FORM 1

SPRUSON & FERGUSON

# COMMONWEALTH OF AUSTRALIA

## PATENTS ACT 1952

596014

## APPLICATION FOR A STANDARD PATENT

American Telephone and Telegraph Company, incorporated in New York, of 550 Madison Avenue, New York, New York, 10022, UNITED STATES OF AMERICA, hereby apply for the grant of a standard patent for an invention entitled:

Code Excited Linear Predictive Vocoder and Method of Operation

which is described in the accompanying complete specification.

Details of basic application(s):-

| Basic Applic. No: | Country: | Application Date: |
|---|---|---|
| 067,649 | US | 26 June 1987 |

The address for service is:-

APPLICATION ACCEPTED AND AMENDMENTS

8 . 2 . 90

Spruson & Ferguson
Patent Attorneys
Level 33 St Martins Tower
31 Market Street
Sydney New South Wales Australia

DATED this TWENTY FOURTH day of JUNE 1988

American Telephone and Telegraph Company

By:

Registered Patent Attorney

TO: THE COMMISSIONER OF PATENTS
OUR REF: 61080
S&F CODE: 63845

REPRINT OF RECEIPT
S000549 24/06/88

5845/3

FORM 8

COMMONWEALTH OF AUSTRALIA

PATENTS ACT 1952-69

DECLARATION IN SUPPORT OF A CONVENTION APPLICATION FOR A PATENT

In support of the Convention application made for a patent for an invention entitled: CODE EXCITED LINEAR PREDICTIVE VOCODER AND METHOD OF OPERATION.
I, Kenneth Graham Johnston of 30 Branscombe Gardens, Winchmore Hill, London N21 3BN, England,
do solemnly and sincerely declare as follows:

1. I am authorized by American Telephone and Telegraph Company, the applicant for the patent to make this declaration on their behalf.

2. The basic application(~~s~~) as defined by Section 141 of the Act was made in the United States of America on the 26th day of June 1987 by Richard Harry Ketchum, Willem Bastiaan Kleijn and Daniel John Krasinski.

3.
Richard Harry Ketchum, 1754C Plymouth Court, Wheaton, Illinois 60187, United States of America,
Willem Bastiaan Kleijn, 238 North Van Nortwick, Batavia, Illinois 60510, United States of America, and
Daniel John Krasinski, 1407 Fairway Drive, Glendale Heights, Illinois 60139, United States of America,
are the actual inventor(s) of the invention and the fact upon which American Telephone and Telegraph Company is entitled to make the application is as follows:

The said Company is the assignee of the actual inventor(s).

4. The basic application referred to in paragraph 2 of this declaration was the first application made in a Convention country in respect of the invention the subject of the application.

Declared at LONDON

this 26th day of June 1988

K G Johnston

Declarant

To: The Commissioner of Patents.

**(12) PATENT ABRIDGMENT** **(11)** Document No. **AU-B-18384/88**
**(19) AUSTRALIAN PATENT OFFICE** **(10)** Acceptance No. **596014**

(54) Title
**CODE EXCITED LINEAR PREDICTIVE VOCODER**

International Patent Classification(s)
(51)[4] **G10L 009/14** **H03M 001/04** **H03M 007/02** **H03M 007/26**

(21) Application No. : **18384/88** (22) Application Date : **24.06.88**

(30) Priority Data

| (31) Number | (32) Date | (33) Country |
|---|---|---|
| **067649** | **26.06.87** | **US UNITED STATES OF AMERICA** |

(43) Publication Date : **05.01.89**

(44) Publication Date of Accepted Application : **12.04.90**

(71) Applicant(s)
**AMERICAN TELEPHONE AND TELEGRAPH COMPANY**

(72) Inventor(s)
**RICHARD HARRY KETCHUM; WILLEM BASTIAAN KLEIJN; DANIEL JOHN KRASINSKI**

(74) Attorney or Agent
**SPRUSON & FERGUSON**

(56) Prior Art Documents
**AU 18378/88 G10L**

(57) Claim

1. A method of encoding speech grouped into frames of speech, each frame comprising a plurality of samples, said method comprising, for each successive frame:

analysing the frame to determine a set of linear prediction coefficients;

selecting an excitation vector from a plurality of overlapping candidate vectors stored in a table; and

communicating the coefficients and the location of the selected vector in the table; the selecting step comprising:

deriving an excitation vector from the frame by a transformation using the linear prediction coefficients such that the result of operating on the excitation vector with a finite impulse response filter defined by the linear prediction coefficients provides an approximation to the frame;

comparing each of the candidate vectors in turn with the derived vector by recursively deriving for each candidate vector an error value indicative of the magnitude of the difference between the results of operating with the filter on the candidate vector and the derived vector;

and selecting the candidate vector for which the error value is least.

S & F Ref: 61080

FORM 10

COMMONWEALTH OF AUSTRALIA

PATENTS ACT 1952

**COMPLETE SPECIFICATION**

596014

(ORIGINAL)

FOR OFFICE USE:

Class Int Class

Complete Specification Lodged:
Accepted:
Published:

Priority:

Related Art:

---

Name and Address of Applicant: American Telephone and Telegraph Company
550 Madison Avenue
New York New York 10022
UNITED STATES OF AMERICA

Address for Service: Spruson & Ferguson, Patent Attorneys
Level 33 St Martins Tower, 31 Market Street
Sydney, New South Wales, 2000, Australia

Complete Specification for the invention entitled:

Code Excited Linear Predictive Vocoder and Method of Operation

The following statement is a full description of this invention, including the best method of performing it known to me/us

5845/4

# CODE EXCITED LINEAR PREDICTIVE VOCODER AND METHOD OF OPERATION

## Technical Field

This invention relates to low bit rate coding and decoding of speech and in particular to an improved code excited linear predictive vocoder.

## Background and Problem

Code excited linear predictive coding (CELP) is a well-known technique. This coding technique synthesizes speech by utilizing encoded excitation information to excite a linear predictive (LPC) filter. This excitation is found by searching through a table of candidate excitation vectors on a frame-by-frame basis.

LPC analysis is performed on the input speech to determine the LPC filter. The analysis proceeds by comparing the outputs of the LPC filter when it is excited by the various candidate vectors from the table or codebook. The best candidate is chosen based on how well its corresponding synthesized output matches the input speech. After the best match has been found, information specifying the best codebook entry and the filter are transmitted to the synthesizer. The synthesizer has a similar codebook and accesses the appropriate entry in that codebook, using it to excite the same LPC filter.

The codebook is made up of vectors whose components are consecutive excitation samples. Each vector contains the same number of excitation samples as there are speech samples in a frame. The vectors can be constructed in one of two ways. In the first method, disjoint sets of samples are used to define the vectors. In the second method, the overlapping codebook, the vectors are defined by shifting a window along a linear array of excitation samples.

The excitation samples used in the vectors in the CELP codebook can come from a number of possible sources. One particular example is Stochastically Excited Linear Prediction (SELP) method, which uses white noise, or random numbers, as the samples. Another method is to use an adaptive codebook. In such a scheme, the synthetic excitation determined for the present frame is used to update the codebook for future frames. This procedure allows the excitation codebook to adapt to the speech.

A problem with the CELP techniques for coding speech is that each excitation set of information in the codebook must be used to excite the LPC filter and then the excitation results must be compared utilizing an error criterion. Normally, the error criterion used is to determine the sum of the squared difference between the original and the synthesized speech samples resulting from the excitation information for each set of information. These calculations involve the convolution of each set of excitation information stored in the codebook with the LPC filter. The calculations are performed by using vector and matrix operations of the excitation information and the LPC filter. The problem is the large number of calculations, approximately 500 million multiply-add operations per second for a 4.8 Kbps vocoder, that must be performed.

