The Preliminary Results of A Mandarin nictation Machine
Based Upftn Chinese Natural Language Analysis

Lin-shan Lee***, Chiu-yu Tseng**'***, KJ. Chen**, and James Huang****

* Dept of Computer Science and Information Engineering, National Taiwan University, Taipei, Taiwan, Rep. of China. Tel: (02) 392-2444

** The Institute of Infonnation Science, Academia Sinica, Taipei, Taiwan, Rep. of China.

*** The Institute of History and Philosophy, Academia Sinica, Taipei, Taiwan, Rep.of China.

****Dept of Modern Languages and Linguistics, Cornell University, Ithaca, N.Y, USA.

Abstract

This paper describes the preliminary results of the first research effort toward a Mandarin dictation machine in the world for the input of Chinese characters to computers. Considering the special characteristics of Chinese language, syllables are chosen as the basic units for dictation. The machine is divided into two subsystems. The first is to recognize the syllables using speech signal processing techniques. Because every syllable can represent many different characters with completely different meaning, the second subsystem then identifies the exact characters from the syllables and corrects the errors in syllable recognition by first forming all possible words from the syllables then finding out one combination of the words which is grammatically valid in a sentence. The preliminary test results indicate that such a dictation machine is not only practically attractive, but technically achievable.

1. Introduction

Today, the input of Chinese characters into computers is still a very difficult and unsolved problem, which is the basic motivation for the development of a Mandarin dictation machine. We define the scope of the research by the following limitations. The input speech is in the form of isolated syllables instead of continuous speech (The choice of syllables as the dictation unit will be discussed in detail later.). The machine is speaker dependent. The first stage goal of this system is to have 90% correction for the sentences in the Chinese textbooks of the primary schools in Taiwan, Rep. of China. The errors can be found by the user on the screen and corrected from the keyboard. Such a performance is still much more efficient than any of the currently existing input systems. Also, only a small dictionary for demonstration purpose is to be established in the first stage. The machine will work well for sentences formed by the words in the dictionary, otherwise the new additional words have to be keyed into the dictionary. To our knowledge, this is the first research effort toward a Mandarin dictation machine in the world.

The dictation machine is divided into two subsystems. The first one is to recognize the syllables using speech signal processing techniques, but this is not very helpful at all because in general every syllable can represent many different characters and can possibly form different multi-syllabic words with syllables on its right or left. Therefore the second subsystem is to identify the correct characters from the syllables by forming correct words which is grammatically valid in a sentence and carefully considering the characteristics of the Chinese natural language.

IL____Considerations for the Special Structure of Chinese Language

There are at least 80 thousands of commonly used words in Chinese. Therefore the words can not be used as the dictation units. There are at least 20 thousands of commonly used Chinese charaters, each character is mono-syllabic. Each word is composed of from one to several characters. A nice feature is that the total number of different syllables in Mandarin speech is only about 1300. If we use the 1300 syllables as the dictation units, all the words or characters will be covered. However, the small number of syllables implies another difficult problem, that is, many different characters will share the same syllable. This is why we need the second subsystem. Based on the above observations on the special structure of Chinese language, the use of syllables as the dictation unit becomes a very natural choice.

Another very special important feature of Mandarin Chinese language is the tones for the syllables. Every character is assigned a tone in general. There are basically four different tones. It has been shown * that the primary difference for the four tones is in the pitch contours, and the tones are essentially independent of the other acoustic properties of the syllables. If the differences among the syllables due to lexial tones are disregarded, only 411 syllables are required to represent all the pronunciations for Mandarin Chinese. This means the recognition of the syllables can be divided into two parallel procedures, the recognition of the tones, and of the 411 syllables disregarding the tones.

