Office de la Propriété Intellectuelle du Canada
Un organisme d'Industrie Canada

Canadian Intellectual Property Office
An agency of Industry Canada

CA 2176103 C 2002/07/16

(11)(21) 2 176 103

(12) BREVET CANADIEN
CANADIAN PATENT

(13) C

(22) Date de dépôt/Filing Date: 1996/05/08

(41) Mise à la disp. pub./Open to Public Insp.: 1996/11/13

(45) Date de délivrance/Issue Date: 2002/07/16

(30) Priorité/Priority: 1995/05/12 (114628/1995) JP

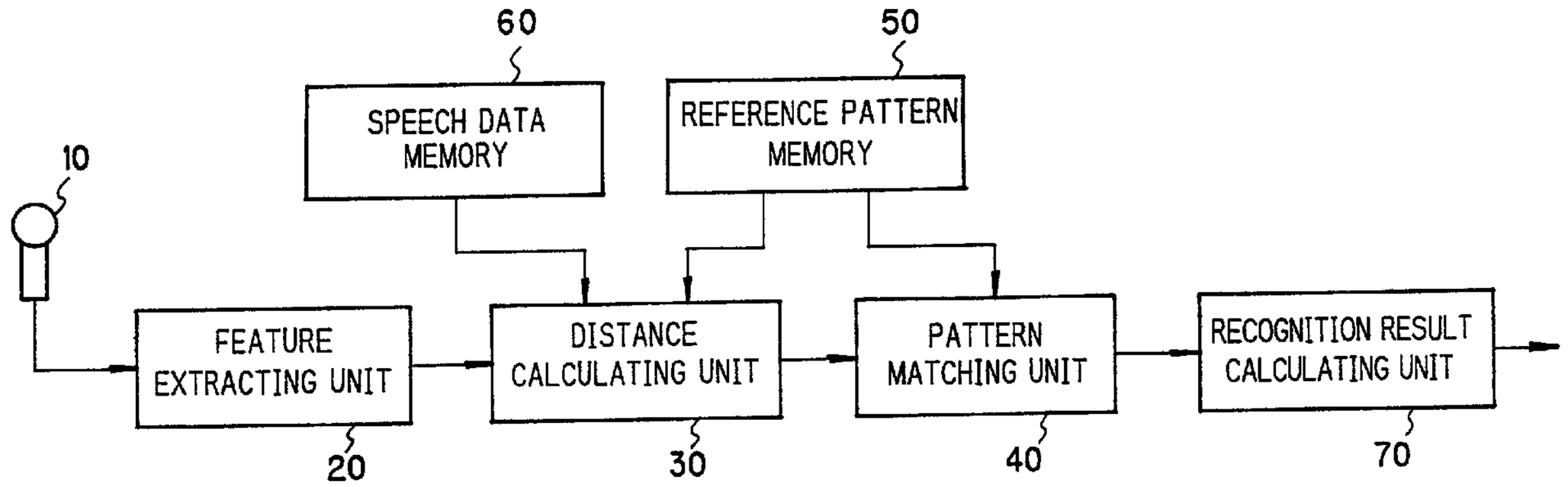
(51) Cl.Int.^{6}/Int.Cl.^{6} G10L 5/06

(72) Inventeur/Inventor:
Iso, Ken-Ichi, JP

(73) Propriétaire/Owner:
NEC CORPORATION, JP

(74) Agent: SMART & BIGGAR

(54) Titre : DISPOSITIF DE RECONNAISSANCE DE LA PAROLE
(54) Title: SPEECH RECOGNIZER



(57) Abrégé/Abstract:
In a speech data memory 60 speech data and symbol trains thereof are stored, and in a reference pattern memory 50 sets each of a given partial symbol train of a word presented for recognition and an index of speech data with the symbol train thereof containing the partial symbol train in the speech data memory 60 are stored. Speech recognition operation is executed on the basis of the read out data from the speech data memory 60 and the reference pattern memory 50.

