International Symposium on Chinese Spoken
Language Processing (ISCSLP 2004)
Hong Kong
December 15-18, 2004

CHINESE LARGE-VOCABULARY NAME RECOGNITION SYSTEM USING CHARACTER DESCRIPTION AND SYLLABLE SPELLING RECOGNITION

Nick J.C. Wang, Ching-Ho Tsai, Patrick Huang and Jia-Lin Shen

Department of Man-Machine Interface, R&D Center,
Delta Electronics Inc., Taipei
nick.jc.wang@delta.com.tw

ABSTRACT

The large-vocabulary name recognition technique is one of the challenging tasks in the application of Chinese speech recognition technology. It can be applied on long-list automatic attendant systems and automatic directory assistance systems. A Chinese name has usually two to three characters with each character pronounced as a single syllable. It is a high perplexity task to recognize a word from a long-list of candidates, like more than three hundred thousand unique names in our experiments, given a very short utterance like one to two seconds of speech. Two novel approaches under an interactive framework are proposed in this paper to aid the recognition of a Chinese name: Character Description Recognition (CDR) and Syllable Spelling Recognition (SSR). Together with our robust finite-state recognizer given a graph-structured syllable lexicon for the full names, we achieved a very promising name recognition success rate, 94.5%, in our system-initiative dialogue system.

1. INTRODUCTION

Chinese language is very different from purely phonetic writing systems of Western languages. First of all, Chinese makes a sharp distinction between its written language and its spoken language. It integrates both meaning and pronunciation information in its written forms – characters (‘字’) [1]. A small portion of them is pictograms or ideograms, which account for the most of the radicals, fundamental components, of Chinese characters. Most Chinese characters are radical-phonetic compounds or radical-radical compounds. For example, the character for ‘mother’ (‘媽’ in Chinese *ma1*) is a radical-radical compoud and consists of a radical componenet meaning ‘femal’ (‘女’ in Chinese *nü3*) in the left side and a phoneic component meaning ‘horse’ (‘馬’ in Chinese *ma3*) in the right side. Another example character for ‘trust’ (‘信’ in Chinese *xin4*) is a radical-phonetic compoud and consists of a radical component meaning ‘a person’ (‘人’ in Chinese *ren2*) in the left side and another radical component meaning ‘speech’ (‘言’ in Chinese *yan2*) in the right side. In the traditional Chinese character set of printed contemporary writing system, there are about three thousand commonly used characters. Characters with the same initial radical are categorized into the same word group in a modern dictionary. In Chinese, a word/phrase (‘詞’, which is a unit of meaning) is composed of one or more characters, like the word ‘辨識’ (recognition) of two characters.

Secondly, Mandarin Chinese is a spoken language with pronunciation based on monosyllabic units and pitch shapes. Each character is pronounced as a single tonal syllable. There are about 408 base syllables, which can be voiced mainly with four different pitch shapes (tones): high-level tone (陰平), rising tone (陽平), low-falling-raising tone (上聲), and falling tone (去聲), besides the neutral tone (輕聲). Each syllable is constructed after the pattern: “optional initial consonant followed by vowel followed by optional nasal”. Zhuyin (also known as “Bopomofo” system) and Hanyu-Pinyin are two of the most popular phonetic transcription systems for Mandarin Chinese. The former is the official standard in Taiwan, the latter in Mainland China. They are both phonetic systems to transcribe each Mandarin syllable sound into a sequence of phonetic symbols.

The Chinese name is made up of a family name, which is always placed first, followed by a given name. There are over 700 different Chinese family names, but as few as twenty cover a majority of Chinese people [2]. The variety in Chinese names therefore depends greatly on given names rather than family names. The great majority of Chinese family names have only one character, but there are a few with two. Most Chinese given names have two characters, but some have only one character and few with more than two. The pronunciation of a Chinese name is probably as short as one to two seconds of speech Because of the short duration of its pronunciation, a long list of approaching to millions names will possess very high confusion complexity in speech recognition. If there are 1,000,000 personal names in choices, its perplexity is 1,000,000 per three-syllable word duration, or 100 per syllable duration. This complexity might be analogous to recognition of a three-syllable command over a long list of a million commands. Besides the high perplexity property, there are heavy homonym problems. Hence, its difficulty is surprisingly high if only the spoken full name is observed.

