Chapter 3

VOICE INTERACTION ONLINE

Terumi Miyazoe[1] and Terry Anderson[2]
[1]Tokyo Denki University, Tokyo, Japan
[2]Athabasca University, Athabasca, Canada

ABSTRACT

This chapter examines the potential for voice activities to enhance online learning. Although research related to online writing tools (such as e-mail, discussion threads, blogs, and wikis) is growing and the use of synchronous and asynchronous voice tools (such as internet phone, *VoiceThread* and multimodal web conferencing) has developed rapidly, little is known about the effects of systematically implementing these voice tools in formal educational settings (Millard, 2010). This chapter first provides a brief overview of the available online voice systems, the design principles of online voice interaction, and prior research and frameworks regarding voice interaction in online and blended learning. It then reports on a study of a systematic application of an asynchronous voice recording system integrated into a learning management system in an undergraduate blended-learning English course. The research found that the voice system was adequately usable and was associated with a positive change in the students' perceptions of speaking English over the semester. Half of the students were reluctant to talk to an online program, whereas the other half were willing to do so. Regular use of online voice assignments encouraged high attendance rates for the in-person classes. This suggests the potential power of asynchronous online voice interaction to help support students' learning.

INTRODUCTION

The application of online voice systems as an integral component of teaching practice is a new area in urgent need of exploration in the online education field (Kenning, 2010). In this chapter, online voice systems are defined as online systems with the ability to record, save and transmit audio data. This chapter uses the term *voice* to highlight the particular feature of aural and oral data generated by human speech, as distinguished from more general audio

data generated by nonhuman devices. The following sections provide the research background by presenting three topics: 1) an overview of the major online voice systems currently available, 2) a framework for designing online voice interaction in online and blended learning, and 3) a review of relevant prior research on the use of online voice systems for educational purposes. Next, we present the results of a research study conducted by the authors that investigates asynchronous online voice systems use in an undergraduate English course to enhance speaking ability.

FRAMEWORK OF ONLINE VOICE INTERACTION

Voice Interaction and Time Factors

We currently have a number of choices for tools that support online voice interaction in our daily lives. Internet phone may be the most familiar. Voicemail, multi-user voice chat, and audio/video messaging are also available. It has been customary and useful for educational purposes to categorize online communication tools as either *synchronous* (same time) or *asynchronous* (different time). For example, voice messaging (similar to e-mail messaging) is asynchronous because messages are received after the initial act of sending them is completed and can be responded to any time in the future, whereas voice chat is synchronous because the voice chat and responses happen with near instantaneous turn-taking. However, with increasingly faster internet connectivity and ever-greater PC capacities, the distinction between the two systems is becoming less evident from a technological perspective, but remains so from an educational perspective in which the length of time delay can be adjusted to either increase spontaneity or increase opportunity for reflection.
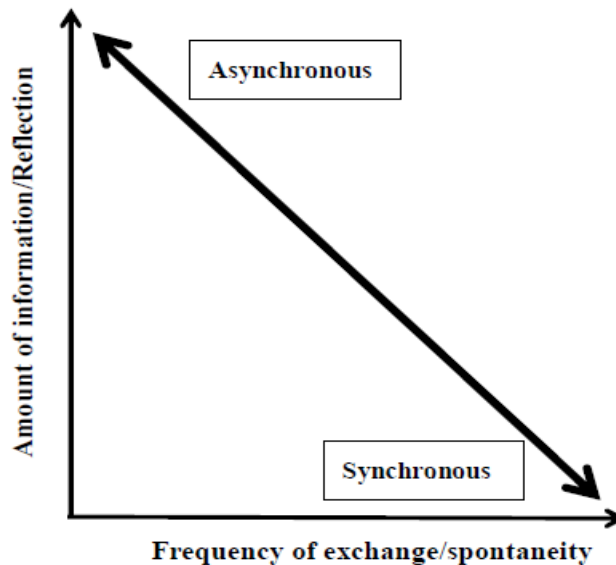

Figure 1. Time concept in online communication.

Figure 1 is a tentative framework that expresses the increasingly blurred relationship between synchronous and asynchronous online communication tools. The vertical axis shows the amount of information sent, and the horizontal axis shows the frequency of exchange. When we need to increase the frequency of exchange, we usually shorten each message to ease the processing burden and reduce the system load for both the sender and the receiver. On the other hand, the more information we try to send or share, the more likely we are to rely on asynchronous communication tools (i.e., e-mails, podcasts, and video sharing) because large files are often difficult to send and receive. Therefore, both the frequency of exchange and the amount of information determine which communication mode, synchronous or asynchronous, is most practical. That is, the distinction between synchronous and asynchronous has led to a tendency to use a specific online system more in certain situations.

Online Voice Systems

Recently, voice systems are being integrated into the suite of online teaching and learning tools known as learning management systems (LMS). This integration provides systematic design convenience of access and management over site. Some systems are commercially available (for example, *Wimba*: http://www.wimba.com/), and some are free-to-use Web 2.0 tools (for example, *Voicethread*: http://voicethread.com/ and others such as *NanoGong*: http://gong.ust.hk/nanogong/ are open source tools that can be integrated into open source LMS. There are also institution-specific voice systems for specific needs, such as the audiographic chat system *Lyceum* at Open University UK.

Figure 2. Lyceum with Whiteboard (provided by Hampel, 2010).

Lyceum offers a multimodal learning environment that provides several tools, voice chat, text chat, polling, a whiteboard, concept mapping, and word processing (Hampel, Felix, Hauck, and Coleman, 2005; Vetter and Chanier, 2006). *Lyceum* was discontinued recently in favor of a commercial web conferencing system (*Elluminate*: http://www.elluminate.com/). Figure 2 is a screenshot of a synchronous session of a German class using the whiteboard feature in *Lyceum*. We can see several student attendees in the right lower corner, along with one student speaking and another text chatting on the lower bottom.

Figure 3 is a screenshot of *Lyceum* with a concept map. On the lower left, we can see that one student "raises a hand" to signal his/her desire to contribute to the discussion. Participants can add entries on the same concept map screen simultaneously from different places. Though it was not equipped with a video function, *Lyceum* paved the way for today's synchronous web conferencing systems.

VoiceThread is an asynchronous tool set that provides both free and priced services to meet the specific needs of users. Its free version provides text, pictures, slides, and videos on the VoiceThread.com website that users can share and comment on in both voice and text modes. Users may also draw or doodle in the video. Voice threads can be embedded into other social network services, including an open-source learning management system called Moodle. Figure 4 shows an example of embedding a photo picture slide with four people's voice comments onto Moodle. The limitation of this action is that any user can comment on or delete another's comment.

VoiceThread's priced option allows greater control of contributors' identities by matching VoiceThread accounts to Moodle accounts and allows the grading of voice assignments using Moodle Gradebook. Figure 5 shows an example of a course instructor's grading of students' voice assignments. Students can also view and comment on one another's work within this program.
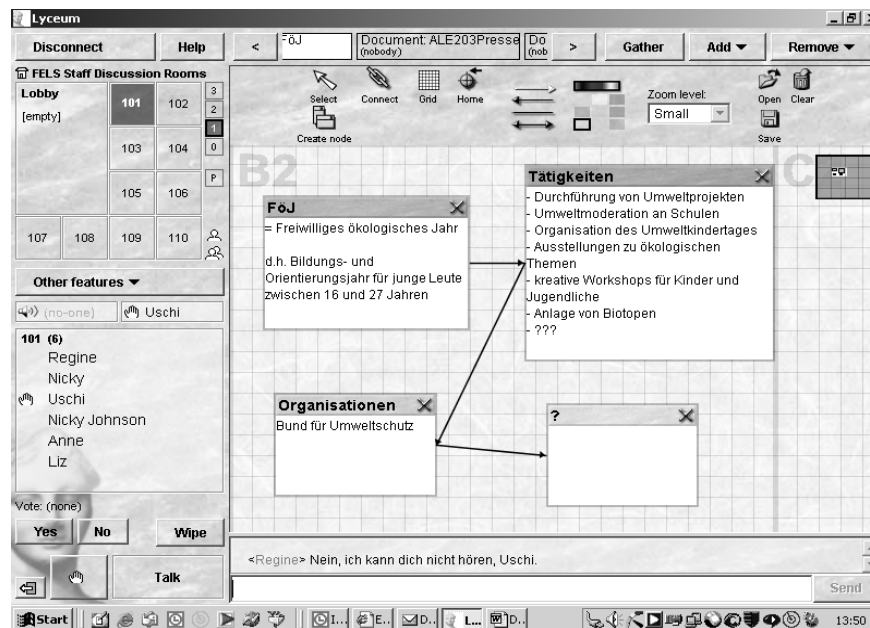
Figure 3. Lyceum with concept map (provided by Hampel, 2010).

Figure 4. VoiceThread voice comments on slides and videos.

Figure 5. VoiceThread voice assignments on Moodle.