ABSTRACT OF THE DISCLOSURE

In a speech data memory 60 speech data and symbol trains thereof are stored, and in a reference pattern memory 50 sets each of a given partial symbol train of a word presented for recognition and an index of speech data with the symbol train thereof containing the partial symbol train in the speech data memory 60 are stored. Speech recognition operation is executed on the basis of the read out data from the speech data memory 60 and the reference pattern memory 50.

SPEECH RECOGNIZER

BACKGROUND OF THE INVENTION

The present invention relates to improvements in speech recognizer reference patterns.

As a method of realizing speech recognizer which are capable of ready alteration of vocabularies presented for recognition, a method which uses context-dependent phone reference patterns has been extensively utilized. In this method, a reference pattern of a given word presented for recognition can be produced by connecting context-dependent phone reference patterns of corresponding phone expressions. A context-dependent phone reference pattern of each phone (which is designated as a set of three elements, i.e., a preceding phone, the subject phone and a succeeding phone), is produced by making segmentation of a number of pieces of speech data collected for training in phone units, and averaging selectedly collected phones that are in accord inclusive of the preceding and succeeding phones. Such method is described in, for instance, Kai-Fu Lee, IEEE Transactions on Acoustics, Speech, and Signal Processing, 1990, Vol. 38, No. 4, pp. 599-609. In this method, a speech data base that is used for producing a context-dependent phone reference pattern, is provided separately from the speech recognizer, and it is used only when

producing the reference pattern.

Fig. 5 shows a case when producing a context-dependent phone reference pattern from speech data corresponding to a phone train "WXYZ" in the speech data base. Referring to Fig. 5, "X (W, Y)" represents a context-dependent phone reference pattern of the phone X with the preceding phone W and the succeeding phone Y. When identical context-dependent phones appear in different parts of speech data, their average is used as the reference pattern.

In the case where a phone reference pattern is produced by taking the contexts of the preceding and succeeding one phone into considerations by the prior art method, including the case shown in Fig. 5, even if there exist speech data in the speech data base that contain the same context as the phone in a word presented for recognition inclusive of the preceding and succeeding two phones, are not utilized at all for recognition. In other words, in the prior art method a reference pattern is produced on the basis of phone contexts which are fixed when the training is made. In addition, the phone contexts to be considered are often of one preceding phone and one succeeding phone in order to avoid explosive increase of the number of combinations of phones. For this reason, the collected speech data bases are not effectively utilized, and it has been

impossible to improve the accuracy of recognition.

SUMMARY OF THE INVENTION

An object of the present invention is therefore to provide a speech recognizer capable of improving speech recognition performance through improvement in the speech reference pattern accuracy.

According to the present invention, there is provided a speech recognizer comprising: a speech data memory in which speech data and symbol trains thereof are stored; a reference pattern memory in which are stored sets each of a given partial symbol train of a word presented for recognition and an index of speech data with the expression thereof containing the partial symbol train in the speech data memory; a distance calculating unit for calculating a distance between the partial symbol train stored in the reference pattern memory and a given input speech section; and a pattern matching unit for selecting, among possible partial symbol trains as divisions of the symbol train of a word presented for recognition, a partial symbol train which minimizes the sum of distances of input speech sections over the entire input speech interval, and outputting the distance sum data at this time as data representing the distance between the input speech and the word presented for recognition.

In a specific embodiment, the distance to be

calculated in the distance calculating unit is the distance between a given section corresponding to the partial train of symbol train expression of speech data stored in the speech data memory and the given input speech section.

According to a concrete aspect of the present invention, there is provided a speech recognizer comprising: a feature extracting unit for analyzing an input speech to extract a feature vector of the input speech; a speech data memory in which speech data and symbol trains thereof are stored; a reference pattern memory in which are stored sets each of a given partial symbol train of a word presented for recognition and an index of speech data with the expression thereof containing the partial symbol train in the speech data memory; a distance calculating unit for reading out speech data corresponding to a partial train stored in the reference pattern memory from the speech data memory and calculating a distance between the corresponding section and a given section of the input speech; a pattern matching unit for deriving, with resect to each word presented for recognition, a division of the subject word interval which minimizes the sum of distances of the input speech sections over the entire word interval; and a recognition result calculating unit for outputting as a recognition result a word presented for recognition, which gives

the minimum one of the distances between the input speech data output of the pattern matching unit and all the words presented for recognition.

