



EUROPEAN PATENT SPECIFICATION

Date of publication of patent specification :
01.02.95 Bulletin 95/05

Int. Cl.⁶ : **G10L 5/06, G10L 3/00**

Application number : **90120020.4**

Date of filing : **18.10.90**

Speech recognition system.

Priority : **19.10.89 JP 272846/89**

Date of publication of application :
24.04.91 Bulletin 91/17

Publication of the grant of the patent :
01.02.95 Bulletin 95/05

Designated Contracting States :
DE FR GB

References cited :
EP-A- 0 286 035
GB-A- 2 183 880
US-A- 4 473 904

Proprietor : **MATSUSHITA ELECTRIC
INDUSTRIAL CO., LTD.**
1006, Oaza Kadoma
Kadoma-shi, Osaka-fu, 571 (JP)

Inventor : **Morii, Shuji**
2-11-16, Jingumae
Shibuya-ku, Tokyo (JP)
Inventor : **Hiraoka, Shoji**
7-6-5, Tsuchihashi
Miyamae-ku, Kawasaki (JP)

Representative : **Pellmann, Hans-Bernd,**
Dipl.-Ing. et al
Patentanwaltsbüro
Tiedtke-Bühling-Kinne & Partner
Bavariaring 4
D-80336 München (DE)

EP 0 423 800 B1

Note : Within nine months from the publication of the mention of the grant of the European patent, any person may give notice to the European Patent Office of opposition to the European patent granted. Notice of opposition shall be filed in a written reasoned statement. It shall not be deemed to have been filed until the opposition fee has been paid (Art. 99(1) European patent convention).

Description**BACKGROUND OF THE INVENTION**

5 The present invention relates generally to speech recognition systems, and more particularly to such a speech recognition system for operation of an apparatus through speech recognition.

As a system for operating an apparatus placed at a remote side through a speech is known a banking service system as disclosed in "Electronic Technique, Vol. 25, No. 1, P 43 to 46, for example. As illustrated in Fig. 1, this system is arranged such that a speech inputted through a telephone set 51 or the like is transmitted
10 through a public line 52 or the like up to a speech recognition apparatus 53 at the central processing equipment side where the inputted speech is recogized and the recognition result is supplied to a task control apparatus. Another approach involves, as illustrated in Fig. 2, recognizing an inputted speech with a speech recognition apparatus 62 incorporated into a user side terminal unit 61 and coding the recognition result with a coder 63 built in the same terminal unit 61, the coded signal being supplied through a transmission line 64 to a decoder
15 65 and then supplied to a task control apparatus 66 placed at the central processing equipment side.

There are problems which arise with such types of speech recognition systems, however, in that the former is affected by the transmission property of the telephone line 52 such as the frequency range limitation of the user's speech and further affected by the line noises introduced during the transmission so as to generally reduce the recognition performance of the speech recognition apparatus 53, and the latter prevents the reduction of the speech recognition rate due to the transmission because of no transmission of the speech itself
20 through the telephone line 52 or the like, but having extreme difficulty to perform change of the vocabulary to be recognized and change of the operating procedure at the central processing equipment side to result in lack of flexibility concurrently with increasing the cost of the terminal side apparatus because the speech recognition apparatus 62 is disposed at the user's terminal unit 61 side.

25 Speech recognition systems employing phoneme estimation are known, for example, from EP-A-0 286 035.

SUMMARY OF THE INVENTION

30 It is therefore an object of the present invention to provide a speech recognition system which is capable of improving the speech recognition rate by preventing the affection of the line noises and others and further freely setting the vocabulary to be recognized and the operating procedure at the central processing equipment side to provide flexibility.

In accordance with the present invention, there is provided a speech recognition system comprising: means
35 responsive to an input of a speech from an external device for recognizing phonemes or syllables constituting the inputted speech to output them as a symbol sequence; means coupled to the extracting means for coding the symbol train and outputting the coded symbol train; means for transmitting the coded symbol train; means coupled through the transmitting means to the coding means for decoding the coded symbol train to restoring it to the original symbol train; and means responsive to the decoded symbol train from the decoding means
40 for recognizing a word or a sentence on the basis of the decoded symbol train.

BRIEF DESCRIPTION OF THE DRAWINGS

45 The present invention will be described in further detail with reference to the accompanying drawings, in which:

Figs. 1 and 2 are block diagrams showing conventional speech recognition systems;

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

50 Fig. 4 is an illustration for describing one example of a word dictionary to be used in the Fig. 3 speech recognition system;

Fig. 5 is a graphic illustration for describing an inputted speech signal and a phoneme recognition; and

Fig. 6 is a block diagram showing a speech recognition system according to a second embodiment of this invention.

