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

(11) Publication number: **0 381 507 A2**

(12) EUROPEAN PATENT APPLICATION

(21) Application number: **90301081.7**

(51) Int. Cl.5: **G10L 3/00**

(22) Date of filing: **01.02.90**

(30) Priority: **02.02.89 JP 22512/89**
02.02.89 JP 22522/89
01.07.89 JP 168310/89

(43) Date of publication of application:
08.08.90 Bulletin 90/32

(84) Designated Contracting States:
DE FR GB

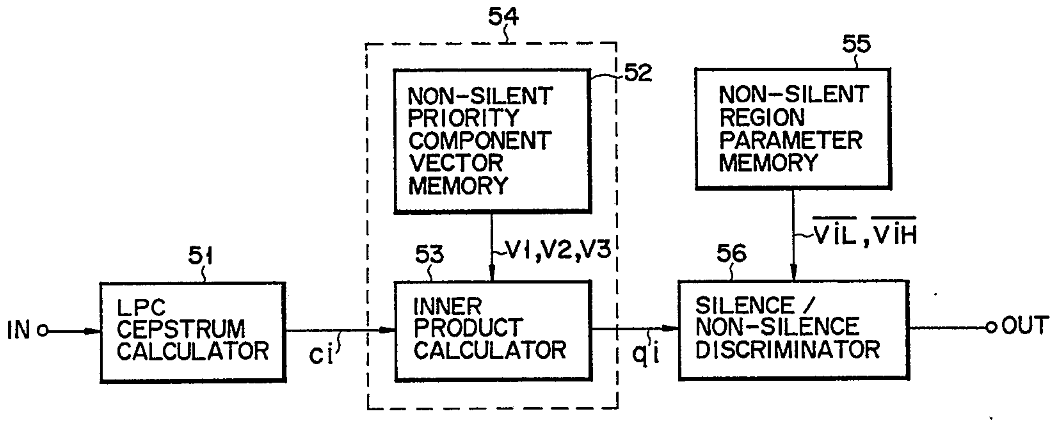
(71) Applicant: **KABUSHIKI KAISHA TOSHIBA**
72, Horikawa-cho Saiwai-ku
Kawasaki-shi Kanagawa-ken 210(JP)

(72) Inventor: **Akamine, Masami, c/o Intellectual Property Div.**
Kabushiki Kaisha Toshiba, 1-1 Shibaura 1-chome
Minato-ku, Tokyo 105(JP)

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

(54) Silence/non-silence discrimination apparatus.

(57) A speech signal is input to an LPC cepstrum calculator (51) and the LPC cepstrums of the speech signal for each frame are calculated as characteristic parameters. The cepstrum is input to a characteristic parameter projection circuit (54) including an inner product calculator (53) and a memory (52) storing first to third priority component vectors that are obtained by applying a priority component analysis to the LPC cepstrums of the non-silent parts of the speech. The inner product calculator (53) calculates inner products of the cepstrum vector and the priority component vectors stored in the priority component vector memory (52) to obtain a projected point of the LPC cepstrums in a vector space formed by the first to third priority component vectors. The output of the inner product calculator (53) is supplied to a silence/non-silence discriminator (56) to which a non-silent region parameter memory (55) storing parameters defining a non-silent region in the non-silent priority component vector space. The silence/non-silence discriminator (56) determines if the speech is silent or non-silent based on whether the projected point is within the non-silent region.


54
52
NON-SILENT PRIORITY COMPONENT VECTOR MEMORY
55
NON-SILENT REGION PARAMETER MEMORY
V1,V2,V3
$\overline{\text{VIL}}$, $\overline{\text{VIH}}$
51
53
56
IN
LPC CEPSTRUM CALCULATOR
ci
INNER PRODUCT CALCULATOR
qi
SILENCE / NON-SILENCE DISCRIMINATOR
OUT

F I G. 5

EP 0 381 507 A2

Silence/non-silence discrimination apparatus

The present invention relates to a silence/non-silence discrimination apparatus adaptable for an ATM (Asynchronous Transfer Mode) communication system in which only a non-silent part of a speech signal is divided into cells before transmitted, a recorder for recording only the non-silent part of the speech signal, and a circuit for extracting a recognition frame as a basic technique of speech recognition.

In the apparatus for processing only the non-silent part of the speech signal, if the silence/non-silence discrimination is not exacted, a transmitted speech is interrupted or an error rate of speech recognition increases. In the ATM communication system, it is impossible to effectively use the communication line. For this reason, an accurate silence/non-silence discrimination has been required. To cope with this, there is a proposal as disclosed in Unexamined Publication Japanese Patent Application No. 60-200300. This proposal is a silence/non-silence discrimination apparatus which can detect a non-silent part of the speech signal having a low level, such as a word head consonant, with a lessened failure of its discrimination even when a signal level varies due to change of ambient conditions and an ambient noise level is large.

Fig. 1 shows a block diagram of the silence/non-silence discrimination apparatus as disclosed in the above-identified Japanese Application. A speech signal input through a microphone, for example, is supplied to an energy extraction circuit 5 and a spectrum extraction circuit 6. The energy extraction circuit 5 includes a smoothing circuit and extracts a power (logarithmic power) as a characteristic parameter of the speech signal every frame period of a predetermined time duration, which is an unit of time for silence/non-silence discrimination. The spectrum extraction circuit 6 includes three types of band-pass filters of low frequencies (250 to 600 Hz), medium frequencies (600 to 1500 Hz), and high frequencies (1500 to 4000 Hz), and three smoothing circuits respectively coupled with the output terminals of those filters. The circuit 6 also extracts a power (logarithmic power) for each frequency band as a characteristic parameter of the speech signal every frame period. The energy extraction circuit 5 and the spectrum extraction circuit 6 form a characteristic parameter extraction circuit 13.

The output signals of the energy extraction circuit 5 and the spectrum extraction circuit 6 are supplied to a multiplexer 7. The multiplexer supplies the signal power from the energy extraction circuit 5 and the frequency band powers from the spectrum extraction circuit 6 to a silence/non-silence discriminator 8 in a time division manner. The discriminator 8 discriminates each frame of the speech signal as being silent or non-silent. Incidentally, the non-silent frame includes a voiced speech frame and a non-voiced speech frame. The discriminator 8 is connected with a threshold value memory 9 and a standard pattern memory 10. The memory 9 stores two threshold values E1 and E2 that are used for determining if the frame is silent or non-silent on the basis of the power. The memory 10 stores a coefficient of a linear discrimination function, which is used for determining if the frame to be detected is a silent frame or a non-voiced speech frame, a coefficient of a linear discrimination function, which is used for determining if the frame is a silent frame or a voiced speech frame, a standard pattern for determining if the frame is a silent frame or a non-voiced speech frame, and a standard pattern for determining if the frame is a silent frame or a voiced speech frame. These threshold values, the coefficients, and the standard patterns are previously obtained by utilizing the statistical feature of a speech signal containing silent frame, voiced speech frame, non-voiced speech frame, that is generated under the condition for using the silence/non-silence discrimination apparatus, and stored in the memories.

The discriminator 8 produces a signal denoting the determination result, and supplies it to a detector 11 for detecting the candidate frames of the starts and the ends of the non-silent part of the speech on the basis of the determination result for each frame. The result of the detection is supplied to a non-silent detector 12 where the start and end of the non-silent part are finally determined.

An operation of the above mentioned prior silence/non-silence discrimination apparatus will be described. A speech signal of each frame is transformed into a power LPW by the energy extraction circuit 5. The same is also transformed into a power LPi (i is a parameter indicative of a frequency band and is any of i to 3) of each frequency band by the spectrum extraction circuit 6. The discriminator 8 determines if the frame is silent or non-silent, by using those four parameters LPW and LPi, the threshold values E1 and E2 and the coefficients of the linear discrimination functions stored in the memories 9 and 10.

For the determination, the two threshold values E1 and E2, and the power LPW are first compared in the following way:

If LPW > E1, it is discriminated that the frame is non-silent,
if LPW < E2, it is discriminated that the frame is silent, and
if E2 ≦ LPW ≦ E1, it is discriminated that the property of the frame is "indefinite".

When the discrimination is "indefinite", another determination is made by using a following discrimination function value FX.

$$FX = \sum_{i=1}^{3} Ai\ (Lpi - \overline{LPi}) \quad \ldots(1)$$

where Ai is the coefficient of the linear discrimination function and $\overline{LPi}$ is the standard pattern both stored in the memory 10.

The function value FX is negative for the silent frame, and is positive for the non-silent frame including the voiced speech frame and the non-voiced speech frame. Calculation is made of the FX when the coefficient Ai and the standard pattern $\overline{LPi}$ are those for determining if the frame is the silent frame or the non-voiced speech frame, and calculation is made of the FX when the coefficient Ai and the standard pattern $\overline{LPi}$ are those for determining if the frame is the silent frame or the voiced speech frame. When either of the calculated FXs is positive, it is determined that the frame is the non-silent frame. In other cases, it is determined that the frame is the silent frame.

The prior apparatus makes the silence/non-silence determination on the basis of a difference between a spectral shape extracted by the spectrum extraction circuit 6 and each of standard spectral shapes of silent frame, non-voiced speech frame, voiced speech frame, for every frame. Therefore, the apparatus may reliably discriminate the property of the speeches of small energy level, such as silent consonants and non-silent consonants. The powers of three frequency bands, low frequencies (250 to 600 Hz), medium frequencies (600 to 1500 Hz), and high frequencies (1500 to 4000 Hz), are used as the characteristic parameters of spectral shapes. However, the selection of the characteristic parameters has no theoretical basis and the number of the characteristic parameters is relatively small. These lead to incorrect silence/non-silence discrimination, failure of detecting the non-silent signal, and increase of noise.

Consider a case where a spectrum of a non-voiced speech and that of noise are shaped as indicated by a solid line and a broken line in Fig. 2. In this case, when Ai (i = 1 to 3) = 1, the function value FX has the same value for both the non-voiced speech and the noise, although the both spectral shapes are greatly different from each other. This results in incorrect silence/non-silence discrimination. The incorrect discrimination is due to the small number of characteristic parameters defining the spectral shape and the improper selection of the parameters. Further, since the parameter selection is lack of the theoretical basis, the selection must be made in a trial-and-error manner. Therefore, much time and labor are consumed during selection of the parameters, but the results are frequently incorrect. If the number of the characteristic parameters is increased, a frequency of the erroneous discrimination-determination is reduced, but the work to calculate the discrimination functions value FX as given by the equation (1) is increased. The above mentioned Japanese Application describes that the discrimination function may be replaced by the Mahalanobis's distance. However, if the Mahalanobis's distance is used, the calculation work is further increased.

Accordingly, an object of the present invention is to provide a silence/non-silence discrimination apparatus which can accurately discriminate a silent part and a non-silent part of a speech signal, with a simple construction.

To achieve the above object, there is provided a silence/non-silence discrimination apparatus comprising means for obtaining a plurality of characteristic parameters from a speech signal; means for projecting the plurality of characteristic parameters onto a priority component vector space of the characteristic parameters of a given type of speech signal, the dimension of the priority component vector space being smaller than the number of the plurality of characteristic parameters; and means for discriminating whether the speech signal is silent or non-silent based on the position of the projected point of the plurality of characteristic parameters.

According to the present invention, a priority component analysis is applied to the characteristic parameters. Therefore, the number of characteristic parameters can be reduced while minimizing the loss of the amounts of information that are possessed by the original characteristic parameters. The silence/non-silence discrimination apparatus according to the present invention has an excellent discrimination accuracy, and has a simple construction. With the priority component analysis, the statistical nature of the characteristic parameters is reflected on the silence/non-silence discrimination. This eliminates the try-and-error process that is essential to the prior apparatus to obtain optimum parameters.

Additional objects and advantages of the invention will be set forth in the description which follows, and in part will be obvious from the description, or may be learned by practice of the invention. The objects and advantages of the invention may be realized and obtained by means of the instrumentalities and combinations particularly pointed out in the appended claims.

This invention can be more fully understood from the following detailed description when taken in conjunction with the accompanying drawings, in which:

The accompanying drawings, which are incorporated in and constitute a part of the specification, illustrate presently preferred embodiments of the invention and, together with the general description given above and the detailed description of the preferred embodiments given below, serve to explain the principles of the invention.

Fig. 1 is a block diagram showing a prior silence/non-silence discrimination apparatus;

Fig. 2 is a graph showing spectral shapes for explaining the operation of the apparatus of Fig. 1;

Fig. 3 is a block diagram for explaining a scheme of a silence/non-silence discrimination according to the present invention;

Fig. 4 is a block diagram showing a speech cell generation apparatus which includes a first embodiment of a silence/non-silence discrimination apparatus according to the present invention;

Fig. 5 is a block diagram showing the first embodiment of the silence/non-silence discrimination apparatus according to the present invention;

Fig. 6 is a flowchart for explaining a sequence of procedural steps to obtain data to be stored in a non-silent priority component vector memory used in the first embodiment;

Fig. 7 is a graph showing a non-silent region in the non-silent priority component vector space, which provides a reference for the silence/non-silence discrimination in the first embodiment;

Fig. 8 is a flowchart for explaining an operation of the silence/non-silence discrimination of the first embodiment;

Fig. 9 is a block diagram showing a second embodiment of a silence/non-silence discrimination apparatus according to the present invention;

Fig. 10 is a flowchart for explaining an operation of the silence/non-silence discrimination of the second embodiment;

Fig. 11 is a block diagram showing a third embodiment of a silence/non-silence discrimination apparatus according to the present invention;

Fig. 12 is a block diagram showing a characteristic parameter projection circuit in the third embodiment;

Fig. 13 is a block diagram showing a detection circuit in the third embodiment;

Fig. 14 is a flowchart for explaining an operation of the silence/non-silence discrimination of the third embodiment;

Fig. 15 is a block diagram showing a fourth embodiment of a silence/non-silence discrimination apparatus according to the present invention;

Fig. 16 is a flowchart for explaining an operation of the silence/non-silence discrimination of the fourth embodiment;

Fig. 17 is a block diagram showing a fifth embodiment of a silence/non-silence discrimination apparatus according to the present invention;

Fig. 18 is a block diagram showing a first example of an FIR (Finite Impulse Response) used in the fifth embodiment;

Fig. 19 is a block diagram showing a second example of the FIR filter in the fifth embodiment;

Fig. 20 is a block diagram showing a matching circuit in the fifth embodiment;

Fig. 21 is a flowchart showing a sequence of procedural steps for obtaining a reference pattern to be stored in a reference pattern memory of the matching circuit in the fifth embodiment;

Fig. 22 is a flowchart for explaining an operation of the silence/non-silence discrimination of the fifth embodiment;

Fig. 23 is a block diagram showing a sixth embodiment of a silence/non-silence discrimination apparatus according to the present invention;

Fig. 24 is a block diagram showing a matching circuit in the sixth embodiment; and

Fig. 25 is a flowchart for explaining an operation of the silence/non-silence discrimination of the sixth embodiment.

Preferred embodiments of a silence/non-silence discrimination apparatus according to the present invention will be described with reference to the accompanying drawings. A scheme of a silence/non-silence discrimination according to the present invention is illustrated in Fig. 3.

Firstly, the characteristic parameters of a speech signal are previously obtained by a known method. The characteristic parameters, that are expressed by using the spectrum in the prior art, may be expressed by using the LPC (linear Predictive Coding) cepstrum, signal power, the number of zero-crossings, linear predictive coefficient, auto-correlation function, DFT (Digital Fourier Transformation) coefficient, and any of their combinations. In the present invention, selection of the number of and the kind of characteristic parameters is not required, and therefore it is preferable to obtain the largest possible number of and many kinds of characteristic parameters.

Secondly, the number of characteristic parameters is reduced to such an extent that the reduction does not adversely affect the accuracy of the silence/non-silence discrimination. To effect this, the characteristic parameters are transformed into another type of parameters. Then, the number of the transformed parameters is reduced. The transformation is made such that when the transformed parameters, after the number of the parameters is reduced, are inversely transformed into the original characteristic parameters, an error between the inversely transformed characteristic parameters and the original characteristic parameters is minimized.

The scheme of a reduction of the parameters will be described in more detail with reference to Fig. 3. L number of original characteristic parameters are expressed by xi (i = 1 to L). A vector represented by the characteristic parameters xi as elements is expressed as X. A transformation employed is an orthogonal transformation. A transformation matrix is expressed as "A". The transformed characteristic parameters are expressed by yi (i = 1 to L). A vector represented by the characteristic parameters yi as elements is expressed as Y. Y indicates a vector formed of such characteristic parameters that of the transformed characteristic parameters yi, N number of parameters yj (j = 1 to N) are left, and the remaining (L - N) number of characteristic parameters (where N < L) are set to 0. An error vector "e", which is caused by reducing the number of characteristic parameters, is a difference between the vector X of the original characteristic parameters and a vector $A^{-1}\tilde{Y}$ that results from the inverse transformation of the transformed characteristic parameter vector $\tilde{Y}$ with the reduced number of parameters, and is mathematically expressed as follows.

$$\begin{aligned} e &= X - A^{-1}\tilde{Y} \\ &= A^{-1}(Y - \tilde{Y}) \qquad (2) \end{aligned}$$

The error due to the parameter reduction can be minimized by using such a transformation as to minimize the square mean value of the above error $\sigma r^2 = E\,[e^t e]$ (where t is a transposition of a matrix and E [ ] indicates an expected value).

The transformation to minimize the square mean value of the difference expressed by the equation (2) is known as a KL transformation, or a transformation whose transformation matrix is a matrix "A" having an eigen vector of an auto-correlation matrix of the parameters xi as a row vector. The eigen vector is equivalent to a priority component vector resulting from the priority component analysis of the parameters xi. In a descending order of the eigen values, the eigen vectors are made to correspond respectively to a first priority component vector, a second priority component vector, and so on.

Assuming that M number of characteristic parameter vector is expressed as Xi (i = 1 to M), the transformation to minimize the square mean value of the difference expressed by the equation (2), for each vector Xi, is defined by an eigen vector that is obtained by the priority component analysis of an auto-correlation matrix of the characteristic parameter vector as given by the following equation.

$$\begin{aligned} \overline{R} &= \frac{1}{M}\sum_{i=1}^{M} XiXi^t \\ &= \frac{1}{M}\sum_{i=1}^{M} \begin{bmatrix} xi_1 \\ xi_2 \\ \vdots \\ xi_L \end{bmatrix} [xi_1,\ xi_2,\ \ldots\ xi_L] \\ &= \frac{1}{M}\sum_{i=1}^{M} \begin{bmatrix} xi_1^2 & xi_1xi_2 & \ldots & xi_1xi_2 \\ xi_2xi_1 & xi_1^2 & \ldots & xi_2xi_L \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ \vdots & & & \vdots \\ xi_Lxi_1 & xi_Lxi_2 & \ldots & xi_L^2 \end{bmatrix} \qquad \ldots(3) \end{aligned}$$

where xi1, xi2, ... xiL are elements of the characteristic parameter vector Xi.

