IJREAT International Journal of Research in Engineering & Advanced Technology, Volume 1, Issue 4, Aug-Sept, 2013
ISSN: 2320 - 8791
www.ijreat.org

Comparison Between Native and Non-Native Speech Recognition for Telugu Language

K.Subramanyam
Assistant Professor, Department of IT
R.V.R & J.C College of Engineering, A.P, INDIA.

ABSTRACT

Communication using speech is inherently natural, with this ability of communication unconsciously acquired in a step-by-step manner throughout life. In order to explore the benefits of speech communication in devices, there have been many research works performed over the past several decades. As a result, automatic speech recognition (ASR) systems have been deployed in a range of applications, including automatic reservation systems, dictation systems, navigation systems, etc. Due to increasing globalization, the need for effective interlingual communication has also been growing. However, because of the fact that most people tend to speak foreign languages with variant or influent pronunciations, this has led to an increasing demand for the development of non native ASR systems. In other words, a conventional ASR system is optimized with native speech; however, non-native speech has different characteristics from native speech. That is, non-native speech tends to reflect the pronunciations or syntactic characteristics of the mother tongue of the non-native speakers, as well as the wide range of fluencies among non-native speakers. Therefore, the performance of an ASR system evaluated using non-native speech tends to severely degrade when compared to that of native speech due to the mismatch between the native training data and the nonnative test data. A simple way to improve the performance of an ASR system for non-native speech would be to train the ASR system using a non-native speech database; though in reality the number of non-native speech samples available for this task is not currently sufficient to train an ASR system. Thus, techniques for improving non-native ASR performance using only small amount of non-native speech are required. A new pronunciation modeling method is proposed as a means of improving the performance of non-native speech recognition. In addition, the performance of a non-native ASR system

adopting the proposed method is evaluated and compared to those employing conventional pronunciation model adaptation methods.

Parameters	Range
Speaking Mode	Isolated words to continuous speech
Speaking Style	Read speech to spontaneous speech
Enrollment	Speaker-dependent to Speaker-independent
Vocabulary	Small (< 20 words) to large (> 20,000 words)
Language Model	Finite-state to context-sensitive
Perplexity	Small (< 10) to large (> 100)
SNR	High (> 30 dB) to low (< 10 dB)
Transducer	Voice-cancelling microphone to telephone

FIG 1: Typical parameters used to characterize the capability of speech recognition systems

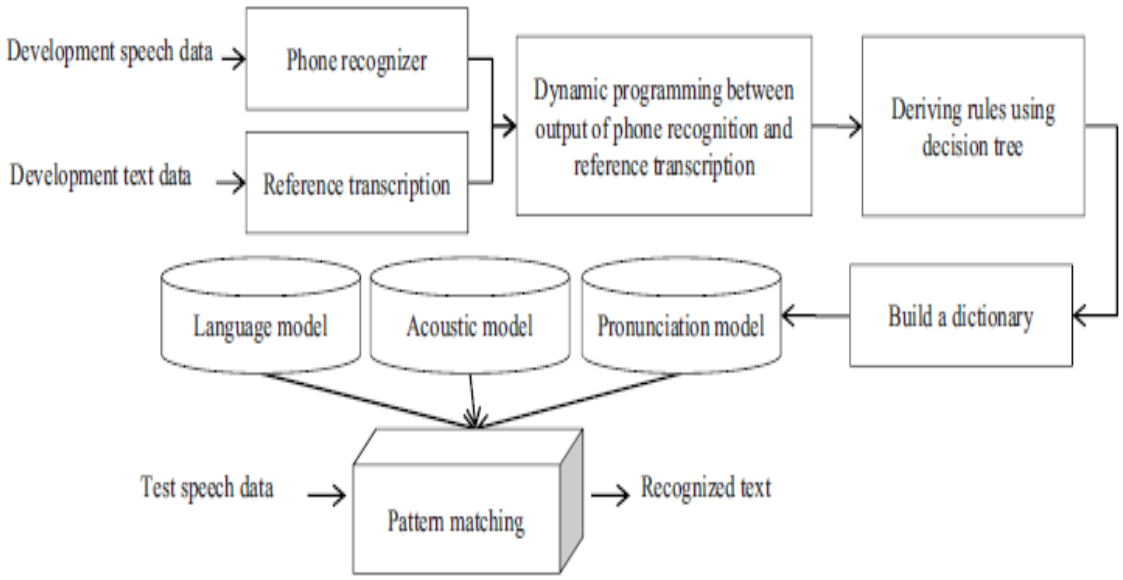

Fig. 2. Procedure of the proposed pronunciation variation modeling method based on an indirect data-driven approach applied to non-native ASR.

