The further development of CSIEC project driven by application and evaluation in English education

Jiyou Jia and Weichao Chen

Address for correspondence: Dr. Jiyou Jia, associate professor, Department of Educational Technology, School of Education, Beijing University. Beijing, 100871, China. Email: jjy@pku.edu.cn

Abstract

In this paper, we present the comprehensive version of CSIEC (Computer Simulation in Educational Communication), an interactive web-based human–computer dialogue system with natural language for English instruction, and its tentative application and evaluation in English education. First, we briefly introduce the motivation for this project, survey the related works and illustrate the system structure with flow diagram. Then we describe its pedagogical functions, especially free chatting and chatting on a given topic. We summarise the free Internet usage within 6 months and introduce its integration into English classrooms, as well as the formal evaluation results of the integration. The evaluation findings show that the chatting function has been improved and fully used by the users, and the application of the CSIEC system in English instruction can motivate the learners to use English and enhance their learning process. Lastly, we discuss the application-driven approach of system development and draw some conclusions for future improvement.

Motivation, underlying theories and related works

Discrepancy between the great demand of English instruction and the actual poor situation

With the rapid economic, cultural and educational internationalisation of China, the importance of cross-cultural communication is gaining more and more attention. English, as an international language, is treated as a key tool for the development and cultivation of such a communicational ability. Therefore, English language is now listed as one of the three core courses in China’s elementary and secondary education and as a compulsory course in higher education. English learning is being emphasised more and more in this country.

Unfortunately, some problems exist in the English language education in China. First of all, it is widely recognised that one of the best ways to learn a foreign language is through spoken dialogue with a native speaker, and frequent communication in English plays an important role in English learning. However, that is not a practical method in
the classroom, especially in China and other countries with English as a foreign language, because of the one-to-one student–teacher ratio it implies. A number of factors ranging from the lack of time to shyness or limited opportunity for quality feedback hamper using the target language (Fryer & Carpenter, 2006). The language environment and few qualified English teachers in China cannot supply enough chance of authentic talking. So it is quite common that schoolteachers complain of working burdens and do not have enough time to communicate in English with students (Liang, 2005). Second, although learning English through communication and application has been emphasised recently, English education in China is traditionally examination oriented. In addition to those challenges, grammar teaching and learning is important in China’s English education because Chinese differs greatly from English in grammar (Liu, 2005). Without basic grammar knowledge, the students cannot make great progress, as they mostly only practise English in school time and cannot learn it unconsciously and spontaneously from the social environment.

In order to solve these problems and to meet the demand of English learning, new educational technologies have been applied in English instruction, eg, multimedia computer programs, web-based learning and CMC (Computer-Mediated Communication), etc. However, most of the current computer programs cannot realise an essential instructional method yet, ie, communication and interaction with the human user in natural language, which plays a significant role in language learning, as many language theories and practice have argued. In CMC, it is difficult for the normal users to find a long-time talking partner or a pen pal. A potential solution to this problem is to apply computer-spoken dialogue systems to role-play a conversational partner. If we could design an interactive web-based system that could chat with the English learners anytime anywhere, their great demand for learning partners could be fulfilled. Such a system should aim at helping the learners improve their skills of using English through frequent chatting with them in English.

Motivated by the great demand for English instruction, we, in 2001, began to design a system that can communicate with the Internet users in English. Our design principle is application and evaluation oriented. So long as the system is applicable, we put it into free use in the Internet and get user feedback. Moreover, we apply and evaluate the system in English classes in middle school and in universities. Cooperating with the English teachers, we try to integrate the system with English instruction more closely. Through systematic application and evaluation, we get more suggestions and critics, which can direct our research more effectively.

Underlying theories
Situated learning and constructivist learning theories are the underpinning theoretical backgrounds of the idea to create a virtual chatting partner for language learners. The theoretical framework of situated learning (Barab & Duffy, 2000; Brown, Collins & Duguid, 1989) concerns about the learning’s reliance on the cognitive conditions, emphasises creating suitable situation to motivate the learning activities and encourages the learners to practise in the related social activities. When applied in language
learning, the situated learning theory stresses the language acquaintance through communication in the proper situation (Asta, 2007; Mondada & Doehler, 2005). The constructivist learning theory (Jonassen, 1994; Von Glasersfeld, 1996; Tenenbaum, Naidu, Jegede & Austin, 2001) underlines learning by doing and learning in situated practice. By language learning, the immersion in the target natural language is central to the language construction (Kaufman, 2004; Wang, 2004; Zhang, 2005).

