vowel identification in continuous speech", pages 546—548. Further, the invention relates to a method of analyzing a speech signal and producing audible synthesized speech as defined in the precharacterizing part of claim 4.

It has long been recognized that analog speech signals contain numerous redundant sounds so as to make such signals not suitable for efficient data transmission. In a direct human interaction situation this inefficiency is tolerable. The technical requirements to cope with inefficient speed transmission though become infeasible due to cost, time and the increased memory storage which is rendered necessary because of the inefficiency. A need exists for a system and a method which can take an analog speech signal and translate it into a digital form which is reconstructable after transmission or storage. This type of device is generally referred to as a "vocoder".

Intrinsic to the recognition of an analog speech is the use of a methodology which breaks the analog speech into its component parts which may be compared to some library for identification. Numerous methods and apparatuses have evolved so as to approximate the human speech and to model it. These modeling techniques include the vocoder, linear predictive filters, and other devices.

One such method of analyzing the analog speech was discussed by James L. Flanagan in the article "Automatic Extraction of Formant Frequencies from Continuous Speech" first printed in J. Acoust. Soc. Am., Vol. 28, pp. 110—118, January 1956, incorporated hereinto by reference.

In the article, Flanagan discusses two electronic devices which automatically extract the first three formant frequencies from continuous speech. These devices yield continuous DC output voltages whose magnitudes as functions of time represent the formant frequencies of the speech. Although the formant frequencies are in an analog form, use of an analog-digital (AD) converter readily transforms these formant frequencies into digital form which is more suitable for use in an electronic environment.

Another method was discussed by H. K. Dunn in his article "Methods of Measuring Vowel Formant Bandwidths" J. Acoust. Soc. Am., Vol. 33, pp. 1737—1746, December 1961, incorporated hereinto by reference. In the article, Dunn discloses the use of spectrums of real speech and the use of an artificial larynx in an application to real subjects.

It is clear therefore that an efficient methodology and apparatus for transformation of an analog speech signal to an approximating digital form does not exist. The mere recognition of formants or the use of diphones in the synthesis of the perceived speech is inaccurate and does not allow for quality recordation and transmission of data representative of the original speech signal.

The problems described above can be solved by use of the features contained in the characterizing parts of claims 1 and 4.

Since the human ear acts as a filtering mechanism and also due to the inherent redundancy of the spoken language, any errors which are generated in the selection of the optimal phoneme match are minimized. For example, assume that the phoneme recognizer incorrectly matched the spoken phoneme "SH" to the phoneme "CH" in the phrase "We will be taking a cruise on the ship". This results in the phrase becoming "We will be taking a cruise on the chip". Although the transmitted phoneme sequence is not a perfect match, the total phrase is still intelligible to the listener since the human ear and the mental process filter out this incorrect phoneme. The human ear and mental process have developed over the years to compensate for variations in pronunciations and the incorrect usage of words.

Some applications of this allophone vocoder device are found in a digital dictating machine, a store and play telephone, voice memos, multichannel voice communications, voice recorded exams, etc. In the situation of a dictating machine, the erroneous matching of the phonemes is more visible than in the synthesized speech situation; but it provides a rough draft or first cut to the document so as to be edited later.

This invention uses a phoneme-to-allophone matching algorithm, such that the quality of synthesized speech is vastely improved since allophones more closely map the human utterances.

This vocoder accepts the analog speech input and matches it to a set of phoneme templates; the phonemes each contain a phoneme code which is compressed into a sequence of phoneme codes and communicated via a channel. This channel should be as noise free as possible so as to provide accurate transmission. The sequence of phonemes is received and then translated to an analogous allophone sequence and synthesized through known electronic synthesis means.

One such means is discussed in U.S. Patent No. 4,209,836 issued to Wiggins, Jr., et al on June 24, 1980, incorporated hereinto by reference. This speech synthesis integrated circuit device uses a linear predictive filter in its generation of the synthesized speech.

The control of the data within the synthesizer is well known in the art. One such method for communicating digital speech data and control of the memory for storing the data is disclosed in U.S. Patent No. 4,234,761 issued to Wiggins, Jr., et al on November 18, 1980.

In an embodiment of the invention as defined in claim 3, the phoneme recognizer contains an automatic gain control (AGC). The phoneme rec-

ognizer receives the voice input and automatically controls the gain of the voice and sends a signal to the formant-determining means for analysis and formant extraction. The algorithm operates on the formants and features of the utterance requiring the detection of the phoneme boundary within the speech. The detected phoneme is matched to a phoneme in a library of phoneme templates. Each phoneme template has a corresponding identification code. The selected identification code is sequentially packed and transmitted via a transmission channel to a receiver.

The transmission channel may be either a wired or wireless communication medium. Ideally the transmission channel is as noiseless as possible so as to reduce errors.

The phoneme-to-allophone synthesizer receives the phoneme codes from the channel. The algorithm converts the phoneme sequence into an analogous allophone sequence and thereby produces quality speech. In the phoneme-to-allophone synthesizer, a control means sequentially directs a library of allophone characteristics to be communicated to a speech synthesizer.

The use of an efficient formant-determining means is beneficial. A formant is a frequency component in the spectrum of speech which has large amplitude energy. It also has a resonant frequency of the pitch and a voiced sound. This resonant frequency is a multiple of a fundamental frequency. The first formant occurs between 200 to 850 Hertz (Hz), and the second formant occurs between 850 and 2,500 Hz, the third formant occurs between 2,500 and 3,500 Hz.

The invention together with its particular embodiments and ramifications will be more fully explained by the following drawings and their accompanying descriptions.

FIGURE 1 is a block diagram of an embodiment of the invention illustrating the data compression and transmission capabilities of the invention.

FIGURE 2a is a block diagram of the communication relationship of the invention.

FIGURES 2b and 2c illustrate the recognition side and the synthesis side respectively of the embodiment illustrated in FIGURE 2a.

FIGURE 3 is an embodiment of the invention utilized to generate indicia representative of the analog speech signal.

FIGURE 4 is illustrative of the determination of the bandwidth associated with a particular formant.

FIGURE 5 is a flow chart of an embodiment determining the formant of the analog speech signal.

FIGURE 6 illustrates a method of determining indicia so as to define a particular formant structure of an analog speech signal.

FIGURE 7 illustrates an encoding scheme for the indicia.

FIGURE 8 illustrates a translational operation of a phoneme to either an allophone or alphanumeric characters.

FIGURE 9 is an example of a decisional tree operating upon the encoded indicia as representated in FIGURE 7.

FIGURES 10a and 10b illustrate the translation of phonemes-to-allophones.

Drawings in Detail

FIGURE 1 illustrates in block diagram the capabilities of an embodiment of the invention.

Analog speech 101 is picked up by the microphone 102 and transmitted in analog form to the analog to digital (A/D) converter 103. Once the signal has been translated into digital form, it is converted to a perceived phoneme via the conversion means 104. Each perceived phoneme is communicated to the comparator 105 and referenced to templates in the library 106 so that a match is obtained. Once a matched phoneme is determined, its code is communicated via the bus 107 to either the phoneme sequencer 108, the storage means 109, or the transmitter 110.

The code sequence which matches to the phoneme sequence totally identifies the analog speech 101. This code sequence is more susceptible to being packed or for storage than the original analog speech 101 due to its digital nature.

The phoneme sequencer 108 utilizes the code communicated via the bus 107 to obtain the appropriate phoneme from the library 106. This phoneme from the library 106 has associated with it a set of allophone characteristics which are communicated to the synthesizer 114. The synthesizer 114 communicates an analog signal to operate speaker 115 in the generation of speech 116. Through the use of the phoneme-to-allophone translation as effectuated by the phoneme sequencer 108, with the aid of library 106, a more intelligible and higher quality speech 116 is generated. This translation ability permits the encoding of the data in a phoneme base so as to facilitate a lower bit per second transmission rate and thus requires less time and storage medium for the recordation of the original analog speech 101.

Alternatively, the phoneme codes are stored via storage means 109 for later retrieval. This later retrieval is optionally used by the phoneme sequencer 108, synthesizer 114, and speaker 115 sequence to again synthesize the phoneme sequence in allophone form for generation of speech 116. Optionally, the storage means 109 communicates the phoneme codes to the phoneme-to-alphabet converter 111 which translates the phonemes to their equivalent alphanumeric parts. Once the phonemes have been translated to the alphanumeric parts, such as in ASCII code, they are readily transmitted to the printer 112 so as to produce a paper copy 113 of the original analog speech 101.

This branch of the operation, the storage

means 109, phoneme-to-alphabet converter 111, and printer 112, allows the invention to generate printed text from a speech input so as to permit an automatic dictating device.

Another alternative is for the phoneme codes from the bus 107 to be communicated to a transmitter 110. The transmitter generates signals 117 representative of the phoneme codes which are perceived by a remote unit 120 at its receiver 118.

The remote unit 120 contains the same capabilities as the local unit 121. This entails the transmission of the phoneme code via a bus 119 from the receiver 118. Again, once the phoneme code is transmitted via the bus 119, it is susceptible for the remote storage means 109' or the remote sequencer 108'. In another embodiment of the invention the phoneme codes transmitted via the bus 119 are also communicatable to a remote transmitter, not shown.