NanoGong is the online audio communication system developed by the Hong Kong University of Science and Technology. It offers two versions of a stand-alone system called Gong and NanoGong, the latter of which can be integrated into Moodle and can work like

other Moodle assignments for submission and grading. Since it became available free of charge in June 2006, it has been used by thousands of people around the world (Rossiter and Lam, 2007; The Gong Project, 2010a; The Gong Project, 2010b). Using an HTML editor, NanoGong seamlessly permits voice activities that Moodle originally provided in text mode, such as voice forums and voice blogging, with only a few clicks. Figure 6 shows an example of a student's voice blog with a picture and narration.

In addition to NanoGong, Moodle voice modules are under development. These include Voice Record (voice recording for grading) and Voiceshadow (voice training assignments) that specifically target language acquisition (Daniels, 2011). These modules will be discussed further in the chapter's Research section.

Three notable points about voice systems conclude this section. First, there are multiple ways to use voice tools for pedagogical purposes, such as providing voice feedback, discussion and debate, voice blogging, and voice storytelling. Many of these functions have previously been text-based but have become voice-based with improved online speed and quality. It is reasonable to expect that additional text-based functions will become voice-based (including video). We should not be overwhelmed by new technologies and terminology—educators have been using these types of interaction for a long time. Second, voice activity implementation requires new technologies, such as microphones, web cameras, and video recorders, in addition to the minimum equipment of a PC with an internet connection. Although these technical capacities are built into most new computers, these technological requirements could be a nonnegligible factor that limits the application of online voice interaction in practice. Third, though a stand-alone voice system is also useful, LMS integration has the special practical value of allowing more systematic course design, management, integration, ease of use and evaluation. Furthermore, especially when LMS is already a stable aspect of online teaching and learning, it helps to provide a converging point for various interfaces to induce greater learner engagement and less confusion.
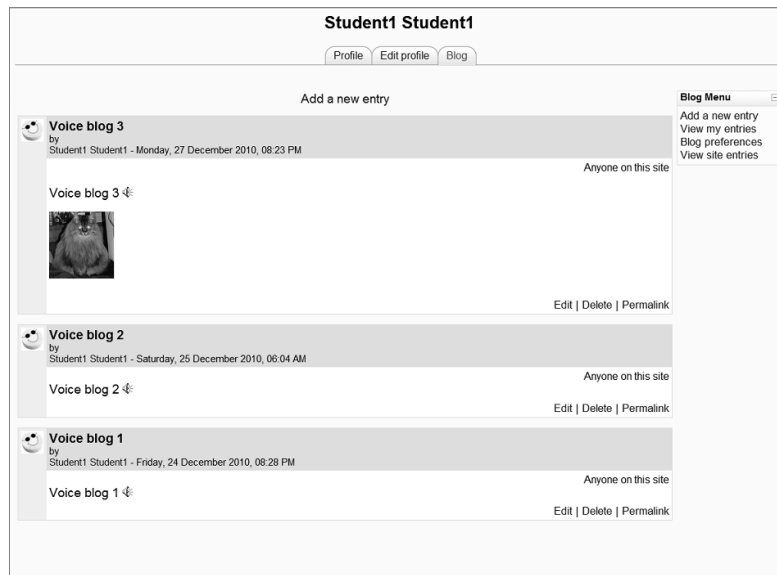

Figure 6. Voice blog using NanoGong.

Voice Interaction in Instructional Design

Figure 7 shows the Modes of Interaction Typology, a framework for interaction in instructional design posited by Randy Garrison and Terry Anderson (2003). It was originally introduced in Anderson and Garrison (1998) (Moore and Anderson, 2003). Garrison and Anderson, the pioneering figures in online and distance education, are based at the University of Calgary and Athabasca University in Canada. Their typology is unique because it originated in distance education and developed to cover online learning as a whole. Garrison and Anderson's model encompasses the critical elements of the interaction prototype posited by Moore (1989) in the United States, which defines three interaction pairs, *student-student*, *student-teacher*, and *student-content*, in learning. However, Garrison and Anderson extend this model by pointing out that *teacher-content*, *teacher-teacher*, and *content-content* interactions are also required to provide a significant learning experience. The typology is introduced here because the dimension of *content* is particularly important when considering online learning. There is no doubt that self-learning with online programs and digital materials accounts for an indispensably large part of the learning process, in addition to the human interaction between *teacher* and *student*.

In combination with the *Modes of Interaction Typology*, Figure 8 provides a framework for designing voice interaction by learning modes (face-to-face, solely online, and blended learning, which combines face-to-face and online instruction) along the parameters of synchronous/asynchronous online communication tools in instructional design. Cross marks (×) signify lower effectiveness, single circles (○) signify higher effectiveness, triangles (△) signify moderate effectiveness, and stars (☆) indicate very high effectiveness. Because the focus of this chapter is voice interaction, content-content interaction is exempted from the table; however, because online programs often collect, analyze, and perform actions without human intervention, it is possible that digital voice data may "talk" to each other as intelligent robots in the very near future.
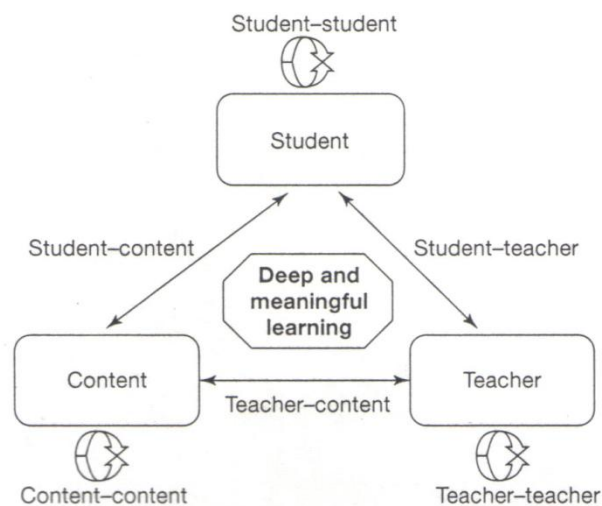

Figure 7. Modes of interaction typology (Garrison and Anderson, 2003, p.43).

Voice interaction in face-to-face mode, such as oral lectures, discussions, and presentations, is efficiently executed without the intervention of online systems, though there are exceptional situations. The rationale is that when people are in the same physical place, speaking face-to-face is the easiest way to communicate. In solely online mode, when teacher(s) and student(s) are physically far apart, synchronous interaction such as voice chat and web conferencing is often planned to incorporate face-to-face learning features. However, especially when participants access the system from different time zones, asynchronous communications such as voice messaging and archived video viewing are preferable to avoid time conflicts. With blended learning, it is reasonable to plan oral interaction as a face-to-face mode (synchronous) during class meetings and online components (asynchronous) outside the class to take advantage of the merits each mode provides. Additionally, it is reasonable to use asynchronous online communication to avoid the difficulties of participating at the same time from different places outside the class. In both solely online learning and the online portion of blended learning, *student-content* and *teacher-content* interactions are highly effective and efficient beyond synchronicity and asynchronicity because both students and teachers can learn and work independently at their own pace.

The key here is that the best voice system for a given purpose differs depending on numerous factors, such as learning mode, available technology, and learner needs and preferences (Miyazoe and Anderson, 2010a). There is no need to assume that synchronous voice communication is the best mode for speaking. Rather, higher value may arise from the asynchronous capability that online voice systems offer for online learning.

Learning mode	Time factor					
	Synchronous	←		→	Asynchronous	
Face-to-face	---	---	---	---	---	---
Online	Student-student	△			Student-student	○
	Student-teacher	△			Student-teacher	○
	Teacher-teacher	△			Teacher-teacher	○
			Student-content	☆		
			Teacher-content	☆		
Blended (online portion)	Student-student	×			Student-student	○
	Student-teacher	×			Student-teacher	○
	Teacher-teacher	×			Teacher-teacher	○
			Student-content	☆		
			Teacher-content	☆		

Figure 8. Suitability of voice interaction patterns in different learning modes.

Voice Interaction and Learning Theories

It is well-known that the three major learning theories (behaviorism, cognitivism, and constructivism) have defined the epistemology of learning (Driscoll, 2005), and each continues to influence how we think and act in education. In brief, behaviorism considers learning as the observable change in behaviors and expected learning outcomes that are conditioned by the reinforcement of a stimulus-response association; cognitivism looks at the process of learning and considers learning the internal change in the learner's perception; and constructivism views learning as constructed by the learner in relation to their unique and social contexts (Kearsley, 2010; Miyazoe, 2008).

The use of technology has changed in accordance with changes in beliefs about how learning occurs. Computer-assisted instruction (CAI) and programmed instruction (PI) follow behaviorism models; multimedia learning and the examination of the effects of different media are a reflection of cognitivism; and the emphasis on the authenticity of learning with collaboration and interaction, as represented by computer-supported collaborative learning (CSCL), is in line with constructivism (Mochizuki, 2010; Nishimori, 2010; Shigeta, 2010).

Figure 8 expresses the relativity of the three learning theories to the learner's knowledge and the difficulty of a given task (Ertmer and Newby, 1993). It shows that behaviorism, cognitivism, and constructivism are not exclusive, but coexist to support different types of learning. For learners who are less prepared developmentally, it is beneficial to reduce task complexity and use a more behaviorist task design. For more advanced learners, tasks of higher complexity involving more interactive elements would be advantageous to induce higher learning outcomes.

