OPEN VOCABULARY CHINESE NAME RECOGNITION WITH THE HELP OF CHARACTER DESCRIPTION AND SYLLABLE SPELLING RECOGNITION

Ching-Ho Tsai[2], *Nick J.C. Wang*[1, 2], *Patrick Huang*[2] *and Jia-Lin Shen*[2]

Graduate Institute of Communication Engineering, National Taiwan University[1] &
Department of Man-Machine Interface, R&D Center, Delta Electronics Inc.[2],
Taipei, Taiwan, R.O.C.
nick.jc.wang@delta.com.tw

ABSTRACT

The open vocabulary name recognition technique is one of the very challenging tasks in the application of automated Chinese speech recognition technology. It can be applied on free name input for any telephony speech application and automatic directory assistance systems. A Chinese name usually has two to three characters, with each pronounced as a single syllable. It is a high perplexity task to recognize a word from a long-list of candidates, like more than billions of names in our experiments. Two novel approaches under an interactive framework were proposed in our previous paper for five hundred thousand names and applied in the paper for open vocabulary Chinese name recognition: 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, 86.7%, 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 (for example, ‘媽’ in Chinese *ma1* is a combination of ‘女’ a radical and ‘馬’ in Chinese *ma3*) or radical-radical compounds (for example, ‘信’ is a combination of ‘人’, a person, and ‘言’, speech). In addition, a Chinese 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 (tones). Each character is pronounced as a single tonal syllable. There are about 408 base syllables, which can be voiced with four major tones, besides the neutral tone (輕聲): high-level tone (陰平), rising tone (陽平), low-falling-raising tone (上聲), and falling tone (去聲). 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 consisted 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. Most Chinese family names have only one character, but a few with two. Most Chinese given names have two characters, but a few of them have one or 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, a long list of approaching to billions of names will possess very high confusion in speech recognition. If there are 1,000,000,000 personal names for choice, its perplexity could be more than 1,000,000 character sequences per three-syllable word duration, or more than 100 characters per syllable duration. This complexity might be analogous to recognition of a three-syllable command over a long list of a billion 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 billion 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, which are explained in the following two sections. Section 4 elaborates our interactive strategy for acquiring the information about the name spelling. After that comes the description of the 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).

In the human operated Chinese name recognition system, like in the popular 104-directory assistance service provided by Chunghwa Telecom in Taiwan, after speaking a Chinese name, people are used to describe the character when encountering homonym problem on Chinese character identification.

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 frequently used 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 about 60,000 descriptions for 4,615 distinct characters was used in our experiments. Two CDR finite-state recognizers were composed of descriptions for about 600 family name characters and for 4,615 given name ones, respectively.

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 one to three Zhuyin symbols and a tone marker. People are used 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 comprises 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.

Two SSR finite-state recognizers were composed of tonal syllable-spelling descriptions for about 600 family name characters and for 4,615 given name ones, respectively. 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 have not heard of the spelled Chinese syllable recognition system before our publications.

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 asked to describe the character. In case 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.

In order to keep the friendliness, we adopt the framework of the above-mentioned strategy used by the human operator 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 Chinese free name recognition system for a Mandarin DA system. To our knowledge, the system is the first one of naturally speaking open-vocabulary name recognition systems over all languages.

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].

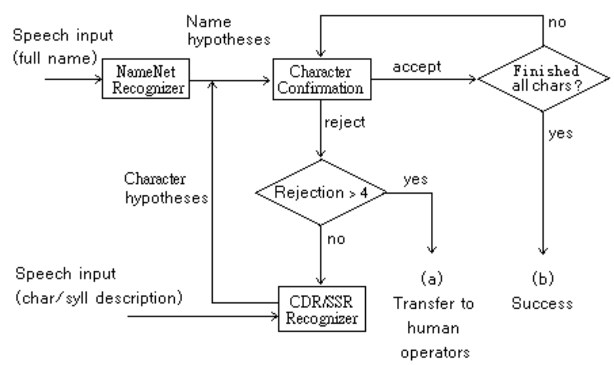

Figure 1. Illustration of the interactive Chinese name recognition procedures

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. After that, a rescoring scheme for tone recognition based on the syllable string hypotheses was performed. 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 backing-off recognition process as in 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. A database of more than five hundred thousand names was used for the training of the language probabilities of the NameNet FSN and the syllable-to-character generation process.

