論文 / 著書情報
Article / Book Information

論題(和文)	
Title(English)	An Analysis of Accent Effects on Mandarin Large Vocabulary Continuous Speech Recognition
著者(和文)	楊 冬, 岩野 公司, 古井 貞熙
Authors(English)	Dong Yang, Koji Iwano, Sadaoki Furui
出典(和文)	日本音響学会2008年春季講演論文集, Vol. , No. 3-10-2, pp. 455-456
Citation(English)	, Vol. , No. 3-10-2, pp. 455-456
発行日 / Pub. date	2008, 3

An Analysis of Accent Effects on Mandarin Large Vocabulary Continuous Speech Recognition *

◎ Dong Yang, Koji Iwano, Sadaoki Furui (Tokyo Institute of Technology)

1 Introduction

Due to the vast region, huge population and historical reasons, there is a large variety of accented Mandarin Chinese spoken in China. Among various speaker variability sources that affect the performance of automatic speech recognition (ASR), accent is one of the most important factors [1]. Accent effects are very complex and it is unclear how accents affect ASR performance and how we can solve the problem, especially for situations in which accents are significantly affected by speakers' dialects.

In this paper, we analyze accent effects on speech recognition using two techniques. One approach lies in eigenvoice framework [2]; since speakers with different accents are acoustically distinct, there must be some eigenvoices which carry accent information and it is expected that we can differentiate speakers from various accents by projecting them using eigenvoices. Another approach is through error analysis in ASR of accented speech before and after adaptation; we measure the effects of accents on ASR errors, and apply the MLLR adaptation method to reduce the errors caused by various factors including accents. In other words, we analyze how much the MLLR method can reduce the mismatch between a general acoustic model and accented speech.

2 System Setup

Our training speech data is METI-Mandarin corpus collected in our lab, which contains 14 hours phonetically balanced speech data, being recorded from 40 standard Mandarin speakers (20 male and 20 female speakers). We use a toneless phoneme set [4] and train an acoustic model using typical HMM structure [3] with 25 dimension features, composed of MFCC features plus its derivatives and delta energy. A language model is trained from a part of the "Xinhua" news text in "Chinese Gigaword" corpus which contains about 289M words. Our evaluation speech data is collected by asking 14 speakers, several of whom are with accents, to read 50 sentences each. A 20,000 word dictionary is generated based on Mandarin CALLHOME lexicon released by LDC.

Table 1 Three types of confusions in Mandarin accented speech

confusion type	phones	examples
flat vs retroflex	(/s/, /ʂ/), (/ts^h/, /tʂ^h/), (/ts/,/tʂ/)	si, shi ci, chi zi, zhi
front-nasal vs back-nasal	(/n/,/ŋ/)	yin, ying
lateral vs nasal	(/n/./l/)	na. la

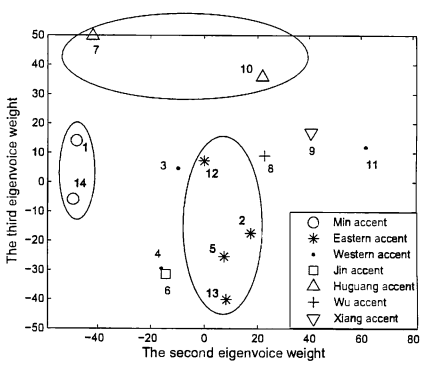

Fig. 1 Eigenvoice analysis results for the speakers

3 Eigenvoice analysis for accented speech

The eigenvoice analysis is widely used for speaker characterization. In the eigenvoice framework, a speaker's voice is represented by a supervector that is composed by concatenating mean vectors of all hidden Markov model Gaussian distributions. Principal component analysis is performed for supervectors by a group of speakers, and the resulting eigenvectors are called eigenvoices. Since every supervector can be represented as a linear combination of the eigenvoices, each supervector can be represented by a weighting vector. Our focus is to use eigenvoices to represent accent variations. We can easily imagine that male and female variation is indicated by the first eigenvoice vector. Therefore, we selected the second and the third eigenvoice vectors to plot the speech data. Figure 1 shows distribution of speakers in a plane consisting of the second and the third eigenvoice weights, where we can find that speakers with the same accent are closely located.