DETAILED DESCRIPTION OF THE INVENTION

Referring now to Fig. 3, there is illustrated a speech recognition system according to a first embodiment of the present invention. Although the speech recognition is generally performed by using words, syllables,

phonemes and others as basic units for recognition, in this invention, the syllables, phonemes or the like, which are units to allow the expression of a sentence and a word, are used as the basic units. The embodiment will be described in terms of one case of using phonemes which are minimum and indispensable phonological units for description of a given speech.

In Fig. 3, the speech recognition system illustrated at numeral 1 comprises a phoneme recognizing section 2 which recognizes an inputted speech and converts it into a phoneme-symbol sequence, each phoneme being a basic unit of the inputted speech. The phoneme-symbol train is supplied to a coder 4 to be coded. The coded phoneme-symbol train is supplied through a transmission line 5 to a decoder 6 which in turn decodes the coded phoneme-symbol train. The decoded phoneme-symbol train is supplied to a word and sentence recognizing section 7 for recognizing a word and a sentence making up the speech. The word and sentence recognizing section 7 is also coupled to a word dictionary 8 storing a phoneme notation. The word and sentence recognizing section 7 performs the matching between the phoneme-symbol train outputted from the decoder 6 and the phoneme notation stored in the word dictionary 8. The output of the word and sentence recognizing section 7 is supplied to a task control apparatus 2 which performs applications of the banking service, information retrieval and others. In this embodiment, the task control apparatus 2 gives instructions for the speech recognition system 1, for example, selection a different dictionary to change a word to be recognized (one dictionary has a group of words which can be recognized with one speech and the word to be recognized is changeable by selection of one of dictionaries), and start of the recognition.

Here, the phoneme recognizing section 3 and the coder 4 are placed at the user side and the decoder 6, the word and sentence recognizing section 7 and the word dictionary 8 are placed at the central processing equipment side which is remotely disposed from the user side.

Fig. 4 shows one example of the contents of the word dictionary 8 which are mentioned with phoneme symbols. In Fig. 4, the "word" column shows Japanese Kanji (Chinese) characters corresponding to, the respective word dictionary items, but not used for the actual recognition. With this arrangement, an operation will be described hereinbelow. The following table 1 shows the kinds of the phonemes of the Japanese language used.

TABLE 1

30	Vowel	/a/, /i/, /u/, /e/, /o/, Devocalization of /i/ and Devocalization of /u/
35	Semivowel & Contracted Sound	/j/, /w/ Unvoiced Contracted Sound
40	Consonant	/m/, /n/, /ŋ/, /b/, /d/, /g/, /r/, /z/, /h/, /s/, /c/, /p/, /t/, /k/
45	Syllabic Nasal	/N/
50	Soundless double consonant	/Q/

A speech is inputted as an electric signal through a microphone, a handset and or the like to the phoneme recognizing section 3 in order to recognize the uttered phoneme. For example, in response to utterance of "SHI-BUYA", the speech signal takes a signal as illustrated by (a) in Fig. 5 and, as obvious from the above-mentioned table 1, the phoneme symbol train becomes "sibuja" as illustrated by (b) in Fig. 5. According to the current speech recognition technique, it is impossible to obtain 100% phoneme recognition rate, and hence the phoneme train contains errors. The recognized phoneme symbol train is supplied to the coder 4 so as to be coded and outputted in order to be suitable for the transmission line 5. In the case that the transmission line 5 is a

general public telephone line, the coding is performed in accordance with the frequency shift keying (FSK) system, the phase shift keying (PSK) system or the like. It is also appropriate to use a digital line such as a bus-structure network (Ethernet) as the transmission line 5. The decoder 6 performs a reverse process of the coding with respect to the signal transmitted through the transmission line 5 so as to restore it to the original phoneme symbol train. The word and sentence recognizing section 7 performs a matching of the phoneme symbol train from the decoder 7 with the phonemes of the respective dictionary items in the word dictionary 8 illustrated in Fig. 4. In the case of word recognition, the word number for the word most similar thereto, i.e., "001" in this embodiment, is outputted as the recognition result to the task control apparatus 2. Here, the word dictionary 8 can be constructed with a plurality of groups so as to be selectively used for every speech recognition process in order to limit the vocabulary. In the case of sentence recognition, it is required to additionally use syntax information, word-semantic information and others.

