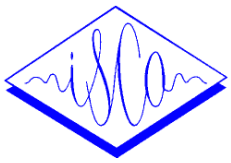

First International Conference on
Spoken Language Processing (ICSLP 90)
Kobe, Japan
November 18-22, 1990

A KNOWLEDGE–BASED UNDERSTANDING SYSTEM FOR THE CHINESE SPOKEN LANGUAGE

Yanghai Lu Beiqian Dai

Department of Electronics Engineering
University of Science & Technology of China
Hefei, Anhui, P.R. China

ABSTRACT

This paper describes a Chinese spoken language understanding system USTC–2. Various kinds of knowledge such as acoustic–phonetics, vocabulary, syntax and semantics are represented and utilized in the system. It's constructed as an expert system based on frame representation capable of metting the needs of various tasks. In USTC–2 system, segmentation of input speech is carried out utilizing the energy contour, LPC variance and syllable length. Recognition is performed by a matching algorithm which constrains the search on the basis of morphological knowledge concerning both the part of word–class and and constituent syllales of a given word . The use of morpohological knowledge not only reduces the amount of computation but also increases the correct rate of syllable recognition in spoken Chinese sentences. An analysis method starting from KEYWORDS was proposed in speech understanding. This method combines syntastic analysis and semantic analysis together and has strong ability of correcting errors. The system discriminates homonymous syllables and homonymous words in each processing step.

1. INTRODUCTION

Spoken language understanding is one of the most important function required for an intelligent computer. However, there exist some difficult problems in this field . Ambiguilty due to homonyms constitutes a major obstacle in its realization. Since there are many homonymous syllables and homonymous words in the Chinese language. Various kinds of knowledge are nesessary for their disambiguation. In USTC–2 system, various kinds of knowledge such as acoustic–phonetics, morphology, syntax and semantics are represented and utilized. USTC–2 system is improved of the USTC–1 system, which was developed by University of Science Technology of China in 1988 . [1] The improvements are mainly on the following aspects:

(1) It is constructed as an expert system based on frame represetation capable of meeting the needs of various tasks.

(2) It incorporates a grammer of natural Chinese language capable of accepting almost all kinds of simple sentences.

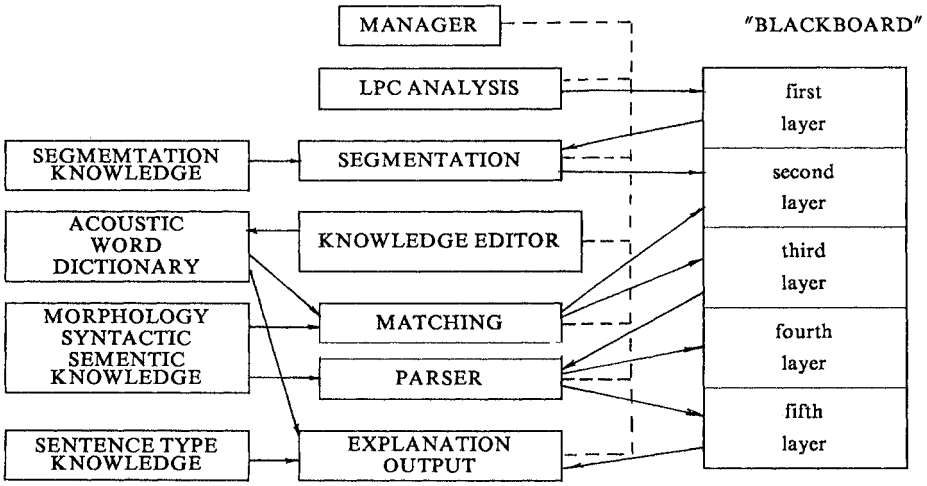

Figure 1

(3) It adopts a new parsing method capable of discriminating homonyms and correcting errors.

In section 2 , we discussed the overall structure of USTC–2 system. We give the essential performance of USTC–2 system in Section 3.

