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

(11) Publication number: **0 138 954**
B1

(12) # EUROPEAN PATENT SPECIFICATION

(45) Date of publication of patent specification: **26.10.88**

(51) Int. Cl.[4]: **G 10 L 5/00**

(21) Application number: **84901491.5**

(22) Date of filing: **12.03.84**

(86) International application number:
PCT/US84/00367

(87) International publication number:
WO 84/04194 25.10.84 Gazette 84/25

(54) **SPEECH PATTERN PROCESSING UTILIZING SPEECH PATTERN COMPRESSION.**

(30) Priority: **12.04.83 US 484231**

(43) Date of publication of application:
02.05.85 Bulletin 85/18

(45) Publication of the grant of the patent:
26.10.88 Bulletin 88/43

(84) Designated Contracting States:
DE FR GB

(56) References cited:
US-A-3 598 921
US-A-3 715 512
US-A-4 216 354
US-A-4 280 192

IEEE TRANSACTIONS ON ACOUSTICS SPEECH AND SIGNAL PROCESSING, vol. ASSP-30, no. 5, October 1982, pages 770-780, IEEE, New York, US; D.Y. WONG et al.: "An 800 Bit/s vector quantization LPC vocoder"

ICASSP '81, IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING, vol. 2, 30th-31st March - 1st April 1981, Atlanta, Georgia, US, pages 804-807, IEEE, New York, US; P. MABILLEAU et al.: "Medium band speech coding using a dictionary of waveforms"

(73) Proprietor: **AMERICAN TELEPHONE AND TELEGRAPH COMPANY**
550 Madison Avenue
New York, NY 10022 (US)

(72) Inventor: **ATAL, Bishnu, Saroop**
138 Knollwood Drive
Murray Hill, NJ 07974 (US)

(74) Representative: **Johnston, Kenneth Graham et al**
AT&T (UK) LTD. AT&T Intellectual Property Division 5 Mornington Road
Woodford Green Essex, IG8 OTU (GB)

(56) References cited:
ICASSP '82, IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING, vol. 3, 3rd-5th May 1982, Paris, FR, pages 1565-1569, IEEE, New York, US; S. ROUCOS et al.: "Segment quantization for very-low-rate speech coding"
IEEE TRANSACTIONS ON COMMUNICATIONS, vol. COM-30, no. 4, April 1982, pages 674-686, IEEE, New York, US; V.R. VISWANATHAN et al.: "Variable frame rate transmission: a review of methodology and application to narrow-band LPC speech coding"

Courier Press, Leamington Spa, England.

EP 0 138 954 B1

(56) References cited:
ICASSP '83, IEEE INTERNATIONAL CONFERENCE ON ACOUSTICS, SPEECH AND SIGNAL PROCESSING, vol. 1, 14th-16th April 1983, Boston, Massachusetts, US, pages 81-84, IEEE, New York, US; B.S. ATAL: "Efficient coding of LPC parameters by temporal decomposition"

Description

This invention relates to speech processing and, particularly, to the compression of speech patterns and to the synthesis of speech patterns from such compressed patterns.

It is generally accepted that a speech signal requires a bandwidth of at least 4 kHz for reasonable intelligibility. In digital speech processing systems such as speech synthesizers, recognizers, or coders, the channel capacity needed for transmission or memory required for storage of the digital elements of the full 4 kHz bandwidth waveform is very large. Many techniques have been devised to reduce the number of digital codes needed to represent a speech signal. Waveform coding such as Pulse Code Modulation (PCM), Differential Pulse Code Modulation (DPCM), Delta Modulation or adaptive predictive coding result in natural sounding, high quality speech at bit rates between 16 and 64 kbps. The speech quality obtained from waveform coders, however, degrades as the bit rate is reduced below 16 kbps.

An alternative speech coding technique disclosed in U.S. Patent 3,624,302 utilizes a small number, e.g., 12—16, of slowly varying parameters which may be processed to produce a low distortion replica of a speech pattern. Such parameters, e.g., Linear Prediction Coefficient (LPC) or log area, generated by linear prediction analysis can be spectrum limited to 50 Hz without significant band limiting distortion. Encoding of the LPC or log area parameters generally requires sampling at a rate of twice the bandwidth and quantizing each resulting frame of log area parameters. Each frame of a log area parameter can be quantized using 48 bits. Consequently, 12 log area parameters each having a 50 Hz bandwidth results in a total bit rate of 4800 bits/sec.

Further reduction of bandwidth decreases the bit rate, but the resulting increase in distortion interferes with the intelligibility of speech synthesized from the lower bandwidth parameters. It is well known that sounds in speech patterns do not occur at a uniform rate and techniques have been devised to take into account such nonuniform occurrences. U.S. Patent 4,349,700 discloses arrangements that permit recognition of speech patterns having diverse sound patterns utilizing dynamic programming. U.S. Patent 4,038,503 discloses a technique for nonlinear warping of time intervals of speech patterns so that the sound features are represented in a more uniform manner. These arrangements, however, require storing and processing acoustic feature signals that are sampled at a rate corresponding to the most rapidly changing feature in the pattern. It is an object of the invention to provide an improved speech representation and/or speech synthesis arrangements having reduced digital storage and processing requirements.

In IEEE Transactions on Communications, vol COM-30, No 4, April 82, pages 674—686, Viswanathan and others disclose a method as set out in the preamble of claim 1. A speech pattern is analysed by linear prediction encoding and the LPC parameters are transmitted only when their values have changed sufficiently over the interval since their preceding transmission.

In the invention as set out in the claims high compression efficiency is combined with accurate reproduction by a coding procedure based on the individual sounds constituting the speech pattern and determined at the centroids of the individual sounds.

Description of the drawing

Fig. 1 depicts a flowchart illustrating the general method of the invention;

Fig. 2 depicts a block diagram of a speech pattern coding circuit illustrative of the invention;

Figs. 3—8 depict detailed flowcharts illustrating the operation of the circuit of Fig. 2;

Fig. 9 depicts a speech synthesizer illustrative of the invention;

Fig. 10 depicts a flow chart illustrating the operation of the circuit of Fig. 9;

Fig. 11 shows a waveform illustrating a speech event timing signal obtained in the circuit of Fig. 2; and

Fig. 12 shows waveforms illustrative of a speech pattern and the speech event feature signals associated therewith.

General description

It is well known in the art to represent a speech pattern by a sequence of acoustic feature signals derived from a linear prediction or other spectral analysis. Log area parameter signals sampled at closely spaced time intervals have been used in speech synthesis to obtain efficient representation of a speech pattern. In accordance with the invention, log area parameters are transformed into a sequence of individual sound or speech event feature signals $\Phi_k(n)$ such that the log area parameters

$$y_i(n) = \sum_{k=1}^{m} a_{ik}\Phi_k(n)$$

$$1 \leq n \leq N \quad 1 \leq i \leq p$$

The speech event feature signals $\Phi_k(n)$ are sequential and occur at the speech event rate of the pattern which is substantially lower than the the log area parameter frame rate. In equation (1), p is the total number of log area parameters $y_i(n)$ determined by linear prediction analysis. m corresponds to the number of speech events in the pattern, n is the index of samples in the speech pattern at the sampling rate

of the log area parameters, $\phi_k(n)$ is the kth speech event signal at sampling instant n, and a_{ik} is a combining coefficient corresponding to the contribution of the kth speech event function to the ith log area parameter. Equation (1) may be expressed in matrix form as

$$Y=A\Phi \tag{2}$$

where Y is a p×N matrix whose (i,n) element is $y_i(n)$, A is a p×m matrix whose (i,k) element is a_{ik}, and Φ is an m×N matrix whose (k,n) element is $\phi_k(n)$. Since each speech event k occupies only a small segment of the speech pattern, the signal $\phi_k(n)$ representative thereof should be non-zero over only a small range of the sampling intervals of the total pattern. Each log area parameter $y_i(n)$ in equation (1) is a linear combination of the speech event functions $\phi_k(n)$ and the bandwidth of each $y_i(n)$ parameter is the maximum bandwidth of any one of the speech event functions $\phi_k(n)$. It is therefore readily seen that the direct coding of $y_i(n)$ signals will take more bits than the coding of the $\phi_k(n)$ switch event signals and the combining coefficient signals a_{ik} in equation (1).

Fig. 1 shows a flow chart illustrative of the general method of the invention. In accordance with the invention, a speech pattern is analyzed to form a sequence of signals representative of log area parameter acoustic feature signals. It is to be understood, however, that LPC, Partial Autocorrelation (PARCOR) or other speech features (see, e.g., U.S. patent 3,624,302) may be used instead of log area parameters. The feature signals are then converted into a set of speech event representative signals that are encoded at a lower bit rate for transmission or storage.

With reference to Fig. 1, box 101 is entered in which an electrical signal corresponding to a speech pattern is low pass filtered to remove unwanted higher frequency noise and speech components and the filtered signal is sampled at twice the low pass filtering cutoff frequency. The speech pattern samples are then converted into a sequence of digitally coded signals corresponding to the pattern as per box 110. Since the storage required for the sample signals is too large for most practical applications, they are utilized to generate log area parameter signals as per box 120 by linear prediction techniques well known in the art. The log area parameter signals $y_i(n)$ are produced at a constant sampling rate high enough to accurately represent the fastest expected event in the speech pattern. Typically, a sampling interval between two and five milliseconds is selected.

