Exploring the Internet for English oral proficiency development

Tsuo-Lin Chiu and Hsien-Chin Liou
Department of Foreign Languages and Literature
National Tsing Hua University, g925256@alumni.nthu.edu.tw

The Internet-based technologies were regarded as weak in providing opportunities for oral English practice as inherently the computer is not smart enough to provide intelligent interaction in authentic contexts, compared with the enhancement of other language skills with technologies. The paper first surveys available recent technologies by pointing out their strengths and limitations for teaching of speaking. Then, three evaluation cases (Payne & Whitney, 2002; Payne & Ross, 2005; Chiu, 2006) that used either online chat or speech recognition for oral proficiency development are highlighted because they have incorporated strong theoretical underpinnings into technology use. The results of the three studies indicated that the use of Internet-based technologies can indeed facilitate aspects of L2 oral development. Last, new perspectives such as maximized use of formulaic language, planning in task-based instruction, and intelligibility-based pronunciation instruction are discussed within the scope of technology use. The paper provides significant insights for English teachers who plan to make good use of Internet technologies in speaking-related courses, or researchers who take interest in examining oral proficiency development in technology-enhanced language learning environments.

INTRODUCTION

The Internet-based technologies were regarded as weak in providing
opportunities for oral English practice as inherently the computer is not smart enough to provide intelligent interaction in authentic contexts, compared with the enhancement of other language skills with technologies. With the advances of new speech technologies and innovative perspectives of technology use, the Internet nowadays can offer English teachers promising choices for oral proficiency development. The paper first surveys available technologies such as online chat, Internet phone (e.g., Skype), and speech recognition by pointing out their strengths and limitations for teaching of speaking. Then, three evaluation cases that used either online chat or speech recognition for oral proficiency development are highlighted because they have incorporated strong theoretical underpinnings into technology use. The studies with online chat (Payne & Whitney, 2002; Payne & Ross, 2005) used Levelt’s language production model (1989), combined with concepts from cognitive psychology, to explore L2 oral proficiency development via synchronous computer-mediated-communication environment. The project with speech recognition (Chiu, 2006) investigated 49 EFL college learners’ perception and oral production after learning from a speech recognition-based conversation environment with the theoretical underpinnings of speech acts. The results of the three studies indicated that the use of Internet-based technologies can indeed facilitate L2 oral development. Last, new perspectives such as maximized use of formulaic language, planning in task-based instruction, and intelligibility-based pronunciation instruction are discussed within the scope of technology use. The paper aims to provide significant insights for English teachers who plan to make good use of Internet technologies in speaking-related courses, or researchers who take interest in examining oral proficiency development in technology-enhanced language learning environments.
AVAILABLE TECHNOLOGIES FOR ORAL PROFICIENCY DEVELOPMENT

Advanced technology nowadays has made the Internet an alternative for the development of speaking-related skills. The sections below, categorized by the mode of speech interaction provided, would introduce some of these promising technologies. Their strengths and weaknesses for teaching and learning of speaking are also discussed.

Person-to-person (text)

Instant messaging is well-known for its convenience to communicate with people and its potentials for developing oral skills. Popular instant messaging services include MSN Messenger, Yahoo Messenger, AOL Instant Messenger, ICQ, etc. Studies in written forms of computer-mediated-communication (CMC) have found slower rates of conversation and more languages produced in chatroom than in face-to-face settings (Kern, 1995; Waschauer, 1996), more equal participation among students (Waschauer, 1996), and students’ more positive attitudes toward the target language (Waschauer, 1996). However, there are also concerns about whether the oral skills gained from the written mode of communication would eventually transfer to face-to-face communication. More critical perspectives are to question whether this form of communication taken place in computer-mediated environments is really culturally-neutral for learners of different cultural backgrounds and to heed the impact of distinct styles in CMC environments on learners’ literacy development (Kern, 2006).