The remote unit 120 utilizes the phoneme codes in the same manner as the local unit 121. The phoneme codes are utilized by the remote sequencer 108' in conjunction with the data in the remote library 106' to generate an analogous allophone sequence which is communicated to the remote synthesizer 114'. The remote synthesizer 114' controls the operation of the remote speaker 115' in generating the speech 116'. The remote unit 120 also has the option of storing the phoneme code at the remote storage means 109' for later use by the remote sequencer 108' or the phoneme to alphabet converter 111'. The phoneme-to-alphabet converter 111' translates the phoneme code to its equivalent alphanumeric symbols which are communicated to the printer 112' to generate a paper copy 113'.

It is clear from this embodiment of the invention that the analog speech is translated to a phoneme code which is more susceptible to storage or for manipulation as a data string. The phoneme code permits easy storage, transmission, generation of a printed copy or eventual synthesis by translation to an analogous allophone sequence.

FIGURE 2a illustrates, in block form, an embodiment of the invention which receives the analog speech input and results in a speech output.

In the embodiment of FIGURE 2a, the original analog speech signal input 201 is communicated to a phoneme recognizer 202 which generates a sequence of phonemes 203 via a communication channel 204. The sequence of phonemes 205 is communicated to a phoneme-to-allophone synthesizer 206 which translates the phoneme sequence into its analogous allophone sequence so as to generate the speech output 207. It should be noted that the phoneme recognizer 202 and the phoneme-to-allophone synthesizer 206 are alternatively in the same unit or are remote one from the other. In this context the communication channel 204 is either a hard wired device such as bus or a telephone line or a radio transmitter with receiver.

FIGURE 2b illustrates an embodiment of the phoneme recognizer 202 illustrated in FIGURE 2a.

The analog speech signal input 201 is communicated to an automatic gain control circuit (AGC) 208 so as to regulate the speech signal into a certain desirable balance. The formant tracker 209 breaks the analog signal into its formant components which are stored in a random access memory (RAM) 210. Although in this embodiment the use of a RAM 210 is illustrated, it is contemplated that any suitable storage means could be employed. The formants stored in RAM 210 are communicated to the phoneme boundary detection means 211 so as to group the formants into perceived phoneme components. Each perceived phoneme is communicated to the recognition algorithm 212 which utilizes the phoneme templates from the library 213 which is comprised of known phonemes. A best match is made between the perceived phoneme from the phoneme boundary detection means 211 and the templates found in the phoneme template library 213 by the recognition algorithm 212 so as to generate a recognized phoneme code 214.

As noted earlier, a best match is obtained, even if not a perfect recognition, since the natural filtering of the human ear and the error correction of the mental processes of the listener minimize any errors generated by the recognition algorithm 212. The recognition algorithm 212 provides a continuous sequence of phoneme codes so that a blank or non-recognized phoneme does not exist in the sequence. A blank for a non-recognition determination only results in an increase in the noise of the invention.

FIGURE 2c illustrates an embodiment of a phoneme-to-allophone synthesizer 206.

The sequence of phoneme codes 205 is communicated to the controller 215. The controller 215 utilizes these codes and its prompting of the read only memory (ROM) 217 to communicate to the speech synthesizer 216 the appropriate bit sequence indicative of the analogous allophone sequence. This data communicated from the ROM 217 to the speech synthesizer 216 establishes the parameters necessary for the modulation of the speaker 218 in the generation of the synthesized speech.

The speech synthesizer is chosen from a wide variety of speech synthesis means, including, but not limited to, the use of a linear predictive filter.

FIGURE 3 is a block diagram of an embodiment of the invention which generates indicia representative of the analog speech.

This indicia is representative of the perceived phoneme and is used in finding a best match or optimal match with the template in the library. The automatic gain control circuit (AGC) 301 communicates an analog speech signal to the pitch tracker 302 and the integration means 304, 314, and 324. The pitch tracker 302 generates a fundamental frequency F0.

For each formant determiner 308, 318, and 328 a respective set of integers is determined for which the fundamental frequency F0, when multiplied by the integer falls within the formant range. The respective sets of integers are broadened to

include an overlap in the sets so that the entire formant is defined. As an example, if the fundamental frequency F0 is 200 Hz, the integer set for the first formant may contain (0, 1, 2, 3, 4); the second formant integer set contains (4, 5, 6, 7); the third formant integer set contains (7, 8, 9).

The formant determiner 308 accepts the fundamental frequency F0 and utilizes it with an integer value from the integer set for n in the sinusoidal oscillator 303. The sinusoidal oscillator 303 generates a sinusoidal signal, s(t), which is centered at the product n and the fundamental frequency. The sinusoidal signal is communicated to the integrator 304 which integrates the product of the sinusoidal signal s(t) and the analog speech signal, f(t) over the chosen frequency of the formant. This integration by the integrator 304 creates a convolution of the analog speech signal f(t).

This operation involving the generation of a sinusoidal signal by the sinusoidal oscillator 303 and the communication thereof to the integrator 304 is continued for all integer values within the integer set by the incrementer 306. The value of n which generates the maximum amplitude from the integrator 304 is chosen by the determinator 305. This optimal value, N', is used to generate the first formant F1 defined by $F1 = N' \times F0$. This product additionally is determinative of the bandwidth BW1, of the first formant and the pair F1 and BW1 are communicated via channel 307.

In like fashion the formant determiners 318 and 328 generate a sinusoidal signal via the sinusoidal oscillators 313 and 323 respectively and subsequently integrate by the integrators 314 and 324 so as to obtain the optimal values M' and K', 315 and 325 respectively.

The indicia BW1, F1, BW2, F2, BW3, F3, and F0, represent the perceived phoneme indicia from the analog speech from the AGC circuit 301. This perceived indicia is used to match the perceived phoneme to a phoneme template in a library so as to obtain a best match.

FIGURE 4 indicates the relationship of the bandwidth to the optimal formant.

Once the optimal integer value N' is determined, its amplitude is plotted relative to the surrounding integers. The independent axis 402 contains the frequencies as dictated by the product of the integer value with the fundamental frequency. The dependent axis 403 contains the amplitude generated by the product in the convolution with the analog speech signal. As illustrated, the optimal value N' generates an amplitude 404. By utilizing the surrounding data points 405, 406, 407, and 408, a bandwidth BW1 is determined for the appropriate optimal value N'.

The use of this bandwidth forms another indicia for determining the perceived phoneme relationship to the phoneme templates of the library. Similar analysis is done for each formant.

FIGURE 5 is a flow chart of an embodiment for determining the optimal formant positions.

The algorithm is started at 501 and a fundamental frequency, F0, 502 is determined. This fundamental frequency is utilized to optimize on N 503. The optimization on N 503 entails the initialization of the N value 504 followed by the sinusoidal oscillation based at the product of N F0 505. The frequency convolver 506 generates the convolution of the fundamental frequency F0 and the inputted analog speech signal over the chosen frequency of the formant. The convolution is optimized at 507 wherein if it is not the optimal value, the N value is incremented at 508 and the process is repeated until an optimal N value is determined. At the optimization of N, the algorithm proceeds to optimize on the value of M 513 and then to optimize on the value K 523. The optimization on N 503, the optimization of M 513, and the optimization of K 523 are identical in structure and performance.

In this embodiment three formant frequency ranges are utilized to define the human language. It has been found that three ranges accurately described the human speech, but this methodology is either extendable or contractable at the will of the designer. No loss in generality is encountered when the algorithm is extended to apply to a single formant or similarly to extend to more than three formants.

FIGURE 6 graphically illustrates another methodology for the encoding of the analog speech signal in the formants.

The analog speech signal 608 is plotted over the independent axis 601 of frequency. The dependent axis 602 is the amplitude. Within the first formant 603, the frequency range lies between 200 and 700 Hz. The second formant 604 has a frequency range of 850 to 2500 Hz; and the third formant 605 has a frequency range of 2700 to 3500 Hz. A method similar to the methodology discussed in FIGURE 3 and FIGURE 5 is used to determine the location of the maximum amplitude within the formant range. These maxima yield a distance between maxima, 606 and 607 respectively. The distance, d_1, between the optimal first and second formants is used to characterize the perceived phoneme for matching to a phoneme template. This methodology allows two integer values d_1 and d_2 to describe what previously necessitated the use of three integer values (for the first, second and third formants).

FIGURE 7 is an embodiment of the encoding scheme for establishing a word for matching to the phoneme template.

The data word 701 in this example is an 8 bit word but any length of word which is capable of adequately describing the perceived phoneme is acceptable. In this embodiment the 8 bits are broken up into four basic components, 702, 703, 704, and 705.

The first component 702 is indicative of a pause or no pause situation. Hence if b_0 is set to a value of 1, a pause has been perceived and the appropriate steps will therefore be taken; similarly a 0 at b_0 indicates lack of a pause. A similar relationship exists at bit b_1, 703, which indicates a voiced or unvoiced phoneme. Bits b_2—b_3, 704, indicate the

contour of the analog speech signal; its assigned value indicates a level slope, a positive slope or a negative slope. Bits b_4—b_7, 705, indicate a mixture of the relative energy, relative pitch, first distance, and second distance. Bits b_4—b_7, 705, are encoded so that their value indicates the characteristics of the perceived phoneme relating to the formant distances. Bits b_4—b_7 are encoded to communicate the distances between the maximums within each formant as illustrated in Figure 6. From table 706, each value within the range of bits b_4—b_7 absolutely defines the two distances.

FIGURE 8 illustrates the translation of the phoneme code sequence into its appropriate allophone sequence or alternately its alphanumeric counterpart.

