

Description

The present invention relates to a speech recognition apparatus and method capable of obtaining high recognition apparatus using a small number of learning patterns.

5 An information I/O operation by speech is very natural to man and superior in view of a man-machine interface. For this reason, various speech I/O systems have been studied. Almost all available speech recognition apparatuses are those for recognizing word speech, and are arranged as shown in Fig. 1.

Acoustic analyzing section 1 converts an utterance into an electrical signal using a microphone arranged therein, and acoustically analyzes the electrical signal using BPF (band pass filtering) analysis or
10 LPC (Linear Prediction Coding) analysis. Beginning frame/ending frame (end) detector 2 detects a word speech interval of the analyzed signal. Standard pattern dictionary 3 prestores standard patterns of words to be recognized. Pattern matching section 4 computes a similarity or distance (e.g., Euclidean distance, a sum of absolute values in DP (Dynamic Programming) matching, and the like), and the like. Determining
15 section 5 determines the recognition results using similarity or distance values computed in pattern matching section 4. For example, section 5 obtains a category name of a standard pattern having the highest similarity as a recognition result of the input speech.

However, in speech recognition based on the pattern matching method, a displacement (pattern deformation) along a time base between an input speech pattern and a prestored standard pattern poses a problem. The conventional system overcomes the displacement along the time base by linear expansion/compression or nonlinear expansion/compression such as Dynamic Programming.
20

Furthermore, a subspace method has been proposed. Such subspace method is described, for example, in IEEE ICASSP'78 (International Conference on Acoustic Speech and Signal Processing, by M. Jalanko and T. Kohonen, pp561-564, April 1978). In the subspace method, an orthogonalized dictionary is created based on learning patterns acquired in advance, and speech recognition is performed using the
25 dictionary. Fig. 2 shows the arrangement of the subspace method. Acoustic analyzing section 1 and end detector 2 have the same arrangements and functions as those in Fig. 1. Sampling point extracting section 6 extracts the predetermined number of sampling points obtained by equally dividing a speech section detected by end detector 2, and obtains a standard pattern represented by (the number of feature vectors $\times$ the number of sampling points). A predetermined number of such standard patterns are acquired in units of
30 categories (word recognition, syllable) to be recognized, and are stored in pattern storage section 7. Gram Schmidt orthogonalizing section (to be referred to as GS orthogonalizing section) 8 creates orthogonalized dictionary 9 using the predetermined number of (three or more) standard patterns stored in storage section 7 as follows.

Assume that an m th learning pattern of a given category is given as a_m , and learning patterns generated
35 three times are used.

i) First learning data a_1 is given as dictionary b_1 of a first axis, and the following relation is registered in orthogonalized dictionary 9:

$$b_1 = a_1 \quad (1)$$

40

ii) The following computation is performed based on second learning data a_2 using a GS orthogonalizing equation:

$$45 \quad b_2 = a_2 - \{ (a_2 \cdot b_1) b_1 \} / \|b_1\|^2 \quad (2)$$

When $\|b_2\|$ is larger than a predetermined value, $\|b_2\|$ is registered in orthogonalized dictionary 9 as b_2 of a second axis. In equation (2), $(\cdot)$ indicates an inner product; T, and $\| \cdot \|$, norm.

iii) The following computation is performed based on third learning data a_3 :

50

$$b_3 = a_3 - \sum_{m=1}^2 \{ (a_3^T \cdot b_m) b_m \} / \|b_m\| \quad \dots \quad (3)$$

55

When $\|b_3\|$ is larger than a predetermined value, $\|b_3\|$ is registered in orthogonalized dictionary 9 as dictionary b_3 of a third axis. However, if the dictionary of the second axis is not yet obtained,

computation of equation (2) is performed.

Operation i) through iii) are performed for each category, and thus preparing the orthogonalized dictionaries.

Similarity computing section 10 computes the following equation between each orthogonalized dictionary 9 created as described above and each input speech pattern X:

$$S_i = \sum_{r=1}^{K_i} (X \cdot b_{i,r})^2 \quad \dots (4)$$

10

As a result, a similarity with each orthogonalized dictionary $b_{i,r}$ of category i is computed. Input speech pattern X is recognized in accordance with the computed similarity. Note that orthogonalized dictionaries $b_{i,r}$ of category i are normalized in advance. K_i indicates the number of dictionaries (number of axes).

However, in the method using the GS orthogonalization, a deviation born by each orthogonal axis is not clear. More specifically, in orthogonalization, a variety of orthogonal axes can be considered, and a pattern deviation changes depending on the orthogonal axis. For this reason, standard patterns $\{b_{i,1}, b_{i,2}, b_{i,3}\}$ of category i of orthogonalized dictionary computed as described above do not always well represent those of category i .

It is an object of the present invention to provide a speech recognition apparatus and method, which can create an orthogonalized dictionary which well represents standard patterns with a small number of learning patterns, and can improve recognition performance.

To solve this object the present invention provides a speech recognition apparatus and method as specified in claim 1, 2 and 5, respectively.

According to the speech recognition system of the present invention, the orthogonalized dictionary is generated to absorb a deviation of the speech pattern along the time base. Therefore, speech recognition can be performed without being influenced by the deviation of the input speech pattern along the time base. As a result, the recognition performance can be improved. Since orthogonalized dictionary 9 is generated using an orthogonalizing time filter, a high-performance orthogonalized dictionary can be efficiently generated with a small number of learning patterns.

Since an additional dictionary orthogonal to the orthogonalized dictionary is generated, a pattern deviation other than that along the time base described above can be efficiently absorbed, and recognition processing can be performed. Thus, the recognition performance can be further improved.

Three or more types of filters are used to obtain an average pattern of acquired learning patterns, and the average pattern is smoothed along both the time base and the frequency axis, thereby obtaining a dictionary of a first axis. Thus, deviations of a speech pattern along the time base and the frequency axis can be effectively absorbed. Furthermore, since the average pattern is differentiated along the time base to obtain a dictionary of a second axis, a displacement of the speech pattern along the time base can be effectively absorbed. Since the averaged pattern is differentiated along the frequency axis to obtain a dictionary of a third axis, the displacement of the speech pattern along the frequency axis can be effectively absorbed.

In this manner, since the orthogonalized dictionary which absorbs pattern deviations along the time base and the frequency axis can be generated, each dictionary pattern of the orthogonalized dictionary can correspond to displacements caused by the deviations, thus greatly contributing to improvement in recognition performance. In addition, dictionaries of the second and third axes are obtained based on a dictionary pattern (first axis) generated by the average pattern absorbing pattern deviations along the time base and the frequency axis, thereby generating the orthogonalized dictionary. Thus, a pattern deviation born by each orthogonal axis of the orthogonalized dictionary itself does not become unclear unlike in the conventional system. As a result, a high-performance orthogonalized dictionary can be generated by effectively using a small number of learning patterns.

