Using Speech Recognition for Real-Time Captioning and Lecture Transcription in the Classroom

Rohit Ranchal, *Member*, *IEEE*, Teresa Taber-Doughty, Yiren Guo, Keith Bain, Heather Martin, J. Paul Robinson, *Member*, *IEEE*, and Bradley S. Duerstock

Abstract—Speech recognition (SR) technologies were evaluated in different classroom environments to assist students to automatically convert oral lectures into text. Two distinct methods of SR-mediated lecture acquisition (SR-mLA), real-time captioning (RTC) and postlecture transcription (PLT), were evaluated in situ life and social sciences lecture courses employing typical classroom equipment. Both methods were compared according to technical feasibility and reliability of classroom implementation, instructors' experiences, word recognition accuracy, and student class performance. RTC provided near-instantaneous display of the instructor's speech for students during class. PLT employed a user-independent SR algorithm to optimally generate multimedia class notes with synchronized lecture transcripts, instructor audio, and class PowerPoint slides for students to access online after class. PLT resulted in greater word recognition accuracy than RTC. During a science course, students were more likely to take optional online quizzes and received higher quiz scores with PLT than when multimedia class notes were unavailable. Overall class grades were also higher when multimedia class notes were available. The potential benefits of SR-mLA for students who have difficulty taking notes accurately and independently were discussed, particularly for nonnative English speakers and students with disabilities. Field-tested best practices for optimizing SR accuracy for both SR-mLA methods were outlined.

Index Terms—Speech recognition, educational technology, electronic learning, multimedia systems, notetaking

1 Introduction

Speech recognition (SR) technology has a burgeoning range of applications in education from captioning video and television for the hearing-impaired, voice-controlled computer operation, and dictation. Some of the most popular commercial applications of SR are for dictation and other hands-free writing tasks with software applications, such as Dragon NaturallySpeaking (Nuance Communications) and IBM ViaVoice. The commercial SR tools are commonly said to achieve 98 percent accuracy but the accuracy of spontaneous speech cannot be assessed in the same way for a number of reasons [1]. Moreover, efforts lead by the Liberated Learning Consortium (LL) (www.liberatedlearning.com) and its members for the past decade have demonstrated that standard commercially available SR software was unsuitable for real-time captioning or transcription of speech [1]. Most commercial SR software applications were developed for dictation with punctuation, not for transcribing extemporaneous speech, which is structurally and grammatically different from written prose. Transcripts produced from a continuous unbroken stream of text are additionally difficult to read and interpret without punctuation or formatting [2], [3]. Alternative SR software applications for real-time captioning or speech transcription have been developed to parse speech into individual transcribed statements using verbal cues rather than having to specify punctuation. Line breaks were introduced using pauses and interjections, such as "um," "ah," and "uh" [3].

These SR technologies have been applied to automatically transcribe instructor's lecture and process the transcription to acquire near verbatim lecture transcripts for students [4], [5], [6]. The benefits of producing lecture transcripts have shown to enhance both learning and teaching. Students could make up for missed lectures as well as to corroborate the accuracy of their own notes during the lectures they attended. Coupled with a recorded audio/video lecture track and copies of the lecture slides, students could recreate the lecture material for replicating the lecture at their own learning pace. These lecture transcripts and additional multimedia recordings also enable instructors to review their own teaching performance and lecture content to assist them to improve individual pedagogy [4]. Likewise, SR has been used for quickly searching certain keywords to retrieve specific text or video lecture content [7], [8].

In this study, we compared the classroom implementation, reliability, and academic performance impact of two

- *R. Ranchal is with the Department of Computer Science, Purdue University, West Lafayette, IN 47906. E-mail: rranchal@purdue.edu.*
- *T. Taber-Doughty and Y. Guo are with the Department of Educational Studies, Purdue University, West Lafayette, IN 47906. E-mail: {tabert, yiren}@purdue.edu.*
- *K. Bain and H. Martin are with the Liberated Learning Consortium, Saint Mary's University, 923 Robie St., Halifax, NS B3H3C3, Canada. E-mail: {keith.bain, heather.martin}@smu.ca.*
- *J.P. Robinson is with the Department of Basic Medical Sciences, Purdue University, West Lafayette, IN 47906. E-mail: jpr@cyto.purdue.edu.*
- *B.S. Duerstock is with the Weldon School of Biomedical Engineering and School of Industrial Engineering, Purdue University, West Lafayette, IN 47906. E-mail: bsd@purdue.edu.*