Solution

A method according to the invention of encoding speech grouped into frames of speech, each frame comprising a plurality of samples, comprises, for each successive frame: analysing the frame to determine a set of linear prediction coefficients; selecting an excitation vector from a plurality of overlapping candidate vectors stored in a table; and communicating the coefficients and the location of the selected vector in the table; the selecting step comprising: deriving an excitation vector from the frame by a transformation using the linear prediction coefficients such that the result of operating on the excitation vector with a finite impulse response filter defined by the linear prediction coefficients provides an approximation to the frame; comparing each of the candidate vectors in turn with the derived vector by recursively deriving for each candidate vector an error value indicative of the magnitude of the difference between the results of operating with the filter on the candidate vector and the derived vector; and selecting the candidate vector for which the error value is least.

Apparatus according to the invention for encoding speech grouped into frames of speech, each frame comprising a plurality of samples, comprises: means for analysing each frame to determine a set of linear prediction coefficients; means for selecting an excitation vector from a plurality of overlapping candidate vectors stored in a table; and means for communicating the coefficients and the location of the selected vector in the table; the selecting means comprising: means for deriving an excitation vector from the frame by a transformation using the linear prediction coefficients such that the result of operating on the excitation vector with a finite impulse response filter defined by the linear prediction coefficients provides an approximation to the frame; and means for comparing each of the candidate vectors in turn with the derived vector by recursively deriving for each candidate vector an error value indicative of the magnitude of the difference between the results of operating with the filter on the candidate vector and the derived vector and selecting the candidate vector for which the error value is least.

Brief Description of the Drawings

Some embodiments of the invention will now be described with reference to the ~~accompany~~ accompanying drawings, in which:

FIG. 1 illustrates, in block diagram form, analyzer and synthesizer sections of a vocoder embodying this invention;

FIG. 2 illustrates, in graphic form, the formation of excitation vectors from codebook 104 using the virtual search technique;

FIGS. 3 through 6 illustrates, in graphic form, vector and matrix operations used by the vocoder of FIG. 1;

FIG. 7 illustrates, in greater detail, adaptive searcher 106 of FIG. 1;

FIG. 8 illustrates, in greater detail, virtual search control 708 of FIG. 7; and

FIG. 9 illustrates, in greater detail, energy calculator 709 of FIG. 7.

Detailed Description

FIG. 1 illustrates, in block diagram form, a vocoder. Elements 101 through 112 represent the analyzer portion of the vocoder, whereas, elements 151 through 157 represent the synthesizer

J.

portion of the vocoder. The analyzer portion of FIG. 1 is responsive to incoming speech received on path 120 to digitally sample the analog speech into digital samples and to group those digital samples into frames using well-known techniques. For each frame, the analyzer portion calculates the LPC coefficients representing the format characteristics of the vocal tract and searches for entries from both the stochastic codebook 105 and adaptive codebook 104 that best approximate the speech for that frame along with scaling factors. The latter entries and scaling information define excitation information as determined by the analyzer portion. This excitation and coefficient information is then transmitted by encoder 109 via path 145 to the synthesizer portion of the vocoder illustrated in FIG. 1. Stochastic generator 153 and adaptive generator 154 are responsive to the codebook entries and scaling factors to reproduce the excitation information calculated in the analyzer portion of the vocoder and to utilize this excitation information to excite the LPC filter that is determined by the LPC coefficients received from the analyzer portion to reproduce the speech.

Consider now in greater detail the functions of the analyzer portion of FIG. 1. LPC analyzer 101 is responsive to the incoming speech to determine LPC coefficients using well-known techniques. These LPC coefficients are transmitted to target excitation calculator 102, spectral weighting calculator 103, encoder 109, LPC filter 110, and zero-input response filter 111. Encoder 109 is responsive to the LPC coefficients to transmit the latter coefficients via path 145 to decoder 151. Spectral weighting calculator 103 is responsive to the coefficients to calculate spectral weighting information in the form of a matrix that emphasizes those portions of speech that are known to have important speech content. This spectral weighting information is based on a finite impulse response LPC filter. The utilization of a finite impulse response filter will be shown to greatly reduce the number of calculations necessary for performing the computations performed in searchers 106 and 107. This spectral weighting information is utilized by the searchers in order to determine the best candidate for the excitation information from the codebooks 104 and 105.

Target excitation calculator 102 calculates the target excitation which searchers 106 and 107 attempt to approximate. This target excitation is calculated by convolving a whitening filter based on the LPC coefficients calculated by analyzer 101 with the incoming speech minus the effects of the excitation and LPC filter for the previous frame. The latter effects for the previous frames are calculated by filters 110 and 111. The reason that the excitation and LPC filter for the previous frame must be considered is that these factors produce a signal component in the present frame which is often referred to as the ringing of the LPC filter. As will be described later, filters 110 and 111 are responsive to the LPC coefficients and calculated excitation from the previous frame to determine this ringing signal and to transmit it via path 144 to subtracter 112. Subtracter 112 is responsive to the latter signal and the present speech to calculate a remainder signal representing the present speech minus the ringing signal. Calculator 102 is responsive to the remainder signal to calculate the target excitation information and to transmit the latter information via path 123 to searcher 106 and 107.

The latter searchers work sequentially to determine the calculated excitation also referred to as synthesis excitation which is transmitted in the form of codebook indices and scaling factors via encoder 109 and path 145 to the synthesizer portion of FIG. 1. Each searcher calculates a portion of the calculated excitation. First, adaptive searcher 106 calculates excitation information and transmits this via path 127 to stochastic searcher 107. Searcher 107 is responsive to the target excitation received via path 123 and the excitation information from adaptive searcher 106 to calculate the remaining portion of the calculated excitation that best approximates the target excitation calculated by calculator 102.

Searcher 107 determines the remaining excitation to be calculated by subtracting the excitation determined by searcher 106 from the target excitation. The calculated or synthetic excitation determined by searchers 106 and 107 is transmitted via paths 127 and 126, respectively, to adder 108. Adder 108 adds the two excitation components together to arrive at the synthetic excitation for the present frame. The synthetic excitation is used by the synthesizer to produce the synthesized speech.

The output of adder 108 is also transmitted via path 128 to LPC filter 110 and adaptive codebook 104. The excitation information transmitted via path 128 is utilized to update adaptive codebook 104. The codebook indices and scaling factors are transmitted from searchers 106 and 107 to encoder 109 via paths 125 and 124, respectively.

Searcher 106 functions by accessing sets of excitation information stored in adaptive codebook 104 and utilizing each set of information to minimize an error criterion between the target excitation received via path 123 and the accessed set of excitation from codebook 104. A scaling factor is also calculated for each accessed set of information since the information stored in adaptive codebook 104 does not allow for the changes in dynamic range of human speech.

The error criterion used is the square of the difference between the original and synthetic speech. The synthetic speech is that which will be reproduced in the synthesis portion of FIG. 1 on the output of LPC filter 117. The synthetic speech is calculated in terms of the synthetic excitation information obtained from codebook 104 and the ringing signal; and the speech signal is calculated from the target excitation and the ringing signal. The excitation information for synthetic speech is utilized by performing a convolution of the LPC filter as determined by analyzer 102 utilizing the weighting information from calculator 103 expressed as a matrix. The error criterion is evaluated for each set of information obtained from codebook 104, and the set of excitation information giving the lowest error value is the set of information utilized for the present frame.

After searcher 106 has determined the set of excitation information to be utilized along with the scaling factor, the index into the codebook and the scaling factor are transmitted to encoder 109 via path 125, and the excitation information is also transmitted via path 127 to stochastic searcher 107. Stochastic searcher 107 subtracts the excitation information from adaptive searcher 106 from the target excitation received via path 123. Stochastic searcher 107 then performs operations similar to those performed by adaptive searcher 106.

