Review of Speech-to-Text Recognition Technology for Enhancing Learning

Rustam Shadiev[1], Wu-Yuin Hwang[2], Nian-Shing Chen[3] and Yueh-Min Huang[1*]

[1]Department of Engineering Science, National Cheng Kung University, No.1, University Road, Tainan 70101, Taiwan // [2]Graduate Institute of Network Learning Technology, National Central University, No.300, Jhongda Road, Jhongli 32001, Taiwan // [3]Information Management Department, National Sun Yat-sen University, No.70, Lienhai Road, Kaohsiung 80424, Taiwan // rustamsh@gmail.com // wyhwang@cc.ncu.edu.tw // nschen@mis.nsysu.edu.tw // huang@mail.ncku.edu.tw

[*]Corresponding author

ABSTRACT

This paper reviewed literature from 1999 to 2014 inclusively on how Speech-to-Text Recognition (STR) technology has been applied to enhance learning. The first aim of this review is to understand how STR technology has been used to support learning over the past fifteen years, and the second is to analyze all research evidence to understand how Speech-to-Text Recognition technology can enhance learning. The findings are discussed from different perspectives as follows: (a) potentials of STR technology, (b) its use by specific groups of users in different domains, (c) quantitative and/or qualitative research methodology used, and (d) STR technology implications. Some STR literature review showed that in earlier stage of development, the STR technology was applied to assist learning only for specific users, i.e., students with cognitive and physical disabilities, or foreign students. Educators and researchers started to apply STR technology in a traditional learning environment to assist broader group of users, while STR technology has been rapidly advancing over the years. The review revealed a number of distinct advantages of using STR for learning. That is, STR-generated texts enable students to understand learning content of a lecture better, to confirm missed or misheard parts of a speech, to take notes or complete homework, and to prepare for exams. Furthermore, some implications over the STR technology in pedagogical and technological aspects were discussed in the review, such as the design of technology-based learning activities, accuracy rate of the STR process and learning behaviors to use STR-texts that may limit the STR educational value. Thus, the review furthermore discussed some potential solutions for the future research.

Keywords

Speech-to-text recognition technology, Literature review, Supporting and enhancing learning, Group of users

Introduction

Recent evidence suggests that some challenges and limitations exist in physical and online synchronous learning environments that still require attention to solve them (Camiciottoli, 2005; Miller, 2007; Chen, Ko, Kinshuk, & Lin, 2005; Huang & Chiu, 2014; Neilsen, 2009; Nisbet & Spooner, 1999; Shadiev, Hwang, & Huang, in press; Wang, Chen, & Levy, 2008). For example, on an academic event, information is usually addressed through audio channels so that students with learning or physical disabilities, foreign students, and other at risk populations are challenged to understand the content (Camiciottoli, 2005; Lee, 2011; Miller, 2007; Nisbet & Spooner, 1999). Furthermore, one of the most common concerns reported in relation to online learning literature is the poor audio quality due to restricted internet bandwidth availability and traffic congestion (Chen et al., 2005; Wang et al., 2008). These problems can hinder students' understanding of a delivered speech, and this may hamper students from engaging in classroom participation and interaction (Camiciottoli, 2005; Miller, 2007; Chen et al., 2005; Wang et al., 2008).

According to related literature, abovementioned problems can be solved by adopting some assistive media-to-text recognition technologies, such as writing-to-text, image-to-text, diagram-to-text, text-to-speech, speech-to-text, and handwriting-to-text. For example, Speech-to-Text Recognition (STR) technology synchronously transcribes text streams from speech input and shows them on a whiteboard or students' computer screens (Alapetite, Andersen, & Hertzum, 2009; Fichten et al., 2000; Hwang, Shadiev, Kuo, & Chen, 2012; Jones, 2005; Konur, 2007; Kuo, Shadiev, Hwang, & Chen, 2012; Shadiev, Hwang, & Huang, 2013). It is suggested that STR-generated texts can greatly help students attain a better understanding of a lecture, do simultaneous note-taking during lectures, and complete homework (Hwang et al., 2012; Kuo et al., 2012; Shadiev et al., 2013). Furthermore, it is argued that STR-generated text can be employed as an additional text confirmation of what is being said, and it aids comprehension in case when listeners are students with learning or physical disabilities, foreign students, and other at risk populations (Shadiev et al., 2013; Wald & Bain, 2008).

The pedagogical usefulness of STR-technology application to enhance students' learning was emphasized in several studies. The following are a few examples. The Speech Recognition in Schools Project (Nisbet & Wilson, 2002; Nisbet, Wilson, & Aitken, 2005) helped students to overcome difficulties in reading, writing, and spelling. The project presented significant improvements in some students' basic reading, writing, and spelling skills with the support of STR. Wald and Bain (2008) developed STR applications to assist deaf students and non-native speakers to be involved in lectures. According to their research, students perceived that text generated by STR could improve learning if its accuracy is fairly good (Colwell, Jelfs, & Mallett, 2005; Wald & Bain, 2008). Ryba et al. (2006) examined the application of STR in a university lecture theatre attended by students who were native and non-native speakers of English. A non-native English lecturer delivered a course about information system to the participants by using STR. The participants reported that the system was a potentially useful instructional support mechanism; however, a greater accuracy in lecture text vocabulary recognition of the system needs to be achieved. Shadiev et al. (2013) applied STR technology to assist non-native English participants to learn at a seminar in English. It was found that most participants perceived that transcripts were useful for learning. Moreover, nineteen learning strategies to use transcripts were discovered, and participants with different learning achievements demonstrated different learning behaviors to use transcripts. Hwang et al. (2012) and Kuo et al. (2012) employed the STR technology in teaching and learning activities in online synchronous learning environment. Compared to students who did not use transcripts, it was found that students who used transcripts showed improvement on homework accomplishments and post-test results (Hwang et al., 2012; Kuo et al., 2012).

This study aims to review previous STR technology relevant literature and how it can enhance learning. STR technology was mostly used to assist specific groups of students (i.e., students with learning or physical disabilities or foreign students) in order to guarantee them the equal access to learning. However, as time passed by, the target users involved into research on STR technology has got broader. That is, nowadays STR technology is adopted to assist not only students with special needs but also general population of students for more educational purposes, such as enhancing students' understanding of a presented learning content during and after academic activities as well as offering students guidance to accomplish reflective writing and homework. Furthermore, due to recent improvement of STR technology, particularly its accuracy rate, the technology is also adopted to support collaborative learning activities with multiple participants speaking simultaneously, such as group discussions or students' oral presentations. Therefore, this study particularly summarized STR development history and its usage by specific group of users. First, this study looks into how STR technology has been used in education over the past fifteen years by reviewing relevant research. Second, this study demonstrates how effective STR technology can be to enhance learning for different groups of users, such as students with learning or physical disabilities, foreign students, online students, and students who study in physical environment. This study further highlights findings on STR technology and proposes several suggestions for future research. The following research questions were addressed in this review:

- How has Speech-to-Text Recognition technology been used in learning over the past fifteen years?
- What learning activities can make the most of Speech-to-Text Recognition technology and bring out the best learning outcomes to enhance learning?

Methodology

The literature from 1999 to 2014 inclusively were searched using the search terms such as speech-to-text, voice-to-text, speech recognition, transcription, and learning from ACM Digital Library, EBSCO Discovery Service, ERIC, PsychINFO, and Social Sciences Citation Index databases. A total of 42 selected articles were reviewed. Primary data source for this review include peer-reviewed journal articles, conference proceedings, and frequently cited books. The references provide a complete list of all the articles reviewed for this project (marked with an asterisk). The publications reviewed are organized into four dimensions that address (a) potentials of STR technology, (b) its use by specific groups of users in different domains, (c) research findings from studies using quantitative and/or qualitative methodologies, and (d) issues and considerations of applying STR technology. These categories provide an organizational framework to understand how STR technology has been used in learning, and to explore any research evidence in terms of how Speech-to-Text Recognition technology can enhance learning.