As seen from the equation (3), the auto-correlation matrix R is an auto-correlation matrix Ri of each characteristic parameter vector Xi which is given by the following equation (4) and is averaged in an L^2 dimension, that is, a center-of-gravity.

$$Ri = \begin{bmatrix} xi_1{}^2 & xi_1xi_2 & \ldots & xi_1xi_2 \\ xi_2xi_1 & xi_1{}^2 & \ldots & xi_2xi_L \\ & & & \\ xi_Lxi_1 & xi_Lxi_2 & \ldots & xi_L{}^2 \end{bmatrix} \qquad \ldots(4)$$

If a single auto-correlation matrix $\bar{R}$ represents auto-correlation matrices Ri (i = 1, 2, ... M) of the M number of characteristic parameter vectors Xi, then the auto-correlation matrix R is a matrix to minimize the square mean error E of the matrices Ri and $\bar{R}$, that is give by the following equation (5).

$$E = \frac{1}{M} \sum_{i=1}^{M} \sum_{k=1}^{L} \sum_{\ell=1}^{L} (\bar{R}(k,\ell) - Ri(k,\ell)) \qquad \ldots(5)$$

The reason for this follows. The above relation is partially differentiated with respect to $\bar{R}$(k, l), and let the result of the differentiation be 0. Then, the following equations (6) and (7) are obtained.

$$\sum_{i=1}^{M} (\bar{R}(k,\ell) - Ri(k,\ell)) = 0 \qquad \ldots(6)$$

$$\bar{R}(k,\ell) = \frac{1}{M} \sum_{i=1}^{M} Ri(k,\ell) \qquad \ldots(7)$$

where $\bar{R}$(k, l) and Ri(k, l) are the (k, l) elements of the auto-correlation matrices $\bar{R}$ and Ri.

By KL transforming the characteristic parameter vector Xi by using such a transformation matrix $\bar{R}$, the priority component analysis is realized.

An operation to KL transform L the number of characteristic parameters xi (i = 1 to L), and then to reduce the number of the transformed characteristic parameters is equivalent to project the characteristic parameter vector X onto an N-dimension priority component vector space with the coordinate axes being the first to N-th priority component vectors. With this projection, the number of the original characteristic parameters may be reduced while minimizing the error due to the number of parameter reduction.

Fig. 4 shows the block diagram of the speech cell generation device which is used in the ATM communication system and incorporates the silence/non-silence discrimination apparatus based on the above mentioned silence/non-silence discrimination system. A speech signal is supplied to a sound encoder 41 and a noise encoder 42, where it is encoded. The coding rates of the encoders 41 and 42 are different from each other, and the coding rate or the bit rate of the encoder 41 is higher than that of the encoder 42. An ADPCM (Adaptive Differential Pulse Code Modulation) coding system is used for the coding system of the encoders 41 and 42. One of the output signals of the encoders 41 and 42 is supplied through a selector 45 to a cell generation circuit 46. When receiving the coded speech signal, the circuit 46 generates corresponding cells. In response to a signal representative of the discrimination result derived from a silence/non-silence discrimination device 43, the selector 45 selects the output of the sound encoder 41 when the non-silent signal is detected and the output of the noise encoder 42 when the silent signal is detected. The noise is encoded and transmitted in order to impart a natural feeling to the transmitted speech. The transmission of noise little degrades the efficiency of the line usage, because the bit rate thereof is low. Therefore, when the silence/non-silence discrimination device 43 detects a silent part of the speech signal, the selector 45 connects to the noise encoder 42.

The noise (silent frame) is transmitted to the receiver only in the initial stage, e.g., when connection to the communication line starts, and subsequently the transmission of the noise is stopped, whereas the transmitted noise is repetitively reproduced in the receiver. When the transmitter detects a change in the noise, the noise is transmitted again. Furthermore, only the sound signal is transmitted while the noise is not transmitted. Accordingly, in this case, unnaturalness involved in the transmitted sound must be tolerated. Also in this case, if necessary, white noise may be inserted in the receiving side.

Fig. 5 is a block diagram showing the first embodiment of the silence/non-silence discrimination apparatus according to the present invention. In this embodiment, the LPC cepstrum is used for the characteristic parameter of the speech signal. Therefore, an LPC cepstrum calculator 51 is coupled with an input terminal for the speech signal. The calculator 51 calculates the LPC cepstrums c_i (i = 1, 2, ... L) of the speech signal for each frame of a fixed time. The number L of the parameters indicates an order of analysis, e.g., is set to 16. In this invention, the number of the parameters is reduced after the priority component analysis. Accordingly, the number L may be more than 16. For the cepstrum calculation, reference is made to Alan V. Oppenhiem and Ronald W. Shafer, "Digital Signal Processing" (Prentice Hall Inc., NJ, 1975).

It is assumed that a vector formed by the cepstrums c_1, c_2, ... c_L is C. The vector C is input to an inner product calculator 53. The inner product calculator 53, together with a non-silent priority component vector memory 52, forms a characteristic parameter projection circuit 54. The memory 52 stores first to third priority component vectors V1, V2, and V3 that are obtained by applying a priority component analysis to the LPC cepstrums of the non-silent parts of the speech signal collected under the condition for using the silent/non-silent discrimination apparatus. Here, the element of the priority component vector Vi (i = 1 to 3) is denoted as vij (j = 1, 2, ... L).

A sequence of procedural steps to obtain the non-silent priority component vectors to be stored in the non-silent priority component vector memory 52 is shown in the form of a flowchart in Fig. 6. In step #1, learning speech data are collected under the condition for using the discrimination apparatus. In step #2, only the non-silent data are extracted from all of the collected speech data. In step #3, the LPC cepstrums of the non-silent data are calculated. In step #4, the priority component analysis is applied to the LPC cepstrums. More exactly, an auto-correlation matrix of the LPC cepstrum vector is calculated. In step #5, the eigen values and the eigen vectors of the matrix are calculated. In step #6, the eigen vectors corresponding to the eigen values in the descending order from the largest absolute value of the eigen values to the smallest absolute value are set to first, second, ... N-th (here, N = 3) priority component vectors. As a result, the non-silent priority component vectors V1, V2, and V3 are obtained.

Returning to Fig. 5, the inner product calculator 53 calculates the inner product of the cepstrum vector C and the priority component vector Vi to obtain a projected point Q of the LPC cepstrum spectrum vector C (= c_1, c_2, ... c_L)^t in a three dimensional vector space formed of the first to third priority component vectors V1, V2, and V3, as given in the following way.

$$qi = \sum_{j=1}^{L} cjvij \qquad \ldots(8)$$

where qi is a component of the projected point Q in the Vi direction.

The output signal of the inner product calculator 53 is supplied to a silence/non-silence discriminator 56. A non-silent region parameter memory 55 storing parameters defining a non-silent region in the non-silent priority component vector space is also connected to the discriminator 56. Assuming that the non-silent region takes the form of a rectangular parallelepiped as shown in Fig. 7, the region defining parameters are $\overline{V_{1L}}$, $\overline{V_{1H}}$, $\overline{V_{2L}}$, $\overline{V_{2H}}$, $\overline{V_{3L}}$, and $\overline{V_{3H}}$defining the upper and lower limits along the direction of the coordinate axes. Those parameters are previously obtained by statistically processing the LPC cepstrums of the non-silent parts and the silent parts (including noises) of the speech signal that are collected under the condition for using the silence/non-silence discrimination apparatus. The discriminator 56 determines if the frame is silent or non-silent based on whether the projected point is within the non-silent region of the rectangular parallelepiped of Fig. 7. Only when $\overline{V_{1L}} \leqq q_1 \leqq \overline{V_{1H}}$, $\overline{V_{2L}} \leqq q_2 \leqq \overline{V_{2H}}$, and $\overline{V_{3L}} \leqq q_3 \leqq \overline{V_{3H}}$, the discriminator 56 determines that the frame is non-silent. In other cases, it determines that the frame is silent. A sequence of the procedural steps to determine if the frame is silent or non-silent by the discriminator 56 is shown in Fig. 8.

In the above description, the determination of the silent or non-silent frame depends on whether or not the projected point is within the non-silent region in the non-silent priority component vector space. Alternatively, it may be done by using a distance between the center-of-gravity of the non-silent region and the projected point Q. In this case, the center-of-gravity G of the non-silent region is set at coordinates (g_1,

g_2, g_3). A distance D as given below is compared with a predetermined threshold value Th. If D ≦ Th, it is determined the frame is non-silent, and if D > Th, it is determined that the frame is silent.

$$D = \sum_{i=1}^{3} Ai(q_i - g_i)^2 \qquad ...(9)$$

where Ai is a weighting coefficient.

According to the first embodiment of the silence/non-silence discrimination apparatus, the L number of characteristic parameters are projected onto the vector space defined by the non-silent priority component vectors. The determination as to whether the frame is silent or non-silent depends on whether or not the projected point is within the non-silent region. This brings about the following advantages. The number of the characteristic parameters used for the actual determination is reduced. The calculation work is reduced, accordingly. The circuit for the determination has a simple construction. Since the parameters are projected onto the priority component vector space, the reduction of the number of the parameters little damages the accuracy of the silence/non-silence discrimination. Since the discrimination depends on the region, even when the non-silent region and the silent region occupy special regions in the priority component vector space, the high accurate silence/non-silence discrimination is ensured. In the silence/non-silence discrimination based on the distance, the distance definition determines a shape of the non-silent region. For example, if in the equation (9), Ai = 1 (i = 1 to 3), the region satisfying D ≦ Th is inside a sphere. Thus, in the distance-dependent discrimination, it is impossible to flexibly select a shape of the non-silent region. On the other hand, the discrimination dependent on the region allows the non-silent region to take any shape.

The first embodiment is not limited to the above description. It is possible to project the characteristic parameters onto the silent priority component vectors, not the non-silent priority component vectors. It is possible to discriminate whether or not the projected point is within the silent region instead of whether or not the projected point within the non-silent region. The LPC cepstrums as the characteristic parameters may be replaced with any of the spectrum, signal power, the number of zero-crossings, linear predictive coefficient, auto-correlation function, and DFT coefficient, and any of their combinations which are used in the prior art. The specific figures, such as the number of the characteristic parameters and the dimension of the priority component spectrum space, may be appropriately selected.

Fig. 9 is a block diagram showing a second embodiment of a silence/non-silence discrimination apparatus according to the present invention. An LPC cepstrum calculator 62 is connected to the input terminal, and calculates the LPC cepstrums c_i (i = 1, 2, L) of an input speech signal for each frame, as in the first embodiment. The cepstrums are supplied to inner product calculators 63 and 64. The inner product calculators 63 and 64 are respectively coupled with a non-silent priority component vector memory 65 and a silent priority component vector memory 66. The calculators 63 and 64, and the memories 65 and 66 form a characteristic parameter projection circuit 67. Thus, in the second embodiment, the characteristic parameters (cepstrums) are projected onto both the non-silent priority component vector space and the silent priority component vector space. The memory 65 stores first to third priority component vectors that are obtained by applying a priority component analysis to the LPC cepstrums of the non-silent parts of the speech signal collected under the condition for using the discrimination apparatus, as in the first embodiment. The memory 66 stores first to third priority component vectors that are obtained by applying a priority component analysis to the LPC cepstrums of the silent parts of the speech signal collected under the condition for using the discrimination apparatus. The inner product calculators 63 and 64 each obtain a projected point Q of the LPC cepstrum vector C in a three dimensional vector space formed by the first to third priority component vectors V1, V2, and V3, like the inner product calculator 53 in the first embodiment.

The output signals of the inner product calculators 63 and 64 are supplied to a non-silence detector 68 and a silence detector 69, respectively. The detectors 68 and 69 are respectively coupled with a non-silent region parameter memory 70 and a silent region parameter memory 71. The memory 70 stores the parameters defining a non-silent region in the non-silent priority component vector space. The memory 71 stores the parameters defining a silent region in the silent priority component vector space. The parameters define the upper and lower limits along the coordinate axes, as in the first embodiment. Also as in the first embodiment, the non-silent detector 68 compares the coordinates of the projected point Q with those parameters, and produces a detection signal of "1" level when the projected point exists within the non-silent region. The silence detector 69 compares the the coordinates of the projected point Q with those parameters, and produces a detection signal of "1" level when the projected point exists within the silent region. The output signals of the detectors 68 and 69 are supplied to a silence/non-silence discriminator 72. The discriminator 72 finally determines that the speech frame is silent or non-silent in the following way.

The frame is determined as being:

(1) Non-silent when the output signal of the non-silence detector 68 is "1" level and the output signal of the silence detector 69 is "0" level;

(2) Silent when the output signal of the non-silence detector 68 is "0" level and the output signal of the silence detector 69 is "1" level;

(3) Non-silent when the output signal of the non-silence detector 68 is "1" level and the output signal of the silence detector 69 is "1" level; and

(4) Non-silent when the output signal of the non-silence detector 68 is "0" level and the output signal of the silence detector 69 is "0" level.

Thus, only when both the detectors 68 and 69 determine that the frame is silent, the discriminator 72 determines that the frame is silent. In other cases, the discriminator 72 determines that the frame is non-silent. A flow of the determination is shown in Fig. 10.

As described above, according to the second embodiment, the silence/non-silence determination is based on the positions of the projected points in both the non-silent vector space and the silent vector space. As a result, even if the input speech signal has an LPC cepstrum pattern different from that of the non-silent parts of the speech signal previously collected for obtaining the non-silent priority component vectors, and it is detected by the non-silent detector 68 that the frame is silence, the discriminator 72 can finally determine that the frame is non-silent, so long as the silence detector 69 detects the silence. Therefore, the silence/non-silence discrimination apparatus according to the second embodiment can prevent it from failing to detect the non-silent components.

It is evident that the second embodiment may also be modified, like the first embodiment. That is, it is possible to detects the silence/non-silence using the silent region in the non-silent priority component vector space as the reference region or using the non-silent region in the silent priority component vector space as the reference region.

In the first and the second embodiments as mentioned above, the silence/non-silence discrimination is done depending on whether or not the projected point of the cepstrum vector onto the non-silence/silence priority component vector space is within the non-silent/silent region. Since non-silent speech have many categories, there are some non-silent speeches that are not discriminated by the first and the second embodiments. For example, the non-silent characteristic parameters of a vowel are different from those of a consonant. The same thing is true between male voice and female voice. Further, the characteristic parameters differ even in the same consonant, based on the phoneme. Therefore, to improve the silence/non-silence discrimination accuracy, the priority component vectors are obtained for a plurality of categories, silence or non/silence is determined for each category, and the final determination is made on the basis of the result of the silence/non-silence determination for all the categories.

It is clear that the larger the number of the categories is, the higher the discrimination accuracy is. Use of a large number of categories as required in the speech recognition technique would increase the size of the discrimination apparatus. Therefore, it is necessary to limit the number of categories to a proper one. How to classify the parameters into categories and to limit the number of the categories will be described.

The characteristic parameters of the non-silent part of the speech signal is classified into a predetermined number of categories, and an auto-correlation matrix representative of each category is obtained. More exactly, so-called LBG algorithm is used.

A number of characteristic parameter vectors Xi (i = 1 to M) of the non-silent parts of the speech signal that are collected under the condition for using the discrimination apparatus are obtained. The auto-correlation matrix Ri of the characteristic parameter vector Xi is calculated by using the equation (4). By applying the LBG algorithm to a training vector whose element is the row vector of the matrix Ri, a predetermined number of representative vectors $\overline{Aj}$ (j = 1 to N) and a partition P($\overline{Aj}$) are obtained. The characteristic parameter vector Xi, which is used for obtaining the auto-correlation matrix Ri belonging to the partition P($\overline{Aj}$), is regarded as a member of the j-th category. Further, the matrix Rj formed of elements of the representative vectors Aj is regarded as an auto-correlation matrix representative of the j-th category. The LBG algorithm is discussed in detail by Y. Linde, A. Buzo, and R. M. Gray, "An Algorithm for Vector Quantitizer Design", IEEE Troms. COM-28, No. 1(January 1980), pp. 84 - 95.

Description that follows is how to classify the characteristic parameters into a plurality of categories and how to obtain priority component vectors for each category and a reference region for the silence/non-silence discrimination in the priority component vector space.

Firstly, a plurality of LPC cepstrum vectors Ci (i = 1, 2, ... M) previously collected under the condition for using the discrimination apparatus, are obtained. Then, an auto-correlation matrix of the LPC cepstrum vectors Ci is calculated by using the following equation. Let a P^2 dimension vector whose elements are the row vectors of the matrix Ri be a training vector Ti.

$$Ri = \begin{bmatrix} ci_1{}^2 & ci_1ci_2 & \dots & ci_1ci_p \\ ci_2ci_1 & ci_2{}^2 & \dots & ci_2ci_p \\ : & : & & \\ : & : & & \\ ci_pci_1 & ci_pci_2 & \dots & ci_p{}^2 \end{bmatrix} \qquad \dots(10)$$

$Ti = (ci_1{}^2, ci_1ci_2, \dots ci_1ci_p, ci_2ci_1, ci_2{}^2, \dots ci_2ci_p, ci_pci_1, ci_1ci_2, \dots ci_p{}^2)^t$ (11)

where ci_1, ci_2, ... ci_p are elements of the LPC cepstrum vector Ci. The training vector Ti is obtained by obtaining N number of representative vectors Yj (j = 1, 2, ... N) and a partition P(Aj) by using an LBG algorithm in the following manner.

Step 1: Initial setting

The values of the following items are initially set:

The number N of the representative vectors, threshold value ϵ of a distortion (square mean value of an error between the representative vector and each vector), initial value Ao of the representative vector, and an initial value of the training vector Ti (i = 1, 2, ... M). Let m be equal to 0, m = 0, and D1 is set with a large value.