Person-to-person (speech)

Internet phone services allow users to speak via Internet at no cost. Popular
Internet phone services provided included Skype, Google Talk, MSN Messenger, and Yahoo, etc. It offers more authentic interaction between person to person compared with the written mode of computer mediated communication. Video-conferencing capability is allowed by most of the Internet phone services for one to one video-audio interaction via web camera (webcam). It provides an alternative for distant language learners. Paralinguistic cues such as facial expressions, which are often perceived by learners as improving understanding, are maintained in video-conferencing. In addition, both Internet phone and video-conferencing are better performed with a wideband capability. Drawbacks of these technologies often include time lag, fizz, and dropout.

**Person-to-computer (text/sound)**

A. L. I. C. E. Artificial Intelligence Foundation ([http://www.pandorabots.com/pandora/talk?botid=f5d922d97e345aa1](http://www.pandorabots.com/pandora/talk?botid=f5d922d97e345aa1)) allows users to engage in text-chat with a virtual character similar to real-life conversation to certain extent. The computer can respond in both text and synthesized voice. Learners could practice formulaic interactions, such as greeting, asking for personal information, or leave-taking, with the virtual character. It seems to be suitable for learners who lacks experiences of talking to native-speaker or who are not ready to speak to practice basic linguistic structures. Authenticity is not guaranteed for more in-depth conversations.

**Person-to-computer (speech)**

Recently the application of Automatic Speech Recognition (ASR) technology shows potentials in enhancing language learning in the areas of speaking and pronunciation. Researchers in the field have evidenced the advantages of applying
ASR during speaking interaction (Bernstein & Najmi & Ehsani, 1999; Egan, 1999; Ehsani & Knodt, 1998; Eskenazi, 1999). Speech recognition technology provides an alternative for practicing both pronunciation and speaking. Software with speech recognition capability includes MY ET (*MY ENGLISH TUTOR*, [http://www.myet.com](http://www.myet.com)), TraciTalk ([http://www.hkbu.edu.hk/~sall/english/material/cdrom/tracitalk.html](http://www.hkbu.edu.hk/~sall/english/material/cdrom/tracitalk.html)), and *CandleTalk* ([http://candle.cs.nthu.edu.tw](http://candle.cs.nthu.edu.tw)). The technology shows potentials in developing oral skills, either pronunciation or speaking, based on different instructional designs. It provides an opportunity for learners who could not talk to real people. It may be useful in directing learners’ attention on persistent production errors (Derwing, Munro, and Carbonaro, 2000). However, spontaneous speech is still not allowed and its recognition errors, which might lead to false feedbacks, are inevitable.

THREE EVALUATION CASES

Three evaluation cases that used either online chat or speech recognition for oral proficiency development are highlighted because they have incorporated strong theoretical underpinnings into technology use.

**Online Chat Based on Cognitive Psychology Theories**

The two studies with online chat (Payne & Whitney, 2002; Payne & Ross, 2005) used Levelt’s language production model (1989), combined with concepts from cognitive psychology, to explore L2 oral proficiency development via synchronous computer-mediated-communication environments.

Levelt’s (1989) model is the most widely adapted model for depicting L2 or bilingual language production. It has indicated the importance of short-term storage of information in language production, in which the language production process has been depicted, but this aspect has not been fully developed. Working Memory theory
(Baddeley & Hitch, 1974) could further explain an individual’s capacity for temporarily maintaining verbal and visual-spatial information in memory for performing judgments based on the immediate conditions. Research in first and second language has shown that individual differences in Working Memory capacity are related to the aspects such as vocabulary development, the ability to contextual clues for learning new words, and language proficiency. Moreover, the two Working Memory concepts related to the task of conversational exchange are executive capacity, which is the ability to maintain language input and calculate the retrieval of syntactic and semantic information from memory, and verbal span, which is the ability to temporarily maintain phonological information. Exploring the relationship between these two concepts and L2 oral language oral proficiency development may shed light on the impact of memory limitations on the task as conversing in a foreign language.