TIL The Overall System Structure

Based on the considerations described above, the overall system structure for the Mandarin dictation machine is shown in Fig.l. The system is basically divided into two subsystems. The first is to recognize the syllables, and the second is to transform the scries of syllables into the characters. For the first subsystem of syllable recognition, the corresponding syllable (disregarding the tones) and the tone are then recognized independently in parallel. Because errors always happen, we therefore have to provide information for confusing syllables, and confusing tones. For the second subsystem we need to first form multi-syllabic words from each syllables. To use the above example, although there are many characters all correspond to the syllable [guo-2] and many to [iu-3], there is only one multi-syllabic word " Qfltl (Mandarin)" has the pronunciation" [quo-2] [iu-3]", etc.. But this doesn't solve the problem well. First, mono-syllabic words, such as [ni-3], [shr-4], [i-2], [jia-4], [huei-4], [tieng-1] can't be identified in the above way. They even form ambiguous multi-syllabic words, for example, the syllable [i-2] can combine with the syllable to its lett lshr-4J to torm a wrong word " : (suitable, [shr-4] |i-2D", or with the syllable to its right [jia-4] to form a wrong word " \%W (Move, [i-2] [jia-4])". The problem becomes even worse when errors occur in the recognized syllable, for example, if the syllable [ni-3] is incorrectly recognized as the syllable [ni-4], it can be combined with the syllable [shr-4] to its right to form a wrong word " (Bad Situation, [ni-4] [shr-4])". The next operation of syntactic analysis then serves as a filter to rule out all ungrammatical combinations of multi-syllabic and mono-syllabic words and only a single syntactically valid sentence will be obtained and appear on the screen as the output text.

TV. The Recognition of the Syllables

The recognition of the 411 syllables in the first subsystem is in fact very difficult^' . This is because the 411 syllables consist of about 37 confusing sets, each of which has from about 4 to 19 confusing syllables2,3, A good example is the A-set: {[a], [ba] , [pa], [ma], [fa], [da], fta], [na], [la], [ga], [ka], [ha], [ja], [cha], [sha] [tza], [tsa], [sa]). If we use 200 of these syllables to train a commercially available speech recognizer product with 200 words of recognition vocabulary designed for English words, the recognition rate is on the order of 54% - 60% only, although the rate is about 92% for English words. An initial / final two-phase recognition approach is thus specially designed2,3 to recognize these very confusing syllables, whose block diagram is in Fig.2. We first detect the final part and recognize the final from a total of 38 different finals. Once the final is determined, we then try to recognize the initial preceding the final among at most 19 candidate initials.

The recognition of the finals is also very difficult because there are again many confusing final sets, such as ([a], [ia], [ua]j, etc. In our research it was found that using multi-section vector quantization (MSVQ) techniques with branch-and-bound algorithm operated on the filter bank coefficients can give very good results while maintain relatively simple hardware/software implementation[4]. Preliminary result of recognition rate of 93.4 % has been achieved for speaker-dependent case. The recognition of the initials is even more difficult because they are relatively short, unstable, and confusing. A special two-phase recognition approach is therefore developed to recognize the initials[2,3]. The first phase classifies the possible unknown initials into several small groups using formant parameters. The second phase then recognizes the exact initial. The preliminary recognition rate is found to achieve 90.6% for speaker dependent case. The correct recognition of the tone is also difficult. The primary problem is the confusion caused by the third tone. We use here the tone recognition method developed by J.-C. Lee[5], in which the sum and difference of the log pitch frequencies for adjacent frames are taken as feature parameters. Vector quantization (VQ) codebooks and Hidden Macov Models (HMM) are developed to recognize the tones. The recognition rate for the tones is found to be 98.3%. Combining the results for the final, initial and tone recognition, the total recognition rate for the syllables is only on the order of 82.3%.

V. Word Formation from Syllables

This is the first part of the second subsystem. The output characters and words for the above example, [ni-3] [shr-4] [i-2] [jia-4] [huei-4] [tieng-1] [guo-2] [iu-3] [de-5| [dian-4] [nan-3], is shown in Fig.3. First, all multi-syllabic words which can be found in the dictionary will be obtained, such as " MR (Suitable, [shr-4] [i-2])[M], " (Move, [i-2] [jia-4])[M], Mandarin, [guo-2] [iu-3])", " (Computer, [dian-4] [nau-3])". Secondly, all syllables which can not form multi-syllabic words with adjacent syllables must correspond to mono-syllabic words, therefore the most frequently used and second frequently used (if any) mono-syllabic words will be found, such as " fa (You)" and " 81 (Plan)" for [ni-3], " % (Can)" for [huei-4], etc. Thirdly, if some of the syllables in the multi-syllabic words obtained in the first step correspond to very frequently used mono-syllabic words, such as " il (Are)" and " $ (Affair)" for [Shr-4], etc., they are also given. In the dictionary, the words are organized in a special structure called word trees such that all words having the same first syllable will be organized in the same tree, and all words in the same tree having the same second syllable will be organized under the same node, etc. to improve the searching speed. In order to reduce the total number of words which have to be stored in the dictionary, some additional word formation rules are designed to generate new words". For example, a verb followed by the character " i$ " ([guo-4]) simply represent the past tense, such as " &£ " ([chr-1], eat) and " M " forms " Jfc j§ " ([chr-1] [guo-4], ate), etc. Some additional word deletion rules are also designed such that redundant or impossible words can be automatically deleted without going to the syntactic level". One example rule is as follows. Any shorter words within a longer word should be deleted. For example, the shorter words " f=i M " ([tai-2] lwan-1], Taiwan) and " !:*£ " ([da-4] [shiue-2], University) will both be deleted if the longer word " feM Af " ([tai-2] [wan-1) [da-4] [shiue-2], Taiwan University) exists. Another example rule tells that all words which can not be connected by other words on both sides (except beginning and ending words) should be deleted.