After the log area parameter signals are stored, the times of occurrence of the successive speech events in the pattern are detected and signals representative of the event timing are generated and stored as per box 130. This is done by partitioning the pattern into prescribed smaller segments, e.g., 0.25 second intervals. For each successive interval having a beginning frame n_b and an ending frame n_e, a matrix of log area parameter signals is formed corresponding to the log area parameters $y_i(n)$ of the segment. The redundancy in the matrix is reduced by factoring out the first four principal components so that

$$y_i(n)=\sum_{m=1}^{4}[c_{im}u_m(n)] \tag{3}$$

and

$$u_m(n)=\sum_{i=1}^{p}[\beta_{im}y_i(n)] \tag{4}$$

The first four principal components may be obtained by methods well known in the art such as described in the article "An Efficient Linear Prediction Vocoder" by M. R. Sambur appearing in the Bell System Technical Journal Vol. 54, No. 10, pp. 1693—1723, December 1975. The resulting $u_m(n)$ functions may be linearly combined to define the desired speech event signals as

$$\phi_k(n)=\sum_{m=1}^{4}[b_{km}u_m(n)] \tag{5}$$

by selecting coefficients b_{km} such that each $\phi_k(n)$ are most compact in time. In this way, the speech pattern is represented by a sequence of successive compact (minimum spreading) speech event feature signals $\phi_k(n)$ each of which can be efficiently coded. In order to obtain the shapes and locations of the speech event signals, a distance measure

$$\theta(L)=[\Sigma_n(n-L)^2\phi^2(n)/\Sigma_n\phi^2(n)]^{1/2} \tag{6}$$

is minimized to choose the optimum $\phi(n)$ and its location is obtained from a speech event timing signal

$$v(L)=[\Sigma_n(n-L)\phi^2(n)]/\Sigma_n\phi^2(n) \tag{7}$$

In terms of equations 5, 6, and 7, a speech event signal $\phi_k(n)$ with minimum spreading is centered at each negative zero crossing of v(L).

Subsequent to the generation of the v(L) signals in box 130, box 140 is entered and the speech event signals $\phi_k(n)$ are accurately determined using the process of box 130 with the speech event occurrence signals from the negative going zero crossings of v(L). Having generated the sequence of speech event representative signals, the combining coefficients a_{ik} in equations (1) and (2) may be generated by minimizing the mean-squared error

$$E=\sum_n[y_i(n)-\sum_{k=1}^{M} a_{ik}\phi_k(n)]^2 \qquad (8)$$

where M is the total number of speech events within the range of index *n* over which the sum is performed. The partial derivatives of E with respect to the coefficients a_{ik} are set equal to zero and the coefficients a_{ik} are obtained from the set of simultaneous linear equations

$$\sum_{k=1}^{M} a_{ik}\sum_n \phi_k(n)\phi_r(n)=\sum_n y_i(n)\phi_r(n) \qquad (9)$$

$$1\leq r\leq M$$

$$1\leq i\leq P$$

Detailed description

Fig. 2 shows a speech coding arrangement that includes electroacoustic transducer 201, filter and sampler circuit 203, analog to digital converter 205, and speech sample store 210 which cooperate to convert a speech pattern into a stored sequence of digital codes representative of the pattern. Central processor 275 may comprise a microprocessor such as the Motorola type MC68000 controlled by permanently stored instructions in read only memories (ROM) 215, 220, 225, 230 and 235. Processor 275 is adapted to direct the operations of arithmetic processor 280, and stores 210, 240, 245, 250, 255 and 260 so that the digital codes from store 210 are compressed into a compact set of speech event feature signals. The speech event feature signals are then supplied to utilization device 285 via input output interface 265. The utilization device may be a digital communication facility or a storage arrangement for delayed transmission or a store associated with a speech synthesizer. The Motorola MC68000 integrated circuit is described in the publication *MC68000 16 Bit Microprocessor User's Manual,* second edition, Motorola, Inc., 1980 and arithmetic processor 280 may comprise the TRW type MPY-16HJ integrated circuit.

Referring to Fig. 2, a speech pattern is applied to electroacoustic transducer 201 and the electrical signal therefrom is supplied to low pass filter and sampler circuit 203 which is operative to limit the upper end of the signal bandwidth to 3.5 KHz and to sample the filtered signal at an 8 KHz rate. Analog to digital converter 205 converts the sampled signal from filter and sampler 203 into a sequence of digital codes, each representative of the magnitude of a signal sample. The resulting digital codes are sequentially stored in speech sample store 210.

Subsequent to the storage of the sampled speech pattern codes in store 210, central processor 275 causes the instructions stored in log area parameter program store 215 to be transferred to the random access memory associated with the central processor. The flow chart of Fig. 3 illustrates the sequence of operations performed by the controller responsive to the instructions from store 215.

Referring to Fig. 3, box 305 is initially entered and frame count index *n* is reset to 1. The speech samples of the current frame are then transferred from store 210 to arithmetic processor 280 via central processor 275 as per box 310. The occurrence of an end of speech sample signal is checked in decision box 315. Until the detection of the end of speech pattern signal, control is passed to box 325 and an LPC analysis is performed for the frame in processors 275 and 280. The LPC parameter signals of the current frame are then converted to log area parameter signals $y_i(k)$ as per box 330 and the log area parameter signals are stored in log area parameter store 240 (box 335). The frame count is incremented by one in box 345 and the speech samples of the next frame are read (box 310). When the end of speech pattern signal occurs, control is passed to box 320 and a signal corresponding to the number of frames in the pattern is stored in processor 275.

Central processor 275 is operative after the log area parameter storing operation is completed to transfer the stored instructions of ROM 220 into its random access memory. The instruction codes from store 220 correspond to the operations illustrated in the flow chart of Figs. 4 and 5. These instruction codes are effective to generate a signal v(L) from which the occurrences of the speech events in the speech pattern may be detected and located.

Referring to Fig. 4, the frame count of the log area parameters is initially reset in processor 275 as per box 403 and the log area parameters $y_i(n)$ for an initial time interval n_1 to n_2 of the speech pattern are transferred from log area parameter store 240 to processor 275 (box 410). After determining whether the

end of the speech pattern has been reached in decision box 415, box 420 is entered and the redundancy of the log area parameter signals is removed by factoring out the first four principal components $u_i(n)$, i=1,...,4 as aforementioned.

The log area parameters of the current time interval are then represented by

$$y_i(n)= \sum_{m=1}^{4} c_{im}u_m(n) \tag{10}$$

from which a set of signals

$$u_m(n)= \sum_{i=1}^{16} \beta_{im}y_i(n) \tag{11}$$

are to be obtained. The $u_i(n)$ signals over the interval may be combined through use of parameters b_i, i=1,...,4, in box 425 so that a set of signals

$$\phi_k(n)= \sum_{m=1} [b_{km}u_m(n)] \tag{12}$$

are produced such that ϕ_k is most compact over the range n_1 to n_2. This is accomplished through use of the $\theta(L)$ function of equation 6. A signal v(L) representative of the speech event timing of the speech pattern is then formed in accordance with equation 7 in box 430 and the v(L) signal is stored in timing parameter store 245. Frame counter *n* is incremented by a constant value, e.g., 5, on the basis of how close adjacent speech event signals $\phi_k(n)$ are expected to occur (box 435) and box 410 is reentered to generate the $\phi_k(n)$ and v(L) signals for the next time interval of the speech pattern.

When the end of the speech pattern is detected in decision box 415, the frame count of the speech pattern is stored (box 440) and the generation of the speech event timing parameter signal for the speech pattern is completed. Fig. 11 illustrates the speech event timing parameter signal for the an utterance exemplary message. Each negative going zero crossing in Fig. 11 corresponds to the centroid of a speech event feature signal $\phi_k(n)$.

Referring to Fig. 5, box 501 is entered in which speech event index l is reset to zero and frame index *n* is again reset to one. After indices l and *n* are initialized, the successive frames of speech event timing parameter signal are read from store 245 (box 505) and zero crossings therein are detected in processor 275 as per box 510. Whenever a zero crossing is found, the speech event index l is incremented (box 515) and the speech event location frame is stored in speech event location store 250 (box 520). The frame index *n* is then incremented in box 525 and a check is made for the end of the speech pattern frames in box 530. Until the end of speech pattern frames signal is detected, box 505 is reentered from box 530 after each iteration to detect the subsequent speech event location frames of the pattern.

Upon detection of end of the speech pattern signal in box 530, central processor 235 addresses speech event feature signal generation program store 225 and causes its contents to be transferred to the processor. Central processor 275 and arithmetic processor 280 are thereby adapted to form a sequence of speech event feature signals $\phi_k(n)$ responsive to the log area parameter signals in store 240 and the speech event location signals in store 250. The speech event feature signal generation instructions are illustrated in the flow chart of Fig. 6.

Initially, location index l is set to one as per box 601 and the locations of the speech events in store 250 are transferred to central processor 275 (box 605). As per box 610, the limit frames for a prescribed number of speech event locations, e.g., 5, are determined. The log area parameters for the speech pattern interval defined by the limit frames are read from store 240 and are placed in a section of the memory of central processor 275 (box 615). The redundancy in the log area parameters is removed by factoring out the number of principal components therein corresponding to the number of prescribed number of events (box 620). Immediately thereafter, the speech event feature signal $\phi_L(n)$ for the current location L is generated.

