Web-based Speech Recognition-Enhanced English Conversation for College Learners

Tsuo-Lin Chiu (邱作麟)  Yu-Li Yeh (葉由俐)  Hsien-Chin Liou (劉顯親)
Tsing Hua University
g925256@oz.nthu.edu.tw  ylyeh@mx.nthu.edu.tw  hcliu@mx.nthu.edu.tw

With growing emphases on communicative competence, the teaching of sociocultural knowledge and speech acts are essential for learners to speak appropriately according to various contexts. It is commonly believed that sociocultural knowledge and speech acts are well developed by learning in native-speaking contexts and having face-to-face conversations with native speakers. Unfortunately, for most EFL (English as a Foreign Language) learners, these conditions are rarely easy to attain. In recent years, one promising field in Computer Assisted Language Learning (CALL) is the application of Automatic Speech Recognition (ASR) technology for assisting learners to engage in meaningful speech interactions. A web-based conversation environment, CandleTalk, has been developed to help EFL learners receive sociocultural strategies training and thus develop communicative competence. CandleTalk is equipped with a speech recognition engine that judges whether learners provide appropriate input. It employed 6 six speech acts as the focus of sociocultural strategies training. Local information is used as the content of dialogues. An effectiveness study was conducted to two groups of college freshmen students, 29 English majors and 20 non-English majors, to see their perception toward CandleTalk and their learning outcomes after a period of practicing. Two sets of tests using the format of Discourse Completion Test (DCT) and an evaluation questionnaire were used for collecting data. The results of the study will implicate the efficacy of applying ASR technology for the learning of conversations and the teaching of speech acts via the training of sociocultural strategies.

INTRODUCTION

A web-based English conversation environment CandleTalk was developed in National Tsing Hua University under a project CANDLE (Corpora And NLP for Digital Learning of English, Liou, et al., 2003, http://candle.cs.nthu.edu.tw) for teaching appropriate use of speech acts as focus of sociocultural strategies training (Savignon & Sysoyev, 2002). CandleTalk adopted advanced Automatic Speech Recognition (ASR) technology for learners to participate in seemingly real life English conversation. A study was conducted to see how EFL college freshmen react to this ASR-based English conversation environment and its effectiveness on the teaching of speech acts as the focus of sociocultural strategies.

Savignon and Sysoyev (2002) proposed sociocultural strategy training to develop EFL learners’ sociocultural aspects of communicative competence. They claimed that English learners need to develop this competence for engaging in conversations with people from unfamiliar sociocultural backgrounds. There were two main principles in their definitions: having a dialogue with people from other cultures and acting as representatives of learners’ own culture. As for the implementation, Savignon and Sysoyev claimed that there are no fixed ways of developing it although they did provide us with examples. In CandleTalk, the dialogues we designed were following the two main principles mentioned and trying to create a context for EFL learners in Taiwan to have an opportunity to receive sociocultural strategy trainings.

Sociocultural competence is also important in EFL learners’ performance of speech acts. Cohen (1994) proposed two abilities for successful speech acts performance which were
sociolinguistic ability and sociocultural ability. The first one was concerning the speaker’s skills at selecting proper linguistic forms for speech act performances in the given language to realize the aim of communication. The second one was concerning the speaker’s skills at selecting proper speech acts based on various sociocultural factors, such as the culture involved, the age and sex of the speaker, their social class, or their roles and status on the interaction. Therefore, we could say that for performance of a contextually appropriate speech act certain amount of sociocultural competence is needed. That is the reason we chose teaching speech acts as the focus of sociocultural strategies training in **CandleTalk**.

In recent years Computer Assisted Language Learning (CALL) has become a tempting alternative for language education. One of the promising fields in CALL is the application of Automatic Speech Recognition (ASR) technology for language learners in the areas of speaking, reading, and pronunciation. Researchers in the field have indicated the advantages of applying ASR for being a private tutor during speaking interaction (Bernstein & Najmi & Ehsani, 1999; Egan, 1999, Ehsani & Knodt, 1998; Eskenazi, 1999; LaRocca & Morgan & Bellinger, 1999; Wachowicz & Scott, 1999). ASR-based speaking programs often allow learners to engage in simulated face-to-face conversation with virtual characters. In addition, programs equipped with ASR technology allow learners to practice speaking without embarrassment. It is also suggested that learners should feel at ease while learning to speak (Eskenazi, 1999). Thus, for EFL learners in Taiwan, programs with ASR technology can assist those who are shy or rarely have chances to speak with English speakers. Still, ASR technology has not been advanced enough for perfect recognition. Programs using ASR technology should have ways to cope with recognition errors to avoid giving learners inappropriate feedback (Eskenazi, 1999)

