DEVELOPMENT AND EVALUATION OF AN INTERACTIVE ENGLISH CONVERSATION LEARNING SYSTEM WITH A MOBILE DEVICE USING TOPICS BASED ON THE LIFE OF THE LEARNER

KAE NAKAYA

Graduate School of Decision Science and Technology, Tokyo Institute of Technology
Ookayama 2-12-1, Meguro-ku, Tokyo, 152-8552, Japan
knakaya@mr.hum.titech.ac.jp
http://www.mr.hum.titech.ac.jp

MASAO MUROTA

Graduate School of Decision Science and Technology, Tokyo Institute of Technology
Ookayama 2-12-1, Meguro-ku, Tokyo, 152-8552, Japan
murota@mr.hum.titech.ac.jp
http://www.mr.hum.titech.ac.jp

Motivating Japanese EFL (English as a foreign language) learners is essential in order for them 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. The system has three features: an application that learners can use anywhere and anytime, topics based on the interests and lives of learners, and pseudo-interactive and agreeable English conversation. The results of an experiment and analysis revealed that the topics based on the interests and lives of learners might be effective for talking in English with relaxed from the beginning and for talking with motivation, while a pseudo-interactive and agreeable English conversation style might work well in helping learners evaluate themselves to speak more fluently, express themselves better, and speak in a more relaxed manner as they continue to practice speaking.

Keywords: English conversation learning; lifelog; second language acquisition; mobile learning.

1. Introduction

Japanese EFL (English as a foreign language) learners often have low motivation. A Benesse Corporation survey of junior high school students revealed that 60% of the students thought of themselves as having poor English skills (Benesse Corporation, 2009). Similarly, a number of surveys of Japanese college students revealed that more than 65% of students disliked or had low motivation toward learning English (Maeda, Tsumura, Koiso, & Kagada, 2010; Tsumura, 2011).

In addition, Japanese EFL learners feel strong anxiety towards speaking English in the classroom, which causes hesitation in speaking (Hojo, 1996). As a result, these learners have low motivation to express their opinions. The National Institute for

Educational Policy Research revealed that more than 60% of junior high school students did not like to speak or write their opinions (National Institute for Educational Policy Research, 2012) and so were not active in English conversation learning.

In order to motivate learners to learn English conversation, increasing their willingness to communicate (WTC) is important. According to the heuristic model of variables influencing WTC by MacIntyre, Clèment, Dörnyei, and Noels, WTC directly affects the frequency of second language (L2) use (MacIntyre, Clèment, Dörnyei, & Noels, 1998). Thus, more active communication in L2, and by extension, improvement in speaking ability, is achieved by increasing the learner's WTC.

Willingness to communicate is influenced by two immediate situational factors, "desire to communicate with a specific person" and "state communicative self-confidence" (MacIntyre et al., 1998).

"Desire to communicate with a specific person" refers to the situation in which a person wants to communicate with a person he/she knows, a person who has a similar opinion, a person he/she meets frequently, or a person to whom he/she is attracted. 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 want to talk more.

"State communicative self-confidence" is affected by whether an individual has experienced the situation or developed knowledge about the topic at the moment of conversation. People are able to communicate with confidence on topics they know well. However, when the situation is unfamiliar, their confidence is reduced. In the case of Japanese EFL learners, the effect of L2 communication confidence on WTC is particularly strong (Yashima, 2002).

English conversation classes in Japan do not, however, sufficiently promote learners' WTC. Learners only practice and imitate conversation examples from textbooks during class. The class materials provide passive learning, and so the learners tend to only reproduce the contents and not express themselves. The materials are not as interactive and attractive as learning with actual 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 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 (Pino, 2009).

In the present paper, we developed a system based on the model by MacIntyre et al., which aims at increasing the motivation of Japanese EFL learners. The system aimed at promoting "state communicative self-confidence" and "desire to communicate with a specific person".

In order to achieve these objectives, we incorporated three features into the system: an application that learners can use anywhere and anytime, topics based on the interests and lives of learners, and pseudo-interactive and agreeable English conversation. An application that learners can use anywhere and anytime aims at encouraging learners to

speak in English more frequently. Topics based on the interests and lives of learners aim at helping learners to communicate with confidence. In order to offer topics based on the interests and lives of learners, we prepare topics based on learners' lifelogs, and topics based on common situations faced by learners. Pseudo-interactive and agreeable English conversation aims at encouraging the desire to communicate with a specific person.

For an application that learners can use anywhere and anytime, we developed an Android application that learners can use to speak whenever and wherever they want. Japanese EFL learners rarely speak English in daily life because Japan is a linguistically homogeneous nation. It is necessary for them to get more opportunity to practice English conversation, not only in the classroom, but also in daily life continuously. Mobile devices are suitable for learning during small intervals of daily life (Sharples, 2000; Sharples, Taylor, & Vavoula, 2005), can adapt to learner diversity and encourage learners' motivation due to their flexibility (Uden, 2007). They enable learners to learn anywhere and anytime. Therefore, the use of a portable device offers learners more opportunity to speak English.

In topics based on the interests and lives of learners, we expected that if learners are given a topic that is related to their own lives or their interests, they might be able to speak about the topic in English with confidence. Learners are sometimes unfamiliar with general topics, and, as a result, are reluctant to speak. However, learners can speak more comfortably and with greater confidence using the proposed system because of the personalized nature of interaction.

In pseudo-interactive and agreeable English conversation, we expect that learners will want to talk more with the system if the system and the learners are able to speak with their own voices interactively and agreeably, as in a real 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.

2. Related Research

2.1. *Pseudo-interactive English conversation materials*

ELIZA (Weizenbaum, 1966) is a computer program that offers pseudo-conversation. ELIZA 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 using robots with artificial intelligence (SpeakGlobal, 2012). To communicate, the robots use speech recognition and speech synthesis technology. In this manner, robots can speak 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 speak directly with the robot without having to use the keyboard.

However, neither ELIZA nor SpeakGlobal offers interaction based on the learner's life. Since the pattern matching of ELIZA is primitive, the program can offer only general

responses, and the interaction often becomes unnatural. With respect to SpeakGlobal, the robots can only speak about general topics and cannot adapt to the diversity of learners. In contrast, in order to offer interactive English conversation learning that aims at motivating learners to speak English more frequently, the proposed system features topics based on the learners' daily life.

2.2. *Mobile learning with personalized or contextual contents*

Mobile learning has a great deal of potential to provide learners with contextual or personalized materials for language learning. Numerous researches of mobile learning have been conducted in an attempt to tailor materials to a diverse set of learners in order to improve their English vocabulary or reading ability and to motivate them to learn English. Chen and Chung (2008), and Chen and Hsu (2008) presented a personalized mobile English learning system by adapting the materials of the system to individual vocabulary or reading skill. de Jong, Specht, and Koper (2010), Edge, Searle, Chiu, Zhao, and Landay (2011), Huang, Huang, Huang, and Lin (2011), and Ogata et al. (2011) offered contextual vocabulary learning contents that are related to learners' location. Peterson, Markiewicz, and BjØrnebekk (2009) utilized individual parameters, such as learners' skill, age, experience, interests, and so on.