A speech recognition system according to a second embodiment of this invention will be described hereinbelow with reference to Fig. 6, where parts corresponding to those in Fig. 3 are marked with the same numerals. In Fig. 6, the speech recognition system indicated by a dotted line and illustrated at numeral 1 is included in a dialogue or interaction system comprising a terminal apparatus 11 and a central apparatus 12 which are coupled through a transmission line 5 to each other. The speech recognition system 1 comprises a phoneme recognizing section 3 responsive to an inputted speech, a coder 4 coupled to the phoneme recognizing section 3, a decoder coupled through the transmission line 5 to the coder 4, a word and sentence recognizing section 7 and a word dictionary 8. Of these sections of the speech recognition system 1, the speech recognizing section 3 and the coder 4 are placed at the terminal apparatus 11 side and the decoder 6, the word and sentence recognizing section and the word dictionary 8 are disposed at the central apparatus 12 side. Further, at the central apparatus 12 side are disposed a task control apparatus 2 coupled to the word and sentence recognizing section 7 and another coder 13 coupled to the task control apparatus 2, and at the terminal apparatus 11 side are disposed another decoder 14 coupled through the transmission line 5 to the coder 13 and a terminal control section 15 coupled to the decoder 14.

An operation of the above-mentioned arrangement will be described hereinbelow. As well as the above-described first embodiment, a pronounced speech by a user at the terminal apparatus 11 side is recognized by the speech recognition system 1. The operation of the task control apparatus for the recognition result is transmitted through the coder 13, transmission line 5 and decoder 14 to the terminal control section 15 which in turn delivers it to the user with a speech or letters through an indicator, a loud speaker or the like. After the operation of the task control apparatus, a speech is again introduced into the phoneme recognizing section 3 of the speech recognition system 1. Here, a recognition start command for the speech recognition system 1 is transmitted from the task control apparatus 2 to the word and sentence recognizing section 7 and further through the terminal control section 15 to the phoneme recognizing section 3. With the above-described arrangement, it is possible to provide a flexibility because the recognition vocabulary process and the operation procedure can be effected at the central processing apparatus side.

According to the above-described first and second embodiments, phonemes expressing a speech is recognized and a symbol train is coded and transmitted through a transmission means to a central processing apparatus. The central processing apparatus decodes it and recognizes and outputs the corresponding word or sentence. Thus, as compared with the direct transmission of a speech, it is possible to prevent reduction of the speech recognition rate due to line noises and others and further possible to recognize a word speech and a sentence speech transmitted from a remote place. Moreover, as compared with the Fig. 2 conventional system, it is possible to reduce the cost of the terminal apparatus to be disposed at the user side.

It should be understood that the foregoing relates to only preferred embodiments of the present invention, and that it is intended to cover all changes and modifications of the embodiments of this invention herein used for the purposes of the disclosure. For example, although in the above-described embodiments phonemes are used as the basic units of a language to be recognized, the present invention is not limited thereto and it is also appropriate to use syllables as the basic units. In addition, although the description is made in connection with Japanese language, the recognition with respect to languages other than the Japanese language can be made if recognition is performed in accordance with phonemes or others corresponding thereto.

A speech recognition system for recognizing a speech to be inputted so as to operate a given apparatus in accordance with the recognized speech. The speech recognition system includes a phoneme recognizing section responsive to input of a speech from an external device for extracting phonemes constituting the inputted speech to output them as a symbol train. The symbol train from the phoneme recognizing section is supplied to a coder for coding the symbol train and outputting the coded symbol train through a transmission line to a decoder for decoding the coded symbol train to restoring it to the original symbol train. The decoded symbol train is inputted to a word and sentence recognizing section which in turn recognizes a word or a sentence on the basis of the decoded symbol train using a word dictionary.

Claims

1. A speech recognition system (1) comprising:
 - means (3) responsive to an input of a speech from an external device for extracting phonemes or syllables constituting the inputted speech to output them as a symbol train;
 - means (4) coupled to said extracting means for coding said symbol train and outputting the coded symbol train;
 - means (5) for transmitting the coded symbol train;
 - means (6) coupled through said transmitting means to said coding means for decoding the coded symbol train to restoring it to the original symbol train; and
 - means (7,8) responsive to the decoded symbol train from said decoding means (6) for recognizing a word or a sentence on the basis of the decoded symbol train.

