

An EFL CALL Teaching Aid for Self-directed Learning Using the AmiVoice® Speech Recognition System: Automatic Scoring of the Fluency of Learners' Utterances

Yasushige ISHIKAWA
Department of British and American Studies
Kyoto University of Foreign Studies
Kyoto, 615-8558, Japan

and

Mutsumi KONDO
Graduate School of Foreign Language Education and Research
Kansai University
Suita, 564-8680, Japan

and

Craig SMITH
Department of British and American Studies
Kyoto University of Foreign Studies
Kyoto, 615-8558, Japan

and

Katsuzi NAGAI
Advanced Media, Inc.
6F Sunshinecity Bunkakaikan, 3-1-4 Higashi-Ikebukuro, Toshima-ku
Tokyo, 170-8630, Japan

ABSTRACT

At Kyoto Junior College of Foreign Languages on the premises of Kyoto University of Foreign Studies, Hospitality English, a new first-year English as a Foreign Language course was introduced in April 2008. Hospitality English, a one semester, one-and-half-hour per week course was designed to develop specific English spoken language skills to help students gain employment in the hotel industry in Japan. The context of all the language skill lessons was situational; and an English as a Foreign Language (EFL) CALL teaching aid for self-directed learning using the AmiVoice® speech recognition system which included video scenes of hotel employee-guest interactions in a Kyoto hotel enriched the situational language learning context. This course was linked to a hotel internship program which students in the course

were expected to apply for as part of their English and Service Industry Studies. This paper reports on the development and refinement of the EFL CALL teaching aid and on the automatic scoring of the fluency of learners' utterances.

Keywords: EFL, Automatic scoring, Shadowing, Self-directed study with CALL, Learner autonomy, Spoken language communication, Hospitality English

1. INTRODUCTION

The White Paper on Tourism published by the Japanese Ministry of Land, Infrastructure, Transport and Tourism claims that increasing the number of tourists who visit Japan from other countries is a key to improving the Japanese economy. The government aims to reach a target of 10,000,000 visitors per year

by 2010. The number of tourists from other countries was over 8,300,000 in 2007 and that number shows signs of increasing [1]. In response to this, a number of universities and junior colleges in Japan have set up faculties and departments that aim to produce work-ready graduates for the tourism sector of the service industry. Among Japanese cities, Kyoto, in which Kyoto Junior College of Foreign Languages is located, is one of the most popular tourist destinations for domestic and foreign tourists. Over 9,300,000 visitors spend one day or longer in Kyoto every year. Thus, it is essential for hotel staff to be able to communicate efficiently in English with tourists from a wide variety of countries and also to have the ability to provide services appropriately to foreign hotel guests in English, that is to say, 'hospitality' English language skills are required.

Kyoto Junior College of Foreign Languages initiated Hospitality English, a new first-year English course in April 2008. Hospitality English, a one semester, one-and-a-half-hour per week course aims to develop students' English spoken language skills in order to help students successfully apply for and complete a competitive hotel internship program at hotels in Kyoto. Participation in the internship program is very helpful if students choose to apply for career positions in the hotel industry after graduation.

The Hospitality English course design included three learning task components: 1. Students engaged in self-directed outside-of-class learning with the use of an English as a Foreign Language (EFL) CALL teaching aid for with the AmiVoice® speech recognition system which included video scenes of key communication events between hotel employees and hotel guests; and learning tasks to support grammar, vocabulary, and pronunciation learning. 2. Conventional classroom EFL learning tasks were used by teachers to teach key features of language, to allow students to practice, and to give students instructive feedback. 3. Simulations of hotel communication situations were conducted in the classroom during visits by native speakers of English. The simulations were intended to bridge the CALL and the conventional classroom learning tasks with actual work situations.

2. LEARNER AUTONOMY WITH THE CALL TEACHING AID

Holec [2] describes learner autonomy: 'when the learner is willing and capable of taking charge of his/her own learning, the learner should be capable of determining the objectives; defining the contents and

the progressions; selecting methods and techniques to be used; monitoring the procedure of acquisition..., evaluating what has been acquired.' Furthermore, Little [3] argues that 'autonomy is a capacity—for detachment, critical reflection, decision-making, and independent action. It presupposes, but also entails that the learner will develop a particular kind of psychological relation to the process and content of his learning.' In other words, autonomy is a taking control of one's own learning so that learners may reflect on their learning processes, evaluate themselves on the progress of their own learning and then, adjust their learning strategies accordingly. The term 'autonomy' has come to be used to describe: 1. self-directed learning situations, 2. a set of autonomous skills which can be applied in self-directed learning, 3. an inborn capacity for individual learning which may be suppressed by institutional education, 4. the exercise of learners' responsibility for their own learning, and 5. the right of learners to determine the direction of their own learning [4].