Although those researches of mobile learning have conducted vocabulary or reading skill that are necessary to acquire second language, Kukulska-Hulmes (2007) referred to necessity of more researches for improving speaking and listening skill. Speaking skill is especially difficult for non-native country learners to improve because they have little opportunity to talk in target language. Cheng, Hwang, Wu, Shadiev, and Xie (2010) proposed a mobile and online system that enables learners to collaborate one another in English practice activities to improve English abilities including speaking. The learners posted their experience of the location by using mobile learning system. After that they discussed the posts and experience in the classroom using online system. Liu (2009) and Liu et al. (2010) revealed that the learners could improve speaking and listening ability and retain their motivation of English learning by English learning activity based on the learners' location. One of the practices that the learners employed was talking about location-based topics with a virtual learning tutor. The learners took a picture of the 2-D bar code beside their location using a mobile device and sent it to the server. The server recognized where they were and the virtual learning tutor offered conversation topics that were related to the learners' location. They could practice contextual English conversation as if they were talking with another person.

However, few researches investigated the mobile learning system that offers English conversation materials based on each learner's daily life to improve learners' speaking ability. While location-based topic which is one of the contextual materials offers learners practice of vocabulary or conversation which is used where they are, conversation practice about learners' own daily life enables learners to talk about themselves in second language. This is more personalized material and learners can practice as if they are talking in daily life. Learners could feel that that kind of materials

is like more real conversation and their motivation might increase. Cheng et al. (2010) also mentioned the necessity of offering language learning contents based on learners' daily life. Therefore we proposed the system that collected learners' lifelog and adapt the learning contents to the lifelog.

3. Design of English Conversation Learning with an Android Application

Figure 1 shows the overall system design. The proposed system provides greater opportunity for English conversation to learners using an Android application (Figure 1(A)), offers topics based on the interests and lives of learners using Twitter and two programs on a server (Figure 1(B)), and offers pseudo-interactive and agreeable English conversation using an English conversation database and Google voice search (Figure 1(C)).

With respect to opportunity (Figure 1(A)), in the present study, we developed an Android application that allows learners to access the system anywhere and anytime. This portability provides learners with greater opportunities to practice English conversation.

Conversation learning occurs through the interaction between a learner and the system. Figures 2 and 3 show the Android application screen used by learners during conversation.

First, after the learner launches the application, the application vocally asks a question based on the learner's life. 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,
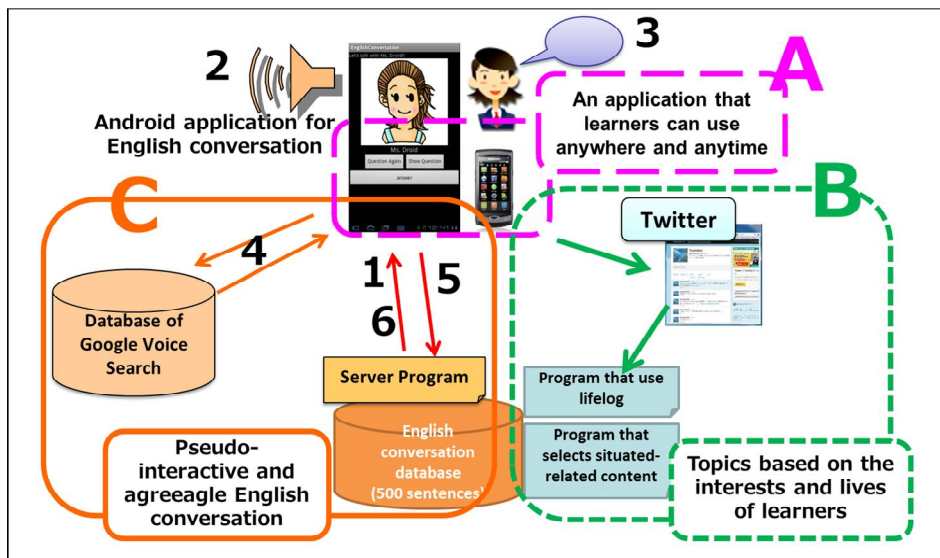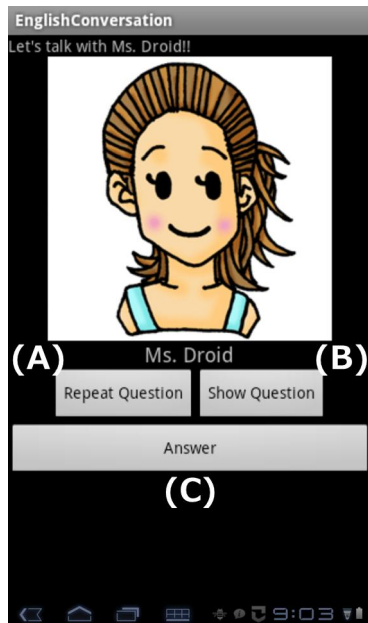

Figure 1. Overall system design.
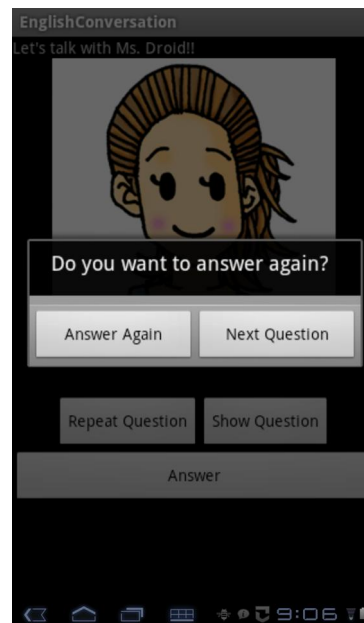
Figure 2. Basic mode screen.

Figure 3. Confirmation mode screen.

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 during the confirmation mode (Figure 3).

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.

The conversation that the system provides has two features: topics based on the interests and lives of learners, and pseudo-interactive and agreeable English conversation. The first topics based on the interest and lives of learners aim at increasing communication confidence (Figure 1(B)), which is described in the next section. The second pseudo-interactive and agreeable English conversation aims at encouraging communication with a specific person (Figure 1(C)), as described in Section 5.

Table 1. Details of conversation topics and sub-topics based on the lifelog of the learner.

Conversation main topics (sub-topics)
classes, homework (doing, finishing), sports (training, having a match, watching a match, sports team, favorite sports), job hunting, research, conference, laboratory, part-time job, daily life (holiday, sick, oversleep, New Year), TV game, TV program, travel (travel plan, sightseeing experience), cooking, pets, presentation (preparation, finishing, listening to), house-moving (new house, place where they want to live), seminar, shopping (recent purchases, places where they usually go shopping), music (concert, karaoke), mobile phone, movies, diet, comic books, programming, favorite books

4. Topics that Our Research Designed

4.1. *Topics based on the interests and lives of learners*

4.1.1. *Design of topics based on the interests and lives of learners*

In order to offer topics based on the interests and lives of learners, the system collects the lifelog of each learner in advance and offers an appropriate conversation topic based on each lifelog. We defined 25 conversation topics that might be relevant to the life of the learner. We defined these topics as the main topics, some of which have sub-topics. Table 1 lists the main topics and sub-topics. Each main topic or sub-topic has from one to two conversation sessions. One conversation session consists of from 11 to 19 questions or comments. Consequently, there are 37 conversation sessions containing 420 sentences. Seven to 17 sentences were prepared for each session.