2. OVERALL STRUCTURE OF USTC–2 SYSTEM

Figure 1 illustrates the block diagram of USTC–2 system. The details of each block in figure 1 will be described following.

A. LPC Analysis

The speech signal is passed through band–pass filter with passband from 200 Hz to 4.5 kHz, and sampled at 8 kHz rate with 12 bits. At first, the sampled speech s(n) is preemphasized by a first–order digital network $H(z)=1-0.95z$. The resulting signal is then blocked into frames of size $N=300$ samples, with adjacent frames seperated by $L=150$ samples. A ten–order LPC analysis is then performed. At the same time, the short time energy and zero–crossing of each frame are calculaled. The log energy of each frame is appended the eleventh parameter of vector for each. [2] Those calculations above is implemented on TMS320 Board in nearly real–time.

B. Segmentation

For continuous speech recognition, a matching method based on word–template is not pratical. This is because the number of words is too large. In USTC–2 system, the syllable is chosen as the basic units of recognition . The syllable plays a very important role in Chinese language. Most syllables are uttered clearly in a medium–speed utterance. Experiments show following features between two adjacent syllables in a medium–speed utterance:

(1) The energy exists a valley between two syllables.
(2) The zero–crossing exists a peak between two syllables.
(3) The LPC parameters change violently between two syllables.
(4) The length of syllable doesn't vary out of the range from a half of the standard length to a double.

In SEGMENTATION KNOWLEDGE SOURCE, there are four kinds of knowledge. They are about energy, zero–crossing, LPC variance and syllable length. The SEGMENTATION procedure combines the four kinds knowledge to decide synthetically the beginning and the end of syllable . After the segmentation was done, a syllable energy normalization was performed on the entire utterance string. The purpose of the energy normalization is to adjust the peak energy is at or close 0 dB for each syllable.

C. Knowledge Representing

The main objects are all kinds of Chinese spoken sentences, not continuous digit speech. We all know that men need knowledge not only when they want to understand sentences but also they want to recognize sentence. Similarly, a computer also need knowledge in order to understand sentences and get a highly recognition rate. There exists the problem of knowledge representing in every intelligent expert system. In USTC–2 system, many kinds of knowledge such as acoustic–phenetics, vocabulary, syntax and semamtics are represented. According to the character and applying way of knowledge, the knowledge can be represented in different way. There are chiefly two representing ways in USTC–2 system. One is framework, the other is production rule.

Framework is a kind of universal data architecture in expert systems. It is suitable to represent the knowledge which discribes regular events. A framework consists of several solts which discribe each aspect of events. Each solts has its name and its value. In USTC–2 system, the acoutic–phonetic dictionary and word dictionary are organized in framework. The acoustic–phonetic dictionary is mainly about the knowledge of Chinese syllable. The entire acoustic–phenetic dictionary is a framework system in which each syllable refer to a framework. Each framework consists of two solts. According to the need of the task, users can increase or decrease the number of syllables and can change the value of any solt. The word dictionary is mainly about the knowledge of vocabulary. The entire word dictionary also consists of a framework system. Each word refers to a framework system too. Each framework consists of six solts. According to the need of task, users can also increase or decrease the number of word and can change the value of any solts.

The frameworks mainly describe objects and matters, but it needs a flexible representing way to inference with matter. PRODUCTION RULE is one of those representing way. Its form is as follows:

IF condition, THEN decision

PRODUCTION RULEs mainly represent the knowledge of rule which has complex logic. Each source of knowledge which is organized in PRODUCTION RULE is an independent computer procedure and each knowledge source consists of many PRODUCTION RULEs, or consists of many CONDITION–ACTION pares. These actions can get the middle results in "BLACKBOARD", and can generate or modify the hypthesis in "BLACKBOARD". In USTC–2 system, MORPHOLOGY KNOWLEDGE, SYNTACTIC KNOWLEDGE and SEMANTIC KNOWLEDGE are represented in this way.

