Speed Compensation for Improving Thai Spelling Recognition with a Continuous Speech Corpus

Chutima Pisarn and Thanaruk Theeramunkong

Sirindhorn International Institute of Technology,
131 Moo 5 Tiwanont Rd., Bangkadi, Muang, Phathumthani 12000, Thailand
{chutimap, thanaruk}@siit.tu.ac.th

Abstract. Spelling recognition is an approach to enhance a speech recognizer to cope with incorrectly recognized words and out-of-vocabulary words. This paper presents a general framework for Thai speech recognition, enhanced with spelling recognition. To implement Thai spelling recognition, Thai alphabets and their spelling methods are analyzed. Based on hidden Markov models, we propose a method to construct a Thai spelling recognition system using an existing continuous speech corpus. To compensate for speed differences between spelling utterances and continuous speech utterances, the adjustment of utterance speed is taken into account. Our system achieves up to 87.37% correctness and 87.18% accuracy with the mix-type language model.

1 Introduction

Currently several works on automatic speech recognition (ASR) for continuous speech are undergoing development, for systems that rely on dictionaries and those that can recognize out-of-vocabulary circumstances. In the situation of misrecognition and out-of-vocabulary words, a practical and efficient solution to assist the ASR is to equip a system with a spelling recognition subsystem, in which users can spell out a word, letter by letter. Spelling recognition is a challenging task with high interest for directory assistance services, or other applications where a large number of proper names or addresses are handled. Many works that focused on spelling recognition were widely developed in several languages, for instance, English, Spanish, Portuguese and German. In [1], hypothesis-verification Spanish continuous spelled proper name recognition over the telephone was proposed. Several feature sets were investigated in models of neural networks. In their succeeding work [2], three different recognition architectures, including the two-level architecture, the integrated architecture and the hypothesis-verification architecture, are analyzed and compared. In [3], a Portuguese subject-independent system for recognizing an isolated letter was introduced. The system is simulated to recognize speech utterances over a telephone line using the Hidden Markov Model (HMM). A number of experiments were made over four different perplexity language models. In [4], Mitchell and Setlur proposed a fast list matcher to select a name from the name list that was created from an *n*-best letter recognizer on spelling over the telephone line recognition task. In [5], an

A. Aagesen et al. (Eds.): INTELLCOMM 2004, LNCS 3283, pp. 100–111, 2004.

integration approach was proposed to combine word recognition with spelling recognition in a user-friendly manner as a fall-back strategy. As a German city name recognizer, the system was applied to directory assistance services.

Unlike other languages, spelling in Thai has several styles. One of them is similar to spelling in the English language, i.e., */h-@@4//m-@@0//z-aa0/* of "หมา" corresponding to */d-ii0//z-oo0// g-ii0/* for "dog". There are three additional methods in Thai spelling, where some syllables are inserted to make it more clear for the hearer. One method is to spell out a letter followed by its representative word's utterance. Another method is to mix between the former two types. The third method is to spell out a set of letters that form a syllable followed by its corresponding pronunciation. Thus far spelling recognition for Thai language has not been explored. One of main reasons is that there is no standard corpus for this purpose. Creating a corpus of spelled utterances is a time consuming task. Fortunately we have a normal speech corpus. In this work, we use the NECTEC-ATR Thai Speech Corpus, a standard continuous Thai speech corpus, for our spelling recognition system. Another objective of this work is to examine how a spelling system can be implemented using a normal Thai continuous speech corpus. That is, as the preliminary stage, we investigate the performance of spelling recognition using such an existing corpus.

This paper is organized as follows. In section 2, language characteristics in Thai are introduced. Section 3 presents our recognition framework. Four spelling styles for Thai words are discussed in section 4. The experimental results and their analysis are shown in section 5. Finally, a conclusion and our future works are given in section 6.

2 Thai Language Characteristics

2.1 Thai Alphabets

Theoretically, the Thai language has 69 alphabet symbols which can be grouped into three classes of phone expression; consonant, vowel and tone. There are 44, 21, and 4 alphabet symbols for consonants, vowels, and tones, respectively. Some Thai consonant symbols share the same phonetic sound. Because of this, there are only 21 phones for Thai consonants. On the other hand, some vowels can be combined with other vowels, resulting in 32 possible phones. However, in practice, only 18 alphabet symbols in the vowel class are used. There are 5 tones in Thai, including one without an alphabet symbol. Concludingly, there are 66 practical alphabet symbols as shown in Table 1.