Findings of this review were organized into two particular aspects. The first aspect is STR methodology and approach. That is, this review aimed to understand how the STR technology has been applied to support learning. Findings in this aspect are reported based on STR technological development. In earlier stages, STR technology was

not as well-developed as it is now. One major issue was how to generate a satisfactory accuracy rate of transcripts from a speaker's speech. Therefore, earlier attempts were made to apply STR technology only for particular groups of users, such as student with cognitive or physical disabilities. Afterwards, a lot of studies of how computers can assist language learning were carried out with applications of STR technology. Finally, STR technology was developed more mature and reliable; the accuracy rate of recognition voice into text became higher and even STR could transcribe multiple speakers at the same time. Thus, some experts applied STR in a traditional classroom during lectures or collaborative learning activities on other fields of knowledge. The other aspect is potentials and findings of STR technology to facilitate learning. That is, this review attempted to analyze all research evidence that how Speech-to-Text Recognition technology can enhance learning. Findings in this aspect centered on applications of the STR technology to support learning of different groups of users in traditional and online learning environments.

The usage of the STR in learning

This section analyzes findings from other studies regarding how the STR technology was applied in learning in the past fifteen years. Findings in this section are classified into the following categories: the usage of STR to assist learning of students (1) with cognitive or physical disabilities, (2) online students, (3) non-native speakers, and students (4) in traditional learning environment, and (5) in collaborative academic activities. Main findings of this section are summarized in Appendix.

Students with cognitive or physical disabilities

According to Lee (2011), Neilsen (2009), Nisbet and Wilson (2002), Nisbet et al. (2005), and Zhili, Wanjie and Jiacheng (2010), many students, who need additional support, have difficulties in reading, writing, or spelling, due to motor difficulties, visual impairment, or specific learning difficulty. Elliot, Foster, and Stinson (2002) suggested that students with hearing impairments rely on either reading lips or watching an interpreter to access to what the instructor spoke. It is extremely difficult for these students to focus their visual attention on note-taking and the instructor (or interpreter) simultaneously. Therefore, it was suggested to apply assistive technologies, such as a speech-to-text support service, to enhance computer-assisted learning for students with different types of disabilities.

In the Speech Recognition in Schools Project (Nisbet & Wilson, 2002; Nisbet et al., 2005), STR technology was used by secondary school students with special educational needs for one semester. Forty students with reading, writing or spelling difficulties from different schools in Scotland participated in this project. The project provided students with Dragon Naturally Speaking or IBM ViaVoice speech to text recognition software, and technical support and training delivered on site. Besides, a training pack was designed by the project for students to learn how to use one or other kinds of software.

IBM ViaVoice software was applied to assist students with hearing impairments to listen to lectures in three Canadian high schools (Leitch, 2008) and one university in UK (Wald, 2010; Wald & Bain, 2008). In the studies of Leitch (2008), Li et al. (2011), Wald (2008), Wald (2010), and Wald and Bain (2008), teachers were engaged in training and testing the STR technology for a minimum period of two weeks. When accuracy rates reached a certain satisfactory level of at least more than 80%, teachers employed STR technology when giving lectures. During this time, teachers displayed text to students. Most students in the studies of Leitch (2008), Li et al. (2011), Wald (2010), and Wald and Bain (2008) used STR-generated texts as an additional resource to verify and clarify what has been said by the lecturer as well as to take their own notes and argue their own opinions.

Elliot et al. (2002) carried out another study with USA high school and college students who were deaf and hard of hearing. Students were provided with notes from a speech-to-text support service called C-Print. The purpose was to help students to fill in gaps in their understanding of what transpired during class. This STR system produced a real-time transcripts displayed for students on their personal or laptop computers to access the information. A C-Print notes include as much information as possible, generally providing almost all of the meanings of the spoken lecture content. After class, notes were saved and edited so students could use them in paper or electronic format. The study with high school students lasted for approximately ten weeks and with college students for the period between ten to sixteen weeks. Analyzing students' learning behaviors, Elliot et al. (2002) found that high school students typically read the notes only, while college students used multiple study strategies with the notes.

Online students

Network traffic congestion can cause poor quality of audio communication in a synchronous cyber classroom. Under such condition, students are not able to hear a speaker clearly. This issue was viewed as one technological challenge. It negatively affects online teaching and learning activities as it hinders students' understanding of a delivered speech, and it also hampers students from engaging in classroom participation and interaction (Chen et al., 2005; Hwang et al., 2012; Kuo et al., 2012; Wang et al., 2008). To address this issue, Hwang et al. (2010), Hwang et al. (2012), Kuo et al. (2011), Kuo et al. (2012), and Shadiev (2011) employed Windows Speech Recognition in the Microsoft Operating System for STR tools to support various teaching and learning activities (lectures, oral presentations, and discussions) and communication for students in a synchronous learning environment. This application in Microsoft Operation system is the most available tool to get for the students and teachers to participate in the experiment. Way et al. (2008) argued that this application is similar to other open-source products that are easy to use and functions well, and it is available at no additional cost for users. The teacher and students of these two studies were given a training session of STR for three to six weeks before using it for teaching and learning activities. They dictated two articles over their local STR, and the content of the articles are related to the target teaching and learning activities. Thus, the STR technology could "learn" a speaker's voice and terminology for specific field during the training, and then achieve a certain level of accuracy rate when being applied in the activities. A speaker's speech was transcribed by the STR technology into text which was displayed simultaneously to students on their computer screens. Thus, the students could listen to a speaker and read the transcripts at the same time. More importantly, STR-generated text was saved for further revision to fix some recognition errors, and the students could obtain a nearly verbatim transcript to study it after the activities and to complete summary writing tasks.

Non-native speakers

According to Camiciottoli (2005), and Wu and Alrabah (2009), non-native speaker students have difficulties listening to content of an academic speech in a language other than their mother tongue (Bennett, Hewitt, Mellor, & Lyon, 2007; Camiciottoli, 2005; Wu & Alrabah, 2009). Bain et al. (2005) defined this problem as "accessibility issue." Many of non-native speaker students flounder in delivering speeches in a foreign language and they need to make extra efforts when attempting to comprehend these speeches (Miller, 2007). Studies conducted with non-native speaker students revealed that they are silent and, generally, they engage in limited classroom participation and interaction (Bain et al., 2005; Camiciottoli, 2005; Miller, 2007). For this reason, speakers who give a speech to a non-native speaking audience should be aware of the potential obstacles and consider the need of delivery method variation. Li et al. (2011) and Wald and Bain (2008) proposed application of the speech-to-text recognition technology as a potentially reliable tool for non-native speaker students to better understand a speech given in a foreign language.

Ryba et al. (2006) applied IBM's ViaVoice and Viascribe STR technology in an university lecture theater with one hundred sixty participants. Half of the participants were native speakers of English while the other half were not. Three 2-hour lectures lasting over a three-week period were delivered to the participants by a non- English speaking lecturer; yet, the lectures were given in English. The lecturer was trained to use the STR system and then applied it during the lectures. Spoken lectures were transcribed into text through STR technology, and then the text was displayed on a large screen in front of the lecture theatre so that students could both see and hear the lecture. After the lecture, the STR-generated texts could be saved and edited, punctuations were inserted, recognition errors were corrected and redundancies removed. The STR-generated text was accessible via the internet.

Coniam (1999) carried out a study in which Dragon Naturally Speaking STR system was employed to assist students to learn English as a second language. A small group of very competent second language subjects participated in the study. First, the subjects had the system recognized their own voices during 45 minutes by reading a training text of 3800 words. Next, the subjects read a text into the voice recognition software and this voiced text was analyzed to compare with that generated from voice of another small group of native speakers.

Shadiev et al. (in press) and Shadiev et al. (2013) applied Windows Speech Recognition built in the Microsoft Operating System to an eight-week graduate seminar program on advanced learning technologies. Seminar in English was executed once a week for non-native English participants. In the program, the participants needed to give a speech that would be graded and the STR technology generated transcripts from the speech. In order to

achieve a good accuracy rate of the STR, every participant was trained to use the system beforehand. Besides, transcripts were edited to fix some recognition errors and nearly verbatim transcripts were projected on the whiteboard along with the Microsoft Power Point slides of a presenter. Furthermore, transcripts were available online so that participants could study them during and after the seminar.