The excitation information in adaptive codebook 104 is excitation information from previous frames. For each frame, the excitation information consists of the same number of samples as the sampled original speech. Advantageously, the excitation information may consist of 55 samples for a 4.8 Kbps transmission rate. The codebook is organized as a push down list so that the new set of samples are simply pushed into the codebook replacing the earliest samples presently in the codebook. When utilizing sets of excitation information out of codebook 104, searcher 106 does not treat these sets of information as disjoint sets of samples but rather treats the samples in the codebook as a linear array of excitation samples. For example, searcher 106 will form the first candidate set of information by utilizing sample 1 through sample 55 from codebook 104, and the second set of candidate information by using sample 2 through sample 56 from the codebook. This type of searching a codebook is often referred to as an overlapping codebook.

As this linear searching technique approaches the end of the samples in the codebook there is no longer a full set of information to be utilized. A set of information is also referred to as an excitation vector. At that point, the searcher performs a virtual search. A virtual search involves repeating accessed information from the table into a later portion of the set for which there are no samples in the table. This virtual search technique allows the adaptive searcher 106 to more quickly react to transitions from an unvoiced region of speech to a voiced region of speech. The reason is that in unvoiced speech regions the excitation is similar to white noise whereas in the voiced regions there is a fundamental frequency. Once a portion of the fundamental frequency has been identified from the codebooks, it is repeated.

FIG. 2 illustrates a portion of excitation samples such as would be stored in codebook 104 but where it is assumed for the sake of illustration that there are only 10 samples per excitation set. Line 201 illustrates that the contents of the codebook and lines 202, 203 and 204 illustrate excitation sets which have been formed utilizing the virtual search technique. The excitation set illustrated in line 202 is formed by searching the codebook starting at sample 205 on line 201. Starting at sample 205, there are only 9 samples in the table, hence, sample 208 is repeated as sample 209 to form the tenth sample of the excitation set illustrated in line 202. Sample 208 of line 202 corresponds to sample 205 of line 201. Line 203 illustrates the excitation set following that illustrated in line 202 which is formed by starting at sample 206 on line 201. Starting at sample 206 there are only 8 samples in the code book, hence, the first 2 samples of line 203 which are grouped as samples 210 are repeated at the end of the excitation set illustrated in line 203 as samples 211. It can be observed by one skilled in the art that if the significant peak illustrated in line 203 was a pitch peak then this pitch has been repeated in samples 210 and 211. Line 204 illustrates the third excitation set formed starting at sample 207 in the codebook. As can be seen, the 3 samples indicated as 212 are repeated at the end of the excitation set illustrated on line 204 as samples 213. It is important to realize that the initial pitch peak which is labeled as 207 in line 201 is a cumulation of the searches performed by searchers 106 and 107 from the previous frame since the contents of codebook 104 are updated at the end of each frame. The statistical searcher 107 would normally arrive first at a pitch peak such as 207 upon entering a voiced region from an unvoiced region.

Stochastic searcher 107 functions in a similar manner as adaptive searcher 106 with the exception that it uses as a target excitation the difference between the target excitation from target excitation calculator 102 and excitation representing the best match found by searcher 106. In addition, search 107 does not perform a virtual search.

A detailed explanation is now given of the analyzer portion of FIG. 1. This explanation is based on matrix and vector mathematics. Target excitation calculator 102 calculates a target excitation vector, t, in the following manner. A speech vector s can be expressed as

$$s = Ht + z .$$

The H matrix is the matrix representation of the all-pole LPC synthesis filter as defined by the LPC coefficients received from LPC analyzer 101 via path 121. The structure of the filter represented by H is described in greater detail later in this section and is part of the subject of this invention. The vector z represents the ringing of the all-pole filter from the excitation received during the previous frame. As was described earlier, vector z is derived from LPC filter 110 and zero-input response filter 111. Calculator 102 and subtracter 112 obtain the vector t representing the target excitation by subtracting vector z from vector s and processing the resulting signal vector through the all-zero LPC analysis filter also derived from the LPC coefficients generated by LPC analyzer 101 and transmitted via path 121. The target excitation vector t is obtained by performing a convolution operation of the all-zero LPC analysis filter, also referred to as a whitening filter, and the difference signal found by subtracting the ringing from the original speech. This convolution is performed using well-known signal processing techniques.

Adaptive searcher 106 searches adaptive codebook 104 to find a candidate excitation vector r that best matches the target excitation vector t. Vector r is also referred to as a set of excitation information. The error criterion used to determine the best match is the square of the difference between the original speech and the synthetic speech. The original speech is given by vector s and the synthetic speech is given by the vector y which is calculated by the following equation:

$$y = HL_i r_i + z,$$

where $L_i$ is a scaling factor.

The error criterion can be written in the following form:

$$e = (Ht + z - HL_i r_i - z)^T (Ht + z - HL_i r_i - z). \quad (1)$$

In the error criterion, the H matrix is modified to emphasize those sections of the spectrum which are perceptually important. This is accomplished through well known pole-bandwidth widing technique. Equation 1 can be rewritten in the following form:

$$e = (t - L_i r_i)^T H^T H (t - L_i r_i). \quad (2)$$

Equation 2 can be further reduced as illustrated in the following:

$$e = t^T H^T Ht + L_i r_i^T H^T HL_i r_i - 2L_i r_i^T H^T Ht. \quad (3)$$

The first term of equation 3 is a constant with respect to any given frame and is dropped from the calculation of the error in determining which $r_i$ vector is to be utilized from codebook 104. For each of the $r_i$ excitation vectors in codebook 104, equation 3 must be solved and the error criterion, e, must be determined so as to chose the $r_i$ vector which has the lowest value of e. Before equation 3 can be solved, the scaling factor, $L_i$ must be determined. This is performed in a straight forward manner by taking the partial derivative with respect to $L_i$ and setting it equal to zero, which yields the following equation:

$$L_i = \frac{r_i^T H^T Ht}{r_i^T H^T Hr_i} . (4)$$

The numerator of equation 4 is normally referred to as the cross-correlation term and the denominator is referred to as the energy term. The energy term requires more computation than the cross-correlation term. The reason is that in the cross-correlation term the product of the last three elements needs only to be calculated once per frame yielding a vector, and then for each new candidate vector, $r_i$, it is simply necessary to take the dot product between the candidate vector transposed and the constant vector resulting from the computation of the last three elements of the cross-correlation term.

The energy term involves first calculating $Hr_i$ then taking the transpose of this and then taking the inner product between the transpose of $Hr_i$ and $Hr_i$. This results in a large number of matrix and vector operations requiring a large number of calculations. The following technique reduces the number of calculations and enhances the resulting synthetic speech.

In part, the technique realizes this goal by utilizing a finite impulse response LPC filter rather than an infinite impulse response LPC filter as utilized in the prior art. The utilization of a finite impulse response filter having a constant response length results in the H matrix having a different symmetry than in the prior art. The H matrix represents the operation of the finite impulse response filter in terms of matrix notation. Since the filter is a finite impulse response filter, the convolution of this filter and the excitation information represented by each candidate vector, $r_i$, results in each sample of the vector $r_i$ generating a finite number of response samples which are designated as R number of samples. When the matrix vector operation of calculating $Hr_i$ is performed which is a convolution operation, all of the R response points resulting from each sample in the candidate vector, $r_i$, are summed together to form a frame of synthetic speech.