When we apply this relativist thinking to voice interaction, *student-content* interaction, such as practicing and training-focused aspects of speaking, is more suitable for learners of developing stages.
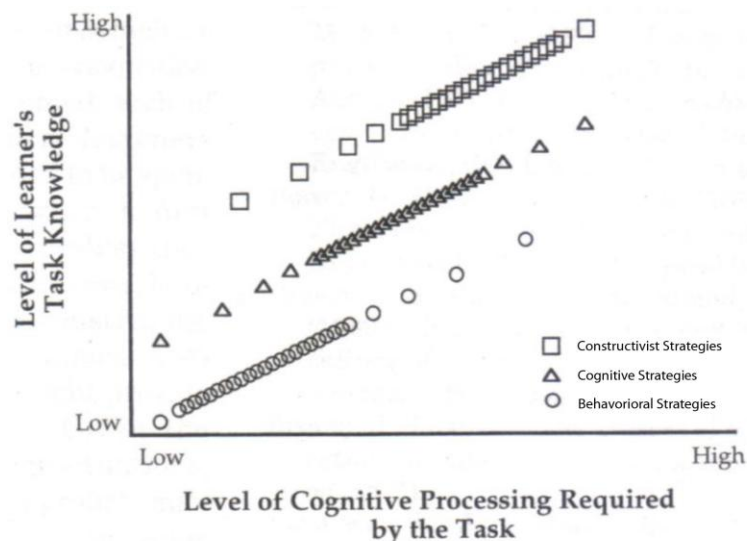

Figure 9. Comparison of the instructional strategies associated with the behavioral, cognitive, and constructivist viewpoints based on the learner's task knowledge level and the level of cognitive processing required by the task (Ertmer and Newby, 1993, p. 69).

Additionally, the simple design of online learning materials helps reduce confusion. Further, when the learning stages become more advanced, better learning outcomes can be achieved by using multimedia applications that suit the learner's cognitive load, such as asynchronous video viewing and voice recording, and that allow *student-content* interaction and delayed *student-teacher* interaction. Even at advanced stages, a highly complex interaction design that mixes *student-content*, *student-teacher*, and *student-student* interactions, a multimedia environment of text, graphics, and audio and synchronous engagement that demands frequent responses within a short time can provide further challenges. In sum, the most suitable voice interaction design depends on the specific context and its specific needs. For each behaviorism, cognitivism, and constructivism, the optimal instructional design shall be different (Suzuki, 2006).

Voice Interaction Studies

It is no surprise that many of the studies and educational applications that apply voice elements to online education are found in the fields of language education and distance education. This results from the necessity of including speaking and listening activities in an integrated language acquisition course plan (Hampel, 2003). Conscious efforts have been made in distance education to bridge the physical and psychological gap between distance learners and teachers, also called *transactional distance* (Moore, 1993), using new technologies.

Among numerous learner factors, this study highlights *speaking anxiety* as a key concept. *Lyceum* was implemented in 2002 (Hampel, 2003) for language courses such as English, German, Spanish, and French at Open University UK, and numerous studies have been done using *Lyceum* by Open University UK scholars (Kenning, 2010). Speaking anxiety is relevant to online voice interaction in two ways. The first is the speaking anxiety that learners potentially experience when talking online without sharing the same physical space with associated interpersonal clues (de los Arcos, Coleman, and Hampel, 2009; Hampel, Felix, Hauck, and Coleman, 2005; Hauck and Hurd, 2005; Pichette, 2009). The second is the more general speaking anxiety that language learners generally express when speaking (Lamy and Hampel, 2007). In other words, the first is the uncertainty of speaking to a machine, with limited response, and the second is the fear of speaking in a foreign language. The issue of speaking anxiety was important when we implemented the online voice interaction used for this chapter's research context (Japan) because Japanese learners are said to show even higher speaking anxiety rates than learners of different nationalities (Cutrone, 2009; Miyazoe and Anderson, 2011; Williams and Andrade, 2008).

Rosell-Aguilar (2007) also approached the issue of the new tutor role in audiographic conferences and concluded that individual tutors' social and human traits ("personality, warmth, and ability," p. 81) become even more important than their traditionally valued technical and pedagogical abilities. The study by Rosell-Aguilar (2006) is suggestive because it illustrates the possible conflict arising when first implementing a new technology. Rosell-Aguilar examined tutors' perceptions using *Lyceum* with beginners' Spanish courses and found that the tutors thought the tool provided good opportunities for speaking the language. This result is in contrast to prior internal surveys by Greenberg (2003) and Coleman (2003) that report negative perceptions (in Rosell-Aguilar, 2006). Rosell-Aguilar explains that this is

because actual experience in using the tool could gradually change people's minds. Vetter andChanier (2006) used *Lyceum* in English for Specific Purposes (ESP) with French false-beginner learners of varying proficiencies (similar to this chapter's research context). They found that *Lyceum* was effective for ensuring students' equal participation and that the combination of voice and text modalities helped students regain self-confidence in speaking. Many of the papers which used *Lyceum* noted the communication difficulties arising from its lack of paralinguistic visual clues (such as facial expressions and body language). This may be the reason for the decision to shift from *Lyceum* to *Elluminate,* which also provides video transmission.

VoiceThread.com offers a shared space to learn ideas about how to effectively use *VoiceThread* in teaching practice (http://voicethread.com/community/library/). EDUCAUSE summarized the potential of *VoiceThreads* in early 2009 (EDUCAUSE, 2009). The work of Burden and Atkinson (2008) is of particular value because it provides a comprehensive framework for the "affordances of *VoiceThread*" based on a study by McLoughlin and Lee (2007). Oughton and Rogers (2010) reported their experience with implementing *VoiceThread* in videotaped role-play sessions to demonstrate counseling skills in a counseling course, using video sharing with peer feedback (text-based, if the users did not have a microphone). The study acknowledged the usefulness of video sharing among students as well as its merits for creating stimulating online learning environments, although some of the students had difficulty using the technology. Augustsson (2010) implemented *VoiceThread* in a blended-mode social psychology class to provide flexibility and transparency in students' participation and to foster collaborative reflection. McCormack (2010) analyzes voice threads from the reflective learning perspective, using miscue analysis and semi-structured interview. He found the tool fostered higher engagement and reflection. Using a content analysis method for 50 randomly chosen, publicly accessible voice threads, Millard (2010) examined the benefits of *VoiceThread* in socio-constructivist teaching and learning but concluded that there was little evidence to support interaction among users. However, these results may have changed if the data had been collected within a formal course or especially within an LMS community created specifically for educational purposes, as demonstrated in Figure 5.

Research studies using the Moodle voice modules with *NanoGong* are also available, though some are published in Chinese (because the module was developed by a Hong Kong project team) and therefore beyond the authors' reach. Kumai and Daniels (2010) developed and tested voice-shadowing modules using *NanoGong* on *Moodle*. Shadowing is "a spoken language training method that requires learners to repeat or shadow a presented native utterance as quickly and closely as possible" (Luo, Yamauchi, and Minematsu, 2010, no page number). It is widely practiced in language education. Kumai and Daniels tested the system with sophomore English classes. A post-course questionnaire found that the students enjoyed the shadowing activities, were more motivated to study English, and had higher confidence in their ability to understand English audio data. During the spring semester of 2010, Miyazoe tested a voice-recording module that is under development by Daniels (2011). This testing provided a pilot study that is presented in the next section of this chapter.

PILOT STUDY

A pilot study was undertaken to detect effects on students' learning of an aural and oral intervention in language learning. The pilot study had two objectives: 1) testing the usability of the system that was newly implemented into *Moodle* and 2) testing the practicability of a blended course design with online voice assignments.

An add-on Moodle voice module (a free download from http://moodlemodules. netcourse.org/) *called Voice Record* was chosen for the experiment (Figure 10). This module allowed students to record their voices as audio files and submit them to the course instructor for grading through the LMS. In addition to *Moodle*, students needed the free downloadable Flash player (http://www.adobe.com/) and a microphone for recording.
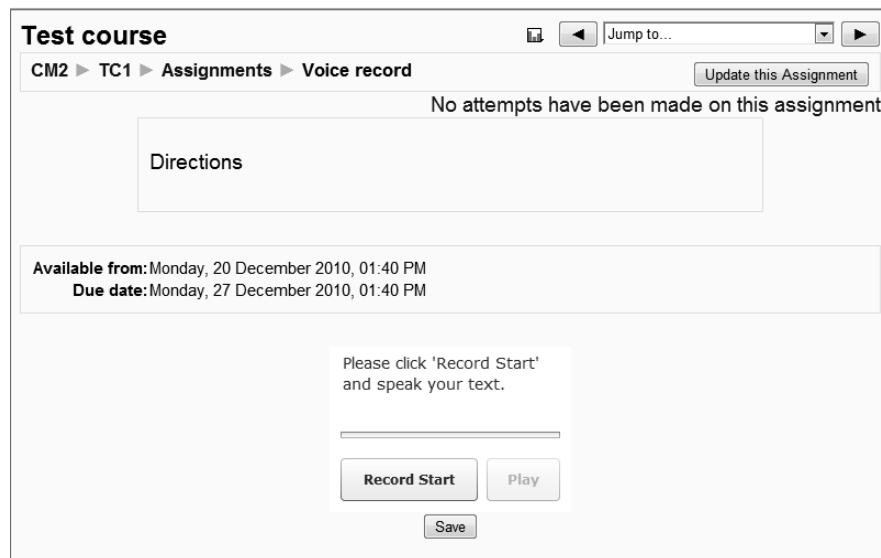

Figure 10. Voice Record Module on Moodle.