In the other study, Shadiev and his colleagues (2014) employed Windows Speech Recognition built in Microsoft Operating System for two lectures in English to assist non-native English participants to better comprehend lectures' content. The difficulty of the first lecture was intermediate level and the other one was advanced level. Both lectures were delivered to participants through computer screens. Participants could see video of the instructor, slides of the lectures, and text generated by the STR. Shadiev et al. (2014) investigated participants' visual attention on STR-generated text by employing eye-tracking technique. How differently effective STR-texts can be to influence participants' learning achievement was also assessed. Besides, Shadiev et al. (2014) compared visual attention and learning behaviors to different characteristics of participants, such as learning ability, learning style preferences, and gender, in using STR-texts. Finally, students' perceptions regarding usefulness of STR-texts for learning were also explored.

In the study of Weggerle, Schmidt, and Schulthess (2009), the spoken lecture was recorded and automatically transcribed by using the Naturally Speaking Software. The quality of the transcription was improved through a subsequent editing and correction process, and the pages from the beamer presentation were added to the time line of the text. STR-generated lecture notes were available as an interactive web application; they were searchable and allowed for individual annotations. The lecture notes were available for students for one semester in Technical Informatics course. Weggerle and colleagues aimed to enhance the learning efficiency for foreign language students and for those who didn't attend the course but allowed to access to the lecture notes.

Students in traditional learning environment

Luppi et al. (2009) argued that the adoption of the STR technology in traditional learning environment has several benefits. One of them is to improve teaching methods and to enhance learning opportunities. For example, by using the STR, teachers can take a proactive, rather than a reactive approach to teach students with different learning styles. It provides educators with a practical means of making their teaching accessible and improves the quality of instruction in the process (Luppi et al., 2009).

Ranchal et al. (2013) adopted IBM ViaScribe and IBM Hosted Transcription Service to assist university students in the lectures of life and social sciences courses. Two distinct methods of the STR-mediated lecture acquisition, such as real-time captioning and post lecture transcription, were evaluated in the study. The instructor underwent initial voice training to develop a voice profile for the systems and to improve STR accuracy before starting real-time captioning or post-lecture transcription. During class, the STR processed verbal information into textual captions and streamed them on a screen or students' computers. Students received drafts, unedited transcripts during lectures. However, errors in the transcribed text were then corrected and it was available for students after class.

Ryba et al. (2006) applied IBM ViaVoice and Viascribe systems to listen to lectures about information systems in an university lecture theatre with more than one hundred students. First, the lecturer underwent training to develop a voice profile for the system and to achieve a high level of accuracy through inputting dialogue and vocabulary into the system. Then, the STR systems transcribed lectures into texts which were displayed on a large screen in front of the lecture theatre. Besides, the edited STR-generated texts were delivered to students after the lecture via the internet. Ryba et al. (2006) explored students' perceptions of using STR-texts and to what extent students make use of them. Moreover, the main advantages and limitations of using STR-texts were investigated.

Goddard, Kaplan, Kuehnle, and Beglau (2007) applied Read & Write GOLD system as cognitive prosthesis to support students' learning needs in general education classrooms. Teachers and students were trained to use the system first. Then students used STR to reach certain level of understanding of learning material that they couldn't understand before. As soon as students caught up the level of their peers' understanding, they reduced or abandoned usage of the STR.

The STR technology applications used for supporting collaborative learning activities

According to Wald and Bain (2008), the STR usage to date focus primarily on situations where there is only one speaker, i.e., one-way lecturing. They argued this is a limited scenario because many occasions involve multiple speakers. Hwang et al. (2012), Kheir and Way (2006), Li et al. (2011), Wald and Bain (2008), and Zschorn, Littlefield, Broughton, Dwyer, and Hashemi-Sakhtsari (2003) suggested that STR technology can be used for multiple speakers environment as well; it may aid in producing transcripts from discussions, meetings, and other collaborative activities.

The STR applications were developed in studies of Fiscus, Ajot, and Garofolo (2007), Wald and Bain (2008), and Zschorn et al. (2003) for multiple-speakers environment.

In the study of Wald and Bain (2008), IBM ViaScribe system was employed when lectures were being given in traditional classroom environment. The lecturer gave a speech on a particular topic and the STR technology simultaneously generated text from the speech which was displayed on the screen for students to read. Students' questions to the lecturer were repeated by the lecturer to the STR and were transcribed on the screen.

Zschorn et al. (2003) developed Automatic Transcriber of Meetings prototype to use in order to automatically create records and transcripts of a discussion during the meeting. The prototype creates transcripts that include all the attendee, agenda, highlights and utterances information of the meeting.

The STR technology, developed by Augmented Multiparty Interaction with Distance Access, IBM, International Computer Science Institute and SRI International and Karlsruhe University, was employed by Fiscus et al. (2007) at small conference room meetings, interactive lectures in a small meeting room, and coffee breaks from lecture meetings. Conference meetings consisted of primarily goal-oriented, decision-making exercises and varied from moderated meetings to group consensus-building meetings. Conference meetings were highly interactive and multiple participants contributed to the information flow and decision-making. Lecture meetings consisted of educational events where a single lecturer briefed audiences on a particular topic. While the audience occasionally participated in question and answer periods, the lecturer predominately controlled the meeting. Coffee breaks from lecture meetings consisted of excerpts selected from lecture meetings where the participants took a coffee break during the recording.

Fiscus et al. (2007), Li et al. (2011), Wald and Bain (2008), and Zschorn et al. (2003) focused primarily on technological aspects of STR technology. That is, these studies developed STR systems and tried to improve STR accuracy rate, but they didn't put much emphasis on evaluating the effects of systems on learning achievement and other pedagogical issues such as systems' practicality or functionality to enhance learning.

Kuo et al. (2011), Kuo et al. (2012), and Shadiev (2011) applied Windows Speech Recognition built in Microsoft Operating System for individual oral presentations and group discussions of native speakers of Chinese students in a synchronous cyber classroom. The effectiveness of applying STR on learning performance was analyzed. Students in the study of Kuo et al. (2011), Kuo et al. (2012), and Shadiev (2011) Kuo et al. (2012) participated in individual oral presentations and group discussions, in which STR technology generated transcripts and it was shown on students' computer screens. STR-generated transcripts were used by students during and after learning activities.

Potentials and findings of the STR to facilitate learning

This section illustrates related literature review on potentials and findings of the STR to facilitate learning in terms of different users and learning environments. Besides, what considerations associated with STR technology application are reported in this section. Main findings of this section are summarized in Appendix.

Students with cognitive or physical disabilities

Nisbet and Wilson (2002), and Nisbet et al. (2005) evaluated STR application in a classroom by using Pupil Evaluation Questionnaire which was completed by students with special needs, in collaboration with teachers. Students' responses to the questionnaire were analyzed and results showed that 70% of the students intended to

continue using STR for learning purpose. According to students, STR technology served as an effective tool to write and record their speech. In some cases, the application of STR has enhanced students' basic reading, spelling and writing skills. For example, the system could read out their STR-text and play recorded audios so that students could compare the read-out of STR-text to playback to identify misrecognitions. Moreover, recordings of the student's dictations were saved, so that students could correct them later with the help of a teacher.

Leitch (2008), Wald (2008), Wald (2010), and Wald and Bain (2008) conducted a survey on students with hearing impairments. The survey data analysis revealed that STR-generated lecture transcriptions helped students to understand lectures content better. Besides, students believed that transcriptions could improve their learning.

To reveal advantages of STR in learning, Elliot et al. (2002) interviewed students with hearing impairments. High school students claimed that reviewing notes helped them to fill in gaps of their understanding of what transpired during class. College students mentioned that STR notes were useful for test preparation and for traditional academic purposes, e.g., background material for research papers.