The additional dictionary orthogonal to the orthogonalized dictionary is generated, and is added to the orthogonalized dictionary. Therefore, the additional dictionary can effectively absorb pattern deviations other than those along the time base and the frequency axis described above so as to perform recognition processing. Thus, recognition performance can be greatly improved.

Other objects and features of the present invention will be apparent from the following description taken in connection with the accompanied drawings, in which:

Fig. 1 is a block diagram showing a speech recognition system of a conventional pattern matching method;

Fig. 2 is a block diagram of a speech recognition system of a conventional subspace method;

Fig. 3 is a block diagram showing an embodiment of a speech recognition system according to the present invention;

Fig. 4 is a flow chart showing processing of an orthogonalizing time filter section shown in Fig. 3;

Fig. 5 is a flow chart showing reorthogonalization processing of the orthogonalizing time filter section shown in Fig. 3;

Fig. 6 is a flow chart showing another processing of the orthogonalizing time filter section shown in Fig. 3;

Fig. 7 is a flow chart showing still another processing of the orthogonalizing time filter section shown in Fig. 3;

Fig. 8 is a block diagram showing another embodiment of a speech recognition system according to the present invention;

Fig. 9 is a flow chart showing processing of generating an additional dictionary in the embodiment shown in Fig. 8;

Fig. 10 is a graph showing test samples for evaluating speech recognition performance in the embodiment shown in Fig. 8;

Fig. 11 is a block diagram showing still another embodiment of a speech recognition system according to the present invention;

Fig. 12 is a flow chart showing processing of an orthogonalizing time frequency filter section shown in Fig. 11;

Fig. 13 is a flow chart showing another processing of the orthogonalizing time frequency filter section shown in Fig. 11; and

Fig. 14 is a block diagram showing still another embodiment of the present invention.

Fig. 3 is a block diagram showing an embodiment of a speech recognition system according to the present invention. The same reference numerals in Fig. 3 denote the same parts as in the conventional speech recognition systems shown in Figs. 1 and 2.

The characteristic feature of this embodiment is use of orthogonalizing time filter section 11 as means for generating an orthogonalized dictionary. In the conventional speech recognition system, orthogonalized dictionary 9 is generated by GS orthogonalizing section 8 using learning patterns stored in pattern storage section 7. In this invention, orthogonalizing time filter section 11 is used in place of GS orthogonalizing section 8. Orthogonalizing time filter section 11 is constituted by a plurality of filters for executing at least smoothing processing and differentiating processing, e.g., a plurality of orthogonalizing time filters. Filter section 11 can comprise, e.g., a digital signal processor. Assume that in this embodiment, a learning pattern acquired in pattern storage section 7 is given as a data series which consists of 16 acoustically analyzed feature vectors represented by j ($= 0, 1, \dots, 15$), and is sampled at 18 sampling points which equally divide a speech section into 17 subintervals given by k ($= 0, 1, \dots, 16$).

If the m th one of three learning patterns which are acquired for category i is given as $a_{m(j,k)}$, filter section 11 generates orthogonalized dictionary 9 as follows. The following description will be made with reference to the flow chart shown in Fig. 4.

First, in step 21, filter section 11 computes average pattern $A_{(j,k)}$ from learning pattern $a_{m(j,k)}$ of category i using equation (5):

$$A(j,k) = (1/3) \sum_{m=1}^3 a_m(j,k) \quad \dots (5)$$

$$[j = 0, 1, \dots, 15, k = 0, 1, 2, \dots, 16]$$

In step 23, filter section 11 substitutes average pattern $A_{(j,k)}$ computed in step 21 into equation (6) to compute dictionary $b_{1(j,k)}$ of the first axis, and registers it in orthogonalized dictionary 9. Dictionary $b_{1(j,k)}$ can be computed by smoothing average pattern $A_{(j,k)}$ along the time base. Filter section 11 normalizes dictionary $b_{1(j,k)}$ and registers it as dictionary data of the first axis serving as reference data of orthogonalized dictionary 9.

$$b_{1(j,k)} = A_{(j,k-1)} + 2A_{(j,k)} + A_{(j,k+1)} \quad (6)$$

$$[j = 0, 1, \dots, 15, k = 0, 1, \dots, 15]$$

In step 25, filter section 11 substitutes average pattern $a_{(j,k)}$ into equation (7) to compute dictionary $b_{2(j,k)}$ of a second axis, and normalizes the computed dictionary. Thereafter, filter section 11 registers the

dictionary in orthogonalized dictionary 9. Dictionary $b_{2(j,k)}$ can be computed by differentiating average pattern $A_{(j,k)}$ along the time base.

$$b_{2(j,k)} = -A_{(j,k-1)} + A_{(j,k+1)} \\ [j = 0, 1, \dots, 15, k = 0, 1, \dots, 15] \quad (7)$$

Dictionary $b_{2(j,k)}$ of the second axis computed in this manner is not perfectly orthogonal to dictionary $b_{1(j,k)}$ of the first axis. Therefore, reorthogonalization processing in step 27 shown in Fig. 5 can be performed for dictionary $b_{2(j,k)}$. More specifically, dictionary $b_{2(j,k)}$ is reorthogonalized using equation (8), and reorthogonalized dictionary data $B_{2(j,k)}$ is normalized. Then, the normalized data can be registered in orthogonalized dictionary 9 as the dictionary of the second axis.

$$B_{2(j,k)} = b_{2(j,k)} - (b_{2(j,k)} \cdot b_{1(j,k)})b_{1(j,k)} \quad (8)$$

However, if such reorthogonalization processing is not performed, satisfactory recognition performance can be obtained using dictionary $b_{2(j,k)}$ of the second axis computed as described above.

The processing operations in steps 21, 23, and 25 are repetitively executed, thus generating orthogonalized dictionary 9.

Fig. 6 shows another processing of generating orthogonalized dictionary 9 by filter section 11.

More specifically, in step 29, acquired learning patterns $a_{m(j,k)}$ are substituted in equation (9) to obtain dictionary $b_{1(j,k)}$ of the first axis, and the dictionary is normalized and registered in orthogonalized dictionary 9.

$$b_{1(j,k)} = (1/3) \sum_{m=1}^3 \{ (a_{m(j,k-1)} + 2 \cdot a_{m(j,k)} + a_{m(j,k+1)}) \} \\ [j = 0, 1, \dots, 15, k = 0, 1, \dots, 15] \quad \dots (9)$$

Filter section 11 substitutes learning patterns $a_{m(j,k)}$ in equation (10) in step 31 to compute dictionary $b_{2(j,k)}$ of the second axis, and registers it in orthogonalized dictionary 9.

$$b_{2(j,k)} = (1/3) \sum_{m=1}^3 \{ -a_{m(j,k-1)} + a_{m(j,k+1)} \} \\ [j = 0, 1, \dots, 15, k = 0, 1, \dots, 15] \quad \dots (10)$$

The processing operations shown in steps 29 and 31 are repeated a plurality of times corresponding to the number of categories, thereby generating the orthogonalized dictionary. In the method of generating the orthogonalized dictionary shown in Fig. 6, dictionary $b_{1(j,k)}$ of the first axis smoothed along the time base and dictionary $b_{2(j,k)}$ of the second axis differentiated along the time base can be directly calculated from the predetermined number of learning patterns $a_{m(j,k)}$ acquired without temporarily computing average pattern

$A_{m(j,k)}$.