From the above discussion, it is clear that it would be very difficult, even for people, to identify a person among close to a million ones by solely hearing his/her name. In this paper, two innovative approaches are proposed to resolve the heavy homonym problems in Chinese name recognition: Character Description Recognition and Syllable Spelling Recognition. The following two sections explain them. Section 4 describes our interactive strategy for acquiring the information about the name spelling, followed by the section describing experimental setup and result. The conclusion is made in the end.

2. CHARACTER DESCRIPTION RECOGNITION

As mentioned above, the Chinese characters are mostly radical-phonetic or radical-radical compounds. People might describe a Chinese character by its radical or phonetic components. For example, character '李' might be described by "木子李", where '木' and '子' are both radical components of '李'. Some character might be described by their strokes of writing, like character '王' being described as "三橫一豎王" ("three horizontal bars and one vertical bar *wang2*"). Another major category of descriptions is to use a word, a phrase, or a named entity to indicate the target character following the pattern: "詞 (word/phrase/name) 的 ('s) 字 (character)", similar to English pattern: "alphabet as in word". Taking some examples, like "孔子的孔" using a well-known person name "孔子" (Confucius), "香港的香" using a place name "香港" (Hong Kong), and "興高采烈的采" using an idiomatic phrase "興高采烈" (cheery).

We collected words, phrases and name entities from several resources. The major part was extracted from the Academia Sinica Balanced Corpus (ASBC), in which there were more than 200 thousands unique words. For about a hundred high frequency family names, a list of descriptions by eleven persons was collected. A portion of characters was manually collected of their radical components. Some popular personal, location and organization names were also collected for CDR use. A total collection of 90,626 descriptions for 4,615 distinct characters was used in our experiments.

The CDR finite-state recognizer was dynamically composed with a portion of these 90,526 descriptions, according to the character position and dialogue contextual information. It received character description utterances to produce a list of character hypotheses.

3. SYLLABLE SPELLING RECOGNITION

There are totally about 1300 tonal syllables in Chinese. Zhuyin transcription system was adopted in our preliminary study of syllable spelling recognition. It uses 37 special symbols to represent the Mandarin sounds: 21 consonants and 16 vowels. A syllable usually contains two or three Zhuyin symbols, as well as a tone marker. People use to spell Mandarin syllable sounds after the pattern: "Zhuyin symbol sequence followed by syllable followed by tone marker phrase followed by optional syllable" in speaking. The SSR finite-state grammar comprised the above spelling and another faster alternative pattern. Taking syllable "jian3" as an example, the SSR grammar included both "'ㄐ' 'ㄧ' 'ㄢ' 'jian3' '三聲' 'jian3'" and "'ㄐ' 'ㄧㄢ' 'jian3' '三聲' 'jian3'". The latter description contains a compound 'ㄧㄢ' as an alternative way in spelling syllable.

The SSR finite-state recognizer was composed with this kind of tonal syllable-spelling descriptions. It produced a list of syllable hypotheses, with highest scores among all, as the recognition output. While SSR was used to recognize a character, a list of syllable hypotheses might be asked to confirm, followed by the confirmation of the possible character hypotheses – the homonyms of the chosen syllable. Similar approach can be easily applied to SSR for Pinyin transcription system.

Earlier spelled Western word recognition researches can be found in [3][4][5][6][7]. But the authors do not know of spelled Chinese syllable recognition system until this publication.

4. INTERACTIVE CHINESE NAME RECOGNITION

The user of a Chinese real-world directory assistance (DA) service is often asked to speak the person name of enquiry as the first step. The human operator would then confirm the enquired name character by character. Once the confirmation of a character failed, the user would be ask to describe the character. In case of the possible characters are heavily confusable, the operator would probably straightly ask the user to describe it, instead of to confirm a long list of character hypotheses.

We adopt the framework of the above-mentioned strategy used by the human operator, in order to keep the friendliness, for example, in an automatic directory assistance system. Such a strategy implies a combination of at least three major components in our system: full name recognition, character confirmation mechanism, and character description recognition. With additional syllable spelling recognition and syllable confirmation mechanism, we established our first large-vocabulary name recognition system for a Mandarin DA system.