The minimization of equation (6) to determine $\phi_L(n)$ is accomplished by forming the derivative

$$\frac{\partial 1n\theta(L)}{\partial^b r}=\tfrac{1}{2}2\left[\sum_{n=n_1}^{n_2}(n-L)^2\phi(n)\frac{\partial\phi(n)}{\partial b_r}\Big/\sum_{n=n_1}^{n_2}(n-L)^2\phi^2(n)\right] \tag{13}$$

where

$$\phi(n)= \sum_{i=1}^{m} b_iu_i(n) \tag{14}$$

m is the prescribed number of speech events and r can be either 1, 2,..., or m. The derivative of equation (13) is set equal to zero to determine the minimum and

$$\sum_{n=n_1}^{n_2}(n-L)^2\phi(n)\frac{\partial\phi(n)}{\partial b_r}\Big/\sum_{n=n_1}^{n_2}(n-L)^2\phi^2(n)$$

$$=\sum_{n=n_1}^{n_2}\phi(n)\frac{\partial\phi(n)}{\partial b_r}\Big/\sum_{n=n_1}^{n_2}\phi^2(n) \qquad (15)$$

is obtained. From equation (14)

$$\frac{\partial\phi(n)}{\partial b_r}=u_r(n) \qquad (16)$$

so that equation (15) can be changed to

$$\sum_{n=n_1}^{n_2}(n-L)^2\phi(n)u_r(n)$$

$$=\left[\sum_{n=n_1}^{n_2}(n-L)^2\phi^2(n)\Big/\sum_{n=1}^{n_2}\phi^2(n)\right]\left[\sum_{n=n_1}^{n_2}\phi(n)u_r(n)\right] \qquad (17)$$

$\phi(n)$ in equation (17) can be replaced by the right side of equation 14. Thus,

$$\sum_{n=n_1}^{n_2}(n-L)^2\sum_{i=1}^{M}b_iu_i(n)u_r(n)$$

$$=\lambda\left[\sum_{n=1}^{N}\sum_{i=1}^{M}b_iu_i(n)u_r(n)\right] \qquad (18)$$

where

$$\lambda=\sum_{n=n_1}^{n_2}(n-L)^2\phi^2(n)\Big/\sum_{n=n_1}^{n_2}\phi^2(n)=\text{min. value }\theta(L) \qquad (19)$$

Rearranging equation (18) yields

$$\sum_{i=1}^{M}b_i\sum_{n=n_1}^{n_2}(n-L)^2u_i(n)u_r(n)$$

$$=\theta(L)\sum_{i=1}^{M}b_i\sum_{n=n_1}^{n_2}u_i(n)u_r(n) \qquad (20)$$

Since $u_i(n)$ is the principal component of matrix Y,

$$\sum_{n=n_1}^{n_2}u_i(n)u_r(n)=0 \qquad i\neq r$$

$$=1 \qquad i=r \qquad (21)$$

equation (20) can be simplified to

$$\sum_{i=1}^{M}b_iR_{ir}=b_r\theta(L) \qquad (22)$$

where

$$R_{ir}=\sum_{n=n_1}^{n_2}(n-L)^2u_i(n)u_r(n) \tag{23}$$

Equation (22) can be expressed in matrix notation as

$$\vec{Rb}=L\vec{b} \tag{24}$$

where

$$\lambda=\theta(L) \tag{25}$$

Equation 25 has exactly *m* solutions and the solution which minimizes $\theta(L)$ is the one for which λ is minimum. The coefficients $b_1, b_2,\ldots, b_m$ for which $\lambda=\theta(L)$ attains its minimum value results in the optimum speech event feature signal $\phi_L(n)$.

In Fig. 6, the speech event feature signal $\phi_L(n)$ is generated in box 625 and is stored in store 255. Until the end of the speech pattern is detected in decision box 635, the loop including boxes 605, 610, 615, 620, 625 and 630 is iterated so that the complete sequence of speech events for the speech pattern is formed.

Fig. 12 shows waveforms illustrating a speech pattern and the speech event feature signals generated therefrom in accordance with the invention. Waveform 1201 corresponds to a portion of a speech pattern and waveforms 1205-1 through 1205-n correspond to the sequence of speech event feature signals $\phi_L(n)$ obtained from the waveform in the circuit of Fig. 2. Each feature signal is representative of the acoustic characteristics of a speech event of the pattern of waveform 1201. The speech event feature signals may be combined with coefficients a_{lk} of equation 1 to reform log area parameter signals that are representative of the acoustic features of the speech pattern.

Upon completion of the operations shown in Fig. 6, the sequence of speech event feature signals for the speech pattern is stored in store 255. Each speech event feature signal $\phi_l(n)$ is encoded and transferred to utilization device 285 as illustrated in the flow chart of Fig. 7. Central processor is adapted to receive the speech event signal encoding program instruction set stored in ROM 235.

Referring to Fig. 7, the speech event index l is reset to one as per box 701 and the speech event feature signal $\phi_l(n)$ is read from store 255. The sampling rate R_l for the current speech event feature signal is selected in box 710 by one of the many methods well known in the art. For example, the instruction codes perform a Fourier analysis and generate a signal corresponding to the upper band limit of the feature signal from which a sampling rate signal R_l is determined. As is well known in the art, the sampling rate need only be sufficient to adequately represent the feature signal. Thus, a slowly changing feature signal may utilize a lower sampling rate than a rapidly changing feature signal and the sampling rate for each feature signal may be different.

Once a sampling rate signal has been determined for speech event feature signal $\phi_l(n)$, it is encoded at rate R_l as per box 715. Any of the well-known encoding schemes can be used. For example, each sample may be converted into a PCM, ADPCM or Δ modulated signal and concatenated with a signal indicative of the feature signal location in the speech pattern and a signal representative of the sampling rate R_l. The coded speech event feature signal is then transferred to utilization device 285 via input output interface 265. Speech event index l is then incremented (box 720) and decision box 725 is entered to determine if the last speech event signal has been coded. The loop including boxes 705 through 725 is iterated until the last speech event signal has been encoded ($l>l_F$) at which time the coding of the speech event feature signals is completed.

The speech event feature signals must be combined in accordance with equation 1 to form replicas of the log area feature signals therein. Accordingly, the combining coefficients for the speech pattern are generated and encoded as shown in the flow chart of Fig. 8. After the speech event feature signal encoding, central processor 275 is conditioned to read the contents of ROM 225. The instruction codes permanently stored in the ROM control the formation and encoding of the combining coefficients.

The combining coefficients are produced for the entire speech pattern by matrix processing in central processor 275 and arithmetic processor 280. Referring to Fig. 8, the log area parameters of the speech pattern are transferred to processor 275 as per box 801. A speech event feature signal coefficient matrix G is generated (box 805) in accordance with

$$g_{kr}=\sum_n\phi_k(n)\phi_r(n) \tag{26}$$

and a Y−ϕ correlation matrix C is formed (box 810) in accordance with

$$c_{lr}=\sum_n y_l(n)\phi_r(n) \tag{27}$$

The combining coefficient matrix is then produced as per box 815 according to the relationship

$$A=G^{-1}C \tag{28}$$

The elements of matrix A are the combining coefficients a_{lk} of equation 1. These combining coefficients are encoded, as is well known in the art, in box 820 and the encoded coefficients are transferred to utilization device 285.

In accordance with the invention, the linear predictive parameters sampled at a rate corresponding to the most rapid change therein are converted into a sequence of speech event feature signals that are encoded at the much lower speech event occurrence rate and the speech pattern is further compressed to reduce transmission and storage requirements without adversely affecting intelligibility. Utilization device 285 may be a communication facility connected to one of the many speech synthesizer circuits using an LPC all pole filter known in the art.

The circuit of Fig. 2 is adapted to compress a spoken message into a sequence of coded speech event feature signals which are transmitted via utilization device 285 to a synthesizer. In the synthesizer, the speech event feature signals and the combining coefficients of the message are decoded and recombined to form the message log area parameter signals. These log area parameter signals are then utilized to produce a replica of the original message.

Fig. 9 depicts a block diagram of a speech synthesizer circuit illustrative of the invention and Fig. 10 shows a flow chart illustrating its operation. Store 915 of Fig. 9 is adapted to store the successive coded speech event feature signals and combining signals received from utilization device 285 of Fig. 2 via line 901 and interface circuit 904. Store 920 receives the sequence of excitation signals required for synthesis via line 903. The excitation signals may comprise a succession of pitch period and voiced/unvoiced signals generated responsive to the voice message by methods well known in the art. Microprocessor 910 is adapted to control the operation of the synthesizer and may be the aforementioned Motorola-type MC68000 integrated circuit. LPC feature signal store 925 is utilized to store the successive log area parameter signals of the spoken message which are formed from the speech event feature signals and combining signals of store 915. Formation of a replica of the spoken message is accomplished in LPC synthesizer 930 responsive to the LPC feature signals from store 925 and the excitation signals from store 920 under control of microprocessor 910.