Manuscript received 25 Sept. 2012; revised 4 Mar. 2013; accepted 28 May 2013; published online 16 June 2013.
For information on obtaining reprints of this article, please send e-mail to: lt@computer.org, and reference IEEECS Log Number TLT-2012-09-0127.
Digital Object Identifier no. 10.1109/TLT.2013.21.

different methods of SR-mediated lecture acquisition (SR-mLA). Both SR-mLA techniques were employed using conventional educational technology found in contemporary university lecture rooms. The first method of SR-mLA provided real-time captioning (RTC) of an instructor's lecture speech using a client-server application for instant viewing during class on a projection screen or directly to the students' laptop personal computers (PCs). The second SR-mLA method, postlecture transcription (PLT), employed a digital audio recording of the instructor's lecture to provide transcripts, which were synchronized with the audio recording and class PowerPoint slides for students to view online or download after class. Both of these SR-mLA approaches—real-time captioning and postlecture transcription, respectively, were compared as case studies using a cognitive walk-through approach with instructors. The more robust PLT method, as demonstrated by greater word recognition accuracy and more flexible use of multimedia class notes, was compared in both social and life sciences courses. Best practices for attaining superior recognition accuracy were created to translate to other classroom environments and for use with other SR algorithms.

Lastly, PLT was evaluated in a life sciences course as a pilot study to gauge student usage and reaction and assess student class performance when the multimedia class notes were present and when unavailable. Previous studies using some type of SR technology to convert the instructor's speech to text were limited to assessing teacher and student attitudes toward lecture transcription or captioning. These studies were conducted in courses other than science, technology, engineering, and mathematics (STEM) and did not attempt to quantitatively measure the effects of providing SR-based lecture notes on student class performance [4], [5], [6].

Though SR has been used for academic purposes related to text searching and retrieval of lecture content [7], [8], to our knowledge class performance using SR-mLA has not been quantitatively evaluated before. Doing so in a STEM course presents unique challenges during lectures, such as using substantial scientific nomenclature and technical terminology and slides with significant graphical content. The relevance of SR-mLA tools for instructors and students, particularly those with special needs, were discussed in depth for future study.

2 Background

2.1 Notetaking for Students with Special Needs

Though audio recording of class notes is easily achieved with a student tape recorder or a classroom recording system provided through the school, acquiring the actual text transcripts of class lectures is vitally important. The addition of text transcripts versus purely audio enhances accessibility [9], [10]. Pure audio recordings can be effective for nonnative English speakers and students with learning disabilities who may simply need more time to comprehend the instructor's oral lecture. However, many students with disabilities cannot write lecture notes without human assistance. Thus, SR-mLA as a means of providing lecture notes is critically important for students who have difficulty taking lecture notes effectively themselves. Students with hearing impairments or nonnative English speakers cannot process the audio of oral lectures. Students with mobility or visual impairments cannot easily translate the teacher's speech into text. Students with learning disabilities may have difficulty with both aspects of lecture-to-lecture note translation [1], [2], [3]. The automatic SR-mLA approaches do not depend on the current standard of employing a captionist to caption instructor's speech into text during class or notetaking services for writing lecture notes [11], [12], [13], [14]. Indeed, class notetaking assistance is one of the most requested school accommodation for SWDs [11], [15]. SWDs often have problems with either the comprehension of the instructor or with the act of quickly and tirelessly writing notes. For students with significant hearing loss, captionists are traditionally employed to interpret and transcribe what the instructor is lecturing [11], [12], [13], [16]. Likewise, students with mobility (i.e., quadriplegia/tetraplegia), learning (dyslexia, attention deficit disorders), or sensory (low vision, hard of hearing) impairments may not be able to take notes by themselves and must rely on hired notetakers to acquire lecture notes [12], [17], [18]. Some students with learning disabilities may find it difficult to process information presented orally [19], [20]. Studies showed students without disabilities recorded up to 70 percent more lecture information than students with learning disabilities [21]. Blind students who are completely reliant upon audio recordings of lectures must constantly fast-forward and rewind to find specific lecture topics. Although there are some novel techniques for efficiently interacting with audio only, this is traditionally a very time-consuming method as opposed to searching for specific keywords in a text document [22].