Each topic comprises several conversation sessions (sub-topics). For example, the sports topic has five conversation sub-topics: training, having a match, watching a match, sports team, and favorite sports.

We abbreviate topics based on the interests and lives of learners to lifelog-based topics from this Section to Section 7.4 except for the titles of Sections.

4.1.2. *How to offer topics based on the interests and lives of learners*

The system offers topics based on the lifelog of each learner.

Figure 4 shows the process of selecting an appropriate topic based on a lifelog. Main topics and sub-topics have a number of keywords. The system selects the first sentence of the topic having keywords that are most closely related to the lifelog.

Table 2 reveals the learners to which the system should offer the first sentence. For example, if the system chooses the sports-match topic for a learner having id 1, the system stacks id 1 (learner id) and id 52 (id of the first sentence of the sports-match topic) into a learner-topic stack table (Table 2).
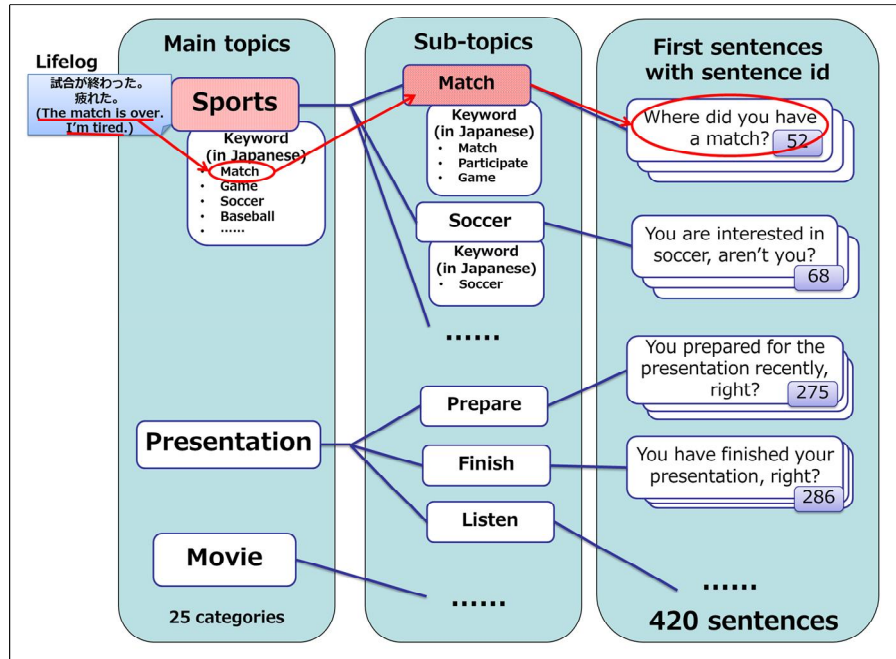
Figure 4. Relation between topics, sub-topics, first sentences, and keywords.

The details of how the system selects the topic are described in the following. The server program receives tweets from the learner's Twitter account every day. The system then calculates the degree of similarity between the keywords of the topics and sub-topics and every tweet using the "Match()" function of MySQL. Subsequently, the system selects the topic or sub-topic with the highest score and assigns it to that particular tweet. If no topics were similar to that tweet, then the system always chooses one of the two normal "Daily Life" topics which we prepared. The system then stacks the first sentence id of the topic and the learner id into the learner-topic stack table. When the learner launches the application, the server program refers to the topic stack table for the topic and sends the first sentence of the topic to the application. (The process after launching is explained in Section 5).

Table 2. Learner-topic stack table.

Learner id	First sentence id
1	52
2	275
3	68
4	1

The system collects in advance the tweets from each learner as their lifelog. In the present study, the system requires learners only to submit a tweet to collect the learners' lifelog with minimal effort from the learner.

4.2. *General topics based on situations that learners often experience*

4.2.1. *Design of general topics based on situations that learners often experience*

In order to observe topics based on the interests and lives of learners, we prepared 10 topics based on situations that learners often experience, such as conversations in a restaurant or airport. (Table 3 lists all of the topics.) These topics are among the materials included in traditional English conversation practice texts in Japanese education. We defined these topics as general topics. Each topic has one conversation session, which consists of from 11 to 19 questions or comments.

The order of questions and comments is pre-defined. The system first explains to the learner the situation behind the conversation and then offers a situation-related question to which the learner replies. The system and the learner then continue the conversation.

4.2.2. *How to offer general topics based on situations that learners often experience*

The system randomly presents two topics a day. The system preferentially selects the topic that has been offered to the learner the least number of times.

5. Pseudo-interactive and Agreeable English Conversation

5.1. *Design of pseudo-interactive and agreeable English conversation*

In this system, in order to offer pseudo-interactive and agreeable English conversation, we pre-define the order of sentences in a session. Figure 5 shows one example of the order of sentences in a conversation (Figure 5(1)). The proposed system offers questions, comments, and explanations about situations behind the conversation.

Moreover, since the accuracy of voice recognition in Android is not acceptable for longer sentences, the system has a branch for the next sentence if the learner's reply includes "Yes" or "No" (Figure 5(2)), which is correctly recognized. If the learner replies "Yes" to the question, then another question is asked.

Many sentences are designed as questions (Figure 5 (3), (4)). These aim to push

Table 3. Details of conversation topics based on situations that learners often encounter.

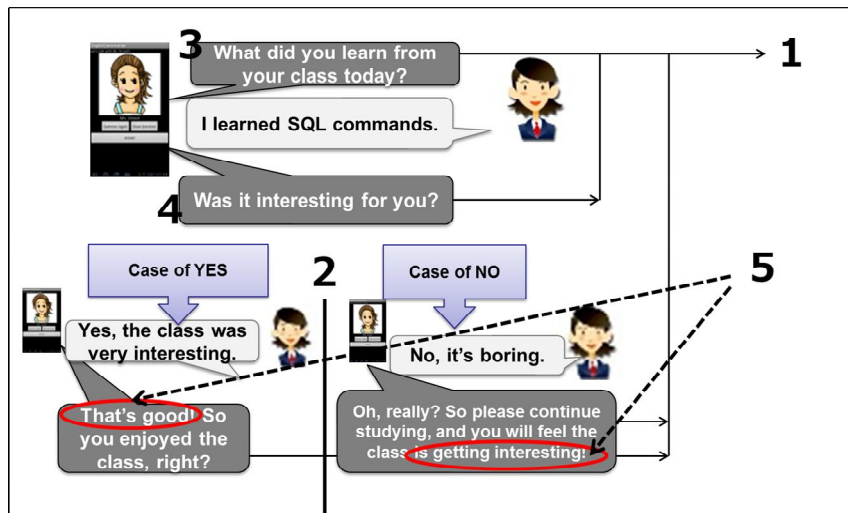| Conversation topics |
| --- |
| ordering food in a restaurant, reserving a hotel room, telling a doctor how you feel, talking in an office, talking during a sightseeing tour, self-introduction, talking with staff in a shopping mall, talking in an airport, talking in a bank or post office, talking about a party invitation |