In principle, the study found that the voice system was adequately usable and practicable for teaching. However, some students needed technical support to learn how to use the system. The study also showed that the students did not show strong reluctance to "talk to a computer," and no significantly negative effects on their learning process due to the fact that the online voice assignments were observed. Because of this, further research using the *Voice Record* module on *Moodle* was considered worth pursuing.

PURPOSE lOF THE MAIN STUDY

Based on the implications of the literature review and the pilot study, the current research sought to answer specific questions about the functionality of the online voice system and its learning outcomes. Specifically, the study aimed to assess the following: 1) the voice recording system's usability, 2) the students' anxiety about speaking to the online voice program, 3) the effects of online voice assignments on attendance in face-to-face classes in

the blended course design, 4) the use of different types of online voice activities for speaking improvement, and 5) changes in the students' perceptions of speaking English associated with using the voice system.

It worth noting that in the context of this chapter's research, that is, a context involving beginner English students in which English is not spoken (that is, English as foreign language context), provides a strong environment to examine the effectiveness of an online voice system because the students are more likely to form a naturally separated group with little English exposure than those who study in an English-speaking context (that is, English as second language context).

METHODS

The Participants

Of 100 students registered for a required oral English course in a Japanese university, 86 (82 male and 4 female) completed the pre-course survey and 65 (57 male, 2 female, 6 no answer) completed the post-course survey. Fifty students per class was the maximum enrollment allowed, so the initial enrollment reached the maximum registration capacity. This course was specifically for students repeating the course, which should have been completed during their freshman year; therefore, the students were sophomores, juniors, and seniors. The oral English course is usually taught by a native English speaker; however, a Japanese instructor was assigned to the repeater classes to avoid disadvantaging students who may have experienced strong speaking anxiety during English-only instruction.

All the students majored in sciences, such as engineering, information, and technology. As a school requirement, all of them possessed a portable notebook, and nearly all of them had internet access at their domicile (a few students who lived in apartments reported that they did not have an internet connection). In addition, they had access to school computer rooms anytime outside of classes.

Online Voice Assignment Design in the Blended Course

The course objective was to acquire basic oral English for daily conversation and life management in English-speaking contexts. Each course met 13 to 15 times in a semester. Including national holidays, one semester usually lasts approximately 18 weeks.

The course adopted a blended design consisting of weekly face-to-face meetings and online voice assignments. The voice assignments were prepared on *Moodle* with the free add-on *Voice Record* module installed. With written permission from the publisher, model audio files recorded by native speakers and related to the course textbook were uploaded to the Moodle course. Students were required to listen to the model, practice, and record their voices to the system for grading. The students could make multiple files for each assignment (and could delete their own files) to practice more than once.

Improvements were made to the voice assignment structure based on the results of the pilot study. The same *Voice Record* format was used to keep the format of the intervention

constant; however, assignments offered more variation and required two formats of recording. In addition to the reading aloud of short text segments, as in the pilot, a brief monologue speech on a provided topic relevant to class content was added. Both were chosen from the class activities of the week and served as a review of the face-to-face class. Furthermore, to facilitate voice assignment completion, five microphones were made available in one of the school computer rooms, and a demonstration session showing how to use the voice recording system was set up at the beginning of the semester to familiarize students with the system.

The rationale for using listen-and-record assignments for the students' developmental stages in this study comes from the major hypotheses in language acquisition studies: 1) a *comprehensible input* step (Krashen, 1985) that involves listening to an audio model recorded by native speakers with a content level that is manageable and appropriately challenging to the students; 2) a *pushed output* step (Schmidt, 1983; Swain, 1985) that involves the forced production of reading aloud or mini-speeches at a level manageable and appropriately challenging to the students; 3) through these process, the students increase their awareness of the gap between the model and their own production; that is, they experience a *noticing* process by which they become aware of their own errors and modify them (Schmidt, 1990). To address possible speaking anxiety, the self-study mode, with recordings submitted only to the instructor, was used as a *confidence-building* process (Williams and Andrade, 2008) to help lower the *affective filter* (Krashen, 1982), allowing students to process more linguistic corpora. Because this course was blended, further collaboration and interaction among students was planned for the face-to-face class meetings.

Instrumentation

The system and course evaluations were measured by pre- and post-course online questionnaires (Appendices A and B). Both times, the survey included the same question regarding the students' perceptions of speaking English to determine any changes resulting from the course experience. A System Usability Scale (SUS) (Brooke, 1996; Tullis and Albert, 2008) consisting of ten short questions was used to measure the Voice Record system's usability. Additionally, specific questions regarding demographics and the utility of the voice assignments were posed for further correlational analysis.

Data were analyzed using SPSS and EXCEL software. For the SUS usability score, a five-point scale, which ranged from 1 to 5, was converted to a 0 to 4 scale. For positively keyed odd number items (1, 3, 5, 7, and 9), the score contribution was the scale position minus 1. For negatively keyed even number items (2, 4, 6, 8, 10), the score contribution was 5 minus the scale position. Therefore, the SUS score ranged from 0 to 100, with the item score contribution at 2.5 (Brooke, 1996). SPSS Text Analytics for Surveys was used to analyze students' answers in Japanese to the survey's open-ended question regarding English writing. The analysis procedure followed the grounded theory qualitative analysis methodology (Strauss and Corbin, 1990) to reduce risks of analyzer subjectivity (Miyazoe and Anderson, 2010b).

Procedures

The pre- and post-course questionnaires were distributed as web links using the online survey tool (http://www.surveymonkey.com/) linked to the Moodle course. The students were also asked to cooperate with the teacher who is also the first researcher in this study, orally during face-to-face classes. The questionnaires were submitted anonymously, and students who read the research objectives and agreed to the possible publication of the results were prompted to answer the questions. Both the pre- and post-course surveys were available for two weeks during the semester.

RESULTS AND DISCUSSION

Results that produced insights for further research and practice are selectively reported. Where the data distribution is similar to a normal distribution, the mean scores are considered representative and the numeric information is reported in tables; however, when the distribution takes an irregular form and a different interpretation is useful, a graphic representation is also provided.

Pre-Course Perceptions towards Speaking Online

Table 1 summarizes the results of the pre-course survey to highlight the students' perceptions of speaking English in comparison to other parameters. In the table "hate" is scored 1, "dislike" is scored 2, "neither" is scored 3, "like" is scored 4, and "love" is scored 5. When comparing the four skills of reading, listening, speaking, and writing English, students reported the least positive feelings toward speaking (2.50 points average), suggesting the presence of speaking anxiety. This result is understandable, given that the class was an oral course specifically designed for repeaters. It is possible that students' dislike of speaking English was linked to the fact that they failed the course before and lost the confidence that they could manage speaking skills in the same way as other skills. Either way, their reported dislike of speaking English is consistent with the other results of the students' pre-course perceptions of speaking English, which are further examined below.

Table 1. Perceptions towards Different English Skills and PC/Internet Use

	N	Minimum	Maximum	Mean	Std. Deviation
Reading	86	1	4	2.71	.852
Listening	86	1	5	2.65	.943
Speaking	86	1	5	2.50	.979
Writing	86	1	5	2.71	.879
PC and internet	86	1	5	3.86	.972

Figure 11 highlights the learner characteristics in this study by comparing the results of the students' attitudes toward speaking English and toward PC and internet usage. The contrast between their like of PC/internet use (light gray bars) and their dislike of speaking English (dark gray bars) is significant. This confirms both the observations from the pilot study and the experiences of teachers of other courses in the same university. Furthermore, it underpins the rationale of the research concept behind the effectiveness of implementing ICT in oral classes; students' enjoyment of ICT may help reduce their speaking anxiety, leading to a higher completion rate for this oral English course.
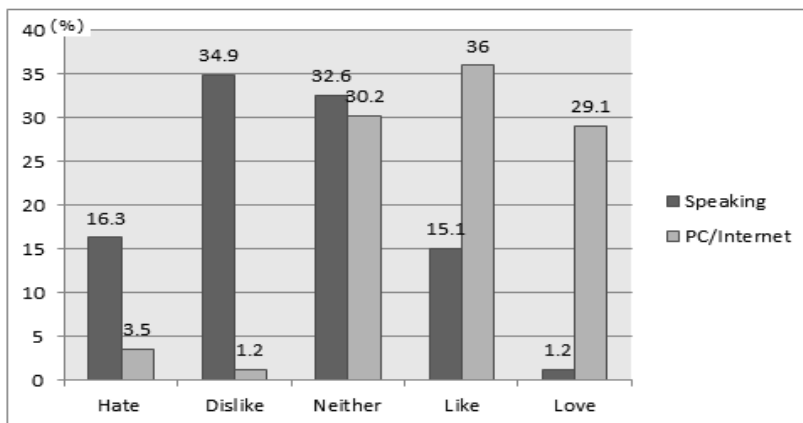

Figure 11. Students' perceptions towards speaking English and PC/Internet use.