**Related works**

A chatbot is a program that can pretend to ‘chat’ with a human user in a certain natural language. Brennan (2006) defined a chatbot as ‘an artificial construct that is designed to converse with human beings using natural language as input and output’. There are several alternative names for chatbot-like systems: human–computer dialogue system, machine conversation system, virtual agent, dialogue system, chatterbot, etc.

ELIZA (Weizenbaum, 1966), created in the 1960s, is the first chatbot. It used keywords to analyse input sentence and created its response based on reassembly rules associated with a decomposition of the input. But it held no memory of the conversation and so could not achieve any form of targeted collaboration or negotiation. However, the syntactical way of NLP (Natural Language Processing) exemplified by ELIZA has been developed significantly from the 1960s up to now, leading to the development of various chatbots. With the improvement of NLP, machine-learning techniques and decision-making capabilities, chatbots have become more practical and have also been applied in education.

Knill, Carlsson, Chi and Lezama (2004) used a chatbot for maths education. They concluded that allowing students to interact with a calculus chatbot added to the variety of instructional tools; however, technology such as computer algebra systems, multimedia presentations or chatbots could only serve as amplifiers but cannot replace a good guide.

Fryer and Carpenter (2006) suggested that the chatbot could be used to practise a language. Two hundred eleven students with English as native language used two well-known bots in class. Most of them enjoyed using the chatbots and also generally felt more comfortable conversing with the bots than a student partner or teacher. They believed that now it is possible for language teachers to bring this technology into the foreign language-learning classroom as a permanent tool for language practice. However, they did not present empirical research to prove their conclusion.

Seneff (2006) described several multilingual dialogue systems specifically designed to address the need for language learning and teaching. Students can engage in dialogue with the computer either over the telephone or through audio/typed input at a web page, i.e, speak with the chatbot. Several different domains were developed, in which a student’s conversational interaction was assisted by a software agent functioning as a
tutor, providing the student with translation assistance at any time. Nevertheless, no statistical data demonstrated the usefulness of the supposed dialogue systems.

Abu Shawar and Atwell (2007) developed algorithms for adapting or retraining a chatbot with a corpus to chat in the language and topic of the training corpus. They stated that the evaluation feedback from language learners and teachers indicated that these adaptive chatbots offered a useful autonomous alternative to traditional classroom-based conversation practice. Unfortunately, no concrete evaluation data was presented in their paper.

Kerly, Hall and Bull (2007) discussed the development and capabilities of both conversational agents and intelligent tutoring systems, in particular, open learner modelling. They described a Wizard-of-Oz experiment to investigate the feasibility of using a chatbot to support negotiation. The experiment result showed that most students liked the chatbot and that the chatbot helped them understand their learner model. They concluded that a fusion of the two fields could lead to developing negotiation techniques for chatbots and the enhancement of the open learner model, and this technology, if successful, could have widespread application in schools, universities and other training scenarios.

From the related works surveyed earlier, we can conclude that the usage of chatbot systems in education is drawing more and more attention from researchers in related fields. This trend confirms our determination to further the development of the CSIEC system and its application in English education.

**Current system structure: symbolic NLP**

We attempt the full syntactical and semantic analysis of the user inputs. The current CSIEC system at version 9 consists mainly of the following components, which are illustrated in Figure 1.

Http request parser resolves the user request from the http connection and gets some parameter values: input text, scenario topic, agent character, speech speed, spelling and grammar checker, etc.

English parser parses the user input text into NLML (Natural Language Markup Language). In fact, NLML is a phrase and dependency tree in XML form and structurally labels the grammar elements (phrases), their relations and other linguistic information in English sentences. NLML parser parses the NLML of the user input into NLOMJ (Natural Language Object Model in Java), which represents the grammatical elements and their dependence relations with the sentence ontology in the working memory (Jia, Ye & Mainzer, 2004). Through NLOMJ, the declarative sentence is retrieved and decomposed into atomic facts, which consist of only one subject and one verb phrase.
NLDB (Natural Language Database) stores the historical discourse (input and output), the user atomic facts in the form NLML, the robot atomic facts, which are also expressed in NLML, and other data.