The phoneme sequence 801 is broken into its phoneme codes such as phoneme code 802. The phoneme code 802 distinctly describes a particular phoneme 807. This phoneme 807 is either printed as at 805 in its ASCII alphanumeric character or it is translated to its analogous allophone sequence when it is taken in conjunction with the surrounding phoneme codes 803 and 804.

The allophone sequence 806 is generated through the knowledge of the target phoneme 807 and its relationship to its surrounding phonemes. In this context, the phonemes which precede, 803, and follow, 804, the target phoneme 802 are retained in memory so as to generate the appropriate allophone sequence 806.

FIGURE 9 illustrates the characteristics of an embodiment of a decisional tree which determines the best approximation of the phoneme template in matching the perceived phoneme.

The decisional tree is broken up into multiple stages 901, 902, etc. Each stage of the tree breaks the perceived phoneme into a feasible and infeasible matches. As the perceived phoneme is further broken into feasible and infeasible states, the infeasible state becomes absorbing and the feasible state decreases so that eventually a single phoneme template is the only possible choice. Hence, the final stage of the tree must consist of as many nodes as there are templates.

The original decision 903 is made on whether the first bit, b_0, is either set or not set. If the first bit is set, transition is made to node 905; the nodes which follow node 904, B1, are ignored. This determination on the b_0 level results in translating the available phoneme templates into an infeasible set, those lying exclusively behind node 904 and a feasible set, those lying behind node B2, 905. A similar determination is made for each component part of the indicia. In this example, another separation is made on b_1 and then on the value of b_2—b_3. This separation into nodes is continued until a final or terminating node is encountered which uniquely identifies the phoneme template chosen.

Movement is acceptable laterally between nodes such as between nodes E1, 908, and E2, 909 via the ray 907. This movement is permissible so long as a cycle is not thereby created. In this context ray 910 indicates a cycle between D1 and C1. For example, a sequence containing C1-D1-C1-D1-C1 is not acceptable since it is a cycle. This sequence causes a never ending cycle which results in a decision never being made. The one qualification of the tree illustrated in this embodiment is that a decision must eventually be reached.

The algorithm illustrated in FIGURE 9 is but one embodiment to identify the best match between the perceived phoneme and the phoneme template. Another approach is to generate a comparison value for each phoneme template relative to the perceived phoneme and then choose the optimal value accordingly. This approach requires more computation and a longer time for its operation.

FIGURES 10a and 10b illustrate a phoneme to allophone transformation wherein a phoneme is translated to its analogous allophone sequence.

In FIGURE 10a, a list of the rules in defining the allophone is set forth. As illustrated "b", 1001 illustrates a blank or a word boundary. The different symbols illustrated indicate different allophonic characteristics which are attachable to a phoneme. The syllables are broken by a period ".", 1002. These allophonic rules are combined with the phonemes to generate the appropriate allophone sequence.

FIGURE 10b illustrates how the phoneme "CH", 1003, translates into an appropriate allophone sequence. Depending upon the preceding and the following phoneme, the phoneme "CH" is either a "b CH", 1004, as in "chain" or lies within a word as illustrated by "CH", 1005, as in "bewitching".

Each phoneme maps into a unique allophone sequence. This allophone sequence is determined through knowledge of the preceding phoneme and the following phoneme within the phoneme sequence.

The invention as described herein details the use of a voice recognition system which translates the analog speech signal into a phoneme sequence which is more susceptible to compaction, storage, transmission, or translation to an analogous allophone sequence for speech synthesis. The phoneme perception allows for an unlimitable vocabulary to be used and also for a best match to be generated. The use of a best match is acceptable since the human ear acts as a filtering mechanism and the human brain ignores random noise so as to also filter the synthesized speech. The synthesized speech is enhanced dramatically through the translation of the phoneme sequence to an analogous allophone sequence. The stored phoneme sequence is susceptible to being translated to an alphanumeric sequence or for transmission via the radio or telephone lines.

This invention makes it possible for a direct speech to text dictating machine to be implemented and also can be advantageously employed to produce a highly efficient speech data transmission rate.

Claims

1. A vocoder system comprising: conversion means (104) for analyzing digital speech data representative of an analog speech signal to generate perceived phonemes representative of components of said digital speech data, memory means (106; 213) having digital speech data stored therein representative of a plurality of respective reference phonemes, and comparator means (105; 212) operably coupled to said conversion means (104) and to said memory means (106; 213) for selecting digital speech data representative of a particular reference phoneme from said memory means (106; 213) as the closest match for respective perceived phonemes of said digital speech data, said conversion means (104) being effective to generate perceived phonemes representative of all components of said digital speech data, said memory means (106; 213) being provided with encoded digital speech data stored therein including phoneme codes representative of a plurality of respective reference phonemes and with digital speech data stored therein representing sets of allophone characteristics, each set being associated with a reference phoneme, characterized in that means (108) are provided that are operably coupled to said comparator means (105; 212) and said memory means (106; 213) for forming a phoneme code sequence of a plurality of said phoneme codes, that said phoneme code sequence-forming means (108) communicates in dependence upon a respective phoneme code sequence of connected phoneme codes digital speech data representing a specific allophone characteristic out of the set of allophone characteristics of the memory means (106; 213) to speech synthesizer means (114), and in that the synthesizer means (114) are connected with audio means (115) for converting the analog speech signal of the synthesizer means into audible synthesized speech.

2. A system as set forth in claim 1, further characterized in that said conversion means (104) comprises means (209) for determining a series of formants from the analog speech signal and located within a frequency spectrum in which the analog speech signal is found, means operably coupled to said formant determining means for generating indicia representative of said series of formants corresponding to said analog speech signal, and means (211) for identifying phoneme boundaries within said analog speech signal and for identifying a set of said indicia representative of said formants within each of the identified phoneme boundaries, and that the comparator means (212) are operably coupled to said phoneme boundary identifying means (211) and said memory means (213) for establishing an optimal match between said set of indicia and said plurality of reference phonemes as represented by the encoded digital speech data stored in said memory means (213) by selecting the one of said plurality of reference phonemes most nearly characterized by said set of indicia to provide a phoneme code representative of said optimal match.

3. A system as set forth in either of claims 1 or 2, further characterized by automatic gain control circuit means (208) for receiving an analog speech input and generating said analog speech signal therefrom, said automatic gain control circuit means (208) being operably coupled to the input of said formant-determining means (209).

4. A method of analyzing a speech signal and producing audible synthesized speech comprising: providing an analog speech signal, identifying phoneme components of said analog speech signal, comparing each of the phoneme components with a plurality of reference phonemes, obtaining the closest match from said plurality of reference phonemes to each of the identified phoneme components, generating analog signals representative of synthesized speech based upon the closest matches of reference phonemes, producing audible synthesized speech from said analog signals and providing a plurality of phoneme codes representative of the reference phonemes comprising all of the recognized phonemes in a given spoken language for comparison with each of the phoneme components as identified from said analog speech signal to provide respective phoneme codes as the closest match thereto, characterized by forming a phoneme code sequence of connected phoneme codes, translating each formed phoneme code sequence into a respective allophone characteristic selected from a set of allophone characteristics belonging to a specific phoneme, and generating said analog signals representative of synthesized speech by means of the translated allophone sequence.

5. A method as set forth in claim 4, further characterized in that the identification of phoneme components of said analog speech signal comprises determining a series of formants within said analog signal, generating indicia representative of said series of formants, identifying phoneme boundaries within said analog signal, and identifying a set of indicia within each of said phoneme boundaries to define respective phoneme components of said analog speech signal by respective sets of indicia.

6. A method as set forth in claim 5, further characterized by subjecting said analog signal to automatic gain control prior to the determination of said series of formants therefrom.

7. A method as set forth in either of claims 5 or 6, further characterized in that the establishment of the closest match includes assigning a comparison value for each reference phoneme relative to said set of indicia, and choosing the optimal comparison value from the plurality of assigned comparison values, thereby selecting the reference phoneme most nearly characterized by said set of indicia.

8. A method as set forth in any of claims 4 to 7, further characterized by normalizing said analog speech signal by setting the pitch and speed thereof in accordance with the voice of a user

prior to the identification of said phoneme components thereof such that the subsequent translation of said phoneme code sequence to said allophone sequence enables the audible synthesized speech produced therefrom to more nearly approximate the original analog speech signal.

