A Learner Adaptive Chinese Pronunciation Education System for Japanese[*]

◎Minyi Ma, Keikichi Hirose, Nobuaki Minematsu (University of Tokyo)

1 Introduction

With the rapid recent emergence of China as a major player in the global economy, there is an increased urgency to find ways to help foreign speakers to learn Chinese. Our research objective is to develop a CALL system to help Japanese people to improve their Chinese pronunciation.

In this paper, we will first present the reason why we developed a learner level adaptive CALL system, then will introduce main modules of our system, and at last, will tell our system evaluation result and our future work.

2 Adaptive CALL System

2.1 Learner Level Adaptive

We collected Mandarin speech data from Japanese speakers and analyzed features of their utterances Most findings are presented in our previous ASJ paper[1]. To conclude in a word, problems of different level—beginner or advanced learner, are different. Beginners are usually poor at phonemes and tones, which make their utterances hard to be understood; while, advanced learners may manage to pronounce phonemes well, but when they speak sentences, there are problems in tones and durations that make their utterances sounds unnatural. Thus, we proposed our learner level adaptive CALL system to provide different pronunciation practice emphases to satisfy different levels of learners.

Our system's form structure is shown in Figure 1. There are three main modules in our system: Learner Level Judgment Module, Word-level Practice Module and Sentence-level Practice Module. Judgment module is to judge learner's pronunciation level and recommend leaner to enter proper practice course. Word-level Practice Module puts emphases on phoneme training and word-level tone training. Its aim is to help beginner level learner to pronounce clearly and correctly in order to make their pronunciation

easy to be understood. Sentence-level Practice Module puts emphases on tone training and duration training on sentences level. Its aim is to help advanced learner to pronounce naturally and thus gradually reach a native level.
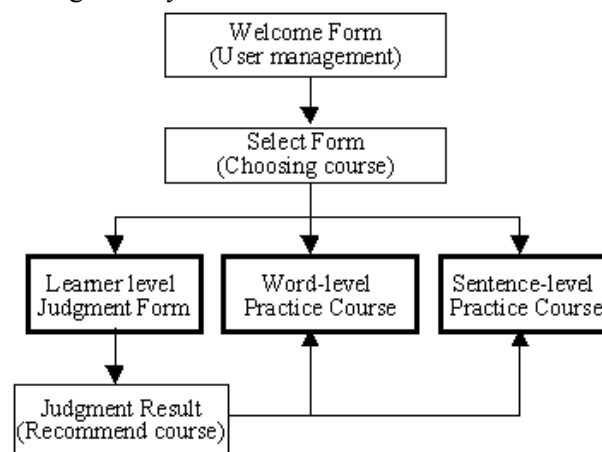

Fig. 1 System Structure

2.2 Learner Level Judgment Module

Learner level judgment is one of our core modules, and it represents the main innovation of our system. In this module, our system will ask leaner to record 10 short sentences and through analyzing these recording, the system will give a judgment report of learner's pronunciation.

The 10 short sentences to be recorded in this module are carefully designed that, a) they include all Chinese 36 vowels and 21 consonants; b) they include most of tone combinations; c) they are all short sentences with no punctuation mark (uttered without pause); d) they are all easy to be understood.

In the judgment report, we are not only intended to provide a result of level judgment like "beginner" or "advanced learner", but also to tell the learner what are the problems of his/her pronunciation and how to overcome his/her weak points. Therefore, the system will provide the following aspects of information in the report:

a. Display HMM recognition results
b. Give evaluation score
c. Recommend practice course

---

d. Tell hints for practice

We tried to design the pronunciation evaluation algorithm to reach a judgment result as closer as possible to human scoring. Finally, we decided to calculate following indices to get machine score:

1. Strictness. According to the recognition result of HMM, we calculate the percentage of wrong phonemes in order to estimate how necessary should the learner have phoneme practice.

2. Segment Duration Scores. Insertions, deletions or substitutions of phonemes will result in duration differences, so we use the simplest approach—a measure of Rate of Speech (ROS) to compute the average number of phonemes per unit of time. The ROS shows learner's performance of phoneme durations and indicates how necessary should the learner have duration practice.

3. Likelihood-based GOP. For each sentence, we first extract MFCC features, then use continuous speech recognizer to finish forced alignment and unconstrained phoneme loop, in order to calculate GOP (Goodness of Pronunciation). The GOP score analyzes a whole utterance and returns scores for each phoneme in this utterance. By setting a threshold (used like a filter), it becomes possible to generate a seriously mispronounced phoneme list of learner. And then by calling a guidance dictionary, the system can provide practice hints to let learner focus on practices of more noteworthy phonemes.

2.3 Word-level Practice Course Module

In this module, first there are initial hints to remind learner for every word, such as suggesting which phoneme to take care or how to pronounce some particular tone correctly. Then the learner can record his pronunciation and listen to teacher's reference utterance. After learner records his pronunciation, system will detect phoneme errors and tone errors. According to detected errors, the text of hint will be changed. For example, if user mispronounced 'sh' as 's', then text of how to pronounce raised tongue phoneme will be shown. Also, system will modify user's utterance and generate corrective audio feedback in learners' own voice using speech modification to help them understand the

problems of their pronunciations[2].

2.4 Sentence-level Practice Course Module

In this module, the generally practice process is similar to word-level practice module. But since the practice emphases are duration and sentence-level tone, the initial hints would be how to segment sentence into small words correctly, and tone changing information. System will also show teacher's model utterances and generate corrective audio feedback.

3 System Evaluation

We evaluated our system on following aspects:

a) Comparing to human result, the system judgment of phoneme correction is about 90%.

b) According to a subjective listening test, the audio feedback generated by system shows good improvement. Our system is providing useful feedback information to help learner to practice.

c) All testers felt the hints and judgment offered system are proper and helpful. According to evaluation score marked by system, all testers are making progress.

4 Future Work

Up to now, only simply evaluation is finished. After we manage to make our system more stable, we are going to contact class of learners to try our system to get more tryout reports.

Difficulty of pronunciation is different depending on phonemes. So when calculating GOP, we are planning to set a different threshold for each phoneme. This may help enhance evaluation precision.

In the judgment module, we did not consider tone evaluation at the moment. How to combine tone recognition is also one of our future works.

Reference

[1] Minyi Ma, Hiroko Hirano, Keikichi Hirose, Nobuaki Minematsu, "Proposal of Adaptive CALL system for Japanese Learners of Mandarin", Proc. Autumn Meeting of ASJ, 1-Q-6, pp.333-334 (2008-9)

[2] Frederic Gendrin, Keikichi Hirose and Nobuaki Minematsu, "Corrective feedback for accent pattern CALL systems using speech modification", IEICE Technical Report, SP2002-161, pp.1-6 (2003-1).