The synthesizer operation is directed by microprocessor 910 under control of permanently stored instruction codes resident in a read only memory associated therewith. The operation of the synthesizer is described in the flow chart of Fig. 10. Referring to Fig. 10, the coded speech event feature signals, the corresponding combining signals, and the excitation signals of the spoken message are received by interface 904 and are transferred to speech event feature signal and combining coefficient signal store 915 and to excitation signal store 920 as per box 1010. The log area parameter signal index l is then reset to one in processor 910 (box 1020) so that the reconstruction of the first log area feature signal $y_1(n)$ is initiated.

The formation of the log area signal requires combining the speech event feature signals with the combining coefficients of index l in accordance with equation 1. Speech event feature signal location counter L is reset to one by processor 910 as per box 1025 and the current speech event feature signal samples are read from store 915 (box 1030). The signal sample sequence is filtered to smooth the speech event feature signal as per (box 1035) and the current log area parameter signal is partially formed in box 1040. Speech event location counter L is incremented to address the next speech event feature signal in store 915 (box 1045) and the occurrence of the last feature signal is tested in decision box 1050. Until the last speech event feature signal has been processed, the loop including boxes 1030 through 1050 is iterated so that the current log area parameter signal is generated and stored in LPC feature signal store 925 under control of processor 910.

Upon storage of a log area feature signal in store 925, box 1055 is entered from box 1050 and the log area index signal l is incremented (box 1055) to initiate the formation of the next log area parameter signal. The loop from box 1030 through box 1050 is reentered via decision box 1060. After the last log area parameter signal is stored, processor 910 causes a replica of the spoken message to be formed in LPC synthesizer 930.

The synthesizer circuit of Fig. 9 may be readily modified to store the speech event feature signal sequences corresponding to a plurality of spoken messages and to selectively generate replicas of these messages by techniques well known in the art. For such an arrangement, the speech event feature signal generating circuit of Fig. 2 may receive a sequence of predetermined spoken messages and utilization device 285 may comprise an arrangement to permanently store the speech event feature signals and corresponding combining coefficients for the messages and to generate a read only memory containing said spoken message speech event and combining signals. The read only memory containing the coded speech event and combining signals can be inserted as store 915 in the synthesizer circuit of Fig. 9.

Claims

1. A method for compressing speech patterns including the steps of: analyzing (101, 110, 120) a speech pattern to derive a set of signals ($y_l(n)$) representative of acoustic features of the speech pattern at a first rate, generating (130, 140, 150) a sequence of coded signals representative of said speech pattern in response to said set of acoustic feature signals at a second rate less than said first rate, characterized in that the generating step includes: generating (420, 425) a sequence of signals ($\Phi_k(n)$) each representative of an individual sound of said speech pattern, each being a linear combination of said acoustic feature signals;

determining (510) the time frames of the speech pattern at which the centroids of individual sounds occur in response to said set of acoustic feature signals; generating (625) a sequence of individual sound feature signals ($\Phi_{L(l)}(n)$) jointly responsive to said acoustic feature signals and said centroid time frame determination; generating (805—815) a set of individual sound representative signal combining coefficients (a_{ik}) jointly responsive to said individual sound representative signals and said acoustic feature signals; and forming said coded signal responsive to said sequence of individual sound feature signals (715) and said combining coefficients (820).

2. A method for compressing speech patterns, as claimed in claim 1, wherein the step of determining the time frames of the speech pattern at which the centroids of individual sounds occur comprises producing (430) a signal (v(L)) representative of the timing of the individual sounds in said speech pattern responsive to the acoustic feature signals of the speech pattern, and detecting each negative going zero crossing in said individual sound time signal.

3. A method for compressing speech patterns as claimed in claim 1 or claim 2 wherein said coded signal forming step comprises generating (710) a signal representative of the bandwidth of each speech representative signal; sampling said speech event feature signal at a rate corresponding to its bandwidth representative signal; coding (715) each sampled speech event feature signal; and producing a sequence of encoded speech event coded signals at a rate corresponding to the rate of occurrence of speech events in said speech pattern.

4. A method for compressing speech patterns as claimed in any of the preceding claims wherein, said acoustic feature signals are linear predictive parameter signals representative of the speech pattern.

5. A method for compressing speech patterns as claimed in claim 4 wherein said linear predictive parameter signals are log area parameter signals representative of the speech pattern.

6. A method for compressing speech patterns as claimed in claim 4 wherein said linear predictive parameter signals are partial autocorrelation signals representative of the speech pattern.

7. Apparatus for compressing speech patterns, including means (210, 215, 225, 280) for analyzing a speech pattern to derive a set of signals representative of acoustic features of the speech pattern at a first rate and means (220—260) for generating a sequence of coded signals representative of said speech pattern in response to said set of acoustic feature signals at a second rate less than said first rate, characterized in that the generating means includes: means (220) for generating a sequence of signals ($\Phi_k(n)$) each representative of an individual sound of said speech pattern, each being a linear combination of said acoustic feature signals and determining the time frames of the speech pattern at which the centroids of individual sounds occur in response to said set of acoustic feature signals, means (230) for generating a set of individual sound representative signal combining coefficients (a_{ik}) jointly responsive to said individual sound representative signals and said acoustic feature signals, means (225) for generating a sequence of individual sound feature signals ($\Phi_{L(l)}(n)$) jointly responsive to said acoustic feature signals and said centroid time frame determination, and means (235) for forming said coded signal responsive to said sequence of individual sound feature signals and said combining coefficients.

8. Apparatus for compressing speech patterns as claimed in claim 7, wherein the means for determining the time frames of the speech pattern at which the centroids of individual sounds occur comprises means (220) for producing a signal representative of the timing of the individual sounds in said speech pattern responsive to the acoustic feature signals of the speech pattern, and detecting each negative going zero crossing in said individual sound time signal.

9. Apparatus for compressing speech patterns as claimed in claim 7 or claim 8, wherein the means for forming a signal comprises means (part of 235) for generating a signal representative of the bandwidth of each speech representative signal; means (part of 235) for sampling each individual sound articulatory configuration representative signal in said speech pattern at a rate corresponding to its bandwidth signal; means (part of 235) for coding each individual sound articulatory configuration representative signal; and means (part of 235) for producing a sequence of said coded individual sound articulatory configuration representative sample signals at a rate corresponding to the individual sound articulatory configuration representative signal bandwidths.

10. Apparatus as claimed in any of claims 7 to 9, wherein the means for analyzing a speech pattern comprises means (210, 215, 275, 280) for generating a set of linear predictive parameter signals representative of the acoustic features of the speech pattern.

11. Apparatus as claimed in any of claims 7 to 10 including means (285 or 910—930) for generating a speech pattern from the coded signal.

Patentansprüche

1. Verfahren zur Kompression von Sprachmustern mit den Verfahrensschritten:

Analysieren (101, 110, 120) eines Sprachmusters zur Ableitung eines Satzes von Signalen ($y_l(n)$), die akustische Merkmale des Sprachmusters bei einer ersten Rate darstellen;

Erzeugen (130, 140, 150) einer Folge von codierten, das Sprachmuster darstellenden Signalen unter Ansprechen auf den Satz akustischer Merkmalssignale mit einer zweiten Rate, die kleiner als die erste Rate ist, dadurch gekennzeichnet, daß der Verfahrensschritt der Erzeugung einer Folge von codierten Signalen umfaßt:

Erzeugen (420, 425) einer Folge von Signalen ($\phi_k(n)$), die je einen individuellen Laut des Sprachmusters darstellen und je eine lineare Kombination der akustischen Merkmalssignale sind;

Bestimmen (510) der Zeitrahmen des Sprachmusters, in denen die Schwerpunkte individueller Laute auftreten, und zwar unter Ansprechen auf den Satz akustischer Merkmalssignale;

Erzeugen (625) einer Folge von individuellen Lautmerkmalssignalen ($\phi_{L(l)}(n)$ unter gemeinsamen Ansprechen auf die akustischen Merkmalssignale und die Schwerpunkt-Zeitrahmenbestimmung;

Erzeugen (805—815) eines Satzes von individuelle Laute darstellenden Signalkombinationskoeffizienten (a_{ik}) unter gemeinsamen Ansprechen auf die individuelle Laute darstellenden Signale und die akustischen Merkmalssignale; und

Bilden der codierten Signale unter Ansprechen auf die Folge von individuellen Lautmerkmalssignalen (715) und die Signalkombinationskoeffizienten (820).

2. Verfahren zur Kompression von Sprachmustern nach Anspruch 1, bei dem die Bestimmung der Zeitrahmen des Sprachmusters, in denen die Schwerpunkte individueller Laute auftreten, die Erzeugung (430) eines die zeitliche Lage der individuellen Laute in dem Sprachmuster darstellenden Signals (v(L)) unter Ansprechen auf die akustischen Merkmalssignale des Sprachmusters und die Feststellung jeder negativ gerichteten Nullkreuzung des Individuallaut-Zeitlagesignals umfaßt.

3. Verfahren zur Kompression von Sprachmustern nach Anspruch 1 oder 2, bei dem die Bildung des codierten Signals umfaßt:

Erzeugen (710) eines Signals, das die Bandbreite jedes Sprache darstellenden Signals darstellt;

Abtasten des individuelle Laute darstellenden Signals mit einer Rate, die seinem Bandbreite darstellenden Signal entspricht;

Codieren (715) jedes abgetasteten Laut darstellenden Signals; und

Erzeugen einer Folge von codierten Lautcodesignalen mit einer Rate, die der Rate für das Auftreten von Lauten in dem Sprachmuster entspricht.