Slower rate of conversation and opportunities for refreshing memory traces were found to be prominent features of the languages produced in synchronous computer-mediated-communication (SCMC) environments. In the first study Payne and Whitney (2002) thus attempted to test the hypothesis that SCMC with its features could indirectly improve L2 oral proficiency by using Levelt’s (1989) model as a basis, augmented with the two Working Memory concepts. Participants were 58 Spanish learners divided into two experimental groups and two control groups. They assumed that the Working Memory may prove to be a useful construct for predicting what types of learners may benefit from SCMC. They hypothesized that learners with lower Working Memory capacities might benefit from a conversational environment where processing demands are reduced, but where the tasks and interactions are the same. There are two research questions in their study: (1) whether L2 oral proficiency can be indirectly developed through chatroom interaction in the target language; (2)
whether individual differences in Working Memory capacity can effectively predict the rate of L2 oral proficiency development for different types of learners in a chatroom setting.

Three major results were found: first, the experimental groups outperformed the control groups, suggesting that the participants spending half of their instructional time in a synchronous online environment were advantaged in their oral proficiency development over those meeting face-to-face. Second, the results suggested that executive capacity appears to have no relationship with oral proficiency development. Third, learners with lower phonological capacity were disadvantaged relative to others in the control group but were not so disadvantaged in the experimental group, suggesting that chatroom environment may be especially beneficial for students with lower ability to maintain verbal information. In addition to the major findings reported, they also found that the turn-taking rules in the chatroom are suitable for stimulating discussion among students, and students are able to produce more language in chatroom settings. Compared with face-to-face interaction, learners in the chatroom setting can not resort to paralinguistic compensation strategies so that they are pushed to experiment with the language by testing the meaning of lexical items used and unfamiliar syntactical patterns. The chatroom setting also provides the opportunity for monitoring learners’ own language and the language of others.

The second study (Payne & Ross, 2005) about the use of the chatroom on developing L2 oral proficiency, carried on the same theoretical framework as Payne and Whitney (2002), was to further examine 24 Spanish learners’ patterns of language use as evidenced in the chat transcripts with an emphasis on two aspects: repetition and relexicalization, and to better understand the interplay between individual differences in working memory capacity, SCMC, and cross-modality transfer of skill from chatting to oral proficiency development.
The participants were divided into an even split of high- and low-span groups based on two tasks, which were a nonword repetition task and a reading span measure, making working memory span an independent variable. The results revealed that the use of repetition and relexicalization in L2 discourse declined in frequency over time, which could be possibly explained by the inverse relation between repetition/relexicalization and proficiency. Furthermore, it was found that there was an apparent difference in the chatting style of the low phonological working memory students, suggesting that they were taking advantage of the reduced cognitive burden by the chatroom to produce more extensive and complicated utterances. Therefore, although the study showed that individual differences in working memory could not be found to explain the drop in the frequency of repetition and relexicalization, the researchers claimed that the findings of this study lend further support to the notion that the chatroom may provide a unique form of support to certain types of learners in developing oral proficiency, which was consistent with the results of Payne and Whitney (2002).

**Speech Recognition Based on the Speech Act View**

The project with speech recognition (Chiu, 2006) investigated 49 EFL (English as a Foreign Language) college learners’ perception and oral production after learning from a speech recognition-based conversation environment based on the theoretical underpinnings of speech acts. Speech acts are the units of words that speakers use for achieving communicative functions (Austin, 1962). Successful speech act performance should be both sociolinguistically and socioculturally appropriate, which means that the linguistic units chosen should be conforming to the contexts (Cohen & Olshtain, 1994). Non-native speakers tend not be able to acquire the norm of successful speech acts performance, and their performance of speech acts tends to be
verbose, lengthy, or repetitive (Blum-Kulka et al., 1989, Cohen & Olshtain, 1981). Thus, explicit instruction on appropriate use of speech acts is needed.

With carefully designed materials, ASR-based speaking programs show potentials to engage learners in simulated face-to-face conversation with virtual characters, and they also allow learners to feel at ease while learning to speak. Previous studies on commercial software MYET showed that the use of speech recognition technology is helpful for pronunciation training for beginning learners (Chen & Chiu, 2005; Tsai, 2003). However, no evaluation study was reported on the use of speech recognition technology on helping learners develop other speaking-related skills. Thus, Chiu (2006) investigated the effectiveness of applying an ASR-supported conversation environment, CandleTalk (http://candle.cs.nthu.edu.tw, under speaking), on improving EFL learners’ speech act performance. There were two research foci: first, to explore the application of ASR technology in CandleTalk on improving students’ oral production in English; second, to understand students’ perceptions toward the use of ASR-based instruction in CandleTalk.