Patentansprüche

1. Vocodersystem mit einer Umsetzungseinrichtung (104) zum Analysieren von digitalen, ein analogisches Sprachsignal repräsentierenden Sprachdaten für die Erzeugung wahrgenommener Phoneme, die Komponenten der digitalen Sprachdaten repräsentieren, Speichereinrichtungen (106; 213), in denen digitale Sprachdaten gespeichert sind, die mehrere zugehörige Referenzphoneme repräsentieren, und Vergleichseinrichtungen (105; 212), die mit der Umsetzungseinrichtung (104) und den Speichereinrichtungen (106; 213) verbunden sind und aus den Speichereinrichtungen (106; 213) digitale Sprachdaten auswählen, die ein bestimmtes Referenzphonem als engste Anpassung an jeweils wahrgenommene Phoneme der digitalen Sprachdaten repräsentiert, wobei die Umsetzungseinrichtung (104) wahrgenommene Phoneme erzeugt, die alle Komponenten der digitalen Sprachdaten repräsentieren und wobei die Speichereinrichtungen (106; 213) codierte digitale Sprachdaten einschließlich von mehrere jeweilige Referenzphoneme repräsentierenden Phonemcodes enthalten sowie ferner digitale Sprachdaten enthalten, die Gruppen von Allophonmerkmalen repräsentieren, wobei jede dieser Gruppen einem Referenzphonem zugeordnet ist, dadurch gekennzeichnet, daß Mittel (108) vorgesehen sind, die wirkungsmäßig mit den Komparatoreinrichtungen (105; 212) und den Speichereinrichtungen (106; 213) verbunden sind und eine Phonemcodefolge aus mehreren Phonemcodes erzeugen, daß die Mittel (108) zur Erzeugung der Phonemcodefolgen abhängig von einer jeweiligen Phonemcodefolge aus verbundenen Phonemcodes digitale Sprachdaten, die ein besonderes Allophonmerkmal aus der Gruppe von Allophonmerkmalen aus den Speichereinrichtungen (106; 213) repräsentieren, zu einer Sprachsyntheseeinrichtung (114) übertragen, und daß die Sprachsyntheseeinrichtung (114) mit einer Audioeinrichtung (115) verbunden ist, die das analoge Sprachsignal aus der Syntheseeinrichtung in hörbare, synthetisch gebildete Sprache umsetzt.

2. System nach Anspruch 1, dadurch gekennzeichnet, daß die Umsetzungseinrichtung (104) Mittel (209) zur Bestimmung eine Reihe von Formanten aus dem analogen Sprachsignal innerhalb eines Frequenzspektrums, in dem das analoge Sprachsignal liegt, ferner Mittel enthält, die wirkungsmäßig mit den Mitteln zum Bestimmen der Formanten verbunden sind und Kennzeichen erzeugen, die die Reihe von Formanten entsprechend dem analogen Sprachsignal repräsentieren, und darüber hinaus Mittel (211) enthält, die innerhalb des analogen Sprachsignals Phonemgrenzen identifizieren sowie eine Gruppe der Kennzeichen identifizieren, die die Formanten innerhalb jede der identifizieren Phonemgrenzen repräsentieren, und daß die Komparatoreinrichtung (212) wirkungsmäßig mit den Mitteln (211) zum Identifizieren der Phonemgrenzen und mit den Speichereinrichtungen (213) verbunden ist, damit eine optimale Anpassung zwischen der Gruppe von Kennzeichen und den mehreren Referenzphonemen repräsentiert durch die in den Speichereinrichtungen (213) gespeicherten digitalen Sprachdaten erzeugt wird, indem aus den mehreren Referenzphonemen dasjenige ausgewählt wird, das durch die Gruppe von Kennzeichen am nächsten charakterisiert wird, damit ein die optimale Anpassung repräsentierender Phonemcode geliefert wird.

3. System nach Anspruch 1 oder 2, gekennzeichnet durch eine Verstärkungsregelschaltungseinrichtung (208), die eine analoge Spracheingabe empfängt und daraus das analoge Sprachsignal erzeugt, wobei die Verstärkungsregelschaltungseinrichtung (208) mit dem Eingang des Formantbestimmungsmittels (209) verbunden ist.

4. Verfahren zum Analysieren eines Sprachsignals und zum Erzeugen einer synthetisch gebildeten hörbaren Sprache, bei welchem ein analoges Sprachsignal erzeugt wird, Phonemkomponenten des analogen Sprachsignals identifiziert werden, jede der Phonemkomponenten mit mehreren Referenzphonemen verglichen wird, die engste Anpassung aus den mehreren Referenzphonemen an jede der identifierten Phonemkomponenten erhalten wird, synthetisch gebildete Sprache repräsentierende Analogsignale auf der Basis der engsten Anpassungen an die Referenzphoneme erzeugt werden, aus den Analogsignalen hörbare synthetisch gebildete Sprache erzeugt wird und mehrere Phonemcodes geliefert werden, die die Referenzphoneme repräsentieren, die alle erkannten Phoneme in einer gegebenen ausgesprochenen Sprache für einen Vergleich mit jede der Phonemkomponenten gemäß der Identifizierung aus dem analogen Sprachsignal enthalten, damit jeweils Phonemcodes als die engste Anpassung daran erhalten werden, dadurch gekennzeichnet, daß eine Phonemcodefolge aus verbundenen Phonemcodes erzeugt wird, daß jede gebildete Phonemcodefolge in ein entsprechendes Allophonmerkmal übersetzt wird, das aus einer Gruppe von Allophonmerkmalen ausgewählt ist, die zu einem speziellen Phonem gehören, und daß die die synthetisch gebildete Sprache repräsentierenden Analogsignale mit Hilfe der übersetzten Allophonfolge erzeugt werden.

5. Verfahren nach Anspruch 4, dadurch gekennzeichnet, daß bei der Identifizierung der Phonemkomponenten des analogen Sprachsignals eine Reihe von Formanten innerhalb des analogen Signals bestimmt wird, die Reihe von Formanten repräsentierende Kennzeichen erzeugt werden, innerhalb des analogen Signals Phonemgrenzen identifiziert werden und eine Gruppe von Kenn-

zeichen innerhalb jede der Phonemgrenzen identifiziert wird, damit jeweilige Phonemkomponenten des analogen Sprachsignals durch jeweilige Gruppen von Kennzeichen definiert werden.

6. Verfahren nach Anspruch 5, dadurch gekennzeichnet, daß das Analogsignal einer Verstärkungsregelung unterzogen wird, bevor die Bestimmung der Reihen von Formanten durchgeführt wird.

7. Verfahren nach Anspruch 5 oder 6, dadurch gekennzeichnet, daß die Erfassung der engsten Anpassung das Zuordnen eines Vergleichswertes für jedes Bezugsphonem relativ zu der Gruppe von Kennzeichen sowie das Auswählen des optimalen Vergleichswertes aus den mehreren zugeordneten Vergleichswerten beinhaltet, wodurch das Referenzphonem ausgewählt wird, das von der Gruppe von Kennzeichen am nächsten charakterisiert wird.

8. Verfahren nach einem der Ansprüche 4 bis 7, dadurch gekennzeichnet, daß das analoge Sprachsignal normiert wird, indem seine Tonhöhe und seine Geschwindigkeit entsprechend der Stimme eines Benutzers vor der Identifizierung der Phonemkomponenten der Stimme eingestellt werden, so daß die anschließende Übersetzung der Phonemcodefolge in die Allophonfolge eine genauere Annäherung der daraus erzeugten hörbaren synthetisch gebildeten Sprache an das originale analoge Sprachsignal ermöglicht.

Revendications

1. Système de codeur vocal comportant: un dispositif de conversion (104) destiné à analyser des données numériques de parole représentant un signal analogique de parole pour produire des phonèmes perçus représentant des composantes desdites données numériques de parole, un dispositif de mémoire (106; 213) dans lequel sont mémorisées des données de parole numérique représentant plusieurs phonèmes de référence et un dispositif comparateur (105; 212) couplé en fonctionnement avec ledit dispositif de conversion (104) et avec ledit dispositif à mémoire (106; 213) pour sélectionner des données numériques de parole représentant un phonème de référence particulier dans ledit dispositif à mémoire (106, 213) comme l'adaptation la plus proche pour des phonèmes perçus respectifs desdites données numériques de parole, ledit dispositif de conversion (104) ayant pour effet de produire des phonèmes perçus représentant toutes les composantes des données de parole numériques, ledit dispositif de mémoire (106; 213) étant prévu de manière que des données numériques de parole codées qui y sont mémorisées comprennent des codes de phonèmes représentant plusieurs phonèmes de référence respectifs et avec des données de parole numériques qui y sont mémorisées représentant des ensembles de caractéristiques d'allophones, chaque ensemble étant associé avec un phonème de référence, caractérisé en ce qu'un dispositif (108) est prévu, couplé en fonctionnement avec ledit dispositif comparateur (105; 212) et ledit dispositif de mémoire (106; 213) pour former une séquence de codes de phonèmes de plusieurs desdits codes de phonèmes, en ce que ledit dispositif (108) de formation de séquence de codes de phonèmes communique en fonction d'une séquence de codes de phonèmes respective de données de parole numériques de codes de phonèmes connectées représentant une caractéristique d'allophones specifique dans le groupe de caractéristiques d'allophones du dispositif de mémoire (106; 213) avec un dispositif synthétiseur de parole (114) et en ce que le dispositif synthétiseur (114) est connecté avec un dispositif audible (115) pour convertir le signal analogique de parole du dispositif synthétiseur en une parole synthétisée audible.

2. Système selon la revendication 1, caractérisé en outre en ce que ledit dispositif de conversion (104) comporte un dispositif (209) destiné à déterminer une série de formants à partir du signal analogique de parole et situé dans un spectre de fréquence dans lequel se trouve le signal analogique de parole, un dispositif couplé en fonctionnement avec ledit dispositif de détermination de formants pour produire des indices représentant lesdites séries de formants correspondant audit signal analogique de parole, et un dispositif (211) destiné à identifier des limites de phonèmes dans ledit signal analogique de parole et à identifier un ensemble desdits indices représentant lesdits formants dans chacune des limites de phonèmes identifiées et en ce que le dispositif comparateur (212) est couplé en fonctionnement avec ledit dispositif d'identification de limite de phonèmes (211) et ledit dispositif de mémoire (213) pour établir une adaptation optimale entre ledit ensemble d'indices et lesdits plusieurs phonèmes de référence tels que représentés par les données de parole numériques codées mémorisées dans ledit dispositif de mémoire (213) en sélectionnant l'un desdits plusieurs phonèmes de référence le mieux caractérisé par ledit ensemble d'indices pour produire un code de phonème représentant ladite adaptation optimale.