Other objects and features will clarified from the following description with reference to attached drawings.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a block diagram showing the basic construction of this embodiment of the speech recognizer;

Figs. 2 to 4 are drawings for explaining operation of the embodiment of the speech recognizer of Fig. 1; and

Fig. 5 is a drawing for explaining a prior art speech recognizer in a case when producing a context-dependent phone reference pattern from speech data corresponding to a phone train "WXYZ" in the speech data base.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT

Now, an embodiment of the speech recognizer according to the present invention will be described with reference to the drawings. Fig. 1 is a block diagram showing the basic construction of this embodiment of the speech recognizer. Referring to Fig. 1, a feature extracting unit 20 analyzes an input speech inputted from a microphone 10, extracts a feature vector and supplies the extracted feature vector train to a distance calculating unit 30. The

distance calculating unit 30 reads out speech data corresponding to a partial train stored in a reference pattern memory 50 from a speech data memory 60 and calculates the distance between the corresponding section and a given section of the input speech. A pattern matching unit 40 derives, with respect to each word presented for recognition, a division of the subject word interval which minimizes the sum of distances of the input speech sections over the entire word interval. A recognition result calculating unit 70 outputs as the recognition result a word presented for recognition, which gives the minimum one of the distances between the input speech data output of the pattern matching unit 40 and all the words presented for recognition.

The operation of the embodiment of the speech recognizer will now be described in detail with reference to Figs. 2 to 4 in addition to Fig. 1. According to the present invention, a number of pieces of speech data and speech context phone expressions thereof are prepared and stored in the speech data memory 60. A reference pattern of a word to be recognized is produced as follows:

(1) Partial trains of phone symbols of a word presented for recognition, are prepared such that they have given lengths (without overlap or missing), as shown in Fig. 2.

(2) Then, as shown in Fig. 3, all speech data portions with phones containing a partial symbol train among the speech data in a speech data base are all picked up.

A combination of possible partial symbol trains as divisions of a word presented for recognition and corresponding speech data portions, is stored as a reference pattern of the word presented for recognition in the reference pattern memory 50. The distance between the input speech data in the pattern matching unit 40 and each word presented for recognition, is defined as follows.

(a) A specific division of the word presented for recognition is selected from the reference pattern memory 50. The phone symbol train of the word presented for recognition is denoted by W, and the division of the symbol train into N partial symbol trains is denoted by $\omega(1),\omega(2)...,\omega(N)$.

(b) From the speech data stored in the speech data memory 6, with the symbol train containing partial symbol trains each defined by a selected division, a given segment of the speech is derived as an acoustical segment of that partial symbol train (Fig. 3).

Among the speech data with the symbol train thereof containing partial symbol trains $\omega(n)$, a k-th speech data portion is denoted by $A[\omega(n),k]$,

(k = 1 to K(n)). The acoustical segment in a section of the speech data from time instant σ till time instant τ, is denoted by A[ω(n),k,σ,τ].

(c) As shown in Fig. 4, distance between that obtained by connecting acoustical segments and the input speech, is calculated in accordance with the sequence of partial symbol trains in the pattern matching unit 40 by DP matching or the like.

Denoting the acoustical segment in a section of the input speech from the time instant s till the time instant t by X[s,t], the distance D is given by the following formula (1).

$$D = \sum^{N}_{n=1} d(X[s(n),t(n)],A[\omega(n),k,\sigma(n),\tau n)] \quad \ldots (1)$$

where d is the acoustic distance which is calculated in the distance calculating unit 30.