In the above description, orthogonalized dictionary 9 is computed up to the second axis. Furthermore, a dictionary of a third axis and thereafter may be generated by quadratic differentiation, or the like. In this case, as a learning pattern, one obtained by extracting not 18 points described above but 20 points or more can be used.

This case will be explained with reference to the flow chart shown in Fig. 7.

In step 33, averaged pattern $A_{(j,k)}$ is computed from learning patterns $a_{m(j,k)}$ of category i using equation (5) as in step 21 shown in Fig. 4.

In step 35, averaged pattern $A_{(j,k)}$ obtained in step 33 is substituted in equation (11) to compute dictionary $b_{1(j,k)}$ of the first axis, and the dictionary is normalized and registered in orthogonalized dictionary 9.

$$b_{1(j,k)} = A_{(j,k-2)} + 4 \cdot A_{(j,k-1)} + 6 \cdot A_{(j,k)} + 4 \cdot A_{(j,k+1)} + A_{(j,k+2)} \\ [j = 0, 1, \dots, 15, k = 0, 1, \dots, 15] \quad (11)$$

In step 37, average pattern $A_{(j,k)}$ is substituted in equation (12) to obtain dictionary $b_{2(j,k)}$ of the second axis.

$$b_{2(j,k)} = -A_{(j,k-2)} - 2^*A_{(j,k-1)} + 2^*A_{(j,k+1)} + A_{(j,k+2)} \\ [j = 0, 1, \dots, 15, k = 0, 1, \dots, 15] \quad (12)$$

In step 39, average pattern $A_{(j,k)}$ is substituted in equation (13) to obtain dictionary $b_{3(j,k)}$ of the third axis.

$$b_{3(j,k)} = -A_{(j,k-2)} - 2^*A_{(j,k-1)} + 3^*A_{(j,k)} - 2^*A_{(j,k+1)} - A_{(j,k+2)} \\ [j = 0, 1, \dots, 15, k = 0, 1, \dots, 15] \quad (13)$$

Table below shows test data which were obtained such that 30 words of speech data consisting of numeric speech inputs and personal names were acquired 15 times from five males and three females, and the acquired data corresponding to 5 times were used for learning and those corresponding to 10 times were used for evaluating recognition performance.

Table

Method	Object to be Recognized	
	Numeric + Personal	Numeric
Averaging Method	93.6%	95.5%
Multi template Method	94.4	96.6
Subspace Method 4 axes	94.3	97.3
Differential Orthogonalizing Method 2 axes	95.4	97.6
Differential Orthogonalizing Method 4 axes	96.1	98.3

As can be seen from these test data, when an orthogonalized dictionary comprising orthogonalizing time filters is used, recognition performance can be improved more than a case wherein an orthogonalized dictionary is generated by the GS orthogonalizing method.

From the above test data, according to the speech recognition system and method of the present invention wherein the orthogonalized dictionary is generated using differential filters for compensating for a displacement along the time base are used so as to perform speech recognition, high recognition performance can be with a small number of learning patterns.

Fig. 8 is a block diagram showing another embodiment of a speech recognition system of the present invention.

The characteristic feature of this embodiment is that a GS orthogonalizing section is arranged in addition to the embodiment shown in Fig. 3, so that a dictionary orthogonal to the orthogonalized dictionary generated by orthogonalizing time filter section 11 is generated as an additional dictionary, and the additional dictionary is added to orthogonalized dictionary 9. Note that the same reference numerals in Fig. 8 denote the same parts as in Fig. 3, and a detailed description thereof will be omitted. Processing of GS orthogonalizing section 8 for generating an additional dictionary orthogonal to the orthogonalized dictionary from learning patterns stored in pattern storage section 7 will be described hereinafter with reference to the flow chart shown in Fig. 9.

More specifically, in step 41, section 8 substitutes learning patterns $a_{m(j,k)}$ acquired in pattern storage section 7 into the GS orthogonalizing equation (equation (14)).

$$b_{p+m} = a_m - \sum_{n=1}^p [(a_m^T \cdot b_n) b_n] / \|b_n\|^2 \quad \dots \quad (14) \\ [n = 0, 1, \dots, p, m = 0, 1, \dots, M]$$

(where p is the number of axes of the orthogonalized dictionary)

In step 43, section 8 checks if $\|b_{p+m}\|$ computed in step 41 is larger than a predetermined value. If YES in step 43, $\|b_{p+m}\|$ is registered in orthogonalized dictionary 9 as an additional dictionary.

The processing operations in steps 41, 43, and 45 are repetitively executed a plurality of times corresponding to the number of categories.

- 5 The additional dictionary is generated by GS orthogonalizing section 8 in addition to the orthogonalized dictionary constituted by using orthogonalizing time filters, and is registered in the orthogonalized dictionary, so that recognition processing can be performed while pattern deviations other than those along the time base can be effectively absorbed.

- 10 From test data shown in Fig. 10, in the complex similarity method (KL expansion), if the number of times of utterance is increased, a rate of decrease in error rate is larger than that in differentiation/orthogonalization. However, for utterance of about 5 times, the orthogonalized dictionary computed by differentiation/orthogonalization is preferably used since it is superior in view of recognition performance.

- 15 Fig. 11 is a block diagram showing still another embodiment of a speech recognition system according to the present invention. The same reference numerals in Fig. 11 denote the same parts as in Fig. 3, and a detailed description thereof will be omitted. The characteristic feature of this embodiment is that orthogonalizing time frequency filter section 12 consisting of three or more types of orthogonalizing time frequency filters is used in place of orthogonalizing time filter section 11 shown in Fig. 3.

Processing of orthogonalizing time frequency filter section 12 for generating an orthogonalized dictionary will be described below with reference to the flow chart shown in Fig. 12.