Voice Recording System Usability

Tables 2 and 3 summarize the usability results for the voice recording system. With 127 students from two semesters of pilot and main studies, the mean usability was 54 out of a possible 100, meaning that the system was fairly easy to use. However, the wide difference between minimum and maximum scores and the relatively large standard deviation signifies that students' perceptions varied widely. Some students felt that using the system was easy, but other students found it difficult to use. It is notable that from Fall 2009 (the pilot study) to Spring 2010 (the main study), the mean SUS score improved by 5.5 points. It is reasonable to think that more effective demonstration guidance and possibly increased peer support helped the students learn the system better. It is also possible that the venue change, from a traditional classroom to the computer room, fostered psychological proximity and interest to do the voice assignments.

Table 2. Usability of Voice Recording System (SUS Results)

SUS total scores	N	Minimum	Maximum	Mean	Std. Deviation
Fall 2009	62	27.5	100.0	51.57	13.620
Spring 2010	65	27.5	87.5	57.00	13.522
Total	127	27.5	100.0	54.35	13.788

Table 3. Usability of Voice Recording System (Specific Features)

	N	Minimum	Maximum	Mean	Std. Deviation
Sound quality	123	1	5	3.09	1.064
Waiting time	123	1	5	3.31	1.087
PC reluctance	123	1	5	3.30	1.286

Table 3 summarizes the results of three specific features offered by the voice system developer. In the table "strongly disagree" is scored 1, "disagree" is scored 2, "neutral" is scored 3, "agree" is scored 4, and "strongly agree" is scored 5 to the given question statements. With a system that required connection to a Flash server, the waiting time for the sound quality and the connection seemed acceptable. The results are linked to the technology environments where each student completed the voice assignment and are difficult to generalize. However, there may be room to improve the system further, especially when the system is used for speaking practice.

Figure 12 summarizes the responses to the statement "I felt no reluctance to talk to the computer." Unlike the majority of other response results, which took an approximately normal distribution shape, 41.3% (17.5% + 23.8%) of the students felt strong or moderate levels of reluctance to talk to a PC, whereas 47.6 % (28.6% + 19%) felt no strong reluctance. That is, the number of students with negative feelings about having to talk to a computer was almost equal to the number that felt positive about it. This resulted in an average score of 3.30, as shown in Table 3.
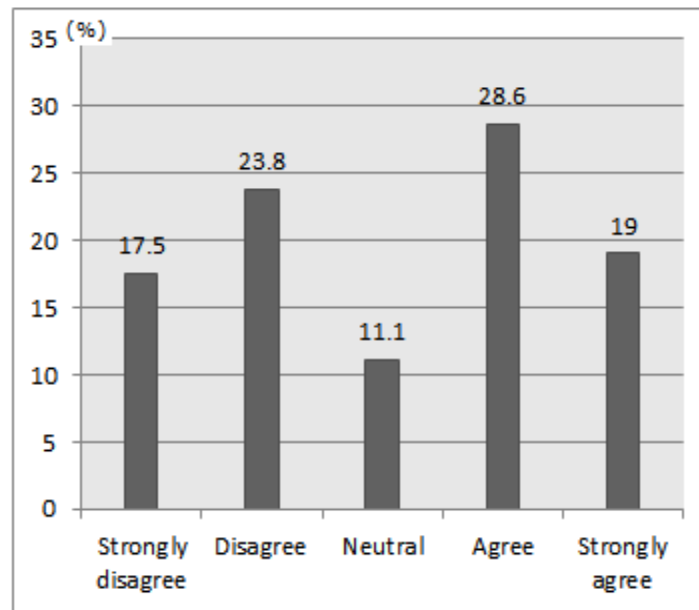

Figure 12. Reluctance to talk to a PC.

Preferences of Voice Recording Formats

Figures 13 and 14 summarize the results of questions regarding the preferred voice assignment formats and students' ideal format balance for the course design. Figure 13 shows that more than 60% of the students preferred the reading text aloud format to the topic-based short monologue speech. Figure 14 shows that 50.8% (38.1% + 12.7%) of them preferred reading text aloud, compared with the 11.1% who preferred the monologue format.
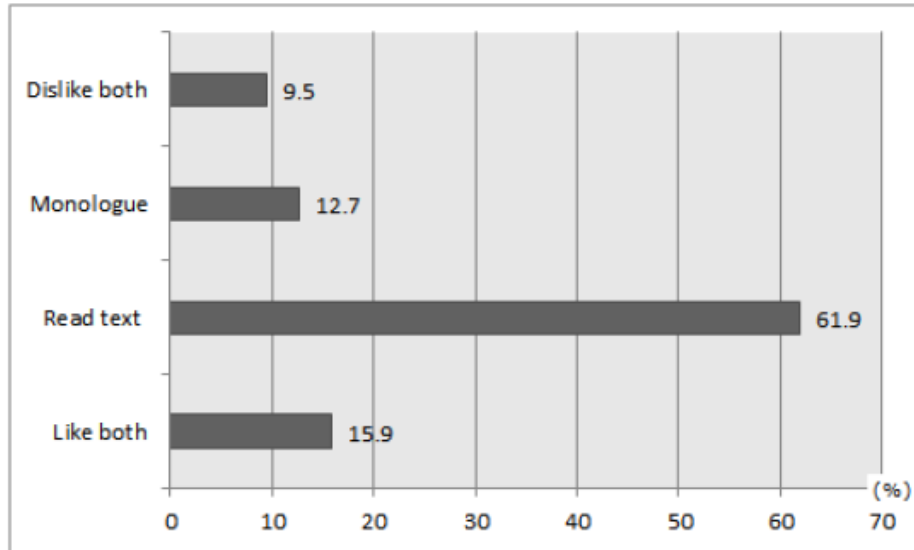

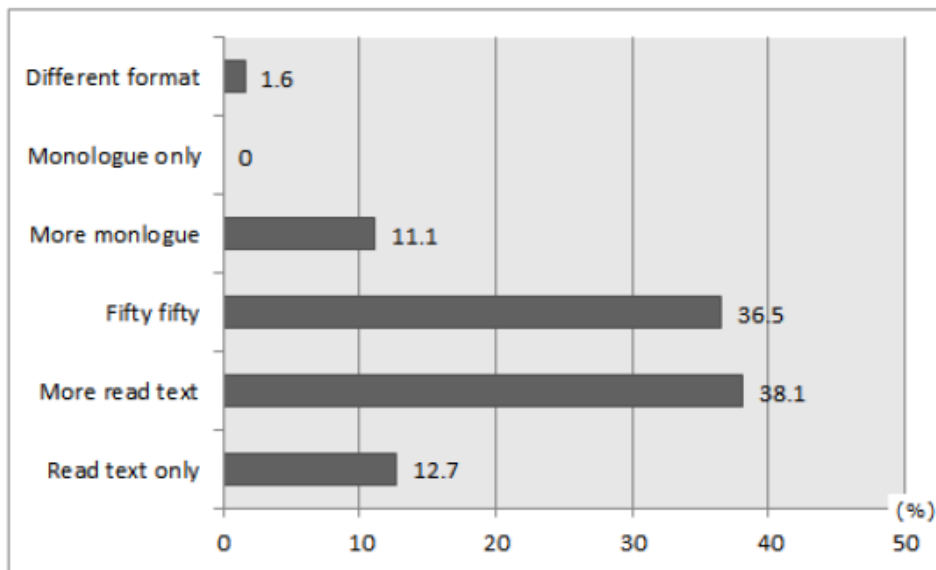Figure 13. Students' preferred voice recording formats.

Figure 14. Students' preferred voice recording format balance.

These results are understandable given that the study participants were not yet capable of speaking English. For many of them, reading text aloud would be more comfortable than having to create new speech on a provided topic. However, it is encouraging that 47.6% (11.1% + 36.5%) of the students accepted the more challenging speech format for the benefit of their learning rather than taking the easier road of reading text aloud.

Utility of Voice Assignments

Table 4 summarizes the students' responses regarding the utility of voice assignments. The students reported that voice assignments offered a fair level of utility to improve their oral skills. In the table "strongly disagree" is scored 1, "disagree" is scored 2, "neutral" is scored 3, "agree" is scored 4, and "strongly agree" is scored 5, to the given question statements. Although the mean score for the reading text aloud format was slightly higher than for monologue speech, a paired-sample t-test showed that the responses were not significantly different. The statement regarding motivation specifically asked whether the students felt that the voice assignments helped them continue the course. This question was asked because the voice assignments were designed to take advantage of students' positive attitudes toward ICT and to use it to motivate students to keep coming to class. Though the mean score of 3.27 for this item is not impressive, further analysis with other factors in the following section revealed responses closer to those we expected.