For continuity, it is necessary to meet a condition given as:

$$s(1) = 1$$
$$s(2) = t(1)+1$$
$$s(3) = t(2)+1$$
$$\cdot$$
$$s(N) = t(n)+1$$
$$t(N) = T \quad \ldots (2)$$

where T is the time interval of the input speech.

(d) By making the division of the symbol train into all possible partial symbol trains in step (c) and obtaining all possible sections (s,t,σ,τ) in the

step (b), a partial symbol train which gives a minimum distance is selected in a step (c), and this distance is made as the distance between the input speech and the word presented for recognition.

The recognition result calculating unit 70 provides as the speech recognition result a word presented for recognition giving the minimum distance from the input speech in the step (d) among a plurality of words presented for recognition. In the above way, the speech recognizer is operated. It is possible of course to use the recognition results obtainable with the speech recognizer according to the present invention as the input signal to a unit (not shown) connected to the output side such as a data processing unit, a communication unit, a control unit, etc.

According to the present invention, a set of three phones, i.e., one preceding phone, the subject phone and one succeeding element, is by no means limitative, but it is possible to utilize all speech data portions of words presented for recognition with identical phone symbol train and context (unlike the fixed preceding and succeeding phones in the prior art method) that are obtained through retrieval of the speech data in the speech data base when speech recognition is made. As for the production of acoustical segments, what is most identical with the input speech is automatically

determined at the time of the recognition. It is thus possible to improve the accuracy of reference patterns, thus providing improved recognition performance.

Changes in construction will occur to those skilled in the art and various apparently different modifications and embodiments may be made without departing from the scope of the invention. The matter set forth in the foregoing description and accompanying drawings is offered by way of illustration only. It is therefore intended that the foregoing description be regarded as illustrative rather than limiting.

74479-17

CLAIMS:

1. A speech recognizer comprising:

a speech data memory in which speech data and symbol trains thereof are stored;

a reference pattern memory in which are stored sets each of a given partial symbol train of a word presented for recognition and an index of speech data with the expression thereof containing the partial symbol train in the speech data memory;

a distance calculating unit for calculating a distance between the partial symbol train stored in the reference pattern memory and a given input speech section; and

a pattern matching unit for selecting, among possible partial symbol trains as divisions of the symbol train of a word presented for recognition, a partial symbol train which minimizes the sum of distances of input speech sections over the entire input speech interval, and outputting the distance sum data at this time as data representing the distance between the input speech and the word presented for recognition.

2. The speech recognizer according to claim 1, wherein the distance to be calculated in the distance calculating unit is the distance between a given section corresponding to the partial train of symbol train expression of speech data stored in the speech data memory and the given input speech section.

3. A speech recognizer comprising:

a feature extracting unit for analyzing an input speech to extract a feature vector of the input speech;

a speech data memory in which speech data and symbol trains thereof are stored;

a reference pattern memory in which are stored sets each of a given partial symbol train of a word presented for recognition and an index of speech data with the expression thereof containing the partial symbol train in the speech data memory;

a distance calculating unit for reading out speech data corresponding to a partial train stored in the reference pattern memory from the speech data memory and calculating a distance between the corresponding section and a given section of the input speech;

a pattern matching unit for deriving, with respect to each word presented for recognition, a division of the subject word interval which minimizes the sum of distances of the input speech sections over the entire word interval; and

a recognition result calculating unit for outputting as a recognition result a word presented for recognition, which gives the minimum one of the distances between the input speech data output of the pattern matching unit and all the words presented for recognition.

for recognition, which gives the minimum one of the distances between the input speech data output of the pattern matching unit and all the words presented for recognition.