Table 1. Thai alphabets: consonants, vowels and tones

Basic Classes	Alphabets in each class
Consonant	ก,ข,ฃ,ค,ฅ,ฆ,ง,จ,ฉ,ช,ซ,ฌ,ญ,ฎ,ฏ,ฐ,ฑ,ฒ,ณ,ด,ต,ถ,ท,ธ,น,บ,ป,ผ,ฝ,พ,ฟ,ภ,ม,ย,ร,ล,ว,ศ,ษ,ส,ห,ฬ,อ,ฮ
Vowel	อั, อะ, อิ, อา, อิ, อี, อึ, อื, อุ, อู, เอ, แอ, โอ, อำ, ไอ, ใอ, ฤ, อ์
Tone	อ่, อ้, อี๊, อ๋

2.2 Thai Syllable Characteristics and Phonetic Representation

In the Thai language, a syllable can be separated into three parts; (1) initial consonant, (2) vowel and (3) final consonant. The phonetic representation of one syllable can be expressed in the form of $/C_i\text{-}V^T\text{-}C_f/$, where C_i is an initial consonant, V is a vowel, C_f is a final consonant and T is a tone which is phonetically attached to the vowel part. Following the concept presented in [6], there are 76 phonetic symbols and 5 tone symbols applied in this work as shown in Table 2.

Table 2. Phonetic symbols grouped as initial consonants, vowels, final consonants and tones

Initial Consonant (C_i)		Vowel (V)	Final Consonant (C_f)	Tone (T)
Base	Cluster			
p,t,c,k,z,ph, th,ch,k,h,b, br,bl,d,dr,m, n,ng,r,f,fr,fl, s,h,w,j	*pr,phr,pl,phl ,tr,thr,kr,khr, kl,khl,kw, khw*	*a,aa,i,ii,v,vv,u, uu,e,ee,x,xx,o, oo,@,@@,q, qq,ia,iia,va, vva,ua,uua*	*p^,t^,k^,n^,m^,n^, g^,j^,w^,f^,l^,s^, ch^,jf^,ts^*	0 Mid 1 Low 2 Falling 3 High 4 Rising

Some initial consonants are cluster consonants. Each of them has a phone similar to that of a corresponding base consonant. For example, *pr*, *phr*, *pl*, and *phl* are similar to *p*. Naturally phones, especially those in the vowel class, are various in their duration. In Thai language, most vowels have their pairs: a short phone and its corresponding long phone. For example, the vowel pair *a* and *aa* have a similar phone but different durations. The other vowel pairs are *i-ii, v-vv, u-uu, e-ee, x-xx, o-oo, @-@@, q-qq, ia-iia, va-vva,* and *ua-uua*.

3 Our Framework

Figure 1 presents our recognition framework designed for a Thai continuous speech recognition system that incorporates a conventional recognizer with a spelling recognition subsystem. The whole process can be divided into two modules; (1) training module and (2) recognition module.

In the training module, waveforms of continuous speech utterances in a corpus are transformed to feature vectors by a signal quantization technique. The derived feature vectors are used for training a set of acoustic models. In the system, two language models are equipped; one model stands for traditional word recognition whereas the other is used for spelling recognition. The traditional language model is trained by transcriptions in the text corpus while the spelling language model is trained by sequences of letters in a proper name corpus.

In the recognition module, the two well-trained models, the acoustic model and the traditional language model, together with a pronunciation dictionary are applied to

recognize a new utterance yielding a set of hypothesis results. Each hypothesis candidate is then checked to determine whether this hypotheis is valid or not. If all hypothesse are invalid, the system will turn to the spelling recognition subsystem.