Online students

Hwang et al. (2010) and Hwang et al. (2012) carried out an experiment, and its results showed that, in an online synchronous learning environment, students who used transcripts (the experimental group) showed a more moderate improvement in their performance than students who did not use transcripts (the control group) on homework accomplishments. However, once the students in the experimental group familiarized themselves with the STR-generated texts and used them as learning tools, they significantly outperformed the control group students in post-test results. Results of the other experiment, carried out by Kuo et al. (2011) and Kuo et al. (2012) in an online synchronous learning environment showed that students who used transcripts performed significantly better than those who did not use transcripts in writing essays, intermediate tests, and post-test evaluations. Furthermore, experimental students in both studies (Hwang et al., 2012; Kuo et al., 2012) perceived that STR system was easy to use and useful for academic activities in online synchronous cyber classrooms. Yet, students expressed their positive willingness to use STR system for learning in the future. According to interviews with experimental students, STR-texts were useful during and after academic activities to understand presented topics, to catch up on missed/misheard parts in a speech, to take notes, and to complete homework. However, it was the STR low recognition accuracy rate when recognizing homophones that students viewed as the one limitation which required attention (i.e., the words with the same pronunciation but different meanings).

Non-native speakers

The participants in the study administered by Ryba et al. (2006) claimed that STR technology has a potential to be an instructional support mechanism, and there were a number of perceived benefits associated with the STR use. Most non-native speaker students, due to their language barrier and mishearing some important parts of the instructor's speech, admitted that STR-texts were useful during lectures to follow the instructor and to clarify and to understand lecture content.

Experimental results in the study of Coniam (1999) showed that transcripts generated from speeches of second language speakers by using STR were with significantly lower accuracy rate than those generated from speeches of native speakers. These results were consistent in line with native speakers' scores; that is, the highest accuracy scores were achieved at the lowest level of analysis, the word level, and the lowest scores at the t-unit, or sentence level of analysis. Furthermore, Coniam (1999) concluded that STR technology is still at early stage of development in terms of accuracy and single-speaker dependency.

Results obtained by Shadiev et al. (in press) and Shadiev et al. (2013) revealed that non-native English participants took advantage of nineteen learning strategies to use STR-generated transcripts during and after seminars in English. Transcripts were used to understand seminar's topics, to answer seminar's questions, and to complete summary writing tasks. However, participants employed learning strategies differently. That is, some participants used transcripts effectively by studying them thoroughly, and they used most important parts of transcripts to write summaries along with their own elaborated ideas. On the other hand, some participants performed meaningless

learning behaviors as they studied transcripts superficially and employed copy-and-paste method to complete summary writing tasks. As a result, those participants who employed meaningful learning strategies to use STR-generated transcripts received higher scores for their summaries than those who used undesirable learning strategies. Finally, Shadiev et al. (in press) and Shadiev et al. (2013) found that most non-native English participants perceived that available STR-generated transcripts were useful for their learning during and after a seminar. However, low STR accuracy rate was a problem proposed by some participants. This problem caused their negative perceptions and slightly decreased their perceived acceptance to use STR in the future. Those participants admitted that there was not enough time to receive STR technology training. Furthermore, as non-native speakers of English, they may have strong accent in pronouncing some words or stumbled over them, and this caused many errors in STR-generated text when speaking to STR.

By using eye-tracking technique to explore non-native English speaker students' visual attention to STR-generated text, Shadiev et al. (2014) found that students relied on STR-texts more than on video of the instructor and Power Point slides during lectures in English. Shadiev and his colleagues concluded that STR-texts were useful during the lectures as to aid learning. Students made a greater use of STR-texts to enhance their comprehension of the lectures content. Shadiev et al. (2014) found that all students, no matter what level of their English as a foreign language (EFL) ability, learning style preference and gender are, learned with the aid of STR-texts. However, STR-texts significantly helped to enhance learning performance of participants with low level of EFL ability. Shadiev et al. (2014) argued that participants of low EFL ability took better advantage of STR-texts while being engaged in perceptual processing during listening. For example, some participants admitted that reading STR-texts could help them understand lecture content better. Some participants mentioned that STR-texts could help them to locate new and unfamiliar vocabulary. Results of this study also revealed that participants tended to gaze on all areas of interest, i.e., video of the instructor, Power Point slides and STR-texts during an intermediate-level lecture, but more on STR-texts. Furthermore, results showed that participants tended to gaze mostly on STR-texts during an advanced-level lecture. Shadiev et al. (2014) explained this finding out of difficulty of the lectures; as difficulty of the lecture is higher, participants paid their visual attention to STR-texts more in order to comprehend the lecture content better.

Weggerle et al. (2009) found that introducing STR into the classroom had several positive learning benefits. For example, pronunciation and correct grammar of the lecturer improved substantially and thus, improved students learning. A transcribed text from lectures was voted by students to be very valuable for exam preparation. However, Weggerle et al. (2009) reported that employing STR technology in their study rarely offered a recognition rate of more than 80 percent, and the delay involved in real time transcription was disturbing. According to the literature on STR (Hwang et al., 2010; Kheir & Way, 2006; Wald, 2010), text generated under such circumstances becomes unhelpful and meaningless for students' learning.

Students in traditional learning environment

Ranchal et al. (2013) concluded that during a science course in traditional classroom, students could benefit from having both, real-time lecture transcriptions and post lecture transcriptions. When lecture transcripts were available, students were able to pay more attention to the instructor instead of focusing on recording complete class notes, and with the lecture transcripts, they could review the lecture material for several times. Besides, students were able to take notes, make comments and remarks, and look for specific text by searching keywords and time periods. However, Ranchal et al. (2013) found that students who had access to post lecture transcriptions received higher scores on the quiz than those who received real-time transcriptions only. Moreover, overall class grades of students who received post lecture transcriptions were higher.

Results of the class survey in the study of Ryba et al. (2006) revealed that more than 30% of students used STR-texts to learn information systems in a traditional classroom. Ryba et al. (2006) further found that more than 40% of students tend to use STR-texts. In the survey, students mentioned that STR-text helped them to understand the lecture, confirming what was missed in the lecture, and to take notes. However, most students claimed that the accuracy rate of STR technology was not precise enough, and text generated with many errors could distract their attention from the lecture.

Goddard et al. (2007) surveyed their participating teachers and primary general education students about benefits of Read & Write GOLD system. From the survey, it was found that the system benefits students to write, to edit, and to rewrite. According to the teachers, students' writing improved after they started using the system. Students heard and

recognized obvious errors that they, at first, did not believe they had made as the system read exactly what the students had written. Editing was not a struggle as the software was reading students' work to them. Teachers reported that spending more time with writing, editing, and rewriting improved the final product. Furthermore, teachers continuously reported that all students were engaged in using the software throughout the year, not just short-term interest.

The STR technology applications in collaborative academic activities

Kuo et al. (2012) found that STR technology is a potential tool to facilitate collaborative learning activities, such as oral presentations and group discussions, and it can also improve their overall learning performance. Experimental results in the study of Kuo et al. (2012) and Shadiev (2011) revealed that students who used STR-generated texts (the experimental group) performed far better than those who did not (the control group) in writing essays, intermediate tests and post-tests. Furthermore, according to results, most students perceived that STR was a useful aid when prepare for oral presentations and essays writing. However, there was a problem that it was difficult to attain a high recognition accuracy rate of STR during group discussion. Therefore, students who got transcripts with low accuracy rate and experienced delay in STR-text generation did not perceive STR as an easy tool to use, and found it not so useful for group discussions. One reason of having a low accuracy rate was due to a speed of students' speech. When a student spoke too slowly, the STR application recognized one spoken word as two. Conversely, when the student spoke too quickly, the STR application recognized two spoken words as one. Furthermore, it was not easy to attain a fluent speech (i.e., when the speech has to be delivered moderately fluent and accurate) during group discussion so that the STR generated texts with low accuracy rate. In addition, students mentioned that their speech became more spontaneous during group discussion which also resulted in low accuracy of transcriptions content. Due to these issues, students couldn't make argumentative discourse with the goal to acquire knowledge but were engaged in idea exchange only.