The H matrix representing the finite impulse response filter is an N + R by N matrix, where N is the frame length in samples, and R is the length of the truncated impulse response in number of samples. Using this form of the H matrix, the response vector Hr has a length of N + R. This form of H matrix is illustrated in the following equation 5:

$$H = \begin{bmatrix} h_0 & 0 & . & . & 0 \\ h_1 & h_0 & . & . & . \\ . & . & . & . & . \\ h_R & h_{R-1} & . & . & . \\ 0 & h_R & . & . & h_0 \\ . & 0 & . & . & h_1 \\ . & . & . & . & . \\ . & . & . & h_R & h_{R-1} \\ 0 & 0 & . & 0 & h_R \end{bmatrix} . \quad (5)$$

Consider the product of the transpose of the H matrix and the H matrix itself as in equation 6:

$$A = H^T H . \quad (6)$$

Equation 6 results in a matrix A which is N by N square, symmetric, and Toeplitz as illustrated in the following equation 7.

$$A = \begin{bmatrix} A_0 & A_1 & A_2 & A_3 & A_4 \\ A_1 & A_0 & A_1 & A_2 & A_3 \\ A_2 & A_1 & A_0 & A_1 & A_2 \\ A_3 & A_2 & A_1 & A_0 & A_1 \\ A_4 & A_3 & A_2 & A_1 & A_0 \end{bmatrix} . \quad (7)$$

Equation 7 illustrates the A matrix which results from $H^TH$ operation when N is five. One skilled in the art would observe from equation 5 that depending on the value of R that certain of the elements in matrix A would be 0. For example, if R = 2 then elements $A_2$, $A_3$ and $A_4$ would be 0.

FIG. 3 illustrates what the energy term would be for the first candidate vector $r_1$ assuming that this vector contains 5 samples which means that N equals 5. The samples $X_0$ through $X_4$ are the first 5 samples stored in adaptive codebook 104. The calculation of the energy term of equation 4 for the second candidate vector $r_2$ is illustrated in FIG. 4. The latter figure illustrates that only the candidate vector has changed and that it has only changed by the deletion of the $X_0$ sample and the addition of the $X_5$ sample.

The calculation of the energy term illustrated in FIG. 3 results in a scalar value. This scalar value for $r_1$ differs from that for candidate vector $r_2$ as illustrated in FIG. 4 only by the addition of the $X_5$ sample and the deletion of the $X_0$ sample. Because of the symmetry and Toeplitz nature introduced into the A matrix due to the utilization of a finite impulse response filter, the scalar value for

FIG. 4 can be easily calculated in the following manner. First, the contribution due to the $X_0$ sample is eliminated by realizing that its contribution is easily determinable as illustrated in FIG. 5. This contribution can be removed since it is simply based on the multiplication and summation operations involving term 501 with terms 502 and the operations involving terms 504 with term 503. Similarly, FIG. 6 illustrates that the addition of term $X_5$ can be added into the scalar value by realizing that its contribution is due to the operations involving term 601 with terms 602 and the operations involving terms 604 with the terms 603. By subtracting the contribution of the terms indicated in FIG. 5 and adding the effect of the terms illustrated in FIG. 6, the energy term for FIG. 4 can be recursively calculated from the energy term of FIG. 3.

This method of recursive calculation is independent of the size of the vector $r_i$ or the A matrix. These recursive calculations allow the candidate vectors contained within adaptive codebook 104 or codebook 105 to be compared with each other but only requiring the additional operations illustrated by FIGS. 5 and 6 as each new excitation vector is taken from the codebook.

In general terms, these recursive calculations can be mathematically expressed in the following manner. First, a set of masking matrices is defined as $I_k$ where the last one appears in the kth row.

$$I_k = \begin{bmatrix} 1 & 0 & . & . & . & 0 \\ 0 & 1 & . & . & . & . \\ . & . & . & . & . & . \\ . & . & . & 1 & 0 & \\ . & . & . & 0 & 0 & . \\ . & . & . & . & . & . \\ 0 & . & . & . & . & 0 \end{bmatrix}. \qquad (8)$$

In addition, the unity matrix is defined as I as follows:

$$I = \begin{bmatrix} 1 & 0 & . & . & . & . \\ 0 & 1 & 0 & . & . & . \\ . & 0 & 1 & 0 & . & . \\ . & . & 0 & 1 & 0 & . \\ . & . & . & 0 & 1 & 0 \\ 0 & . & . & . & 0 & 1 \end{bmatrix}. \qquad (9)$$

Further, a shifting matrix is defined as follows:

$$S = \begin{bmatrix} 0 & 1 & 0 & . & . & 0 \\ 0 & 0 & 1 & . & . & . \\ . & . & . & . & . & . \\ . & . & . & . & 0 & 1 \\ 0 & . & . & . & 0 & 0 \end{bmatrix} . \qquad (10)$$

For Toeplitz matrices, the following well known theorem holds:

$$S^T AS = (I-I_1)\, A\, (I-I_1). \qquad (11)$$

Since A or $H^TH$ is Toeplitz, the recursive calculation for the energy term can be expressed using the following nomenclature. First, define the energy term associated with the $r_{j+1}$ vector as $E_{j+1}$ as follows:

$$E_{j+1} = r_{j+1}^T H^T Hr_{j+1} \ . \quad (12)$$

In addition, vector $r_{j+1}$ can be expressed as a shifted version of $r_j$ combined with a vector containing the new sample of $r_{j+1}$ as follows:

$$r_{j+1} = Sr_j + (I-I_{N-1})\, r_{j+1} \ . \quad (13)$$

Utilizing the theorem of equation 11 to eliminate the shift matrix S allows equation 12 to be rewritten in the following form:

$$E_{j+1} = E_j + 2\left[r_{j+1}^T (I-I_{N-1})\, H^T HSr_j - r_j^T (I-I_1)\, H^T HI_1 r_j\right]$$
$$-r_j^T I_1 H^T HI_1 r_j + r_{j+1}^T (I-I_{N-1})\, H^T H\, (I-I_{N-1})\, r_{j+1} \ . (14)$$

It can be observed from equation 14, that since the I and S matrices contain predominantly zeros with a certain number of ones that the number of calculations necessary to evaluate equation 14 is greatly reduced from that necessary to evaluate equation 3. A detailed analysis indicates that the calculation of equation 14 requires only 2Q+4 floating point operations, where Q is the smaller of the number R or the number N. This is a large reduction in the number of calculations from that required for equation 3. This reduction in calculation is accomplished by utilizing a finite impulse response filter rather than an infinite impulse response filter and by the Toeplitz nature of the $H^tH$ matrix.

Equation 14 properly computes the energy term during the normal search of codebook 104. However, once the virtual searching commences, equation 14 no longer would correctly calculate the energy term since the virtual samples as illustrated by samples 213 on line 204 of FIG. 2 are changing at twice the rate. In addition, the samples of the normal search illustrated by samples 214 of FIG. 2 are also changing in the middle of the excitation vector. This situation is resolved in a recursive manner by allowing the actual samples in the codebook, such as samples 214, to be designated by the vector $w_i$ and those of the virtual section, such as samples 21? of FIG. 2, to be denoted by the vector $v_i$. In addition, the virtual samples are restricted to less than half of the total excitation vector. The energy term can be rewritten from equation 14 utilizing these conditions as follows:

$$E_i = w_i^T H^T H w_i + 2 v_i^T H^T H w_i + v_i^T H^T H v_i \quad . \qquad (15)$$

The first and third terms of equation 15 can be computationally reduced in the following manner. The recursion for the first term of equation 15 can be written as:

$$w_{j+1}^T H^T H w_{j+1} = w_j^T H^T H w_j - 2 w_j^T (I - I_1) H^T H I_1 w_j - w_j^T I_1 H^T H I_1 w_j \quad ;(16)$$

and the relationship between $v_j$ and $v_{j+1}$ can be written as follows:

$$v_{j+1} = S^2 (I - I_{p+1}) v_j + (I - I_{N-2}) v_{j+1} \quad . \qquad (17)$$

This allows the third term of equation 15 to be reduced by using the following:

$$H^T H v_{j+1} = S^2 H^T H v_j + S^2 H^T H (I_p - I_{p+1}) v_j + (I - I_{N-2}) H^T H S^2 (I - I_{p+1}) v_j + H^T H (I - I_{N-2}) v_{j+1} . (18)$$

The variable p is the number of samples that actually exists in the codebook 104 that are presently used in the existing excitation vector. An example of the number of samples is that given by samples 214 in FIG. 2. The second term of equation 15 can also be reduced by equation 18 since $v_i^T H^T H$ is simply the transpose of $H^T H v_i$ in matrix arithmetic.