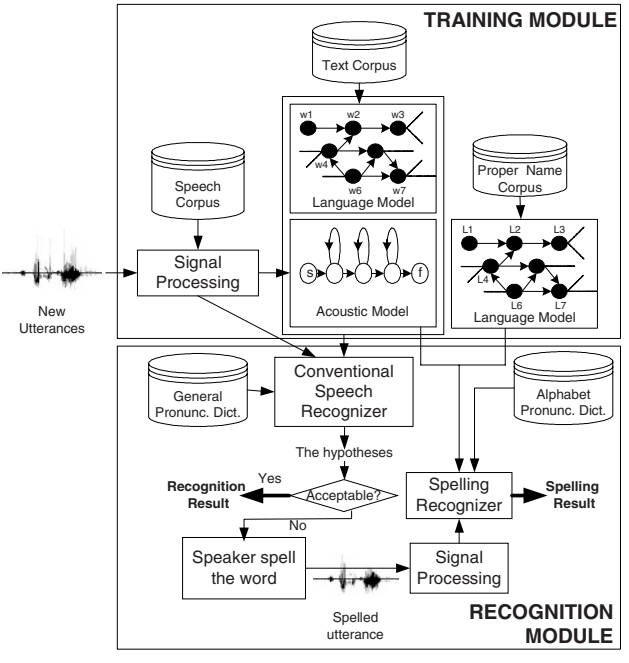

Fig. 1. Our Recognition framework

At this stage, the user is asked to spell the word letter-by-letter. The utterance of spelling is then fed to the signal-processing module to convert the waveform to feature vectors. In this work, as our preliminary stage, we focus on the spelling recognition subsystem. We use the acoustic models trained by normal continuous speech utterances because we lack a spelling corpus. Incorporating the acquired acoustic models with a trained spelling language model and an alphabetic pronunciation dictionary, spelling results can be obtained.

4 Spelling Styles for Thai Words

4.1 Basic Pronunciation of Thai Alphabets

As referred in section 2.1, there are three basic classes of Thai alphabet symbols. Pronouncing Thai alphabet symbols in different classes results in different styles. The consonant class alphabet symbols can be uttered in either of the following two styles. The first style is simply pronouncing the core sound of a consonant. For example, for the alphabet symbol 'ก', its core sound can be represented as the syllable phonetic

/k-@@0/. Normally, some consonants share the same core sound such as '**ค**', '**ค**', '**ฆ**' have the same phonetic sound */kh-@@0/*. In such a case, the hearer may encounter alphabet ambiguity. To solve this issue, the second style is generally applied by uttering a core sound of the consonant followed by the representative word of that consonant. Every consonant has its representative word. For example, the representative word of '**ก**' is "**ไก**" (meaning: "chicken", sound: */k-a1-j^/*), and that of '**ข**' is "**ไข**" (meaning: egg, sound: */kh-a1-j^/*). To express the alphabet '**ก**' using this style, the sound */k-@@0/+/k-a1-j^/* is uttered.

Expressing symbol in the vowel class is quite different from that of the consonant class. There are two types of vowels. First-type vowels can be pronounced in two ways. One way is to pronounce the word "**สระ**" (meaning: "vowel", sound: */s-a1/ /r-a1/*), followed by the core sound of the vowel. The other way is to simply pronounce the core sound of the vowel. For second-type vowels, they are uttered by calling their names. The vowel symbols of each type are listed in Table 3. As the last class, tone symbols can be pronounced by calling their names.

Table 3. Two types of vowels

Type	Vowels
The first-type	อะ, อา, อิ, อี, อึ, อื, อุ, อู, เอ, แอ, โอ, อำ, ไอ, ใอ
The second-type	อั, อี่, อ์, ฤ