3. Système selon selon la revendication 1 ou 2, caractérisé en outre par un circuit de commande de gain automatique (208) destiné à recevoir une entrée de parole analogique et à produire ledit signal analogique de parole, ledit circuit de commande de gain automatique (208) étant couplé en fonctionnement avec l'entrée dudit dispositif de détermination de formants (209).

4. Procédé d'analyse d'un signal de parole et de production d'une parole synthétisée audible consistant: à produire un signal analogique de parole, à identifier des composantes de phonèmes dudit signal analogique de parole, à comparer chacune des composantes de phonèmes avec plusieurs phonèmes de référence, à obtenir l'adaptation la plus proche desdits plusieurs phonèmes de référence avec chacune des composantes de phonèmes identifiées, à produire des

signaux analogiques représentant une parole synthétisée basée sur l'adaptation la plus proche des phonèmes de référence, à produire une parole synthétisée audible à partir desdits signaux analogiques et à produire plusieurs codes de phonèmes représentant les phonèmes de référence comprenant tous les phonèmes reconnus dans un langage parlé donné pour une comparaison avec chacune des composantes de phonèmes identifiées à partir du signal analogique de parole pour produire des codes de phonèmes respectifs comme leur plus proche adaptation, caractérisé par la formation d'une séquence de codes de phonèmes de codes de phonèmes associés, de traduction de chaque séquence de codes de phonèmes formés en une caractéristique d'allophones respective sélectionnée dans un ensemble de caractéristiques d'allophones appartenant à un phonème spécifique et de production desdits signaux analogiques représentant la parole synthétisée au moyen de la séquence d'allophones traduits.

5. Procédé selon la revendication 4, caractérisé en outre en ce que l'identification des composantes de phonèmes dudit signal analogique de parole consiste à déterminer une série de formants dans ledit signal analogique, à produire des indices représentant ladite série de formants, à identifier des limites de phonèmes dans ledit signal analogique et à identifier un ensemble d'indices dans chacune desdites limites de phonèmes pour définir des composantes de phonèmes respectives dudit signal analogique de parole par des ensembles respectifs d'indices.

6. Procédé selon la revendication 5, caractérisé en outre en ce qu'il consiste à soumettre ledit signal analogique à une commande automatique de gain avant la détermination de ladite série de formants.

7. Procédé selon la revendication 5 ou 6, caractérisé en outre en ce que l'établissement de l'adaptation la plus proche consiste à affecter une valeur de comparaison pour chaque phonème de référence par rapport audit ensemble d'indices, et à choisir la valeur de comparaison optimale parmi les plusieurs valeurs de comparaison affectée, en sélectionnant ainsi le phonème de référence le plus proche caractérisé par ledit ensemble d'indices.

8. Procédé selon l'une quelconque des revendications 4 à 7, caractérisé en outre en ce qu'il consiste à normaliser ledit signal analogique de parole en établissant la hauteur et la vitesse en fonction de la voix d'un utilisateur avant l'identification desdites composantes de phonèmes de manière que la traduction suivante de ladite séquence de codes de phonèmes en ladite séquence d'allophones permette que la parole synthétisée audible produite soit une approximation la plus proche du signal de parole analogique original.

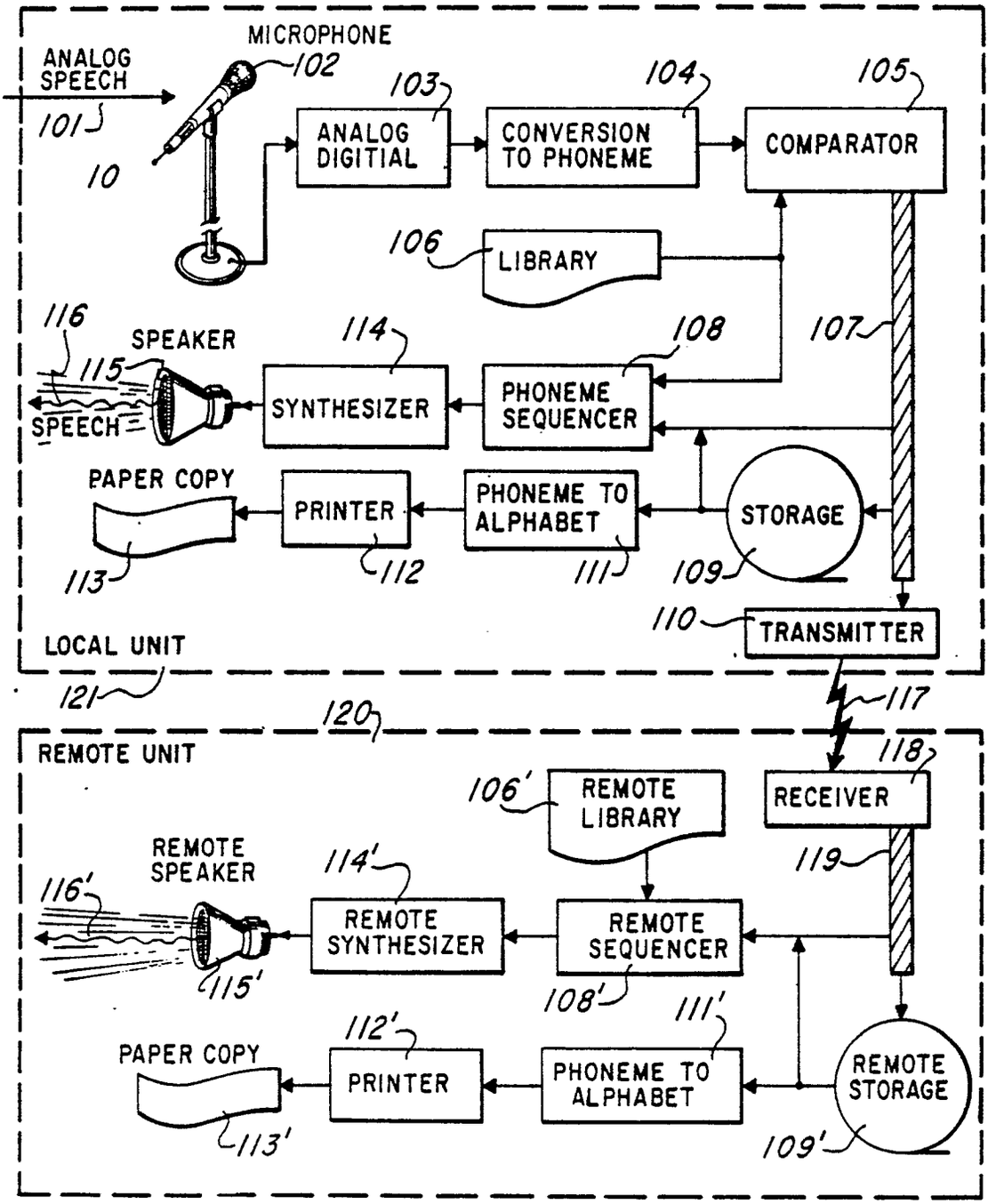

MICROPHONE
ANALOG SPEECH
101
102
10
103
ANALOG DIGITIAL
104
CONVERSION TO PHONEME
105
COMPARATOR
106
LIBRARY
107
116
SPEAKER
115
SPEECH
114
SYNTHESIZER
108
PHONEME SEQUENCER
PAPER COPY
113
PRINTER
112
PHONEME TO ALPHABET
111
STORAGE
109
110
TRANSMITTER
LOCAL UNIT
121
120
117
REMOTE UNIT
118
106'
REMOTE LIBRARY
RECEIVER
119
REMOTE SPEAKER
116'
114'
REMOTE SYNTHESIZER
REMOTE SEQUENCER
115'
108'
111'
112'
PAPER COPY
PRINTER
PHONEME TO ALPHABET
REMOTE STORAGE
113'
109'

