Investigation of Learners' Performance in English Conversation Mobile Learning System Using Learners' Own Daily Life Topics

Kae Nakaya
Tokyo Institute of Technology, Japan
knakaya@mr.hum.titech.ac.jp

Masao Murota
Tokyo Institute of Technology, Japan
knakaya@mr.hum.titech.ac.jp

Abstract: We have developed an English conversation mobile learning system for Japanese EFL (English as a foreign language) learners. We have already reported development of English conversation learning mobile system and effectiveness of the system for improving learners' WTC (willingness to communicate).. In this research, we focused on the effectiveness of the system for learners' performance. During the experiment, we collected response time and answer time as learners' performance. Response time is how long it takes for a leaner to start talking as an answer after a question or comment from the system. Answer time is the duration of the learner's answer. The seven-day experiment and results of the analysis revealed that the participants could decrease response time by continuous English conversation practices with lifelog-based topics. We had already concluded that this kind of practices with the system might be effective for improving some of learners' WTC. Consequently, we concluded that improving learners' WTC scores resulting from the practices with the system might be effective for decreasing response time.

Keywords: *English Conversation Learning, Second Language Acquisition, Mobile Learning, Learners' Performance*

(This page is used for the printed version of the proceeding. The body should start from the next page)

INTRODUCTION

Some of Japanese EFL (English as a foreign language) learners cannot speak English fluently because they are obsessed by speaking English perfectly and become nervous. Therefore they feel strong anxiety towards speaking English in the classroom, which causes hesitation in speaking (Hojo, 1996).

In order to make them feel relaxed and to motivate them to speak English more fluently, increasing their willingness to communicate (WTC) is important. According to the heuristic model by MacIntyre et al. (1998), WTC directly affects the frequency of second language (L2) use. Therefore, more active communication in L2, which leads to improvement in speaking ability, can be achieved by increasing the learner's WTC.

English conversation classes in Japan do not, however, sufficiently promote learners' WTC. Learners only practice and imitate conversation examples from textbooks during the class. Many teachers provide passive learning, and then the learners tend to only reproduce and memorize the contents. The learners have fewer opportunities to speak their opinions and express themselves in English.

In addition, because the topics used in the English conversation classes in Japan are fixed and far from students' real daily life, the classes are not enough interactive and attractive to promote their WTC. Pino (2009) suggested that, in order to encourage learners to speak English more frequently, teachers should choose topics that are related to the learners' own experiences and avoid the ritual domain. Yashima (2002) also revealed that in order to increase Japanese EFL learners' WTC, they need to have confidence in their L2 communication. In other words, English conversation teachers should offer topics that learners can talk on with confidence in a relaxed manner.

In order to offer English conversation learning opportunities that achieve both making learners feel relaxed and promote their WTC, we have already developed English conversation learning system using mobile devices and evaluated learners' motivation by the system (Nakaya et al. 2013). The system is based on the heuristic model by MacIntyre et al (1998) and offers pseudo-interactive and agreeable English conversation with lifelog-based topics. With this system learners can have conversations with a computer agent on their own daily life topics as if they are talking with another person. The system could make learners talk in English in a relaxed manner with motivation. Furthermore, the learners evaluated themselves to speak more fluently, express themselves better, and speak in a more relaxed manner as they continue to practice speaking.

In this paper, we focused on learners' performance on the system. In order to clarify whether or not an increase of learners' WTC score by the system is effective for learners' performance, we measured response time and answer time as learners' performance. Response time is how long it takes for a leaner to start talking as an answer after a question or comment from the system. Answer time is the duration of the learner's answer. Furthermore we investigated the relationship between WTC scores and learners' performance.

SYSTEM DESIGN

Overall system design

Figure 1 shows the overall system design. The system consists of an Android application that learners use (Figure 1(A)), a sub-system for offering lifelog-based topics (Figure 1(B)), and a sub-system for offering pseudo-interactive and agreeable English conversation (Figure 1(C)).

Using Android application (Figure 1(A)), learners can practice English conversation anywhere and anytime due to its mobilibility. Therefore it enables learners to practice more often.

A sub-system for lifelog-based topics (Figure 1(B)) consists of two server programs and Twitter. Twitter is an external program and we can collect and utilize learners' tweet by using external API. Our sub-system collects each learner's lifelog from Twitter and extracts the learner's interests or their daily life. After that the server programs automatically select topics that are similar to the extracted data. By this process it offers topics based on the interests and lives of learners (hereafter lifelog-based topics).

