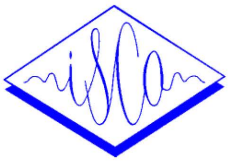

European Conference on
Speech Technology
Edinburgh, Scotland, UK
September 1987

A New Architecture of Large Vocabulary Word Recognition for Chinese Character

Zhao Guotian

Speech Technology Lab.
Dept. of Radioengineering
Harbin Institute of Technology
Harbin, P.R.China

ABSTRACT

The speech input system of Chinese Character for office automation needs a large vocabulary spoken word recognition. For a word recognition system the recognition time is proportional to the vocabulary size in it. So it is difficult to achieve real time operation of the system. We propose a new word recognition system architecture using severel word recognizers which have full recognition function and template memory separately. The system's vocabulary is the sum of the recognizer's vocabulary's and the recognition time is near to a recognizer's. The central processor fetches the numbers and similarity parameters and decides which is correct input word. If the correct input word is behind the first, a select key will be typed.

We have organized several such system vocabulary in which is over 2,000 - 3,000 words and they are on trial in several organization now.

10.21437/ECST.1987-81

I. Introduction

Recently, Chinese Information Processing (CIP) have gone into a new stage. China is a great nation and there are many people using Chinese character in it. The research of CIP is strategic significance and a urgent task to rise the efficiency of computer operation and secretary work. The large amount of work in CIP, such as the approach of Chinese character coding system, has being done. Several hundred of coding systems have been developed in China since 1978, but the problem of typing in Chinese character remains. These coding systems need skillful men to operate and it isn't easy to type in machine as English. The load of operater is too heavy. These troubles are avioded if the input way of utterance is made use of. The way is more nature than typing the key. So the voice input of Chinese character is a important area of CIP and is noticeable now. It is well known, Chinese is a tonal language with large vocabulary and Chinese character is monosyllable. Chinese words are built facilly in modern literary, so can't say how many words there are eventhough ignoring the dialect. Therefore Chinese say phrase often instead of term word.

But the official and business documents the vocabulary is not very large and the words are more normalized. We have done a lexical statistics using a database including some goverment documents, reports, announcements, official statements, notices and speeches of goverment heads. The conclusion

was that two thousand of words satisfies the 97% of requirement [1].

To realize that Chinese character is inputed into machine by speech recognition method, it is necessary to weigh the recent technique power of speech recognition.

In accordance with the difference of the template used, the system for automatical spoken word recognition may be classified as two. They are the recognition system using word or phoneme reference template [2]. Fig.1 is a principle block of the system, where the word templates are used. The templates have been getten by the way that all words are trained one by one, so that the accuracy is higher. Now this type of the recognition system is widely applied. The response time is proportional to the vocabulary size because the input utterence is matched against each of templates by the same accuracy and precise pattern matching takes more computation, which is one of lacks. At present Dynamic Programming (DP) is most accurat and robust recognition algorithm. The number of words which most currently available systems can recognize in real time is limited to a few hundred for the lack. The recent recognition technique can not provide a thousands word recognizer yet. And recognizer using phoneme template wich can process hundreds words is not available yet too.

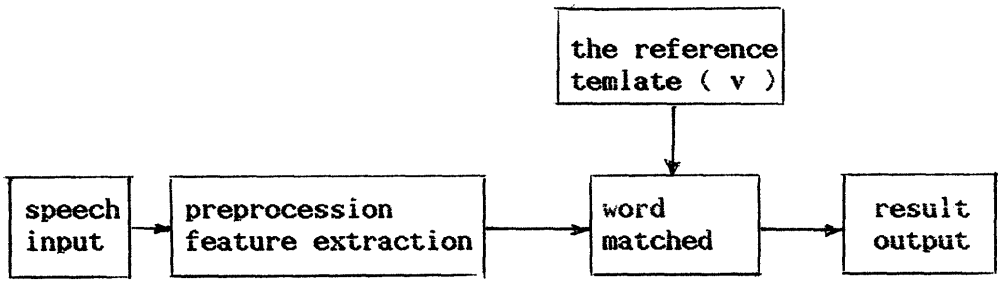

Fig.1

We have not recommended the utterance of monosyllable to input Chinese Characters because there are many homophonic characters for one monosyllable. It implies that operator have to type a key several time to select a character.

II. THE RECOGNITION PRINCIPLE

To use the high accuracy algorithm, this paper presents a new architecture using parallel processing the input utterance simultaneously. It requires each recognizer is a independent subsystem which have all matching function and own temolate memory. And the system requires all subsysytem are same to ease operating for main computer. The response time and accuracy of the whole system depend on those of single recognizer. The vocabulary may be increased multiplicatively. According to the new scheme where the input utterance is parallel processed in the form of multirecognizer (in different template subset), a function block diagram of the recognition system pro-

posed may be shown by Fig.2 . The speech recognition subsystem in Fig.2 is a speech recognizer using word reference template(see Fig.1). Then the recognition result of every recognizer was send into main computer, and they are ranked as the candidates and shown in the prompt line located lowest on the screen. If the wanted word is located in the first position , the next utterance will input it into text. If it is located following, the operator have to type a numeral key according to the rank to put it into text. If the wanted word is not located in the prompt line, the operator have to type a numeral key with a number biger than possible rank and re-utter the word.