Patentansprüche

1. Einrichtung zur Spracherkennung (1), mit:
 - einer Einrichtung (3), die auf eine Spracheingabe von einer äußeren Einrichtung anspricht, um Laute oder Silben, die die eingegebene Sprache bilden, zu extrahieren und sie als eine Symbolfolge auszugeben;
 - einer Einrichtung (4), die mit der Extrahiereinrichtung zum Codieren der Symbolfolge und zum Ausgeben der codierten Symbolfolge gekoppelt ist;
 - einer Einrichtung (5) zum Übertragen der codierten Symbolfolge;
 - einer Einrichtung (6), die über die Übertragungseinrichtung mit der Codiereinrichtung gekoppelt ist, um die codierte Symbolfolge zu decodieren und sie in die ursprüngliche Symbolfolge zurückzuwandeln; und
 - einer Einrichtung (7, 8), die auf die decodierte Symbolfolge von der Decodiereinrichtung (6) anspricht, um ein Wort oder einen Satz auf der Grundlage der decodierten Symbolfolge zu erkennen.

Revendications

1. Un dispositif (1) pour la reconnaissance de la parole, comprenant :
 - un moyen (3) réagissant à une entrée se présentant sous forme de parole, venant d'un dispositif externe, en vue d'extraire des phonèmes ou syllabes constituant la parole ayant été introduite, en vue de les émettre sous forme d'un train de symboles;
 - un moyen (4), couplé audit moyen d'extraction, pour coder ledit train de symboles et émettre ledit train de symboles codé;
 - un moyen (5) pour transmettre le train de symboles codé;
 - un moyen (6), couplé, par l'intermédiaire dudit moyen de transmission, audit moyen de codage, pour décoder le train de symboles codé, afin de le remettre sous sa forme originelle de train de symboles;
 - et
 - des moyens (7, 8), réagissant au train de symboles décodé venant dudit moyen de décodage (6), en vue d'identifier un mot ou une phrase, d'après le train de symboles décodé.