4. Verfahren zur Kompression von Sprachmustern nach einem der vorhergehenden Ansprüche, bei dem die akustischen Merkmalssignale lineare Voraussageparametersignale sind, die das Sprachmuster darstellen.

5. Verfahren zur Kompression von Sprachmustern nach Anspruch 4, bei dem die linearen Voraussageparametersignale logarithmische Bereichsparametersignale sind, die das Sprachmuster darstellen.

6. Verfahren zur Kompression von Sprachmustern nach Anspruch 4, bei dem die linearen Voraussageparametersignale partielle Autokorrelationssignale sind, die das Sprachmuster darstellen.

7. Vorrichtung zur Kompression von Sprachmustern mit einer Einrichtung (210, 215, 225, 280) zum Analysieren eines Sprachmusters zwecks Ableitung eines Satzes von Signalen, die akustische Merkmale des Sprachmusters bei einer ersten Rate darstellen, und einer Einrichtung (220—260) zum Erzeugen einer Folge von codierten, das Sprachmuster darstellenden Signalen unter Ansprechen auf den Satz akustischer Merkmalssignale mit einer zweiten Rate, die kleiner als die erste Rate ist, dadurch gekennzeichnet, daß die Erzeugungseinrichtung umfaßt:

eine Einrichtung (220) zur Erzeugung einer Folge von Signalen ($\phi_k(n)$), die je einen individuellen Laut des Sprachmusters darstellen und je eine lineare Kombination der akustischen Merkmalssignale sind, und zur Bestimmung der Zeitrahmen des Sprachmusters, in denen die Schwerpunkte individueller Laute auftreten, und zwar unter Ansprechen auf den Satz akustischer Merkmalssignale,

eine Einrichtung (230) zur Erzeugung eines Satzes von individuelle Laute darstellenden Signalkombinationskoeffizienten (a_{ik}) unter gemeinsamen Ansprechen auf die individuelle Laute darstellenden Signale und die akustischen Merkmalssignale,

eine Einrichtung (225) zur Erzeugung einer Folge von individuellen Lautmerkmalssignalen ($\phi_{L(l)}(n)$) unter gemeinsamem Ansprechen auf die akustischen Merkmalssignale und die Schwerpunkt-Zeitrahmenbestimmung, und

eine Einrichtung (235) zur Bildung der codierten Signale unter Ansprechen auf die Folge von individuellen Lautmerkmalssignalen und die Signalkombinationskoeffizienten.

8. Vorrichtung zur Kompression von Sprachmustern nach Anspruch 7, bei der die Einrichtung zur Bestimmung der Zeitrahmen des Sprachmusters, in denen die Schwerpunkte individueller Laute auftreten, eine Einrichtung (220) zur Erzeugung eines die zeitliche Lage der individuellen Laute in dem Sprachmuster darstellenden Signals unter Ansprechen auf die akustischen Merkmalssignale des Sprachmusters und zur Feststellung jeder negativ gerichteten Nullkreuzung in dem Individuallaut-Zeitlagesignal umfaßt.

9. Vorrichtung zur Kompression von Sprachmustern nach Anspruch 7 oder 8, bei der die Einrichtung zur Bildung des codierten Signals umfaßt:

eine Einrichtung (Teil von 235) zur Erzeugung eines Signals, das die Bandbreite jedes Sprache darstellenden Signals darstellt;

eine Einrichtung (Teil von 235) zur Abtastung jedes individuelle Laute darstellenden Signals in dem Sprachmuster mit einer Rate, die seinem Bandbreitesignal entspricht;

eine Einrichtung (Teil von 235) zur Codierung jedes individuelle Laute darstellenden Signals; und

eine Einrichtung (Teil von 235) zur Erzeugung einer Folge der codierten, individuelle Laute darstellenden Abtastsignale mit einer Rate, die der Bandbreite des individuelle Laute darstellenden Signals entspricht.

10. Vorrichtung nach einem der Ansprüche 7 bis 9, bei der die Einrichtung zur Analyse eines Sprachmusters eine Einrichtung (210, 215, 275, 280) zur Erzeugung eines Satzes linearer Voraussageparametersignale aufweist, die die akustischen Merkmale des Sprachmusters darstellen.

11. Vorrichtung nach einem der Ansprüche 7 bis 10 mit einer Einrichtung (285 oder 910—930) zur Erzeugung eines Sprachmusters aus dem codierten Signal.

Revendications

1. Un procédé pour comprimer des configurations de parole, comprenant les opérations suivantes: on analyse (101, 110, 120) une configuration de parole pour élaborer, à une première cadence, un ensemble de signaux ($y_l(n)$) représentatifs de caractéristiques acoustiques de la configuration de parole, on génère (130, 140, 150) une séquence de signaux codés représentatifs de la configuration de parole, sous la dépendance de l'ensemble précité de signaux de caractéristiques acoustiques, à une seconde cadence inférieure à la première cadence, caractérisé en ce que l'opération de génération comprend: la génération (420, 425) d'une séquence de signaux ($\phi_k(n)$), chacun d'eux étant représentatif d'un son individuel de la configuration de parole, et chacun d'eux étant une combinaison linéaire des signaux de caractéristiques acoustiques; on détermine (510) les trames temporelles de la configuration de parole dans lesquelles apparaissent les centroïdes de sons individuels, sous la dépendance de l'ensemble de signaux de caractéristiques acoustiques; on génère (625) une séquence de signaux de caractéristiques de sons individuels ($\phi_{L(l)}(n)$), sous la dépendance conjointe des signaux de caractéristiques acoustiques et de la détermination des trames temporelles de centroïdes; la génération (805—815) d'un ensemble de coefficients de combinaison de signaux représentatifs de sons individuels (a_{lk}), sous la dépendance conjointe des signaux représentatifs de sons individuels et des signaux de caractéristiques acoustiques; et la formation de signal codé sous la dépendance de la séquence de signaux de caractéristiques de sons individuels (715) et des coefficients de combinaison (820).

2. Un procédé pour comprimer des configurations de parole selon la revendication 1, dans lequel l'opération de détermination des trames temporelles de la configuration de parole dans lesquelles apparaissent les centroïdes de sons individuels, comprend la génération (430) d'un signal (v(L)) représentatif des instants d'apparition des sons individuels dans la configuration de parole, sous la dépendance des signaux de caractéristiques acoustiques de la configuration de parole, et la détection de chaque passage par zéro en sens négatif dans le signal d'instants d'apparition de sons individuels.

3. Un procédé pour comprimer des configurations de parole selon la revendication 1 ou la revendication 2, dans lequel l'opération de formation d'un signal codé comprend la génération (710) d'un signal représentatif de la largeur de bande de chaque signal représentatif de la parole; l'échantillonnage du signal de caractéristiques d'événement de parole à une cadence qui correspond à son signal représentatif de la largeur de bande; le codage (715) de chaque signal de caractéristiques d'événement échantillonné; et la génération d'une séquence de signaux codés d'événement de parole, à une cadence qui correspond à la cadence d'apparition des événements de parole dans la configuration de parole.

4. Un procédé pour comprimer des configurations de parole selon l'une quelconque des revendications précédentes, dans lequel les signaux de caractéristiques acoustiques sont des signaux de paramètres de prédiction linéaire, représentatifs de la configuration de parole.

5. Un procédé pour comprimer des configurations de parole selon la revendication 4, dans lequel les signaux de paramètres de prédiction linéaire sont des signaux de paramètre d'aire logarithmique, représentatifs de la configuration de parole.

6. Un procédé pour comprimer des configurations de parole selon la revendication 4, dans lequel les signaux de paramètres de prédiction linéaire sont des signaux d'autocorrélation partielle, représentatifs de la configuration de parole.

7. Appareil pour comprimer des configurations de parole, comprenant des moyens (210, 215, 225, 280) pour analyser une configuration de parole de façon à élaborer, à une première cadence, un ensemble de signaux représentatifs de caractéristiques acoustiques de la configuration de parole, et des moyens (220—260) pour générer une séquence de signaux codés représentatifs de la configuration de parole, sous la dépendance de l'ensemble de signaux de caractéristiques acoustiques, à une seconde cadence qui est inférieure à la première cadence, caractérisé en ce que les moyens de génération comprennent: des moyens (220) destinés à générer une séquence de signaux ($\phi_k(n)$), chacun d'eux étant représentatif d'un son individuel dans la configuration de parole, et chacun d'eux étant une combinaison linéaire de signaux de caractéristiques acoustiques, et pour déterminer les trames temporelles de la configuration de parole dans lesquelles apparaissent les centroïdes de son individuels, sous la dépendance de l'ensemble de signaux de caractéristiques acoustiques, des moyens (230) destinés à générer un ensemble de coefficients de combinaison de signaux représentatifs de sons individuels (a_{lk}), sous la dépendance conjointe des signaux représentatifs de sons individuels et des signaux de caractéristiques acoustiques, des moyens (225) destinés à générer une séquence de signaux de caractéristiques de sons individuels ($\phi_{L(l)}(n)$), sous la dépendance conjointe des signaux de caractéristiques acoustiques et de la détermination de trames temporelles de centroïdes, et des moyens (235) destinés à former le signal codé sous la dépendance de la séquence de signaux de caractéristiques de sons individuels et des coefficients de combinaison.