Figure 5. Example of conversation with topic "Class".

learners' output. One of the problems in second language acquisition might be that when learners output, they cannot utilize second language knowledge that they have already known. By output of a second language, learners can pay attention to grammatical structure and recognize what they don't know or what part of their knowledge they cannot control, i.e. output encourage learners to improve syntactic processing (Swain, 2005). Moreover, if the learners output over many times, they can create greater automaticity (Gass, 2003). Thus, they can output with little effort, control the knowledge, and acquire second language. Therefore, we designed many sentences as questions so as the learners were given more opportunity to output and could control their knowledge that they have already acquired.

In addition, other sentences include phrases that are intended to cheer up the learner and put them at ease (Figure 5(5)). In particular, phrases that are intended to cheer up the learner are often presented after a learner says "No". This is very important because Japanese EFL learners need to learn to express their opinions without reluctance but they have significantly less tendency to approach argument situations than Americans (Prunty, Klopf, & Ishii, 1990). Japanese people generally tend to be shy and hesitate to clearly express their opinions, which may prevent fluent, relaxed conversation. In the present study, we aimed to make learners express themselves in a relaxed manner and to increase "desire to communicate" of learners by offering agreeable responses. Moreover, the system did not indicate learners' grammatical or speech errors during the conversation, so that the learners would not worry about mistakes too much. Too frequent correction might cause anxiety and prevent "desire to communicate".

5.2. *How to offer pseudo-interactive and agreeable English conversation*

5.2.1. *Detail of the process between the server and the application*

Figure 1 shows the pseudo-interactive and agreeable English conversation process. The application receives a sentence from the English conversation database and sends the learner's reply to the database. When the application accepts the learner's speech input, the application receives the text of the input by voice recognition and Google voice search.

The process is as follows. First, when the learner launches the Android application, the server of the system presents a question or comment (Figure 1(1)). The application then vocalizes the question or comment through speech synthesis (Figure 1(2)). The learner listens to the question and immediately answers (Figure 1(3)). The application then accepts the learner's speech input using Google voice search (Figure 1(4)) and sends the input to the server (Figure 1(5)). The server then sends the application a new question or comment based on the learner's reply (Figure 1(6)). Learners practice English conversation by repeating this process.

5.2.2. *Design of tables in the English conversation database*

We created two tables in the English conversation database. One includes and saves each sentence of the application (English content table, Table 4), and the other defines the order of sentences (sentence order table, Table 5). Moreover, the system uses the learner-topic stack table (Table 2) just before presenting pseudo-interactive and agreeable English conversation.

The server presents a sentence according to the learner-topic stack table (Table 2), the order table (Table 5), and the English content table (Table 4). For example, if the learner with id 4 launches the application, the system refers to the learner-topic stack table (Table 2). The first sentence id for the learner with id 4 is 1. Consequently, the server refers to the English content table (Table 4) and sends sentence with id 1, "What classes did you have today? Please tell me one of them." to the application. The system then accepts the learner's speech input from the application and refers to the order table for the id of the

Table 4. Partial list of English conversation sentences.

Sentence id	English sentence
1	What classes did you have today? Please tell me one of them.
2	What did you learn from your class today?
3	Was it interesting for you?
4	That's good! So you enjoyed the class, right?
5	Oh, really? So please continue studying, and you will feel the class is getting interesting!
6	How many students take the same class?

Table 5. Partial sentence order table.

Id	Keyword	Next id
1	-	2
2	-	3
3	Yes	4
3	No	5
4	-	6
5	-	6

next question. The order table (Table 5) shows that the question following question with id 1 is question with id 2. The server then presents question with id 2.

If the next question has a branch, the server checks the learner's speech input and presents a new question depending on the learner's input. For example, the server presents question with id 3, and the learner replies "No, it's not interesting for me". After accepting the learner's speech input, the server refers to the order table, which shows that the question following question with id 3 is either question with id 4 or question with id 5. The server then checks the learner's input and whether the input includes the word "No". The question following question with id 3 with "No" as a keyword is question with id 5. Consequently, the server presents question with id 5: "Oh, really? So please continue studying, and you will feel the class is getting interesting!" (Table 4).

6. Evaluation

6.1. *Procedure of the experiments*

We conducted a seven-day experiment, in an attempt to observe improvements in attitude during English conversation. We observed the attitude by collecting reflection items regarding on the participants' motivation and questionnaire items regarding on their impression of the system's English sentences.

The experimental process is shown in Figure 6. In process 1, we conducted a conversation session and collected reflection items to observe the participants' pre-willingness. In process 2 the participants practiced seven-day learning using the application and then we collected reflection items and questionnaire items after every conversation session. We interviewed the participants in process 3.

The participants were split into two groups. One group was presented with topics based on learners' lifelogs, and the other group was presented with general topics. After using the application to practice English conversation, we analyzed the scores of reflection items and questionnaire items.
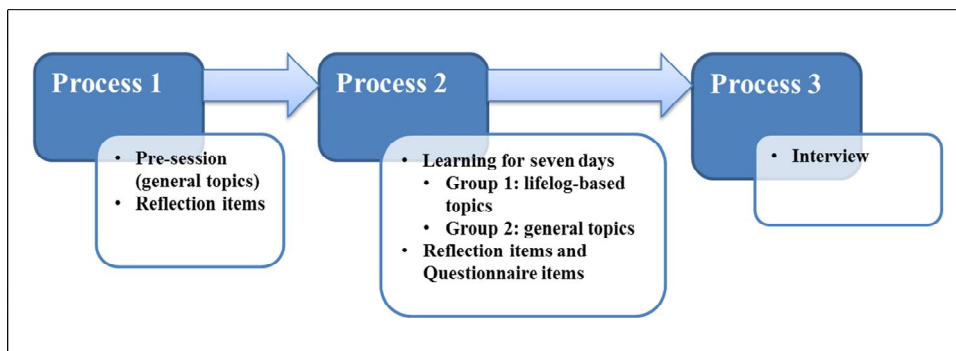
Figure 6. Process of experiment.

6.2. *Reflection items and questionnaire items*

In order to clarify the effectiveness of the proposed system, a questionnaire using a five-point Likert scale on the willingness to engage in English conversation and on the impression of the presented sentences was presented after each conversation session.

We defined reflection items as items related to willingness, which include fun, anxiety, expressing oneself, motivation, relaxation, and fluency.

We also defined questionnaire items as relating to the participants' impression of English sentences, which include interests, daily life, friendliness, difficulty in understanding, and difficulty in replying.