World model contains commonsense knowledge, which is the basis for response generation and logical inference. It is now represented by WordNet (Fellbaum, 1998), a very large and popularly used lexical ontology.

CR (Communicational Response) mechanism comprehensively takes into account the user facts stored in NLDB, the world model, and the personality of the user expressed in the previous dialogue and that of the robot itself selected by the user.

Scenario dialogue handler creates the robot output corresponding to the user input within a given scenario. Scenario show handler creates the random robot–robot talk show scripts within a given scenario. Scenario DB stores the robot–robot talk show scripts and human–robot dialogue scripts, which are manually written by a designer, for example, an English language teacher.
Microsoft agent script formatting transforms the output text from CR, scenario show handler or scenario dialogue handler into VB scripts, considering the selected agent character and speaking speed. Browser/server interface processes the http request from the client machine and responds with the robot output, either in text or with VB script.

**Basic functions of the system**

*Multimodal user interface and selectable chatting pattern*

Human–computer dialogue with natural language is the most specific function of the CSIEC system. However, like human being’s authentic dialogue situation, the Internet users have various preferences for the dialogue simulation. In order to adapt to variant user preferences, the CSIEC provides several user interfaces and dialogue patterns.

First, the user can chat with the robot either through text or via voice. The users can hear synthesised voice and watch the avatar performance through Microsoft agent technology. They can also speak to the robot through a microphone with the help of a third-party program such as the ViaVoice from IBM Corp, Burlington, MA, USA.

Second, the robot can check the spelling and grammar of the user’s input upon the user’s request.

Third, the chat topic between the user and the robot can be either free (unlimited) or given (limited). The unlimited dialogue simulation does not specify the dialogue topic and content. It benefits users whose English is fluent or who are at least good at written English, such as college and university students and graduates, as well as users who are extroversion or conversational. However, users whose English (at least the written English) is poor, or whose characteristic is introversion, have little to chat with the virtual chatting partner. Therefore, their dialogues with the agent are mostly very short. For these users, an instructive dialogue in a specific scenario guided by the agent is more helpful. Additionally, language teachers acknowledge that conversation practice is normally on a specific topic during the learning of topic-specific vocabulary and language. So the limited dialogue is appropriate for the junior learners and school students and often happens in the classroom.

However, in normal human talking, these two chatting patterns are not absolutely separated but often interleave each other. In our system design, we consider this interaction too. In the following two subsections, we introduce these two chatting patterns, as well as their relationship, in more detail.

*Free chatting adaptive to user preference and topic*

In free chatting, the users with different characters and personalities may choose different types of chatting pattern. For example, some users may prefer to chat with someone who will listen to them quietly most of the time, while some others may hope the chatting partner can tell them stories, jokes or news. For the sake of user dialogue personalisation, we have designed five Microsoft agent characters that represent different kinds of chatting patterns (Jia, Chen & Hou, 2006). Christine always tells the user...
stories, jokes and world news. Stephan prefers to listen quietly when the users share with him their own experiences. Emina is a curious girl and is fond of asking users all kinds of questions related with the users’ input. Christopher provides comments, suggestions and advice on the user’s input. Ingrid behaves as a comprehensive virtual chatting partner who gives users responses corresponding to both the input text and the discourse context.

Upon user registration to the chatting system, the user’s profile, such as sex, age (birthday), educational level and address (province), is obtained and recorded. So the corresponding chatting topic and content can be generated based on the personal information. Of course, if the user expresses the wish to change the chatting topic during the process of the robot’s narrating comments or asking questions, the robot should terminate this process and transfer to another topic given by the user or by the robot itself. If the user specifies a topic, the robot changes the topic to it. If the user just expresses the wish to change the topic but does not determine a topic, the robot selects one from the waiting topics list that has not been talked about with the given user.

The user’s interests are also expressed in the input texts, eg, the mentioned nouns and verbs in the sentences. Thus, the chatting topic can be triggered by nouns and verbs and their combination. Chatting between the user and the robot can be regarded as guided chatting or chatting in some context. We deal with the chatting on a given topic in two ways. The first one is predefining some comments or asking some questions about this topic. By talking about this topic, only one statement or question will be randomly selected and given out. The second way is to search the topic or related topic in the guided chatting within a given scenario and then transfer the chatting to the guided chatting in a given scenario, which will be introduced more in the next subsection. The arrow from the scenario dialogue handler to the communicational response in Figure 1 indicates this relationship.