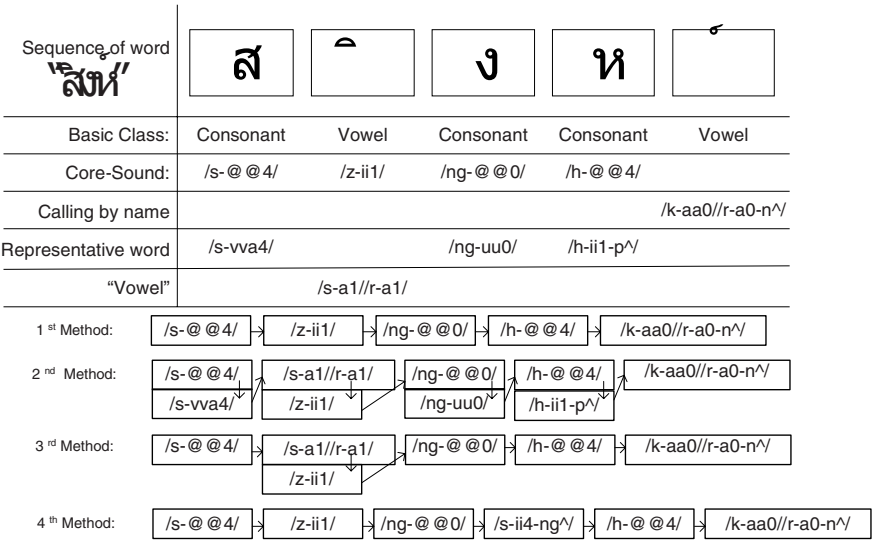

Fig. 2. Four Spelling Methods for the word "สิงห์"

4.2 Thai Word Spelling Methods

Spelling a word is done by uttering alphabet symbols in the word one by one in order. We can refer spelling to a combination of the pronunciation of each alphabet symbol in the word. Only four Thai commonly used spelling methods are addressed. For all methods, the second-type vowels and tones are pronounced by calling their names. The differences are taken place in spelling consonants and the first-type vowels. For the first spelling method, consonants are spelled by using only their core sounds, and first-type vowels are pronounced by their core sound without the word "**สระ**" (/s-a1/ /r-a1/). This spelling method is similar to spelling approach in English language.

For the second method, the representative word of each consonant is pronounced, after its core sound while pronouncing a first-type vowel is to utter the word "**สระ**" and then its core sound. In the third method, the way to pronounce a consonant and a vowel are varied. For instance, the word can be spelled by spelling a consonant using only its core sound but spelling a vowel by pronouncing "**สระ**"(/s-a1//r-a1/) with the vowel's core sound. The last method is to spell out a set of letters that form a syllable and then followed by its corresponding pronunciations. The spelling sequence of alphabets in each syllable starts with initial consonant, vowel, and followed by final consonant (if any) and tone (if any), and then, the sound of that syllable is inserted at the end of this sequence. The examples of these methods in spelling the word "**สงห**" are depicted in Figure 2. Because the second method is the prevalant spelling method in Thai, we concentrate an effort on this method.

5 Experimental Results and Analysis

5.1 Experimental Environment

As mentioned, unfortunately, the corpus for spelling recognition is not available at this time. Therefore, this work applies the NECTEC-ATR Thai Speech Corpus, constructed by NECTEC (National Electronics and Computer Technology Center) incorporated with ATR Spoken Language Translation Laboratories. In Thai language speech recognition, this corpus is often used for continuous speech recognition. This speech corpus is used as the training set for our spelling recognition system. The corpus contains 390 sentences gathered by assigning 42 subjects (21 males and 21 females) to read all sentences for one trial. Thus, there are 16,380 read utterances in total.

At the first place, by the reason of computation time, only utterances of 5 males and 5 females, are used, i.e., totally 3,900 trained utterances. In this work, the performance of spelling recognition using a normal continuous training corpus is investigated. Even when the training corpus has quite different characteristics compared to test utterances, we can expect a reasonable result. For the test utterances, we record 136 spelled proper names pronounced by six other subjects. These 136 proper names are shop names, company names, family names and first names.

The speech signals were digitized by 16-bit A/D converter of 16 kHz. A feature vector used in our experiment is a 39-feature vector, consisting of 12 PLP coefficients and the 0^{th} coefficient, as well as their first and second order derivatives. Therefore, there are 39 elements in total.

There are three bigram language models used in this task; LM1, LM2 and LM3. LM1 is a close-type language model trained using 136 proper names from the test transcription. LM3 is an open-type language model trained using 5,971 Thai province, district, and sub district names. Since LM2 is a mix-type language model, a mixture of the two models, where both 146 proper names and 5,971 location names are used. In this work, we will focus on LM2.