Fig.1

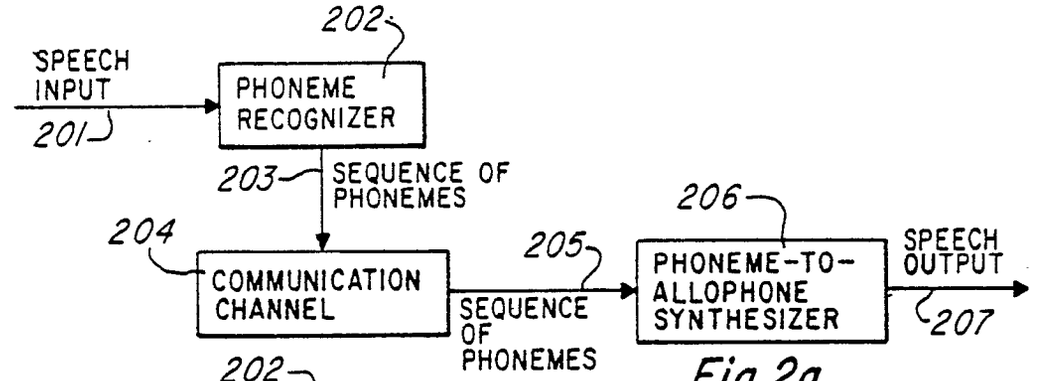


Fig. 2a

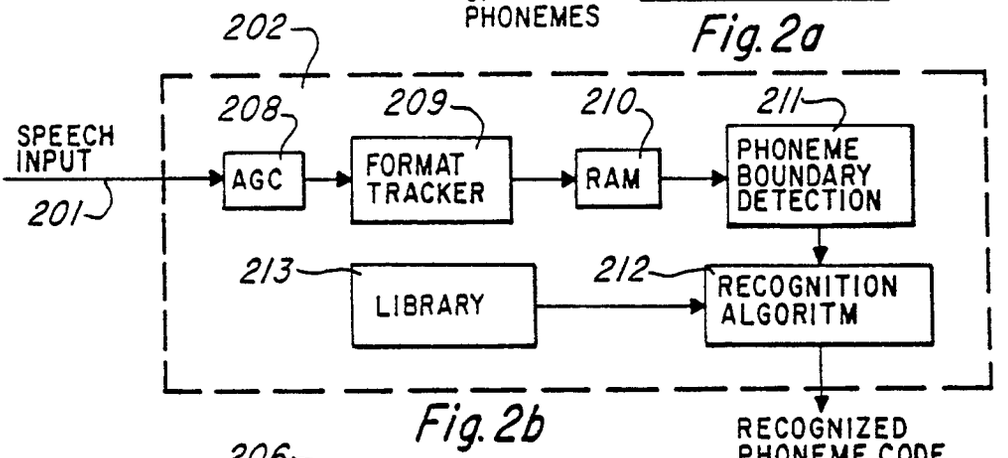


Fig. 2b

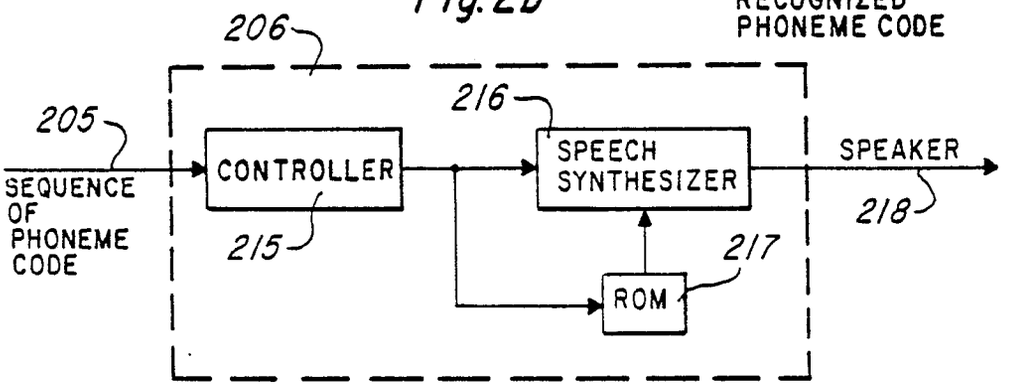


Fig. 2c

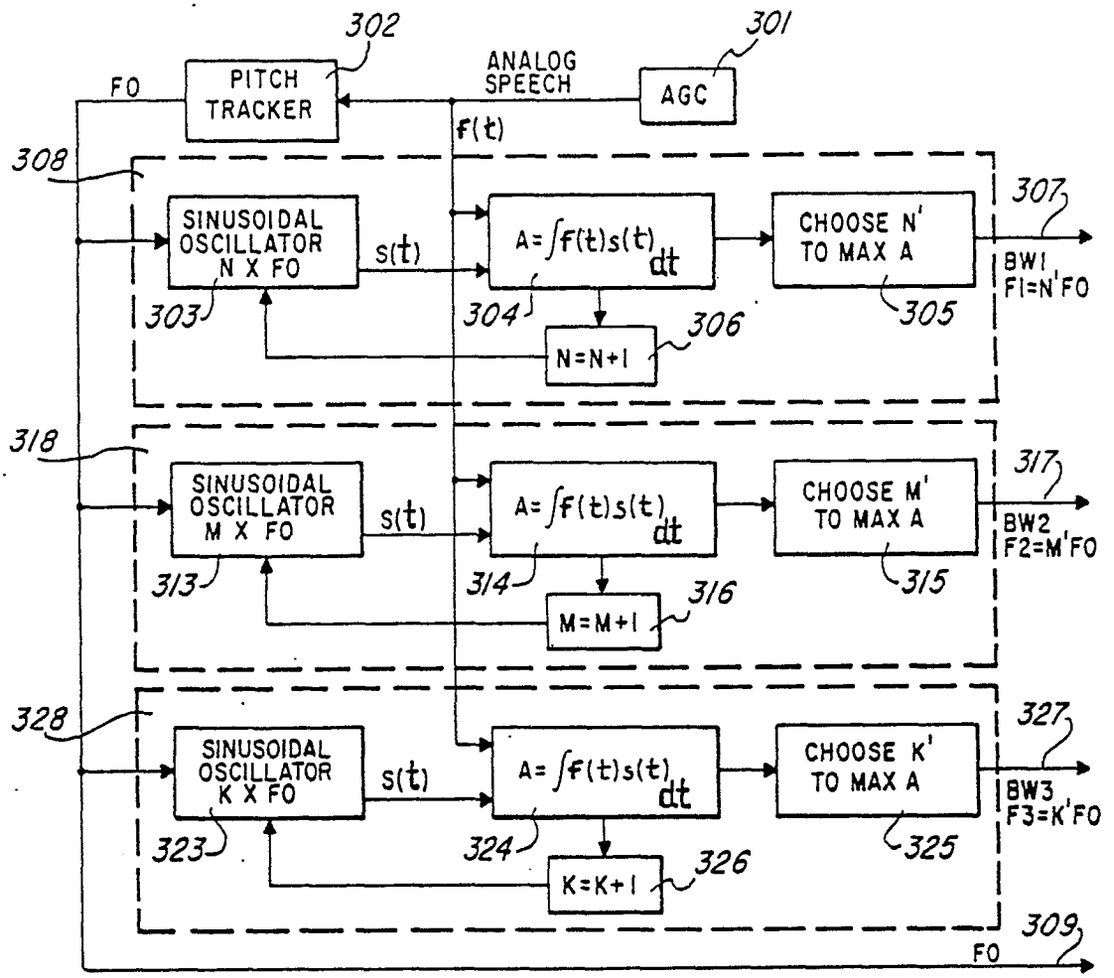

302
301
PITCH TRACKER
ANALOG SPEECH
AGC
F0
f(t)
308
SINUSOIDAL OSCILLATOR N X F0
s(t)
A=∫f(t)s(t)dt
CHOOSE N' TO MAX A
307
BW1
F1=N'F0
303
304
305
N=N+1
306
318
SINUSOIDAL OSCILLATOR M X F0
s(t)
A=∫f(t)s(t)dt
CHOOSE M' TO MAX A
317
BW2
F2=M'F0
313
314
315
M=M+1
316
328
SINUSOIDAL OSCILLATOR K X F0
s(t)
A=∫f(t)s(t)dt
CHOOSE K' TO MAX A
327
BW3
F3=K'F0
323
324
325
K=K+1
326
309
F0