In summary, free chatting is aimed at motivating the users into talking. Therefore, the robot tries to adapt itself to the user’s interest and to launch new topics.

Guided chatting in a given scenario for drill or examination

The dialogue on a given topic requires not only the common sense knowledge but also the domain knowledge for this special scenario, in addition to the explicit usage of the knowledge with the form of natural language. The dialogue should be developed step by step around a red line or a topic for this scenario. It can be described by a complex tree structure with many branches. We use a script to describe the decision tree. The script is made up of lines of dialogue steps, each of which is a branch in the decision tree. Suppose the robot speaks first. In every line, there must be a text output from the robot and its order number in the dialogue. The output may be triggered by specific user input, which we call the prerequisite of this output text. The robot may also expect the user to input certain texts, or some texts with specific semantic or syntactical characters, which we call the expectation of this output text. We write the line in the script with the format:
The ‘Nr.’ and (text) are the two necessary components in every line. The ‘Nr.’ is an integer that indicates the line order in the whole script, whereas the ‘text’ can be any text from the robot, either statement or question, and so on, and is written within closed brackets.

In a script line, the prerequisite and expectation are optional. If they appear, they must be written within closed sharp brackets. If the prerequisite exists and is satisfied, the output text can be given out by the robot. The expectation can be applied to instructional goal. For example, if the user’s input does not satisfy the robot’s expectation, he or she will face the previous robot output again, until the expectation is fulfilled. This dialogue pattern can be used for drills. Another alternative is that the user is given a high mark if his or her input satisfies the robot output; otherwise, a low mark will be given although the robot continues the next dialogue. This pattern can be used in tests or examinations.

The format of the prerequisite is:

\(<\text{Nr, variable 1: value 1, value 2 ... ; variable 2: value 1, value 2 ...}\>)

The format of the expectation is:

\(<\text{variable 1: value 1, value 2 ... ; variable 2: value 1, value 2 ...}\>)

Both are almost the same form. Only the prerequisite needs an order number indicating the expectation of which line this condition fulfils. There may be more than one value for a given variable. This means that if the variable equals any one of the listed values, the condition is fulfilled, ie, the values for a given variable have the relation of logical ‘or’. There may also be more than one variable and its corresponding values. The relation among these variables is the relation of logical ‘and’.

This kind of strict script is difficult to write for human authors, for example, the English teachers who use this program to train the students. Thus, we have designed a Java GUI, ie, DSE (Discourse Script Editor) for editing the scripts step by step more easily. With it, a normal user such as an English teacher need not pay attention to the writing format but to the discourse content and process. Starting from 1, the line number increases automatically. The prerequisites can automatically be selected from the previous expectations. Although the script designer can be somewhat relieved by this script editor, he or she has to spend much time on planning this discourse script between the robot and the human user, just like a film director. This work is not just the language teaching but also the teaching of response strategies through natural language.

*Listening training*

We use the Microsoft agent technology to synthesise the output text because the agent’s voice is lifelike, the agent’s figures, movements, as well as actions, can be designed very
vividly and it can also synchronously display the spoken text, which facilitates the aural understanding and activates the user’s interests. The robot’s reading speed can be adjusted by the users at any time. We have also designed a free web page whose agent can read any texts inputted or pasted by the user.

Different from traditional audio technologies such as audio players, the user confronts with unexpected robot text and voices, just like talking with a real human being. So this function can benefit the user’s listening comprehension and prompt response.

**Talk show of two robots**

This function is designed to aid the user’s chatting on a given topic. With this, the users can watch the talk show of two robots before the human–computer interaction. The talking texts are predefined by the teacher for the specific context or topic and can be readily written by the teachers with any text editor. However, the actual texts for a given meaning can be expressed randomly. So this kind of talk show is different from the monotonous one presented in the traditional video or audio cassette. It will enforce the learner’s spontaneous listening and understanding.

**Application and evaluation**

The CSIEC system is application and evaluation oriented. Since its birth, it has been practically applied in English learning, first, freely accessed by the Internet users and then used by the students under the guide of the English teachers. As described in the previous sections about system structure and functions, this system is characterised by the human–computer interaction with natural language. Its validation and effectiveness can be proved by the free usage from the Internet users and checked by the analysis of users’ chatting records. The special system test before the application is actually the laborious work of Turing test (Turing, 1950) and Loebner prize contest, and cannot be implemented with our very limited human resources.