Table 4. Utility of Voice Assignments

	N	Minimum	Maximum	Mean	Std. Deviation
Read text	63	1	5	3.35	1.034
Monologue	63	1	5	3.29	1.054
Motivation	63	1	5	3.27	1.066

Relationship between Physical and Online Participation

Table 5 summarizes the results of a correlation analysis of voice activity utility, physical/online participation, and system usability. Only the results that showed high correlations are mentioned here.

A significantly high correlation was observed between the belief that the assignments improved English speaking and the utility of read-aloud text and monologue speech (r = .793, p < .01). This means that students who thought one format was useful tended to think the other format was also useful. This is understandable because both were introduced to provide more opportunities to practice speaking, with monologue assignments offering greater challenges than read-aloud text.

A significantly high correlation was observed between physical class attendance and completing online voice assignments (r = .642, p < .01). This means that those who came to face-to-face class meetings tended to complete online assignments also. This positive cycle has been confirmed many times in the authors' previous studies (Miyazoe and Anderson,

2010b). Therefore, in a blended learning course, it is worthwhile to encourage students to either come to class or complete online assignments; one study mode triggers the other to improve study performance in the end.

A high correlation was observed between students' reports of the usefulness of both voice assignment types to improve English speaking and the students' inclination to complete the course (r = .525, p < .01; r = .550, p < .01). This means that although the students were understandably uncertain about the extent to which the online voice assignments actually improved their speaking ability (because no oral English proficiency test was included in the research design due to the large number of constraints), they did feel that voice assignments were useful and motivated them to continue the course.

Finally, a moderately high correlation was observed between the voice recording system usability scores and the frequency of completing voice assignments (r = .323, p < .05). This means that students may have been motivated to complete the voice assignments by factors other than the voice system's ease of use (such as awareness of the assignments' utility or motivation to keep coming to the classes). However, because a moderately high correlation was observed, it is possible that greater usability would encourage even more students to complete the assignments and, consequently be associated with higher attendance at classes and completion of the course.

Table 5. Correlations among Voice Activity Utility, Physical/Online Participation, and System Usability

Spearman's rho		Monologue	Motivation	Attendance	Assignment	SUS total
Read text	Correlation Coefficient	.793**	.525**	.481**	.447**	.391**
	Sig. (2-tailed)	.000	.000	.000	.000	.002
	N	63	63	59	59	63
Monologue	Correlation Coefficient		.550**	.489**	.560**	.411**
	Sig. (2-tailed)		.000	.000	.000	.001
	N		63	59	59	63
Motivation	Correlation Coefficient			.231	.323*	.251*
	Sig. (2-tailed)			.078	.013	.048
	N			59	59	63
Attendance	Correlation Coefficient				.642**	.238
	Sig. (2-tailed)				.000	.069
	N				59	59
Assignment	Correlation Coefficient					.323*
	Sig. (2-tailed)					.012
	N					59

** Correlation is significant at the 0.01 level (2-tailed).
* Correlation is significant at the 0.05 level (2-tailed).

Change in Students' Perceptions towards Speaking English

Figures 15 to 18 summarize the pre-/post-course survey results for the text analysis of the open-ended question regarding the students' perceptions towards speaking English. The question took the format, "How do you feel about speaking English? Answer as thoroughly as you can." Figures 15 and 16 show the network representation linking the core concept categories. Figures 17 and 18 provide bar graph representations of the numerical frequency with which these concept categories appear in the text data. In the network presentations, the circle sizes correspond to the number of students who mentioned the concepts in their responses, and the weights of lines linking the concept nodes represent the number of students whose responses reflected associations between the concepts. In the bar graphs, the percentages signify the percentage frequency of each concept category's appearance, using the concept of English as 100%. That is, among all the students who mentioned the concept of *English*, 73.7% (Figure 17) associated it with the concept of *difficulty*. In this way, we eliminated the responses that did not directly answer the question and gained a more accurate profile of the percentage of students that felt and shared the given concept category. Both Figures 15 and 16 were generated with *speaking* as their node center to make the pre- and post-network presentations comparable.
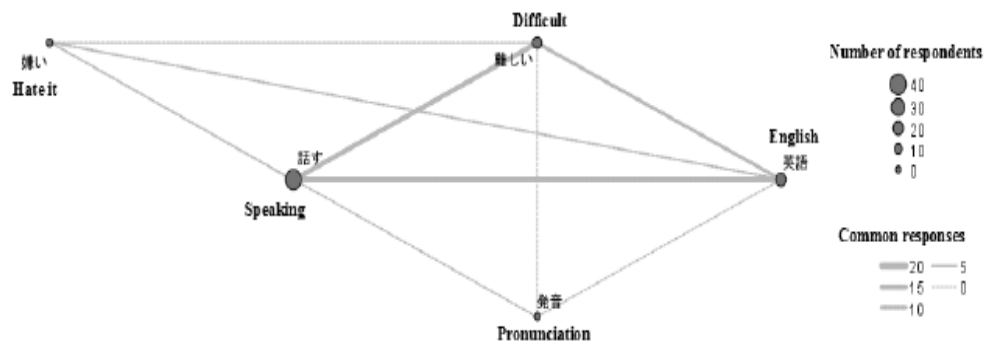

Figure 15. Pre-course perceptions about speaking English.
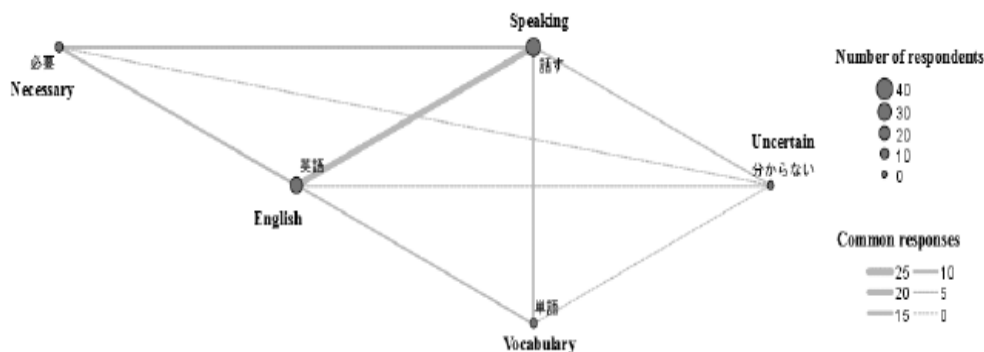

Figure 16. Post-course perceptions about speaking English.

Figure 15 shows the representative concept categories at the beginning of the course semester. Speaking English is associated with difficulty, and six students out of 86 explicitly used the term *hate* to express their feelings about speaking English. Additionally, eight students expressed their difficulty with producing appropriate pronunciation. In all, positive ideas such as fun and usefulness were mentioned by only one or two students; therefore, at this stage, these ideas were considered exceptional and were not included in the pre-course concept network.

Figure 16 shows the representative concept categories at the end of the course. The concept category of *uncertain* referred to their uncertainty about understanding what their interlocutors were saying; six students out of 65 expressed this idea. Fifteen students associated the concept of *necessity* with speaking English, and 12 students stated that increasing their *vocabulary* could improve their speaking ability.
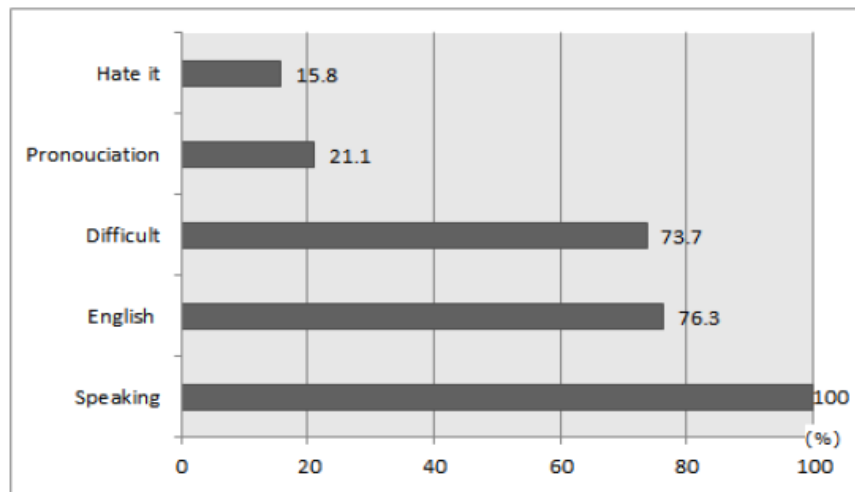

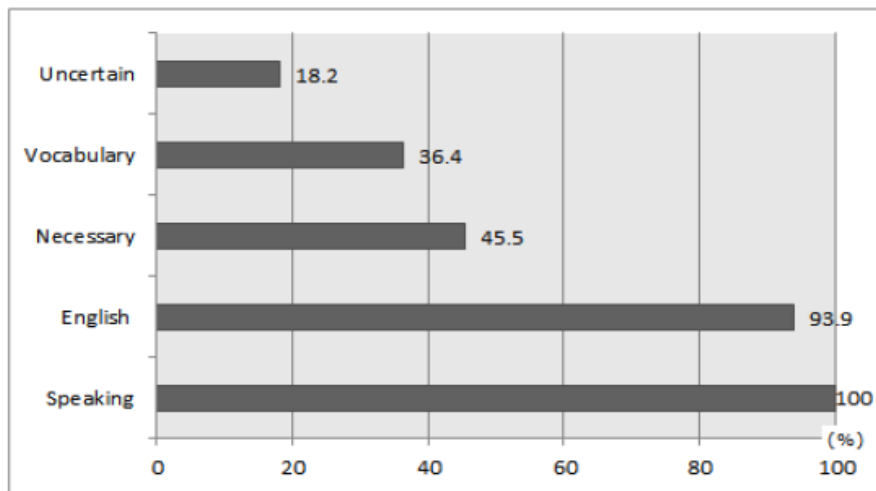Figure 17. Word frequencies in the pre-course survey data.