Step 2: Calculation of Minimum Mean Distortion

Such a partition P(Am) = {Si}, i = 1, 2, ... N so as to provide a minimum mean distortion in a set Am = (Yj: j = 1, 2, ... N) of given representative vectors is obtained by using the training vector Ti. For all of the training vectors Ti belonging to the divided region Si, d(Ti, Yi) < d(Ti, Yj) (where j = 1, 2, ... N) is set up. Here, d(Ti, Yi) is a distortion provided between Ti and Yi, and it is defined as a following square error.

$$d(Ti, Yi) = \sum_{R=1}^{p2} (Ti(R) - Yi(R))^2 \qquad \dots(12)$$

where Ti(R) and yi(R) are elements of the vectors Ti and Yi.

The minimum means distortion due to the partition P(Am) is calculated in the following way.

$$\begin{aligned} D_m &= D_m[\{Am, P(Am)\}] \\ &= \frac{1}{M} \sum_{i=1}^{M} \min[d(Ti, Yi)] \qquad \dots(13) \end{aligned}$$

Step 3: Convergence Check

If $(D_{m-1} - D_m)/D_m < \epsilon$, the processing is stopped, and the Am is used as a set of the representative vectors.

Step 4: Repeat

Let the representative vector set A_{m+1} resulting from the present partition be Am, and let m be m+1.

Then, the procedure returns to step 2.

In this instance, N = 10, ϵ = 0.01, and M = 10000.

Through the procedural processing as mentioned above, ten partitions Si (i = 1 to 10) are obtained and treated as ten categories. An LPC cepstrum vector Cj forming a training vector Tj belonging to the partition Si constitutes a member of the i-th category. Further, a following matrix $\overline{Ri}$ resulting from the rearrangement of the elements of the representative vector Yi serves as a representative auto-correlation matrix of the i-th category.

$$\overline{Ri} = \begin{bmatrix} yi(1) & yi(2) & \ldots\ yi(P) \\ yi(P+1) & yi(P+2) & \ldots\ yi(2P) \\ & & \\ yi\{(P-1)P+1\} & yi\{(P-1)P+2\} & \ldots\ yi(P^2) \end{bmatrix} \quad \ldots(14)$$

In this manner, the auto-correlation matrices of the characteristic parameter vectors are classified into a predetermined number of categories and the representative auto-correlation matrix of each category is obtained. Since the LBG algorithm is used, a square average value of an error, caused when the auto-correlation matrices Ri (i = 1, 2, ... M) is represented or approximated by representative auto-correlation matrices $\overline{Rj}$ (j = 1, 2, ... N) (where N < M), is minimized.

Then, the priority component vector of each category is obtained by applying the priority component analysis to the representative auto-correlation matrix $\overline{Rj}$ of each category. For each category, discrimination regions to determine whether or not characteristic parameters belong to the category is set up in the priority component vector space by projecting the parameter vector belonging to the category onto the priority component vector space formed of the priority component vectors for the category. The silence/non-silence discrimination is performed in a manner that the characteristic parameter vector obtained for each category is projected onto the priority component vector space of each category, and the projected point is compared with the predetermined discrimination region in the vector space.

A block diagram of the third embodiment as mentioned above is shown in Fig. 11. As in the first and second embodiments, the input terminal is coupled to an LPC cepstrum calculator 82, which calculates the LPC cepstrums ci (i = 1, 2, ... L) as characteristic parameters for each frame. The calculated cepstrums are supplied to characteristic parameter projection circuits 84a to 84j, which are respectively provided for categories #1 to #10. In this embodiment, the number of categories is 10. Each circuit 84 makes a priority component analysis for each category. An example of the characteristic parameter projection circuit 84a is shown in Fig. 12. The projection circuit 84a is formed of a vector memory 92 for storing priority component vectors for the category #1 and an inner product calculator 94 for calculating the inner product of the priority component vectors of the category #1 stored in the memory 92 and the characteristic parameter vector. Thus, the projection circuits 84a to 84j respectively project the characteristic parameter vectors onto the priority component vector spaces for the categories #1 to #10, thereby to obtain projected points.

The output signals from the projection circuits 84a to 84j are respectively supplied to detection circuits 86a to 86j. Each of the detection circuit 86a to 86j determines whether the frame is non-silent or silent on the basis of the coordinates of the projected point. An example of the detection circuit 86a is shown in Fig. 13. As shown, the detection circuit 86a is formed of a region parameter memory 102 for storing the parameters defining a non-silent region of a category #1 in a priority component vector space, and a non-silence detector 104. If the priority component vector space is a three dimensional space, the non-silent region takes the form of a rectangular parallelepiped as shown in Fig. 7. The parameters define the upper and the lower limits along the respective coordinate axes. The non-silence detector 104 produces a detection signal of "1" level when the projected point is within the non-silent region, and produces a detection signal of "0" level in other cases. The output signals from the detection circuits 86a to 86j are supplied to a silence/non-silence discriminator 88. When at least one of the detection circuits 86a to 86j produces a detection signal of "1" level, the discriminator 88 decides that the frame is non-silent. A sequence of procedural steps to make the decision is shown in Fig. 14.

The priority component vector of each category is obtained by applying the priority component analysis to the matrix $\overline{Ri}$. The parameters defining the region of each category in a priority component vector space are previously obtained in a manner that for each category, an LPC cepstrum vector belonging to that category is projected onto the priority component vector space of the category.

As seen from the foregoing description, the third embodiment makes a silence/non-silence discrimination in a manner that the characteristic parameters are projected onto the priority component vector space of each category and the final discrimination is made based on the judgments for all the categories. Therefore, the third embodiment has an advantages of an improved accuracy of the silence/non-silence discrimination, in addition to the advantages of the first and the second embodiments. Further, in the third embodiment, the LBG algorithm is used for classifying the characteristic parameter vectors into categories and for obtaining the priority component vectors for each category. Accordingly, the auto-correlation matrix of the M number of characteristic parameters are classified into a smaller number of categories, leading to the improvement of the silence/non-silence discrimination accuracy. Incidentally, the modifications of the first and the second embodiments are allowed also in the third embodiment.

In the above-mentioned embodiments, when the projected point of the characteristic parameters are outside the decision reference region, the inversion of a decision made when the projected point is inside the space is instantly made. In such a case, if a decision based on another method is applied again, the decision accuracy will be further improved. This approach is realized by a silence/non-silence discrimination apparatus shown in Fig. 15, which is a fourth embodiment of the present invention. As in the previous embodiments, a speech signal is input to an LPC cepstrum calculator 122. The calculator 122 calculates the LPC cepstrums ci (i = 1, 2, ..., L) for each frame. The cepstrums are supplied to an inner product calculator 124, which is coupled with a non-silent priority component vector memory 126. The inner product calculator 124 and the non-silent priority component vector memory 126 form a characteristic parameter projection circuit 128. In this embodiment, the characteristic parameters are projected onto a non-silent priority component vector space. The inner product calculator 124 calculates the coordinates of a projected point Q. The non-silent priority component vector memory 126 stores the vectors of first to third priority components that result from a priority component analysis of the LPC cepstrums of a non-silent part of the speech signal as previously collected under the condition for using the discrimination apparatus.

The output signal of the inner product calculator 124 is supplied to a detection circuit 130. The detection circuit 130 is coupled with a non-silent region parameter memory 132 and a silent region parameter memory 134. The memory 132 stores parameters defining a non-silent region in the non-silent priority component vector space. The memory 134 stores parameters defining a silent region in the non-silent priority component vector space. When the projected point Q is inside the non-silent region, the detection circuit 130 decides that the frame is non-silent. When the projected point Q is inside the silent region, the detection circuit 130 decides that the frame is silent. When the projected point Q is outside both the silent region and the non-silent region, the detection circuit 130 decides that the property of the frame is indefinite.

The output signal of the detection circuit 130 is supplied to a silence/non-silence discriminator 136. The output signal of the discriminator 136 is output as a final decision result and is stored in a discrimination result memory 140. The memory 140 stores the discrimination results for at least three past frames. The data derived from the memory 140 is supplied to a conditional probability table 138. The table 138 stores a probability of silence or non-silence as predicted on the basis of the discrimination results for the three past frames, viz., a conditional probability. Assuming that the discrimination result for the present frame is D_i, and the discrimination results for the three past frames are D_{i-1}, D_{i-2}, and D_{i-3}, the conditional probability P is given as follows.

$$P(D_i \mid D_{i-1}, D_{i-2}, D_{i-3}) = \frac{P(D_i, D_{i-1}, D_{i-2}, D_{i-3})}{P(D_{i-1}, D_{i-2}, D_{i-3})} \qquad \ldots(15)$$

where Di is 1 when the i-th frame is non-silent, and is 0 when it is silent. $P(D_i, D_{i-1}, D_{i-2}, D_{i-3})$ and $P(D_{i-1}, D_{i-2}, D_{i-3})$ are previously obtained by probability calculations on the basis of the four consecutive frames and the three consecutive frames. Foe example, P(0, 0, 0) represents the probability in which the past three frames are silent. Those frames are formed in a manner that each frame of the speech signal collected under the condition for using the discrimination apparatus is labeled with silence and non-silence while observing its waveform and spectrum.

A conditional probability that the present frame will be non-silent and a conditional probability that the present frame will be silent are input to the silence/non-silence discriminator 136. When the decision result from the detection circuit 130 denoting that the property of the frame is indefinite, the discriminator 136 compares both the conditional probabilities, and employs the higher of the two. The procedural flow till the final decision is reached is shown in Fig. 16.

According to the fourth embodiment, when the decision using the decision region in the priority component vector space fails to provide a positive decision result, another decision is made on the basis of the conditional probability which is led from the learning data, although each previous embodiment unconditionally provides the negative decision result. That is, the fourth embodiment employs two steps of decisions for the silence/non-silence discrimination, thereby improving the discrimination accuracy. Use of the conditional probability implies use of the knowledge on the speech signal that a transition of the property of the frames, non-silence → silence → non-silence → silence, is a rare case. Therefore, a probability of occurrence of mistaken decisions on the small power consonants of voiced speech or non-voiced speech is reduced. Further, omission of the beginnings and the ends of words and addition of noise infrequently occur. As a matter of course, the modifications of the previous embodiments are allowed also in the fourth embodiment.

While in the previous embodiments, the discrimination of silence/non-silence is based on the position of the projected point of the characteristic parameter, it may be done on the basis of a time variation of the projected point. This approach further improves the discrimination accuracy, and is realized as shown in Fig. 17, which shows a fifth embodiment of a silence/non-silence discrimination apparatus according to the present invention. As in the previous embodiments, an LPC cepstrum calculator 152 is connected to the input terminal. The calculator 152 calculates the LPC cepstrums ci (i = 1, 2, ... L) for each frame. The cepstrums are supplied to an inner product calculator 154, which is coupled with a non-silent priority component vector memory 156. The inner product calculator 154 and the non-silent priority component vector memory 156 form a characteristic parameter projection circuit 158. The non-silent priority component vector memory 156 stores the vectors of first to third priority components that result from a priority component analysis of the LPC cepstrums of the non-silent part of the speech signal as previously collected under the condition for using the discrimination apparatus. The inner product calculator 154 calculates the coordinates of a projected point Q in a three-dimensional space whose axes respectively consist of first to third priority component vectors of the non-silent cepstrums.

The output signal of the inner product calculator 154 is supplied to a finite impulse response (FIR) digital filter 160. Examples of the filter 160 are shown in Figs. 18 and 19. The FIR filter of order of p is shown in Fig. 18 and the FIR filter of order of two is shown in Fig. 19. The filter 160 is for obtaining a change vector $\Delta Q(n) = (\Delta q_1(n), \Delta q_2(n), \Delta q_3(n))$ by filtering a projecting point vector Q(n) in a priority vector space in the n-th frame. In the filter of Fig. 18, $\Delta q_i(n)$ is given as follows.

$$\Delta q_i(n) = \frac{1}{\sigma_i} \sum_{j=1}^{p} a_j q_j(n-j) \qquad \ldots(16)$$

where a_j (j = 1 to p) is a filter coefficient, p is an order of the filter, and σ is a coefficient to normalize a variance of the filter outputs, and is expressed as a standard deviation as follows.

$$\sigma_i = \sqrt{\frac{1}{N} \sum_{j=0}^{N-1} q_i{}^2(n-j)} \qquad \ldots(17)$$

In the filter of Fig. 19, $\Delta q_i(n)$ is given as follows.

$\Delta q_i(n) = \frac{1}{\sigma_i}$ $(x(n) + b_1 x(n-1) + b_2 x(n-2))$ (18)

$x(n) = q_i(n) + a_1 x(n-1) + a_2 x(n-2)$ (19)

The transfer functions H1(z) and H2(z) of the filters of Figs. 18 and 19 are expressed in the following way.

$$H1(z) = \frac{1}{\sigma_i} \sum_{j=0}^{P} a_j z^{-1} \qquad \ldots(20)$$

$$H2(z) = \frac{1}{\sigma_i} \times \frac{1+b_1 z^{-1}+b_2 z^{-2}}{1-a_1 z^{-1}-a_2 z^{-2}} \qquad \ldots(21)$$

The filter coefficients a_j and b_j are previously selected so that the transfer functions H1(z) and H2(z) specify the high-pass filtering. If required, those may be varied in accordance with a signal power.

The output signal of the filter 160 is supplied to matching circuits 162a to 162j. Each matching circuit 162 calculates a similarity expressed in terms of the Euclidean distance, and mathematically expressed by the following equation. The details of the matching circuit 162a is typically illustrated in Fig. 20.

$$Sm = \sum_{i=1}^{3} (\Delta q_i - r_i^{(m)})^2 \qquad \ldots(22)$$

where $R^{(m)} = (r_1^{(m)}, r_2^{(m)}, r_3^{(m)})$ indicates the m-th reference pattern, and is stored in a table 182 in the m-th matching table 182. It is evident that another known similarity may be used.

The reference pattern is previously obtained in accordance with a sequence of steps shown in Fig. 21. In step #41, the speech data in a part of a speech which is generated under the condition for using the discrimination apparatus and is considered as being non-silent is collected as learning data. In step #42, an LPC cepstrum of the learning data is obtained every frame. In step #43, the cepstrum is projected onto a non-silent priority component vector space. In step #44, a plurality of change vectors denoting the change of the projected point with respect to time are extracted. In step #45, the center-of-gravity of those change vectors is calculated, and is used as a reference pattern. The reason why a number of matching circuits 162 are used is that similarities between the reference patterns and the change vector of the projected point are calculated, thereby to improve the discrimination accuracy.

The output signals of the matching circuits 162a to 162j are input to a silence/non-silence discriminator 164. The discriminator 164 compares a preset threshold value with a similarity which is the smallest of ten input similarities. When the minimum similarity is above the threshold value, the discriminator 164 decides that the frame is non-silent. In other cases, the same decides that the frame is silent. A sequence of the procedural steps to make the decisions is illustrated in Fig. 22.

As described above, according to the fifth embodiment, the silence/non-silence determination is made on the basis of the time variation of the projected points. This eliminates the mistaken decision due to the background noise. The position of the projected point of the characteristic parameters is possibly moved due to the noise. However, the change of the projected point with respect to time is relative, not absolute, and hence is insensible to noise. The modifications of the previous embodiments are allowed also in the fifth embodiment.

Fig. 23 shows a block diagram of a silence/non-silence discriminator according to a sixth embodiment of the present invention. An output signal of an LPC cepstrum calculator 202 is supplied to a characteristic parameter projection circuit 208 formed of an inner product calculator 204 and a non-silence priority component vector memory 206. The non-silent priority component vector memory 206 stores the vectors of first to third priority components that result from a priority component analysis of the LPC cepstrums of the non-silent parts of the speech signal as previously collected under the condition for using the discrimination apparatus. The inner product calculator 204 calculates the coordinates of a projected point Q in a three dimensional space whose axes are constituted by first to third priority component vectors of the non-silent LPC cepstrums.

The output signal of the inner product calculator 204 is supplied to an FIR filter 210 and a plurality of detection circuits 212a to 212j. Each detection circuit 212 decides whether the projected point is within or outside the non-silent region for each category. If it is within the region, the detection circuit 212 produces a detection signal of "1" level. In other cases, it produces a detection signal of "0" level. The details of the detection circuit 212a is typically shown in Fig. 24. As shown, the circuit 212a is formed of a non-silent region parameter memory 224 for each category, and a non-silence detector 226. The output signals of the detection circuits 212a to 212j are supplied to a temporary detector 214. When any of the output signals of the detection circuits 212a to 212j has "1" level, the temporary detector 214 temporarily decides that the frame is non-silent. In other cases, it temporarily decides that the frame is silent.

The output signal of the temporarily detector 214 is supplied to an inequality detector 216 which in turn decides if the decision result on the previous frame is equal to the decision result on the present frame. When both the decisions are unequal, the detector 216 produces a detection signal FF of "1" level. When equal, it produces a detection signal FF of "0" level.

The output signal of the FIR filter 210 is supplied to a change detector 218. The detector 218 calculates a change quantity Δ given by the following equation by using a change vector $\Delta Q = (\Delta 1q_1, \Delta q_2, \Delta q_3)$ of the projected point Q for each frame, that is derived from the filter 216. The detector 218 outputs a detection signal CF of "1" level when the change quantity Δ is larger than a predetermined threshold value and a detection signal CF of "0" level when the change quantity Δ is not larger than the threshold value.

$\Delta = W1\Delta q_1^2 + W2\Delta q_2^2 + W3\Delta q_3^2$ (23)

where W1, W2, and W3 are linear weighting coefficients and for which eigen values Wi of an auto-correlation matrix of characteristic parameters in the non-silent part are used.

The output signal of the temporary detector 214, the output signal FF of the inequality detector 216, and the output signal CF of the change detector 218 are supplied to a silence/non-silence discriminator 220. The discriminator 220 first compares the output signal FF of the detector 216 with the output signal CF of the detector 218. When both the output signals are equal, the output signal of the temporary detector 214 is output as a final decision result. When those are unequal, the decision is made in the following way.

If FF = "1" and CF = "0", the temporary decision result of the previous frame is output as the final decision result and the temporary decision result of the present frame is altered to make it equal to that of the previous frame.

If FF = "0" and CF = "1", the temporary decision result of the present frame is inverted, and the inverted one is output as the final decision result. A flow of above decision procedure is shown in Fig. 25.

As seen from the foregoing description, in the sixth embodiment, the position of the projected point of the characteristic parameters in the priority component vector space as well as the change of the projected point are used for the silence/non-silence discrimination. The result is decreased occurrence of mistaken decisions due to noise and improved decision accuracy. The modifications of the previous embodiments are allowed also in the sixth embodiment.