49 participants of the study were divided into a English-major group and a non-English-major group. After about one month treatment, the comparison of pretest and posttest oral performance of a discourse completion test using t-test indicated that learners in general improved their overall performance and their use of speech acts, but not in their comprehensibility of speech. A further investigation on the difference between the two groups revealed that it was the non-English-major group who improved significantly regarding overall performance and speech act use, while the English-major group only had slight but no significant increase in the scores of each category. In addition, regarding learners’ perceptions toward the ASR-based instruction in CandleTalk, it showed that learners held positive attitude toward the instruction with speech recognition technology and they believed that the instruction
with ASR technology would assist their speaking in English. They perceived positively toward the virtual conversation provided by the environment and the inclusion of topics about their home culture into the dialogues. However, they also perceived negatively toward the recognition errors, which are still limited for current speech recognition technology, and the complex recording interface design in the environment.

**Implications of the Three Evaluation Studies on Oral Proficiency Development**

The two studies of online chat (Payne & Whitney, 2002; Payne & Ross, 2005) implicated that L2 oral proficiency can be indirectly developed through chatroom interaction in the target language. Owing to the reduced memory load in the chatroom setting, it is especially beneficial for learners with lower phonological Working Memory capacities. Payne and Whitney recommended that chatroom may well serve as a conversation simulator for foreign language learners. They did not claim that the chatroom setting can replace face-to-face instruction, but it can provide another choice for language learning in which affect and rate of speech are reduced, which might be appealing conditions for certain learners.

The study of speech recognition (Chiu, 2006) could give a number of implications concerning oral proficiency development. Since lack of opportunities for speaking in English is considered a common problem for EFL learners, instructional materials with ASR could help learners who are too shy to speak and provide speaking-opportunities for learners who rarely have chances to speak with native speakers. In addition, the study showed that the application of ASR in language learning could be effective in helping learners practice appropriate use of speech acts. Further, the speech recognizer could give immediate individualized feedback to learners based on their oral performance and thus help relieve teachers’ loads. Given
the large class size in most Asian countries, it is both laborious and time-consuming for teachers to attend to each student’s performance in speaking. Therefore, judicious use of ASR materials is helpful on relieving teachers’ loads and offering learners opportunities for individual learning.

NEW PERSPECTIVES FOR ORAL PROFICIENCY DEVELOPMENT

In this section, new perspectives such as maximized use of formulaic language, planning in task-based instruction, and intelligibility-based pronunciation instruction are discussed within the scope of technology use.

Formulaicity in Language

The definition of formulaicity or formulaic sequence is: “a sequence, continuous or discontinuous, of words or other meaning elements, which is, or appears to be, prefabricated: that is, stored and retrieved whole from memory at the time of use, rather than being subject to generation or analysis by the language grammar” (Wray, 2000, p. 1). The basic concept of formulaicity by Wray (2000, 2002) is to see native speakers’ use of language as formulaic sequences, which are subsets of all the possibilities in a language. Native speakers do not try to exercise all the grammatical possibilities in a language and only a small portion of grammatically acceptable sentences exercised are considered native-like. Moreover, the meaning of a formulaic sequence exercised by native speakers is often processed holistically, instead of componentially. Hence, for non-native speakers who are not sufficiently immersed to a native speech community, they might fail to obtain the full range of expressions made by native speakers, and the consequence is that they might create their own language use which is grammatically correct but non-native like.

Formulaicity in language could serve as a short-cut for processing and a means
of facilitating social interaction (Wray, 2000, 2002). On one hand, formulaic expressions could impose fewer burdens in the processing stage since the speaker could quickly retrieve the prefabricated expressions. On the other hand, during social interaction, the use of formulaic expressions, such as greeting, requesting, are manipulative in nature so that the speaker could anticipate the desired response from the hearer, and meanwhile the hearer is more likely to respond in the desired manner. It lessens the risk of misunderstanding in the side of hearer, and it helps both sides to better achieve the purpose of the conversation.