Figures 7 and 8 show the screens with which learners input the scale. After a pull-down menu for an item is selected from the item selection screen (Figure 7), the

Figure 7. Item selection screen.

Figure 8. Scale input screen.

application presents a scale input screen (Figure 8) and the learner selects the scale for the item. The participants were asked to input the scale for the item about the practice previously.

6.3. *Outline of the experiment*

We conducted the 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 analyzed all of the data except for that for one participant, whose TOEIC score was not appropriate for the present research.

6.4. *Analysis results*

6.4.1. *The analysis method*

We statistically compared the scores of reflection items and questionnaire items of both topics in order to clarify any improvement in willingness due to the effectiveness of lifelog-based topics and pseudo-interactive and agreeable English conversation.

Table 6. Mixed two-way repeated measures ANOVA table for reflection items.

Reflection item	Source of variation	Sum of squares	DOF	Mean square	F-value	p-value
Fun	Between topics	0.021	1	0.021	0.019	0.896
	Among days	3.517	6	0.586	0.531	0.780
	Interaction	4.905	6	0.817	0.741	0.621
Anxiety	Between topics	4.688	1	4.688	0.662	0.453
	Among days	4.517	6	0.753	1.667	0.164
	Interaction	3.232	6	0.539	1.192	0.337
Expressing oneself	Between topics	4.252	1	4.252	1.001	0.363
	Among days	7.707	6	1.284	2.792	0.028*
	Interaction	1.421	6	0.237	0.515	0.792
Motivation	Between topics	0.310	1	0.310	0.218	0.660
	Among days	1.320	6	0.220	0.338	0.912
	Interaction	9.401	6	1.567	2.405	0.051+
Relaxation	Between topics	4.000	1	4.000	1.099	0.342
	Among days	7.753	6	1.292	6.558	0.000****
	Interaction	4.467	6	0.744	3.779	0.006**
Fluency	Between topics	4.511	1	4.511	0.790	0.415
	Among days	14.142	6	2.357	6.165	0.000****
	Interaction	3.183	6	0.530	1.387	0.252

Table 7. Mixed two-way repeated measures ANOVA table for questionnaire items.

Questionnaire item	Source of variation	Sum of squares	DOF	Mean square	F-value	p-value
Interests	Between topics	0.021	1	0.021	0.008	0.934
	Among days	1.058	6	0.176	0.128	0.992
	Interaction	4.405	6	0.734	0.533	0.779
Daily life	Between topics	3.072	1	3.072	0.690	0.444
	Among days	6.994	6	1.166	1.110	0.380
	Interaction	2.014	6	0.336	0.320	0.921
Friendliness	Between topics	7.864	1	7.864	4.900	0.078+
	Among days	1.934	6	0.322	0.405	0.870
	Interaction	3.505	6	0.584	0.734	0.627
Difficulty in understanding	Between topics	5.526	1	5.526	0.909	0.384
	Among days	10.981	6	1.830	7.063	0.000****
	Interaction	1.308	6	0.218	0.841	0.548
Difficulty in replying	Between topics	9.313	1	9.313	0.935	0.378
	Among days	10.737	6	1.790	4.780	0.001***
	Interaction	1.717	6	0.286	0.764	0.604

The results were analyzed by mixed two-way repeated measures ANOVA using two valuables. The between-subjects factor is "topic" which has two levels (lifelog-based and general) and the within-subjects factor is "day" which has seven levels (from 1st day through 7th day). Table 6 and 7 show the results for reflection items and questionnaire items. Figures 9 through 14 show changes in average scores in each reflection item and Figures 15 through 19 show those in each questionnaire item. A detailed analysis of these results is contained in the next subsections.

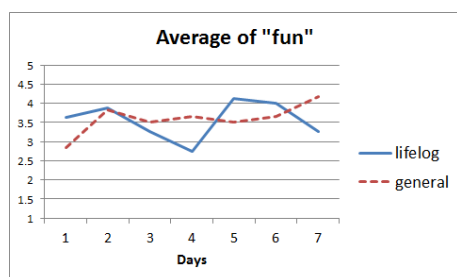Moreover, we statistically compared the reflection item scores of both topics in pre-

Figure 9. Average score of "fun" in reflection items.
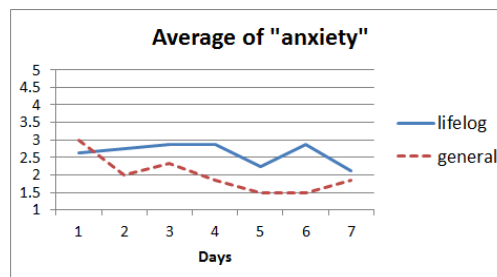

Figure 10. Average score of "anxiety" in reflection items.
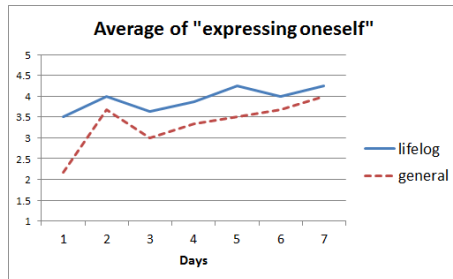
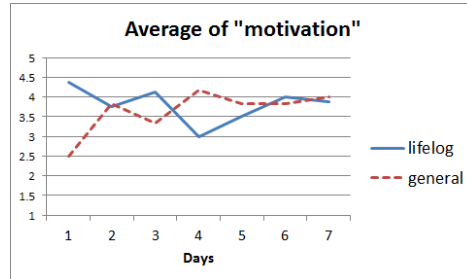Figure 11. Average score of "expressing oneself" in reflection items.

Figure 12. Average score of "motivation" in reflection items.

session to compare the scores between topics. A t-test showed no significant differences ($t(5)=0.434$ for "fun", $t(5)=3.287$ for "anxiety", $t(5)=2.535$ for "expressing oneself", $t(5)=1.767$ for "motivation", $t(5)=0.105$ for "relaxation", $t(5)=1.581$ for "fluency", $p > 0.05$ for all). There were no differences between topics in willingness before 7-day practice.

6.4.2. *The analysis result for effectiveness of topics based on the interests and lives of learners*

We looked at the items that showed significant changes between topics to clarify any improvement resulting from the effectiveness of lifelog-based topics. Significant differences between topics show the influence of topics on willingness.