- 20 In step 51, average pattern $A_{(j,k)}$ is computed from learning patterns $a_{m(j,k)}$ of category i as follows:

$$A_{(j,k)} = (1/3) \sum_{m=1}^3 a_{m(j,k)}$$

25 $[j = 0, 1, 2, \dots, 15, k = 0, 1, 2, \dots, 17]$

In step 53, using average pattern $A_{(j,k)}$ computed described above, dictionary $b_{1(j,k)}$ of the first axis is computed as follows:

$$b_{1(j,k)} = A_{(j-1,k-1)} + A_{(j-1,k)} + A_{(j-1,k+1)} + A_{(j,k-1)} + 2A_{(j,k)} + A_{(j,k+1)} + A_{(j+1,k-1)} + A_{(j+1,k)} + A_{(j+1,k+1)}$$

30 $[j = 0, 1, \dots, 13, k = 0, 1, \dots, 15]$

- 35 The computed dictionary is normalized and registered in orthogonalized dictionary 9. Dictionary $b_{1(j,k)}$ is one obtained by smoothing average pattern $A_{(j,k)}$ along the time base and the frequency axis, and is registered as the dictionary data of the first axis as reference data in orthogonalized dictionary 9.

In step 55, dictionary $b_{2(j,k)}$ of the second axis is computed using average pattern $A_{(j,k)}$ as follows:

$$b_{2(j,k)} = -A_{(j-1,k-1)} + A_{(j-1,k+1)} + \{-A_{(j,k-1)} + A_{(j,k+1)}\} + \{-A_{(j+1,k-1)} + A_{(j+1,k+1)}\}$$

40 $[j = 0, 1, \dots, 13, k = 0, 1, \dots, 15]$

The computed dictionary is normalized, and is registered in orthogonalized dictionary 9. Dictionary $b_{2(j,k)}$ is obtained by differentiating average pattern $A_{(j,k)}$ along the time base.

- 45 Since dictionary $b_{2(j,k)}$ of the second axis computed in this manner is not perfectly orthogonal to dictionary $b_{1(j,k)}$ of the first axis, reorthogonalization processing expressed as the following equation can be performed as needed:

$$B_{2(j,k)} = b_{2(j,k)} - (b_{2(j,k)} \cdot b_{1(j,k)})b_{1(j,k)}$$

- 50 Reorthogonalized dictionary data $B_{2(j,k)}$ can be registered in orthogonalized dictionary 9 as new dictionary $b_{2(j,k)}$ of the second axis. However, even if such reorthogonalization is not performed, satisfactory recognition performance can be obtained by dictionary $b_{2(j,k)}$ obtained as described above.

The processing operations in steps 51, 53, and 55 are repetitively executed for each category, thus generating orthogonalized dictionary 9.

- 55 Fig. 13 is a flow chart showing processing of generating an orthogonalized dictionary when the number of axes is 3. Note that steps 57, 59, and 61 in Fig. 13 are the same as steps 51, 53, and 55 in Fig. 12, and a detailed description thereof will be omitted.

In step 63 in Fig. 13, using average pattern $A_{(j,k)}$, dictionary $b_{3(j,k)}$ of the third axis is computed by the following equation:

$$b_{3(j,k)} = -A_{(j-1,k-1)} - A_{(j-1,k)} - A_{(j-1,k+1)} + A_{(j+1,k-1)} + A_{(j+1,k)} + A_{(j+1,k+1)}$$

5 $[j = 0, 1, \dots, 13, k = 0, 1, \dots, 15]$

The computed dictionary is normalized, and is then registered in orthogonalized dictionary 9.

Dictionary $b_{3(j,k)}$ is obtained by differentiating average pattern $A_{(j,k)}$ along the frequency axis.

Orthogonalized dictionary 9 can be generated by repeating processing in steps 57, 59, 61, and 63 for
10 each category.

Fig. 14 is a block diagram showing still another embodiment of the present invention.

The characteristic feature of this embodiment is that a GS orthogonalizing section is arranged in addition to the embodiment shown in Fig. 11, so that a dictionary orthogonal to the orthogonalized dictionary generated by orthogonalizing time frequency filter section 12 is generated as an additional
15 dictionary, and is added to the orthogonalized dictionary 9.

Note that the same reference numerals in Fig. 14 denote the same parts as in Fig. 11, and a detailed description thereof will be omitted. The processing of GS orthogonalizing section 8 shown in Fig. 14 is performed using equation (14) in the same manner as that shown in Fig. 8, and a detailed description thereof will be omitted.

20

Claims

1. A speech recognition apparatus comprising:

- acoustic analyzing means (1) for analyzing an input speech to obtain an input speech pattern;
- 25 - sampling pattern generating means (6) for generating a sampling pattern by extracting a predetermined number of sampling points obtained by equally dividing the input speech pattern obtained from said acoustic analyzing means (1),
- a time filter section (11), and
- similarity computing means (10),
- 30 characterized in that
- said time filter section (11) comprises
 - (a) means (21) for computing an average pattern $A_{(j,k)}$ of n learning patterns $a_{m(j,k)}$ of word recognition or syllable category and collected by said sampling pattern generating means (6) based on the following equation:

35

$$A_{(j,k)} = (1/n) \sum_{m=1}^n a_{m(j,k)},$$

40

where j denotes a number of feature vectors obtained by acoustic analyzing in said acoustic analyzing means (1), and k denotes a number of sampling points for dividing speech pattern in said sampling pattern generating means (6),

- 45 (b) means (23) for obtaining dictionary data $b_{1(j,k)}$ of a first axis using the average pattern $A_{(j,k)}$ in accordance with the following equation:

$$b_{1(j,k)} = A_{(j,k-1)} + 2A_{(j,k)} + A_{(j,k+1)}, \text{ and}$$

50

- (c) means (25) for obtaining dictionary data $b_{2(j,k)}$ by differentiating average pattern $A_{(j,k)}$ along a second axis in accordance with the following equation:

$$b_{2(j,k)} = -A_{(j,k-1)} + A_{(j,k+1)}, \text{ and}$$

55

- said similarity computing means (10) comprising means for computing similarities between the input speech pattern acoustically analyzed by said acoustic analyzing means (1), and the



In step 63 in Fig. 13, using average pattern $A_{(j,k)}$, dictionary $b_{3(j,k)}$ of the third axis is computed by the following equation:

$$b_{3(j,k)} = -A_{(j-1,k-1)} - A_{(j-1,k)} - A_{(j-1,k+1)} + A_{(j+1,k-1)} + A_{(j+1,k)} + A_{(j+1,k+1)}$$

5 $[j = 0, 1, \dots, 13, k = 0, 1, \dots, 15]$

The computed dictionary is normalized, and is then registered in orthogonalized dictionary 9.

Dictionary $b_{3(j,k)}$ is obtained by differentiating average pattern $A_{(j,k)}$ along the frequency axis.

Orthogonalized dictionary 9 can be generated by repeating processing in steps 57, 59, 61, and 63 for each category.

Fig. 14 is a block diagram showing still another embodiment of the present invention.

The characteristic feature of this embodiment is that a GS orthogonalizing section is arranged in addition to the embodiment shown in Fig. 11, so that a dictionary orthogonal to the orthogonalized dictionary generated by orthogonalizing time frequency filter section 12 is generated as an additional dictionary, and is added to the orthogonalized dictionary 9.