A phone-based HMM is applied as the recognition system. The acoustic units used in this experiment are defined in the same manner as in [6]. All experiments, including automatic transcription labeling, are performed using the HTK toolkit [7]. We evaluate the recognition performance in the terms of correctness and accuracy. The word correctness is the ratio of the number of correct words to the total number of words while the word accuracy is the ratio of the number of correct words subtracted by the number of word insertion errors, to the total number of words. For detail of correct and accuracy, see [7].

5.2 Setting a Baseline

In the first experiment, we investigate the spelling results using the original training and testing data as they are. Using the phone-based HMM, all experiments are performed with context independent considerations. Context-independent means that the recognition of a certain phone does not depend on the phone's preceding or following phones. In this initial stage, for LM2, we can gain 83.38% correctness and 73.28% accuracy, respectively. The low accuracy indicates that there are a large number of insertion errors. We also analyzed the errors and found that many spelling results violated the applied language model (bigram model). Because of this, the weight ratio between the acoustic model and the language model is set to be a low value, forcing the language model more to be important than the acoustic model. The results in the cases that the weight is set to 0.1 and 0.2 gain the highest recognition. However, the weight ratio of 0.1 gains more recognition rate than that of 0.2 in most case. As a result, 85.71% correctness and 85.26% accuracy derived from the LM2 language model with the weight ratio of 0.1 becomes our baseline through this work. The results of various weight ratios are shown in Table 4.

Table 4. The recognition results of the baseline with various weight ratios

Weight ratio		1.0	0.2	0.1	0.05
LM1	Correctness	83.47	90.70	93.08	90.04
	Accuracy	73.87	88.86	92.89	88.35
LM2	Correctness	83.38	86.22	85.71	74.38
	Accuracy	73.28	84.17	85.26	73.92
LM3	Correctness	83.32	85.74	84.03	73.63
	Accuracy	72.79	83.06	83.33	73.07

5.3 Adjusting the Duration

The major difference between the training and the test sets is the duration of the utterances. The speeds of training and test utterances are measured in the form of the number of phones per second. The result indicates the speed of the test set is approximately 1.5 times slower than that of training utterances. To compensate for this duration difference between the training utterance and the test utterance, the time-stretching method [8], [9], [10], a method to stretch a speech signal, by preserving pitch and auditory features of the original signal, is applied in our signal preprocessing. Stretching the training utterances is performed using various scaling factors in order to investigate the effectiveness. Table 5 shows the recognition results of stretched training utterances with various scaling factors. Here, the original test utterances are used.

For all scaling factors, LM1 gains the highest recognition rate while LM3 obtains the lowest one. In principle, stretching training utterances causes the original utterances to be distorted. The more the utterances are stretched, the more distorted utterances we obtain. As a result, stretching train utterances to be 1.25 times of the original one yields the highest recognition rate while stretching them with 1.43 and 1.67 scaling factor causes the recognition rate to drop. The results show that 1.25Train gains higher correctness and accuracy for every language model. The recognition rate of LM2 are 87.37% correctness and 87.18% accuracy, which are improvements of 1.66% and 1.92%, respectively, compared to the baseline.

Table 5. Recognition results of stretched training utterances with various scaling factors

Model		1.25Train	1.43Train	1.67Train
LM1	Correctness	93.92	91.79	84.39
	Accuracy	93.76	91.65	84.01
LM2	Correctness	87.37	85.37	77.47
	Accuracy	87.18	85.19	76.94
LM3	Correctness	85.75	83.84	76.03
	Accuracy	85.41	83.38	75.37

5.4 Each Subject Test Utterance

The recognition results shown in Table 5 are performed using all six subjects' test utterances. We also investigate the recognition rate of an individual subject. The recognition result and the spelling speed of each subject are shown in Table 6. Note that the test utterances of different subjects have different speeds.

From Table 6, correctness and accuracy of all subjects are not very different. However, the spelling speech affects recognition performance. We observe that FS3 has the slowest spelling speed and we can obtain the lowest accuracy from FS3's experiment. This is caused by a relatively high difference between FS3's spelling speed and the average speed of the training utterances. To handle this issue, the two experiments are performed; (1) using stretched training utterances and 2) shrinking the test utter-

ances to investigate the appropriate scaling factor of this subject. Table 7 and Table 8 indicate the recognition rate of FS3's test utterances in the environments of stretching training utterances experiment and shrinking test utterances.