The rate at which searching is done through the actual codebook samples and the virtual samples is different. In the above illustrated example, the virtual samples are searched at twice the rate of actual samples.

FIG. 7 illustrates adaptive searcher 106 of FIG. 1 in greater detail. As previously described, adaptive searcher 106 performs two types of search operations: virtual and sequential. During the sequential search operation, searcher 106 accesses a complete candidate excitation vector from adaptive codebook 104; whereas, during a virtual search, adaptive searcher 106 accesses a partial candidate excitation vector from codebook 104 and repeats the first portion of the candidate vector accessed from codebook 104 into the latter portion of the candidate excitation vector as illustrated in FIG. 2. The virtual search operations are performed by blocks 708 through 712, and the sequential search operations are performed by blocks 702 through 706. Search determinator 701 determines whether a virtual or a sequential search is to be performed. Candidate selector 714 determines whether the codebook has been competely searched; and if the codebook has not been completely searched, selector 714 returns control back to search determinator 701.

Search determinator 701 is responsive to the spectral weighting matrix received via path 122 and the target excitation vector received path 123 to control the complete search codebook 104. The first group of candidate vectors are filled entirely from the codebook 104 and the necessary calculations are performed by blocks 702 through 706, and the second group of candidate excitation vectors are handled by blocks 708 through 712 with portions of vectors being repeated.

If the first group of candidate excitation vectors is being accessed from codebook 104, search determinator communicates the target excitation vector, spectral weighting matrix, and index of the candidate excitation vector to be accessed to sequential search control 702 via path 727. The latter control is responsive to the candidate vector index to access codebook 104. The sequential search control 702 then transfers the target excitation vector, the spectral weighting matrix, index, and the candidate excitation vector to blocks 703 and 704 via path 728.

Block 704 is responsive to the first candidate excitation vector received via path 728 to calculate a temporary vector equal to the $H^THt$ term of equation 3 and transfers this temporary vector and information received via path 728 to cross-correlation calculator 705 via path 729. After the first candidate vector, block 704 just communicates information received on path 728 to path 729. Calculator 705 calculates the cross-correlation term of equation 3.

Energy calculator 703 is responsive to the information on path 728 to calculate the energy term of equation 3 by performing the operations indicated by equation 14. Calculator 703 transfers this value to error calculator 706 via path 733.

Error calculator 706 is responsive to the information received via paths 730 and 733 to calculate the error value by adding the energy value and the cross-correlation value and to transfer that error value along with the candidate number, scaling factor, and candidate value to candidate selector 714 via path 730.

Candidate selector 714 is responsive to the information received via path 732 to retain the information of the candidate whose error value is the lowest and to return control to search determinator 701 via path 731 when actuated via path 732.

When search determinator 701 determines that the second group of candidate vectors is to be accessed from codebook 104, it transfers the target excitation vector, spectral weighting matrix, and candidate excitation vector index to virtual search control 708 via path 720. The latter search control accesses codebook 104 and transfers the accessed code excitation vector and information received via path 720 to blocks 709 and 710 via path 721. Blocks 710, 711 and 712, via paths 722 and 723, perform the same type of operations as performed by blocks 704, 705 and 706. Block 709 performs the operation of evaluating the energy term of equation 3 as does block 703; however, block 709 utilizes equation 15 rather than equation 14 as utilized by energy calculator 703.

For each candidate vector index, scaling factor, candidate vector, and error value received via path 724, candidate selector 714 retains the candidate vector, scaling factor, and the index of the vector having the lowest error value. After all of the candidate vectors have been processed, candidate selector 714 then transfers the index and scaling factor of the selected candidate vector which has the lowest error value to encoder 109 via path 125 and the selected excitation vector via path 127 to adder 108 and stochastic searcher 107 via path 127.

FIG. 8 illustrates, in greater detail, virtual search control 708. Adaptive codebook accessor 801 is responsive to the candidate index received via path 720 to access codebook 104 and to transfer the accessed candidate excitation vector and information received via path 720 to sample repeater 802 via path 803. Sample repeater 802 is responsive to the candidate vector to repeat the first portion of the candidate vector into the last portion of the candidate vector in order to obtain a complete candidate excitation vector which is then transferred via path 721 to blocks 709 and 710 of FIG. 7.

FIG. 9 illustrates, in greater detail, the operation of energy calculator 709 in performing the operations indicated by equation 18. Actual energy component calculator 901 performs the operations required by the first term of equation 18 and transfers the results to adder 905 via path 911. Temporary virtual vector calculator 902 calculates the term $H^T H v_i$ in accordance with equation 18 and transfers the results along with the information received via path 721 to calculators 903 and 904 via path 910. In response to the information on path 910, mixed energy component calculator 903 performs the operations required by the second term of equation 15 and transfers the results to adder 905 via path 913. In response to the information on path 910, virtual energy component calculator 904 performs the operations required by the third term of equation 15. Adder 905 is responsive to information on paths 911, 912, and 913 to calculate the energy value and to communicate that value on path 726.

Stochastic searcher 107 comprises blocks similar to blocks 701 through 706 and 714 as illustrated in FIG. 7. However, the equivalent search determinator 701 would form a second target excitation vector by subtracting the selected candidate excitation vector received via path 127 from the target excitation received via path 123. In addition, the determinator would always transfer control to the equivalent control 702.

The claims defining the invention are as follows:

1. A method of encoding speech grouped into frames of speech, each frame comprising a plurality of samples, said method comprising, for each successive frame:

analysing the frame to determine a set of linear prediction coefficients;

selecting an excitation vector from a plurality of overlapping candidate vectors stored in a table; and

communicating the coefficients and the location of the selected vector in the table; the selecting step comprising:

deriving an excitation vector from the frame by a transformation using the linear prediction coefficients such that the result of operating on the excitation vector with a finite impulse response filter defined by the linear prediction coefficients provides an approximation to the frame;

comparing each of the candidate vectors in turn with the derived vector by recursively deriving for each candidate vector an error value indicative of the magnitude of the difference between the results of operating with the filter on the candidate vector and the derived vector;

and selecting the candidate vector for which the error value is least.

2. A method as claimed in claim 1 including forming a difference vector from the derived excitation vector and the selected excitation vector, selecting a second excitation vector from a second plurality of overlapping candidate vectors stored in a second table by comparing each of the candidate vectors of the second plurality in turn with the difference vector by recursively deriving for each candidate vector an error value indicative of the magnitude of the difference between the results of operating with the filter on the candidate vector and the difference vector, electing the candidate vector for which the error value is least and communicating the location of the selected second excitation vector in the second table.

3. A method as claimed in claim 2 including adding the

selected excitation vector and the selected second excitation vector to form a synthesis vector and updating the first said table with the synthesis vector.

4. A method as claimed in any of the preceding claims wherein, prior to deriving the excitation vector, that portion of the response of the filter for the previous frame to the selected excitation vector or vectors for the previous frame which carries over to the current frame is subtracted from the current frame.

5. A method as claimed in any of the preceding claims wherein each successive one of the overlapping candidate vectors differs from the preceding candidate vector by only a first set of component values, said first set being included in the preceding candidate vector but not in the said one candidate vector, and a second set of component values, said second set being included in the said one candidate vector but not in the preceding candidate vector, and wherein the recursive derivation of the error value comprises calculating the contributions of the said sets of component values to the error value and subtracting the contribution of the first set from and adding the contribution of the second set to the error value for the preceding candidate vector.