D. Knowledge Guided Matching and Parsing

There are 1300 syllables in Chinese language , in which there exist many similar syllables. If matching was done with each syllable and a syllable–lattice was created, not only the amount of matching computation is very large but also there are many similar candidates for the syllable in the syllable–lattice. Many of the candidates cannot be accepted by morphology. So the morphology could be utilized to restrict the matching candidates for reducing the amount of computation. There are two kinds of knowledge in MORPHOLOGY KNOWLEDGE SOURCE. One is a set of language constitution rules. In Chinese language, words can be divided into ten classes, namely, noun, verb, adjective, pronoun, nueral, preposition, adverb, interjection, conjection and auxiliary particle which only exists in Chinese language. There are some rules on language constitution. For example, the word following a adverb must be a verb or an adjective in Chinese. The other in MORPHOLOGY KNOWLEDGE SOURCE is a dictionary of which tells the word–class of a word and the syllable sequence consisting of the word. The MATCHING procedure utilized the kinds of knowledge to restrict the candidates. As a result, a word–lattice was created directly instead of a syllable–lattice.

Table 1 gives the comparison of Morpholgy Guided matching .

TABLE 1 COMPARISON OF M.G.MATCHING AND NORMAL MATCHING

	Average Recognition time for each Syllable	Accuracy of Syllable
K. G. Matching	4 seconds	92%
Normal Matching	20 seconds	85%

The MATCHING procedure outputs the word–lattice with uncertainties due to homonymous words or errors in word recognition. The Parsing task aggregates the word–lattice into a sentence while discriminating homonymous words and correcting errors.

In traditional parsing method, a language transition network was used. It can tell whether a sentence is accepted, but it is not convenient to discriminate homonymous words and to correct errors due to word recognition. So a new parsing method was proposed in this paper, which is called Keyword parsing, and it was adopted in USTC–2 system. We all know every word does not play an equal role in a sentence. Some are important to understand sentences, we called them Keywords. Some are not important. If they are mis–recognized, we may still understand the sentence and maybe correct them. Keyword parsing method is just on this principle. There are many syntactic and semantic rules in SYNTATIC & SEMANTIC KNOWLEDGE SOURCE. According to the syntatic rules, the KEYWORD parser first researches for keywords using TOP–DOWN approach. Then it parses the whole sentence from the keyword. In this parsing procedure, not only the syntatic &semantic knowledge is used but also some special rules for discrimintating homonymous words are used.

The PARSER canceled the words unaccepted by syntax & semantics, discriminated the homonymous and corrected the errors due to word recognition. If the parser can not correct the errors successfully, it will transfer the information to MANAGER. USTC–2 system will recognized the sentence again under the control of MANAGER.

In Chinese language, there are four sentence function types. They are declarative sentences, interrogative sentences, Imperative sentences and execlamatory sentences. After parsing, the function type of the sentence is also obtained.

Finally USTC–2 system outputs Chinese character sentences with punctuation marks.

3. THE PERFORMANCE OF USTC–2 SYSTEM

The system is being developed by computer simulation .The computer system consists of two parts : a signal processor TMS320 board for the LPC analysis of input speech and the main processor IBM–PC / XT for other processing, linked by the 8086–ASM. Overall performance of USTC–2 system is summarized in Table 2.

TABLE 2 USTC–2 SYSTEM PERFORMANCE

Speaker	Dependent
Environment	Computer terminal room
Microphone	Medium–quality, close–speaking
Speaker adaptation	None required
Vocabulary	100 syllable, 300 words
Language constraints	Natural simple sentence
Test data	50 utterance, 10 syllable / utterance average
Accuracy	5% syllable error, 9% sentence

REFERENCES

[1] Xiaodong Wang and Beiqian Dai, "Chinese Speech Understanding System", ICPR–88, ROME , PP.1088–1090 .

[2] W.A. Woods, "Transition Network Grammars for Natural Language Analysis", CACM 13, pp. 591– 606, 1970