Figure 18. Word frequencies in the post-course survey data.

Overall, from the pre-course to post-course surveys, the students' attitudes towards speaking English seem to have changed from negative to more positive and more analytical. At the pre-course stage, negative attitudes, such as difficulty, hatred, and worry about their pronunciation, all resonated with the speaking anxiety syndrome discussed in the above section. On the other hand, at the post-course stage, they admitted the necessity of mastering spoken English and felt they should be able to express their thoughts and understand their interlocutors; for these purposes, they needed to acquire more vocabulary.

FURTHER RESEARCH AND RECOMMENDATIONS

This study aimed to determine 1) the voice recording system usability; 2) the students' speaking anxiety levels related to the online voice program; 3) the effects of online voice assignments on attendance in face-to-face classes in a course with a blended design; 4) the utility of different online voice activity types for speaking improvement; and 5) students' changes in perceptions towards speaking English and the possible learning outcomes.

This study found that the voice system offered sufficient usability for teaching practice. The study found a mixed reaction to the human-interface interaction; that is, half of the students were reluctant to talk to an online program, whereas half of them accepted it. In a blended course design, the regular completion of online voice assignments likely created a positive cycle that encouraged high attendance in the physical classes. Though the students preferred the simple reading-aloud activities to the more challenging monologue speech activities, they seemed to manage both formats with no problems. The students' perceptions of speaking English changed positively, becoming more analytical over the semester. Overall, the study concludes that asynchronous online voice activity is helpful and effective for supporting students' learning.

For future studies regarding online voice interaction, several points should be amended as next steps. Because the voice system showed adequate practicality levels and the instructor became more confident about using the system during its third phase (currently underway during Fall 2010, as this chapter was being written), the instructional design should go one step further and allow the students to share their own voice submissions among classmates. This adds to the *student-content* and *student-teacher* interaction *student-student* elements. Concurrently, the voice activity design included an additional strategic approach to language acquisition (the shadowing approach) to gradually invite the students to elaborate and induce a higher learning outcome.

In the study's fourth stage (assuming that the third stage is successful with few negative effects), we plan to introduce pre-/post-course measurement of the change in students' speaking ability over the semester in 2011. This process needs to establish a nonnegligible research budget, so it will be delayed to ensure that the implementation of online voice interaction will produce meaningful results in students' learning.

Finally, this study recommends an attitude of reflective research and practice cycle, or design-based research (Anderson, 2005; Brown, 1992; Collins, 1992), when implementing new technology to produce higher learning outcomes, as illustrated in the longitudinal process of examining the educational effectiveness of *Lyceum* implementation at Open University UK. This should start with an objective measurement of the system's usability, instructional

design consideration to determine whether the system is meaningful for teaching, a meaningful intervention designed with the practitioners in a real educational context and a learning effects evaluation to determine whether the system produces any positive learning results.

CONCLUSION

This study is significant in the following ways: 1) it provides frameworks to clarify the instructional design factors that must be considered when implementing online voice interaction; 2) it provides a review of the most current online voice interaction research at the time of writing; and 3) it provides its own empirical research to support the effectiveness of asynchronous online voice interaction integrated into an LMS.

The field of online learning is growing, with the paradox that new technologies appear, are adapted, and often disappear quickly. A strong emphasis on research accompanied by constant improvement of systems and revisions of their use in instructional design is necessary to progress beyond the continuous waves of new technologies.

ACKNOWLEDGMENT

BIOGRAPHICAL NOTES

Terumi Miyazoe is an associate professor of e-learning and English at Tokyo Denki University in Japan. She is also an assistant professor of e-learning, learning design, and distance education at Open University Japan. She earned an MA in TESOL (UK) and an MA in Distance Education (Canada) through online distance learning courses taken from Japan. Her research interests include online learning, semiotics, and foreign language education. She holds a PhD in Educational Technology.

Website: http://members3.jcom.home.ne.jp/t.miyazoe/.

Terry Anderson is a professor and the Canadian Research Chair in Distance Education at Athabasca University, Canada's leading online and distance open university. He teaches educational technology courses in Athabasca University's master's and doctorate programs in Distance Education. He has been the research mentor of Terumi Miyazoe since 2005.

Website: http://cde.athabascau.ca/faculty/terrya.php and http://terrya.edublogs.org/.

APPENDIX A

Pre-Course Questionnaire (Extract)

Aim: To determine students' perceptions towards speaking English, in contrast to other English skills and factors

Question Group 1: Demographics (gender and age)

Question Group 2: Perceptions towards speaking English

Q 1: How do you feel about speaking English? Answer as thoroughly as you can.

Q 2: Choose one that best describes your feeling towards English.

Reading: Hate it – Dislike it – Neutral – Like it – Love it

Listening: Hate it – Dislike it – Neutral – Like it – Love it

Speaking: Hate it – Dislike it – Neutral – Like it – Love it

Writing: Hate it – Dislike it – Neutral – Like it – Love it

Question Group 3: Perceptions towards PC and internet use

Q: Choose one that best describes your feeling towards PC and internet use.

Hate it – Dislike it – Neutral – Like it – Love it

APPENDIX B

Post-Course Questionnaire (Extract)

Aim: To assess voice recording system usability, the utility of voice recording assignments and a blended course design, and perceptions towards speaking English

Question Group 1: Demographics (gender and age)

Question Group 2: Usability of the voice recording system

System Usability Scale (SUS) questions*

SUS 1. I think that I would like to use this system frequently.

SUS 2. I found the system unnecessarily complex.

SUS 3. I thought the system was easy to use.

SUS 4. I think I would need the support of a technical person to be able to use this system.

SUS 5. I found the various functions in this system were well integrated.

SUS 6. I found this system was too inconsistent.

SUS 7. I would imagine that most people would learn to use this system very quickly.

SUS 8. I found the system very cumbersome to use.

SUS 9. I felt very confident using the system.

SUS 10. I needed to learn a lot of things before I could get going with this system.

* (Tullis and Albert, 2008, pp. 138-139)

D 1. The quality of the recorded voice was good enough.

D 2. The time needed to wait for the audio file to be recorded/played was appropriate.

D 3. I felt no reluctance to talk to the computer.

Question Group 3: Utility of voice recording assignments

Q 1: In this course, two formats of voice recording assignments (reading text aloud and monologue speech) were used. Choose one response that best describes your feeling.

Like both formats

Like reading text aloud more

Like monologue speech more

Dislike both formats

Q 2: Choose one that best describes your feeling about the following statement: Reading aloud assignments helped to improve my speaking ability.

Strongly disagree – Disagree – Neutral – Agree – Strongly agree

Q 3: Choose one that best describes your feeling about the following statement: Monologue speech assignments helped to improve my speaking ability.

Strongly disagree – Disagree – Neutral – Agree – Strongly agree

Q 4: Choose one that best describes your feeling about the following statement: Voice recording assignments motivated me to keep studying in this course.

Strongly disagree – Disagree – Neutral – Agree – Strongly agree

Q 5: Choose the best balance in frequency between reading text aloud and monologue speech assignments.

Only reading text aloud

More reading text aloud than monologue speech

Half reading text aloud, half monologue speech

More monologue speech than reading text aloud

Only monologue speech

Formats other than reading text aloud and monologue speech (please specify:)

Question Group 4: Perceptions towards speaking English

Q: How do you feel about speaking English? Answer as thoroughly as you can.

Question Group 5: Frequency of attending face-to-face classes and doing online assignments

Q 1: Choose one that best describes your feeling about the following statement: I attended almost all the face-to-face classes (consider that seven times is the midpoint of a total of fourteen meetings).

Strongly disagree – Disagree – Neutral – Agree – Strongly agree

Q 2: Choose one that best describes your feeling about the following statement: I did almost all the voice recording assignments (consider that five times is the middle of a total of ten assignments).

Strongly disagree – Disagree – Neutral – Agree – Strongly agree

REFERENCES

Anderson, T., and Garrison, R. (1998). Learning in a networked world: New roles and responsibilities. In C. Gibson (Ed.), *Distance learners in higher education* (pp. 97-112). Madison, WI: Atwood Publishing.

Anderson, T. (2005). Design-based research and its application to a call centre innovation in distance education. *Canadian Journal of Learning and Technology, 31*(2). Retrieved from http://www.cjlt.ca/index.php/cjlt/article/viewArticle/143/136.

Augustsson, G. (2010). Web 2.0, pedagogical support for reflexive and emotional social interaction among swedish students. *The Internet and Higher Education, 13*(4), 197-205.