In this section, we summarise the Internet usage in the first part and analyse the data collected from the formal evaluation of the system’s English class integration in the second part. Through the evaluation, we want to check if the design goal of the system has been achieved.

**Summative evaluation of free using in Internet**

The Internet users reach the CSIEC web page mainly through search engines because our website has become one of the top five in the searching results of famous search engines such as google.com, yahoo.com and baidu.com by related keywords such as ‘chatbot’, ‘English chatbot’, ‘Online English learning’ in Chinese or in English, although we have not made any large-scale advertisement. The effectiveness and attractiveness of the system’s adaption to English learning in China have been somewhat demonstrated by this practical achievement.
Now with the human–computer dialogues recorded in the database, we make a summarr evaluation of the system’s chatting function from January 20 to June 20, 2007.

**Users’ amount and demographical information**

The users with different user names who accessed the CSIEC during this period count 1783.

During registration, the users are required to select their gender (male/female), education level (pre-school/elementary school /middle school/undergraduate student/graduate), birthday and address (province names in China and other country names). Based on these gathered user data, we investigate the demographical distribution, although the user may randomly and freely select any value of the options.

Among the users, there are more men (1008) than women (775), which is consistent with the gender distribution of Chinese Internet users (China Internet Network Information Centre, 2007).

More than half of the users are undergraduate students. This can be justified by the fact that English is the obligatory course in higher education and that the university and college students find it easier to access the Internet freely. The second largest user population is middle school students. Except for 45 students who are required to use the system in the formal evaluation, there are still 377 (21.1%) free users. They also have mastered basic English knowledge and can talk with the robot in English.

The users come from all provinces in China and other countries like the USA, Japan, Korea, Germany, Vietnam, and so on. The top 10 provinces with the most users are: Beijing (667, 37.4%); Guangdong (168, 9.4%); Shandong (104, 5.8%); Shanghai (82, 4.6%); Jiangsu (86, 4.8%); Liaoning (73, 4.1%); Henan (54, 3.0%); Zhejiang (45, 2.5%); Hebei (43, 2.4%); Hubei (37, 2.1%); and Hunan (37, 2.1%). They altogether constitute the majority (78.2%) of the users. This distribution corresponds to the geographical distribution of Chinese economic development on the one side, and on the other side reflects that the system server located in Beijing cannot supply high-speed response to the users from other locations.

**The duration of chatting**

The chatting quality can be measured by the chatting duration between the user and the robot. To calculate the chatting duration, we define two terms: round and number of the rounds. A round means a user input and a corresponding robot output to the user. Therefore, the total rounds of a given user covers all dialogues between the user and the chatbot and can be used to describe the duration of the user’s chatting with this chatbot.

Sometimes the user cannot receive any response from the chatbot after a long waiting time. The reasons are still not clear. One may be the multi-users’ simultaneous visit of
the server, another may be some bugs in the system design. After scrutinising the user dialogue records, we find 313 such error rounds. Besides these error rounds, there are 48,840 effective rounds with both input and output. We define the natural error rate as the ratio of the error rounds number divided by the sum of error rounds number and effective rounds number. In this study period, the natural error rate is \( \frac{313}{313 + 48,840} \), ie, 0.64%.

The mean of user chatting rounds is 27.4. The number of the rounds from each user varies from 1 to 580. We divide the number of the rounds (duration of the dialogues) into five classes, as Table 1 shows. From it, we see that about 49% of the users chat with the robot briefly (≤10 rounds); about 46% of them chat with it long or longer; and only few, about 5%, chat with it very long (>100 rounds). Compared with our previous finding in Jia (2004), which is listed in the last column of Table 1, the percentage of the brief chatting with the robot has decreased by 21.78%. Proportionally, the percentage of the long and longer chatting has increased.

The distribution of user chatting patterns

The CSIEC system provides multimodal user interface and selectable chatting patterns. Thus, we investigate the distribution of chatting patterns, as shown in Table 2. We can see that 84.66% of the chatting is held with the free chatting pattern, and only 15.34% uses the chatting in a given context. The reason may be that the free users do not understand what is chatting in a given context very well so that most users of the context chatting are the students in our project English classes. Among the chatting for a given context, the text pattern is used almost as frequently as the agent pattern. It can be verified by our team’s assistant and tutoring about the system usage, especially the installation and usage of Microsoft agent characters, in every unit of the English class.