8. Appareil pour comprimer des configurations de parole selon la revendication 7, dans lequel les

moyens de détermination de trames temporelles de la configuration de parole dans lesquelles apparaissent les centroïdes de sons individuels, comprennent des moyens (220) destinés à produire un signal représentatif des instants d'apparition des sons individuels dans la configuration de parole, sous la dépendance des signaux de caractéristiques acoustiques de la configuration de parole, et à détecter chaque passage par zéro de sens négatif dans le signal d'instants d'apparition de sons individuels.

9. Appareil pour comprimer des configurations de parole selon la revendication 7 ou la revendication 8, dans lequel les moyens destinés à former un signal comprennent des moyens (une partie de 235) destinés à générer un signal représentatif de la largeur de bande de chaque signal représentatif de la parole; des moyens (une partie de 235) destinés à échantillonner chaque signal représentatif d'une configuration d'articulation d'un son individuel dans la configuration de parole, à une cadence qui correspond à son signal de largeur de bande; des moyens (235) destinés à coder chaque signal représentatif d'une configuration d'articulation d'un son individuel; et des moyens (une partie de 235) destinés à produire une séquence de signaux d'echantillon représentatifs d'une configuration d'articulation d'un son individuel, à une cadence correspondant aux largeurs de bande des signaux représentatifs d'une configuration d'articulation d'un son individuel.

10. Appareil selon l'une quelconque des revendications 7 à 9, dans lequel les moyens d'analyse d'une configuration de parole comprennent des moyens (210, 215, 275, 280) destinés à générer un ensemble de signaux de paramètres de prédiction linéaire, représentatifs des caractéristiques acoustiques de la configuration de parole.

11. Appareil selon l'une quelconque des revendications 7 à 10, comprenant des moyens (285 ou 910—930) destinés à générer une configuration de parole à partir du signal codé.

FIG. 1

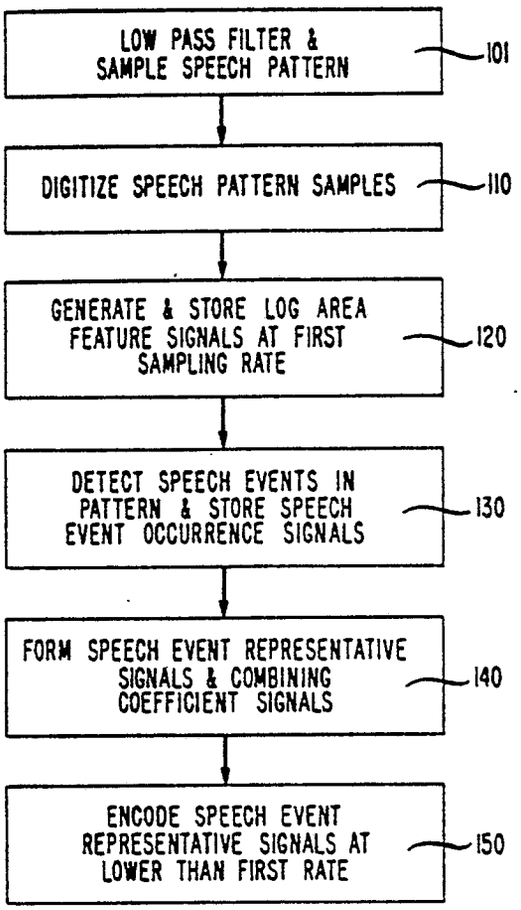


FIG. 3

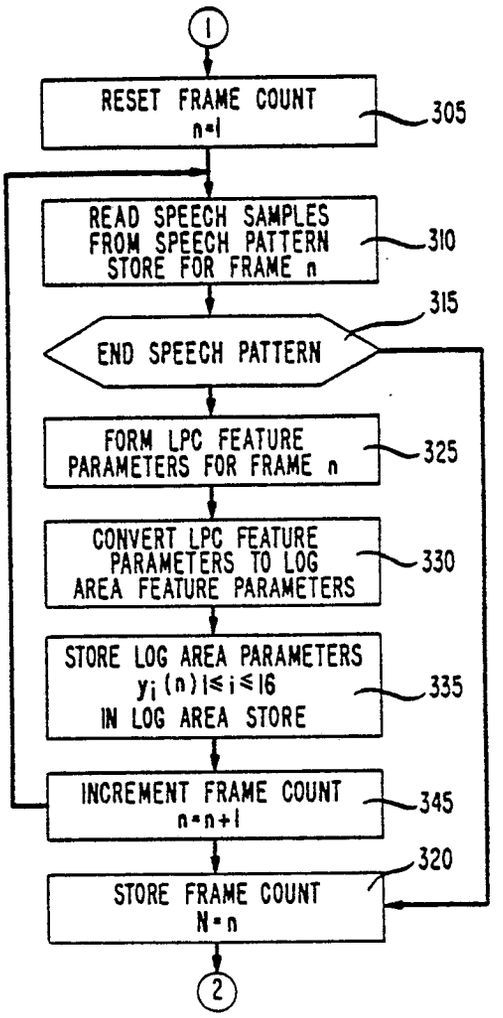
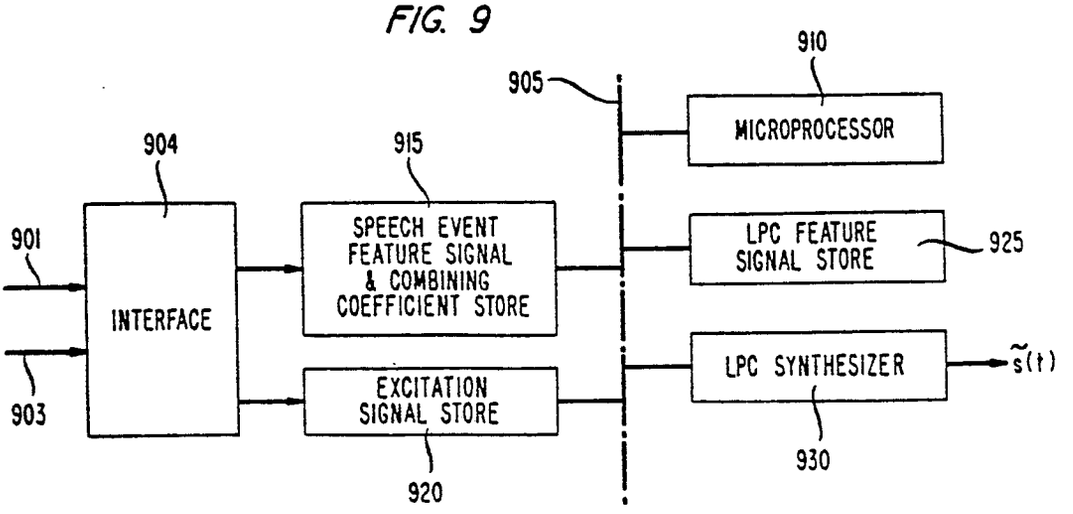