XL—Syntactic Analysis To Identify Correct Words and Characters

This is the second part of the second subsystem. The input is the series of all possible mono-syllabic and multi-ayllabic words corresponding to the recognized syllables. The goal is to find the unique correct combination of these words which form a grammatically valid sentence. The approach is simply to take all possible combinations of the words and parse the sentences in parallel based on a Chinese grammar. The ungrammatical combinations will be ruled out and only the correct solution will be left and shown on the screen. A Syntactic Analysis System for Chinese Sentences (SASC) is employed here to parse the Dossible sentences and rule out ungrammatical sentences'' . In SASC, Chinese sentences are syntactically analyzed from the viewpoints of generative grammar. It uses a bottom-up parser instead of a top-down parser, because the former tends to be more efficient for Chinese sentence analysis. The parser uses charts as global working structures to avoid the inefficiency in duplicating many computations that a top-down parser often suffers when backtracking occurs. The parser parses sentences in such a way that phrases are built up on the chart by starting with their heads and adjoining constituents on the left or right of the heads. For example, according to the phrase structure rule (PSR), " VP. . . . > YzR NP", V-n (transitive verb) is the head of VP(Verb Phrase). When encountering a transitive verb, the parser's action is to try to adjoint the following NP as its object. In order to solve the complicated syntactic phenomena such as passivization, relativization, topicalization, and Ba-transformation in Chinese sentences, a raise-bind mechanism based upon the theory of empty categories is developed**. In this way, the SASC parser will treat the above different transformations very easily in the same way, i.e., if an empty element is inserted into the right position, the syntactic structure will become easy. The parser generates an empty NP inserted into the vacant position where an NP is expected to appear. Using this approach, many syntactically sophisticated sentences can be easily parsed.

VII, Test Results and Conclusion

By combining all the different operations described above to form the complete system, some preliminary tests are performed. The results in general depend on the text. The rate for syllable recognition is on the order of 81% - 83%, but the final rate for dictation (i.e., the correction rate for the characters) is on the order of 88% -90%. In other words, some errors made in acoustic level can be finally corrected in the syntactic level. Further improvements should be made on each part of the system and the project is still under active development.

Acknowledgement

The authors wish to thank many excellent students working with them in Taiwan University in the last few years. Without their diligent efforts such research results are simply impossible.

References

[1] S.-M. Lei, L.-S. Lee, " Digital Synthesis of Mandarin Speech Using Its Special Characteristic ", Journal of the Chinese Institute of Engineers, Vol.6, No.2, Mar. 1983, pp.107-115.

[2] M.-S. Yu, G.-S. Chen, C.-C. Hsiao, C.-Y. Tseng, L.-S. Lee, " A Cascaded Two-phase Approach to Isolated Mandarin Syllable Recognition ", Third European Signal Processing Conference, Sept. 1986, The Hague, The Netherland, pp. 525-528.

[3] M.-S. Yu, G.-S. Chen, C.-C. Hsiao, C.-Y. Tseng, L.-S. Lee, " A Preliminary Approach to Complete Vocabulary Mandarin Syllable Recognition", 1986 International Conference on Chinese Computing, Aug. 1986, Singapore, pp. 168-171.