Note that the same reference numerals in Fig. 14 denote the same parts as in Fig. 11, and a detailed description thereof will be omitted. The processing of GS orthogonalizing section 8 shown in Fig. 14 is performed using equation (14) in the same manner as that shown in Fig. 8, and a detailed description thereof will be omitted.

Claims

1. A speech recognition apparatus comprising:

- acoustic analyzing means (1) for analyzing an input speech to obtain an input speech pattern;
- sampling pattern generating means (6) for generating a sampling pattern by extracting a predetermined number of sampling points obtained by equally dividing the input speech pattern obtained from said acoustic analyzing means (1),
- a time filter section (11), and
- similarity computing means (10),

characterized in that

- said time filter section (11) comprises

(a) means (21) for computing an average pattern $A_{(j,k)}$ of n learning patterns $a_{m(j,k)}$ of word recognition or syllable category and collected by said sampling pattern generating means (6) based on the following equation:

$$A_{(j,k)} = (1/n) \sum_{m=1}^n a_{m(j,k)},$$

where j denotes a number of feature vectors obtained by acoustic analyzing in said acoustic analyzing means (1), and k denotes a number of sampling points for dividing speech pattern in said sampling pattern generating means (6),

(b) means (23) for obtaining dictionary data $b_{1(j,k)}$ of a first axis using the average pattern $A_{(j,k)}$ in accordance with the following equation:

$$b_{1(j,k)} = A_{(j,k-1)} + 2A_{(j,k)} + A_{(j,k+1)}, \text{ and}$$

(c) means (25) for obtaining dictionary data $b_{2(j,k)}$ by differentiating average pattern $A_{(j,k)}$ along a second axis in accordance with the following equation:

$$b_{2(j,k)} = -A_{(j,k-1)} + A_{(j,k+1)}, \text{ and}$$

- said similarity computing means (10) comprising means for computing similarities between the input speech pattern acoustically analyzed by said acoustic analyzing means (1), and the



dictionaries of the first and the second axes generated by said time filter section (11).

2. A speech recognition apparatus comprising:

- acoustic analyzing means (1) for analyzing an input speech to obtain an input speech pattern;
- sampling pattern generating means (6) for generating a sampling pattern by extracting a predetermined number of sampling points obtained by equally dividing the input speech pattern obtained from said acoustic analyzing means (1),
- a time-frequency filter section (12), and
- similarity computing means (10),

characterized in that

- said time-frequency filter section (12) comprises
 - (a) means (57) for computing pattern $A_{j,k}$ of learning patterns $a_{m(j,k)}$ of word recognition or syllable category and collected by said sampling pattern generating means (6) based on the following equation:

$$A(j,k) = (1/n) \sum_{m=1}^n a_m(j,k),$$

where j denotes a number of feature vectors obtained by acoustic analyzing in said acoustic analyzing means (1), and k denotes a number of sampling points for dividing speech pattern in said sampling pattern generating means (6),

(b) means (59) for obtaining dictionary data $b_{1(j,k)}$ of a first axis using the average pattern $A_{j,k}$ in accordance with the following equation:

$$b_{1(j,k)} = A_{j-1,k-1} + A_{j-1,k} + A_{j-1,k+1} + A_{j,k-1} + 2 \cdot A_{j,k} + A_{j,k+1} + A_{j+1,k-1} + A_{j+1,k} + A_{j+1,k+1},$$

and

(c) means (61) for obtaining dictionary data $b_{2(j,k)}$ by differentiating average pattern $A_{j,k}$ along a second axis in accordance with the following equation:

$$b_{2(j,k)} = -A_{j-1,k-1} + A_{j-1,k+1} + \{(-A_{j,k-1} + A_{j,k+1})\} + \{(-A_{j+1,k-1} + A_{j+1,k+1})\}$$

and means (63) for obtaining dictionary data $b_{3(j,k)}$ by differentiating average pattern $A_{j,k}$ along a third frequency axis in accordance with the following equation:

$$b_{3(j,k)} = -A_{j-1,k-1} - A_{j-1,k} - A_{j-1,k+1} + A_{j+1,k-1} + A_{j+1,k} + A_{j+1,k+1} \text{ and}$$

- said similarity computing means (10) comprising means for computing similarities between the input speech pattern acoustically analyzed by said acoustic analyzing means (1), and the dictionaries of the first and the second axes generated by said time filter section (11).

3. An apparatus according to claim 1 or 2, characterized in that said time filter section (11), respectively said time-frequency filter section (12)

(d) obtains the dictionary of the second axis by re-processing of the dictionary data $b_{2(j,k)}$ of the second axis based on an equation expressed as follows:

$$B_{2(j,k)} = b_{2(j,k)} - (b_{2(j,k)} \cdot b_{1(j,k)})b_{1(j,k)}.$$

4. An apparatus according to claim 2, characterized in that said time-frequency filter section comprises at least three types of filters, for obtaining the average pattern from the sampling patterns, smoothing the average pattern along the time base and the frequency axis to generate the dictionary data of the first axis, differentiating the average pattern along the time base to generate the dictionary data of the second axis, and differentiating the average pattern along the frequency axis to generate dictionary data



of the third axis.

5. A speech recognition method comprising the steps of:

- acoustically analyzing an input speech to obtain an input speech pattern;
- generating a sampling pattern by extracting a predetermined number of sampling points obtained by equally dividing the speech pattern obtained from said acoustic analyzing step, and characterized by
- computing an average pattern $A_{j,k}$ of learning patterns $a_{m(j,k)}$ of word recognition of syllable category based on the following equation:

$$A(j,k) = (1/n) \sum_{m=1}^n a_m(j,k),$$

where j denotes a number of feature vectors obtained by acoustic analyzing in said acoustic analyzing step, and k denotes a number of sampling points for dividing speech pattern in said sampling pattern generating step,

- obtaining dictionary data $b_{1(j,k)}$ of a first axis using the average pattern $A_{j,k}$ in accordance with the following equation:

$$b_{1(j,k)} = A_{j-1,k-1} + A_{j-1,k} + A_{j-1,k+1} + A_{j,k-1} + 2 \cdot A_{j,k} + A_{j,k+1} + A_{j+1,k-1} + A_{j+1,k} + A_{j+1,k+1},$$

- thereafter, obtaining dictionary data $b_{2(j,k)}$ by differentiating average pattern $A_{j,k}$ along a second axis in accordance with the following equation:

$$b_{2(j,k)} = -A_{j-1,k-1} + A_{j-1,k+1} + \{(-A_{j,k-1} + A_{j,k+1})\} + \{(-A_{j+1,k-1} + A_{j+1,k+1})\} \text{ and}$$

- obtaining dictionary data $b_{3(j,k)}$ by differentiating average pattern $A_{j,k}$ along a third frequency axis in accordance with the following equation:

$$b_{3(j,k)} = -A_{j-1,k-1} - A_{j-1,k} - A_{j-1,k+1} + A_{j+1,k-1} + A_{j+1,k} + A_{j+1,k+1}, \text{ and}$$

- computing similarities between the acoustically analyzed input speech pattern and the dictionaries of the first and second axes.