As described above, in the silence/non-silence discrimination apparatus according to the present invention, a number of characteristic parameters of the speech signal are calculated. The parameters are projected onto a given priority component vector space whose dimension number is smaller than the number of the calculated characteristic parameters. Discrimination is made as to if the speech signal, more exactly the frame of it, is silent or non-silent, on the basis of the position of the projected point. Therefore, the number of characteristic parameters used for the discrimination may be reduced by utilizing the statistical nature of the parameters, while keeping the satisfactory accuracy of the discrimination. Further, there is no need for optimizing the number of the parameters and their categorization.

Additional advantages and modifications will readily occur to those skilled in the art. Therefore, the invention in its broader aspects is not limited to the specific details, representative apparatuses, and illustrated embodiments shown and described. Accordingly, departures may be made from such details without departing from the sprit or scope of the general inventive concept as defined by the appended claims and their equivalents.

Claims

1. A silence/non-silence discrimination apparatus comprising:
means for obtaining characteristic parameter from an input speech, characterized by further comprising:
means (54, 67, 84, 128, 158, 208) for projecting said characteristic parameters onto a vector space formed of first to i-th ("i" being a positive integer smaller than the number of said characteristic parameters) priority component vectors of the characteristic parameters of a given type of speech, to obtain a projected point; and
means (56, 72, 88, 136, 164, 220)for discriminating whether the input speech is silent or non-silent based on the position of said projected point.

2. The apparatus according to claim 1, characterized in that
said projecting means comprises means (53) for projecting said characteristic parameters onto the vector space formed of the first to i-th priority component vectors of the characteristic parameters of a non-silent speech, and
said discriminating means comprises means (56) for discriminating that the input speech signal is silent or

not depending on whether or not said projected point is within a non-silent region in said vector space.

3. The apparatus according to claim 2, characterized in that
said projecting means comprises means (53) for calculating inner products of the first to i-th priority component vectors of said non-silent speech and a vector formed of said characteristic parameters, and
said discriminating means comprises means (56) for comparing the inner product with the upper and the lower limits of said non-silent region.

4. The apparatus according to claim 1,
characterized in that
said projecting means comprises first projection means (63) for projecting said characteristic parameters onto the vector space formed of the first to i-th priority component vectors of the characteristic parameters of a non-silent speech to obtain a first projected point, and second projection means (64) for projecting said characteristic parameters onto the vector space formed of first to j-th ("j" being a positive integer smaller than the number of said characteristic parameters obtained) priority component vectors of the characteristic parameters of a silent speech to obtain a second projected point, and
said discriminating means comprises first detection means (68) for detecting whether or not said first projected point is within a predetermined non-silent region in the vector space of the non-silent speech, second detection means (69) for detecting whether or not said second projected point is within a predetermined silent region in the vector space of the silent speech, and third detection means (72), when said first detection means detects said first projected point is within the non-silent region, for detecting that said input speech is non-silent, and when said first detection means detects said first projected point is not within the non-silent region and said second detection means detects said second projected point is within the silent region, for detecting that said input speech is silent, and when said first detection means detects said first projected point is not within the non-silent region and said second detection means detects said second projected point is not within the silent region, for detecting that said input speech is non-silent.

5. The apparatus according to claim 4, characterized in that
said first projection means comprises means (63) for calculating first inner products of said first to i-th priority component vectors of said non-silent speech and a vector formed of said characteristic parameters,
said second projection means comprises means (64) for calculating second inner products of said first to j-th priority component vectors of said silent speech and a vector formed of said characteristic parameters,
said first detection means comprises means (68) for comparing the first inner product with the upper and the lower limits of the non-silent region, and
said second detection means comprises means (69) for comparing the second inner product with the upper and the lower limits of the silent region.

6. The apparatus according to claim 1, characterized in that
said projecting means comprises means (84a to 84j) for projecting said characteristic parameters onto the vector space formed of the first to i-th priority component vectors of the non-silent speech for every categories, and
said discriminating means comprises means (86a to 86j) for detecting whether or not the projected point is within a predetermined non-silent region in said vector space for each category, and means (88), when at least one projected point is within the non-silent region, for detecting that said input speech is non-silent, and when no projected point is within the non-silent region, for detecting that said input speech is silent.

7. The apparatus according to claim 6, characterized in that
said projecting means comprises means (94) for calculating inner products of the first to i-th priority component vectors of the non-silent speeches for every categories and a vector formed of said characteristic parameters, and
said discriminating means comprises means (104) for comparing the inner products for every categories with the upper and the lower limits of the non-silent region.

8. The apparatus according to claim 1, characterized in that
said projecting means comprises means (128) for projecting said characteristic-parameters onto the vector space formed of the first to i-th priority component vector of the characteristic parameters of a nonsilent speech, and
said discriminating means comprises first detection means (130) for detecting whether or not the projected point is within a predetermined non-silent region, second detection means (130) for detecting whether or not the projected point is within a predetermined silent region, and third detection means (136), when said first and second detection means detect that the projected point is within the non-silent region and within the silent region, for detecting that said input speech is non-silent and silent, respectively, when said first and second detection means detect that the projected point is not within the non-silent region and within the silent region, respectively, for calculating a first conditional probability that the input speech is non-silent

and a second conditional probability that the input speech is silent on the basis of the past discrimination result, and discriminating whether the input speech is silent or non-silent based on one of the first and second conditional probabilities which is larger than the other.

9. The apparatus according to claim 8, characterized in that
said projecting means comprises means (124) for calculating inner products of the first to i-th priority component vectors of said non-silent speech and a vector formed of said characteristic parameters,
said first detection means comprises means (130, 132) for comparing the inner product with the upper and the lower limits of said non-silent region, and
said second detection means comprises means (130, 134) for comparing the inner product with the upper and the lower limits of said silent region.

10. The apparatus according to claim 1, characterized in that
said discriminating means comprises means (160) for detecting a change of said projected point with respect to time, and means (162, 164) for detecting whether the input speech is silent or non-silent on the basis of the change of the projected point.

11. The apparatus according to claim 10, characterized in that
said detecting means comprises a high-pass filter (160) coupled for reception with a signal representative of the position of the projected point.

12. The apparatus according to claim 10, characterized in that
said projecting means comprises means (158) for projecting said characteristic parameters onto a vector space formed of first to i-th priority component vectors of a non-silent speech, and
said detecting means (162) comprises means (180) for storing the center-of-gravity of vectors representing the change of the projected point of the characteristic parameters of various types of speeches as reference patterns, means (182) for calculating similarities of said vectors representing the changes of the projected point with each of said reference patterns, and means (164) for detecting if the input speech is silent or non-silent by comparing a minimum value of the said similarities with a given threshold value.

13. The apparatus according to claim 10, characterized in that
said discriminating means comprises means (218) for detecting a change of said projected point with respect to time, and means (220) for detecting whether the input speech is silent or non-silent based on the position of the projected point and the change of said projected point with respect to time.

14. The apparatus according to claim 13, characterized in that
said projecting means comprises means (208) for projecting said characteristic parameters onto a vector space formed of the first to i-th priority component vectors of a non-silent speech, and
said discriminating means comprises first detection means (212) for detecting whether or not the projected point is within a predetermined non-silent region, second detection means (218) for detecting whether or not the detection result of said first detection means is equal to the previous detection result, third detection means (216) for detecting whether or not a quantity of said change of the projected point is above a predetermined value, and fourth detection means (220), when the detection result of said first detection means is unequal to the previous detection result and a change quantity of the projected point is greater than the predetermined value and when the detection result of said first detection means is equal to the previous detection result and a change quantity of said projected point is less than the predetermined value, for making a detection on the basis of the detection result of said first detection means, when the detection result of said first detection means is equal to the previous detection result and a change quantity of said projected point is greater than the predetermined value, for making a detection on the basis of said previous detection result of said first detection means and for replacing the present detection result with said previous result, and when the detection result of said first detection means is unequal to the previous detection result and a change quantity of said projected point is less than the predetermined value, for making a detection on the basis of the inverse of said previous detection result of said first detection means.

15. A silence/non-silence discrimination method characterized by comprising the steps of:
making a priority component analysis of a non-silent speech and/or silent speech, to obtain a predetermined number of priority component vectors of the non-silent speech and/or silent speech;
obtaining region parameters defining a non-silent region and/or a silent region in a vector space formed of said predetermined number of priority component vectors of the non-silent speech and/or silent speech;
obtaining characteristic parameters, larger than said priority component vectors in number, from an input speech;
projecting a vector formed of said characteristic parameters onto said vector space; and
detecting that said speech is silent or non-silent depending on whether the projected point of the vector formed of said characteristic parameters is within or outside said non-silent region and/or said silent region.

16. A silence/non-silence discrimination method characterized by comprising the steps of:
making a priority component analysis of a non-silent speech and/or silent speech, to obtain a predetermined number of priority component vectors of the non-silent speech and/or the silent speech;
obtaining region parameters defining a non-silent region and/or a silent region in a vector space formed of said priority component vectors;
obtaining characteristic parameters, larger than said priority component vectors in number, from an input speech;
projecting a vector formed of said characteristic parameters onto said vector space formed of said priority component vectors;
obtaining a change of said projected point with respect to time;
obtaining similarities between said change of said projected point and changes of the projected points of the characteristic parameters of various non-silent speeches and/or silent speeches with respect to time; and
detecting that said iput speech is silent or non-silent on the basis of the similalities obtained.

17. A speech cell generation apparatus characterized by comprising:
first encoding means for encoding an speech;
second encoding means for encoding the speech at a lower coding rate or bit rate than that of said first encoding means;
means for detecting whether or not said speech is silent or non-silent depending on the position of a projected point of characteristic parameters of said speech signal onto a vector space formed of priority component vectors of a given type of speech; and
means for converting the output signal of said first encoding means into cells when the speech is detected as being non-silent, and converting the output signal of said second encoding means into cells when the speech is detected as being silent.

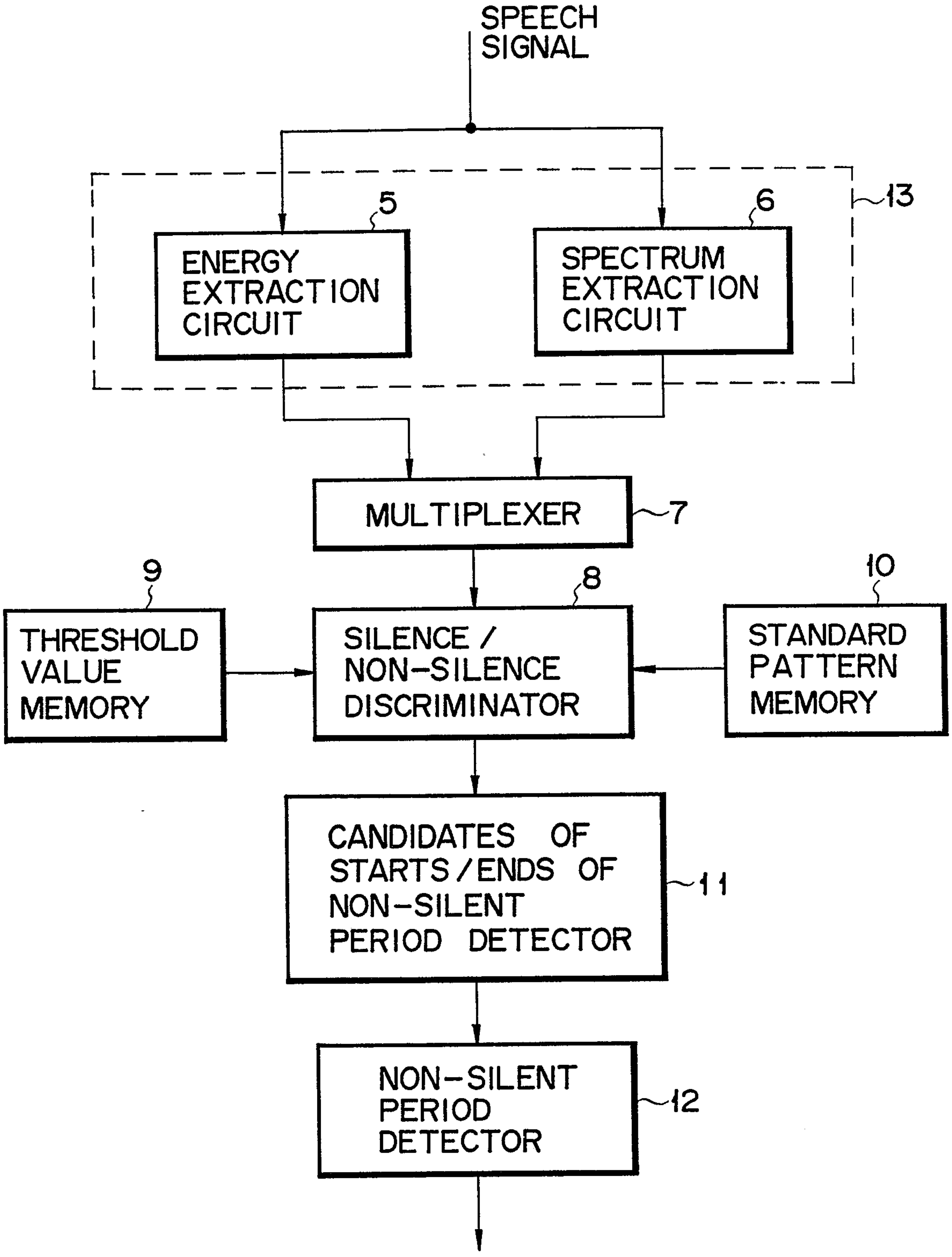

SPEECH SIGNAL
13
5
ENERGY EXTRACTION CIRCUIT
6
SPECTRUM EXTRACTION CIRCUIT
MULTIPLEXER
7
9
THRESHOLD VALUE MEMORY
8
SILENCE / NON-SILENCE DISCRIMINATOR
10
STANDARD PATTERN MEMORY
CANDIDATES OF STARTS / ENDS OF NON-SILENT PERIOD DETECTOR
11
NON-SILENT PERIOD DETECTOR
12

FIG. 1

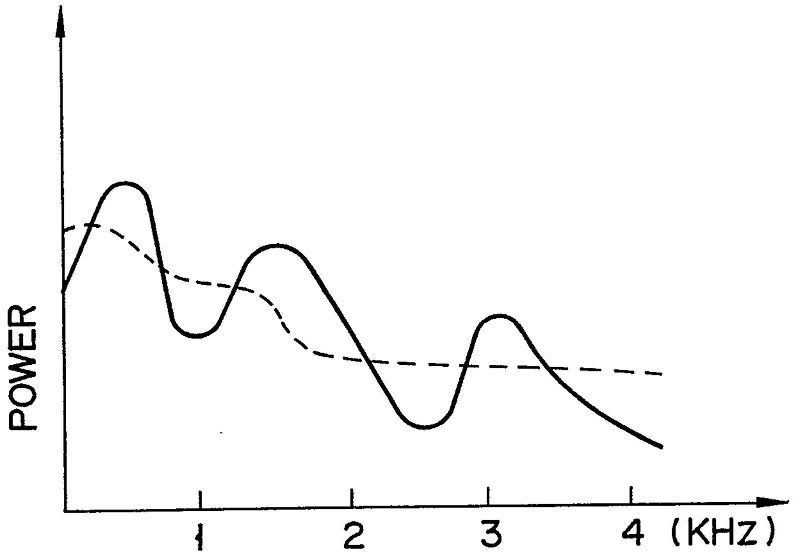

POWER
1
2
3
4 (KHZ)

