Liu, C.-C. et al. (Eds.) (2014). Proceedings of the 22nd International Conference on Computers in Education. Japan: Asia-Pacific Society for Computers in Education

Developing and Evaluating a Test Generation Module to Support Personalized Phoneme-based Training

Yaheng ZOU [a*], Harumi KASHIWAGI[b], Yi SUN[c], Kazuhiro OHTSUKI[a] & Min KANG[a]
[a]*Graduate School of Intercultural Studies, Kobe University, Japan*
[b]*School of Languages and Communication, Kobe University, Japan*
Graduate School of Information Technology, Kobe Institute of Computing, Japan
*zouqqing@gmail.com

Abstract: In this paper, we describe a test generation module to support personalized phoneme-based training in listening learning of English vocabulary. On the basis of the personalized phoneme error patterns detected by an error detection module, the test generation module would generate personalized listening tests. An evaluation experiment was conducted for evaluating the module.

Keywords: CALL System, Data Mining, Adaptive Learning, English Vocabulary Learning, Phoneme Perception

1. Introduction

In the past decade, adaptive learning has become considerable and there is a trend to develop the applications to generate adaptable materials. The trend also influences the field of computer assisted language learning (CALL). Item response theory was used to evaluate individual English vocabulary ability and a personalized mobile English vocabulary learning system has been developed (Chen and Chung, 2008). As enhancing communication skill becomes more and more important in foreign language education, CALL systems have been widely used in Japan as well. By CALL, students' learning log-data can be easily saved in database and the saved log-data gives us the possibility to automatically create and offer contents to students to guide their own learning processes or improve their weak points (Ueno, 2007). However, it has not been enough explored how to build personalized systems for listening learning.

On the other hand, it has been indicated that word recognition is the core process in listening of English (Rost, 2002). In some cognitive models of speech perception such as the cohort model, phoneme perception is considered as one of the key factors. In classroom education, it is well-known that Japanese learners have difficulties in distinguishing some special phonemes such as /r/ and /l/ (Horibe and Furuhashi, 1974) so that the phoneme-based training has been indicated to be an important learning strategy in vocabulary learning. Therefore, we consider that detecting individual weak-points at the phoneme level and conducting personalized phoneme-based training may help the individual enhance the listening comprehension ability in vocabulary learning.

In this paper, we describe a test generation module to support personalized phoneme-based training in English vocabulary learning. For a student, the log-data in word dictation tests are analyzed and the student's error patterns at the phoneme level are detected by the use of the error detection module developed in our previous work (Zou, Kashiwagi, Ohtsuki and Kang, 2013). On the basis of the error patterns, the test module will generate a test with alternative response questions for the student who is requested to discriminate minimal pairs of English words. In addition, an evaluation experiment is conducted in order to evaluate the module.

2. Vocabulary Test System and the Error Detection Module

We built a vocabulary test system in AMP (Apache, MySQL, and PHP) environment implemented the error detection module and the test module. The error detection module was targeted at analyzing the log-data in word dictation tests in that students are requested to input the spellings of English words as English answers and Japanese translations corresponding to the words.

In the error detection module, we have focused phoneme error detection on the detection of misperceptions, e.g. "slips of the ear" (Rost, 2002). We manually checked the wrong spellings and the corresponding right spellings in the past log-data and noted that most of the wrong answers could be categorized into mistakes of misperception. Most of the wrong answers were given as different English words in that only one phoneme or two phonemes are different from the right English words. We have used Japanese translations to examine whether a wrong spelling of a word represents the misunderstanding of the word or not. We suggested that wrong spellings mostly are in agreement with misunderstanding of words because there were only few answers in that the spellings of words were wrong but their Japanese translations were right. Certainly, the suggestion depends on the log-data collected.

According to the implemented algorithm, the error detection module would first decompose a right answer and a corresponding wrong answer into phonemes by the use of the CMU Pronouncing Dictionary (CMU pronouncing dictionary, n.d.), and then extract the phoneme error pair. Finally, the personalized phoneme error patterns would be detected by the use of the Apriori-like method in sequence data mining and be described as /Right-Phoneme/ -/Wrong-Phoneme/.

3. Test Generation Module

On the basis of the detection result by the error detection module, the test generation module would generate a test with alternative response questions in order to help an individual student or group distinguish phonemes that he/she or they mistook. We consider that the more a pair occurred in the results, the more it turns difficult for a student to distinguish the pair. In this work, the top error pair that is most frequently occurred in error patterns is selected to be the training target.

In a test generated by the module, each question consists of two words that are minimal pair and correspond to a phoneme error pair respectively. The words are randomly picked out from the minimal pair database. As an example, for the phoneme pair /r/ → /l/, a minimal pair would be the words "rock" and "lock". One sound file corresponding to each minimal pair would be generated by the use of AT&T Natural Voice Engine.

4. Results in the Evaluation Experiment and Future Work

In order to evaluate the module, we conducted an evaluation experiment. We attempted to find out if the module can generate appropriate tests for training for individual students and if the training is effective to help students enhance their listening comprehension ability in vocabulary learning.