Overview of non-native speech recognition

Recently, speech recognition technology has become more familiar in our lives as numerous applications are increasingly adopting speech recognition systems. However, when these ASR systems are used by non-native speakers, the performance of the system can

rapidly degrade because of the mismatches between the native training data and the non-native test data.

Previously, several works have investigated the characteristics of non-native speech and the effect of non-native speech on ASR performance, some of which tried to explore the differences in characteristics between native and non-native speakers. For examples the duration and the first and second formant frequencies of English vowels spoken by Spanish speakers had different characteristics from those of native English speakers. Moreover, it was found that Spanish accented English was perceived better when the listeners were trained with this form of English. Similarly, it was noticed that the tongue location of the English vowels by nonnative speakers had different characteristics from that of native speakers In addition, unique consonants existed in some languages, such as four emphatic consonants of Arabic, and these unfamiliar Consonants were found to be hard to perceive by non-native speakers. It was then found that when non-native speakers pronounced words containing these unfamiliar consonants, degradation of ASR performance could occur.

Other researchers have attempted to compare the ASR performance of both native and nonnative speech and were shown that the word error rate (WER) of an English ASR system by German speakers was 49.3% whereas that of native English speakers was 16.2%. Moreover, in an ASR system trained by German speakers provided WERs of 18.5% and 34.0% when tested by native German speakers and English speakers, respectively. However, when the same ASR system was trained by English speakers but tested by German speakers, the WER increased from 35.0% to 65.6%. Based on these previous works, it is evident that adjusting for different pronunciation characteristics between native and non-native speakers is crucial for improving the ASR performance of non-native speech.

There have been three major approaches for handling non-native speech for ASR: acoustic modeling, language modeling, and pronunciation modeling approaches. First, acoustic modeling approaches find pronunciation differences and transform and/or adapt acoustic models to include the effects of non-native speech. Second, language modeling approaches deal with the grammatical effects or speaking style of non-native speech. Third, pronunciation modeling approaches derive pronunciation variant rules from non-native speech and apply the derived rules to pronunciation models for non-native speech.

Classification of techniques applied to non-native ASR.

1. *Non-native speech database design*

In order to develop a non-native ASR system and investigate the characteristics of nonnative speech, we first require non-native speech databases

2. *Acoustic modeling approach*

Acoustic modeling approaches are used to adjust acoustic models and thereby improve the recognition performance of non-native speech A simple way of adjusting acoustic models is to train them using a large amount of non-native speech. However, in practice it is rather difficult to collect a sufficient amount of non-native speech; therefore, acoustic models are usually adapted via a conventional acoustic model adaptation method, such as maximum likelihood linear regression (MLLR) and/or maximum a posteriori (MAP) methods. As an alternative, the acoustic models adjusted for non-native speech can also be obtained by interpolating the acoustic models for native speech and the acoustic models for the mother tongue. In other words, the acoustic models trained with two different languages are combined to obtain the acoustic models for non-native speech. However, the most popular way of obtaining the adjusted acoustic models is to apply an adaptation technique with only small amount of adaptation data for non-native speech.

3. *Language modeling approach*

Language modeling approaches deal with the grammatical effects or speaking styles of non-native speech, since non-native speakers tend to make a different sentence structure from native speakers. However, there are relatively few research works in this area, compared to either the acoustic modeling approaches or the pronunciation modeling approaches.

4. *Pronunciation modeling approach*

Pronunciation modeling approaches first derive pronunciation variants from non-native speakers and then apply them to the pronunciation models for non-native speech. Usually, the variant pronunciations for each word are added to the pronunciation models, which is similar to a multiple pronunciation dictionary approach. The pronunciation variants from non-native speakers can be derived by either knowledge-based or data-driven approaches. Note that knowledge based approaches are based on linguistics or phonetic knowledge whereas data-driven approaches automatically derive pronunciation variants from non-native speech data and can be further classified into either a direct method or an indirect method. If many

pronunciation variants are derived, the adapted pronunciation model becomes enlarged, resulting in performance degradation of the ASR system due to the fact that confusability in the pronunciation model is increased. Thus, several confusability reduction methods have also been proposed.