Fig.3

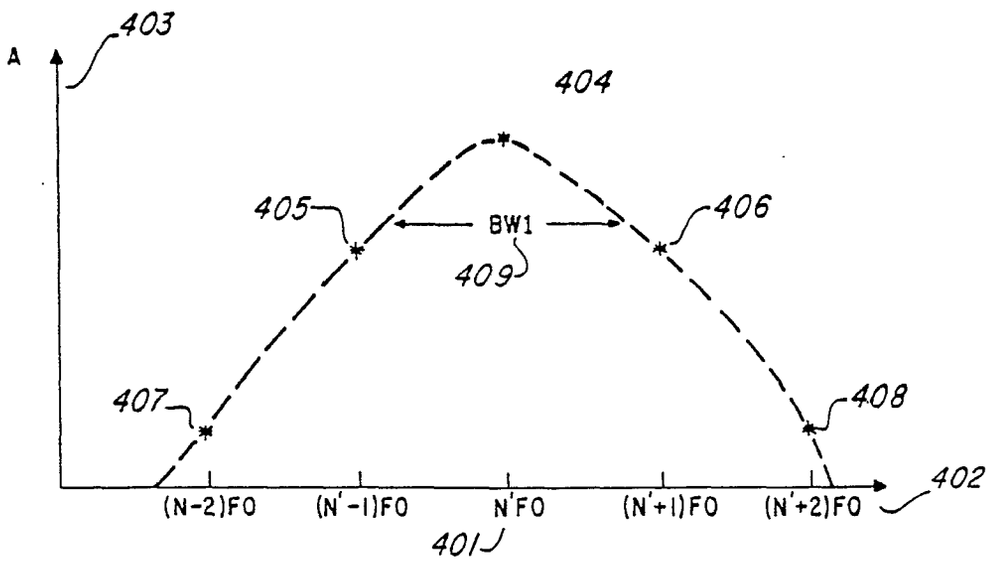

403
A
404
405
406
BW1
409
407
408
(N-2)F0
(N'-1)F0
N'F0
(N'+1)F0
(N'+2)F0
402
401

Fig.4

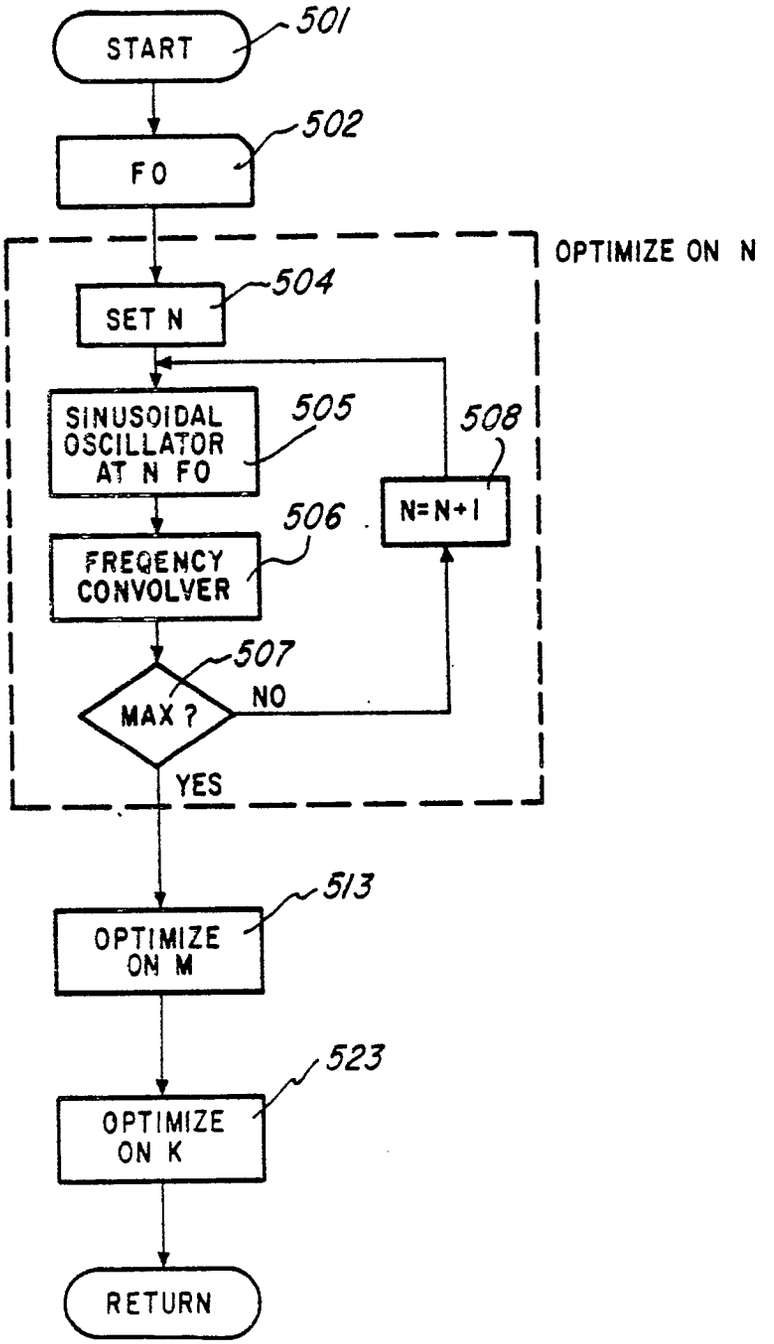

START
501
F0
502
OPTIMIZE ON N
SET N
504
SINUSOIDAL OSCILLATOR AT N F0
505
FREQENCY CONVOLVER
506
MAX ?
507
NO
N=N+1
508
YES
OPTIMIZE ON M
513
OPTIMIZE ON K
523
RETURN

Fig. 5

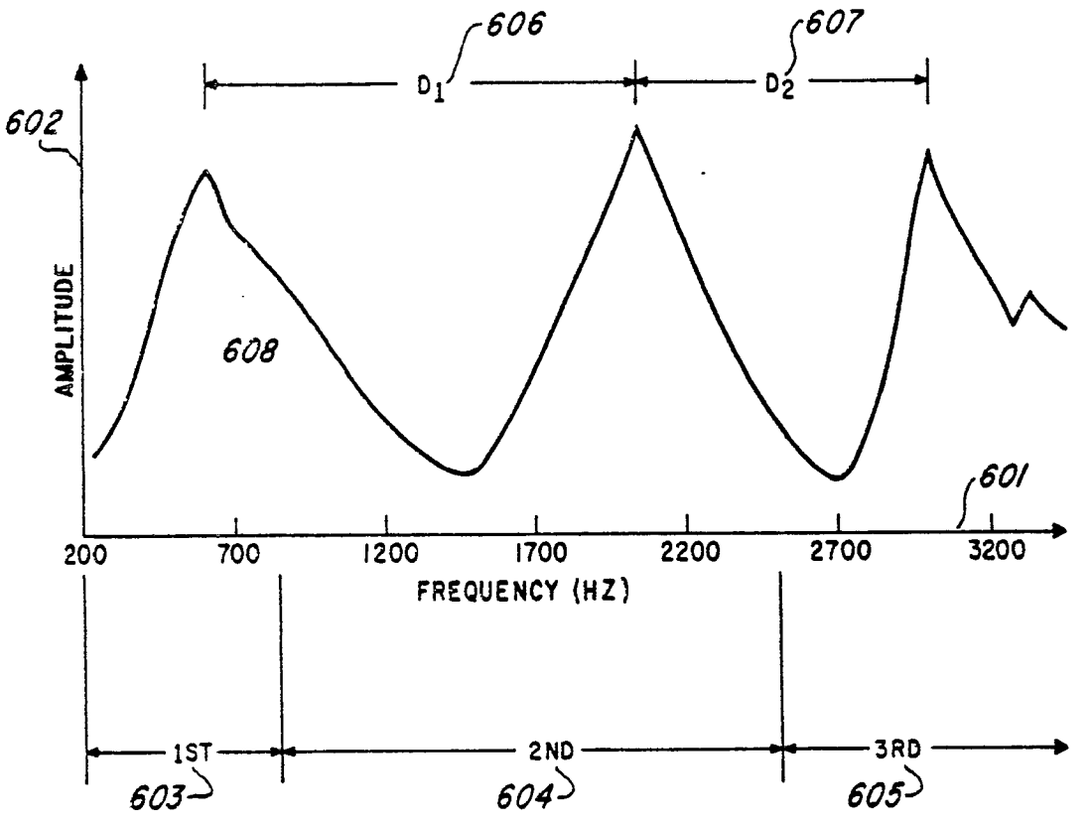
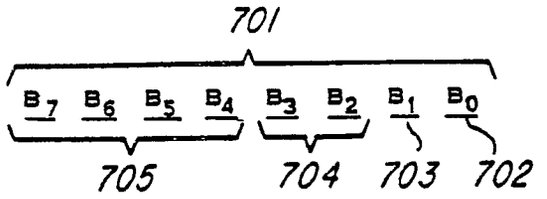

606
607
D1
D2
602
AMPLITUDE
608
601
200
700
1200
1700
2200
2700
3200
FREQUENCY (HZ)
1ST
2ND
3RD
603
604
605

Fig. 6



702 — B_0 = PAUSE/$\overline{\text{PAUSE}}$ { 1 – PAUSE; 0 – NO PAUSE }

703 — B_1 = VOICED/UNVOICED { 1 – VOICED; 0 – UNVOICED }

704 — $B_2 - B_3$ = SLOPE { 3 – NOT ASSIGNED; 2 – NEGATIVE; 1 – POSITIVE; 0 – LEVEL }

705 — $B_4 - B_7$ = FORMANT

706

VALUE	FIRST DISTANCE (HZ)	SECOND DISTANCE (HZ)
0	0–400	0–400
1	0–400	401–800
2	0–400	0–400
3	0–400	OVER 1200
4	401–800	0–400
5	401–800	401–800
6	401–800	801–1200
7	401–800	OVER 1200
8	801–1200	0–400
9	801–1200	401–800
10	801–1200	801–1200
11	801–1200	OVER 1200
12	OVER 1201	0–400
13	OVER 1201	401–800
14	OVER 1201	801–1200
15	OVER 1201	OVER 1200

