Development and Evaluation of Interactive English Conversation Learning System with a Mobile Device Using Topics Based on Learners' Life

Kae NAKAYA[a], Masao MUROTA[a]
[a]*Dept. of Human System Science, Graduate School of Decision Science and Technology, Tokyo Institute of Technology, Japan*
{knakaya, murota}@mr.hum.titech.ac.jp

Abstract: Motivating Japanese EFL (English as a Foreign Language) learners is essential if they are to learn English conversation effectively. We developed an English conversation mobile learning system based on a heuristic model of variables influencing the concept of "Willing to Communicate" by MacIntyre et al, with "interactive English conversation", "topics based on interests and life of learners", and "applications which learners can use anywhere and anytime". The results of experiments showed that the "interactive English conversation" style worked well in making learners feel relaxed and making them speak fluently.

Keywords: English Conversation Learning, Lifelog, Second Language Acquisition, Mobile Learning

Introduction

Japanese EFL (English as a foreign language) learners often have low motivation towards English learning. A Benesse Corporation survey held among junior high school students showed that 60 percent think of themselves as not good at English [1]. Similarly, many surveys of Japanese college students report that more than 65 percent of them dislike or hold a low motivation towards learning English [2][3].

In particular, Japanese EFL learners feel strong anxiety towards speaking English in the classroom, which causes hesitation in speaking [4]. As a result, they have low motivation to express what they think. The National Institute for Educational Policy Research shows that more than 60 percent of junior high school students don't like to speak or write what they think [5]. Therefore, they are not active in English conversation learning.

To motivate learners to learn English conversation, it is important to increase their "willingness to communicate" (WTC). According to the "Heuristic Model of Variables Influencing WTC" by MacIntyre et al, WTC directly affects the frequency of L2 (Second Language) use [6]. Thus, more active communication in the L2, and by extension, improvement in speaking ability, is achieved by increasing the learner's WTC. WTC is influenced by two immediate situational factors, which are the "desire to communicate with a specific person" and the "state of communicating self-confidence".

The "desire to communicate with a specific person" is used to mean that people want to communicate with a person who they know well, who holds similar opinions to them, is somebody who they often meet, or are attracted to. When people enjoy talking, the

conversation becomes interactive. People tend to talk more if the interlocutor agrees with what they say and an immediate reply is given. If the conversation is strongly-interactive and comfortable, people will desire to talk more.

The "state of communicating self-confidence" is affected by whether one has experienced the situation or developed knowledge about the topic at the moment of conversation. People can communicate with self-confidence on topics they know well. However, when the situation is unfamiliar to them, their confidence is reduced. In the case of Japanese EFL learners, the effect of L2 communication confidence on the WTC is particularly strong [7].

English conversation classes in Japan do not, however, do enough to promote learners' WTC. Learners only practice and imitate conversation examples from textbooks during class. The materials of the classes encourage passive learning, so learners tend to only reproduce the contents and not express themselves. The materials are not as interactive and attractive as learning with real interlocutors.

In addition, the topics used in class do not necessarily motivate all learners because English conversation teachers can only offer a few topics in one class. Pino suggests that in order to encourage learners to speak English more, teachers should choose topics that are related to learners' own experiences and avoid ritual domain [8].

In this paper, we developed a system based on the model by MacIntyre et al, which aims to increase motivation of Japanese EFL learners. We propose pseudo-interactive and agreeable English conversation to motivate the "desire to communicate with a specific person". We also propose topic-based English conversation about learners' daily lives to make learners acquire the "state of communicating self-confidence". We propose 2 kinds of contents: topics based on learners' lifelogs, and topics based on common situations learners face.

To encourage learners to talk in English more frequently, we developed an Android application that learners can use to talk whenever and wherever they want. Japanese EFL learners rarely talk in English in daily life because Japan is a linguistically homogeneous nation. Therefore, with respect to this point, (we believe) the use of a portable device is more suitable than a stationary PC.

The targets of this system are bachelor and masters course students who have achieved a TOEIC level C score [9]. The students are supposed to have already learned all basic grammar, but tend to be reluctant to speak in English.

1. Related work

1.1 Pseudo-interactive English conversation materials

ELIZA[10] is a computer program which offers pseudo-conversation. It replies to users using pattern matching and, in some cases, makes users feel as if they are interacting with another person.