A number of programs and studies using ASR were reported in the literature (Bernstein & Najmi & Ehsani, 1999; Eskenazi, 1999; Harless, Zier, & Duncan, 1999; Hsia & Wang & Chuang, 2004; Rypa & Price, 1999; Tsai, 2003). Reviews of the literature showed that most studies were either pilot testing on a small number of subjects or relying on learners’ self-report as measures of the effectiveness of environments. One study (Harless, et al., 1999) was done in military school in which the context is quite different from the context for most EFL learners. Tsai (2003) conducted a pilot testing on 9 junior college students using commercial software MY ET (MY ENGLISH TUTOR, [http://www.myet.com/](http://www.myet.com/)). The data was based on students’ representative scores from the system and the questionnaires. There seems to lack empirical data from a larger scale of testing that examines the effectiveness of ASR on EFL learners’ language learning. Therefore, the current study examined not only learners’ perception toward **CandleTalk** but also its effectiveness on learners’ learning of speech acts as focus of sociocultural strategies. In this paper, preliminary results from learners’ questionnaires after a period of practicing on **CandleTalk** are reported.

**Design and Development of CandleTalk**

**CandleTalk** is a conversation environment supported with continuous ASR systems. A continuous ASR system is suitable for processing longer phrases and for designing sentence-level language activities (Wachowicz & Scott, 1999). The designing features of **CandleTalk** include the employment of advanced ASR technology, intelligent responses for learners’ input, a game-like format, and the integration of cultural information from *Sinorama* ([http://www.sinorama.com.tw](http://www.sinorama.com.tw)) as content of the dialogues. *Sinorama* is a high-quality bilingual magazine in Taiwan, and the content covers a large number of cultural issues about Taiwan and Chinese society.

There are four units in **CandleTalk**, with one or two speech acts incorporated as the focus of sociocultural strategies training in each unit. Six speech acts were included: greeting, parting, requesting, complaining, apologizing, and complimenting. With the help of ASR technology and the design of dialogues, **CandleTalk** aims to help learners recognize the appropriate use of
sociocultural strategies in different situations.

In the situational dialogues of each unit, there are two characters: a Taiwanese and an English native speaker. The learner can play one of the roles in the dialogue. Usually the learner plays the role of a local student in Taiwan, and the computer plays the role of a speaker from an English-speaking country. Since speech recognition for free conversation is still impossible, we designed the dialogues in *CandleTalk* with a guided mode (see figure 1 for an example of the flow chart). There is a pre-patterned route behind each unit. In each unit there are two speakers. The learner plays the role of speaker A, and computer, speaker B. For each turn of the conversation, the learner may be given several choices to initiate or respond to the conversation. Different selections lead the learner to a different route of the dialogue. In addition, a game-like format was designed in *CandleTalk*—in certain point of the dialogue if the learner chooses an inappropriate response in the dialogue, the program stops the conversation, and the learner has to restart the unit. With such design of the dialogues, it may give the learner a simulated real-life conversation.

![Figure 1. A Flow Chart in CandleTalk](image)

Upon entering each unit, the first page is the introduction of the unit. It describes and lists the objectives of the unit and also presents the context of the dialogue. After reading the introduction, the bottom of the page is for learners to choose whether they wish to act as a male or female speaker, and whether they want to talk to a male or female speaker. If they choose to talk to a female speaker, they will hear a female voice, and vice versa. The voices in *CandleTalk* are pre-recorded by two English native speakers with North American accents. Upon entering a unit, they see two characters (see figure 2). The one on the right side represents the learner, and the one on the left side represents the computer. While practicing a unit, for each turn of conversations, learners are offered with one to several choices. After learners select one of the choices, they have to finish the recording procedure. They have to choose the appropriate time for recording, click on the record button, and begin to speak. Having a unit completed, learners see a review page which demonstrates the important expressions concerning the appropriate use of speech acts in the unit. Furthermore, *CandleTalk* is also equipped with a tracking system to keep a record of individual student’s learning process for later self-reflection and evaluation (see figure 3).
The content of dialogues in *CandleTalk* was designed following the principles by Savignon and Sysoyeve (2002) in which learners have to stand for their own cultures and discuss their cultural information in English. Thus, *CandleTalk* adopted local cultural information from the *Sinorama* magazines as the content of the dialogue in each unit to demonstrate how to use English to describe learners’ own culture. In addition, the contexts selected for each unit are relevant to Taiwanese college freshmen’s life experiences, such as living in a school dormitory or going to a night market. There is one topic of conversation for each unit, which are a famous singer (A-Mei), a local resort, online games, and the traditional night market. Those topics selected are supposed to be familiar to college learners and should be quite interesting to them.
THE EFFECTIVENESS STUDY

Participants

A total of 49 college freshmen students, 29 English majors and 20 non-English majors in National Tsing Hua University, participated in the study. Their average year of learning English was 6.8.