6. A method as claimed in claim 5 wherein the first set consists of the first component value of the preceding candidate vector and the second set consists of the last component value of the said one candidate vector.

7. A method as claimed in any of the preceding claims wherein the determination of the error values comprises:

calculating the response matrix for the finite impulse response filter;

calculating a spectral weighting matrix of Toeplitz form by matrix operations on the response matrix;

calculating a cross-correlation value of the derived vector and each candidate vector, using the spectral weighting matrix;

recursively calculating an energy value for each candidate vector as the correlation of the candidate vector with itself using the spectral weighting matrix;

and calculating the value for each candidate vector as the quotient of the cross- ation value and the energy value.

8. A method as claimed in any of the preceding claims including calculating a scaling factor, being the factor by which the selected excitation vector is to be scaled to provide the best approximation to the derived vector, and communicating the scaling factor.

9. Apparatus for encoding speech grouped into frames of speech, each frame comprising a plurality of samples, comprising,

means for analysing each frame to determine a set of linear prediction coefficients;

means for selecting an excitation vector from a plurality of overlapping candidate vectors stored in a table; and

means for communicating the coefficients and the location of the selected vector in the table; the selecting means comprising:

means for deriving an excitation vector from the frame by a transformation using the linear prediction coefficients such that the result of operating on the excitation vector with a finite impulse response filter defined by the linear prediction coefficients provides an approximation to the frame;

and means for comparing each of the candidate vectors in turn with the derived vector by recursively deriving for each candidate vector an error value indicative of the magnitude of the difference between the results of operating with the filter on the candidate vector and the derived vector and selecting the candidate vector for which the error value is least.

10. Apparatus as claimed in claim 9 including means for forming a difference vector from the derived excitation vector and the selected excitation vector,

means for selecting a second excitation vector from a second plurality of overlapping candidate vectors stored in a second table by comparing each of the candidate vectors of the second plurality in turn with the difference vector by recursively deriving for each candidate vector an error value indicative of the magnitude of the difference between the results of operating with the filter on the candidate vector and the difference vector and selecting the candidate vector for which the error value is least and means for communicating the location of the selected second excitation vector in the second table.

11. Apparatus as claimed in claim 10 including means for adding the selected excitation vector and the selected second excitation vector to form a synthesis vector and updating the first said table with the synthesis vector.

12. Apparatus as claimed in any of claims 9 to 11 including means for subtracting that portion of the response of the filter for the previous frame to the selected excitation vector or vectors for the previous frame which carries over to the current frame from the current frame prior to the deriviation of the excitation vector.

13. Apparatus as claimed in any of claims 9 to 12 wherein each successive one of the overlapping candidate vectors differs from the preceding candidate vector by only a first set of component values, said first set being included in the preceding candidate vector but not in the said one candidate vector, and a second set of component values, said second set being included in the said one candidate vector but not in the preceding candidate vector, and wherein the comparing means is arranged to derive the error value by calculating the contributions of the said sets of component values to the error value and subtracting the contribution of the first set from and adding the contribution of the second set to the error value for the ~~preceing~~ preceding candidate vector.

14. Apparatus as claimed in claim 13 wherein the first set consists of the first component value of the preceding candidate vector and the second set consists of the last component value of the said one candidate vector.

15. Apparatus as claimed in any of claims 9 to 14 wherein the comparing means is arranged to determine the error value by:

calculating the response matrix for the finite impulse response filter;

calculating a spectral weighting matrix of Toeplitz form by matrix operations on the response matrix;

calculating a cross-correlation value of the derived vector and each candidate vector, using the spectral weighting matrix;

recursively calculating an energy value for each candidate vector as the correlation of the candidate vector with itself using the spectral weighting matrix;

and calculating the error value for each candidate vector as the quotient of the cross-correlation value and the energy value.

16. Apparatus as claimed in any of claims 9 to 15 including means for calculating a scaling factor, being the factor by which the selected excitation vector is to be scaled to provide the best approximation to the derived vector, and means for communicating the scaling factor.

CMKW/BG

~~25~~. 17 A vocoder as hereinbefore particularly described with reference to what is shown in Figure 1.

~~26~~. 18 A vocoder as hereinbefore particularly described with reference to what is shown in the accompanying drawings.

~~27~~. 19 A method of encoding speech grouped into frames of speech as hereinbefore particularly described with reference to what is shown in the accompanying drawings.

~~28~~. 20 Apparatus for encoding speech grouped into frames of speech, each frame comprising a plurality of samples, as hereinbefore particularly described with reference to what is shown in the accompanying drawings.

DATED this TWENTY NINTH day of NOVEMBER 1989
American Telephone & Telegraph Company