FIG. 2

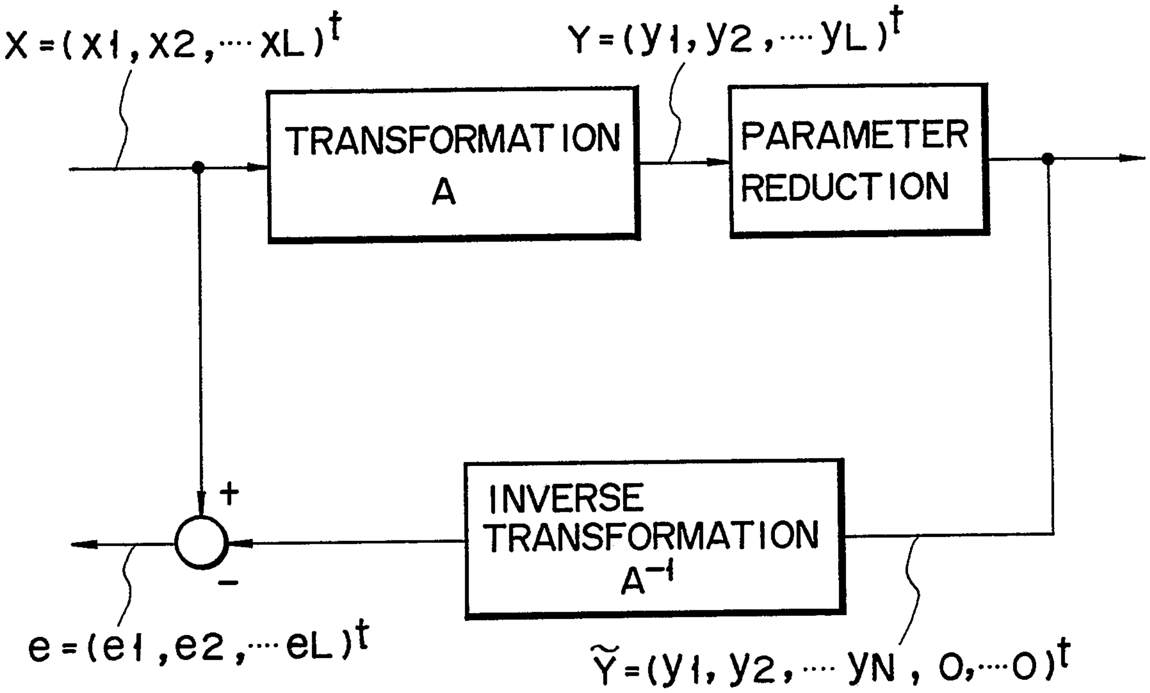
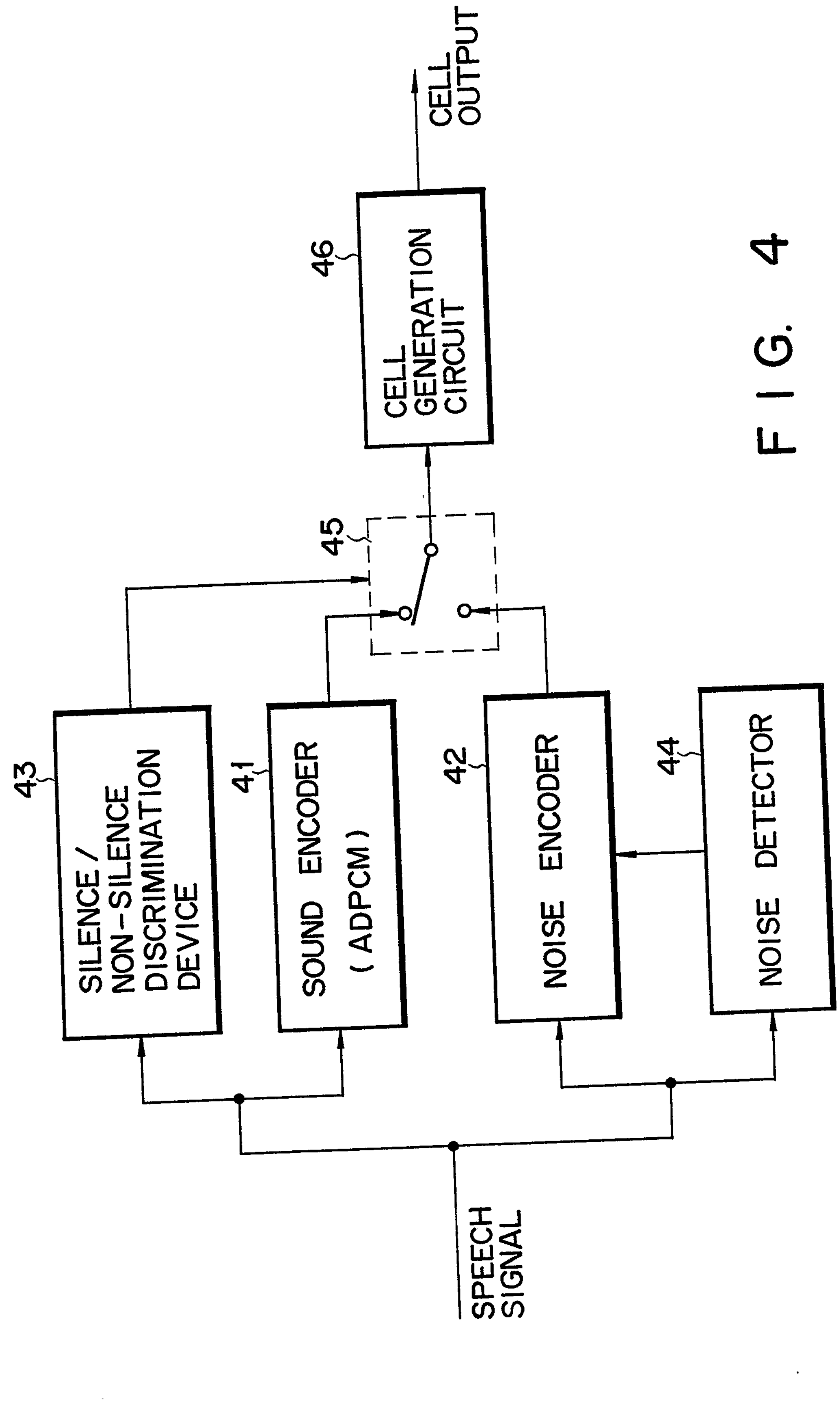
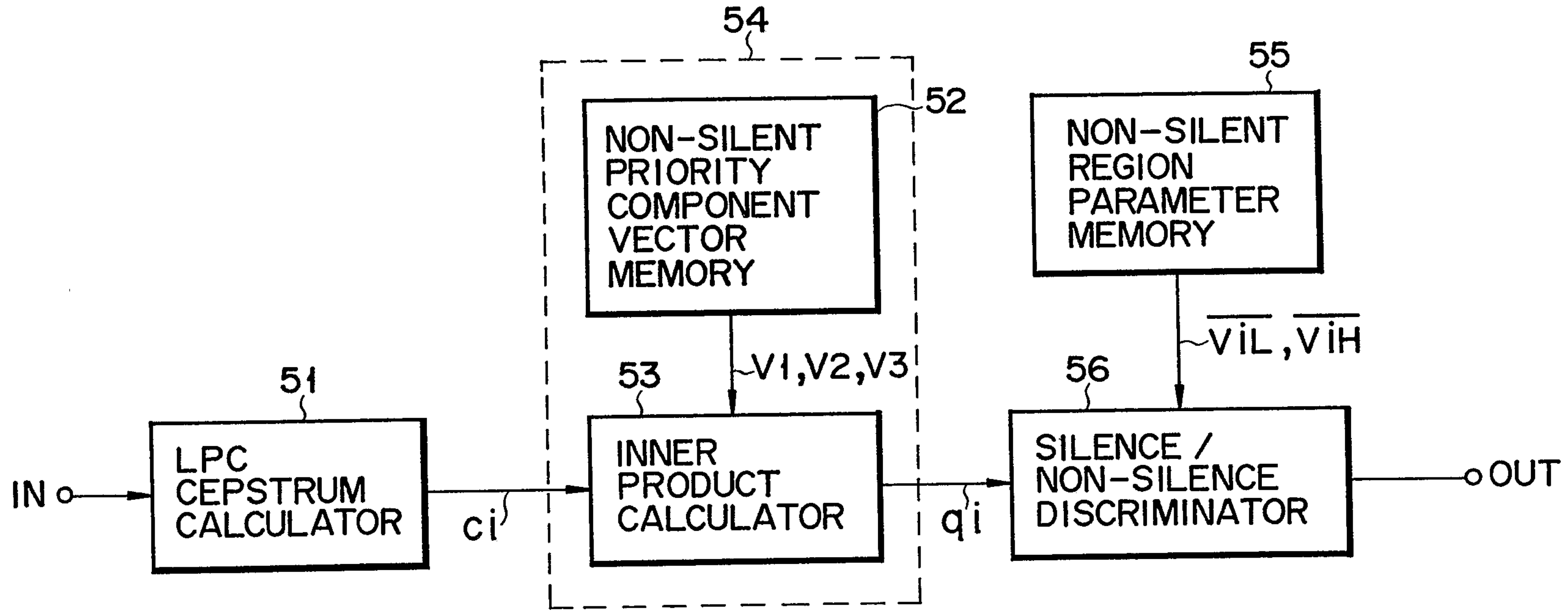

X=(X1, X2, ···· XL)^t
Y=(y1, y2, ···· yL)^t
TRANSFORMATION A
PARAMETER REDUCTION
INVERSE TRANSFORMATION A^-1
+
−
e=(e1, e2, ···· eL)^t
Ỹ=(y1, y2, ···· yN, 0,····0)^t

FIG. 3


SPEECH SIGNAL
43
SILENCE / NON-SILENCE DISCRIMINATION DEVICE
41
SOUND ENCODER (ADPCM)
42
NOISE ENCODER
44
NOISE DETECTOR
45
46
CELL GENERATION CIRCUIT
CELL OUTPUT

F I G. 4


IN
51
LPC CEPSTRUM CALCULATOR
ci
54
52
NON-SILENT PRIORITY COMPONENT VECTOR MEMORY
V1,V2,V3
53
INNER PRODUCT CALCULATOR
qi
55
NON-SILENT REGION PARAMETER MEMORY
$\overline{ViL}$, $\overline{ViH}$
56
SILENCE / NON-SILENCE DISCRIMINATOR
OUT

F I G. 5

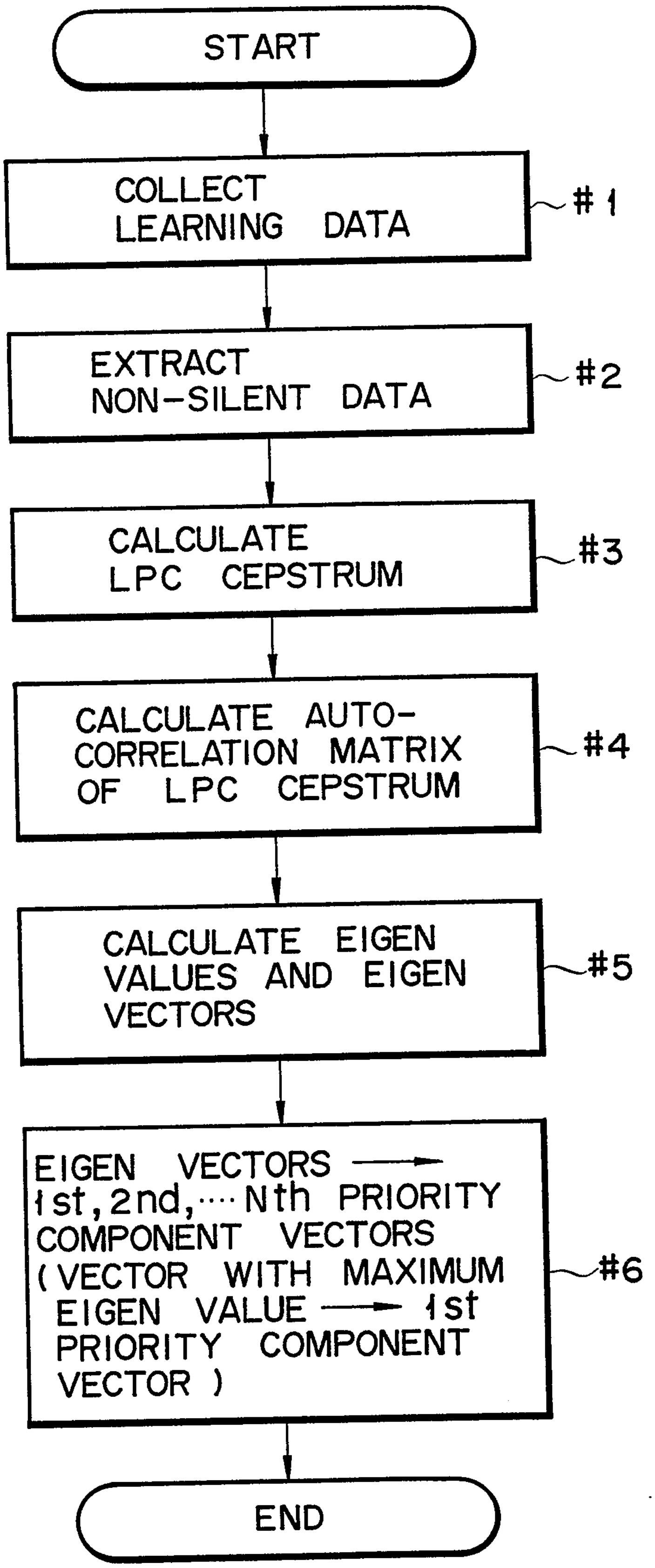
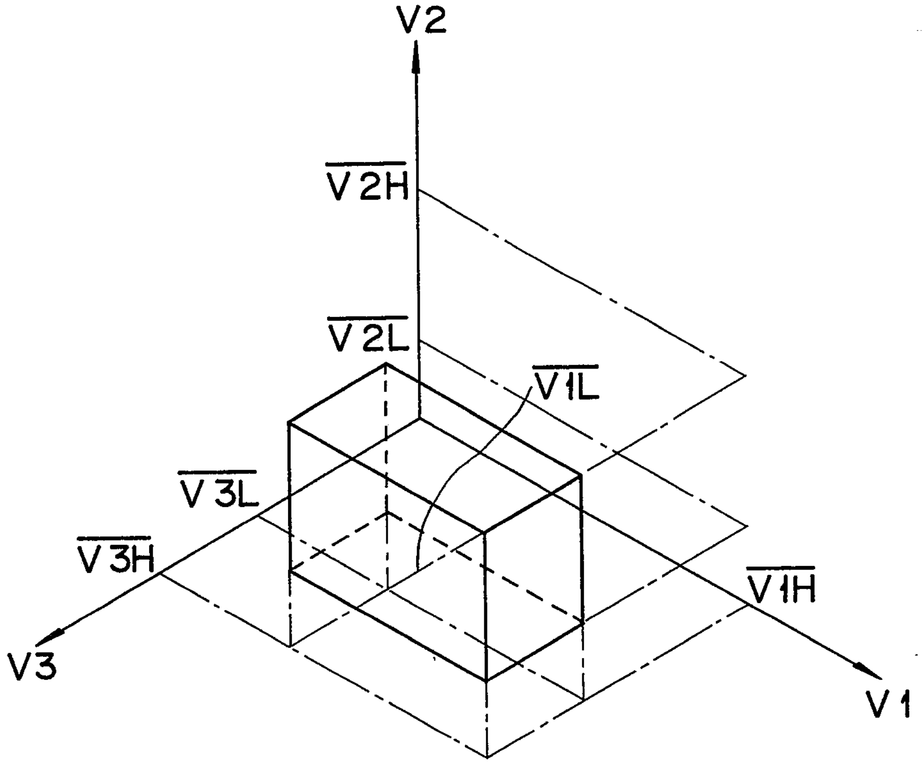
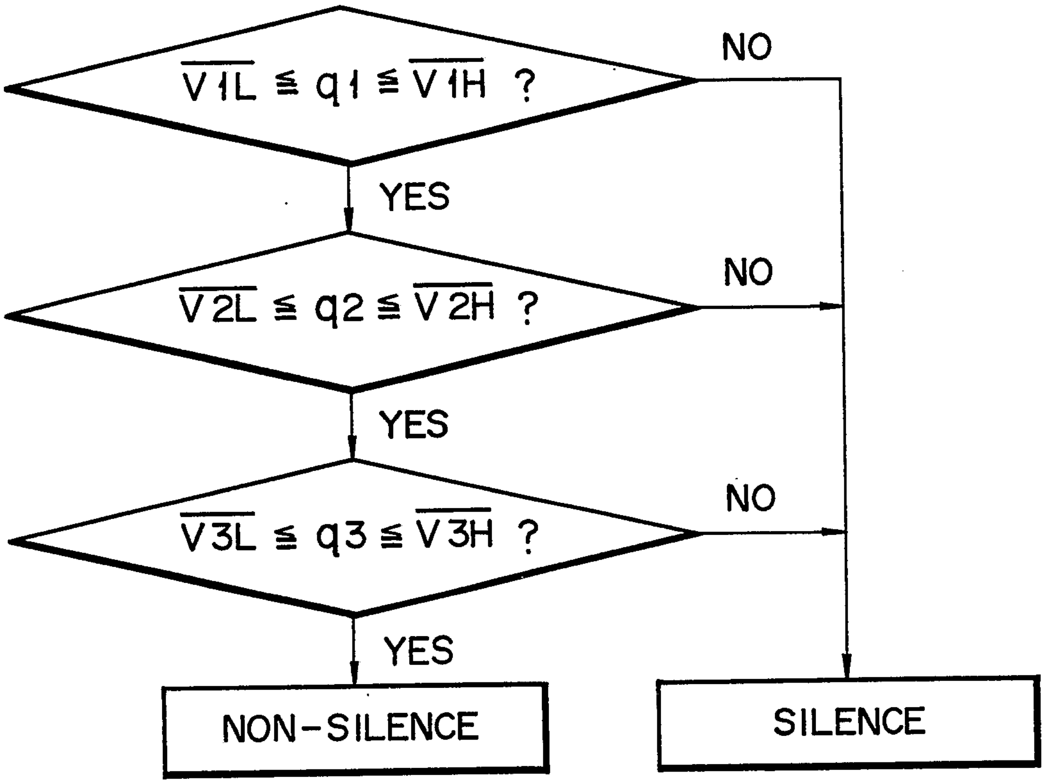

START
COLLECT LEARNING DATA
#1
EXTRACT NON-SILENT DATA
#2
CALCULATE LPC CEPSTRUM
#3
CALCULATE AUTO-CORRELATION MATRIX OF LPC CEPSTRUM
#4
CALCULATE EIGEN VALUES AND EIGEN VECTORS
#5
EIGEN VECTORS → 1st, 2nd, ···· Nth PRIORITY COMPONENT VECTORS (VECTOR WITH MAXIMUM EIGEN VALUE → 1st PRIORITY COMPONENT VECTOR)
#6
END

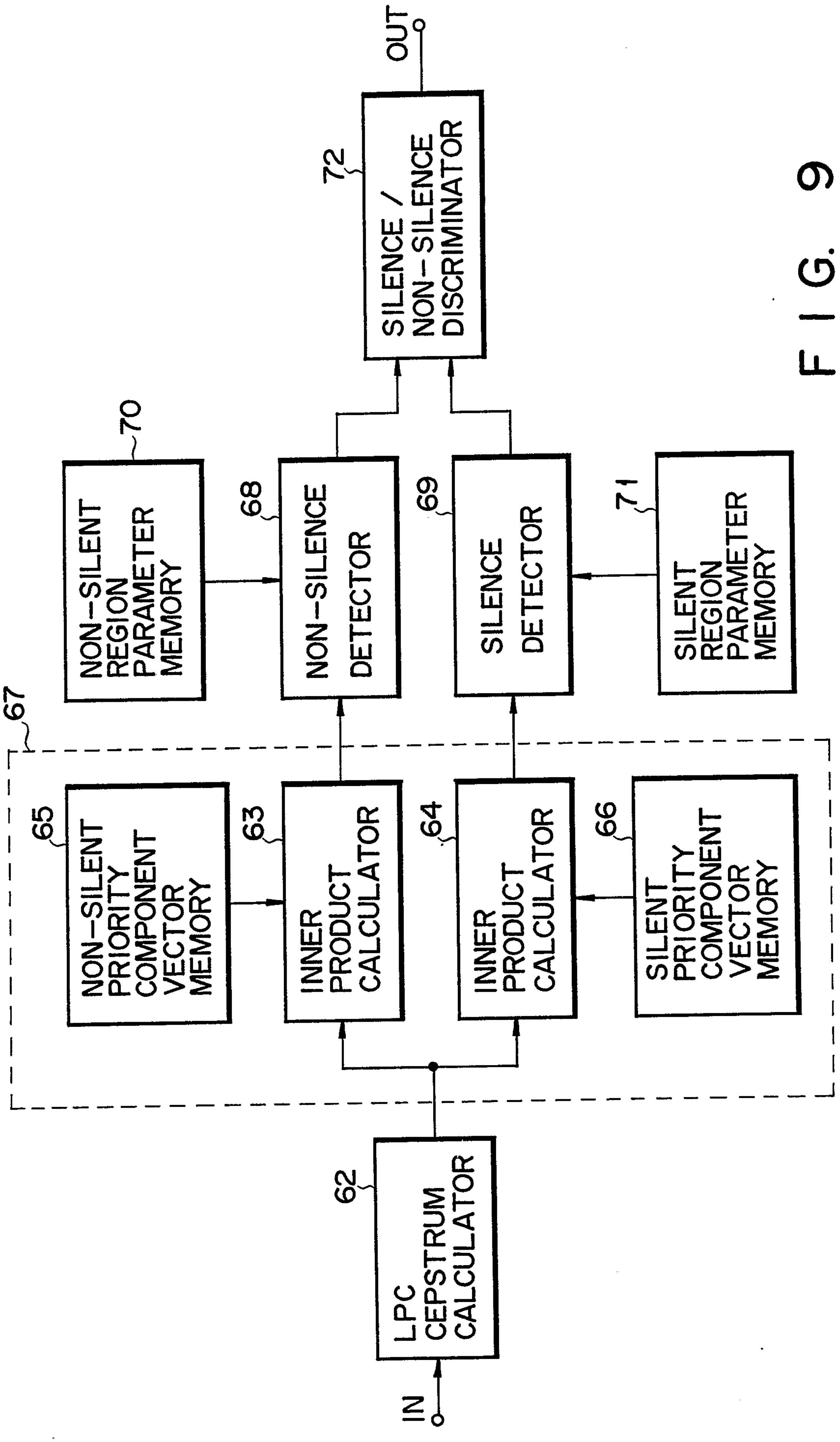
FIG. 6


V2
$\overline{V2H}$
$\overline{V2L}$
$\overline{V1L}$
$\overline{V3L}$
$\overline{V3H}$
$\overline{V1H}$
V3
V1