6. A speech recognition method according to claim 5, characterized by a step of

- obtaining the dictionary of the second axis by re-processing of the dictionary data $b_{2(j,k)}$ of the second axis based on an equation expressed as follows:

$$B_{2(j,k)} = b_{2(j,k)} - (b_{2(j,k)} \cdot b_{1(j,k)})b_{1(j,k)}.$$

Patentansprüche

1. Vorrichtung zur Spracherkennung, welche folgendes umfaßt:

- eine akustische Analyseinrichtung (1) zur Analysierung einer Eingangssprache, um ein Eingangssprachmuster zu erhalten;
- eine Abfragemustergeneratoreinrichtung (6) zur Ableitung einer vorgegebenen Anzahl von Abfragepunkten, welche sich durch gleiche Teilung des aus der akustischen Analyseinrichtung erhaltenen Eingangssprachmusters ergeben,
- einen Zeitfilterabschnitt (11) und



- eine Ähnlichkeitsberechnungseinrichtung (10), dadurch gekennzeichnet, daß der Zeitfilterabschnitt (11) folgendes umfaßt:
 - (a) eine Einrichtung (21) zur Berechnung eines Durchschnittsmusters $A_{(j,k)}$ von n Lernmustern $a_{m(j,k)}$ der Worterkennungs- oder Silbenkategorie, welche von der Abfragemustergeneratoreinrichtung (6) auf Basis der folgenden Gleichung erfaßt wurden:

$$A_{(j,k)} = (1/n) \sum_{m=1}^n a_{m(j,k)}$$

wobei j eine Anzahl von Merkmalvektoren bedeutet, welche durch Analyse der akustischen Analyseinrichtung erhalten werden und k eine Anzahl von Abfragepunkten zur Teilung des Sprachmusters in der Abfragemustergeneratoreinrichtung (6) bedeutet,

(b) eine Einrichtung (23) zum Erhalt von Wörterbuchdaten $b_{1(j,k)}$ einer ersten Achse unter Verwendung der Durchschnittsmuster $A_{(j,k)}$ gemäß der nachstehenden Gleichung:

$$b_{1(j,k)} = A_{(j,k-1)} + 2 \cdot A_{(j,k)} + A_{(j,k+1)}$$

und

(c) eine Einrichtung (25) zum Erhalt von Wörterbuchdaten $b_{2(j,k)}$ durch Differentiation des Durchschnittsmusters $A_{(j,k)}$ entlang einer zweiten Achse gemäß der nachstehenden Gleichung:

$$b_{2(j,k)} = -A_{(j,k-1)} + A_{(j,k+1)}$$

und

- die Ähnlichkeitsberechnungseinrichtung (10) eine Einrichtung zur Berechnung von Ähnlichkeiten zwischen dem akustisch von der akustischen Analyseinrichtung (1) analysierten Eingangssprachmuster und den vom Zeitfilterabschnitt (11) generierten Wörterbüchern der ersten und zweiten Achse umfaßt.

2. Spracherkennungsvorrichtung, welche folgendes umfaßt:

- eine akustische Analyseinrichtung (1) zur Analysierung einer Eingangssprache, um ein Eingangssprachmuster zu erhalten;
- eine Abfragemustergeneratoreinrichtung (6) zur Generierung eines Abfragemusters durch Ableiten einer vorgegebenen Anzahl von Abfragepunkten auf Basis gleicher Teilung des von der akustischen Analyseinrichtung (1) erhaltenen Eingangssprachmusters,
- einen Zeitfrequenzfilterabschnitt (12), und
- eine Ähnlichkeitsberechnungseinrichtung (10), dadurch gekennzeichnet, daß der Zeitfrequenzfilterabschnitt (12) folgendes umfaßt:
 - (a) eine Einrichtung (57) zur Berechnung eines Durchschnittsmusters $A_{(j,k)}$ der Lernmuster $a_{m(j,k)}$ der Worterkennungs- oder Silbenkategorie, welche auf Basis der nachstehenden Gleichung von der Abfragemustergeneratoreinrichtung (6) erfaßt wurden:

$$A_{(j,k)} = (1/n) \sum_{m=1}^n a_{m(j,k)}$$

wobei j die Anzahl der durch die akustische Analyse in der akustischen Analyseinrichtung (1) erhaltenen Merkmalvektoren und k die Anzahl der Abfragepunkte zur Teilung des Sprachmusters in der Abfragemustergeneratoreinrichtung (6) kennzeichnet,

(b) eine Einrichtung (59) zum Erhalt von Wörterbuchdaten $b_{1(j,k)}$ einer ersten Achse unter Verwendung des Durchschnittsmusters $A_{(j,k)}$ gemäß der nachstehenden Gleichung:

$$b_{1(j,k)} = A_{(j-1,k-1)} + A_{(j-1,k)} + A_{(j-1,k+1)} + A_{(j,k-1)} + 2 \cdot A_{(j,k)} + A_{(j,k+1)} + A_{(j+1,k-1)} + A_{(j+1,k)} + A_{(j+1,k+1)}$$



und

(c) eine Einrichtung (61) zum Erhalt von Wörterbuchdaten $b_{2(j,k)}$ durch Differentiation des Durchschnittsmusters $A_{(j,k)}$ entlang einer zweiten Achse gemäß der nachstehenden Gleichung:

$$b_{2(j,k)} = -A_{(j-1,k-1)} + A_{(j-1,k+1)} + \{(-A_{(j,k-1)} + A_{(j,k+1)})\} + \{(-A_{(j+1,k-1)} + A_{(j+1,k+1)})\}$$

und eine Einrichtung (63) zum Erhalt von Wörterbuchdaten $b_{3(j,k)}$ durch Differentiation des Durchschnittsmusters entlang einer dritten Frequenzachse gemäß der nachstehenden Gleichung:

$$b_{3(j,k)} = -A_{(j-1,k-1)} - A_{(j-1,k)} - A_{(j-1,k+1)} + A_{(j+1,k-1)} + A_{(j+1,k)} + A_{(j+1,k+1)}$$

und

- die Ähnlichkeitsberechnungseinrichtung (10) eine Einrichtung zur Berechnung von Ähnlichkeiten zwischen dem akustisch von der akustischen Analyseinrichtung (1) analysierten Eingangssprachmuster und den vom Zeitfilterabschnitt (11) generierten Wörterbüchern der ersten und zweiten Achse umfaßt.