FIG. 1

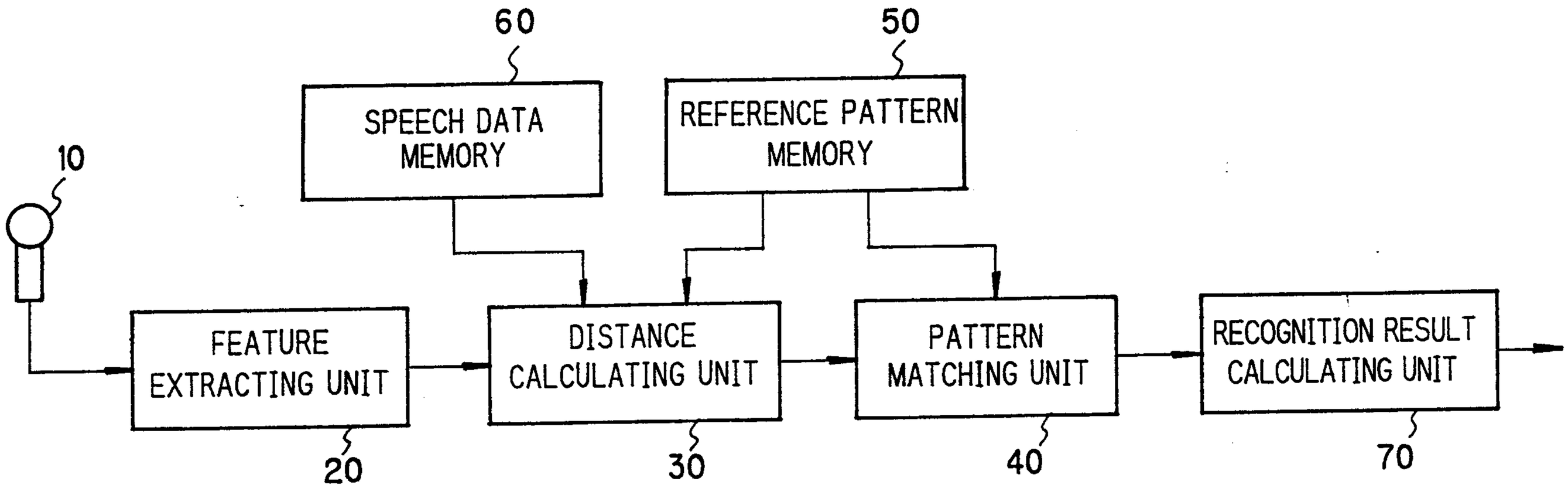
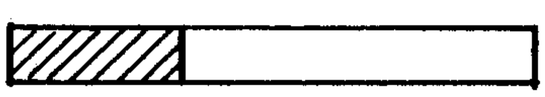
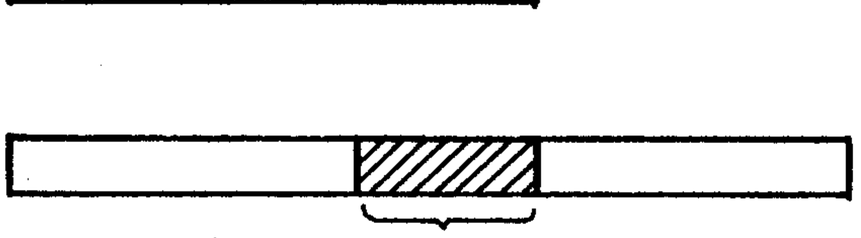


FIG. 2

PHONE SYMBOL OF WORD PRESENTED FOR RECOGNITION	DIVISION FORM
A B C D	(A, B, C, D)
	(A, BC, D)
	(A, B, CD)
	(A, BCD)
	(AB, C, D)
	(AB, CD)
	(ABC, D)
	(ABCD)

FIG. 3

SPEECH DATA INCLUDING TRAIN ELEMENT AB

PHONE SYMBOL	SPEECH DATA
A B P Q R	
S T A B U V	
⋮	⋮

Patent Agents
Smart & [illegible]