* 中国語大語彙連続音声認識における訛りの影響の分析,
楊 冬, 岩野公司, 古井貞熙 (東工大)

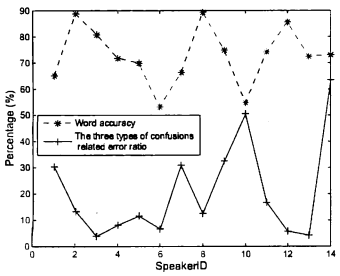

Fig. 2 Baseline word accuracy and ratio of the 3 accent-related errors for each speaker

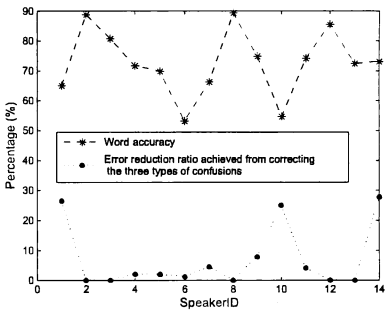

Fig. 3 Baseline word accuracy & error reduction ratio of that by correcting the three types of confusions to the total error reduction

4 Recognition error analysis of accented speech

We focus on the three types of mispronunciations which are the major problems in Mandarin accents, as shown in Table 1.

Under the condition stated in Section 2, which we refer to as baseline, the average word accuracy is 73.1% and the result for each speaker is shown in Figure 2. For each speaker, 30 sentences are evaluated and the remaining 20 sentences are kept for supervised adaptation. Since we do not have a large amount of accented speech data, accent adaptation cannot be performed directly. Therefore, we perform MLLR speaker adaptation and analyze the accent effect implicitly solved by the adaptation process.

The ratio of the recognition errors caused by the three kinds of the accent effects to the total errors is also shown in Figure 2 for each speaker. In the figure, there exists very clear negative correlation between the word accuracy and the accent-related error ratio. This means that accent is a dominant problem in Mandarin Chinese ASR. Speakers 1, 10 and 14, having relatively low word accuracies, have significantly stronger accents than other speakers; in other words, their high recognition error rate is due to the errors caused by the three types of confusions. Speaker 6, also having very low word accuracy, is an exception, since this speaker has relatively few accent related errors. The reason is that, although speaking good Mandarin, this speaker has very different voice from ordinary speakers.

The error reduction by the MLLR adaptation is evaluated from the viewpoints of the accent effects. The ratio of error reduction achieved by correcting the three confusions-related errors observed in the baseline experiment to the total error reduction for each speaker is shown in Figure 3. Except for Speaker 6, it can be observed that there exists negative-correlation between the baseline word accuracy and the accent related error reduction. These results show that the MLLR adaptation is effective in solving the accent mismatch problems.

5 Conclusion

Our analysis and experimental results show that accents greatly affect Mandarin speech recognition, and the MLLR adaptation technique is effective in solving the accent mismatch problems. Further studies will be conducted by using a larger accented speech corpus.

6 Acknowledgement

This work was supported in part by the 21st century COE program "Framework for Systematization and Application of Large-scale Knowledge Resources". The speech corpus used for training the acoustic model was funded by the METI Project "Development of Fundamental Speech Recognition Technology".

References

[1] C. Huang et al.,"Accent issues in large vocabulary continuous speech recognition," International Journal of Speech Technology, 7(2-3), 141-153, 2004.

[2] R. Kuhn et al., "Rapid speaker adaptation in eigenvoice space," IEEE Trans on Speech and Audio Processing, 8(6), 695-707, 2000.

[3] S. Young et al., "Large vocabulary continuous speech recognition," IEEE Signal Processing Magazine, 13(5),45-57, 1996

[4] http://en.wikipedia.org/wiki/pinyin