In the Hospitality English course described in this paper it was anticipated that the integration of the CALL teaching aid with conventional EFL teaching methods would create opportunities for self-directed learning which, in turn, would contribute to the development of autonomous learning skills by allowing students to make decisions about their own learning. Self-directed learning could be expected to enhance students' spontaneous motivation [5], [6] and in addition, it was hoped that the support of the classroom instruction related to the effective use of the CALL aid would help sustain student motivation.

In order to foster students' autonomy with the CALL teaching aid, the following seven steps were implemented in each lesson. The seven steps, as shown in Fig. 1, are: (1) Preparation for Class Meeting 1, (2) Review of the contents of Class Meeting 1, (3) Preparation for Class Meeting 2, (4) Review of the contents of Class Meeting 2, (5) Conducting an achievement test and preparation for simulations of work experiences in Class Meeting 3, (6) Students' self-assessments of their own weak points and the setting of new goals through reflection on the results of the achievement test and the work simulations and (7) the self-directed learning experiences for achieving the new goals.

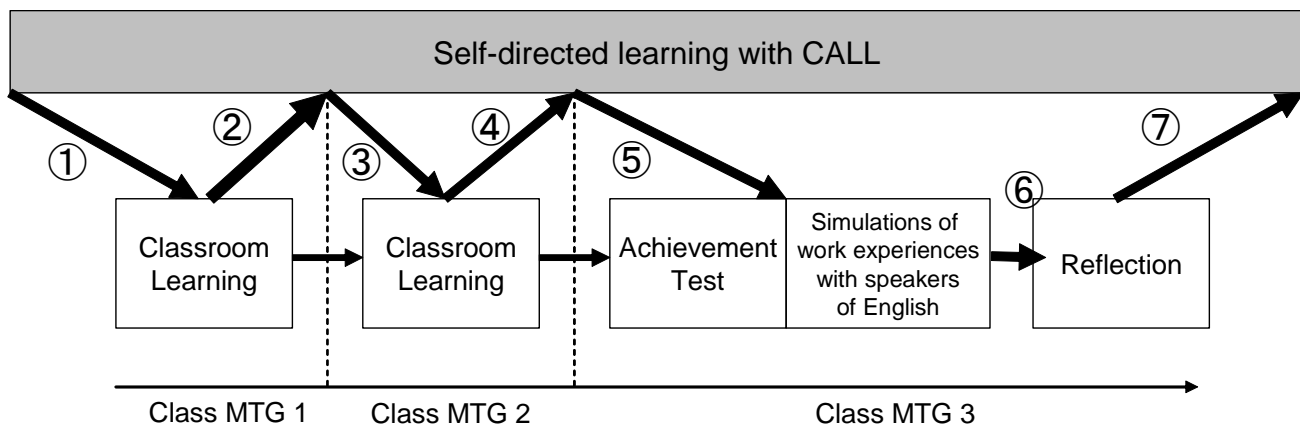


Figure 1: 7 steps to foster autonomy

3. DEVELOPMENT OF THE CALL TEACHING AID

The assumption of the CALL design was that it would be useful for students to repeatedly listen to useful expressions and to say these expressions aloud at the same time. The positive effects of students' listening training would be transferred to their pronunciation ability, and in turn, students' improved pronunciation skills would be transferred to their listening training. Moreover, it was assumed that listening training would be more effective when phrases rather than single words were the learning targets. This is because the phonology and meaning can be more easily inferred when phrases rather than single words are used [7].

The CALL teaching aid was designed to demonstrate an overall model for the students, to give instructions in the use of the CALL teaching aid, and to provide learning tasks to help students develop accurate pronunciation, learn low frequency vocabulary and phrases useful in learning to be a good communicator in attending to customers in the travel and tourism business after graduation. The CALL aid which we developed in 2008 included six skits that were dialogues between a hotel guest from another country and a hotel clerk at various sites in a hotel. There was one review of the six skits. Each skit had eight components and the eight components provided students with opportunities to use the expressions which students had learned through self-directed learning of the skits. Moreover, the aid had a voice guidance facility for students to virtually speak to the aid during their self-directed learning periods. The components in the CALL teaching aid were: vocabulary learning preparation, true/false

comprehension questions, shadowing of the text, fill-in-the-blank spelling questions, listening to a model role play, scripted role playing, vocabulary building, and unscripted virtual role playing.