Many researches have been done for large-vocabulary DA systems. Some of them recognized names with joint decision based on both spelled and spoken names [4] [5] [6] [7]. Most of them started dialogues by asking the spelling of the last name, differing from most of the human operating services. Our approach experimented dialogue strategies starting from asking spoken names, same as in [4], and might provide friendlier experience to the user. Some concentrated on the spoken name recognition itself and improved it by, for example, rescoring with alternative pronunciations [8], speaker adaptation [10] or confidence measurement [11].

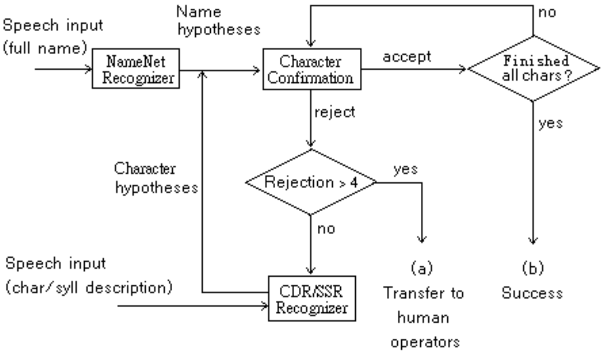


Figure 1. Illustration of the interactive Chinese name recognition procedures

In this paper, a robust and friendly interaction strategy of name recognition procedure, illustrated as Figure 1, was proposed and simulated a human operating DA system. A graph-structured toneless-syllable finite-state recognizer of names, called "Name-Net" recognizer in the paper, was performed in the first stage to provide a list of name hypotheses, given the spoken full name utterance. Character-wise confirmation scheme followed and separated the uncertain characters from the certain ones. Once all characters were confirmed correctly, it finished the whole recognition process. Otherwise, the rejected characters from the confirmation process were transferred to a back-off recognition process as in the below of the Figure 1, which was composed by another finite-state recognizer with CDR and SSR grammars in parallel. System would guide the user to speak for either the character description or the syllable spelling freely to find the most possible character hypotheses, which were as usual sent to the character confirmation process mentioned earlier. After three times failure of character recognition by CDR or SDR, the system would simulate call transfer to a human operator and treat it as a failure of name recognition dialogue.

In general, our framework of name recognition was built mostly based on the well-known problem solving approach "divide-and-conquer" and divided a bigger problem of Chinese name recognition into a sequence of sub-problems, recognition of three characters, which are more solvable with the help of CDR and SSR. Besides recognition, confirmation of names was also divided into confirmation of three characters. The unfriendliness of selecting a correct character from the bottom of a long list of hypotheses was largely reduced in many cases. For example, once there were 27 hypotheses to confirm until reaching the correct one, the user would probably be asked three times around three choices of characters instead of being asked of 27 choices of full name if under the character-wise scheme.

CDR and SSR are described in the above section. The following sub-sections will explain our personal name database and a graph-structured full name finite-state recognizer. The reason to use a parallel combination of CDR and SSR was mostly according to user friendliness concern. Normally, for a commonly used character the user might speak a well-known character description pattern, for example a word or a phrase with the character in it. Otherwise, for an uncommonly used character the user might choose to spell its syllable pronunciation as a simpler back-off description way.

4.1. Personal-Name-533K database

There are more than a billion of Chinese in the world. They are currently the biggest population in the world. The recognizing a Chinese name is probably one of the most difficult tasks. To initiate the study on it, we collected a list of Chinese names from more than five hundred thousand persons, who attended the joint entrance examinations of the universities in Taiwan during years 1994 to 2001. For simplification, the subset of all three-character names was defined as the "Personal-Name-533K" database for our preliminary study. There were totally 533,179 personal names, or unique 326,594 character strings. Its detailed statistics are given in Table 1. As the first characters treated as the family names and the other two characters treated as the given names for simplification, there were totally 562 unique family names, or corresponding 266 unique toneless syllables, and 133,116 unique given names, or corresponding 31,929 unique two-toneless-syllable strings.