Instruments

Two types of instruments were used in the study: tests and questionnaires. Two sets of tests (a pretest and a posttest) were used to measure students’ learning of speech acts as the focus of sociocultural strategies shown in their oral responses in the format of the Discourse Completion Test (DCT, Blum-Kulka, 1982). There was one modification of the DCT; that is, instead of filling in responses in written form, students responded to the situations by speaking. There were ten different situations in both pre- and posttest with different orders. These ten situations were categorized based on the six speech acts: one situation for greeting and parting, and two situations for requesting, apologizing, complaining, and complimenting. Students’ oral responses were recorded through the microphone and recording program in the computer. After the test, the sound files were collected for analysis (See appendix for the test).

Two types of questionnaires, a background questionnaire and an evaluation questionnaire, were administered in the study. The Background questionnaire with 6 items was used mainly concerning students’ English learning experiences and exposures to computers. The Evaluation questionnaire with 21 items was concerning the designing features of the website, effectiveness of each unit, and overall effectiveness of CandleTalk. Specific questions concerning each unit were incorporated into the review page for students to fill in immediately after finishing the unit online.

Procedures

The study was conducted in spring semester, 2005. The two groups completed the background questionnaire, and the pretest recording in the first class. Then they had one hour to practice unit 1 and unit 2 in class, followed by two weeks to practice the two units at home. Then, again they had one hour to practice unit 3 and unit 4 in class and two weeks to practice at home. For home practice, they were required to practice each unit for at least three times. After that, they completed the posttest recording and the evaluation questionnaire.

Data Analysis

Two types of data were collected for the study: questionnaire responses and oral responses from the DCT. The results of questionnaires were coded according to different categorizes, such as the effectiveness of the environment, the interface design, etc., to examine students’ perception toward the designing features of the environment. In addition, the results from the pretest and posttest were analyzed and compared to see whether students improve their use of speech acts. The students’ responses on the recorded sound files of the Discourse Completion Test will be rated based on two criteria: (1) the comprehensibility of the speech and (2) the appropriate use of speech acts. For the first category, a 3-point scoring scale (0=unintelligible, 1=problematic but intelligible, 2=intelligible) was used by two raters: the researcher and a co-researcher (a graduate student in an MA-TEFL program). For the second category, a 4-point scoring scale (see table 1) was used by the same raters. Ten students’ results of pre- and posttest were first rated by an English native speaker, who is an experienced researcher in TESOL, to locate the range finders for rating on the two scales. Disagreements between the two raters were solved by face-to-face discussion. The final score for each
student’s score in either the pretest or the posttest will come from an average of points given by
the two different raters with inter-rater reliability calculated. The results of the DCT responses
are not in time for reporting at this writing. The statistic procedure, Paired T-Test, will be used
to compared their ratings responded in the pretest and the posttest periods.

Table 1. Scoring Scale for Use of Speech Act

<table>
<thead>
<tr>
<th>Score</th>
<th>Descriptions</th>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>No Opinion</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>No answer or totally misuse of a speech act</td>
<td>2 (4%)</td>
<td>18 (37%)</td>
<td>16 (33%)</td>
<td>11 (22%)</td>
<td>2 (4%)</td>
</tr>
<tr>
<td>1</td>
<td>Correct use of speech act based on the question, but incomprehensive, inappropriate, or incomplete</td>
<td>5 (10%)</td>
<td>23 (47%)</td>
<td>14 (29%)</td>
<td>4 (8%)</td>
<td>3 (6%)</td>
</tr>
<tr>
<td>2</td>
<td>Correct use of speech act, appropriate, but containing some non-native features.</td>
<td>Yes</td>
<td>No</td>
<td>20 (40%)</td>
<td>29 (60%)</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Correct use of speech act. Comprehensive and appropriate</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

RESULTS OF THE QUESTIONNAIRES

Preliminary results from the two questionnaires are presented and discussed in three
sections: background information, items concerning 4 units, and the overall effectiveness of the
online environment. As can be seen from Table 2, when asked “I am afraid of speaking
English,” 26% of students agreed (both strongly agree and agree), 33% had no opinion, and
41% disagreed (both strongly disagree and disagree).