Patent Attorneys for the Applicant
SPRUSON & FERGUSON

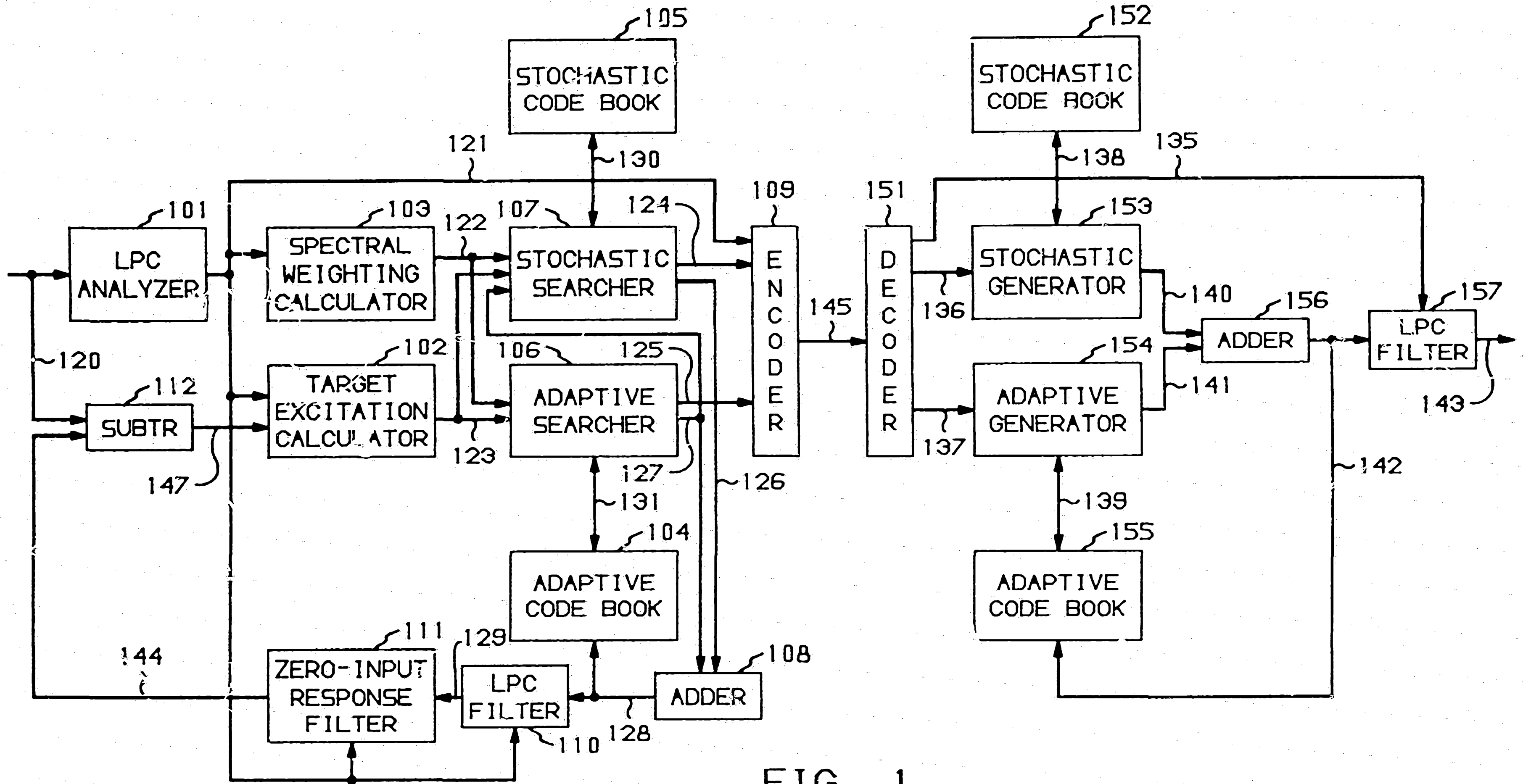


FIG. 1

1/5

18 384/88

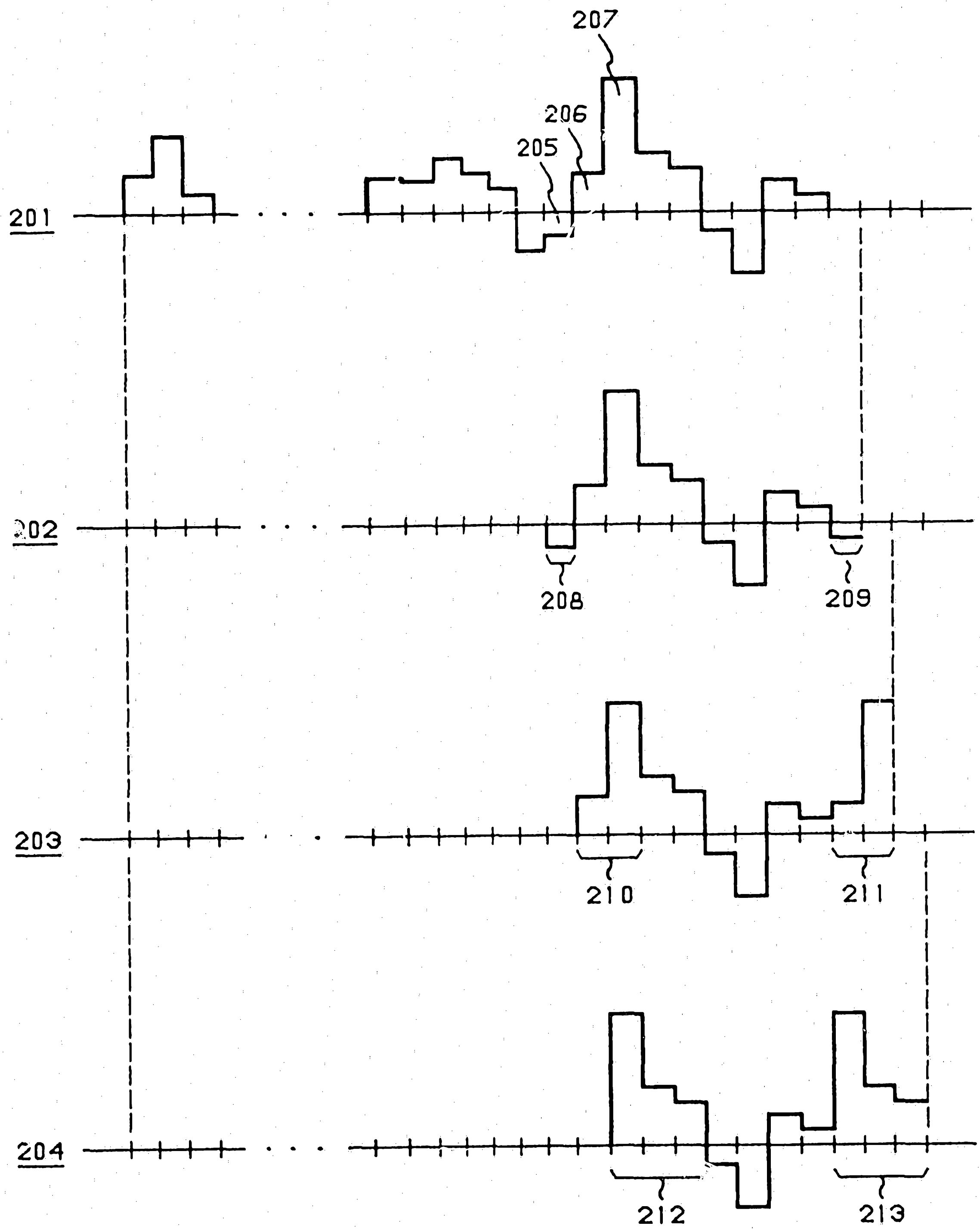


FIG. 2

$$r_1^T H^T H r_1 = \begin{bmatrix} X_0 & X_1 & X_2 & X_3 & X_4 \end{bmatrix} \begin{bmatrix} A_0 & A_1 & A_2 & A_3 & A_4 \\ A_1 & A_0 & A_1 & A_2 & A_3 \\ A_2 & A_1 & A_0 & A_1 & A_2 \\ A_3 & A_2 & A_1 & A_0 & A_1 \\ A_4 & A_3 & A_2 & A_1 & A_0 \end{bmatrix} \begin{bmatrix} X_0 \\ X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}$$

FIG. 3

$$r_2^T H^T H r_2 = \begin{bmatrix} X_1 & X_2 & X_3 & X_4 & X_5 \end{bmatrix} \begin{bmatrix} A_0 & A_1 & A_2 & A_3 & A_4 \\ A_1 & A_0 & A_1 & A_2 & A_3 \\ A_2 & A_1 & A_0 & A_1 & A_2 \\ A_3 & A_2 & A_1 & A_0 & A_1 \\ A_4 & A_3 & A_2 & A_1 & A_0 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \end{bmatrix}$$

FIG. 4

$$r_1^T H^T H r_1 = \begin{bmatrix} \overset{501}{X_0} & X_1 & X_2 & X_3 & X_4 \end{bmatrix} \overset{502}{\underset{504}{\begin{bmatrix} A_0 & A_1 & A_2 & A_3 & A_4 \\ A_1 & A_0 & A_1 & A_2 & A_3 \\ A_2 & A_1 & A_0 & A_1 & A_2 \\ A_3 & A_2 & A_1 & A_0 & A_1 \\ A_4 & A_3 & A_2 & A_1 & A_0 \end{bmatrix}}} \overset{503}{\begin{bmatrix} X_0 \\ X_1 \\ X_2 \\ X_3 \\ X_4 \end{bmatrix}}$$

FIG. 5

$$r_2^T H^T H r_2 = \begin{bmatrix} X_1 & X_2 & X_3 & X_4 & \overset{601}{X_5} \end{bmatrix} \overset{604}{\underset{602}{\begin{bmatrix} A_0 & A_1 & A_2 & A_3 & A_4 \\ A_1 & A_0 & A_1 & A_2 & A_3 \\ A_2 & A_1 & A_0 & A_1 & A_2 \\ A_3 & A_2 & A_1 & A_0 & A_1 \\ A_4 & A_3 & A_2 & A_1 & A_0 \end{bmatrix}}} \underset{603}{\begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ X_4 \\ X_5 \end{bmatrix}}$$