Literature review shows that participants in most studies on STR, no matter what category of users they belong to and no matter what learning environment they learn in, had positive perceptions toward usefulness of STR transcripts for learning. However, Mayer (2008) argued that the same information presented in both auditory and written format makes it redundant and gives rise to a split-attention effect and cognitive load (modality principle). However, the participants still relied on transcripts in written format because of their learning needs, physical/ cognitive abilities, or specific learning environment (Elliot et al., 2002; Hwang et al., 2012; Kuo et al., 2012; Leitch, 2008; Nisbet & Wilson, 2002; Nisbet et al., 2005; Ryba et al., 2006; Shadiev et al., 2013; Shadiev et al., 2014; Wald, 2010; Wald & Bain, 2008). According to Kirsh (2010), and Rogers, Sharp, and Preece (2011), external representations, such as STR-generated texts, greatly extend and support students' ability to carry out cognitive activities (e.g., inference, problem-solving and understanding). One benefit that transcripts offer is on memory. Firstly, transcripts reduce memory workload by providing external tokens for the information that must otherwise be kept in mind. Secondly, transcripts serve as visual retrieval cues for long term memory, evoking relevant information that might not otherwise be retrieved. Finally, transcripts are more "enduring" (visual) text-based content, which goes along with the more "temporary" (oral) speech-based presentation. According to Dual Processing theory (Moreno & Mayer, 2002), redundant information presented in two modes (i.e., visual and oral), and processed aurally and visually can support the recognition and learning of that information. Thus, in the finding of Moreno and Mayer (2002), participants used strategies such as scanning transcriptions when they missed or misheard some parts of a speech. In this way, STR technology can provide much more essential support for students to process aural text with the help of simultaneously displayed transcriptions (Jones & Plass 2002; Ryba et al., 2006).

The STR considerations

Three main issues with respect to the STR technology were pinpointed by teachers and students in the reviewed literature. First issue was reported in Hwang et al. (2012), Kuo et al. (2012), Shadiev et al. (2013), and other related studies and it relates to the usage of the STR technology. It was found that students who did not use the STR technology or used it irregularly perceived STR not a useful aid for learning. The second issue associates with STR process accuracy rate. Most studies report that although STR technology is useful for learning, a greater accuracy in the system's recognition of speech is required. According to Alapetite et al. (2009), Fichten et al. (2000), Jones (2005), Kanevsky et al. (2006), Kheir and Way (2006), Konur (2007), Petta and Woloshyn (2001), and Wald (2010), texts generated with low accuracy recognition rate contain many errors which are incomprehensible and meaningless for learning. Finally, it's the issue that relates to learning behaviors in using STR texts. Shadiev and his colleagues

(2013) noticed that participants in their study performed slack learning behaviors, such as studying transcripts superficially and employing copy-and-paste method to complete summary writing tasks. Performing such learning behaviors, students did not learn much, and as a result, they were scored low on examinations.

Suggestions and implications

To begin with, the literature review suggests that educators and researchers design technology-based teaching and learning activities in a way that encourages users (i.e., instructors and students) to use STR more regularly. Such approach will enable users to identify strengths and limitations of the STR, and then to fully utilize STR for their teaching and learning. For example, Hwang et al. (2012), Kuo et al. (2012), and Shadiev et al. (2013) encouraged and motivated their participants by training them how to use STR technology first and then to use it to complete homework. With such kind of learning activity design, students could identify what advantages and disadvantages of the STR are through real experience with STR technology.

According to Hwang et al. (2012), Jones (2005), Kuo et al. (2012), and Nisbet, Wilson, and Balfour (2008), in order to achieve good detection accuracy rate, the STR application training should last at least one week. Hwang et al. (2012) and Kuo et al. (2012) argued that by using training scripts with content related to the learning material, STR technology can "learn" domain-specific terminologies during the training period and then it can recognize them when learning activities are ongoing. To increase STR process accuracy rate during the training period and academic activities, Nisbet et al. (2008) suggested that we use STR dictionary and correction tool. For example, according to Ranchal et al. (2013), a user can add words that are frequently detected to the dictionary so that STR recognizes those words easier. Besides, Ranchal et al. (2013) claimed that a user can simultaneously correct errors in transcript while speaking to STR by using the STR correction tool. Furthermore, recognition errors can be corrected after the lecture. In this case, the instructor or teaching assistant listens to the lecture audio recording and corrects misrecognized words, inserts missed words, or deletes superfluous wording (Ranchal et al., 2013). If transcripts were generated with high error rates, students in the class can be involved in this work collectively by using an online correction tool and sharing the workload among several people. The STR correction tool can also help to train STR against a word that consistently misrecognized; for that, a user has to record a pronunciation of how he/she says that word. Hwang et al. (2012) and Kuo et al. (2012) also suggested that it is feasible to apply a set of strategies during the training on STR technology. Such strategies involve sharing issues related to the STR process with peers, finding possible solutions together, preparing a script with main points of a speech and making rehearsal with a script and STR technology beforehand. According to Colwell et al. (2005), Hwang et al. (2012), Kheir and Way (2006), Kuo et al. (2012), and Wald (2010), only STR-generated text with reasonable accuracy rate of more than 85 percent is useful and meaningful for students. Kheir and Way (2006) reported that, in their study, the accuracy rate of STR improved from 75 percent, when STR was not trained, to 88 percent after minimal training on STR, to 90 percent after moderate training, and to 91 percent after its dictionary was customized with a domain-specific terminology. Furthermore, Kuo et al. (2012) and Ranchal et al. (2013) suggested that speakers try to adapt to the STR recognition capacity by speaking with moderate speed and volume, less spontaneity, and better fluency. Microphone should also be positioned correctly to avoid "breathiness." Nisbet et al. (2008) suggested a speaker speak clearly to STR and avoid non-lexical utterances (e.g., "huh," "uh," or "erm"). Only the speaker's voice should be reliably recorded; if responding to students' questions, the instructor should repeat questions and then respond (Ranchal et al., 2013). To increase its accuracy rate during discussion, Kuo and his colleagues advised that speakers make speaking sentences shorter and at a moderate pace of one sentence after another, and to locate and correct errors in the transcript simultaneously while speaking to STR. Ranchal et al. (2013) recommended speakers to take breaks periodically if lectures are long to check the reliability of the STR system. Based on these findings, it is suggested that offering users a set of guidelines on how to train and speak to STR more efficiently can achieve better STR accuracy rate and make transcripts more useful and meaningful for learning.

To avoid students' slack learning behaviors, it is suggested that, besides providing learning material, participants need to be instructed about how to use effective learning strategies to use STR-texts. Learning strategies to use STR-generated texts were proposed in Nisbet & Wilson, 2002; Nisbet et al., 2005; Ryba et al. (2006), Shadiev et al. (2013), and Shadiev et al. (2014). These strategies can facilitate participants during and after an academic activity to understand content of a presented topic better, to answer questions, and to complete summary writing tasks. Shadiev

et al. (2013) and Shadiev et al. (2014) suggested some more advanced strategies to the ones reviewed in related literature. Two of them are 1) to use a transcript to ask questions, to give comments or to have discussion with others and 2) to compare a transcript with a student's summary in order to confirm that a summary includes all main points of a speech.

Finally, it is suggested that STR technology can be applied in a learning environment not only with a single speaker but with multi speakers as well. In this case, individual learning, such as using lecture transcript to involve in a speech, taking notes and completing homework, will be enhanced. That is, after individual learning, students can share and discuss their opinions about the topic, correct each other's misconceptions, and enhance their own understanding of a topic by using STR technology. However, some issues need to be considered with respect to STR technical and pedagogical process in such learning scenario. For example, one is how to make STR correctly recognize speech input made by multiple speakers with different speech characteristics (e.g., articulation, pronunciation, speech rate) and then distinguish that input in a STR-generated transcript by each speaker and with orderly timeline when it was spoken (Fiscus et al., 2007; Li et al., 2011; Wald & Bain, 2008; Zschorn et al., 2003). Another issue is how to design collaborative learning activities that facilitate students to fully utilize STR-generated texts for learning (Hwang et al., 2012; Kuo et al., 2012).

Conclusions

The following conclusions can be drawn from literature review. First, it is fairly clear that STR technology was applied to aid learning in different ways based on the progress of STR technological development. That is, the earlier studies employed this technology only to assist learning of particular groups of users, such as students with cognitive and physical disabilities, online or foreign students, due to low accuracy rate and a delay in STR process. However, afterwards, research addressing abovementioned technological limitations emerged; as a result, STR technology improved and became more reliable. Then, STR was employed to aid learning of students in a traditional learning environment during and after individual and collaborative learning activities. Second, the literature considered STR technology beneficial to extend learning during and after learning activities. There is widespread consensus in the literature about the number of distinct advantages of STR-texts, such as enabling students to better understand content of academic activities, to confirm missed parts of a speech, to take notes, to complete homework, and to prepare for exams. However, some arguments over the STR technology considerations that limit educational value of the technology still exist. The literature review showed how those considerations can be addressed by employing various approaches to increase the effectiveness of STR application on learning.