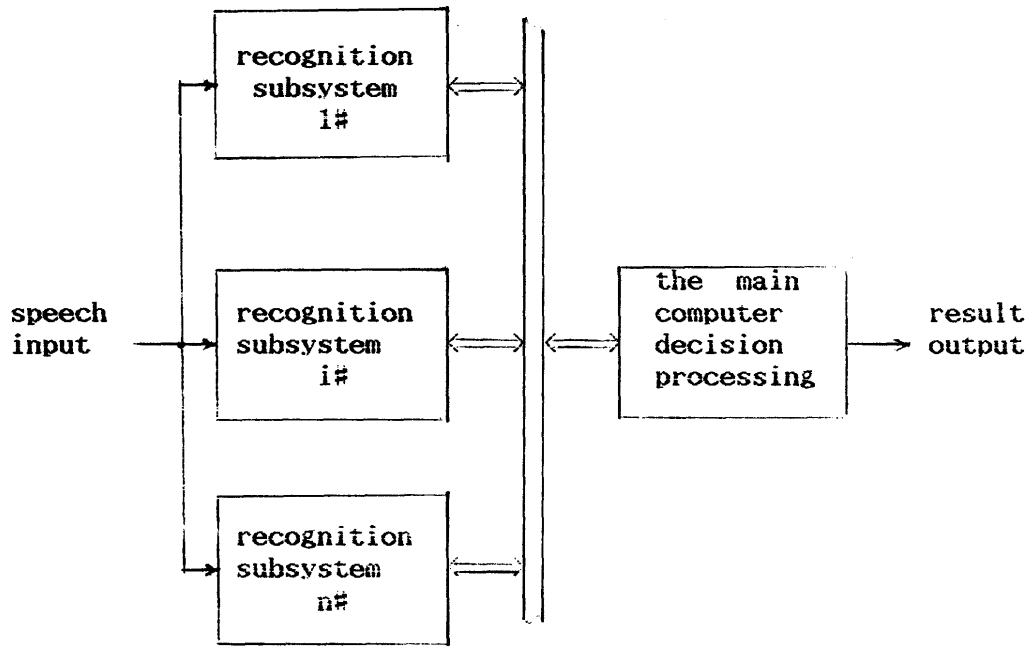

Fig.2 Function block diagram of the recognition system

III. THE EXPERIMENTAL SYSTEM

To test the efficiency of scheme above, we have built several speech input systems with intellectualized word recognizer. The main computer is a 16-bit computer IBM PC/XT or AT, or GREAT WALL 0520 in this system. We have done the experimental research with N=4 or 6. The vocabulary involved in this system was a practical one consisting of 2048 or 3072 most frequently occurring words in office-correspondance corpus. The recognizers used here are same independent subsystem.They are connected with main computer through the standard parallel interface(8255), and the templates can be saved on the diskettes of main computer.

Meanwhile, the system allows operator to input Chinese characters by key according to any coding system and to edit the text any time.

IV. CONCLUSION AND DISCUSSION

In our experiments over 95 % of wanted words were input by first utterance and 80 % of the words were input without typing any key. The input rate was about two times that for typing in using any coding system for non-skillful operators.

Through experiment and analysis, the conclusions are as follows:

1.) The architechture scheme proposed here is feasible in practice. Maybe it is one of approaches to increase vocabulary for speech recognition as hierarchical algorithm [4];

2.) Within the recognition system discribed above, the accuracy and responce of the system are limited to those of the subsystem because the subsystem is independently unaffected each other. The accuracy is similar to that of subsystem. Though the main computer for judging takes very short time only, the responce time is yet similar to that of the subsystem;

3.) As the i of the subsystem geting greater, the real-time vocabulary V of the system increase by a factor of i:

$$V = v * i$$

Where v is the vocabulary of the subsystem.

A problem which must be solved in automatic speech recognition is that of finding reasonable automatic ways to reduce newtalker load in the training process. The current work did not address this problem.

REFERENCE

1. Zhao Guotian: "On Input technique of Chinese Character", International Conference on Speech Input/output Technique and Application, London, UK, March 1986

2. S.Sato "Speech Recognition" Trains of IECE Japan. Vol.62, No.2, 1982, Japanese.

3. S.Nagata: "Acoustic Signal Processing LSI", JASJ, Vol.39, No.11, 1983, Japanese.

4. T.Kaneko and N.Rex Dixoon :"A Hierarchical Approach to Large-Vocabulary Discrete Utterance Recognition", IEEE Trans., Vol.ASSP-31, No.5, October, 1983.