Table 1. Statistics of Personal-Name-533K database

	#Person	#Char	#TL_Syll
Full names	533,179	326,594	245,510
Family names	533,179	562	266
Given names	533,179	133,116	31,929

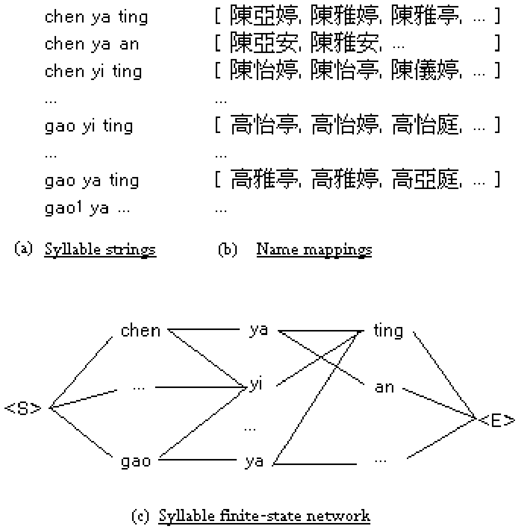
Their character perplexities of the first, second, and third characters are listed in the Table 2, respectively. The "left-context-dependent" perplexities are given in the third column, which are the one given the knowledge of the earlier acquired character(s), a kind of system belief information. In our experiments, we utilized this knowledge by dynamically building sub-grammars for the recognition of the second and third characters. It helped in narrowing down the possible hypotheses and enhancing the recognition accuracies. Especially for the third character, a dramatic reduction of character perplexity from 416.2 to 19.0 was observed. Sometimes, as the possible choices of character were few, the system would straightly ask the user to select it from a list, instead of to speak its descriptions.

Table 2. Character perplexity analyses of name database

	Context-indep	Left-cntxt-dep
1st character	48.4	-
2nd character	349.1	234.8
3rd character	416.2	19.0

4.2. Graph-structured syllable network

Given a list of Chinese personal names, its corresponding syllable transcriptions could be generated automatically, even though a few percentages of these characters may have multiple pronunciations. The syllable strings and corresponding name mappings are shown in Figure 2(a) and 2(b). The syllable strings of names could be automatically arranged as a finite state network, as Figure 2(c), by applying a minimizing procedure, in which syllable nodes with the same left or right context could be shared. For the reason of computational consideration, we chose to build a toneless graph-structured syllable network, named "Name-Net" recognizer in the paper, instead of a tonal one. Earlier researches have experimented syllable-sized units in their recognizers and observed robust performance against phone units [7] [8] [9], especially for large vocabulary experiments [9]. Compared to word units, in the other hand, syllables provide reasonable sized inventory with coverage over the majority of unseen names in English [7].



(a) Syllable strings (b) Name mappings

(c) Syllable finite-state network

Figure 2. Graph-structured syllable network of names

5. EXPERIMENTS

In the experiments, the Delta-Electronics Finite-State-Network speech recognizer was used, which consisted an HMM based acoustic model with context-dependent phones and an FSN based language model. Standard MFCC features with their first derivatives, 26 dimensions in total, and a CMS process for noise reduction were applied. Mandarin telephony speech corpora MAT-2000 [12] were used to train the acoustic model. A Delta-Electronics rule-based Text-To-Speech component was used for automatic response generation [13].

Two hundred testing names were randomly selected from the Personal-Name-533K database. They were separated equally into ten sets, each set for a tester, to evaluate the DA-like name recognition dialogue system. It should be noticed that the testers were mostly unfamiliar with these spoken names, contrary to a real-world system condition that people inquire their familiar names. A notebook computer with Pentium M 1.4G Hz processor and 768 MB RAM was used to simulate the telephone system in the preliminary work. Each tester interacted with the system via a microphone and a keyboard, in which six buttons were used to confirm characters, and inquired his/her given twenty names. The system performed real-timely except as some dynamically building of finite-state network for the left-context-dependent CDR and SSR.

The interaction between a real user and the system was highly dependent on the user. Some tended to use character confirmation list to get the correct one, some tended to use CDR as soon as possible, and some tended to use SSR. The result of testing on the 200 names by the ten testers is displayed in Table 3. The second row shows the actual success rate of the system. The third row counts successes for those cases by Name-Net and CDR, but neither SSR (counted as failures for those cases of using SSR in their dialogues). The fourth row counts successes for those using only Name-Net, but neither CDR nor SSR.