[4] C.-W. Hwang, C.-Y. Tseng, L.-S. Lee, " An Efficient Mandarin Vowel Recognition System Based Upon Multi-Section Vector Quentization and Branch-and-Bound Classification Techniques ", 1986 International Computer Symposium, Dec. 1986, Tainan, Taiwan, Rep. of China.

[5] J.-C. Lee," Mandarin Lexical Tone Recognition Based on Vector Quantization and Hidden Macov Models" Master Thesis, Tsing-Hua University , Taiwan, Rep. of China, May 1986.

[6] C.-G. Chen, K.-J. Chen, L.-S. Lee," A Model for Lexical Analysis and Parsing of Chinese Sentences ", 1986 International Conference on Chinese Computing, Aug. 1986, Singapore, pp.33-40.

[7] L.-J. Lin, K.-J. Chen, James Huang and L.-S. Lee (1986) " SASC: A Syntactic Analysis System for Chinese Sentences ", 1986 International Conference on Chinese Computing, Aug. 1986,Singapore, pp.29-32.

[8] Lin, L.-J., Huang, J., Chen, K.-J., and Lee, L.-S. (1986) " A Chinese Natural Language Processing System Based Upon the Theory of Empty Categories ", Fifth National Conference on Artificial Intelligence, AAAI, Aug. 1986, Philadelphia, PA, USA, pp. 1059-1062.

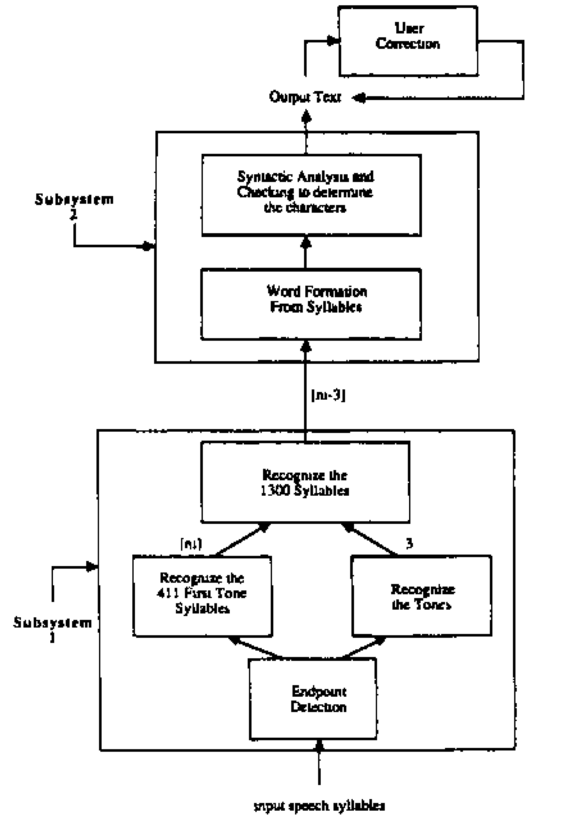

Fig 1 The Detailed Overall System Structure for the Mandarin Dictation Machine.

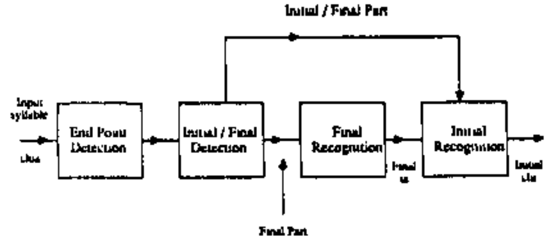

Fig.2. The Initial / Final Two-phase Recognition Scheme

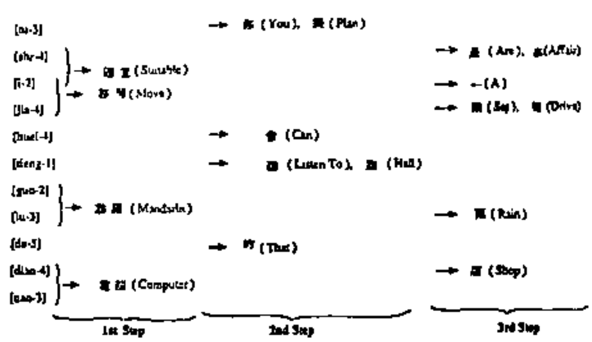

Fig. 3 The Steps in Word Formation from Syllables, only the underlined words are correct.