Brooke, J. (1996). SUS: A quick and dirty usability scale. In P. W. Jordan, B. Thomas, B. A. Weerdmeester and I. L. McClelland (Eds.), *Usability evaluation in industry* (no page number). London: Taylor and Francis.

Brown, A. L. (1992). Design experiments: Theoretical and methodological challenges in creating complex interventions in classroom settings. *Journal of the Learning Sciences, 2*(2), 141-178.

Burden, K., and Atkinson, S. (2008). Evaluating pedagogical affordances of media sharing web 2.0 technologies: A case study. Paper presented at Ascilite 2008, Melbourne. Retrieved from *www.ascilite.org.au/conferences/melbourne08/.../burden-2.pdf.*

Collins, A. (1992). Towards a design science of education. In E. Scanlon, and T. O'Shea (Eds.), *New directions in educational technology* (pp. 15-22). Berlin: Springer.

Cutrone, P. (2009). Overcoming japanese EFL learners' fear of speaking. *University of Reading: Language Studies Working Papers, 1*, 55-63.

Daniels, P. (2011). In Miyazoe T. (Ed.), *Moodle for language education*. Tokyo: Kaibundo. (In press).

de los Arcos, B., Coleman, J. A., and Hampel, R. (2009). Learners' anxiety in audiographic conferences: A discursive psychology approach to emotion talk. *ReCALL, 21*(1), 3-17.

Driscoll, P. M. (2005). *Psychology of learning for instruction* (3rd ed.). Boston: Pearson Education, Inc.

EDUCAUSE. (2009). *7 things you should know about...VoiceThread.* Retrieved from http://net.educause.edu/ir/library/pdf/ELI7050.pdf

Ertmer, P., and Newby, T. J. (1993). Behaviorism, cognitivism, constructivism: Comparing critical features from an instructional design perspective. *Performance Improvement Quarterly, 6*(4), 50-71.

Garrison, D. R., and Anderson, T. (2003). *E-learing in the 21st century: A framework for research and practice*. New York: RoutledgeFalmer.

Hampel, R. (2003). Theoretical perspectives and new practices in audio-graphic conferencing for language learning. *ReCALL, 15*(1), 21-36.

Hampel, R., Felix, U., Hauck, M., and Coleman, J. A. (2005). Complexities of learning and teaching languages in a real-time audiographic environment. *German as a Foreign Language, 3*. Retrieved from http://www.gfl-journal.de/3-2005/hampel_felix_hauck_coleman.html.

Hauck, M., and Hurd, S. (2005). Exploring the link between language anxiety and learner self-management in open language learning contexts. *European Journal of Open, Distance and E-Learning.* Retrieved from http://www.eurodl.org/materials/contrib/2005/Mirjam_Hauck.htm.

Kearsley, G. (2010). *Explorations in learning and instruction: The theory into practice database.* Retrieved from http://tip.psychology.org/.

Kenning, M. (2010). Differences that make the difference: A study of functionalities in synchronous CMC. *ReCALL, 22*(1), 3-19.

Krashen, S. D. (1982). *Principles and practice in second language acquisition*. Oxford: Pergamon.

Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. Longman: New York.

Kumai, N., and Daniels, P. (2010). Development of Moodle voice record and play modules and its application to shadowing practice. *Computer Center Annual Report, 31*, 73-89.

Lamy, M., and Hampel, R. (2007). *Online communication in language learning and teaching.* New York: Palgrave Macmillan.

Luo, D., Yamauchi, Y., and Minematsu, N. (2010). Speech analysis for automatic evaluation of shadowing. Paper presented at the *Second Language Studies: Acquisition, Learning, Education and Technology (CD-ROM).* Retrieved from http://www.gavo.t.u-tokyo.ac.jp/L2WS2010/.

McCormack, V. (2010). Increasing teacher candidate responses through the application of VoiceThread. *International Journal of Arts and Sciences, 3*(11), 160-165.

McLoughlin, C., and Lee, M. J. W. (2007). Social software and participatory learning: Pedagogical choices with technology affordances in the web 2.0 era. Paper presented at Ascilite 2007, Singapore. Retrieved from *www.ascilite.org.au/conferences/singapore07/.../mcloughlin.pdf.*

Millard, M. O. (2010). Analysis of interaction in an asynchronous CMC environment. In *Proceedings of the WebSci10: Extending the Frontiers of Society on-Line,* Raleigh, NC: USA. Retrieved from http://journal.webscience.org/391/.

Miyazoe, T. (2008). Objectivism vs. constructivism re-visited: An interpretation in online and distance learning environments. *Educational Studies, 50*, 189-200. Available at http://ci.nii.ac.jp/naid/110007324650.

Miyazoe, T., and Anderson, T. (2010a). Empirical research on learners' perceptions: Interaction equivalency theorem in blended learning. *European Journal of Open, Distance and E-Learning.* Available at http://www.eurodl.org/.

Miyazoe, T., and Anderson, T. (2010b). Learning outcomes and students' perceptions of online writing: Simultaneous implementation of a forum, blog, and wiki in an EFL blended learning setting. *System, 38*(2), 94-104. Accessible at http://www.sciencedirect.com/.

Miyazoe, T., and Anderson, T. (2011). Anonymity in Blended Learning: Who Would You Like to Be?. *Journal of Educational Technology and Society, in press*. Available at http://www.ifets.info/.

Mochizuki, T. (2010). Learning through discussions: Computer-supported collaborative learning. In Y. Yamauchi (Ed.), *Pedagogy of digital materials* (pp. 41-58). Tokyo: University of Tokyo Press.

Moore, M. G. (1993). Theory of transactional distance. In D. Keegan (Ed.), *Theoretical principles of distance education* (pp. 22-38). London and New York: Routledge.

Moore, M. G. (1989). Editorial: Three types of interaction. *The American Journal of Distance Education, 3*(2), 1-7.

Moore, G. M., and Anderson, G. W. (Eds.). (2003). *Handbook of distance education*. New Jersey: Lawrence Erlbaum Associations, Inc.

Nishimori, T. (2010). Multimedia materials that makes learning context. In Y. Yamauchi (Ed.), *Pedagogy of digital materials* (pp. 25-39). Tokyo: University of Tokyo Press.

Oughton, J. M., and Rogers, M. (2010). Using VoiceThread for counseling role-playing sessions. Paper presented at *the 17th Annual Sloan Consortium International Conference on Online Learning*, Florida. Retrieved from http://sloanconsortium.org/2010aln/presentation/using-voicethread-counseling-role-playing-sessions

Pichette, F. (2009). Second language anxiety and distance language learning. *Foreign Language Annals, 42*(1), 77-93.

Rosell-Aguilar, F. (2006). Online tutorial support in open distance learning through audio-graphic SCMC: Tutor impressions. *The JALT CALL Journal, 2*(2), 37-52.

Rosell-Aguilar, F. (2007). Changing tutor roles in online tutorial support for open distance learning through audio-graphic SCMC. *The JALT CALL Journal, 3*(1-2), 81-94.

Rossiter, D., and Lam, G. (2007). Gong - a voice for the web world. Paper presented at the *Teaching and Learning Symposium 2007: Learning Impact through Teaching Innovation*. Retrieved from celt.ust.hk/tlsymp07/pdf/R3-DavidRossiter.pdf.

Schmidt, R. (1983). Interaction, acculturation and the acquisition of communicative competence. In N. Wolfson, and E. Judd (Eds.), *Sociolinguistics and language acquisition* (pp. 137-174). Rowley, MA: Newbury House.

Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*, 129-58.

Shigeta, K. (2010). Computer assisted instruction adapted to individual differences. In Y. Yamauchi (Ed.), *Pedagogy of digital materials* (pp. 11-24). Tokyo: University of Tokyo Press.

Strauss, A., and Corbin, J. (1990). *Basics of qualitative research: Grounded theory procedures and techniques* (1st ed.). California: SAGE publications, Inc.

Suzuki, K. (2006). Instructional design supported by system approach and learning psychology. In E. Nojima, K. Suzuki and A. Yoshida (Eds.), *Human information science and E-learning* (pp. 91-103). Tokyo: Open University Japan.

Swain, M. (1985). Communicative competence: Some roles of comprehensible input and comprehensible output in its development. In S. Gass, and D. Madden (Eds.), *Input in second language acquisition* (pp. 235-253). Rowley, MA: Newbury House.

The Gong Project. (2010a). *Gong 5*. Retrieved from http://gong.ust.hk/.

The Gong Project. (2010b). *NanoGong 4*. Retrieved from http://gong.ust.hk/nanogong/

Tullis, T., and Albert, B. (2008). *Measuring the user experience*. Amsterdam: Morgan Kaufmann Publishers.

Vetter, A., and Chanier, T. (2006). Supporting oral production for professional purposes in synchronous communication with heterogenous learners. *ReCALL, 18*(1), 5-23.

Williams, K. E., and Andrade, M. R. (2008). Foreign language learning anxiety in Japanese EFL university classes: Causes, coping, and locus of control. *Electronic Journal of Foreign Language Teaching, 5*(2), 181-191.