In the case of using various scaling factors to stretch the training utterances, stretching train utterances to be 1.25, 1.43 and 1.67 times of the original utterances of FS3 outperforms the FS3's baseline for all language models. The 1.43Train achieves the highest recognition rate while stretching with the 1.67 scaling factor causes the recognition drop down. The results show that 1.43Train with LM2 gain 87.33% correctness and 86.70% accuracy, which results in correctness and accuracy gain of 5.98% and 12.60%, respectively, compared to the FS3's baseline results.

Table 6. The baseline results of each subject's test utterances and their spelling speeds

Subject	LM1		LM2		LM3		Speed (Phones/Sec.)
	Corr	Acc	Corr	Acc	Corr	Acc	
FS1	94.51	94.44	87.12	86.84	85.50	85.15	6.72
FS2	94.23	94.09	86.00	85.86	84.45	84.17	6.87
FS3	87.84	86.77	81.98	80.23	80.93	78.54	5.07
MS1	93.67	93.67	85.57	85.22	84.52	84.24	6.51
MS2	93.31	93.24	86.14	86.60	83.74	83.32	5.59
MS3	95.29	95.14	87.47	87.40	85.01	84.59	6.45

Table 7. Recognition of FS3 using stretched training utterances with various scaling factors

Stretch Training Set		1.25Train	1.43Train	1.67Train
LM1	%Correct	91.77	92.75	90.29
	Accuracy	91.34	92.19	89.23
LM2	%Correct	86.14	87.33	84.80
	Accuracy	85.43	86.70	83.81
LM3	%Correct	84.80	85.86	82.97
	Accuracy	83.81	84.24	81.28

Table 8. Recognition of shrinking FS3's utterances with various scaling factors

Shrinking Test Set		0.8Test	0.7Test	0.6Test
LM1	%Correct	92.12	92.89	92.26
	Accuracy	91.77	92.61	92.05
LM2	%Correct	85.38	86.42	84.38
	Accuracy	84.38	85.86	84.10
LM3	%Correct	83.95	83.81	83.53
	Accuracy	83.11	83.32	83.11

We examine the correctness and accuracy when the test utterances are shrunk with various scaling factors. The original training utterances are used for training our system. Similar to the case of stretching training utterances, shrinking utterances of FS3 to be 0.6, 0.7 and 0.8 times of the original, yields better recognition rates. Focusing on the most natural model LM2, the 0.7Test can achieve higher correctness and accuracy than the 0.8Test and 0.6Test. Shrinking the test utterance to be 0.7 times of the original duration can improve the recognition rate of 5.07% for correctness and 11.76% for accuracy, compared to the original test utterances.

5.5 Error Analysis

In the baseline experiment, as a consequence of setting the weight ratio between the acoustic model and the language model to be a low value, forcing the language model to be more important than the acoustic model, the insertion errors are explicitly reduced. At this point, the main errors in this task are substitution errors. Sets of alphabets that cause a lot of substitution errors are {'อ', 'ป', 'ม', 'พ', 'ช', 'จ'}. For instance, the alphabet 'อ' is rather substituted by 'ป', 'ม', 'พ', 'ช', or 'จ'. One potential reason is that these alphabets are pronounced with two syllables sharing the same vowel phone @@ in the first syllable and the phone *aa* in the second syllable.

Another substitution error is caused by the confusion of vowel pairs. Investigating recognition results of the baseline, we found out that the vowel alphabet 'อิ' (sound: */s-a1//r-a1//z-i1/*) is substituted by its long vowel pair 'อี' (sound: */s-a1//r-a1//z-ii0*) as well as the vowel alphabet 'อุ' (sound: */s-a1//r-a1//z-u1/*) is mostly substituted by 'อู' (sound: */s-a1//r-a1//z-uu0*). After compensating for the duration difference between training and test utterances by stretching training utterances to be 1.25Train, these substitution errors are dominantly reduced.