3. Vorrichtung gemäß Anspruch 1 oder 2, dadurch gekennzeichnet, daß der Zeitfilterabschnitt (11) bzw. der Zeitfrequenzfilterabschnitt (12)

- (d) das Wörterbuch der zweiten Achse durch Wiederverarbeitung der Wörterbuchdaten $b_{2(j,k)}$ der zweiten Achse auf Basis einer wird folgt dargestellten Gleichung erstellt:

$$B_{2(j,k)} = b_{2(j,k)} - (b_{2(j,k)} \cdot b_{1(j,k)})b_{1(j,k)}$$

4. Vorrichtung gemäß Anspruch 2, dadurch gekennzeichnet, daß der Zeitfrequenzfilterabschnitt mindestens drei Typen von Filtern zum Erhalt des Durchschnittsmusters, zum Glätten des Durchschnittsmusters entlang der Zeitbasis und der Frequenzachse zur Generierung der Wörterbuchdaten der ersten Achse, zur Differentiation des Durchschnittsmusters entlang der Zeitbasis zur Generierung der Wörterbuchdaten der zweiten Achse und zur Differentiation des Durchschnittsmusters entlang der Frequenzachse zur Generierung der Wörterbuchdaten der dritten Achse umfaßt.

5. Spracherkennungsverfahren, welches folgende Schritte umfaßt:

- akustische Analysierung einer Eingangssprache zum Erhalt eines Eingangssprachmusters;
- Generierung eines Abfragemusters durch Ableitung einer vorgegebenen Anzahl von Abfragepunkten, welche sich durch gleiche Teilung des aus der akustischen Analyseinrichtung erhaltenen Eingangssprachmusters ergeben, und gekennzeichnet durch
- Berechnung eines Durchschnittsmusters $A_{(j,k)}$ von Lernmustern $a_{m(j,k)}$ der Worterkennungs- oder Silbenkategorie auf Basis der folgenden Gleichung:

$$A_{(j,k)} = (1/n) \sum_{m=1}^n a_{m(j,k)}$$

wobei j eine Anzahl von Merkmalvektoren bedeutet, welche durch Analyse der akustischen Analyseinrichtung erhalten werden und k eine Anzahl von Abfragepunkten zur Teilung des Sprachmusters im Schritt der Abfragemustergenerierung bedeutet,

- Beschaffung von Wörterbuchdaten $b_{1(j,k)}$ einer ersten Achse unter Verwendung des Durchschnittsmusters $A_{(j,k)}$ gemäß der nachstehenden Gleichung:

$$b_{1(j,k)} = A_{(j-1,k-1)} + A_{(j-1,k)} + A_{(j-1,k+1)} + A_{(j,k-1)} + 2 \cdot A_{(j,k)} + A_{(j,k+1)} + A_{(j+1,k-1)} + A_{(j+1,k)} + A_{(j+1,k+1)}$$

- anschließende Beschaffung von Wörterbuchdaten $b_{2(j,k)}$ durch Differentiation des Durchschnittsmusters $A_{(j,k)}$ entlang einer zweiten Achse gemäß der nachstehenden Gleichung:



$$b_{2(j,k)} = -A_{(j-1,k-1)} + A_{(j-1,k+1)} + \{(-A_{(j,k-1)} + A_{(j,k)} + A_{(j,k+1)})\} + \{(-A_{(j+1,k-1)} + A_{(j+1,k+1)})\}$$

und

- Beschaffung von Wörterbuchdaten $b_{3(j,k)}$ durch Differentiation des Durchschnittsmusters $A_{(j,k)}$ entlang einer dritten Frequenzachse gemäß der nachstehenden Gleichung:

$$b_{3(j,k)} = -A_{(j-1,k-1)} + A_{(j-1,k)} - A_{(j-1,k+1)} + A_{(j+1,k-1)} + A_{(j+1,k)} + A_{(j+1,k+1)}$$

und

- Berechnung von Ähnlichkeiten zwischen dem akustisch analysierten Eingangssprachmuster und den Wörterbüchern der ersten und zweiten Achse.

6. Verfahren zur Spracherkennung gemäß Anspruch 5, gekennzeichnet durch einen Schritt der Beschaffung des Wörterbuchs der zweiten Achse durch Wiederverarbeitung der Wörterbuchdaten $b_{2(j,k)}$ der zweiten Achse auf Basis der nachstehend ausgedrückten Gleichung:

$$B_{2(j,k)} = b_{2(j,k)} - (b_{2(j,k)} \cdot b_{1(j,k)})b_{1(j,k)}$$

Revendications

1. Appareil de reconnaissance de la parole comprenant :

un moyen d'analyse acoustique (1) pour analyser une parole d'entrée afin d'obtenir un motif de parole d'entrée ;

un moyen de génération de motif d'échantillonnage (6) pour générer un motif d'échantillonnage en extrayant un nombre prédéterminé de points d'échantillonnage obtenus en divisant de façon égale le motif de parole d'entrée obtenu depuis ledit moyen d'analyse acoustique (1) ;

une section de filtre temporel (11) ; et

un moyen de calcul de similarité (10),

caractérisé en ce que :

ladite section de filtre temporel (11) comprend :

(a) un moyen (21) pour calculer un motif moyen $A_{(j,k)}$ des motifs d'apprentissage $a_{m(j,k)}$ d'une catégorie de reconnaissance de mot ou de syllabe, lesquels sont collectés par ledit moyen de génération de motif d'échantillonnage (6) sur la base de l'équation suivante :

$$A_{(j,k)} = (1/n) \sum_{m=1}^n a_{m(j,k)}$$

où j indique un nombre de vecteurs caractéristiques obtenus au moyen d'une analyse acoustique réalisée dans ledit moyen d'analyse acoustique (1) et k indique un nombre de points d'échantillonnages pour diviser un motif de parole dans ledit moyen de génération de motif d'échantillonnage (6) ;

(b) un moyen (23) pour obtenir des données de dictionnaire $b_{1(j,k)}$ d'un premier axe en utilisant le motif moyen $A_{(j,k)}$ conformément à l'équation suivante :

$$b_{1(j,k)} = A_{(j,k-1)} + 2A_{(j,k)} + A_{(j,k+1)} ; \text{ et}$$

(c) un moyen (25) pour obtenir des données de dictionnaire $b_{2(j,k)}$ en différenciant un motif moyen $A_{(j,k)}$ suivant un second axe conformément à l'équation suivante :

$$b_{2(j,k)} = -A_{(j,k-1)} + A_{(j,k+1)} ; \text{ et}$$

ledit moyen de calcul de similarité (10) comprenant un moyen pour calculer des similarités entre le motif de parole d'entrée analysé acoustiquement par ledit moyen d'analyse acoustique (1) et les dictionnaires des premier et second axes générés par ladite section de filtre temporel (11).