Implications drawn on this perspective could be on the explicit teaching or learning of use of speech acts for L2 learners to develop better communicative skills as illustrated above (Chiu, 2006). By focusing on the formulaic usages of speech acts in diverse social encounters, it could help learners facilitate more effective communication. Online tools such as a collocation concordancer (http://candle.cs.nthu.edu.tw/collocation) could serve as a referencing tool for catching the norm of formulaic usages as illustrated below using the British National Corpus academic collocation corpus as illustrated in Table 1.

Table 1: Collocation output using an online referencing tool

<table>
<thead>
<tr>
<th>1. agree with view (15)</th>
</tr>
</thead>
<tbody>
<tr>
<td>You may like to consider how far you agree with this view.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>2. agree in principle (14)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Last year 's iwc meeting agreed in principle a new scientific procedure for sustainable catches of whales, although it was not clear if Norway used the procedure to calculate its catch.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>3. agree with observation (12)</th>
</tr>
</thead>
<tbody>
<tr>
<td>But if the theory was mathematically consistent and always gave predictions that agreed with observations, we could be reasonably confident that it was the right one.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>4. agree with conclusion (10)</th>
</tr>
</thead>
<tbody>
<tr>
<td>I would totally agree with that conclusion and I would say that in the York situation, you should choose the one that is closer to York than is farthest out.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>5. agree with those (10)</th>
</tr>
</thead>
</table>
Cytogenetic analysis always failed with coelomic fluid, but fetal sexing was always successful with fluorescence in-situ hybridisation and polymerase chain reaction, and the results agreed with those obtained from chorionic villi and amniotic fluid in all cases.

Planning Time
The effect of planning time on language learners’ language production has drawn lots of attention, especially in the area of speaking assessment under the task-based instruction framework. Li and Chen (2005) examined the influence of planning on the speaking performance of two groups of adult EFL learners. In their study thirty adult EFL learners were divided into a high proficiency group and a low proficiency group based on their TOFEL scores. The learners completed two tests with similar difficulty levels, and the first test was done with planning time given and another was done without planning time. The major result of the study indicated that planning was more effective for the low proficiency group in which more complicated and accurate utterances were produced in the test. The study implicated that different lengths of planning time should be given to learners of different proficiency levels during speaking assessments. A case of online speaking assessment unit was presented at Chiu, et al. (2006, http://candle.cs.nthu.edu.tw, under Speaking) in which planning time is given to learners before responding to the situations given. Hopefully, with planning time given, the online unit lessens test anxiety and elicits the best students’ oral performance.

Intelligibility-based Pronunciation Instruction
The history of pronunciation instruction has been guided by two leading principles: the nativeness principle and the intelligibility principle (Levis, 2005). The goal of the nativeness principle is to attain an accent of a prestigious inner-circle modal, such as RP (Received Pronunciation) or GA (General American English),
while the goal of the intelligibility principle is to achieve mutually intelligible pronunciation. The nativeness principle is often criticized as unrealistic because few English learners really attain its goal, especially for learners who learn English after puberty. Therefore, regarding the lingua franca status of English today, the intelligibility principle seems to be a more realistic goal for language learning.

The instruction following the intelligibility principles is often sensitive to contexts. Jenkins (2000, 2002) argued that in English as an international language contexts (EIL) learners’ pronunciation does not necessarily need to adapt to native-speaker norms but should adjust their speech to primarily nonnative speakers. Similarly, Derwing and Munro (2005) took mutual intelligibility as primary concern when addressing pronunciation teaching in English as second language contexts (ESL). The difference between intelligibility-based pronunciation instruction in EIL and ESL contexts is that, as addressed by Derwing and Munro, “ESL learners have to make themselves understood to a wide range of interlocutors within a context where their L2 is the primary language for communication and where, in many cases, NSs (Native Speakers) are the majority” (p. 380). Therefore, when the teaching of pronunciation is concerned, it would be more realistic for language instructors to address pronunciation based on the context of instruction.