Table 2 Background Information (total=49, percentage rounded off)

When asked “I am used to learn English with computer,” only 14% of students agreed, 29% had no opinion, and 57% disagreed. It shows that over half of the students are not quite accustomed to learning English with computer. Last, 60% of students reported they heard of the English learning software MT ET, and 40% did not. The reason for asking this question is to understand whether they had an idea of the application of ASR on English learning since MY ET is an ASR-based commercial software in Taiwan. The result shows that over half of them did not know it.

Table 3. Items Concerning 4 Units

<table>
<thead>
<tr>
<th>Rank</th>
<th>Items</th>
<th>Mean in rank order</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The introduction page tells me the focus of the unit</td>
<td>4.07</td>
</tr>
<tr>
<td>2</td>
<td>The introduction page tells me the context of the dialogue</td>
<td>4.01</td>
</tr>
<tr>
<td>3</td>
<td>I learned how to perform appropriate speech acts from the unit</td>
<td>3.98</td>
</tr>
<tr>
<td>4</td>
<td>The review page tells me how to use the dialogues taught in the unit</td>
<td>3.94</td>
</tr>
<tr>
<td>5</td>
<td>The difficulty level of the dialogue in the unit is appropriate</td>
<td>3.87</td>
</tr>
<tr>
<td>6</td>
<td>The topic of the dialogue in the unit is interesting</td>
<td>3.65</td>
</tr>
<tr>
<td>7</td>
<td>The recognition rate in the unit is appropriate</td>
<td>3.02</td>
</tr>
</tbody>
</table>
Table 3 presents the results of specific questions in the evaluation questionnaire concerning the design of the four units. The results have been ranked according to the mean of each item. The mean was calculated according to the scores of each selection (strongly agree=5, agree=4, no opinion=3, disagree=2, and strongly disagree=1). As can be seen from the results, the number one (4.07 out of the total 5.00) and two (4.01) ranks of the items are “The introduction page tells me the context of the dialogue,” and “The introduction page tells me the focus of this unit.” It seems that students generally agreed the introduction page has provided them with a clear picture before they entered the dialogue. Rank number three (3.98) is “I learned how to perform appropriate speech acts from the unit.” It suggests that students generally agreed that they learned how to use the speech act that we intend to teach them in each unit. Rank number four to six are items concerning the review page, the difficulty and the topics of dialogues. The lowest rank is the item “The recognition rate is appropriate in each unit.” It may reveal that the most difficult part for students in each unit is to pass the recognizer. They might have to speak a sentence for several times before move on to the next one. As mentioned in previous research (Hsia, et al., 2004), sometimes it is difficult for the recognizer to understand longer sentences, or the equipment used and noises in the surroundings might affect the voice quality and recognition rate. Thus, the reasons students were blocked by the recognizer might be their poor performance in speaking or occasional recognition errors.

Table 3. Features of the Site

<table>
<thead>
<tr>
<th>Rank</th>
<th>Items</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The responses from the computer are clear</td>
<td>4.16</td>
</tr>
<tr>
<td>2</td>
<td>The directions of how to operate the environment are clear</td>
<td>4.10</td>
</tr>
<tr>
<td>3</td>
<td>The word color and font size are clear and easy to read</td>
<td>3.96</td>
</tr>
<tr>
<td>4</td>
<td>The interface design in the environment is acceptable</td>
<td>3.86</td>
</tr>
<tr>
<td>5</td>
<td>The color and pictures in the environment are appropriate</td>
<td>3.84</td>
</tr>
<tr>
<td>6</td>
<td>The recording operation in the environment is easy. It is easy for me to move around in the environment</td>
<td>3.78</td>
</tr>
</tbody>
</table>

Table 4 and Table 5 show the results of evaluation questionnaire concerning the features of the website and the overall effectiveness of CandleTalk. The means for all the items in the evaluation questionnaire are all above 3, which may suggest that students overall agreed with the items provided. Results were analyzed from two aspects: (1) features of the site and (2) overall effectiveness of the environment.

First of all, we look at items concerning the features of the site (see Table 4). The highest rank is concerning the responses from the computer. It seems that students had no difficulties understanding the speech recorded. In addition, the lowest rank in this category is concerning the recording procedure and the operation of the environment. It may suggest that the flexibility of the environment is not good enough and it might need more explanations for the recording procedure.