Among the free chatting, more users select the text version instead of the agent version. One reason may be that the text pattern is simpler and more convenient than the agent version, as the agent version requires the users to install two files, and the unskilled computer users may encounter some setting problems, which is proven by some users’ feedback complaint that they cannot use the agent version. Another reason may be that

<table>
<thead>
<tr>
<th>Duration of the dialogues</th>
<th>Range of the number of rounds</th>
<th>Number of users</th>
<th>Number of users/total user number (%)</th>
<th>Number of users/total user number (Jia, 2004) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Short</td>
<td>(0, 10]</td>
<td>871</td>
<td>48.85</td>
<td>62.34</td>
</tr>
<tr>
<td>Long</td>
<td>(10, 50]</td>
<td>685</td>
<td>38.42</td>
<td>30.10</td>
</tr>
<tr>
<td>Longer</td>
<td>(50, 100]</td>
<td>136</td>
<td>7.63</td>
<td>4.78</td>
</tr>
<tr>
<td>Very long</td>
<td>(100, 580]</td>
<td>91</td>
<td>5.10</td>
<td>2.78</td>
</tr>
<tr>
<td>Total user number</td>
<td></td>
<td>1783</td>
<td>100.00</td>
<td>100.00</td>
</tr>
</tbody>
</table>
the synthesised agent voice does not sound so good, which is proven by some users’ complaining about the voice quality.

Among the free chatting, the chatting without spelling and grammar check (ca. 66%) is much more used than that with check (ca. 18%). This result reflects that most free users treat the system as a chatting partner, so they would like to chat with it more fluently instead of worrying about grammar and spelling errors. Human–computer chatting is the most unique function of the CSIEC system; therefore, the users like to fully use it.

As there are fewer female users than male users, the rounds of all chatting patterns of the female users are less than that of the male users. But by contextual chatting with text, the difference is minor, i.e., 49.1 versus 50.9%. The mean of chatting rounds amount of the female users is 29.3 and more than the average number of rounds of the male users, 25.9. It seems that female users talk with the robot more than male users do.

As the undergraduate students constitute the largest user population, almost in every chatting pattern of the free chatting with text, their chatting rounds are more than the others’. But by free chatting with agent, the rounds amount distinction between middle school and undergraduate school students is not great. Almost all of the context chatting rounds with an agent are held by the middle school students. This phenomenon happens because the middle school students participating in the formal evaluation are required and also given help to use the chatting with agents for given topics. The same reason also leads to another fact, i.e., the average chatting rounds of middle school students (43.2) is almost twice many as that of the undergraduate students (22.3), and also more than that of the other groups.

Table 2: The distribution of user chatting patterns

<table>
<thead>
<tr>
<th>Chatting pattern</th>
<th>Dialogue rounds</th>
<th>Dialogue rounds/total rounds (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Context chat with an agent</td>
<td>3,945</td>
<td>8.08</td>
</tr>
<tr>
<td>Context chat with text</td>
<td>3,550</td>
<td>7.26</td>
</tr>
<tr>
<td>Free chat with an agent</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Without grammar and spelling check</td>
<td>10,101</td>
<td>12,965</td>
</tr>
<tr>
<td>Check and find grammar error</td>
<td>588</td>
<td>1.20</td>
</tr>
<tr>
<td>Check and find spelling error</td>
<td>339</td>
<td>0.69</td>
</tr>
<tr>
<td>Check but find no grammar or spelling error</td>
<td>1,937</td>
<td>3.97</td>
</tr>
<tr>
<td>Free chat with text</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Without grammar and spelling check</td>
<td>22,362</td>
<td>28,380</td>
</tr>
<tr>
<td>Check and find grammar error</td>
<td>1,442</td>
<td>2.95</td>
</tr>
<tr>
<td>Check and find spelling error</td>
<td>471</td>
<td>0.96</td>
</tr>
<tr>
<td>Check but find no grammar or spelling error</td>
<td>4,105</td>
<td>8.40</td>
</tr>
<tr>
<td>Total</td>
<td>48,840</td>
<td>100.00</td>
</tr>
</tbody>
</table>
User feedbacks
In the foot of almost every web page of the CSIEC system, we leave a feedback text area so that the users can straightforwardly enter their comments, critics and suggestions either in Chinese or in English. Through the detailed scrutinising and analysis of these texts both in English and in Chinese, we hope to find what problems are addressed by the users.