In terms of reflection items, statistically or marginally significant differences between topics are shown partly in "motivation" and "relaxation". In both items, interaction between "topic" and "day" showed statistically or marginally significant changes ($F(6,30) = 2.405$ and $p < 0.1$ for "motivation", $F(6,30) = 3.779$ and $p < 0.01$ for "relaxation"). Therefore we did tests of simple main effects.
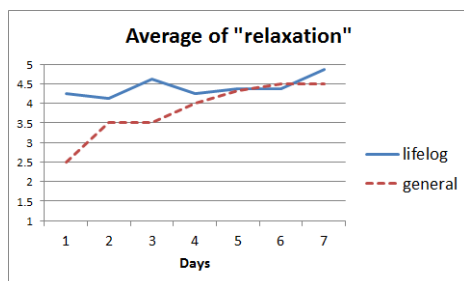

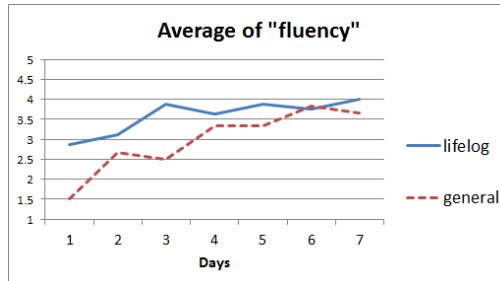Figure 13. Average score of "relaxation" in reflection items.

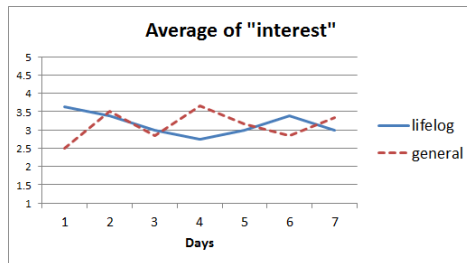Figure 14. Average score of "fluency" in reflection items.

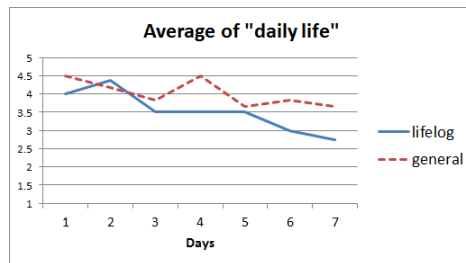Figure 15. Average score of "interest" in questionnaire items.

Figure 16. Average score of "daily life" in questionnaire items.

As for "motivation", a statistically significant difference was shown for "topic" factor on the 1st day (F(1,35) = 7.916, p < 0.01) and the score of lifelog-based topics was higher than that of general topics, as seen in Figure 12. A marginally significant difference was shown for "topic" factor on the 4th day (F(1,35) = 3.065, p < 0.1) and the score of general topics was higher than that of lifelog-based topics, as also seen in Figure 12.

As for "relaxation", a statistically significant difference was shown for the topic factor on the 1st day (F(1,35) = 7.623, p < 0.01) and the score of lifelog-based topics was higher than that of general topics, as seen in Figure 13. A marginally significant difference was shown for the topic factor on the 3rd day (F(1,35) = 3.150, p < 0.1) and the score of lifelog-based topics was higher than that of general topics, as also seen in Figure 13.

In terms of questionnaire items, marginally significant differences between topics are shown in "friendliness" (F(1,5) = 4.900, p < 0.1). The scores of general topics are higher than those of lifelog-based topics, as seen in Figure 17.

6.4.3. *The analysis result for effectiveness of pseudo-interactive and agreeable English conversation*

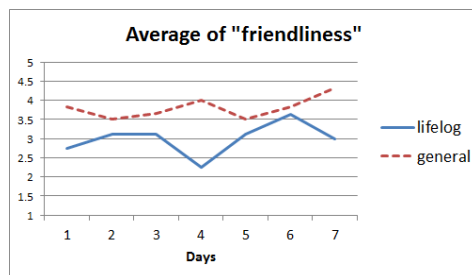We observed the items that showed significant changes for the day factor to clarify any

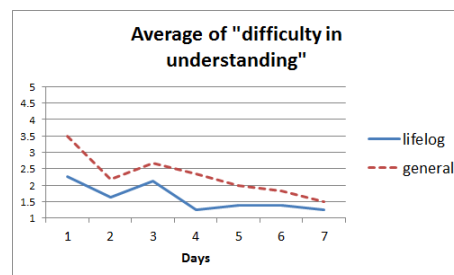Figure 17. Average score of "friendliness" in questionnaire items.

Figure 18. Average score of "difficulty in understanding" in questionnaire items.
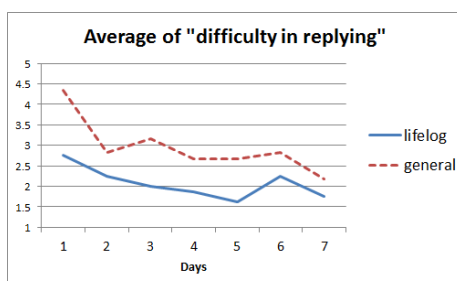
Figure 19. Average score of "difficulty in replying" in questionnaire items.

improvement due to effectiveness of pseudo-interactive and agreeable English conversation. We offered pseudo-interactive and agreeable English conversation to the participants of both topics. Therefore we could clarify the effectiveness of pseudo-interactive and agreeable English conversation by observing changes within days.

In terms of reflection items, significant changes for the day factor are observed in "expressing oneself", "relaxation" and "fluency" ($F(6,30) = 2.792$ and $p < 0.05$, $F(6,30) = 6.558$ and $p < 0.000$, $F(6,30) = 6.165$ and $p < 0.000$). Therefore we assessed the result with Ryan multiple comparison procedure.

As for "expressing oneself", the analysis showed that the 7th day score was significant higher than the 1st day score for both topics, as seen in Figure 11.

As for "relaxation", the analysis showed that in both topics, the 3rd, 4th, 5th, 6th and 7th day scores were significant higher than the 1st day score and 7th day score was also higher than the 2nd day score, as seen in Figure 13. Moreover, according to the simple main effect test in 6.4.2, the result for the day factor of general topics also showed significant changes ($F(6,30) = 9.184$, $p < 0.000$). The result with Ryan multiple comparison procedure showed the 3rd, 4th, 5th, 6th and 7th day scores improved significantly from the 1st day score for general topics, as also seen in Figure 13.

As for "fluency", the 3rd, 4th, 5th, 6th and 7th day scores were significant higher than the 1st day score for general topics, as seen in Figure 14.

In terms of questionnaire items, significant changes for the day factor are observed in "difficulty in understanding" and "difficulty in replying" ($F(6,30) = 7.063$ and $p < 0.000$, $F(6,30) = 4.780$ and $p < 0.001$). Therefore we assessed the result with Ryan multiple comparison procedure.

As for "difficulty in understanding", the 2nd, 4th, 5th, 6th and 7th day scores were significant higher than the 1st day score and the 7th scores were higher than the 3rd score, as seen in Figure 18.

As for "difficulty in replying", the 2nd, 3rd, 4th, 5th, 6th and 7th day scores were significant higher than the 1st day score, as seen in Figure 19.

6.4.4. *Interview results*

After the experiment, we interviewed the participants regarding their impressions of the application and suggestions for improvement. We asked questions to the participants face to face, and they answered freely. Table 8 shows the list questions asked.

Table 8. The list of questions of the interview.