FIG. 9



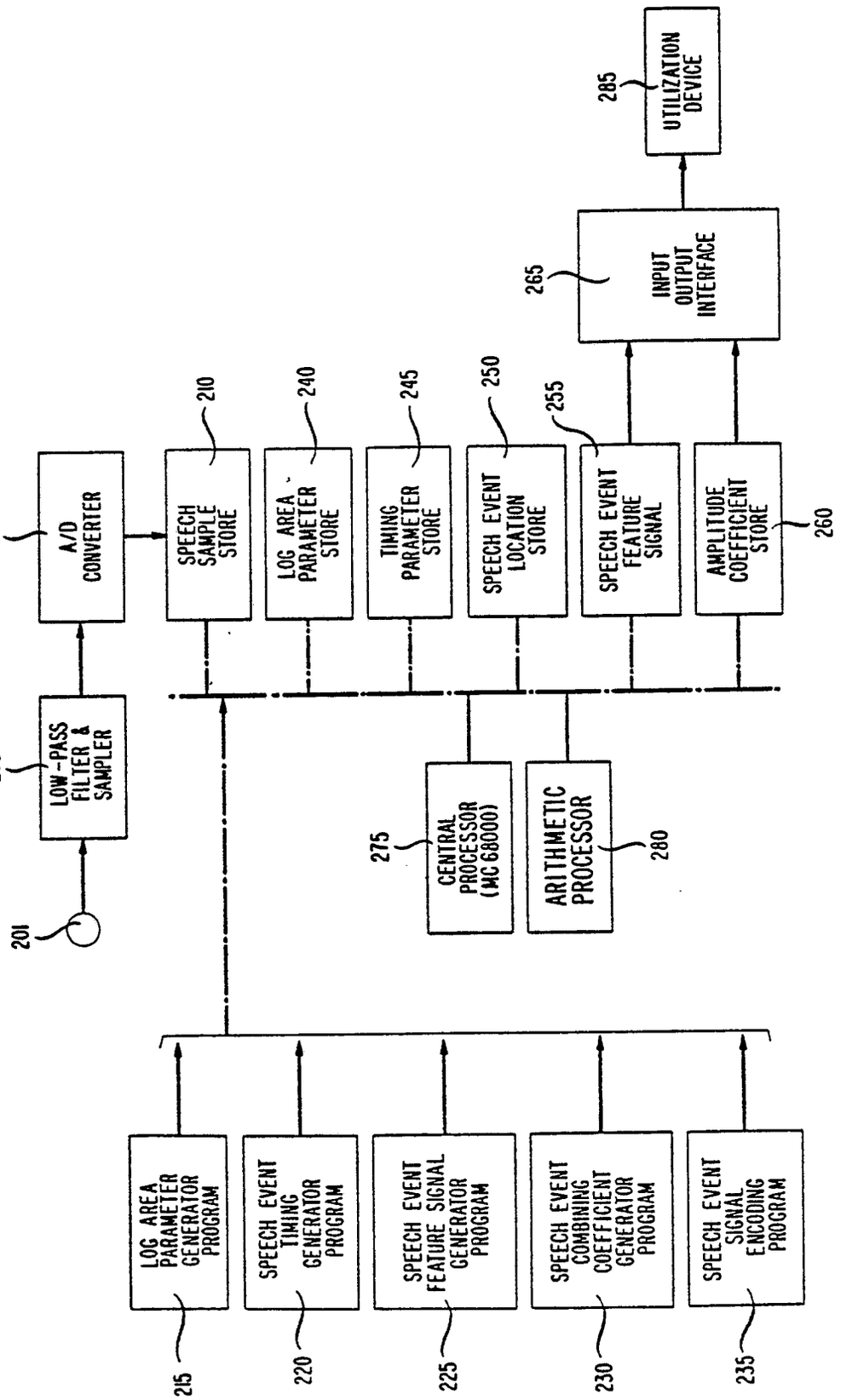

201
203
LOW-PASS FILTER & SAMPLER
205
A/D CONVERTER
210
SPEECH SAMPLE STORE
240
LOG AREA PARAMETER STORE
245
TIMING PARAMETER STORE
250
SPEECH EVENT LOCATION STORE
255
SPEECH EVENT FEATURE SIGNAL
260
AMPLITUDE COEFFICIENT STORE
265
INPUT OUTPUT INTERFACE
285
UTILIZATION DEVICE
275
CENTRAL PROCESSOR (MC 68000)
280
ARITHMETIC PROCESSOR
215
LOG AREA PARAMETER GENERATOR PROGRAM
220
SPEECH EVENT TIMING GENERATOR PROGRAM
225
SPEECH EVENT FEATURE SIGNAL GENERATOR PROGRAM
230
SPEECH EVENT COMBINING COEFFICIENT GENERATOR PROGRAM
235
SPEECH EVENT SIGNAL ENCODING PROGRAM

FIG. 2

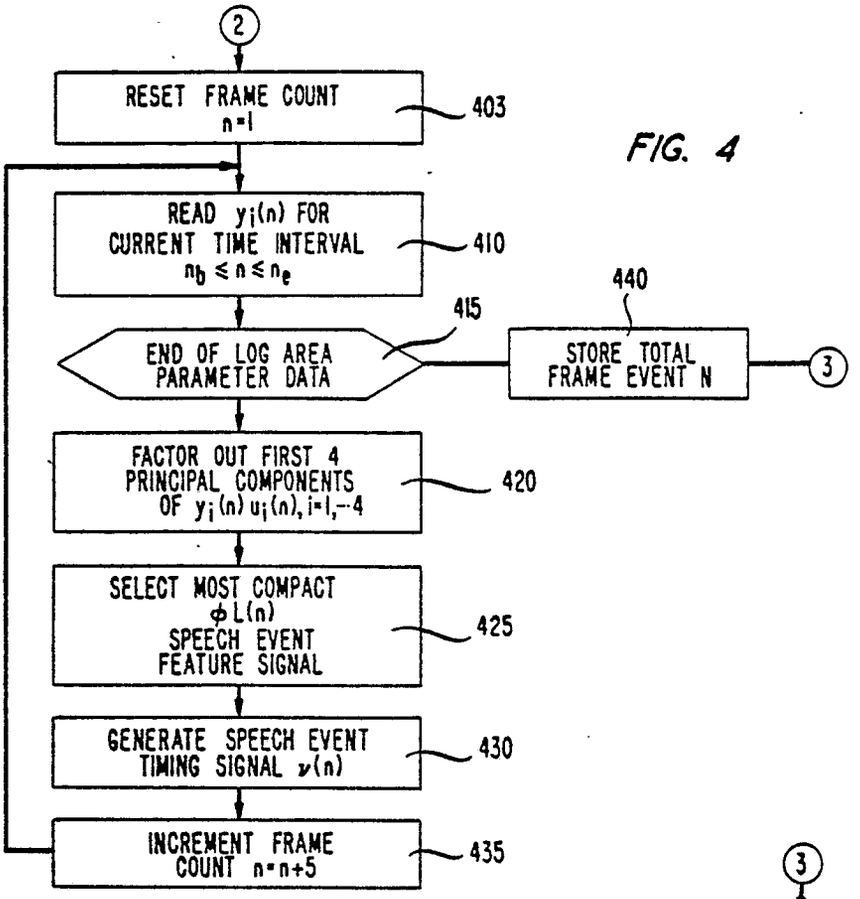

2
RESET FRAME COUNT
n=1
403
READ $y_i(n)$ FOR
CURRENT TIME INTERVAL
$n_b \leq n \leq n_e$
410
415
END OF LOG AREA
PARAMETER DATA
440
STORE TOTAL
FRAME EVENT N
3
FACTOR OUT FIRST 4
PRINCIPAL COMPONENTS
OF $y_i(n)\,u_i(n)$, i=1,--4
420
SELECT MOST COMPACT
$\phi L(n)$
SPEECH EVENT
FEATURE SIGNAL
425
GENERATE SPEECH EVENT
TIMING SIGNAL $\nu(n)$
430
INCREMENT FRAME
COUNT n=n+5
435

FIG. 4

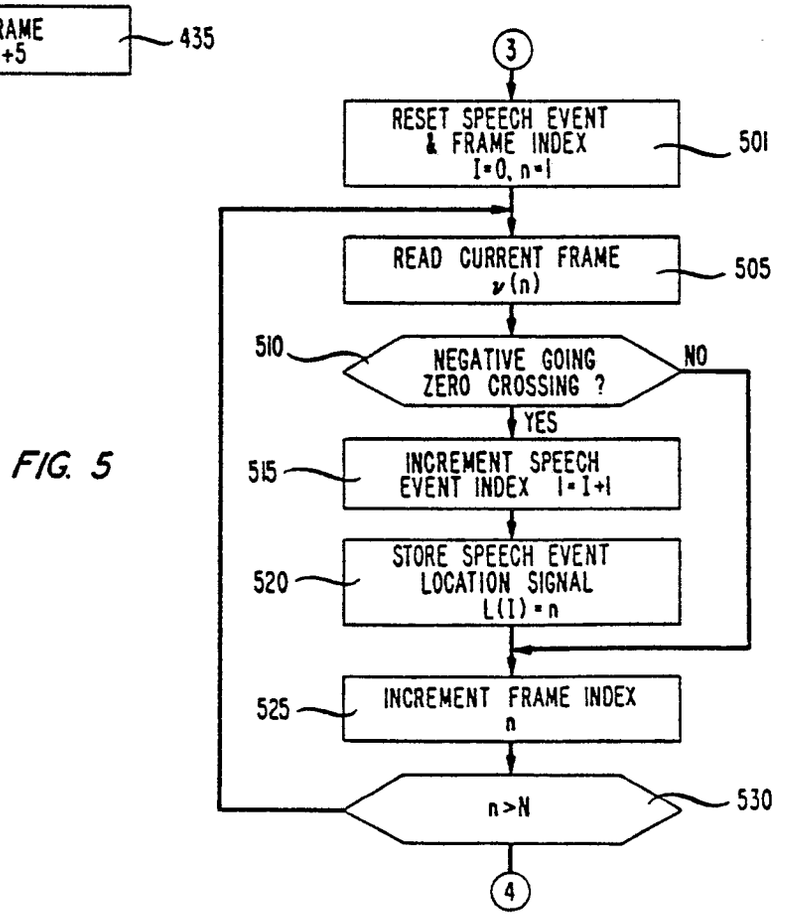

3
RESET SPEECH EVENT
& FRAME INDEX
I=0, n=1
501
READ CURRENT FRAME
$\nu(n)$
505
510
NEGATIVE GOING
ZERO CROSSING ?
NO
YES
515
INCREMENT SPEECH
EVENT INDEX I=I+1
520
STORE SPEECH EVENT
LOCATION SIGNAL
L(I)=n
525
INCREMENT FRAME INDEX
n
n>N
530
4