We designed a pretest-training-posttest experiment. In each pretest, a student needs to answer 20 word dictation questions. Once he/she submitted the answers, then a web page with his/her phoneme error patterns and 10 alternative response questions will be shown to him/her for training. After the training, the posttest with 15 word dictation questions will be presented. Each test for training and corresponding posttest are automatically generated by the test generation module.

In the experiment, we had two third-year students and two fourth-year students take four pretest-training-posttests numbered Test I, II, III and IV in one month period. We randomly chose the subjects from the students who have completed all necessary English courses provided by the university. The words both in the pretests and the posttests of Test I, II and III were randomly selected from the JACET 2000 list and the words in Test IV were those selected from the JACET 4000 list (JACET Vocabulary, n.d.). Students were previously requested to declare what phoneme pairs are difficult to distinguish and answer a paper-based test that measures vocabulary levels in reading. The result of the paper-based test shows that the subjects can understand almost all of words in the JACET 4000 list. This was aimed at decreasing spelling mistakes in the answers of the subjects. In the previous investigation, Student A and Student B declared that the /r/-/l/ pair is the most difficult pair to distinguish while Student C and Student D declared that the /v/-/b/ pair is the most difficult one.

Table 1 shows the auto-generated tests for Student B by the module. It is obvious that all tests for training are appropriately generated by the module although there are some duplicate minimal pairs.

Table 1. Auto-generated Tests For Phoneme-based Training

No.	Error Pairs	Words in Questions
I	/v/-/b/	vow-bow, rebel-revel, dub-dove, vein-bane, bale-veil, bane-vain, vent-bent, bale-veil, vat-bat, beer-veer
II	/r/-/l/	bright-blight, loot-root, glamour-grammar, rite-light, arrive- alive, flee-free, lock-rock, rocker-locker, ray-lay, rob-lob
III	/i/-/i:/	sip-seep, skim-scheme, kill-keel, bid-bead, bin-bean, skim- scheme, slip-sleep, ill-eel, piss-peace, slick-sleek
IV	/i/-/i:/	din-dean, piss-piece, sin-scene, shin-sheen, din-dean, piss- piece, sip-seep, each-itch, fist-feast, hip-heap

We examined the phoneme error pairs in the pretests of the subjects. It is noted that Student A had the same problem in all of the tests that she possibly cannot distinguish /r/ and /l/ very well as she declared. On the other hand, the weak points of the other subjects seem varying with the pretests as shown in Table 1. After the experiment, we manually performed the error detection module to deal with three kinds of accumulative pretest data that consists of the data in Test I and Test II, the data in Test I, II and III, and the data in Test I, II, III and IV, respectively. We noticed that the accumulative pretest data would provide a more stable result about the error pairs of all subjects. For Student A, the most difficult phoneme pair to distinguish is still the /r/-/l/ pair. For Student B, Student C and Student D, the /v/-/b/ pair become the most difficult one.

For investigating the effect of the tests for training, we compared the correct answer rates of the words in the pretests and the corresponding posttests by categorizing the words to the corresponding error pairs. The result by Fisher's exact test explains that there is a significant improvement in the posttests for Student A. However, there is no significant improvement observed for Student B, Student C and Student D. This may stem from whether the module detected the real weak points of the subjects or not. The consistency of the error pairs in the pretest data and the accumulative pretest data assume that the /r/-/l/ pair seems to be the real weak point of Student A. Conversely, for the other subjects, the module possibly has not appropriately estimated the real weak points.

Therefore, we assume that the auto-generated tests for training may help students enhance the perception of phonemes if the training targets are appropriately determined. For the question about how to make the auto-generated tests for training more effective, further investigation is needed.

Acknowledgements

This work was partially supported by JSPS KAKENHI Grant Number 23501108.

References

Chen, C. M. & Chung, C. J. (2008). Personalized mobile English vocabulary learning system based on item response theory and learning memory cycle. *Computers & Education 51*, 624–645

CMU Pronouncing Dictionary (n.d.). Retrieved May 23, 2014, from http://www.speech.cs.cmu.edu/cgi-bin/cmudict

Horibe, N. & Furuhashi, S. (1974). Hierarchy of Aural Perception Difficulties at Several Levels of English Teaching. *JACET Bulletin, 5*, 87-106

JACET Vocabulary (n.d.). Retrieved May 23, 2014, from http://j-varg.sakura.ne.jp/publications/index.html

Rost, M. (2002). Teaching and Researching Listening. *Pearson Education*, Harlow

Ueno, M. (2007). Data mining in e-Learning (in Japanese). *Japan Journal of Educational Technology*, *31*(3), 271-283

Zou, Y. H., Kashiwagi, H., Ohtsuki, K. & Kang, M. (2013). Developing a Learning Support System to Detect Phonemic Errors in Data of Listening Learning Log. *Proceedings of the International Conference on e-Commerce, e-Administration, e-Society, e-Education, and e-Technology*. 1145-1154