Many users still input normal chatting text into this area, such as ‘hello’, ‘what is your name?’, ‘how old are you?’ and other nonsense texts. All of these non-feedback expressions count more than one-half of the obtained ‘feedbacks’. Excluding them, we received 341 lines of real feedback.

Among 79 positive comments, 37 are very simple positive comments such as ‘(very) good’, ‘(very) well’. Six emotionally praised the system without any reason, for example: ‘clever’, ‘I like this’, ‘I love you’, ‘thank you’. Thirty-six express positive comments with reasons, eg, ‘The robot is more advanced than before, and also personalized’, ‘The access speed is faster than before’, ‘The dialogue is fluent. I hope the master to enrich the robot’s language’ and ‘The kind of communication can improve our English’.

Seven feedbacks are very simple negative comments without any reason, such as ‘not good’, ‘just so so’, ‘simple’ and ‘stupid’. Sixteen of them write that the response from the robot in the Internet is (too) slow. Twelve complain that they could neither see the agent animation nor hear the agent’s speech, that the agent character is not good, the agent voice sounds weird, the agent speaks too fast or the agent speaks too slowly. Ten complained that the dialogue in a given scenario is too short. Fifteen complain that the robot only asked the user the same questions, could not answer the user’s question or gave false answers. Twelve complain that the robot talks little or often changes the topic. Thirteen cannot understand what the robot is saying or what the new words mean. They suggest that it is better to be able to read the Chinese translation of these new words. Five report that they found grammatical errors in the robot’s responses.

Moreover, 10 of the 341 feedbacks give both positive and negative comment, like ‘the system is very good, but the robot responds too slowly’.

Twenty-eight users state they are not good at English and want to learn English but do not know how to and hope the system would give some suggestions. Seven ask questions about the system usage, for example, ‘how to select the agent character?’ and ‘how can I review my dialogue with the robot?’

The feedbacks point out either technical problems or content shortcomings, which should be tackled in the further improvement.

Formative evaluation of English class integration
After discussing with the English teachers about the class integration and evaluation of the CSIEC system, we decided the instructional instruments are the talk show by two chatting robots on the one side and the students’ talking in English with one robot for
a given topic on the other side. The main application goal is to facilitate role-playing activities in the classes for English. Before the integration term, we have designed the talk show scripts and human–computer dialogue scripts according to the textbooks, so that the students can use them on time. During the whole term, we formally evaluated it through questionnaires, observations in the classes and surveys with teachers and students focus groups.

Apparently, the English instruction for graduate students and high school students should be much distinct because the graduate students should have a larger vocabulary and more grammatical knowledge than high school students. But in our formal evaluation case, the difference is not obvious. According to the discussion with the teacher and students from the graduate classes, we find that most graduate students have not frequently used English or practised it for at least 2 years. But the experimental high school class consists of students with very solid English basis. Some of them even plan to take the TOEFL (Test of English as a Foreign Language).

Moreover, for the role-playing activities according to the content requirement in the textbook, the vocabulary and grammar rules are limited. Our system can supply the talk show and human–computer dialogue simulation function within a given context, no matter what kind of content it is. Therefore, we apply the system both in graduate student class and in high school student class and collect data from them in order to compare the validation and effectiveness of the system for both cases.

In the first term, 86 graduate students from two English classes taught by the same teacher participated in the study, and the teacher only recommended the students to use the system, but did not require them. According to the course content of 12 teaching units, we designed 25 scenario scripts for the role-play talk show and human–robot chatting. In the second term, 45 high school students in Grade 2 attended the study, where the teacher required the students to use the system together in the computer room. According to the course content of 10 units, we designed 40 scenario scripts for the role-play talk show and human–robot chatting.