At the end of the fall semester in the 2008 academic year, a post-course evaluation was conducted for the seventeen students of Kyoto Junior College of Foreign Languages who participated in our study: eight students took Hospitality English 1, a spring-semester course and nine students took Hospitality English 2. The students' evaluations of the three learning components: self-directed study with CALL, classroom learning, and simulations of work experiences, revealed the following information:

- 1) Students reported that they had not been satisfied with CALL and they had not worked with enthusiasm.
- 2) Students reported that they had worked on the classroom activities with enthusiasm.
- 3) Students believed that the simulations of work experiences with the native speakers of English had been useful to them and had helped improve their language skills.

Through a correlation analysis between the evaluations of the three learning components, the following observations were drawn:

- 4) In both Hospitality English 1 and Hospitality English 2, the rate of the correlation between the classroom activities and simulations of work experiences was high. In other words, the participants felt that the classroom activities were useful preparation for successful participation in the simulations of work experiences.
- 5) In Hospitality English 1, the rate of the

correlation between the classroom activities and the CALL was low [8].

The above results indicate that CALL language teaching aid that we developed in 2008 needs to be improved in order to achieve constructive relationships with the skills that are developed in the other components of the course. Therefore, we have refined the CALL teaching aid.

4. REFINEMENT OF THE CALL TEACHING AID

We refined the 2008 version of the CALL teaching aid by adding two newly-developed components, “Before You Start” and multiple-choice comprehension questions, and by deleting the true/false questions. We also created ten more skits that were dialogues between a hotel guest from another country and a hotel clerk at various sites in a hotel. Moreover, we developed an automatic scoring system to assess the fluency of students’ utterances in the shadowing task and in the role playing task. In the following sections, there is an explanation of the nine components in the refined CALL teaching aid: “Before You Start,” comprehension questions, vocabulary learning preparation, shadowing exercises, fill-in-the-blank questions, listening to a role play, role playing, vocabulary building and virtual role playing.

“Before You Start”

This activity is intended to check how much students know about hospitality English language skills. Students input their own responses to a guest’s questions and check what a professional receptionist in the video clip says in order to recognize differences between their assumptions and professional hospitality English standards. Fig. 2 shows the screenshot of “Before You Start.”



Figure 2: Screenshot of “Before You Start”

Comprehension questions

This language activity is intended to check whether students understand the content of each skit. After watching a video clip of each skit, students are required to answer five questions regarding the contents of the skit. Students are requested to listen to the questions without reading the text. However, students who have low listening ability may read the question by pushing the button “Hints” as Fig. 3 shows. After answering all the five questions, they check their answers with a screen display which indicates failure or success with the figures × and ○.



Figure 3: Screenshot of comprehension questions

Vocabulary learning preparation

Students learn useful vocabulary in each skit. After listening to model pronunciation examples of vocabulary, they say the words aloud. An English speech recognition system developed by Advanced Media, Inc. features a continuous speech recognition system which can evaluate Japanese learners’ English pronunciation. With an acoustic database of Japanese learners’ English pronunciation, a learner’s pronunciation problems can be immediately detected and appropriate feedback on pronunciation errors and an estimate of intelligibility can be provided as Fig. 4 shows.



Figure 4: Screenshot of vocabulary learning preparation

Fill-in-the-blank questions

Students are required to spell out the vocabulary that they had learned in the vocabulary learning preparation task to check whether they have memorized the spelling of the words. Students fill in blanks typing out the last sound that they hear and by checking pop up windows attached to the underlined and italicized parts in the script, students also learn useful expressions which are related to the underlined and italicized parts as shown in Fig. 5.



Figure 5: Screenshot of a pop up window in fill-in-the-blank questions

Listening to a role play

Students are required to say aloud the model phrases as they watch the video clip of each text and to pronounce the phrases correctly. The English speech recognition system assesses students' utterances and a learner's pronunciation problems can be immediately detected by an acoustic database of Japanese learners' English pronunciation. The pronunciation errors in the students' utterances are counted and at the same time the words which were recognized by the speech recognition system are shown in green on a screen display as indicated in Fig. 6.