Question categories	Questions	The results of the interview	
English sentences of the system	Difficulty in listening to what the system vocalized	Not difficult to listening to for all the participants.	
	Difficulty in understanding the sentences or impression of the sentences	Lifelog-based topics	Easy to understand (100 %), but sometimes the reply from the system was obviously unnatural (25%).
		General topics	Easy to understand (100 %).
	The variety of topics during the experiment	Lifelog-based topics	Sometimes similar (100%).
		General topics	Appropriately various (100 %).
	Other topics that the participants want to talk about in English	More detail contents about sports: e.g. about strategy that the sports team took (71 %), more detail contents about food; e.g. what they ate today or where they ate lunch or dinner (57 %), the contents about actors or actresses (28 %).	
Usability	Usability and any improvements of the system	All the participants agree that the system is easy to use.	
Effectiveness of the system	The effectiveness of the system for increasing willingness to communicate in English	Lifelog-based topics	It is effective for fun, motivation, and relaxation if appropriate lifelog-based topics are offered (75 %), only effective for relaxation (25 %). Longer-period practice might increase anxiety (50 %).
		General topics	Effective for expressing by encouraging output (67 %), only effective for relaxation (33 %).
	The differences between the system and traditional English conversation texts	Lifelog-based topics	The system was more effective for expressing oneself (100 %).
		General topics	The system was more effective for practicing English conversation because they had to answer immediately.

The interview revealed that 75 % of the participants using lifelog-based topics described they enjoyed practicing the conversation if the topics that our system offered were really related to the tweet on their Twitter account. A half of the participants said that if they were to use the application over a longer period, their anxiety might decrease.

In addition, the participants of both topics became more relaxed during conversation or more motivated to speak English after seven days of learning using the application, because participants had to immediately think and vocally reply to the system's spoken questions. They had to express their opinion or what they thought immediately.

On the other hand, the participants of lifelog-based topics sometimes felt that the interaction was obviously unnatural. Moreover, the participants often found topics to be the same or too similar. The reason of this is that the participants' tweets were not sometimes similar to topics in the database of the system and it offered "Daily life" topic. Therefore, at first the participants enjoyed the conversation very much because the topic was related to their own life. However, if the topic was presented several times, the participants lost interest.

Most participants wanted the ability to check and improve their English. They wanted the application to check their grammar, record and playback the conversation, or practice unfamiliar vocabulary. Moreover, some of the participants could not express in English what they wanted to say during conversation and wanted the application to recommend example answers when needed.

7. Discussion

7.1. *Overview of the analysis results*

After the analysis of Section 6.4, we concluded that lifelog-based topics and pseudo-interactive and agreeable English conversations, which were the main features of our system, were effective in improving some parts of willingness. Lifelog-based topics were effective for relaxation and motivation during practice English conversations from the beginning of the experiment period. Pseudo-interactive and agreeable English conversation was effective for evaluating oneself to express oneself better and to talk in English in a more relaxed and fluent manner. In the following sections, we discuss the effectiveness of our system in detail, and outline necessary improvements to each feature as well as our future research plans.

7.2. *Topics based on the interests and lives of learners*

According to the analysis of reflection and questionnaire item scores in Subsection 6.4.2, we concluded that lifelog-based topics might be effective for "relaxation" and "motivation" starting early period of the practice.

The "relaxation" scores of the participants using lifelog-based topics were statistically or marginally significantly higher than ones using general topics on the 1st and 3rd days, which is early in the practice period. Participants using lifelog-based topics had the opportunity to speak in English in a relaxed manner from the beginning. Participants

using general topics grew accustomed to practicing English conversation with the system and gradually became more relaxed with each day, but participants using lifelog-based topics had "relaxation" scores of 4 or higher from beginning to end, as seen in Figure 13. Consequently, lifelog-based topics might be effective for relaxation.

The "motivation" scores of the participants using lifelog-based topics were statistically or marginally significantly higher than ones using general topics on the 1st day. This also showed that lifelog-based topics might be effective for "motivation" even if the participants are not used to practicing using the system yet.

However, the method of presenting topics and English sentences based on the learner's life as well as the method of lifelog collection should be improved.

First, the 4th day "motivation" score for reflection items shows that we should improve how to offer lifelog-based topics. The present system offering the lifelog-based topics sometimes selected the same topic during the experiment. On the 4th day, the system only offered lifelog-based topics at a rate of 50 % and offered "Daily life" topics otherwise. Moreover, the half of those lifelog-based topics had already been offered before. Therefore some participants using lifelog-based topics became disinterested. This was the only item that showed a significant decrease over the practice period for lifelog-based topics. "Fun" in reflection items and "interest" and "friendliness" in questionnaire items for lifelog-based topics were lowest on the 4th day, as seen in Figure 9, 15, and 17. These low scores might be due to the same reason. By the end of the experiment, the system was properly offering lifelog-based new topics again (87.5%), and the "motivation" for lifelog-based topics increased once again.

Our system utilized Twitter in order to collect learners' lifelog. The usefulness of the Twitter lifelog is strongly dependent on how learners express aspects of their lives and/or on how often learners post tweets. We must develop other methods for collecting lifelogs and for offering topics that are of interest to learners, or about which learners want to speak. Another possible method is one in which learners can select topics that they would like to talk about if there are no similarities between their lifelogs and the topics in the system's database, i.e. the topic selection would be learner-driven.

Second, we also need to improve English sentences for topics based on learners' lifelog. As seen in Subsection 6.4.2, "friendliness" in questionnaire items for lifelog-based topics was marginally significantly lower than for that of general topics. One of the reasons for this might be that the answers to some questions in the lifelog-based topics do not vary from day to day. For example, in sports topic, the system asks learners "Which of the teams do you like best?" Learners' favorite teams do not change from day to day. Learners might enjoy answering this kind of question at first, but if the question is offered several times, they might become bored because their answers do not change. Consequently, we need to improve the questions to accommodate static topics, so that learners can give variable answers, e.g. "Have you heard anything about your favorite team recently?"

7.3. *Pseudo-interactive and agreeable English conversation*

According to the analysis in Subsection 6.4.3, the pseudo-interactive and agreeable English conversation was effective in making learners feel that they could speak English more fluently and use their English knowledge to express themselves better in a relaxed manner.

As seen in Figures 13 and 14, "relaxation" and "fluency" increased during the first several days rapidly, especially in general topics. This might be attributed to the fact that participants grew used to practicing English conversation using the system.

"Relaxation" and "fluency" continued to increase for the latter half of the experiment period, while "expressing oneself" increased each day from the beginning of the experiment. The reason for this might be that with repeated exposure to the pseudo-interactive and agreeable English conversation component, participants became better able to express themselves. In our system, each time the system presents a question or comment, the learner has to reply immediately and the system responds instantaneously. If the participants repeated this processes numerous times, they could create greater automaticity (Gass, 2003) and the result was an increase in their ability to speak English with little effort. As a result they felt that they could express themselves better, more fluently, and in a relaxed manner.

Consequently, we concluded that pseudo-interactive and agreeable English conversation might effectively encourage learner output and that is why the results of learners' self-evaluations showed increased fluency, expressivity and relaxation.