F I G. 7


$\overline{V1L} \leqq q1 \leqq \overline{V1H}$?
NO
YES
$\overline{V2L} \leqq q2 \leqq \overline{V2H}$?
NO
YES
$\overline{V3L} \leqq q3 \leqq \overline{V3H}$?
NO
YES
NON-SILENCE
SILENCE

F I G. 8


IN
62
LPC CEPSTRUM CALCULATOR
67
65
NON-SILENT PRIORITY COMPONENT VECTOR MEMORY
63
INNER PRODUCT CALCULATOR
64
INNER PRODUCT CALCULATOR
66
SILENT PRIORITY COMPONENT VECTOR MEMORY
70
NON-SILENT REGION PARAMETER MEMORY
68
NON-SILENCE DETECTOR
69
SILENCE DETECTOR
71
SILENT REGION PARAMETER MEMORY
72
SILENCE / NON-SILENCE DISCRIMINATOR
OUT

F I G. 9

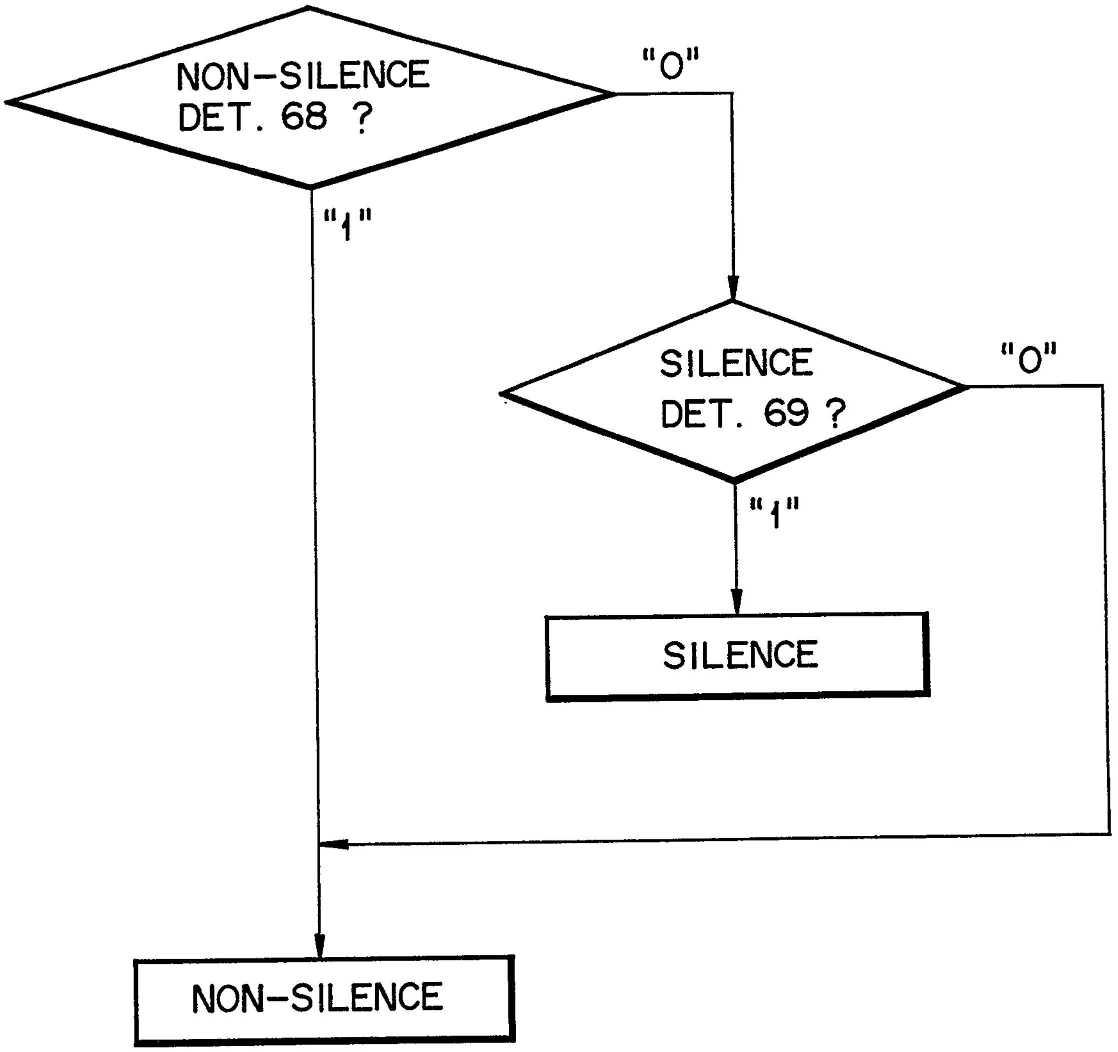
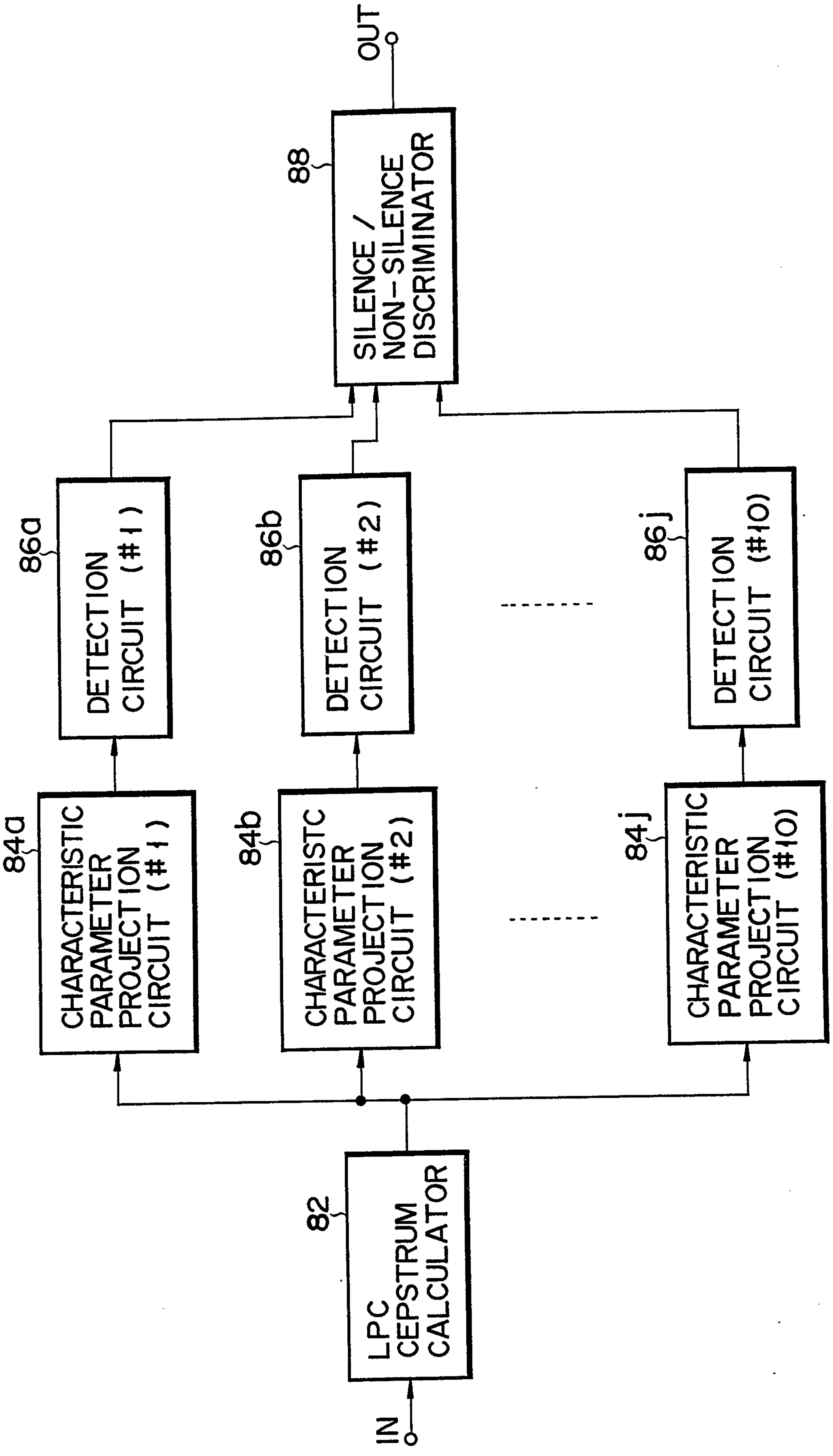
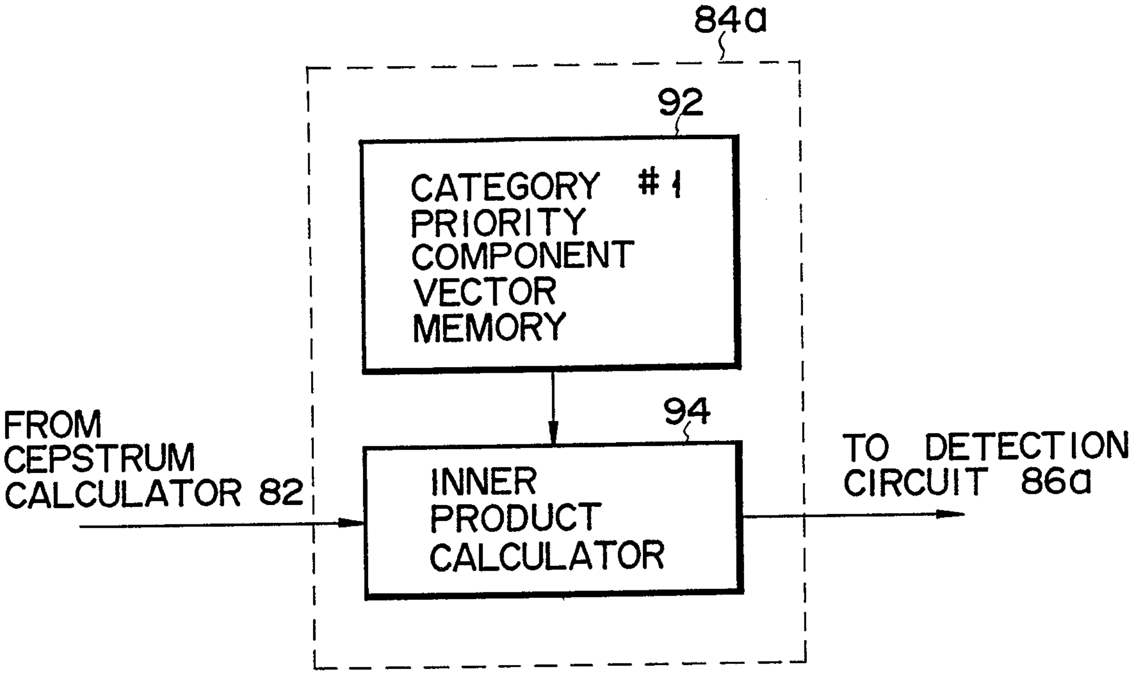


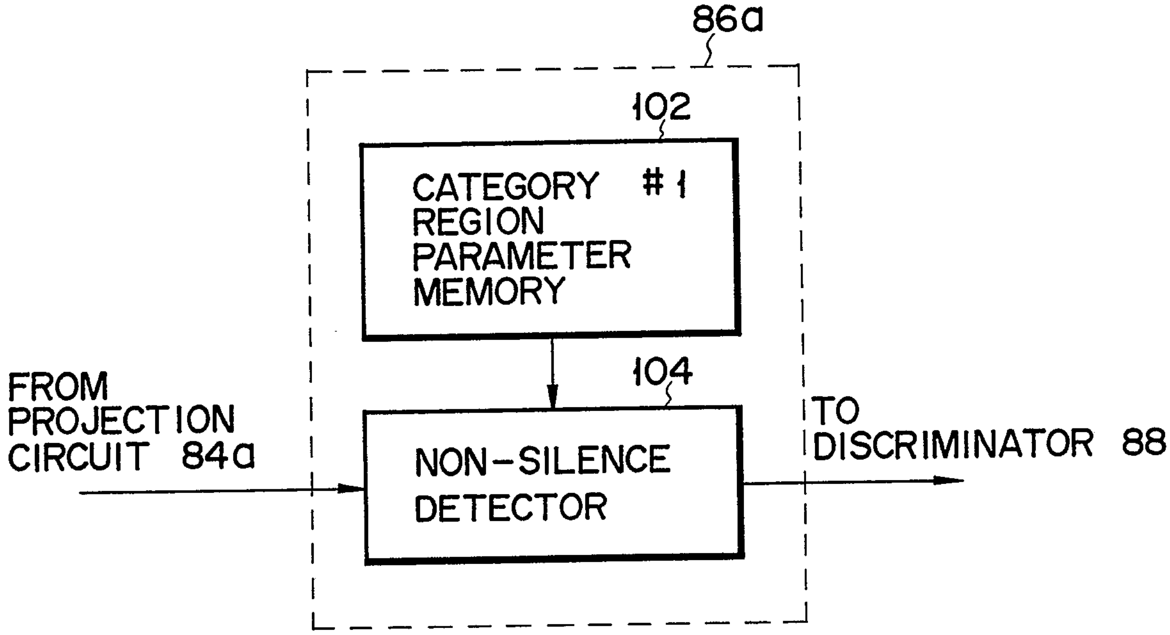
FIG. 10


IN
82
LPC CEPSTRUM CALCULATOR
84a
CHARACTERISTIC PARAMETER PROJECTION CIRCUIT (#1)
84b
CHARACTERISTC PARAMETER PROJECTION CIRCUIT (#2)
84j
CHARACTERISTIC PARAMETER PROJECTION CIRCUIT (#10)
86a
DETECTION CIRCUIT (#1)
86b
DETECTION CIRCUIT (#2)
86j
DETECTION CIRCUIT (#10)
88
SILENCE / NON-SILENCE DISCRIMINATOR
OUT

F I G. 11


84a
92
CATEGORY #1 PRIORITY COMPONENT VECTOR MEMORY
94
INNER PRODUCT CALCULATOR
FROM CEPSTRUM CALCULATOR 82
TO DETECTION CIRCUIT 86a