Figure 6: Screenshot of words recognized by the speech recognition system

Role playing

Students are required to speak all the parts of the hotel clerk. Students are also required to follow the tempo of the utterances of the hotel clerk. The English speech recognition system assesses students' utterances.

Vocabulary building

Students learn vocabulary which will appear in the next language activity, virtual role playing. The English speech recognition system assesses students' utterances.

Virtual role playing

Students speak with a virtual dialogue partner. The English speech recognition system assesses students' utterances. If students' pronunciation is not satisfactory, the response system remains silent.

5. SHADOWING

Considerable attention has been directed towards the practice of shadowing, students' repeated reading aloud of a text as they listen to a narration, in the field of EFL education in Japan. Japanese students tend to use a Japanized English pronunciation. This is, from a cognitive linguistic point of view, because English words have been stored in the long-term memory of the mental lexicon as words which have been transferred into Japanized English. Shadowing does not give our brains enough time to search for vocabulary information in the mental lexicon and requires us to pronounce English words as heard. This leads to the imitation of the sounds heard and to improvements in pronunciation. Shadowing also leads to improvement in listening skills. Listening is composed of perception and apprehension, both of which consume cognitive resources. Shadowing allows more cognitive resources to be provided to apprehension than to perception by automatizing perception through repetitive listening to English sounds and by the storing of English sounds in the mental lexicon. As a result, the practice of shadowing can be expected to improve listening skills [9].

Definition of fluency

The term fluency has been used for teachers to describe non-native performance as well as native performance. Richards & Schmidt [10] define the term fluency as the following:

- a. the ability to produce written and/or spoken language with ease
- b. the ability to speak with a good but not

In second and foreign language teaching, fluency describes a level of proficiency in communication, which includes:

- necessarily perfect command of intonation, vocabulary, and grammar
- c. the ability to communicate ideas effectively
- d. the ability to produce continuous speech without causing comprehension difficulties or a breakdown of communication.

It is sometimes contrasted with accuracy, which refers to the ability to produce grammatically correct sentences but may not include the ability to speak or write fluently.

Fillmore [11] identified four different abilities of fluency and the first among the four abilities is “the ability to talk at length with few pauses, the ability to fill time with talk.” Brumfit [12] defined fluency “the maximally effective operation of the language system so far acquired by the student” According to Nation [13], these definitions suggests that fluency can be related to “the speed and flow of language production.” In order to define these temporal aspects of fluency, Lennon [14] claims that producing “speech at the tempo of native speakers” is equal to speaking English fluently. In our study, the term fluency was defined as the ability to speak native-like English by speaking English at the tempo of native speakers.

The prosody hypothesis

Prosody is the set of the phonological characteristics of the intonational sounds of English. Crystal [15] defines “prosody” as “a term to refer collectively to variations in pitch, loudness, tempo and rhythm.” In our study, we generated the following hypothesis:

Students would be able to acquire prosody, in other words, to improve the pitch, loudness and rhythm of their own utterances by following the tempo of model utterances as closely as possible.

Two perspectives for evaluating learners’ fluency

In the shadowing component, we built into the learning tasks an automatic scoring of learners’ utterances utilizing the AmiVoice[®] speech recognition system. Automatic scoring of learners’ utterances was accomplished by: 1. comparing a model utterance with a learner’s utterance based on time-constrained bottom-up clustering [16] and, 2. detecting learners’ pronunciation errors through segmentation in a CALL system using speech recognition technology [17]. In our study, we evaluated learners’ fluency from the following two perspectives:

- 1) The detection of Japanese phonemes using the database of Japanese English pronunciation of the speech recognition system

- 2) The Words Per Minute rate calculated by the sum of word units in a learner’s utterance detected by a segmentor in the speech system.

Automatic scoring of the fluency of learners’ utterances

Watching a video clip of each skit and listening to utterances from the video clip, students are required to repeat what they hear. They are also required to imitate the tempo of the utterances. After the English speech recognition system assesses the students’ utterances based on the perspective described in Section 5.3 and shows the Words Per Minute rates of both the model sounds and the students’ speech as shown in Fig 7, a screen display of the script is shown in order to indicate whether the students’ pronunciation is correct or not. The words which students pronounce correctly are colored black, while the words which students pronounce incorrectly are colored red as is shown in Fig. 8. Students are able to listen to their own utterances as well as to the model utterances in order for students to compare the differences of the tempo.