Given what was found in the literature, the following are important issues to address in future STR related studies. First, researchers need to begin theorizing the cognitive processes that occur through learning with STR technology. Besides, STR technology needs to be employed based on relevant pedagogical principles for them to be more effective. Second, there is a need to use well-established and reliable outcome measures in future STR studies. For example, the measures used to demonstrate the effects of STR applications should be given careful consideration based on both objective and subjective evidence. More research needs to be conducted in more dynamic and communicative educational settings, such as collaborative teaching and learning with multiple participants speaking to the STR system simultaneously. Besides, whether there will be different learning effects when the STR is applied to learning environments with students of different cultural backgrounds or language families should be investigated. It is possible that there might be impacts on the STR accuracy rate and learning when different cultural backgrounds or language families are concerned. For example, in general, students from oriental cultural background are less active in terms of learning interaction which may influence learning dynamics and outcomes during group discussion. Finally, research should focus on issues that go beyond applications of STR technology. For example, STR technology can be considered from the angle of ergonomics, i.e., to concern the design and arrangement of the technology to make users interact with it more efficiently. Besides, in the future, STR technology can be extended by combining it with other technology, e.g., automatic translation, to simultaneously generate text from a speech and translate it into many languages. Such approach will enable teachers and students to have instant audio-lingual interpretations using their own native languages.

Acknowledgments

This research is partially supported by the "International Research-Intensive Center of Excellence Program" of NTNU and Ministry of Science and Technology, Taiwan, R.O.C. under Grant no. NSC 103-2911-I-003-301, NSC 102-3113-P-006-019-, MOST 103-2511-S-006-007-MY3, and MOST 103-2511-S-006-002-MY3.

References

Alapetite A., Andersen H. B., & Hertzum M. (2009). Acceptance of speech recognition by physicians: A survey of expectations, experiences, and social influence. *International Journal of Human-Computer Studies, 67*(1), 36–49.

Bain K., Basson S., Faisman A., & Kanevsky D. (2005). Accessibility, transcription, and access everywhere. *IBM systems journal, 44*(3), 589–603.

Bennett, S., Hewitt, J., Mellor, B., & Lyon, C. (2007). Critical success factors for automatic speech recognition in the classroom. *In Universal Access in Human-Computer Interaction. Applications and Services* (pp. 224–233). Berlin, Germany: Springer.

Camiciottoli, B. C. (2005). Adjusting a business lecture for an international audience: A case study. *English for Specific Purposes, 24*(2), 183–199.

Chen, N. S., Ko, H. C., Kinshuk & Lin, T. (2005). A model for synchronous learning using the Internet. *Innovations in Education and Teaching International, 42*(2), 181–194.

Colwell, C., Jelfs, A., & Mallett, E. (2005). Initial requirements of deaf students for video: Lessons learned from an evaluation of a digital video application. *Learning, Media and Technology, 30*(2), 201–217.

*Coniam, D. (1999). Voice recognition software accuracy with second language speakers of English. *System, 27*(1), 49–64.

*Elliot, L., Foster, S., & Stinson, M. (2002). Student study habits using notes from a speech-to-text support service. *Exceptional children, 69*(1), 25–40.

Fichten, C. S., Asuncion, J. V., Barile, M., Fossey, M., & Simone, C. (2000). Access to educational and instructional computer technologies for post-secondary students with disabilities: Lessons from three empirical studies. *Journal of Educational Media, 25*(3), 179–201.

*Fiscus, J. G., Ajot, J., & Garofolo, J. S. (2007). The rich transcription 2007 meeting recognition evaluation. *Lecture Notes in Computer Science, 4625*, 373–389.

*Goddard, W., Kaplan, L., Kuehnle, J., & Beglau, M. (2007). *Voice recognition and speech-to-text pilot implementation in primary general education technology-rich eMINTS classrooms.* Retrieved from http://www.emints.org/wp-content/uploads/2012/02/TtS-VRpilot-qualitative.pdf

Huang, Y. M., & Chiu, P. S. (2014). The effectiveness of a meaningful learning-based evaluation model for context-aware mobile learning. *British Journal of Educational Technology*. DOI: 10.1111/bjet.12147

*Hwang, W. Y., Shadiev, R., Kuo, T. C. T., & Chen, N. S. (2012). Effects of speech-to-text recognition application on learning performance in synchronous cyber classrooms. *Journal of Educational Technology & Society, 15*(1), 367–380.

*Hwang, W. Y., Shadiev, R., Kuo, T. C. T., & Chen, N. S. (2010). A study of speech to text recognition and its effect to synchronous learning. In J. Herrington & B. Hunter (Eds.), *Proceedings of world conference on educational multimedia, hypermedia and telecommunications 2010* (pp. 546–555). Chesapeake, VA: AACE.

Jones, D. (2005). Voice recognition: A new assessment tool? *Technology, Pedagogy and Education, 14*(3), 413–427.

Jones, L. C., & Plass, J. L. (2002). Supporting listening comprehension and vocabulary acquisition in French with multimedia annotations. *The Modern Language Journal, 86*(4), 546–561.

Kanevsky, D., Basson, S., Chen, S., Faisman, A., Zlatsin, A., Conrod, S., & McCormick, A. (2006). (2006, June). *Speech transcription services*. Paper presented at the 11th International Conference Speech and Computer, St. Petersburg, Russia.

Kheir, R., & Way, T. (2006, June). *Improving speech recognition to assist real-time classroom note taking*. Paper presented at the 29th Rehabilitation engineering and assistive technology society of North America conference, Atlanta, GA, USA.

Kirsh, D. (2010). Thinking with external representations. *AI & Society: Journal of Knowledge, Culture and Communication, 25*(4), 441–454.

Konur, O. (2007). Computer-assisted teaching and assessment of disabled students in higher education: the interface between academic standards and disability rights. *Journal of Computer Assisted Learning, 23*(3), 207–219.

*Kuo, T. C. T., Shadiev, R., Hwang, W. Y., & Chen, N. S. (2012). Effects of applying STR for group learning activities on learning performance in a synchronous cyber classroom. *Computers & Education, 58*(1), 600–608.

*Kuo, T. C. T., Shadiev, R., Hwang, W. Y. & Chen, N. S. (2011). Effects of applying STR for group learning activities on learning performance in a synchronous cyber classroom. In I. Aedo, N. S. Chen, D. G. Sampson, J. M. Spector, & Kinshuk (Eds), *The 11th IEEE International Conference on Advanced Learning Technologies 2011* (pp. 232–236). Los Alamitos, CA: The IEEE Computer Society Press.

Lee, I. X. (2011). *The application of speech recognition technology for remediating the writing difficulties of students with learning disabilities.* (Unpublished doctoral dissertation). Seattle, WA: University of Washington.

*Leitch, D. (2008). *GIFT Atlantic liberated learning high school pilot project: A study of the transfer of speech recognition technology from university classrooms to high school classrooms.* (Phase III Report). Nova Scotia, Canada: Saint Mary's University press.

*Li Y., Wald M., Wills G., Khoja S., Millard D., Kajaba J., Singh P., & Gilbert L. (2011). Synote: Development of a web-based tool for synchronized annotations. *New Review of Hypermedia and Multimedia, 17*(3), 295–312.

Luppi, E., Primiani, R., Raffaelli, C., Tibaldi, D., & Violi, A. M. (2009). Net4Voice-new technologies for voice-converting in barrier-free learning environments: Development of innovative learning methodologies, experiment and results. *eLearning papers, 13*, 1–13.

Mayer, R. E. (2008). Applying the science of learning: Evidence-based principles for the design of multimedia instruction. *American Psychologist, 63*(8), 760–769.