Table 3. Name recognition success rates

	Success Rate
Name-Net + CDR + SSR	94.5%
Name-Net + CDR	88.5%
Name-Net only	55.5%

Table 4. Success rates in counts of character

	Name-Net	CDR	SSR
#Character	600	128	26
Success Rate	79%	67.2%	73.1%

Table 4 displays the character success rates by Name-Net, CDR, and SSR, respectively. While all 600 characters had been recognized by Name-Net, 79% of them had successfully acquired via confirmation process, without the help of either CDR or SSR. CDR had been performed 128 times as a back-off recognition process, in which 67.2% successfully acquired the character. SSR had been performed only 26 times, in which 73.1% successful. Our earlier experiments also showed similar results for them: character success rate of CDR was 70.4% over 304 trial cases, while that of SSR was 72.5% over 149 cases.

6. CONCLUSIONS

The paper proposed a character-wise divide-and-conquer approach in recognition of a Chinese name among over three hundred thousand unique names. The character description recognition approach provided a robust and friendly way to recognize the precise character, as its monosyllabic pronunciation, less than 0.5-second speech in average, has a lot of confusable homonyms. The syllable spelling recognition further enhanced the robustness and friendliness of the overall system, especially when the user could not pick up a description of a character easily. The resultant online evaluation gave a very positive performance of up to 94.5% name recognition success rate. Besides, we plan to improve the system to approach free name recognition and to do more testing on the improved version of the system.

7. ACKNOWLEDGEMENTS

We would like to thank all colleagues in the speech team of Delta Electronics R&D Center for their grateful support to our development of the name recognition system. Especially, Mr. Gary Wang helped us a lot in our initial phase. Thanks!

8. REFERENCES

[1] "Chinese character," http://en.wikipedia.org/wiki/Chinese_character, in the website of Wikipedia, The Free Encyclopedia, July 2004.

[2] "Chinese name," http://en.wikipedia.org/wiki/Chinese_name, in the website of Wikipedia, The Free Encyclopedia, July 2004.

[3] Hild, H. & Waibel, A., "Recognition of Spelled Names over the Telephone," *Proc. ICSLP,* Philadelphia, October 1996.

[4] Meyer, M., & Hild, H., "Recognition of Spoken and Spelled Proper Names," *Proc. EUROSPEECH,* Rhodes, September 1997.

[5] Seide, F. & Kellner, A., "Towards an Automated Directory Information System," *Proc. EUROSPEECH*, Rhodes, September 1997.

[6] Kellner, A., Rueber, B., & Schramm, H., "Strategies for Name Recognition in Automatic Directory Assistance Systems," *Proc. IVTTA*, Torino, September 1998.

[7] Suchato, A., "Framework for Joint Recognition of Pronounced and Spelled Proper Names," *MS Thesis,* MIT Department of Electrical Engineering and Computer Science, Boston, September 2000.

[8] Bechet, F., DeMori, R., & Subsol, G., "Very Large Vocabulary Proper Name Recognition for Directory Assistance," *Proc. ASRU,* Madonna di Campiglio, December 2001.

[9] Sethy, A., Narayanan, S., & Parthasarthy, S., "A Syllable Based Approach for Improved Recognition of Spoken Names," *Proc. PMLA,* Estes Park, Colorado, September 2002.

[10] Gao, Y.Q., Ramabhadran, B., Chen, J., Erdogan, H., & Picheny, M., "Innovative Approaches for Large Vocabulary Name Recognition," *Proc. ICASSP,* Salt Lake City, Utah, May 2001.

[11] Liao, Y.F. & Rose, G., "Recognition of Long Lists of Chinese Names Uttered Within Whole Utterance," *Proc. ICASSP,* Orlando, Florida, May 2002.

[12] Wang, H. C., Seide, F., Tseng, C. Y. and Lee, L.-S., "MAT-2000 – Design, Collection, and Validation of a Mandarin 2000-Speaker Telephony Speech Database", *Proc. ICSLP,* Beijing, October 2000.

[13] Liau, W. & Shen, L., "Improved Prosody Module in a Text-To-Speech System", *submitted to Conference on Computational Linguistics and Speech Processing XVI,* Taipei, September 2004.