6 Conclusions

We presented a general framework for Thai speech recognition enhanced with spelling recognition. Four styles in spelling Thai words were introduced and discussed. Without a spelling corpus, the spelling recognizer was constructed using a normal continuous speech corpus. To achieve higher correctness and accuracy,we adjusted the ratio of importance between the acoustic model and the language model, making the language models more important than the acoustic models. To compensate for utterance speed among the training and test utterances, the training utterances were stretched and the experiments are performed on six subjects' test utterances. As a result, we gained correctness and accuracy. The experimental results for LM2 indicated a promising performance of 87.37% correctness and 87.18% recognition accuracy after this adjustment. To improve the recognition rate of the worst subject's test utterances, we experimented to find a good scaling factor of stretching the training utterances or shrinking the test utterances. As the result, the system achieved up to 12.60% accuracy improvement over the baseline. An analysis of recognition errors was also done. This work showed that applying a normal continuous speech corpus to train a spelling recognizer yield an acceptable

performance. Our further works are (1) to construct a system that can recognize several kinds of spelling methods, and (2) to explore a way to incorporate spelling recognition to the conventional speech recognition system.

Acknowledgements

The authors would like to thank National Electronics and Computer Technology Center (NECTEC) for allowing us to use the NECTEC-ATR Thai Speech Corpus. We would like to thank Prof. Cercone for many useful comments on earlier draft of this paper. This work has partly been supported by NECTEC under project number NT-B-22-I5-38-47-04.

References

1. San-Segundo, R., Macias-Guarasa, J., Ferreiros, J., Martin, P., Pardo, J.M.: Detection of Recognition Errors and Out of the Spelling Dictionary Names in a Spelled Name Recognizer for Spanish. Proceedings of EUROSPEECH 2001. (2001)
2. San-Segundo, R., Colas, J., Cordoba, R., Pardo, J.M.: Spanish Recognizer of Continuously Spelled Names Over the Telephone. Journal of Speech Communication, Vol. 38. (2002) 287-303
3. Rodrigues, F., Rodrigues, R., Martins, C.: An Isolated Letter Recognizer for Proper Name Identification Over the Telephone. Proceedings of 9th Portuguese Conference on Pattern Recognition. (1997)
4. Mitchell, C.D., Setlur, A.R.: Improved Spelling Recognition using a Tree-based Fast Lexical Match. Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing. 2(1999) 597-600
5. Bauer, J.G., Junkawitsch, J.: Accurate recognition of city names with spelling as a fall back strategy. Proceedings of EUROSPEECH 1999. (1999) 263-266
6. Pisarn, C., Theeramunkong, T.: Incorporating Tone Information to Improve Thai Continuous Speech Recognition. Proceedings of International Conference on Intelligent Technologies 2003. (2003)
7. Young, S., Evermann, G., Hain, T., Kershaw, D., Moore, G., Odell, J., Ollason, D., Povey, D., Valtchev, V., Woodland, P.: The HTK Book (for HTK Version 3.2.1). Cambridge University Engineering Department. (2002)
8. Pallone, G.: Time-stretching and pitch-shifting of audio signals: Application to cinema /video conversion. http://www.iua.upf.es/activitats/semirec/semi-pallone/index.htm
9. Verhelst, W., Roelands, M.: An overlap-add technique based on waveform similarity (wsola) for high quality time-scale modification of speech. Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, Vol 2. (1993) 554-557
10. Wikipedia: The free encyclopedia, Audio time stretching. http://www.ebroadcast.com.au/lookup/encyclopedia/au/Audio_time_stretching.html
11. Anastasakos, A., Schwartz, R., Shu, H.: Duration Modeling in Large Vocabulary Speech Recognition. Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing. (1995) 628-631
12. Thubthong, N., Kijsirikul, B.: Tone Recognition of Continuous Thai Speech under Tonal Assimilation and Declination Effects using Half-Tone Model. Journal of International of Uncertainty, Fuzziness and Knowledge-Based System 9(6). (2001) 815-825

13. Betz, M., Hild, H.: Language Models for a Spelled Letter Recognizer. Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing. (1995) 856-859
14. Jurafsky, D., Martin J.: Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition. Prentice Hall (2000)