Figure 7: Screenshot of screen display of the Words Per Minute rates



Figure 8: Screenshot of the results of students’ shadowing

6. CONCLUSION

This study was limited to a report on the automatic

scoring of the fluency of learners' utterances as well as the development and refinement of the EFL CALL teaching aid. A course incorporating the refined EFL CALL aid was prepared for implementation in the fall semester, 2009. Verification of the prosody hypothesis will be discussed based on the analysis and results of pre- and post-testing, post-course evaluation and interviews with students following the course. This CALL teaching aid, at the present time, is available only as a standalone. Work is currently underway to store the contents of the teaching materials, including exams and assignments, in an e-learning server so that they will be accessible online—not only from the classroom but also from PCs in other locations, both on and off-campus. We also plan to collect data regarding students' access to the server and their scores from the online exams and assignments in order to gain a better understanding of students' progress and to determine what sections of the teaching materials require further improvement.

7. REFERENCES

- [1] Japanese Ministry of Land, Infrastructure, Transport and Tourism. **White Paper on Tourism**. , 2008.
http://www.mlit.go.jp/hakusyo/kankou-hakusyo/h20/kanko20_.html
- [2] H. Holec, **Autonomy and Foreign Language Learning**, Oxford: Pergamon, 1981.
- [3] D. Little, **Learner Autonomy 1: Definitions, Issues and Problems**, Dublin: Authentik, 1991.
- [4] P. Benson, & P. Voller, **Autonomy and Independence in Language Learning**, London: Longman, 1997.
- [5] E. L. Deci, & R. M. Ryan, **Intrinsic Motivation and Self-determination in Human Behavior**, New York: Plenum, 1985.
- [6] E. L. Deci, & R. M. Ryan, "The "What" and "Why" of Goal Pursuits: Human Needs and the Self-determination of Behavior", **Psychological Inquiry**, Vol.11, 2002, pp. 227-268.
- [7] Y. Ikuma, & R. Akahane-Yamada, "An Empirical Study on the Effects of Acoustic and Semantic Contexts on Perceptual Learning of L2 Phonemes", **Annual Review of English Language Education in Japan (ARELE)**, Vol. 15, 2004, pp. 101-108.
- [8] Y. Ishikawa, M. Kondo, & C. Smith, CALL for Hotel Careers: An Innovation in EFL Spoken Language Teaching Methodology. In L. G. Chova, M. Belenguer & C. Torres (Eds.), **INTED2009 Proceedings CD** (pp. 3225-3235). Valencia: International Association of Technology, Education and Development, 2009.
- [9] K. Tamai, **Strategic Effect of Shadowing on Listening Ability**, Tokyo: Kazamashobo, 2005. (in Japanese)
- [10] J. C. Richards, & R. Schmidt, **Longman Dictionary of Language Teaching and Applied Linguistics (3rd Ed.)**, Harlow: Pearson Education, 2002.
- [11] C. J. Fillmore, "On Fluency", In C. J. Fillmore, D. Kempler, & W. S-Y. Wang (Eds.) **Individual Differences in Language Ability and Language Behavior** (pp. 85-101). New York: Academic Press, 1979.
- [12] C. Brumfit, **Communicative Methodology in Language Teaching: The Role of Fluency and Accuracy**, Cambridge: Cambridge University Press, 1984.
- [13] P. Nation, "Improving Speaking Fluency", **System**, Vol. 3, 1989, pp. 377-383.
- [14] P. Linnon, "Investigating Fluency in EFL: A Quantitative Approach", **Language Learning**, Vol. 3, 1990, pp. 387-417.
- [15] D. Crystal, **A Dictionary of Linguistics and Phonetics (6th Ed.)**, Oxford: Blackwell Publishing, 2008.
- [16] N. Shimomura, N. Minematsu, Y. Yamauchi, & K. Hirose, "Automatic Scoring of Language Learners' Utterances Generated Through Shadowing Based on Bottom-up Clustering", **Technical Report of IEICE**, Vol. 108, No.116, 2008, pp. 55-60. (in Japanese)
- [17] S. Ikai, T. Sano, J. Ogata, & Y. Arika, "A CALL System with Segmentation and Evaluation Function of a User Utterance", **SLP40-02**, 2002, pp. 7-12. (in Japanese)