A sub-system for pseudo-interactive and agreeable English conversation (Figure 1(C)) consists of a server program, English conversation database and Google voice search. Google voice search is an external program and we can get text data of learners' speech input by using Android API. The server program first sends the question on lifelog-based topics that is selected the sub-system (Figure 1(B)). After a learner replies, the server program gets text data of the learner's speech input using Google voice search, selects a questions that depends on what the learner replies, and sends the question to the Android application (Figure 1(A)). By repeating this process, it offers pseudo-interactive and agreeable English conversation.

The targets of this system are undergraduate and graduate students who have achieved a TOEIC level C score (Educational Testing Service, 2012). The students should have already learned all basic grammar but tend to be

reluctant to speak English.

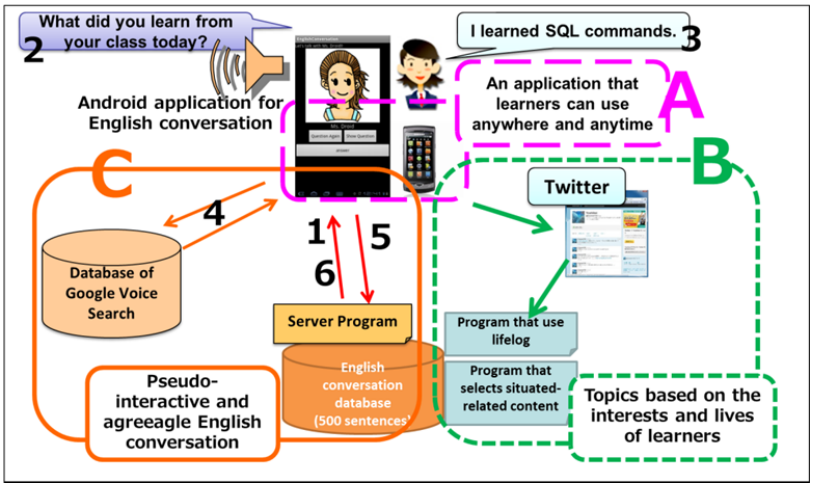

Figure 1. Overall system design

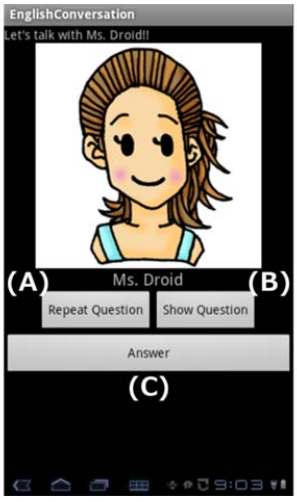

Figure 2. Basic mode screen

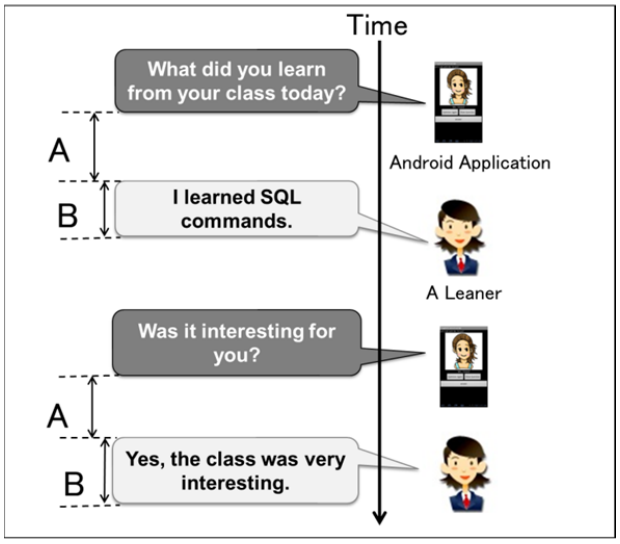

Figure 3. Response time (A) and answer time (B)

English conversation process and data collection using the Android application

Conversation learning occurs through the interaction between a learner and the system. Figure 2 shows the Android application screen used by learners during conversation.

First, after the learner launches the application, the application vocally asks a question on the lifelog-based topic. The learner can replay to the question by pressing the “Repeat Question” button (Figure 2(A)). If the learner wants to read a transcript of the question, pushing the button shown in Figure 2(B) will display the text, and pushing the button a second time will hide the text.