Miller L. (2007). Issues in lecturing in a second language: Lecturer's behaviour and students' perceptions. *Studies in Higher Education, 32*(6), 747–760.

Moreno, R., & Mayer, R. E. (2002). Verbal redundancy in multimedia learning: When reading helps listening. *Journal of Educational Psychology, 94*(1), 156–163.

Neilsen, M. (2009). Supporting struggling writers with the use of voice recognition software in class. *Literacy, 49*(1), 30–39.

*Nisbet, P. D., & Spooner, R. (1999). *Supportive writing technology.* Edinburgh, Scotland: University of Edinburgh CALL Centre.

*Nisbet, P. & Wilson, A. (2002). *Introducing speech recognition in schools.* Edinburgh, UK: CALL Centre, University of Edinburgh.

Nisbet, P., Wilson, A., & Aitken, S. (2005). Speech recognition for students with disabilities. *Proceedings of the Inclusive and Supportive Education Congress, ISEC 2005 Conference*. Delph, UK: Inclusive Technology.

Nisbet, P., Wilson, A., & Balfour, F. (2008). *Introducing speech recognition in schools: Using dragon naturally speaking.* Edinburgh, UK: CALL Centre, University of Edinburgh.

Petta, T. D., & Woloshyn, V. E. (2001). Voice recognition for on-line literacy: continuous voice recognition technology in adult literacy training. *Education and Information Technologies, 6*(4), 225–240.

*Ranchal, R., Taber-Doughty, T., Guo, Y., Bain, K., Martin, H., Robinson, J., & Duerstock, B. (2013). Using speech recognition for real-time captioning and lecture transcription in the classroom. *IEEE Transactions on Learning Technologies, 6*(4), 299 –311.

Rogers, Y., Sharp, H., & Preece, J. (2011). *Interaction design: Beyond human-computer interaction.* Hoboken, NJ: Wiley.

*Ryba, K., McIvor, T., Shakir, M., & Paez, D. (2006). Liberated learning: Analysis of university students' perceptions and experiences with continuous automated speech recognition. *Journal of Instructional Science and Technology, 9*(1), 1–19.

*Shadiev, R. (2011). *A study of speech to text recognition and its effects on learning performance in synchronous cyber classrooms* (Unpublished doctoral dissertation). National Central University, Jhongli, Taiwan.

Shadiev, R., Hwang, W. Y., & Huang, Y. M. (in press). A pilot study of facilitating cross-cultural understanding with project-based collaborative learning activity in online environment. *Australasian Journal of Educational Technology*.

*Shadiev, R., Hwang, W. Y., & Huang, Y. M. (in press). Investigating applications of speech to text recognition for face to face seminar to assist learning of non-native English participants. *Technology, Pedagogy and Education*.

*Shadiev, R., Hwang, W. Y., & Huang, Y. M. (2014). Investigating applications of speech-to-text recognition to assist learning in online and traditional classrooms. *International Journal of Humanities and Arts Computing, 8*(supplement), 179–189.

*Shadiev, R., Hwang, W. Y., & Huang, Y. M. (2013). Investigating learning strategies of using texts generated by Speech to Text Recognition technology in traditional classroom. In Childress et al. (Eds.), *Proceedings of the AECT International Conference on the Frontier in e-Learning Research* (pp.279–286). Taichung, Taiwan: National Central University & AECT.

*Shadiev, R., Huang, Y. M., & Hwang, W. Y. (2014, July). *Investigating visual attention of students with different learning ability on texts generated by speech-to-text recognition.* Paper presented at the 14th International Conference on Advanced Learning Technologies, Athens, Greece.

*Wald, M. (2008). Learning through multimedia: Automatic speech recognition enhancing accessibility and interaction. *Journal of Educational Multimedia and Hypermedia*, *17*(2), 215–233.

*Wald, M. (2010). Synote: Accessible and assistive technology enhancing learning for all students. In K. Miesenberger et al. (Eds.), *ICCHP 2010, LNCS 6180* (pp. 177–184). Berlin, Germany: Springer.

*Wald, M., & Bain, K. (2008). Universal access to communication and learning: the role of automatic speech recognition. *International Journal Universal Access in the Information Society*, *6*(4), 435–447.

*Way, T., Kheir, R., & Bevilacqua, L. (2008). Achieving aceptable accuracy in a low-cost, assistive note-taking, Speech Transcription System. *Proceedings of the IASTED International Conference on Telehealth and Assistive Technologies* (pp. 72–77). Retrieved from http://www.csc.villanova.edu/~tway/publications/wayAT08.pdf

Wang, Y., Chen, N. S., & Levy, M. (2008). The design and implementation of a holistic training model for language teacher education in a cyber face-to-face learning environment. *Computers and Education, 55*(2), 777–788.

*Weggerle, A., Schmidt, P., & Schulthess, P. (2009, November). *Speech to multi-media document transcription for university lectures.* Paper presented at the 2nd International Conference of Education, Research and Innovation, Madrid, Spain.

Wu, S. H., & Alrabah, S. (2009). A cross-cultural study of Taiwanese and Kuwaiti EFL students' learning styles and multiple intelligences. *Innovations in Education and Teaching International, 46*(4), 393–403.

Zhili, L., Wanjie, T., & Jiacheng, X. (2010). A study and application of speech recognition technology in primary and secondary school for deaf/hard of hearing students. *Proceedings of the 4th International Convention on Rehabilitation Engineering & Assistive Technology* (pp. 44–46). Singapore: Singapore Therapeutic, Assistive & Rehabilitative Technologies Centre.

*Zschorn, A., Littlefield, J. S., Broughton, M., Dwyer, B., & Hashemi-Sakhtsari, A. (2003). *Transcription of multiple speakers using speaker dependent speech recognition*. (DSTO Technical Report DSTO_TR_1498). Canberra, Australia: The Defense Science and Technology Organization.

References, which content was reviewed for the analysis and deriving findings, are marked with an asterisk.

Appendix

Research findings on applications of STR to enhance learning

Reference	Research focus	Target group	STR methodology	STR system	General findings
1. Students with cognitive or physical disabilities					
Nisbet & Wilson (2002) Nisbet et al. (2005)	To investigate best practices of STR applications in schools.	Secondary school students with reading, writing and spelling difficulties.	- Students individually used STR system during class; - STR-text was simultaneously displayed to students.	Dragon Naturally Speaking / IBM ViaVoice	- Most students intended to continue using STR for learning purpose; - STR was found as an effective tool to write and record; - In some cases, STR has enhanced basic reading, spelling and writing skills.
Leitch (2008)	To understand whether applications of STR can assist in creating a positive and beneficial learning environment for students.	High school students with hearing impairments.	- The instructor applied STR during lectures; - Lecture transcription was simultaneously displayed to students on a whiteboard/computer screens.	IBM ViaVoice	- STR-texts helped students to understand lectures content better; - Students believed that STR-texts could improve their learning.
Wald (2010) Wald & Bain (2008)	To understand how STR applications may contribute to an improved learning environment for students.	University students with hearing impairments.	- The instructor applied STR during lectures; - Lecture transcription was simultaneously displayed to students on a whiteboard/computer screens.	IBM ViaVoice	- STR-texts helped students to understand lectures content better; - Students believed that STR-texts could improve their learning.
Elliot et al. (2002)	Students' learning strategies to study with STR notes were explored.	High school and college students with hearing impairments.	- The instructor used STR to pre-generate lecture notes; - Lecture notes were delivered to students.	C-Print	- STR-generated notes helped high school students to fill in understanding gaps; - STR notes were useful for college students to prepare for the test and to write research papers.
2. Online students					
Hwang et al. (2010)	The effectiveness of STR applications on students	Open university	- The instructor	Windows Speech	- Experimental students perceived that STR