FIG. 4

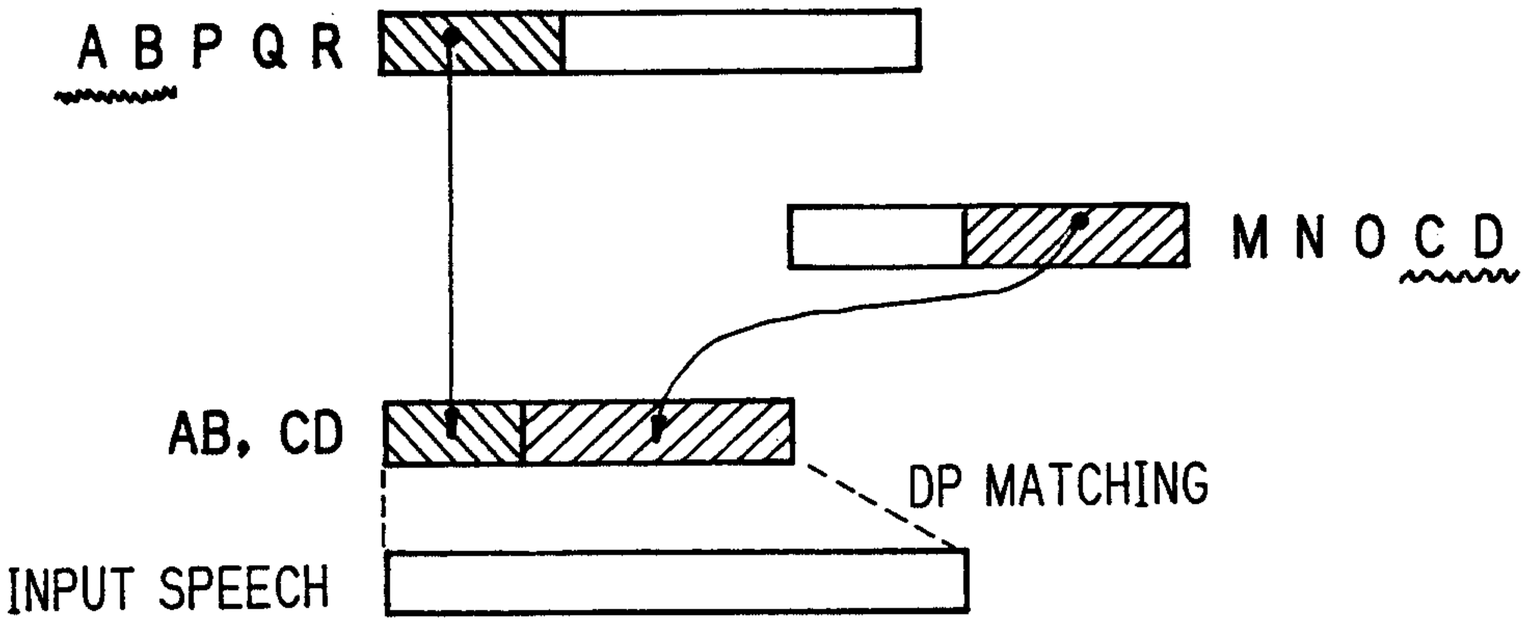
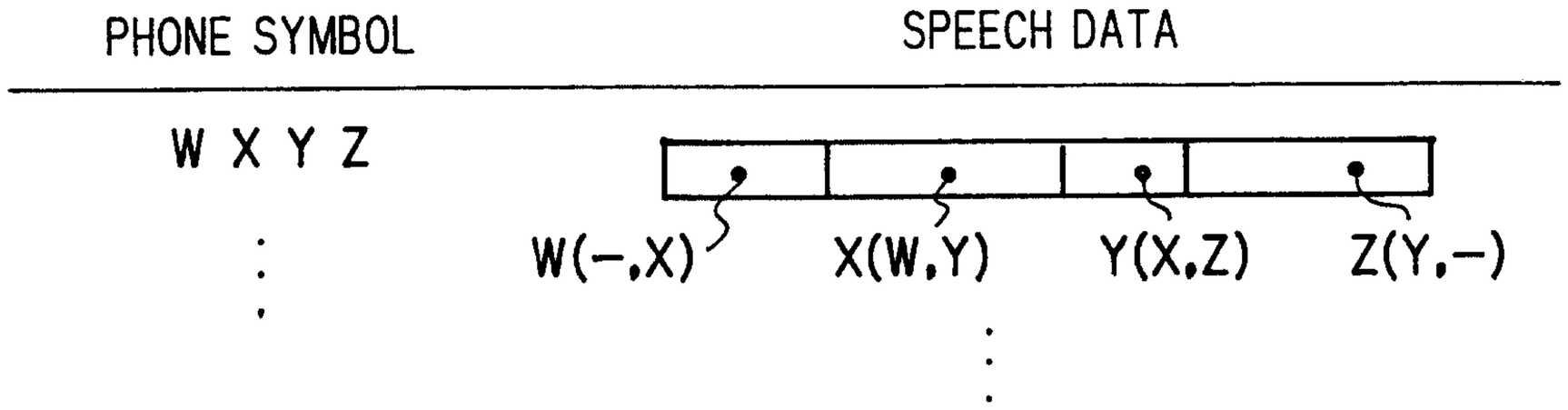
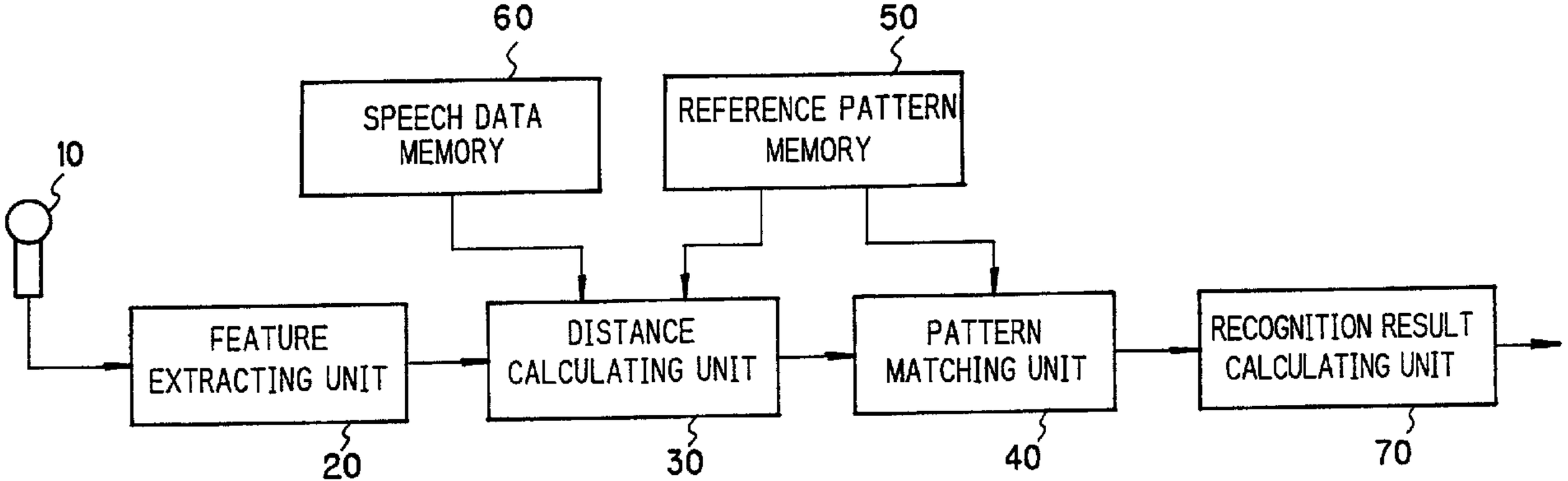


FIG. 5

(PRIOR ART)




60
50
SPEECH DATA MEMORY
REFERENCE PATTERN MEMORY
10
FEATURE EXTRACTING UNIT
DISTANCE CALCULATING UNIT
PATTERN MATCHING UNIT
RECOGNITION RESULT CALCULATING UNIT
20
30
40
70