SpeakGlobal is a website that offers learners interactive English conversation learning with robots using artificial intelligence [11]. To communicate, the robots use speech recognition and speech synthesis technology. This way, robots can talk with their own voice, and learners can reply with theirs, making the conversation feel intimate because of its interactivity. In other words, learners are motivated to talk because they can talk directly with the robot and not have to use the keyboard.

However, both of them cannot offer interaction based on the learner's life. The pattern matching of ELIZA is primitive. Therefore the program can only offer general responses and the interaction often becomes unnatural after some time. As for SpeakGlobal, these

robots can only talk about general topics and cannot adapt to learner diversity. In contrast, in order to offer interactive English conversation learning that aims to motivate learners to speak more English, our system features topics based on learners' lifelogs.

1.2 Language learning environments that have a learner-centered design

Much research has tried to tailor materials to a diverse set of learners in order to motivate them to learn English. These kinds of materials are based on a "learner-centered design" that personalizes each material's content. This research has collected learners' data, such as their profiles, location, time, and learning log as parameters, for the personalization of the learning materials. They proposed learning systems based on leaners' data and offer learning contents that are more related to learners [12][13][14]. Personalization is an important factor in the design of language learning materials.

Recently, mobile learning systems have become a hot topic in this field, with many research papers on mobile learning [12][13][14][15]. Mobile devices enable systems to collect learners' data effectively, while also enabling learners to learn anywhere and anytime, because of the compact nature of these devices.

We have adapted personalization and learner-centered design to English conversation learning because present research in mobile language learning has predominantly focused on improving vocabulary skill.

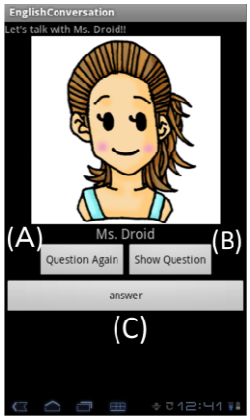

Figure 1 : Basic mode screen

Figure 2 : Confirmation mode screen

2. Design of English conversation learning

2.1 The features of the system

Our system offers a type of English conversation that is closer to in-person conversation. Acting as an interface to the system, the Android application asks a question that the learner answers immediately. Moreover, in the case of "yes" or "no" questions, it includes a branch containing the next question or comment. Learners can therefore feel as if they are talking with the application interactively.

In addition, because conversation topics are related to learners' daily life, the learning is personalized to every learner. If the topics are general, learners are sometimes unfamiliar with them, and, as a result, are reluctant to speak. However, learners can talk more comfortably and with greater self-confidence in our system because of the personalized nature of interaction.

2.2 The procedure of English conversation learning with the system

Conversation learning occurs with the interaction between learner and system. First, after the learner launches the application, the application asks a question based on the learner's life which the learner listens to. The learner can listen again by pressing the "Question Again" button (Figure 1(A)). If the learner wants to read a transcript of what the application says, pushing the button shown in Figure 1(B) will display the transcript onscreen. Pushing the button again will hide the transcript.

After pushing the "answer" button (Figure 1(C)), the learner replies using English speech recognition. The learner has a chance to reply again during the confirmation mode (Figure 2). Third, the application speaks the next question. One instance of conversation learning consists of about 20 interactions. We call this one conversation session.

3. Design of the system

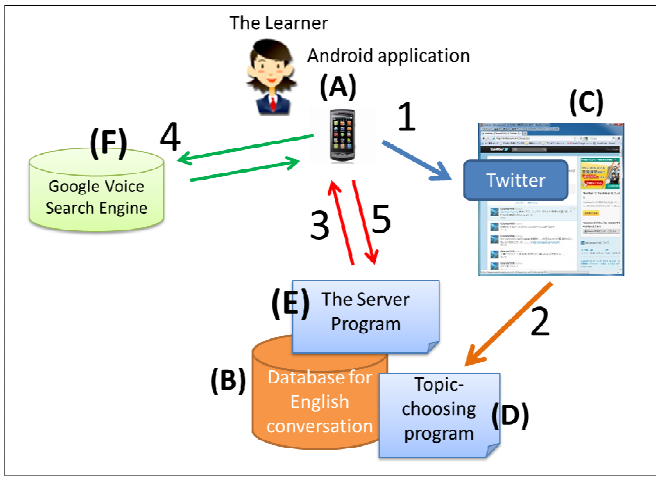

Figure 3 : English conversation learning process

3.1 The outline of the system

Our system offers pseudo-interactive English conversation, and consists of the elements shown in Figure 3. When the learner launches the application, the server sends a question based on the learners' life from the database to the application (Figure 3(3)). The application vocalizes the question using speech synthesis. If the learner replies in English, the application accepts the speech input using Google's Voice Search application (Figure 3(4)) and sends the input to the server (Figure 3(5)). The server then sends the application a new question or comment based on the learners' reply (Figure 3(3)). By repeating this process, the learner can practice English conversation.