FIG. 6

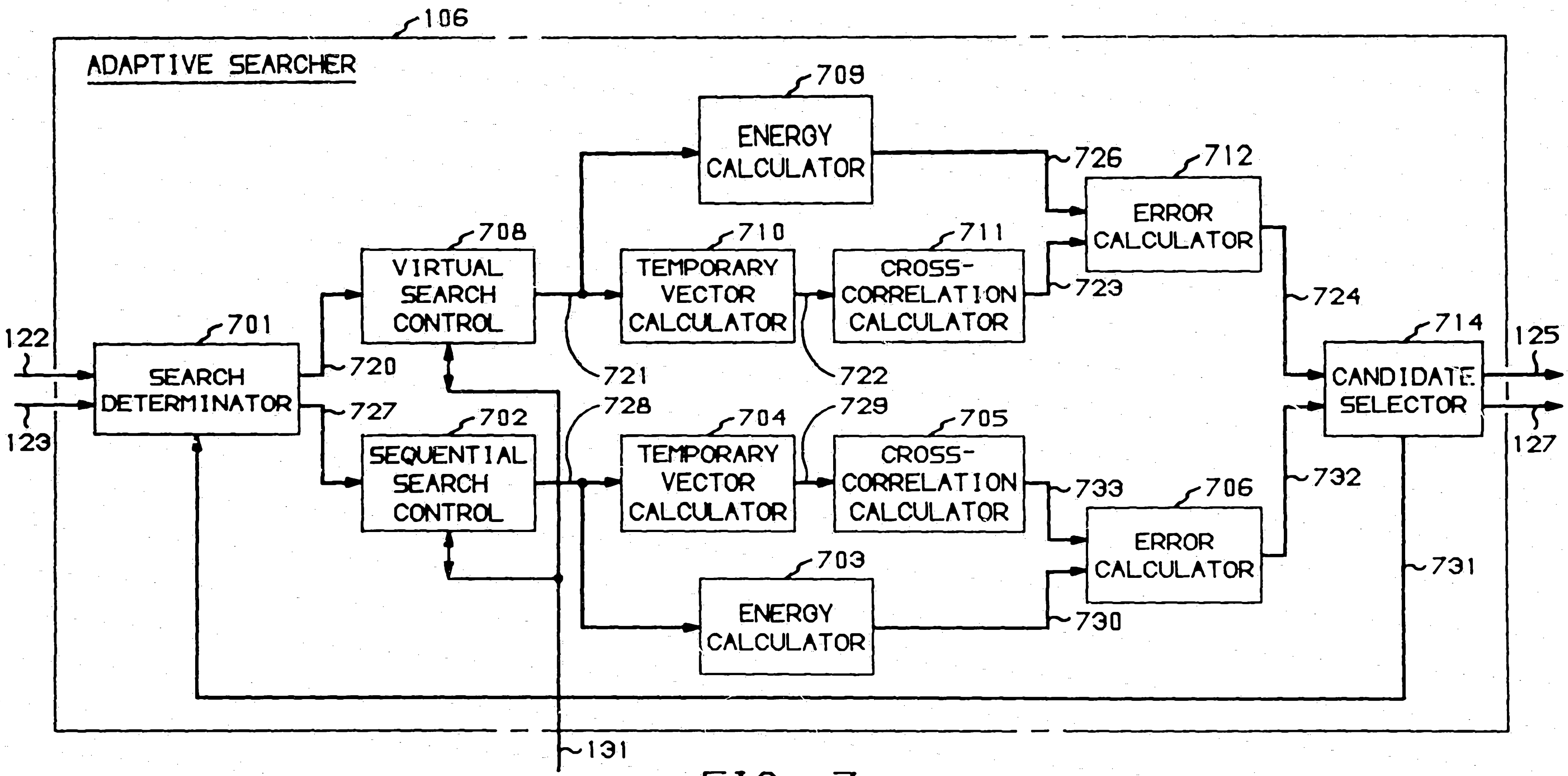
106
ADAPTIVE SEARCHER
709
ENERGY CALCULATOR
726
712
ERROR CALCULATOR
708
VIRTUAL SEARCH CONTROL
710
TEMPORARY VECTOR CALCULATOR
711
CROSS-CORRELATION CALCULATOR
723
724
701
SEARCH DETERMINATOR
122
123
720
721
722
714
CANDIDATE SELECTOR
125
127
727
702
SEQUENTIAL SEARCH CONTROL
728
704
TEMPORARY VECTOR CALCULATOR
729
705
CROSS-CORRELATION CALCULATOR
733
706
ERROR CALCULATOR
732
703
ENERGY CALCULATOR
730
731
131

FIG. 7

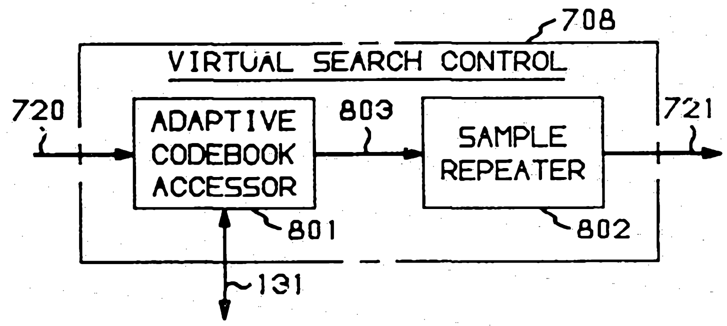
708
VIRTUAL SEARCH CONTROL
720
ADAPTIVE CODEBOOK ACCESSOR
803
SAMPLE REPEATER
721
801
802
131

FIG. 8

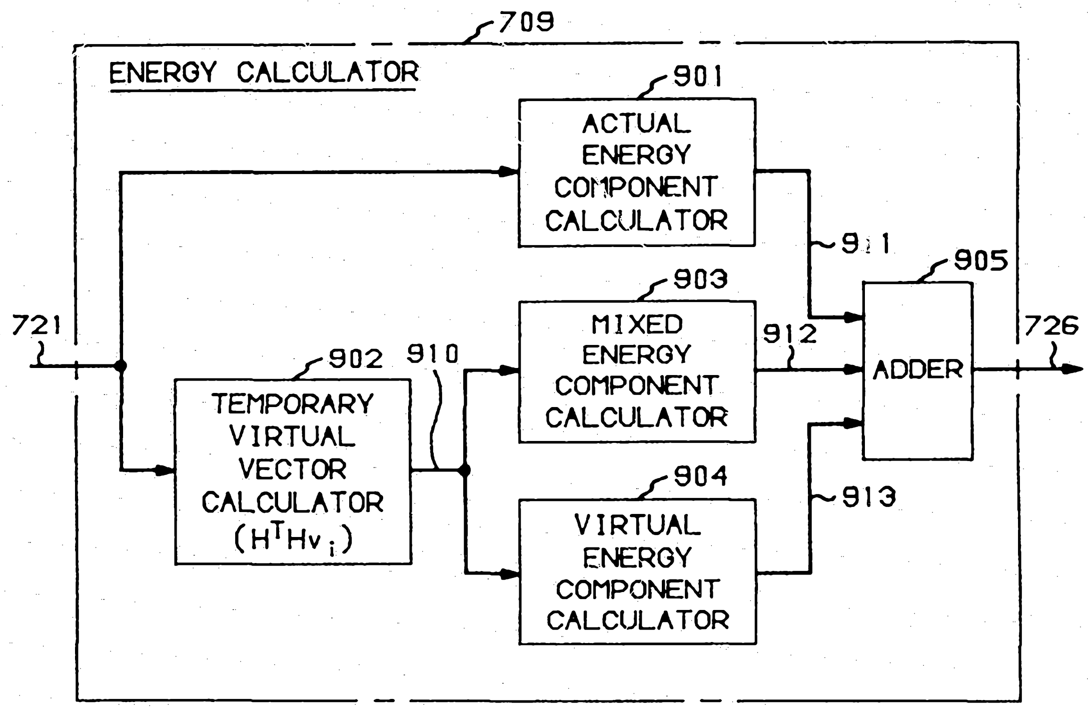
709
ENERGY CALCULATOR
901
ACTUAL ENERGY COMPONENT CALCULATOR
911
905
903
MIXED ENERGY COMPONENT CALCULATOR
912
726
721
902
910
ADDER
TEMPORARY VIRTUAL VECTOR CALCULATOR ($H^T H v_i$)
904
VIRTUAL ENERGY COMPONENT CALCULATOR
913

FIG. 9