5. *Hybrid modeling approach*

Hybrid modeling approaches combine several modeling approaches, as described

Above, to further improve the performance of non-native ASR. In other words, acoustic or pronunciation modeling approaches can be combined in an MLLR and/or MAP adaptation. In particular, Bouselmi et al has proposed several combination schemes for pronunciation and MLLR/MAP acoustic model adaptations. On the other hand, pronunciation variant rules were decomposed into either pronunciation or acoustic variants. After that, pronunciation and acoustic model adaptations were applied to pronunciation and acoustic variants, respectively.

6. *Feature-domain approach*

The feature-domain approach applies a feature adaptation method to compensate for mismatches between training and test conditions; the acoustic models are trained using native speech, but are tested using non-native speech.

The approach which I worked is pronunciation modeling approach.

3 MODELING VARIATION: OVERVIEWOF APPROACHES

An important distinction that is often drawn in modeling pronunciation variation is that between within-word and cross-word variation [21]. The underlying phonetic mechanisms are different in the two and hence the need to address them separately. Approaches to handle cross-word variation have widely employed the use of multi-words [2-6], wherein frequent word clusters are concatenated as one lexical entry. This technique can account only a small portion of cross-word variation, like the variation between words that occur in very frequent sequences. Due to this limitation, other techniques involving rewrite rules based on word context etc have also been proposed like those described in [7-10].

Within word variation is the kind of variation that can be modeled at the level of the lexicon by adding pronunciation variants [11]. On similar lines, this thesis used modeling within-word variations. There are earlier approaches to this problem and differed within two broad phases:

1. Finding the information on variation of pronunciation

2. Integrating this information into ASR

RESULTS

The following are the Results:

Training	Testing	Accuracy	Error Rate
Native1	NonNative1	38.69%	74.78%
Native2	NonNative2	36.52%	63.4%
Native3	NonNative3	33.043%	70.0%
Sand	rav	46.087%	60.0%
Kon1	Kon2	77.391%	29.565%
Rav	Kon1	79.130%	28.696%

After addition to the Dictionary the results are as follows

Training	Testing	Accuracy	Error Rate
Native1	NonNative1	42.39%	72.78%
Native2	NonNative2	37.32%	61.4%
Native3	NonNative3	34.043%	68.83%
sand	rav	47.087%	59.0%
Kon1	Kon2	87.391%	17.565%
Rav	Kon1	92.130%	08.696%

CONCLUSION

The pronunciation modeling approach has given improved accuracy for non-native speech Telugu data.

FUTUREWORK

In this paper I used a limited database of 40 sentences each of 10 speakers. By increasing the database we can improve accuracy more upto 100%. We used static dictionary in this paper and we can extend it by implementing it using a dynamic dictionary. The accuracy can be achieved more by recording voices in noiseless environment.

REFERENCES

[1] Finke Michael and Waibel Alex, "Speaking mode dependent pronunciation modeling in large vocabulary conversational speech recognition", in Proc. Eurospeech, 1995.

[2] H. J. Nock and S. J. Young, "Detecting and improving poor pronunciations for multi words".

[3] M . Ravishankar and M. Eskenazi, "Automatic generation of context-dependent pronunciations", in Proc. Eurospeech '97, (Rhodes, Greece), pp. 2467–2470, 1997.

[4] T. Sloboda and A. Waibel, "Dictionary learning for spontaneous speech recognition," in Proc. ICSLP '96, (Philadelphia), pp. 2328–2331, 1996.

[5] N. Cremelie and J.-P. Martens, "In search of better pronunciation models for speech recognition", Speech Communication, vol. 29, no. 2-4, pp. 115–136, 1999.

[6] Adda-Decker M. and Lamel L., "Pronunciation variants across system configuration", Speech Communication, 1999.

[7] Eric John Fosler-Lussier , "Dynamic Pronunciation Models for Automatic Speech Recognition," Tech. Rep. TR-99-015, University of California, Berkeley, Berkeley, CA, 1999.