Since achieving mutual intelligibility is recognized as an appropriate goal for pronunciation instruction, the next step is to find out the factors contributing to intelligibility. Several factors could affect intelligibility of a speech, such as pronunciation, grammar, or discourse. Some pointed out that prosody was rated as a main factor contributing to accentedness which might affect intelligibility. Munro and Derwing (1995) in their research methodology viewed intelligibility as the ability to transcribe the actual words of an utterance. Later, Munro, Derwing, and Morton (2006) attempted to find out whether it is stimulus properties which decide the intelligibility
of a speech or it is listener factors which decide the intelligibility by investigating 48 participants’ evaluations of four different accented speeches. The results showed that listener factors were relatively minor contributing to the intelligibility of a speech since listeners of the same accent did not exhibit an advantage in understanding their own accent. Properties of the speech are more influential when understanding an L2 speech. The study implicated that the effects of L1 background and experience with a particular type of accent were relatively minor factors in the ability to understand the L2 speech. Further, it was suggested that accented speech is still possible to understand only if the listeners were trained or willing to understand it.

Four factors contributing to successful pronunciation instruction under the scope of technology use were proposed by Neri, et al. (2002) and Eskenazi (1999). These include (1) input, (2) output, (3) feedback, and (4) reduction of stress. A speech recognition-supported environment aiming at pronunciation, for example, first should allow learners to be exposed to large amounts of input in order to construct the language model of their own. The way input was presented to learners should be contextually meaningful so as to better stimulate learners’ intrinsic motivation. Second, merely input is not enough for pronunciation training. Learners need to have opportunities for speech production, and through production of the L2 they could compare their output with the input model to form a correct L2 model. Besides input and output, pertinent and corrective feedback given by the speech recognizer should help learners notice the discrepancies between their output and the L2. Feedback given should be able to pinpoint learners’ errors or offer suggestions. In the cases where scores are given as feedback, then learners should be able to interpret the score received. Last, a stress-free environment is needed to encourage learners for more speech production. An example of an instructional material following the designing principles is given in Chiu, et el. (2006) aiming at training intelligible pronunciation
for EFL learners (http://candle.cs.nthu.edu.tw, “how to compliment” under the Speaking component).

CONCLUSION

The development of teaching speaking using technologies is at its infancy stage. It is hoped that the paper provides significant insights for English teachers who plan to make good use of Internet technologies in speaking-related courses, or researchers who take interest in examining oral proficiency development in technology-enhanced language learning environments. Various innovative technologies are under development and the delivery efficiency, conditioned by either audio or video, will be improved in the near future. We have pointed out different mechanisms available for teaching oral skills where English teachers can make a wise decision on what to choose for their pedagogical purposes.

The three case studies illustrated what teachers could make good use of and what researchers may want to know, with the theoretical underpinnings. Synchronous computer mediated communication is different from the common oral mode humans use to interact. Still, as learners in the chatroom setting can not resort to paralinguistic compensation strategies, they would be pushed to experiment with the language by testing the meaning of lexical items used and unfamiliar syntactical patterns. That is the advantage of using online chat for teaching speaking. The study using speech recognition shows that CALL developers may want to use the technology for skills other than pronunciation.

Three suggested perspectives point out crucial issues for L2 instruction and assessment of oral skills. Maximized use of formulaic language on EFL learners leads to more native-like and fluent oral performance, as a recent study (Liao & Fukuya, 2004) indicates Chinese learners of English tend to avoid formulaic language such as
phrasal verbs. Planning in the task-based teaching tradition may tell teachers how to
evaluate learners’ anxiety in oral skill courses and what factors to consider in order to do
fair oral performance assessment. Again, intelligibility-based pronunciation
instruction opens up a number of sociocultural and pedagogical issues related to the
English-as-a-global language phenomenon and L2 oral proficiency development for
both teachers and researchers to reflect upon.

ACKNOWLEDGEMENTS: The paper is partly supported by grants from National
Science Council research projects (NSC 93-2524-S007-001; NSC 94-2524-S007-001).

References


advances in learning and motivation* (Vol. 8, pp. 47-90), New York: Academic
Press.


Selected papers from the 14th Int’l symposium on English Teaching (pp. 58-69),
Vol. 2. Taipei: Crane.

Chiu, T. L. (2006). *Effects of online conversation materials with the support of speech
recognition technology on college EFL learners*. MA thesis, Tsing Hua
University, Taiwan.