F I G. 12


86a
102
CATEGORY #1 REGION PARAMETER MEMORY
104
NON-SILENCE DETECTOR
FROM PROJECTION CIRCUIT 84a
TO DISCRIMINATOR 88

F I G. 13

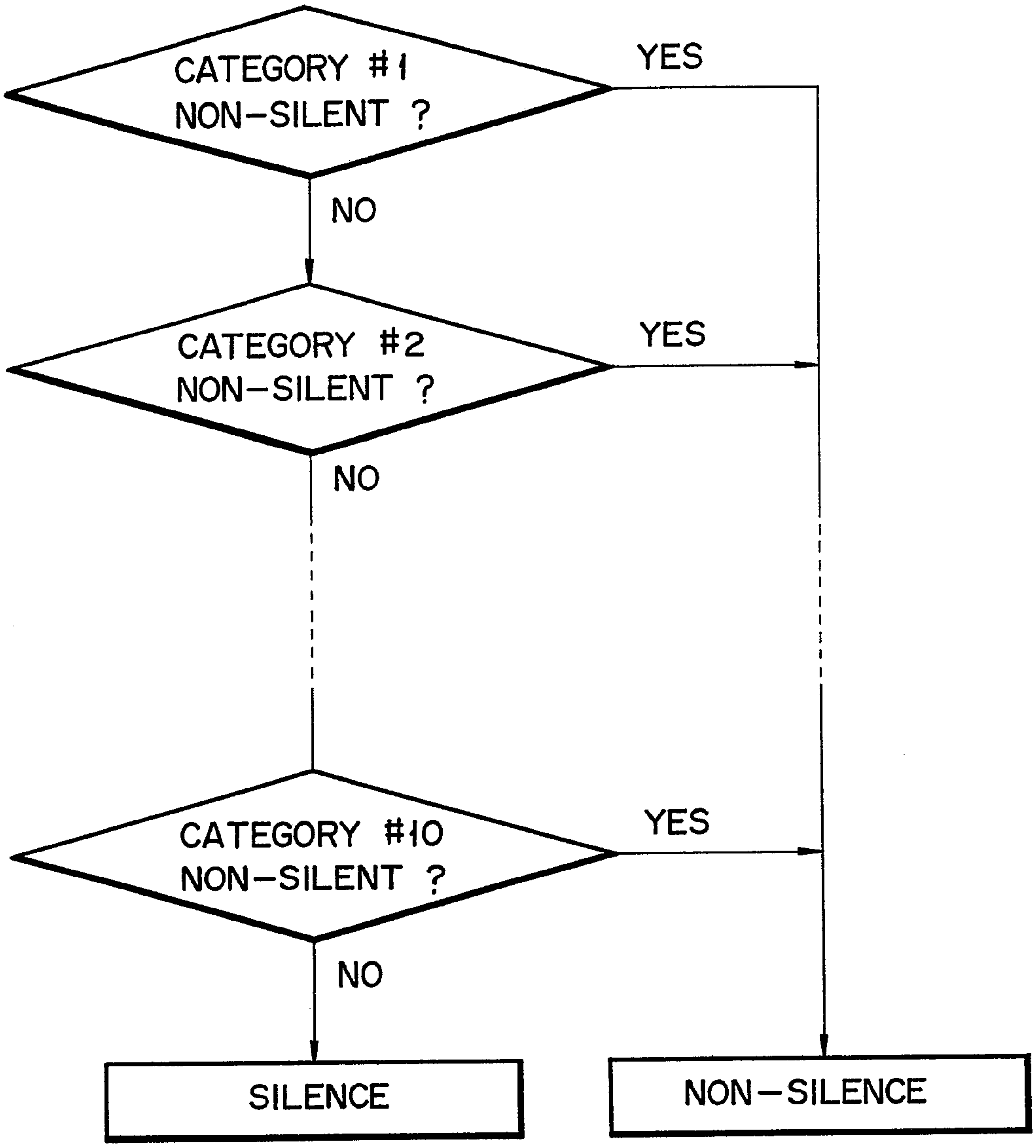

CATEGORY #1
NON-SILENT ?
YES
NO
CATEGORY #2
NON-SILENT ?
YES
NO
CATEGORY #10
NON-SILENT ?
YES
NO
SILENCE
NON-SILENCE

FIG. 14

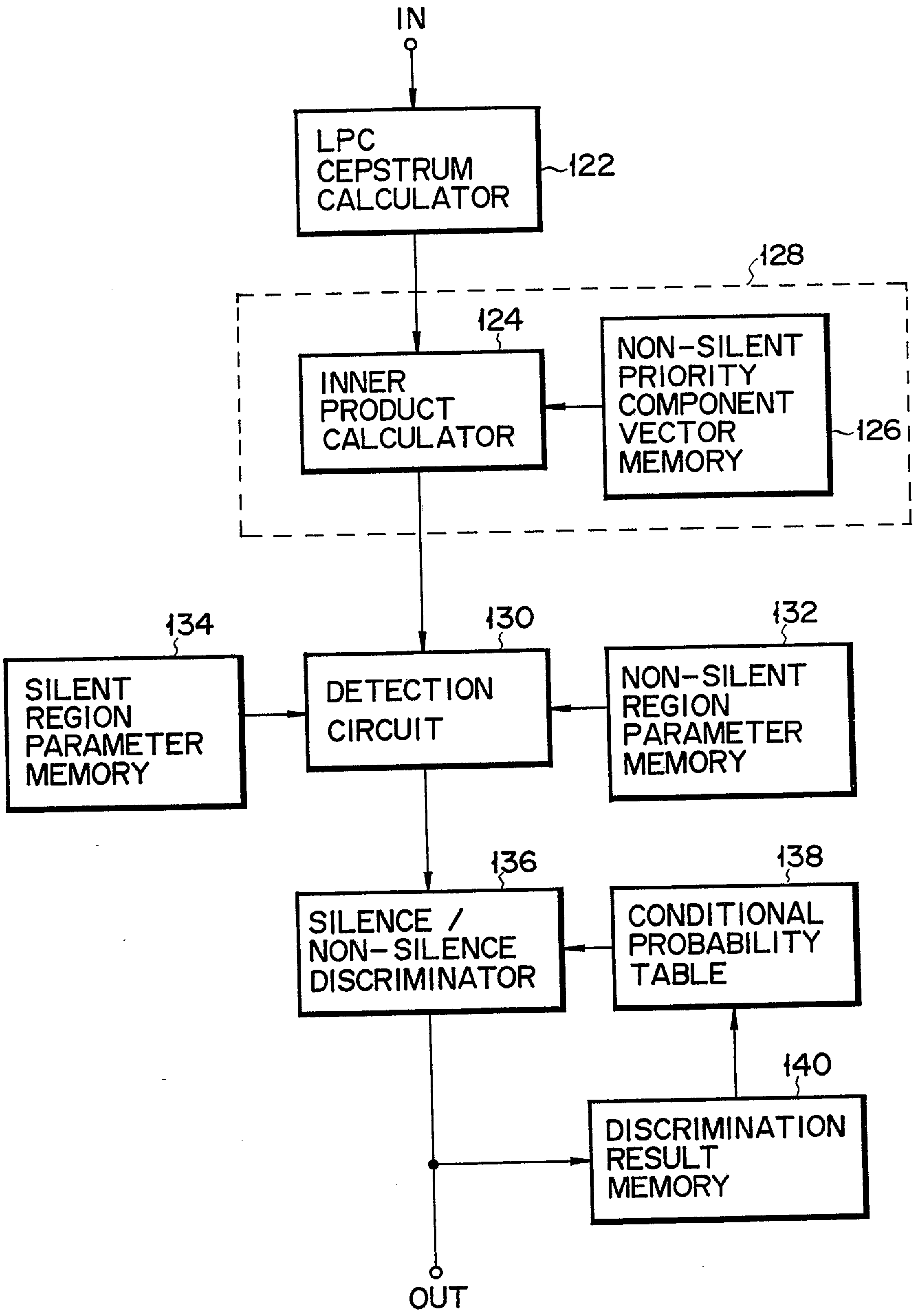

IN
LPC CEPSTRUM CALCULATOR
122
128
124
INNER PRODUCT CALCULATOR
NON-SILENT PRIORITY COMPONENT VECTOR MEMORY
126
134
SILENT REGION PARAMETER MEMORY
130
DETECTION CIRCUIT
132
NON-SILENT REGION PARAMETER MEMORY
136
SILENCE / NON-SILENCE DISCRIMINATOR
138
CONDITIONAL PROBABILITY TABLE
140
DISCRIMINATION RESULT MEMORY
OUT

FIG. 15

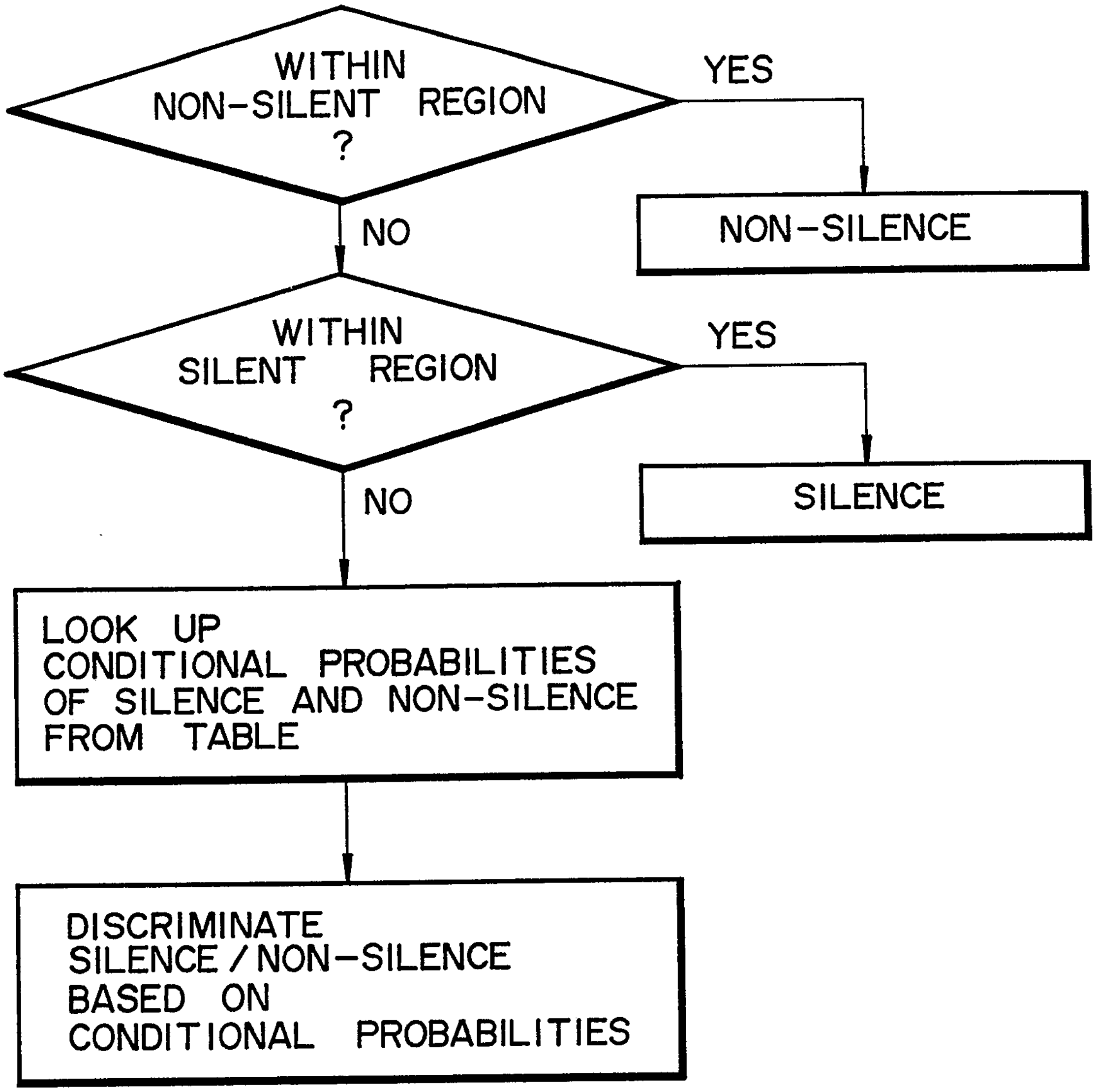
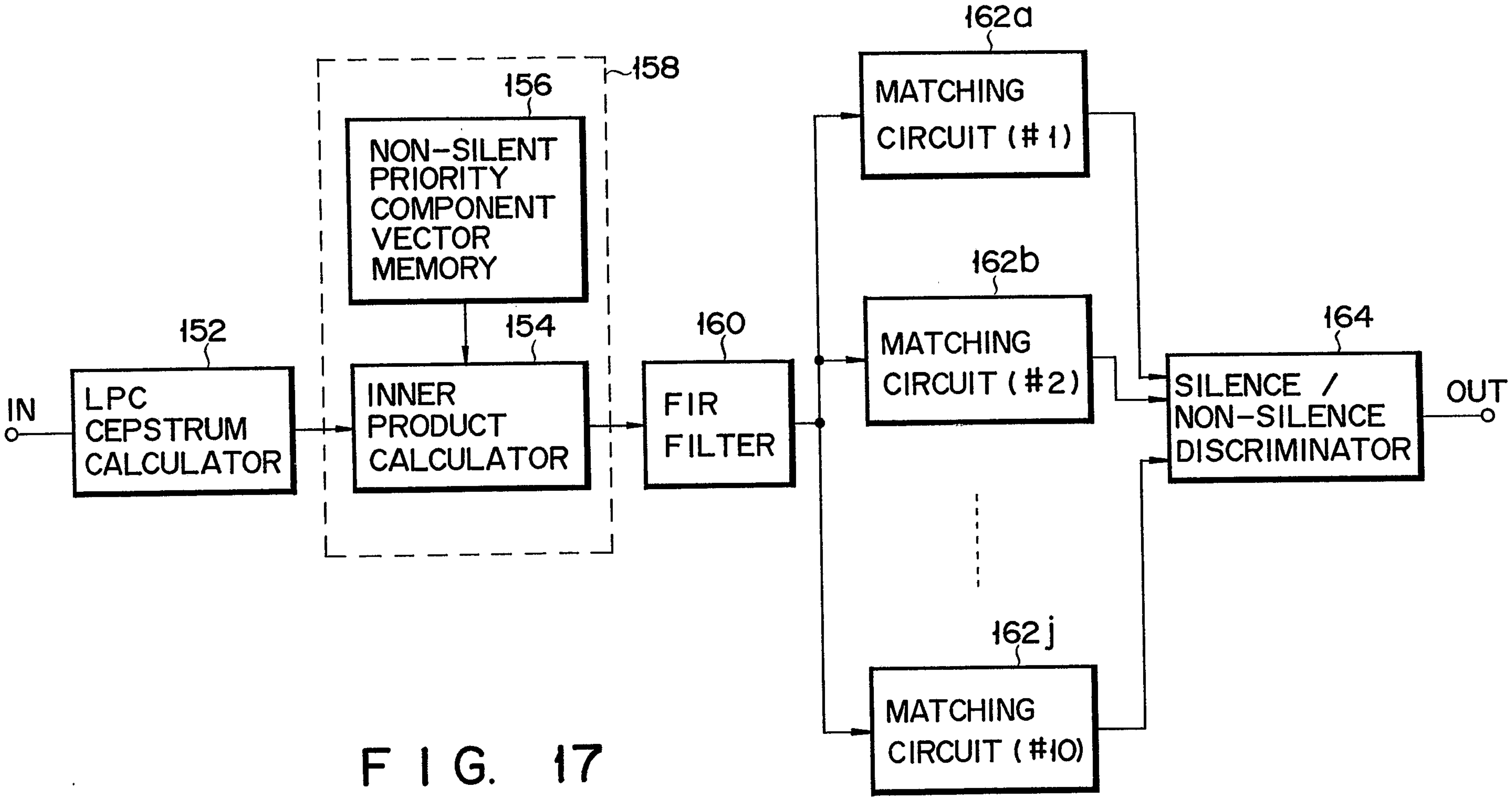

WITHIN
NON-SILENT REGION
?
YES
NON-SILENCE
NO
WITHIN
SILENT REGION
?
YES
SILENCE
NO
LOOK UP
CONDITIONAL PROBABILITIES
OF SILENCE AND NON-SILENCE
FROM TABLE
DISCRIMINATE
SILENCE / NON-SILENCE
BASED ON
CONDITIONAL PROBABILITIES

FIG. 16


IN
152
LPC CEPSTRUM CALCULATOR
158
156
NON-SILENT PRIORITY COMPONENT VECTOR MEMORY
154
INNER PRODUCT CALCULATOR
160
FIR FILTER
162a
MATCHING CIRCUIT (#1)
162b
MATCHING CIRCUIT (#2)
162j
MATCHING CIRCUIT (#10)
164
SILENCE / NON-SILENCE DISCRIMINATOR
OUT

F I G. 17

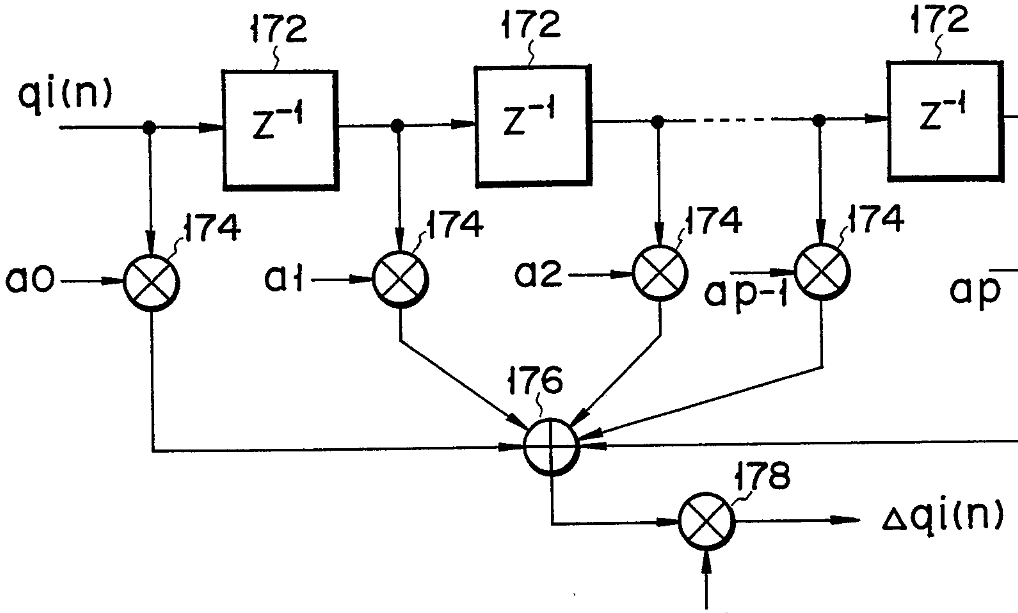

qi(n)
172
172
172
Z^{-1}
Z^{-1}
Z^{-1}
174
174
174
174
174
a0
a1
a2
ap-1
ap
176
178
Δqi(n)
σi