The learner can reply vocally after pushing the “Answer” button (Figure 2(C)). The reply will be sent after speech recognition, and the learner has a chance to reply again.

The application then asks the next question.

One conversation practice session consists of approximately 20 interactions. By repeating this process, learners practice English conversation that is more similar to in-person conversation.

Response time and answer time were collected by the Android application during the conversation practice. Figure 3 shows what response time and answer time are. Response time is how long it takes for a leaner to start talking after the end of the question from the system, which is shown in Figure 3(A). Answer time is the duration of the learner's answer, which is shown in Figure 3(B).

The Android application collected response time by measuring time between the time when the application finished playing the sound of the question and the time when the learner pushes "Answer" button (Figure 2 (C)). The Android application collected answer time by measuring the duration of voice recognition.

METHOD

Materials for the experiment

In order to compare English conversation learning of lifelog-based topics with traditional English conversation learning and to observe the effectiveness of lifelog-based topics, we prepared another 10 topics based on situations that learners often experience, such as conversations in a restaurant or airport. These topics are among the materials included in traditional English conversation practice texts in Japanese education. We defined these topics as general topics.

During the experiment, we collected the item scores about WTC ("fun", "anxiety", "expressing oneself", "motivation", "relaxation", and "fluency") and participants' performance (response time and answer time). The item scores were collected by five-point Likert scale. Response time and answer time were collected by millisecond.

Android mobile phones or tablets for collecting response time and answer time are GALAXY S SC-02B, MEDIAS WP N-06C, ICONIA TAB A100, REGZA Phone T-01C, HTC EVO WIMAX, Xperia arc SO-01C, and IS03.

Experiment

We conducted a seven-day experiment (from January 17th to 23rd, 2012). The participants were eight undergraduate students and graduate students (seven male and one female). The mean age was 22.5 (SD = 1.1). Four participants were presented with lifelog-based topics, and the others were presented with general topics. All the participants practiced two times a day. We called one time practice "session".

After using the application to practice English conversation, we analyzed the response time and answer time and compared the result with the result of WTC scores that we had already analyzed (Nakaya et al, 2013). We analyzed all of the data except for that for one participant, whose TOEIC score was not appropriate for the present research.

Analysis

We statistically compared the response time and answer time of both topics in order to clarify any improvement in the participants' performance due to the effectiveness of lifelog-based topics and pseudo-interactive and agreeable English conversation, which are features of the system.

Before analyzing response time and answer time, we normalized the data. First, we excluded some data as outliers that values were far from the mean value by more than two standard deviations. These outliers might be caused by the participants' distraction that was not related to the experiment. Moreover, we treated response time and answer time as logarithm scale because these data was human response. The process of this exclusion and using logarithmic scale was referred to the way by Juff (1998).

The results were analyzed by mixed three-way repeated measures ANOVA using three valuables. The between-subjects factor is "topic" which has two levels (lifelog-based and general) and the within-subject factors are "day" which has seven levels (from 1st day through 7th day) and "session" which has two levels (the first and second session). Figure 4 and 5 show changes in average response time and answer time by a logarithmic scale.