1	What classes did you have today? Please tell me one of them.
Example learner responses	I had a database class today.
2	What did you learn from your class today?
Response	I learned SQL commands.
3	Was it interesting for you?
Response	Yes, the class was very interesting.
4	That's good! So you enjoyed the class, right?
Response	Yes, it was a fun.
5	How many students take the same class?

Figure 4 Conversation example in the case of "yes"

1	What classes did you have today? Please tell me one of them.
Example learner responses	I had a database class today.
2	What did you learn from your class today?
Response	I learned SQL commands.
3	Was it interesting for you?
Response	No, it's boring.
4	Oh, really? So please continue studying, and you will feel the class is getting interesting!
Response	All right. I'll do my best.
5	How many students take the same class?

Figure 5 Conversation example in the case of "no"

3.2 Design of sentences for the conversation

In this system, we offer "pseudo-interactive English conversation" in three ways. First, we arrange the order of questions or comments in a session to make learners feel as if they are "talking naturally" with the application whatever they speak (an example can be seen in Figure 4 and Figure 5). Second, as the accuracy of voice recognition in Android is not acceptable for longer sentences, the system has a branch for the next question or comment in the case of the learner's reply including "yes" or "no", which are correctly recognized. If the learner replies "yes" to the question (Figure 4 (3)), another question is asked (Figure 4 (4)). Third, the comments include phrases that aim to cheer learners up, and make them feel more relaxed (Figure 4 (4), Figure 5 (4)).

Table 1: List of conversation categories

classes, homework, sports, job hunting, research, conference, laboratory, part-time job, daily life, TV game, TV program, travel, cooking, pets, presentation, house-moving, seminar, shopping, music, mobile phone, movie, diet, comic books, programming, favorite books

3.3 Topics based on the learner's lifelog

We prepared 420 sentences categorized into 25 categories of topics relating to the daily life or interests of the learners (Table 1). Each category comprises of several conversation sessions. The system chooses the appropriate category based on the learner's lifelog, which is collected from the learner's twitter account (Figure 3(2)). The system calculates the degree of similarity between the categories and the obtained lifelog, and then chooses the category with the highest score. In case all scores are zero, the system chooses the "Daily Life" category.

Table 2 : List of conversation topics based on commonly occurring situations

order in a restaurant, reserve a room of a hotel, tell what you feel to a doctor, talk in an office, talk in a sightseeing tour, self-introduction, talk with staff in a shopping-mall, talk in the airport, talk in a bank or post office, talk about party invitation

3.4 Topics based on situations that often happen to learners

We prepared 10 kinds of topics based on situations that often happen to learners, such as conversations in a restaurant or an airport (Table 2). The system offers 2 topics a day at random. First the system explains to the learner the situation behind the conversation. After that, the system offers a situative question and the learner replies.

4. Evaluation

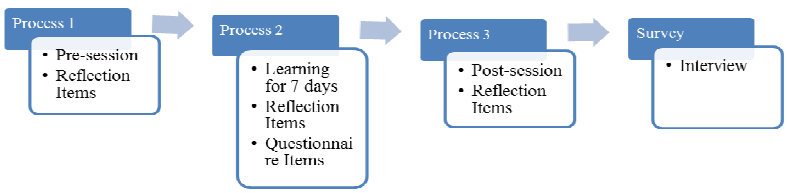

Figure 6 : Experimental procedure

4.1 The aim of the experiments

We conducted an experiment for 7 days, aiming to observe changes in attitude during English conversation. The subjects were split into two groups, one using topics based on learners' lifelogs, and one using situated general topics. After they had used the application to practice English conversation, we measured the differences between the two groups.

4.2 Reflection Items and Questionnaire Items

To observe the 2 kinds of effectiveness, we collected reflections and a questionnaire after every conversation session about willingness of English conversation and about their impression towards the sentences, using a 5-point Likert scale. We defined reflection items concerning willingness as "Reflection Items", which include fun, anxiety, expressing oneself, motivation, relaxation, and fluency. We defined question items about the subjects' impression of English sentences as "Questionnaire Items", which include interests, daily life, friendliness, difficulty with understanding, and difficulty in replying.