Table 3 demonstrates the students’ attitude towards the CSIEC’s application in English instruction based on the collected data from the two questionnaires completed at the

<table>
<thead>
<tr>
<th>Table 3: Students’ attitude towards the Computer Simulation in Educational Communication’s application in English instruction</th>
</tr>
</thead>
<tbody>
<tr>
<td>Enhancing fluency of English</td>
</tr>
<tr>
<td>Graduate students</td>
</tr>
<tr>
<td>Middle school students</td>
</tr>
</tbody>
</table>
end of the experimental term. Because in the high school, the learning of the content in every course unit is stressed by the teacher, an item of ‘reviewing key points in the course units’ is specially added in the questionnaire. All the items were measured with a 5-point Likert agreement scale, ie, the value 5 indicates the maximum best agreement, and 1 indicates no agreement. For the five items of the graduate students, the Cronbach’s alpha is 0.933. So the reliability of these surveyed items is very good. All the correlations between any two variables are significant at the 0.01 level. For the six items of the high school students, the Cronbach’s alpha is 0.741. The reliability of these surveyed items is not very good, but acceptable. Six correlations between two variables are significant at the 0.01 level, and three correlations are significant at the 0.05 level.

From Table 3, we cannot recognise a big difference between the graduate students’ attitude and that of the high school students. It shows that high school students feel that the CSIEC-based English learning can help them with course unit review, make them more confident, improve their listening ability and enhance their interest in language learning. Another item in the questionnaire for the high school students shows 60.5% of the students “like” or “like very much” such a form of English learning, whereas only 2.3% dislike it. Among the high school students 60.5% will continue using the system after class, even without the teacher’s request.

Through the integration and assessment in English class instruction, some new functions have been added to the system according to the students and teachers’ suggestions and comments. They include the adjustment of speaking speed of the agent character, two-robots talk show, etc. Thus, the application and evaluation guide the development of the CSIEC system in the direction of users’ practical learning needs.

**Discussion and conclusion**

The original goal of the system is supplying a virtual chatting partner for English learners. So the chatting is the most fundamental function. The statistical analysis about the users’ behaviour indicates that the users have a preference for chatting without spelling and grammar checking. This fact proves that the users prefer the unique chatting function that is lacking in other systems. So we must continue to reinforce this primary utility.

The chatting quality can be somewhat shown by the chatting length. The increased percentage of the long and longer chatting shows that the free chatting quality of CSIEC is becoming better. The underlying design principle, ie, fully syntactical and semantic analysis of the user input, and communicative response mechanism, as well as the effort of chatting personalisation and adaptation contribute to the chatting quality progress. Certainly, the content analysis of the dialogues should also be carried out furthermore in order to investigate the chatting quality more exactly.

Chatting on a given topic is mainly used by the students in the evaluation study and is also the main function of the whole system the students have used. The formal evaluation results indicate that the application of CSIEC system in English class can better
assist their language learning, eg, enhance the fluency of English, the confidence on English communications and the interest in learning English, help them master practical expressions and improve listening skills. The planned functions, including free chatting and chatting on a given topic, and listening training have been brought into actual pedagogical play.

The CSIEC has been practically applied since its birth. We continue to improve its interface and content according to user feedback, either from free Internet users or from the English class students. Some new functions, such as the robot talk show and the adjustable speaking speed of the agent characters, are originally suggested by the users. The design, implementation, application and evaluation are not separated, but are seamlessly integrated together. This kind of application-driven research can quickly transfer the user’s demands into technical implementation and new emerging technologies into pedagogical application. Our system development procedure is consistent with the theory of design-based research. Design-based research came into birth in the 1990s, with the goals of filling the gap between the practice and the traditional evaluation research about the integration of technology and education, and enhancing the integration of technology into curriculum and learning efficiency through practice-oriented research (Reeves, Herrington & Oliver, 2005). It combines the learning environment design and theoretical development together and stresses research and development through the continuous and upgraded rotation of design, implementation, feedback and analysis (The Design-Based Research Collective, 2003).

Through application and evaluation, we find that, currently, there are still some user requirements that have not been fulfilled well, for example, the system’s stronger ability of natural language understanding and generation, which is the fatal factor influencing the human–computer communication, the lifelike synthesised agent voice and the high response speed, which also have been addressed in the users’ feedback. Solely in NLP, many problems are still hard to be solved, such as the textual ambiguity and entailment (Bar-Haim et al. 2006). How to tackle these problems with current available technologies is still a great challenge to us.

Acknowledgements
We are grateful for the support given by the Konrad-Adenauer-Foundation Germany, Ministry of Education China, Beijing University and Education Committee of Capital Beijing to our projects. We also thank the insightful and significant comments from the anonymous reviewers.

References