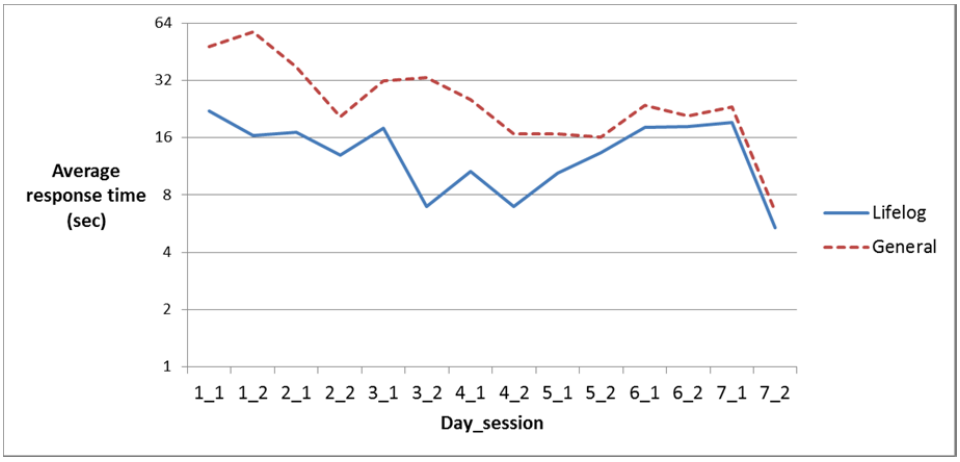

Figure 4. Average response time

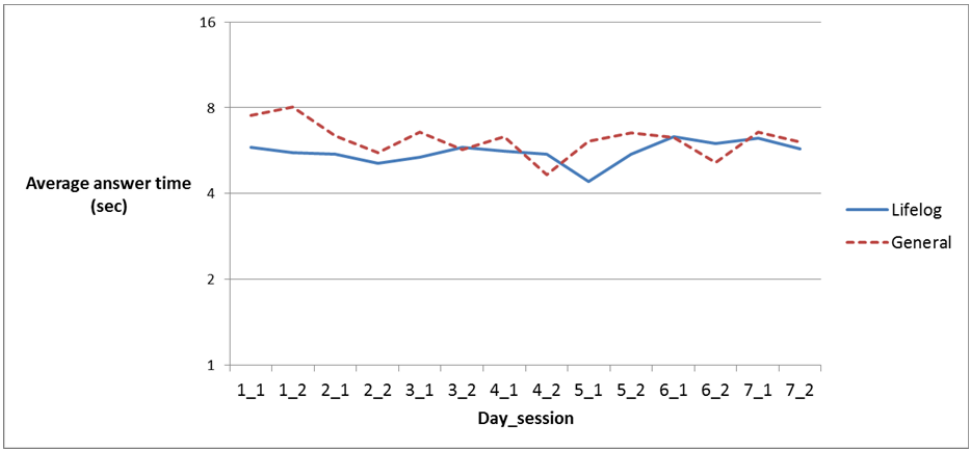

Figure 5. Average answer time

RESULT

Response time

The response time showed marginally significant differences between topics ($F(1,5) = 6.114$, $p < 0.1$). This means lifelog-based topics might be more effective to decrease response time than general topics. The average response time using lifelog topics in 7 days was 14.005 sec, while 26.968 sec for general topics.

The response time also showed statistically or marginally significant changes for within-subject factors.

In the day factor, statistically significant changes were observed ($F(6,30) = 5.090$, $p < 0.01$). Therefore we assessed the result with Ryan multiple comparison procedure. The analysis showed that the average response time in the 4th, 5th, and 7th day was significant lower than that in the 1st day.

In the session factor, marginally significant changes were observed ($F(1,5) = 6.030$, $p < 0.1$). The average response time in the 2nd session was lower than that in the 1st session.

The interaction between topics, days, and sessions did not show any significant differences.

Answer time

There were no significant differences for any factors in analysis of answer time.

DISCUSSION

The effectiveness of the system for response time

(1) The effectiveness of lifelog-based topics

According to significant differences in response time between topics, we concluded that lifelog-based topics might be effective for decreasing response time.

Early in the practice period, response time of the participants using lifelog-based topics was much better than that using general topics, as seen in Figure 4. According to Figure 4, participants especially using general topics grew accustomed to practicing English conversation with the system and they could decrease response time gradually. As a result, their response time tended to be approximated to that of the participants using lifelog-based topics in latter days, though interaction between topics and days didn't show any significant changes.

This result tends to be similar to the result of relaxation score which we had evaluated in the same experiment (Nakaya et al. 2013). In relaxation scores, interaction between topics showed significant changes and the 1st day and 3rd day scores between topics showed significant differences. Relaxation scores of the participants using lifelog-based topics were much better than those using general topics early in the practice period. The scores of the participants using general topics got improved gradually, but the participants using lifelog-based topics had high relaxation scores from beginning to end. We had concluded that high relaxation scores of those using lifelog-based topics from beginning to end might be caused by practice with lifelog-based topics, while improving relaxation scores of those using general topics gradually might be caused by repeating practices with the system. .

Consequently we concluded that practices with lifelog-based topics might be effective for decreasing response time from beginning. Learners using lifelog-based topics might answer faster than those using general topics because lifelog-based topics make them relaxed. They might not feel nervous because the topics are related to their own daily life and the topics are not formal. They might be able to decrease response time more than usual due to lifelog-based topics even if they were not used to practicing with the system.