FIG. 5

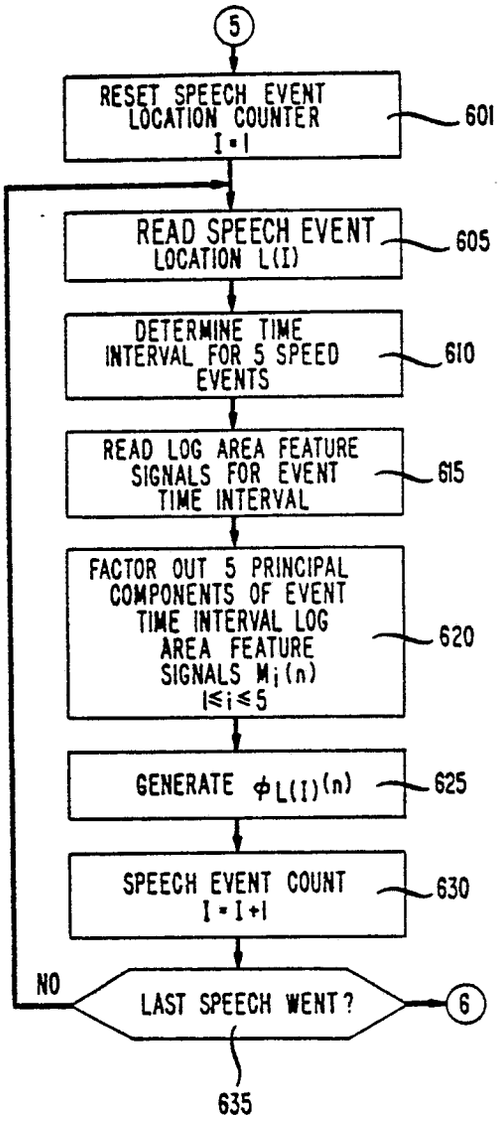

5
RESET SPEECH EVENT LOCATION COUNTER I=1
601
READ SPEECH EVENT LOCATION L(I)
605
DETERMINE TIME INTERVAL FOR 5 SPEED EVENTS
610
READ LOG AREA FEATURE SIGNALS FOR EVENT TIME INTERVAL
615
FACTOR OUT 5 PRINCIPAL COMPONENTS OF EVENT TIME INTERVAL LOG AREA FEATURE SIGNALS $M_i(n)$ $1 \leq i \leq 5$
620
GENERATE $\phi_{L(I)}(n)$
625
SPEECH EVENT COUNT I=I+1
630
NO
LAST SPEECH WENT?
6
635

FIG. 6

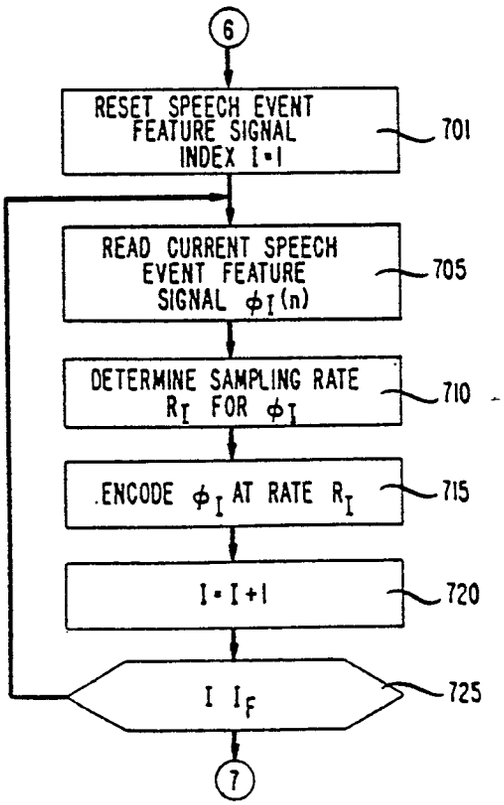

6
RESET SPEECH EVENT FEATURE SIGNAL INDEX I=1
701
READ CURRENT SPEECH EVENT FEATURE SIGNAL $\phi_I(n)$
705
DETERMINE SAMPLING RATE R_I FOR ϕ_I
710
ENCODE ϕ_I AT RATE R_I
715
I=I+1
720
I I_F
725
7

FIG. 7

FIG. 8

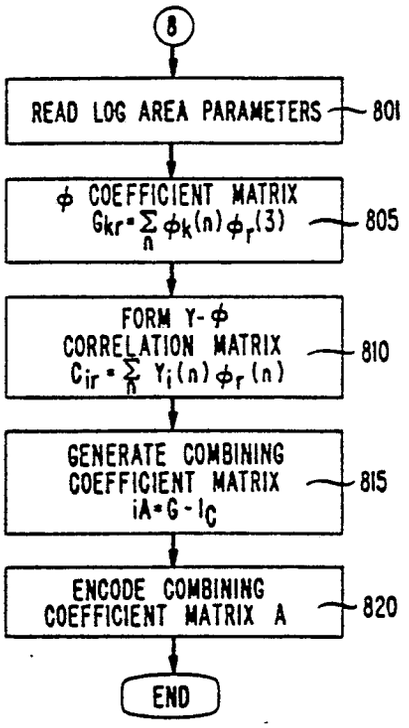

8
READ LOG AREA PARAMETERS
801
φ COEFFICIENT MATRIX
$G_{kr} = \sum_n \phi_k(n)\phi_r(3)$
805
FORM Y-φ CORRELATION MATRIX
$C_{ir} = \sum_n Y_i(n)\phi_r(n)$
810
GENERATE COMBINING COEFFICIENT MATRIX
$iA = G^{-1}C$
815
ENCODE COMBINING COEFFICIENT MATRIX A
820
END

FIG. 10

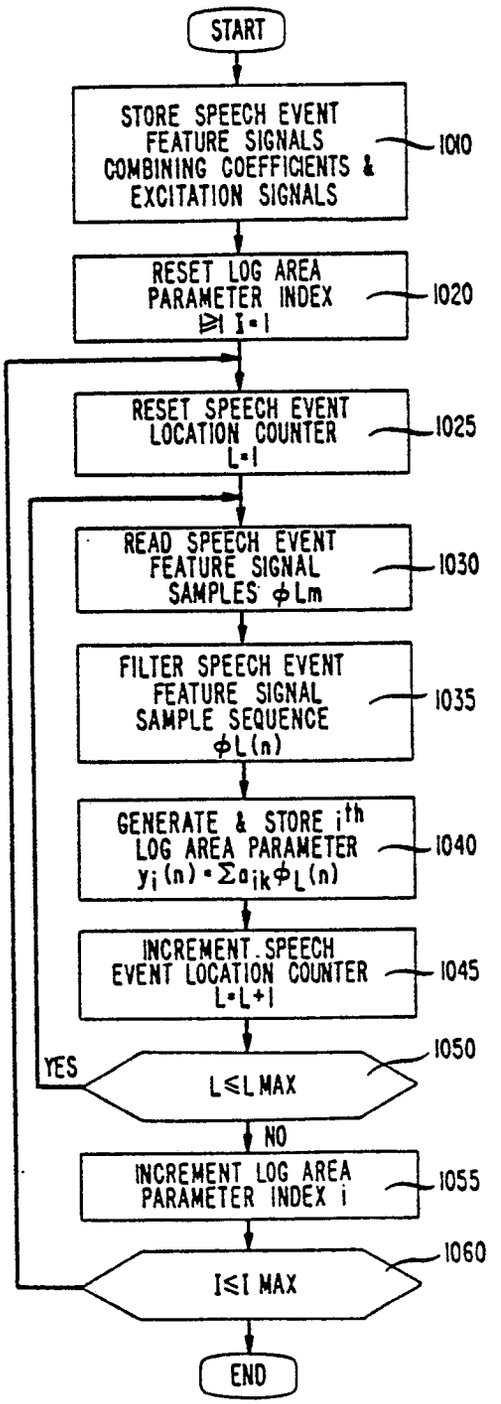

START
STORE SPEECH EVENT FEATURE SIGNALS COMBINING COEFFICIENTS & EXCITATION SIGNALS
1010
RESET LOG AREA PARAMETER INDEX
≥ 1 I = 1
1020
RESET SPEECH EVENT LOCATION COUNTER
L = 1
1025
READ SPEECH EVENT FEATURE SIGNAL SAMPLES ϕ_{Lm}
1030
FILTER SPEECH EVENT FEATURE SIGNAL SAMPLE SEQUENCE
$\phi_L(n)$
1035
GENERATE & STORE i^{th} LOG AREA PARAMETER
$y_i(n) = \Sigma a_{ik}\phi_L(n)$
1040
INCREMENT SPEECH EVENT LOCATION COUNTER
L = L+1
1045
YES
$L \leq L_{MAX}$
1050
NO
INCREMENT LOG AREA PARAMETER INDEX i
1055
$I \leq I_{MAX}$
1060
END

FIG. 11

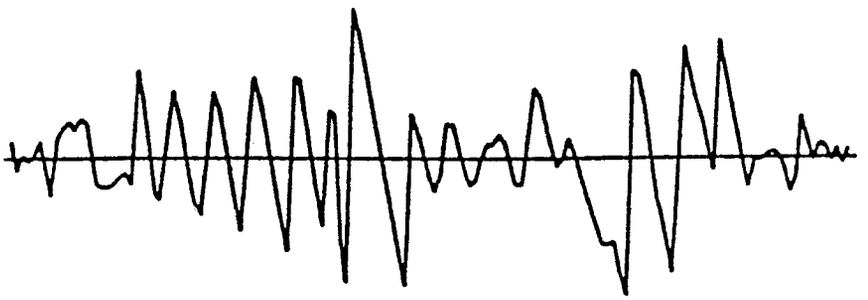


FIG. 12