FIG. 18

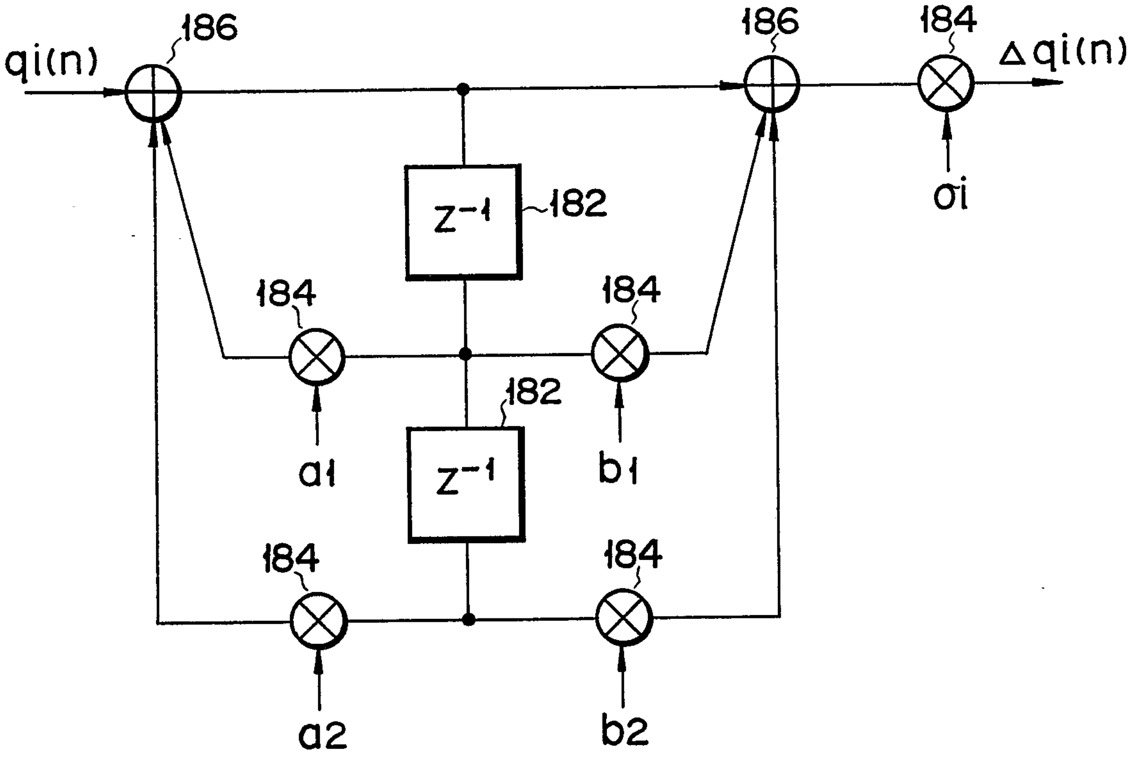

qi(n)
186
186
184
Δqi(n)
σi
Z^{-1}
182
184
184
a1
b1
182
Z^{-1}
184
184
a2
b2

FIG. 19

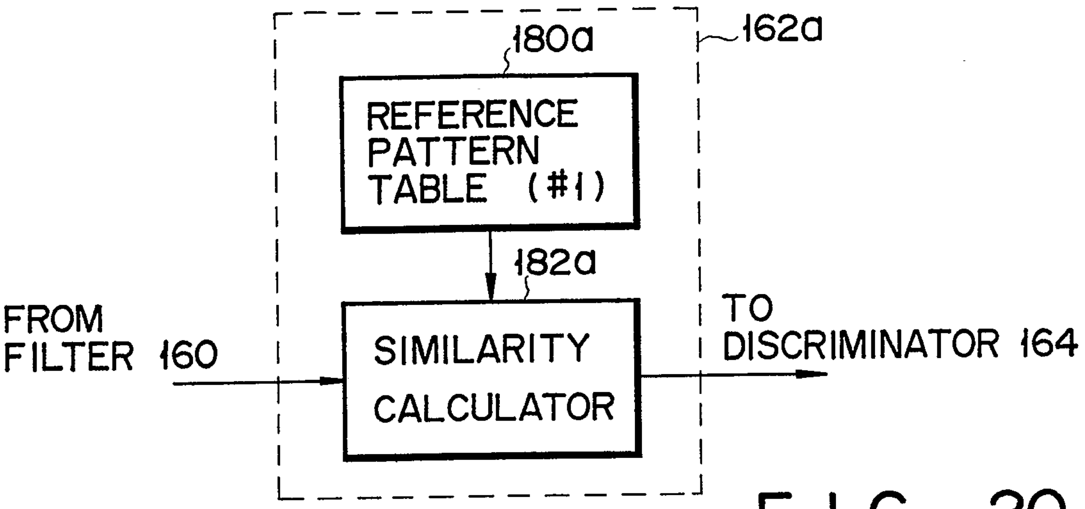

162a
180a
REFERENCE PATTERN TABLE (#1)
182a
FROM FILTER 160
SIMILARITY CALCULATOR
TO DISCRIMINATOR 164

FIG. 20

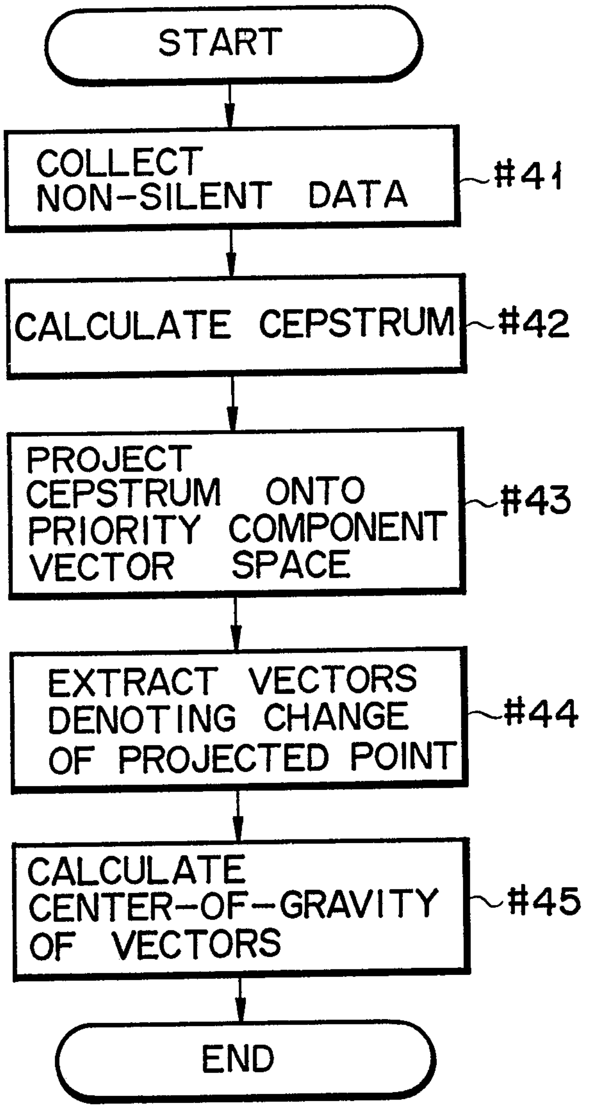
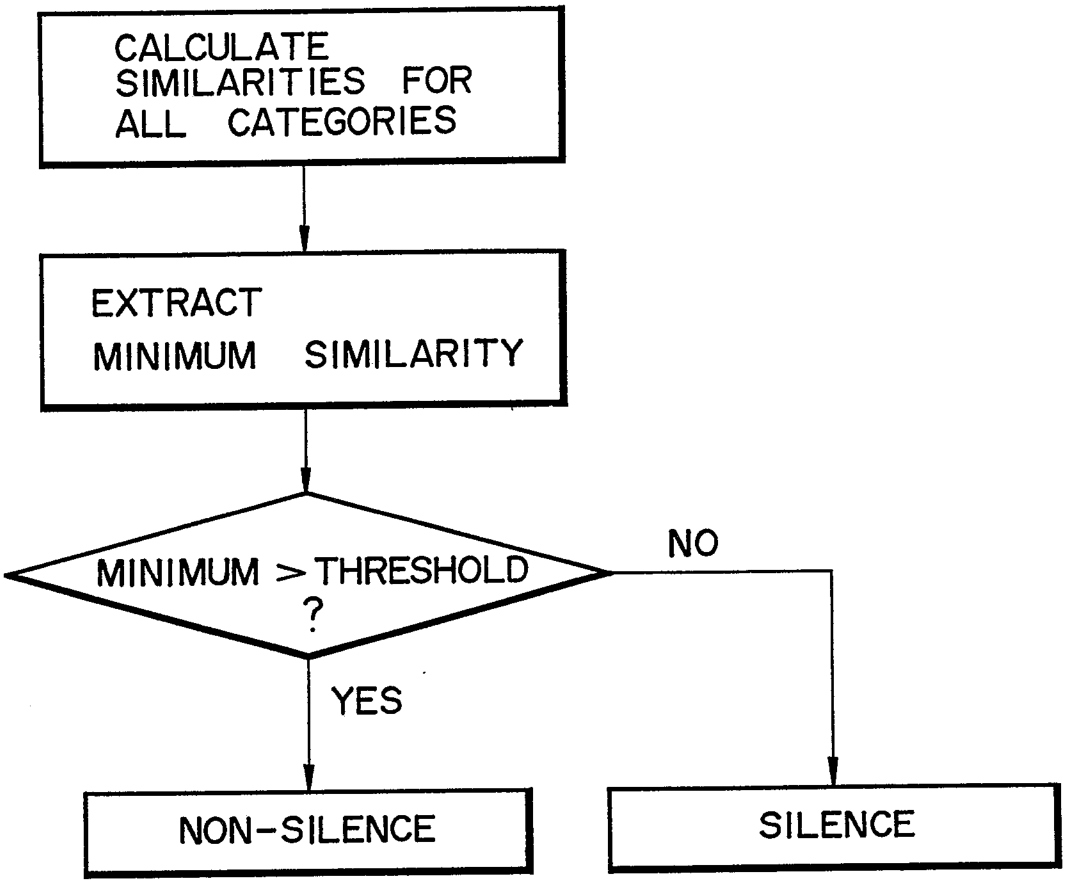

START
COLLECT NON-SILENT DATA
#41
CALCULATE CEPSTRUM
#42
PROJECT CEPSTRUM ONTO PRIORITY COMPONENT VECTOR SPACE
#43
EXTRACT VECTORS DENOTING CHANGE OF PROJECTED POINT
#44
CALCULATE CENTER-OF-GRAVITY OF VECTORS
#45
END

FIG. 21


CALCULATE SIMILARITIES FOR ALL CATEGORIES
EXTRACT MINIMUM SIMILARITY
MINIMUM > THRESHOLD ?
NO
YES
NON-SILENCE
SILENCE

F I G. 22

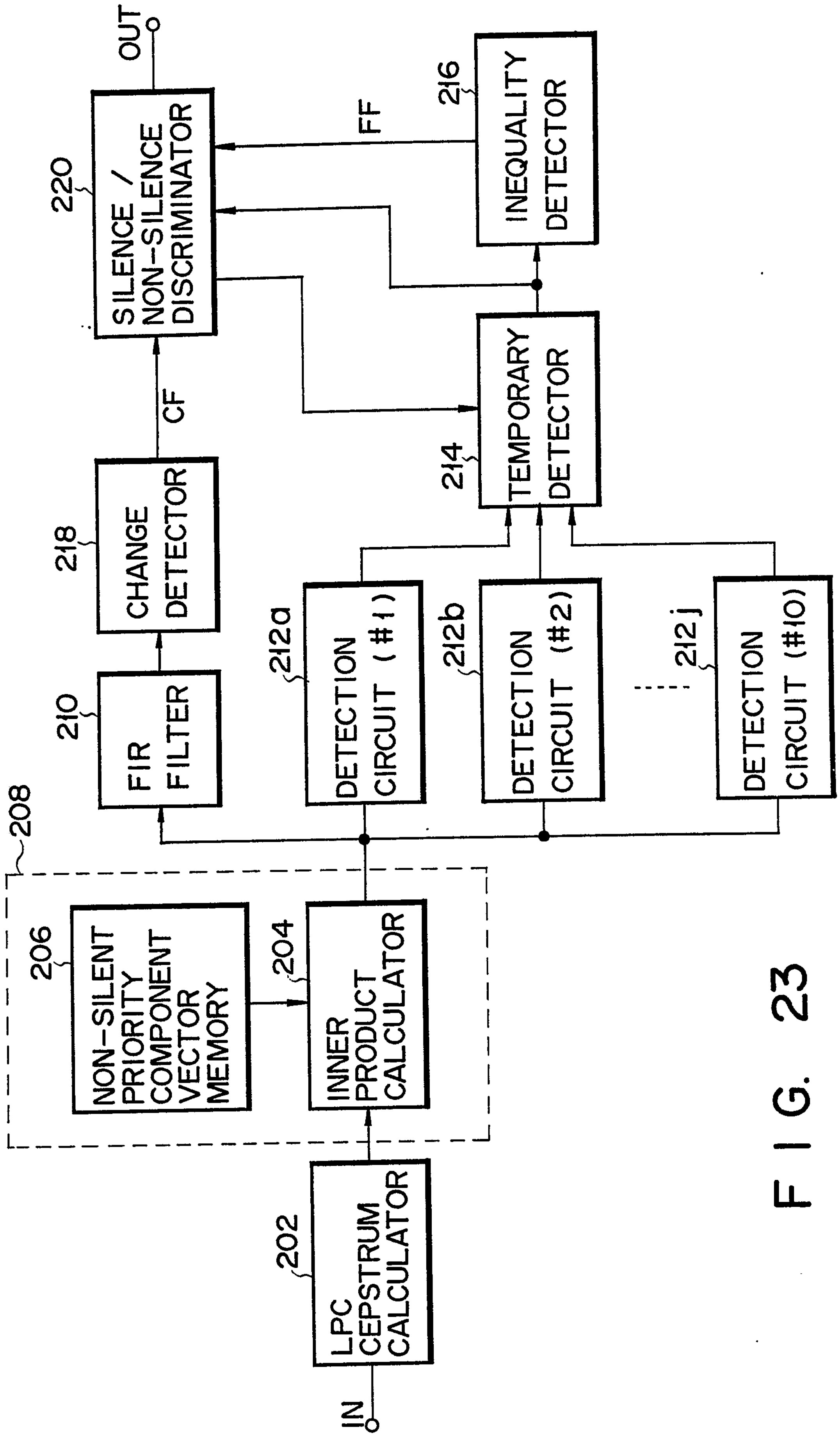

IN
202
LPC CEPSTRUM CALCULATOR
208
206
NON-SILENT PRIORITY COMPONENT VECTOR MEMORY
204
INNER PRODUCT CALCULATOR
210
FIR FILTER
218
CHANGE DETECTOR
CF
220
SILENCE / NON-SILENCE DISCRIMINATOR
OUT
212a
DETECTION CIRCUIT (#1)
212b
DETECTION CIRCUIT (#2)
212j
DETECTION CIRCUIT (#10)
214
TEMPORARY DETECTOR
216
INEQUALITY DETECTOR
FF

FIG. 23

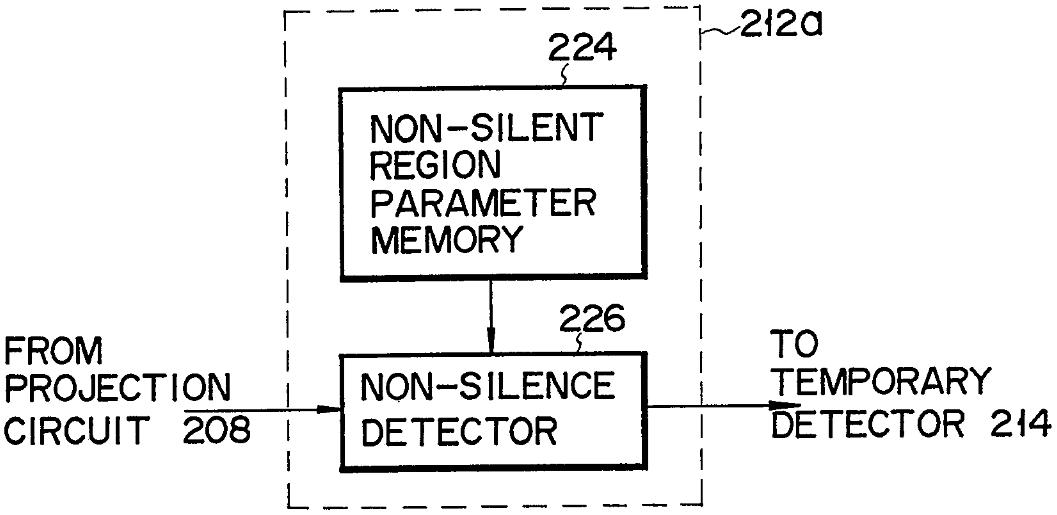

212a
224
NON-SILENT REGION PARAMETER MEMORY
226
FROM PROJECTION CIRCUIT 208
NON-SILENCE DETECTOR
TO TEMPORARY DETECTOR 214

FIG. 24

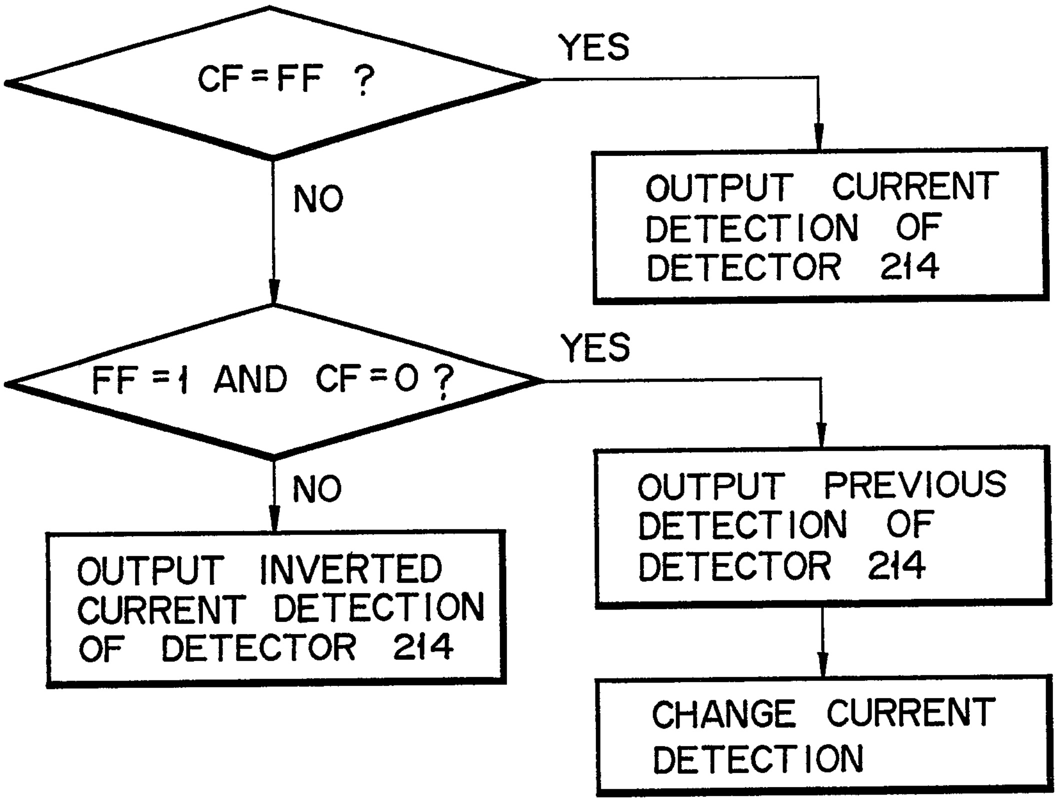

CF=FF ?
YES
NO
OUTPUT CURRENT DETECTION OF DETECTOR 214
FF=1 AND CF=0 ?
YES
NO
OUTPUT PREVIOUS DETECTION OF DETECTOR 214
OUTPUT INVERTED CURRENT DETECTION OF DETECTOR 214
CHANGE CURRENT DETECTION

FIG. 25