(2) The effectiveness of continuous practice using pseudo-interactive and agreeable English conversation

According to statistically or marginally significant decreasing response time in both topics of 4th, 5th, and 7th day compared to that of 1st day, we concluded that continuous practice using pseudo-interactive and agreeable English conversation might be effective for decreasing response time.

These results might be similar to the result of "expressing oneself", "relaxation", and "fluency" scores which we had evaluated in the same experiment (Nakaya et al. 2013). These scores in latter days improved comparing with the scores in the 1st day for both topics.

We had concluded that the improvement of these three scores about WTC might be caused by continuous practice using pseudo-interactive and English conversation. The participants felt relaxed and evaluated themselves to express themselves better and to speak English fluently by repeating the practice numerous times.

Consequently, we concluded that the other element of the system that can make learners answer more immediately in English conversation learning might be offering repeating practice using pseudo-interactive and agreeable English conversation. This kind of English conversation practices might help learners feel relaxed and evaluate themselves to express themselves better and to speak English fluently. As a result they might have confidence in English conversation gradually and they can answer more immediately than before.

Future work for improving response time

Though response time of the participants using lifelog-based topics was lower than that using general topics, response time during the experiment was too long for interactive conversation despite lifelog-based topics, as seen in Figure 4. The 7 day average response time of the participants using lifelog-based topics was 14.005 second and that using general topics was 26.968 second.

The too long response time revealed that the system is not effective enough to make learners respond in English immediately. In the present system, after the question from the system, the participants could answer anytime. They might consider the answer carefully and it took too long to answer. The system should set the limitation of response time so that learners can practice responding immediately.

Future work for measuring answer time more accurately or another learners' performance

In terms of answer time, the design of the system might not be appropriate for accepting longer learners' speech input. The Android voice recognition is not acceptable for long sentences. From the beginning the participants might speak English more than the limitation of accepting by the voice recognition. If the participant speaks long sentences, the voice recognition of Android stops accepting. As a result, they might answer again and modify the answer shortly in order to adapt the answer to the limitation of voice recognition.

Consequently, we need to consider another automatic speech recognition technology for the system.

In addition, answer length partly depends on what kind of questions or comments the system offers. We should consider another data for measuring learners' performance, such as grammatical accuracy of learners' answer, word per minute scores of learners' answer, or some English proficiency tests.

CONCLUSION

In order to clarify effectiveness of our English conversation mobile learning system for learners' performance, we collected response time and answer time as learners' performance during an experiment and evaluated the data. The result revealed that the system might be effective for decreasing response time. The reason might be that improving some of learners' WTC ("expressing oneself", "relaxation", and "fluency") resulting from the features of the system might make learners respond more immediately than usual.

We will improve the method for offering interactive English conversation because the response time of both topics was too long for interactive and natural English conversation. The system needs to set the limitation of response time for learners and we should consider another method how to accept learners' speech input.

In addition, we have to improve the method of accepting learners' speech input and consider another data for observing learners' performance instead of answer time.

REFERENCES

Nakaya, K. and Murota, M. (2013). Development and Evaluation of an Interactive English Conversation Learning System with a Mobile Device Using Topics Based on the Life of the Learner. Research and Practice in Technology Enhanced Learning, 8(1), 65-89.

MacIntyre, P. D., Clèment, R., Dörnyei, Z., and Noels, K. A. (1998). Conceptualizing willingness to communicate in a L2: A situational model of L2 confidence and affiliation, In Modern Language Journal, 82, 545-562.

Hojo, R. (1996). A Study of Students' Anxiety over Classroom English (6). Bull. Joetsu Univ. Educ., 15(2), 495-506.

Pino, C. (2009). Overcoming Japanese EFL Learners' Fear of Speaking. University of Reading Language Studies Working Papers, 1, 55-63.

Yashima, T. (2002). Willingness to Communicate in a Second Language: The Japanese EFL Context. The Modern Language Journal, 86, I, 54-66.

Educational Testing Service. "PROFICIENCY SCALE", Retrieved June 27, 2013, from http://www.toeic.or.jp/toeic/pdf/data/proficiency.pdf

Juff, A. (1998). Main Verb Versus Reduced Relative Clauses Ambiguity Resolution in L2 Sentence Processing. Language Learning, 48(1), 107-147.