FIG. 1 PRIOR ART

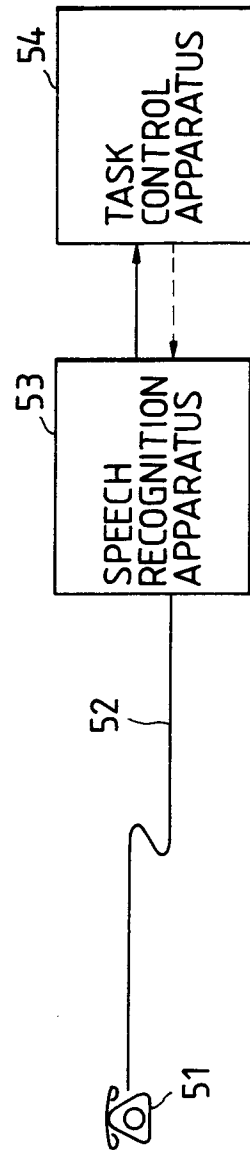


FIG. 2 PRIOR ART

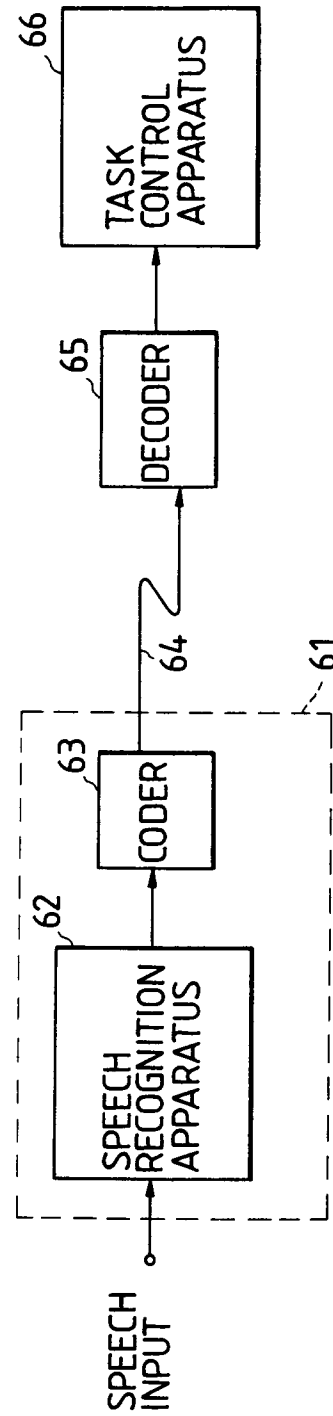


FIG. 3

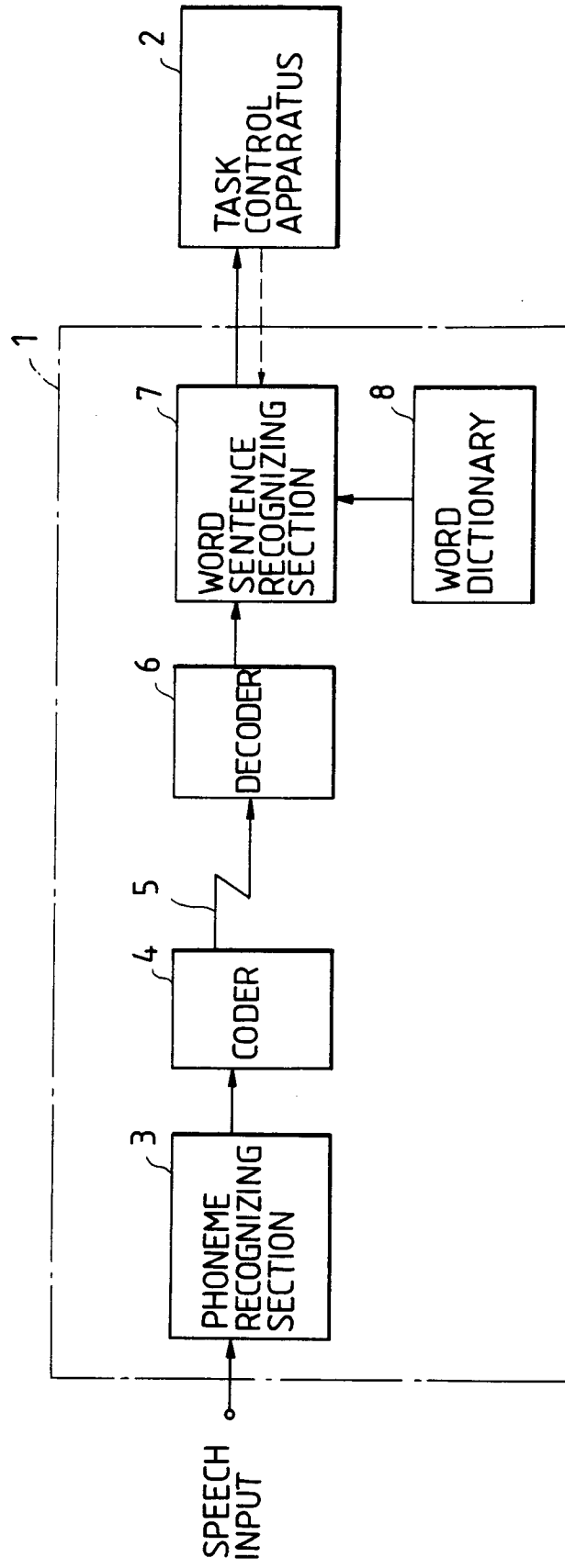


FIG. 4

8

WORD NUMBER	PHONEME DESCRIPTION	WORD
001	sibuya	渋谷
002	harajuku	原宿
003	yoyogi	代々木
004	sinzjuke	新宿
⋮	⋮	⋮

FIG. 5

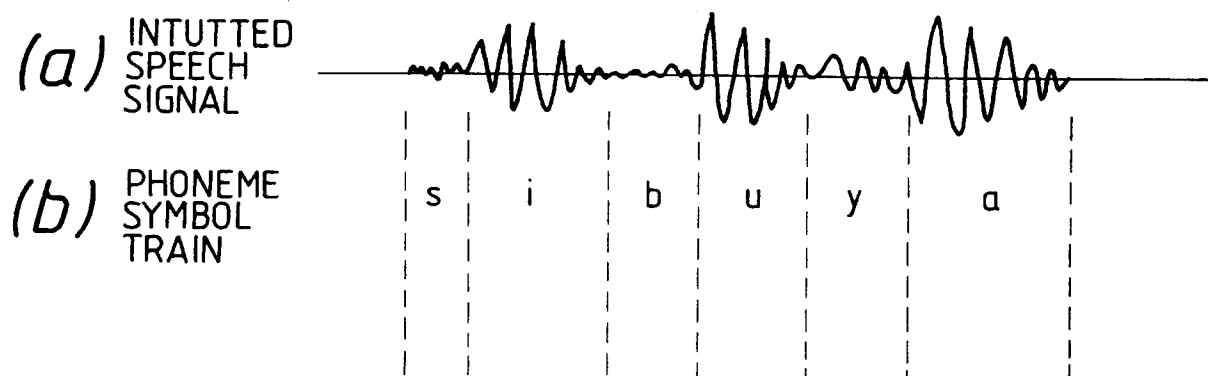


FIG. 6