2. Appareil de reconnaissance de la parole comprenant :

un moyen d'analyse acoustique (1) pour analyser une parole d'entrée afin d'obtenir un motif de parole d'entrée :

un moyen de génération de motif d'échantillonnage (6) pour générer un motif d'échantillonnage en extrayant un nombre prédéterminé de points d'échantillonnage obtenus en divisant de façon égale le motif de parole d'entrée obtenu depuis ledit moyen d'analyse acoustique (1) ;

une section de filtre temps-fréquence (12) ; et

un moyen de calcul de similarité (10),

caractérisé en ce que :

ladite section de filtre temps-fréquence (12) comprend :

(a) un moyen (57) pour calculer un motif-moyen $A_{(j,k)}$ de motifs d'apprentissage $a_{m(j,k)}$ d'une catégorie de reconnaissance de mot ou de syllabe, lesquels sont collectés par ledit moyen de génération de motif d'échantillonnage (6) sur la base de l'équation suivante :

$$A_{(j,k)} = (1/n) \sum_{m=1}^n a_{m(j,k)}$$

où j indique un nombre de vecteurs caractéristiques obtenus au moyen d'une analyse acoustique réalisée dans ledit moyen d'analyse acoustique (1) et k indique un nombre de points d'échantillonnage pour diviser un motif de parole dans ledit moyen de génération de motif d'échantillonnage (6) ;

(b) un moyen (59) pour obtenir des données de dictionnaire $b1_{(j,k)}$ d'un premier axe en utilisant le motif moyen $A_{(j,k)}$ conformément à l'équation suivante :

$$b1_{(j,k)} = A_{(j-1,k-1)} + A_{(j-1,k)} + A_{(j-1,k+1)} + A_{(j,k-1)} + 2 \cdot A_{(j,k)} + A_{(j,k+1)} + A_{(j+1,k-1)} + A_{(j+1,k)} + A_{(j+1,k+1)}$$

et

(c) un moyen (61) pour obtenir des données de dictionnaire $b2_{(j,k)}$ en différenciant un motif moyen $A_{(j,k)}$ suivant un second axe en relation avec l'équation suivante :

$$b2_{(j,k)} = -A_{(j-1,k-1)} + A_{(j-1,k+1)} + \{A_{(j,k-1)} + A_{(j,k+1)}\} + \{-A_{(j+1,k-1)} + A_{(j+1,k+1)}\}$$

et un moyen (63) pour obtenir des données de dictionnaire $b3_{(j,k)}$ en différenciant un motif moyen $A_{(j,k)}$ suivant un troisième axe des fréquences conformément à l'équation suivante :

$$b3_{(j,k)} = -A_{(j-1,k-1)} - A_{(j-1,k)} - A_{(j-1,k+1)} + A_{(j+1,k-1)} + A_{(j+1,k)} + A_{(j+1,k+1)} ; \text{ et}$$

ledit moyen de calcul de similarité comprenant un moyen (10) pour calculer des similarités entre le motif de parole d'entrée analysé acoustiquement par ledit moyen d'analyse acoustique (1) et les dictionnaires des premier et second axes générés par ladite section de filtre temporel (11).

3. Appareil selon la revendication 1 ou 2, caractérisé en ce que respectivement ladite section de filtre temporel (11), ladite section de filtre temps-fréquence (12) (d) obtiennent le dictionnaire du second axe en retraçant les données de dictionnaire $b2_{(j,k)}$ du second axe sur la base d'une équation exprimée comme suit :

$$B2_{(j,k)} = b2_{(j,k)} - (b2_{(j,k)} \cdot b1_{(j,k)})b1_{(j,k)}.$$

4. Appareil selon la revendication 1 ou 2, caractérisé en ce que ladite section de filtre temps-fréquence comprend au moins trois types de filtre pour obtenir le motif moyen à partir des motifs d'échantillonnage, pour lisser le motif moyen suivant la base de temps et l'axe des fréquences afin de générer les données de dictionnaire du premier axe, pour différencier le motif moyen suivant la base de temps pour générer les données de dictionnaire du second axe et pour différencier le motif moyen suivant l'axe des fréquences afin de générer des données de dictionnaire du troisième axe.



5. Procédé de reconnaissance de la parole comprenant les étapes de :

analyse acoustique d'une parole d'entrée afin d'obtenir un motif de parole d'entrée ;

génération d'un motif d'échantillonnage en extrayant un nombre prédéterminé de points d'échantillonnage obtenus en divisant de façon égale le motif de parole obtenu lors de ladite étape d'analyse acoustique, et

caractérisé par :

le calcul d'un motif moyen $A_{(j,k)}$ de motifs d'apprentissage $a_{m(j,k)}$ d'une catégorie de reconnaissance de mot ou de syllabe sur la base de l'équation suivante :

$$A_{(j,k)} = (1/n) \sum_{m=1}^n a_{m(j,k)}$$

où j indique un nombre de vecteurs caractéristiques obtenus au moyen d'une analyse acoustique réalisée lors de ladite étape d'analyse acoustique et k indique un nombre de points d'échantillonnage pour diviser le motif de parole lors de ladite étape de génération de motifs d'échantillonnage.

l'obtention de données de dictionnaire $b1_{(j,k)}$ d'un premier axe en utilisant le motif moyen $A_{(j,k)}$ conformément à l'équation suivante :

$$b1_{(j,k)} = A_{(j-1,k-1)} + A_{(j-1,k)} + A_{(j-1,k+1)} + A_{(j,k-1)} + 2 \cdot A_{(j,k)} + A_{(j,k+1)} + A_{(j+1,k-1)} + A_{(j+1,k)} + A_{(j+1,k+1)}$$

puis l'obtention de données de dictionnaire $b2_{(j,k)}$ en différenciant un motif moyen $A_{(j,k)}$ suivant un second axe conformément à l'équation suivante :

$$b2_{(j,k)} = -A_{(j-1,k-1)} + A_{(j-1,k+1)} + \{-A_{(j,k-1)} + A_{(j,k+1)}\} + \{-A_{(j+1,k-1)} + A_{(j+1,k+1)}\}; \text{ et}$$

l'obtention de données de dictionnaire $b3_{(j,k)}$ en différenciant un motif moyen $A_{(j,k)}$ suivant un troisième axe des fréquences conformément à l'équation suivante :

$$b3_{(j,k)} = -A_{(j-1,k-1)} - A_{(j-1,k)} - A_{(j-1,k+1)} + A_{(j+1,k-1)} + A_{(j+1,k)} + A_{(j+1,k+1)}; \text{ et}$$

le calcul de similarités entre le motif de parole d'entrée analysé acoustiquement et les dictionnaires des premier et second axes.

6. Procédé de reconnaissance de la parole selon la revendication 5, caractérisé par une étape d'obtention du dictionnaire du second axe en retraitant les données de dictionnaire $b2_{(j,k)}$ du second axe sur la base d'une équation exprimée comme suit :

$$B2_{(j,k)} = b2_{(j,k)} - (b2_{(j,k)} \cdot b1_{(j,k)})b1_{(j,k)}$$