Second, we look at items concerning overall effectiveness of the units (see Table 5). The highest and the second highest ranks in this category (4.16, 4.10) are the items “The environment helps me learn how to speak according to different contexts,” and “The recognizer lets me pay attention to my English conversation.” These two items are the main purposes of CandleTalk. It could possibly suggest that students did perceive and agree with the designer’s intentions, but there is still room for improvement of their effectiveness. In addition, Rank
number 3 is the item “The environment fits my expectation of learning English” with a score of 3.59. It seems that although over half of the students reported that they are not used to learn English with computers (see table 2), most of them still agreed that this ASR-based environment fits their expectations. It could suggest that ASR technology still has much potential for language education. Last, the lowest rank in this category is the item “The dialogues in each unit are fluent.” It could also be the results of lower scores in the recognition rate mentioned before. Since some students were constantly blocked by the recognizer, they might consider that the fluency of the dialogues is not good enough.

Table 5. Overall Effectiveness

<table>
<thead>
<tr>
<th>Rank</th>
<th>Items</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>The environment helps me learn how to speak according to contexts</td>
<td>3.80</td>
</tr>
<tr>
<td>2</td>
<td>The recognizer let me pay attention to my English conversation</td>
<td>3.73</td>
</tr>
<tr>
<td>3</td>
<td>The environment fits my expectation of learning English</td>
<td>3.59</td>
</tr>
<tr>
<td>4</td>
<td>The multiple choices in each dialogue make me feel like talking to a real person</td>
<td>3.55</td>
</tr>
<tr>
<td>5</td>
<td>By interacting with computer, it is helpful for me to learn English dialogues</td>
<td>3.51</td>
</tr>
<tr>
<td>6</td>
<td>The dialogues in each unit are fluent</td>
<td>3.37</td>
</tr>
</tbody>
</table>

CONCLUDING REMARKS

In this study, the results from the two questionnaires show that college freshmen in general have positive attitude toward CandleTalk since the means for all the items are above 3. But the results show that there is still room for improvement in some specific aspects. In a conversation environment like CandleTalk it is better to make recording procedures less complicated so that the interaction with computer could be more human-like. Furthermore, the ranks of the items reveal that the recognition rate could be a major problem when facing longer sentences in dialogues. Since the main purpose is for students to communicate, the emphasis should not be put too much on the accuracy of pronunciation or it might lower students’ motivation and the fluency of dialogues. While there are various factors affecting the accuracy of recognizer, the results could possibly implicate that teachers who want to employ ASR technology into their teaching should always take good care of recognition errors. Last, the questionnaire results reveal that students perceived well the intentions of teaching appropriate use of speech acts. It will be better if we could compare their results of pretest and posttest, which is still on progress, to further consolidate the effectiveness of the use of ASR technology on language education.

Acknowledgements: We would like to thank the financial support provided from the National Science Council (NSC 94-2524-S007-001) under the National Science and Technology Program for E-Learning. Programmers, involved research assistants, and participating students are also acknowledged.

REFERENCES


**APPENDIX**

對話完成測驗

1. You happen to meet one of your high school classmates on the street. You two haven’t seen each other for years. You two talk for a while, and you are going to leave. You want to end this conversation. What would you say?

2. You are a college freshman. One day while you are on a school bus, you see one of your classmates standing next to you, and he/she also notices you. You two are not familiar with each other, but it is very impolite to keep silent in that situation. You think it is a good opportunity to make friends with him/her so you are going to break the ice. What would you say?

3. You are going to go abroad for a week. You are worried that nobody can feed your cat while you are abroad. You are going to ask your friend to take care of your cat. What would you say?

4. You are doing your homework on the computer, but suddenly the computer crashes. You
don’t know how to fix the computer. Luckily, your friend is around and he/she is very good at the computer stuff. If you asked him/her for help, what would you say?

5. You completely forget a crucial meeting with your friend. An hour later you call him/her to apologize. The problem is that this is the second time you’ve forgotten such a meeting. What would you say to your friend?

6. You and your friend go out for a drink. You accidentally spill some coffee on your friend’s favorite jacket, which makes him/her very angry. You try to apologize. What would you say?

7. It is not the first time that loud music is heard from your roommate’s computer quite late at night. You can’t fall asleep in such noise. What would you say to him/her?

8. A friend who takes the same course as you at the university refuses to share some important material for the next test. In the past, you helped him/her many times. You see him/her on campus. What would you say to him/her?

9. Your friend has just cooked a wonderful dinner for you. As his/her best friend, what would you say to him/her?

10. Your friend just bought a new cell phone, which you think it looks great. What would you say to your friend?