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, the recognitions 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 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, if under the character-wise scheme, the user would probably be asked three times around three choices of characters instead of being asked of 27 choices of full name. Hence the efficiency of it was probably not worse than the confirmation of the whole name. Besides, the heavy homonym problem in Chinese makes it impossible for distinctive confirmation of the names without describing the characters.

CDR and SSR are described in the above section. The following sub-sections will explain our database of personal names 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, such as 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 backing-off description way.

4.1. Personal-Name-533K database Analysis

There are more than a billion of Chinese in the world. They are currently the biggest population in the world. With so many names to recognize a Chinese name with open vocabulary would be challenging. 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 from 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 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 if further constraints on the database could be used.

Table 2. Character perplexity analyses on the database of Persoanl-Name-533K

	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

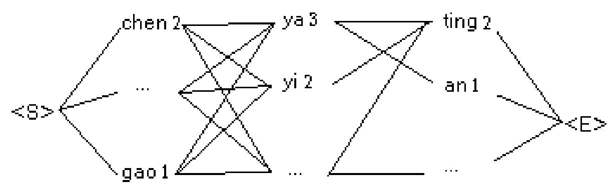

Figure 2. Graph-structured syllable network of names

The tonal syllable graph was constructed with 437 ones for family names and 67,599 double-syllable strings for given names, as illustrated in Figure 2, named "Name-Net" recognizer in the paper. It contained around thirty millions possible name syllable string hypotheses, mapped to more than ten billions of character string hypotheses, covering almost all possible personal names.

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].

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 ETSI robust feature extraction process [12] for noise reduction were applied. Mandarin telephony speech corpora MAT-2000 [13] were used to train the acoustic model. A Delta-Electronics rule-based

Text-To-Speech component was used for automatic response generation [14].

Testing set with 316 name dialogues was performed freely for around 80 testers to speak their familiar person 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 eight buttons were used to confirm characters. The system performed as in real-time.

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 316 names is displayed in Table 3.

The last row shows the actual success rate of the entire system with NameNet recognition and backing-off CDR/SSR recognitions, 86.7%. The second row counts successes for those cases by Name-Net only, 35.1%, while the third row counts successes for those backing-off CDR/SDR recognitions of at most three times, 79.5%.

Table 3. Name recognition success rates analyses

	Success Rate
Name-Net only	35.1%
Backing-off CDR/SSR 3 times	79.5%
Entire system	**86.7%**

Table 4. Character success rates analyses

	Name-Net	CDR	SSR
#Character	944	410	94
Percentage	100%	44%	10%
Success Rate	78.3%	54.6%	60.6%

In the dialogues of our experiments, when the NameNet Recognition fails to recognize any of the spoken characters, the use of CDR and SSR usually can resolve the problem, and the overall success rate could be highly improved, which makes the system more accepted by the users.

Table 4 displays the character success rates by each single process of Name-Net, CDR, and SSR, respectively. During the experimental dialogue for Chinese name recognition, while all 944 characters had been recognized by Name-Net, 78.3% of them had successfully acquired via confirmation process, without the help of either CDR or SSR. Once a character was failed to recognize and confirm, it will be sent to backing-off CDR/SSR process. Hence, the name recognition via NameNet only was counted as failure in Table 3. CDR had been performed 410 times for less than 44% of the characters during the experimental dialogues as a backing-off recognition process, in which 54.6% successfully acquired the character. SSR had been performed only 94 times for around than 10% of the characters, in which 60.6% successful. Our earlier experiments with constraints on Person-Name-500K showed better results due to lower perplexity [15].

6. CONCLUSIONS

The paper proposed a character-wise divide-and-conquer approach to deal with open-vocabulary Chinese name recognition task for more than billions of possible 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 86.7% name recognition success rate.

7. ACKNOWLEDGEMENTS

We would like to thank all colleagues in the speech team of Delta Electronics R&D Center for their great support to our development of the name recognition system. 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] ETSI, "Speech processing, transmission and quality aspects (STQ); distributed speech recognition; advanced front-end feature extraction algorithm; compression algorithms", Tech. Rep. ETSI ES 202 050, ETSI, 2002.

[13] 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.

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

[15] Wang, N., Tsai, C.H., Huang, P., & Shen, J.L., "Chinese Large-Vocabulary Name Recognition System using Chracter Description and Syllable Spelling Recognition", submitted to *Proc. ISCSLP,* 2004.