4.3 The outline of the experiments

We conducted the experiment from January 17th to 23th, 2012. The subjects were 8 bachelor and masters course students (7 male and 1 female). The mean age was 22.5 (*SD*=1.1). Four subjects used the application with lifelog based topics, and the others used one with general situative topics. We analyzed all data except for 1 subject whose TOEIC score didn't match the target of this research.

Table 3 : Mixed two-way factorial ANOVA table for the Reflection Part

Reflection Items	Source of variation	Sum of squares	DOF	Mean square	F-value	p-value
Fun	Between topics	0.0134	1	0.0134	0.0131	0.9132
	Between terms	0.1801	1	0.1801	0.5368	0.4966
	Interaction	0.4301	1	0.4301	1.2822	0.3089
Anxiety	Between topics	0.8932	1	0.8932	0.3890	0.5602
	Between terms	0.8932	1	0.8932	2.7796	0.1563
	Interaction	0.3575	1	0.3575	1.1125	0.3398
Expressing oneself	Between topics	0.9301	1	0.9301	0.7030	0.4400
	Between terms	1.2515	1	1.2515	5.6142	0.0640 +
	Interaction	0.1801	1	0.1801	0.8077	0.4100
Motivation	Between topics	0.3810	1	0.3810	0.7748	0.4190
	Between terms	0.1205	1	0.1205	0.1948	0.6774
	Interaction	0.3348	1	0.3348	0.5411	0.4950
Relaxation	Between topics	0.8218	1	0.8218	0.5833	0.4795
	Between terms	2.2057	1	2.2057	5.9397	0.0589 +
	Interaction	0.4557	1	0.4557	1.2272	0.3184
Fluency	Between topics	0.9301	1	0.9301	0.4933	0.5138
	Between terms	5.5372	1	5.5372	29.8636	0.0028 **
	Interaction	0.5372	1	0.5372	2.8973	0.1495

Table 4 : A part of mixed two-way factorial ANOVA table for the Questionnaire Part

Questionnaire Items	Source of variation	Sum of squares	DOF	Mean square	F-value	p-value
Daily Life	Between topics	0.8932	1	0.8932	1.0400	0.3546
	Between terms	3.0807	1	3.0807	7.3115	0.0426 *
	Interaction	0.4557	1	0.4557	1.0816	0.3460
Difficulty with understanding	Between topics	1.3393	1	1.3393	0.8855	0.3899
	Between terms	2.7515	1	2.7515	11.0984	0.0207 *
	Interaction	0.2515	1	0.2515	1.0144	0.3601
Difficulty in replying	Between topics	2.1488	1	2.1488	0.7243	0.4336
	Between terms	2.1488	1	2.1488	15.1681	0.0115 *
	Interaction	0.2917	1	0.2917	2.0588	0.2108

4.4 The results of the analysis

4.4.1 Reflection Items

We defined the mean value of the 1st and 2nd day as “the value in the beginning” and the mean value of the 6th and 7th days as “the value at the end”. We statistically compared the value at the beginning with the one at the end to observe any improvement in willingness. The result was analyzed by ANOVA with two independent valuables: topics (lifelog-based and situative general) and terms (beginning and end).

Statistically or marginally significant improvements were observed for three Reflection items on the term factor (Table 3). “Fluency” showed improvement with significance ($F(1,5)=29.8636$, $p<.01$). “Express yourself” and “Relax” exhibited a marginally significant improvement ($F(1,5)=5.6142$, $F(1,5)=5.9397$, $p<.10$ for both). All the values of the three items at the end are higher than the ones in the beginning. In addition, we analyzed the value of Reflection Items from the pre- and post-session of general topics. The Wilcoxon signed-rank test for both applications did not reveal any prominent changes.

4.4.2 Questionnaire Items

The analysis showed statistically significant changes at three items on the term factor. “Daily Life”, “Difficulty with understanding”, and “Difficulty in replying” significantly improved ($F(1,5)=7.3115$, $F(1,5)=11.0984$, $F(1,5)=15.1681$, $p<.05$). The values of both topics at the end are lower than the ones in the beginning.