Fig.7

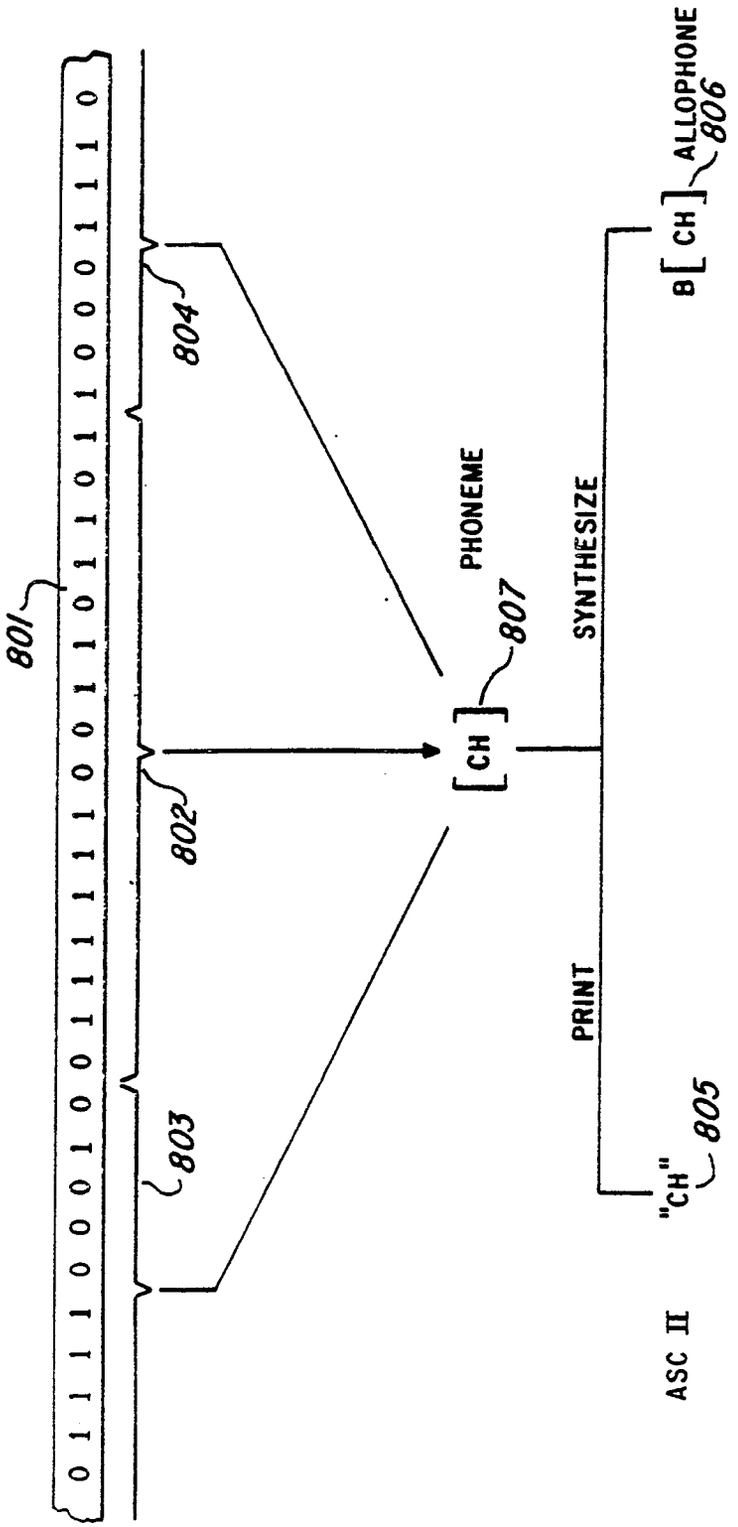

0 1 1 1 1 0 0 0 1 0 0 1 1 1 1 1 1 1 0 0 1 1 0 1 1 0 0 0 1 1 1 1 0
801
802
803
804
[CH]
807
PHONEME
SYNTHESIZE
PRINT
ASC II
"CH"
805
B [CH]
ALLOPHONE
806

Fig. 8

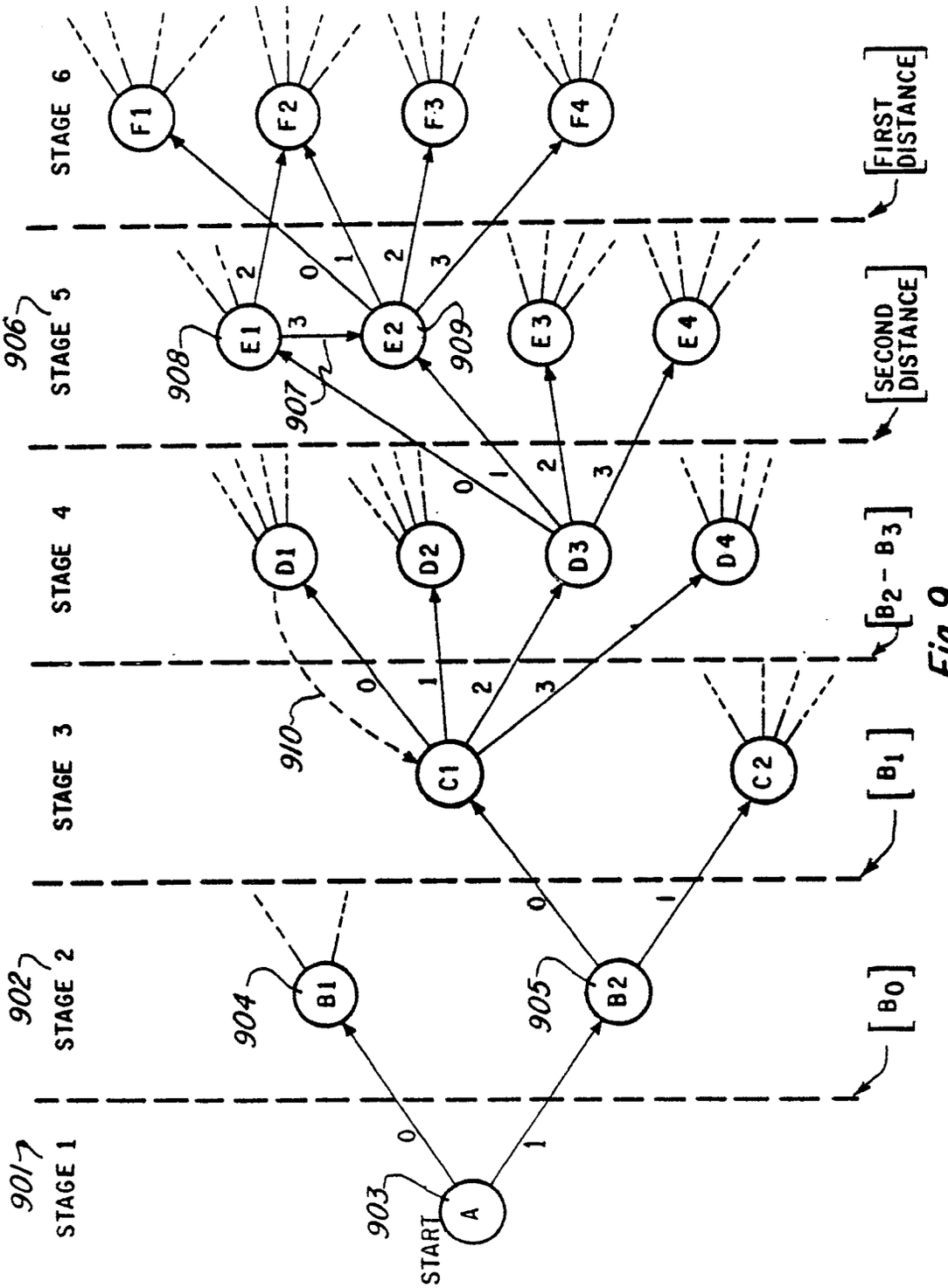

STAGE 1
STAGE 2
STAGE 3
STAGE 4
STAGE 5
STAGE 6
901
902
906
START
A
B1
B2
C1
C2
D1
D2
D3
D4
E1
E2
E3
E4
F1
F2
F3
F4
903
904
905
907
908
909
910
[B0]
[B1]
[B2 - B3]
[SECOND DISTANCE]
[FIRST DISTANCE]

Fig. 9

PHONEME-TO-ALLOPHONE RULES

PHONEME TO ALLOPHONE RULES FOR PRONOUNCING DICTIONARY.

LEGEND:

1001 ~ B = BLANK, WORD BOUNDARY
$ = NASAL CONSONANT M:N:NG
% = ONE VOICED CONSONANT :B: D: G: J: L: M: N: NG: R: THV: V: W: Y: Z: ZH:
! = ONE VOICELESS CONSONANT CH F H K P S SH T THF
^ = ONE CONSONANT
: = ZERO OR MORE CONSONANTS
/ = PRIMARY OR SECONDARY STRESS:^:—:
; = ONE LABIAL CONSONANT :B: F: M: P: V: W:
@ = ONE VELAR CONSONANT :G: K: NG:
* = ONE PALATAL/ALVEALAR CONSONANT :CH: D: J: N: S: SH: T: Y: Z: ZH:
+ = ONE HIGH-FRONT VOWEL :EE: EH: EI: I:
& = ONE CENTRAL VOWEL :AH: ER: UH:
# = ONE BACK-ROUND VOWEL :AW: OO: OW: U:
1002 ~ · = SYLLABLE BOUNDARY :-:^:-:

Fig. 10a

B RULES

B[BR]	= /BR/	AS IN /BROWN/
· [BR]	= /BR/	AS IN /ABRADE/
B[BL]	= /BL/	AS IN /BLACK/
· [BL]	= /BL/	AS IN /OBLIGE/
B[B]	= /B/	AS IN /BIKE/
[B]B	= /B#/	AS IN /GRAB/
[B]	= /B*/	AS IN /ABNIGATE/

1003

CH RULES ~

1004 ~ B[CH]	= /CH-/	AS IN /CHAIN/
[CH]B	= /CH#/	AS IN /ITCH/
1005 ~ [CH]	= /CH*/	AS IN /BEWITCHING/

D RULES

B[DR]	= /DR/	AS IN /DRIVE/
· [DR]	= /DR/	AS IN /REDRESS/
B[D]	= /D-/	AS IN /DIVE/
[D]B	= /D#/	AS IN /RIDE/
[D], UH	= /D!/	AS IN /LADDER/
[D]/I	= /D!/	AS IN /RIDING/
[D]/L	= /D!/	AS IN /LADLE/
[D]/N	= /D!/	AS IN /LADEN/
'D	= /D*/	AS IN /LADY/

Fig. 10b