On the other hand, we have to improve the pseudo-interactive and agreeable English conversation of our system because based on the interviews we conducted with participants, some of those who used the lifelog-based topics, felt that the reply from the system was overly unnatural. The order of the responses during a conversation session was pre-defined and the present system could only recognize and adapt to "yes" or "no" answers. Therefore, the system could only respond to the learner's answers in this way. The conversation flow of lifelog-based topics depends on the learner's answer, while the flow of general topics is stable to some extent and we can presume what learners will say because the situation behind each conversation for a general topic is clear. In lifelog-based topics, the system is limited by the learners' answers.

In order to make the system more adaptive to learners' answers and make the interaction more natural, the system needs to recognize correctly what learners speak and to provide the proper response automatically. Therefore we need to consider automatic speech recognition technology and natural language processing.

In the present research we developed the system as an Android application aimed to offer an application that learners can use anywhere and anytime. However the voice recognition of Android is not acceptable for longer sentences. Many researches have conducted better performance of automatic speech recognition technology, including for mobile devices (Lee et al., 2012). We need to consider utilizing better automatic speech recognition technology to make the system realizing precisely what learners say.

In addition, we have to consider integrating natural language processing into our system in order to make the interaction more natural, especially in lifelog-based topics.

7.4. *Other future work for improving our research*

In the present research, we focused on the willingness to communicate, because increasing learners' motivation is important to support learning of a second language. During the 7-day experiment some parts of willingness to learn English conversation increased. However, the scores that we collected in the present research depended on the participants' subjectivity and therefore the effectiveness of our system for improving English conversation skills is unclear.

The improvement of learners' skills is the final goal of our research. Therefore in future work, we should measure, observe, and analyze the change of their skill after improving our system in terms of what we stated in Section 7.2 and 7.3. It is also necessary to measure the participants' ability before the experiment and analyze the ability scores with ANCOVA by using the ability as covariate to remove the influence of pre-skill or pre-willingness of participants on effectiveness of our system.

We also need to conduct more elaborate experiments, for longer periods and with more participants. First, based on the interviews with participants, half of the participants using lifelog-based topics said that if they were to use the application over a longer period, their anxiety might decrease. Longer term experiments could show more clearly a change in "anxiety" scores. Moreover, in the present research we compared the system of lifelog-based topics with that of general topics, but we should compare our system with traditional English conversation texts with more participants in order to reveal more clearly which of the elements is effective in increasing learners' motivation and their ability to converse in English.

8. Conclusion

We developed a mobile English conversation learning system to motivate Japanese EFL learners. This system has three features: an application that learners can use anywhere and anytime, topics based on the interests and lives of learners, and pseudo-interactive and agreeable English conversation.

The experiment and analysis revealed that the effectiveness of topics based on the interests and lives of learners might lie in being able to speak in English in a more relaxed way from the very beginning of practice and with more motivation. The results also showed that English practice with pseudo-interactive and agreeable English conversations might be effective for evaluating oneself, for expressing oneself better and for speaking in English more fluently and in a more relaxed manner.

In future study, we will improve the method for collecting lifelog and for offering lifelog-based topics during practice and also re-design the pseudo-interactive and agreeable English conversation. At that time we plan to conduct more elaborate studies.

References

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

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.

Chen, C. M., & Hsu, S. H. (2008). Personalized intelligent mobile learning system for supporting effective English learning. *Educational Technology & Society, 11*(3), 153–180.

Cheng, S. C., Hwang, W. Y., Wu, S. Y., Shadiev, R., & Xie, C. H. (2010). A mobile device and online system with contextual familiarity and its effects on English learning on campus. *Educational Technology & Society, 13*(3), 93–109.

de Jong, T., Specht, M., & Koper, R. (2010). A study of contextualised mobile information delivery for language learning. *Educational Technology & Society, 13*(3), 110–125.

Edge, D., Searle, E., Chiu, K., Zhao, J., & Landay, J. A. (2011). MicroMandarin: Mobile language learning in context. In *Proc. Of the CHI 2011 Conference on Human Factors in Information Systems* (pp. 3169–3178).

Educational Testing Service. *Proficiency scale*. Retrieved May 21, 2012, from http://www.toeic.or.jp/toeic/pdf/data/proficiency.pdf

Gass, S. (2003). Input and interaction. In C. Doughty & M. Long (Eds.), *The handbook of second language acquisition* (pp. 224–255). Oxford: Blackwell Publishers.

Hojo, R. (1996). A study of students' anxiety over classroom English (6). *Bull. Joetsu Univ. Educ., 15*(2), 495–506.

Huang, Y. M., Huang, Y. M., Huang, S. H., & Lin, Y. T. (2011). A ubiquitous English vocabulary learning system: Evidence of active/passive attitudes vs. usefulness/ease-of-use. *Computers & Education, 58*, 273–282.

Kukulska-Hulme, A., & Shield, L. (2007). An overview of mobile assisted language learning: Can mobile devices support collaborative practice in speaking and listening? Retrieved April 6, 2013, from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.84.1398&rep=rep1&type=pdf

Lee, H., & Yook, D. (2012). Feature adaptation for robust mobile speech recognition. *IEEE Transactions on Consumer Electronics, 58*(4), 1393–1398.

Liu, T. Y. (2009). A context-aware ubiquitous learning environment for language listening and speaking. *Journal of Computer Assisted Learning, 25*, 515–527.

Liu, T. Y., & Chu, T. Y. (2010). Using ubiquitous games in an English listening and speaking course: Impact on learning outcomes and motivation. *Computers & Education, 55*, 630–643.

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. *Modern Language Journal, 82*, 545–562.

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.

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

Ogata, H., Li, M., Hou, B., Uosaki, N., Uosaki, N., El-Bishouty, M., & Yano, Y. (2011). SCROLL: Supporting to share and reuse ubiquitous learning log in the context of language learning. *Research and Practice in Technology Enhanced Learning, 6*(2), 69–82.

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*, *4*(1), 33–60.

Pino, C. (2009). Overcoming Japanese EFL learners' fear of speaking. *University of Reading Language Studies Working Papers, 1*, 55–63.

Prunty, A. M., Klopf, D. W., & Ishii, S. (1990). Argumentativeness: Japanese and American tendencies to approach and avoid conflict. *Communication Research Reports, 7*(1), 75–79.

Sharples, M. (2000). The design of personal mobile technologies for lifelong learning. *Computers & Education, 34*(3-4), 177–193.

Sharples, M., Taylor, J., & Vavoula, G. (2005). Towards a theory of mobile learning. *Proceedings of mLearn 2005.*

SpeakGlobal. Retrieved May 21, 2012, from http://speakglobal.co.jp/

Swain, M. (2005). The output hypothesis: Theory and research. In E. Hinkel (Ed.), *Handbook of research in second language teaching and learning* (pp. 471–483). New York: Routledge.

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

Uden, L. (2007). Activity theory for desining mobile learning. *International Journal of Mobile Learning and Organisation, 1*(1), 81–102.

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

Yashima, T. (2002). Willingness to communicate in a second language: The Japanese EFL context. *The Modern Language Journal, 86*(1), 54–66.