4.5 The results of the survey

We took a survey of the subjects’ impression and suggestions for improvement of the application after the experiment. The survey showed that the subjects became more relaxed during conversation or motivated to speak English after 7 days of learning with the application because subjects had to think and reply immediately by “speaking” to the system’s spoken questions. Moreover, some subjects said that if they were to use the application over a longer term, it might decrease their anxiety.

On the other hand, sometimes the subjects felt that the interaction was unnatural in an obvious way. Moreover, the subjects that used lifelog-based topics often found topics to be the same or too similar. Most subjects wanted functionality to check and improve their English. They wanted the application to check the grammar of what they spoke, record and playback the conversation, or practice vocabulary they did not yet know.

5. Conclusion and future works

5.1 The effectiveness of interactive conversation by speaking

According to the analysis of Reflection Items and the survey, we conclude that the pseudo-interactive English conversation which was offered in both topics worked effectively in making learners feel that they were speaking English fluently, in a relaxed manner, all using the their own English knowledge. The reason is that every time the system sends a question or comment, and the learner replies immediately, the system replies instantaneously. The subjects repeated this processes many times and became used to expressing themselves in English. The result that showed statistically significant decreases in “difficulty with understanding” and “difficulty in replying” supports this conclusion.

5.2 Necessity to improve the use of lifelog for English conversation learning

We have to improve the way we offer topics and English sentences based on learners' life and improve lifelog collection. This is because the Questionnaire Items of "Daily Life" showed a statistically significant lower change especially around topics based on the lifelog obtained from subjects' tweets. The usefulness of the Twitter lifelog strongly depends on how learners express their life or how often the learners post tweets. The application that offered the lifelog-based topics sometimes chose the same topic. Therefore some subjects became disinterested. We have to not only propose another way of collecting lifelogs but also other ways of offering topics that interest learners or learners want to talk about.

5.3 Support learners in learning English

As many subjects said, the system needs features for learning English, such as grammar checking, offering practice vocabulary for new words, and reflecting on the conversation. These features will motivate learners to not only learn English conversation but also acquire new English skills during conversation learning.

References

[1] Benesse Corporation. (2009). The first survey on junior high school English (in Japanese). http://benesse.jp/berd/center/open/report/chu_eigo/seito_soku/pdf/data_00.pdf(Accessed May 21, 2012)

[2] Maeda, K., Tsumura, S., Koiso, K., Kagada, T. (2010). Survey on English learning motivation –using text mining- (in Japanese). A collection of theses of Osaka University of Commerce, 6(2), 77-90.

[3] Tsumura, S. (2011). Relationship between self-esteem and demotivating factors in English learning. In A. Stewart(Ed.), JALT2010 Conference Proceedings. Tokyo: JALT. 806-815.

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

[5] National Institute for Educational Policy Research. (2012). The survey about writing English (in Japanese). http://www.nier.go.jp/kaihatsu/tokutei_eigo_2/tyousakekka.pdf (Accessed May 21, 2012)

[6] MacIntyre P. D., Clèment R., Dörnyei Z., 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.

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

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

[9] Educational Testing Service. "PROFICIENCY SCALE" , http://www.toeic.or.jp/toeic/pdf/data/proficiency.pdf (Accessed May 21, 2012).

[10] Weizenbaum, Joseph. (1966). ELIZA – a computer program for the study of natural language communication between man and machine. Communications of the ACM, 9, 1, 36-45.

[11] SpeakGlobal. http://speakglobal.co.jp/ (Accessed May 21, 2012)

[12] Ogata, H., Li, M., Hou, B., Uosaki, N. & Yano, Y. (2011). Effectiveness of Ubiquitous Learning Log System. Proc. of the 19th International Conference on Computers in Education (ICCE 2011), Chiang Mai, Thailand. http://www.nectec.or.th/icce2011/program/proceedings/pdf/C4_S7_303S.pdf (Accessed May 21, 2012)

[13] Peterson, S. A., Markiewicz, J. & BjØrnebekk, S. S. (2009). Personalized and Contextualized Language Learning: Choose When, Where and What. Research and Practice in Technology Enhanced Learning, 1-42.

[14] Edge, D., Searle, E., Chiu, K., Zhao, J. & Landay, J. A. (2011). MicroMandarin: Mobile Language Learning in Context. Proc. Of the CHI 2011 Conference on Human Factors in Information Systems, 3169-3178.

[15] Stockwell, G. (2010). Using Mobile Phones for Vocabulary Activities: Examining the Effect of the Platform. Language Learning & Technology, 12(2), 95-110.