Hwang et al. (2012) Shadiev (2011)	learning performance during and after one-way lectures in online synchronous cyber classrooms was investigated.	students	applied STR during lectures; - Lecture transcription was simultaneously displayed to students online; - The instructor provided students with edited transcriptions after lecture.	Recognition in the Microsoft Operating System	system was easy to use and useful for one-way lectures and individual learning; - Most experimental students expressed that they were highly motivated to use STR as a learning tool in the future; - Experimental students performed moderately better compared to control students in homework accomplishments; - Experimental students significantly outperformed control students in post-test results.
Kuo et al. (2011) Kuo et al. (2012) Shadiev (2011)	The effectiveness of STR application on students learning performance during and after collaborative learning activities in online synchronous cyber classrooms was explored.	Open university students	- Students applied STR during collaborative learning activities; - Activities' transcriptions were simultaneously displayed to students online; - The instructor provided students with edited transcriptions after activities.	Windows Speech Recognition in the Microsoft Operating System	- Students who used STR-texts (experimental) outperformed students who did not use STR-texts (control) on essays writing, intermediate test and post-test; - Most experimental students perceived that STR was useful for individual presentations and for essays writing; - Experimental students were willing to use STR system for learning in the future; - Experimental students, who obtained transcripts with low accuracy rate and experienced delay in STR-text generation, perceived STR system wasn't easy to use and useful for group discussions.
3. Non-native speakers					
Ryba et al. (2006)	Perceived benefits of STR applications were examined.	University students (native and non-native speakers)	- The instructor applied STR during lectures; - STR-text was displayed on a whiteboard;	IBM ViaVoice	- STR technology has a potential to be an instructional support mechanism; - Most non-native speaker students admitted that STR-texts were useful to

			- STR- text was edited and available for students after lectures.		understand and clarify lectures content and to follow the instructor.
Coniam (1999)	Potentials of STR applications to enhance students' learning English as a second language was explored.	Second language learners (L2)	- L2 learners generated texts from their voices by using STR system; - STR-texts of L2 learners was analyzed and compared with STR-texts of native speakers.	Dragon Naturally Speaking	- STR-texts of L2 learners were less accurate compared to those of native speakers in each category of analysis; - The highest accuracy scores were achieved at the lowest level of analysis, the word level, and the lowest scores at the t-unit, or sentence level of analysis.
Shadiev et al. (in press) Shadiev et al. (2013)	Students' perceptions toward STR applications, the difference between using STR-texts for writing one-week summaries versus immediate summaries, and learning behaviors to use STR-text were studied.	Graduate students (non-native speakers)	- Students applied STR during seminar; - Transcripts generated during seminar were simultaneously displayed to students online; - A speaker provided students with edited transcriptions after seminar.	Windows Speech Recognition in the Microsoft Operating System	- Nineteen learning strategies to use STR-texts were revealed; - Participants employed learning strategies differently; - Participants scored differently in their summary writing assignments; - Most participants perceived that STR-texts were useful for learning; - Low accuracy rate was a problem proposed by some participants.
Shadiev et al. (2014)	Visual attention on STR-text, how differently effective STR-texts can be to influence learning achievement, and students' perceptions regarding usefulness of STR-texts for learning were investigated. Furthermore, visual attention and learning behaviors to different participants' characteristics (i.e., learning ability, learning style preferences, and gender) to use STR-text were compared.	Graduate and undergraduate students (non-native speakers)	- STR-texts were displayed to students on computer screens during two lectures on intermediate and advanced levels.	Windows Speech Recognition in the Microsoft Operating System	-Participants relied on STR-texts more than on video of the instructor and Power Point slides; - Participants made a greater use of STR-texts to enhance their comprehension of the lectures content; - Participants, no matter what levels of their EFL ability, learning style preference and gender are, learned with the aid of STR-texts; - STR-texts significantly helped to enhance learning performance of low ability participants.
Weggerle	To enhance the learning	University	- The	Dragon	- With the help of STR

et al. (2009)	efficiency by applying STR system during lectures.	students (foreign language students and non-attendants)	instructor used STR system during lectures to generated texts from voice input; - STR-texts were edited and provided to students as lecture notes.	Naturally Speaking	system, the lecturer pronunciation and correct grammar improved substantially and thus, improved students learning; - Students perceived that STR-texts are very valuable tool for exam preparation; - STR recognition rate rarely was obtained more than 80 percent, and the delay in real time transcription was disturbing.
4. Students in traditional learning environment					
Ranchal et al. (2013)	The effectiveness of real-time captioning and post-lecture transcription on learning were evaluated.	University students	- The instructor applied STR system during lectures; - Real-time lecture transcriptions streamed on computer screen; - Edited STR-texts were provided to students after class.	IBM ViaVoice and Hosted Transcriptio n Service	- Students benefited from both, real-time lecture transcriptions and post lecture transcriptions; - Real-time lecture transcriptions helped students to pay more attention to the instructor, to take notes, make comments, remarks and dynamically search for specific lecture keywords and time periods; - Students who had access to post lecture transcriptions received higher scores on the quiz compared to students who received real-time transcriptions; - Overall class grades of students who received post lecture transcriptions were higher.
Ryba et al. (2006)	Students' usage of STR-texts and perceptions toward usefulness of STR-texts for learning were explored. Moreover, the advantages/disadvantages of using STR-texts for learning were investigated.	University students	- The instructor used STR system during lectures; - STR-texts were simultaneously displayed on a whiteboard; - Students obtained edited	IBM ViaVoice	- More than 30% of students used STR-texts for learning; - STR-texts were useful to understand lectures, to confirm what was missed in lectures, and to take notes; - Most students complained that the accuracy rate of STR technology was too low

			STR-texts after lectures.		and text generated with many errors could distract their attention from lectures.
Goddard et al. (2007)	How teachers and students use STR system in classroom to support students' learning needs was investigated.	Primary school students	- Students trained STR system to their voices; - Students spoke to STR system and texts were generated from their voices; - Students' speeches were audio recorded; - System read back generated texts and played back students audio recordings.	Read & Write GOLD	- Writing, editing, and rewriting were classroom benefits of the system for students' compositions; - Students' writing improved after they started using the system; - The system helped students to hear and recognize errors that they made; - Spending more time with writing, editing and rewriting improved the final product.
5. Collaborative learning activities					
Wald and Bain (2008)	To understand how STR system may contribute to an improved learning environment.	University students	- The instructor used STR system during lectures; - STR-texts were simultaneously displayed on a whiteboard; - Students asked questions which were repeated by the lecturer to STR system so questions also appeared transcribed on a whiteboard.	IBM ViaScribe	Focus on technological aspects of STR technology: STR system was developed and researchers attempted to improve its accuracy rate. Effects of the system on learning achievement and system's practicality or functionality in pedagogical aspect were not evaluated.
Zschorn et al. (2003)	To develop and evaluate STR system that produces text and audio records of a discussion during meetings.	General group of users	- As the meeting participants speak, STR system generates texts from voice inputs and segments speeches into	Automatic Transcriber of Meetings prototype	Focus on technological aspects of STR technology: STR system was developed and researchers attempted to improve its accuracy rate. Effects of the system on learning achievement and system's practicality or

			utterances; - STR-texts appear on computer screens.		functionality in pedagogical aspect were not evaluated.
Fiscus et al. (2007)	To design and evaluate the Rich Transcription Meeting Recognition.	General group of users	- As meeting participants speak STR system transcribes voice inputs into texts.	The Rich Transcription 2007 Meeting Recognition	Focus on technological aspects of STR technology: STR system was developed and researchers attempted to improve its accuracy rate. Effects of the system on learning achievement and system's practicality or functionality in pedagogical aspect were not evaluated.
Kuo et al. (2011) Kuo et al. (2012) Shadiev (2011)	The effectiveness of applying STR during collaborative learning activities on learning performance was analyzed.	Open university students	- Students used STR system for oral presentations and group discussions; - Speakers took turns to speak; - STR system generated texts from voice inputs and displayed them simultaneously on computer screens; - STR-texts were available to students after learning activities.	Windows Speech Recognition in the Microsoft Operating System	- Applications of STR could facilitate collaborative learning activities as to improve students overall learning performance; - Students who used STR-texts (experimental) outperformed students who did not use STR-texts (control) in two sessions of writing essays, intermediate test and post-test; - Most experimental students perceived that STR system was useful for individual presentations and for essays writing; - Experimental students expressed their willingness to use STR system for learning in the future; - Experimental students who obtained transcripts with low accuracy rate and experienced delay in STR-text generation did not perceive STR system as easy to use and useful for group discussions.