Nonnative English speaking students struggle with lecture content delivered in auditory format, typically having greater exposure to English language in print form [23], [24]. SR-mLA has been shown to help international students who are nonnative speakers with regard to the language of instruction (in this case English) to acquire accurate lecture notes [5], [25]. Many typical students, regardless of ability or native language skills, can also experience difficulties with notetaking tasks to some degree for certain courses or situations.

2.2 Notetaking in STEM Subjects

Notetaking is a fundamental and ubiquitous learning activity that students are expected to perform and master during their educational development [26], [27], [28]. The benefits of lecture notetaking include for students to organize, summarize, and better comprehend lecture information, recording content for later studying, self-regulated learning through the active process of notetaking, and simply staying attentive during class [26], [29], [30], [31], [32].

STEM courses, in particular, require substantial notetaking due to the large amount of class information and content specific terminology presented during a relatively short time period [33]. Undergraduate students in the United States overwhelmingly (99 percent) stated that they must take notes during science classes [34], [35]. Science courses, such as histology and biochemistry, were selected

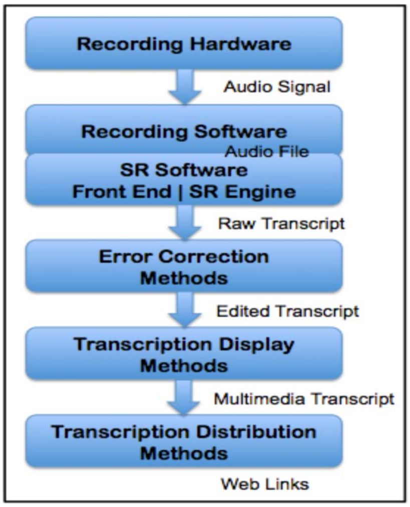


Fig. 1. General SR-mLA methodology.

by students as classes they would most want assistance with notetaking [33]. College students typically record less than 40 percent of the information presented when lecture notetaking [28]. Students cited a tradeoff between taking complete notes and paying full attention to the teacher. This means students can spend much of their time and mental effort on notetaking. About half of undergraduate students try to copy lecture notes literally to ensure accuracy of what was spoken by the instructor, while the other half paraphrase lecture information to ensure they follow the teacher's explanations during class [31].

3 Objectives

The objective of this study was to explore the practical applications of implementing SR technology in the classroom to automatically convert instructors' speech to text. In-classroom captioning and the provision of lecture transcripts can serve a variety of educational or pedagogical purposes, including supplementing students' class notes with accurate SR transcripts, allowing students who have difficulty taking notes by themselves to acquire the lecture content, and permitting instructors to confirm what they said during class.

Two different methods of SR-mLA were compared during different course subjects as a technical feasibility and case study. Our specific objectives were to:

- identify the issues regarding the use of SR-mLA as a standard classroom tool in capturing spoken lecture information,
- compare the technical reliabilities and word recognition accuracies of RTC and PLT,
- explore the effects of SR-mLA on student class grades and voluntary quiz participation and performance in a regular STEM course, and
- investigate the potential benefits of acquiring SR lecture transcripts for instructors and students, particularly for those with special needs.

It was not the purpose of this study to compare the many different SR algorithms or software currently available but the practices of SR usage in lecture capture. We believe the strategies and framework of this study could be applied to include any SR engine that others may wish to employ.

TABLE 1
Comparison of Major Functionalities between ViaScribe and HTS Systems

Features	ViaScribe (RTC)	HTS (PLT)
Recording method	ViaScribe has in-built recording capability through classroom PC that receives audio from wireless headset microphone worn by lecturers.	This system runs software on classroom PC to record lecture audio in a SR-compatible format.
User-dependent/independent SR engine	Initial user-specific profile training is necessary to accustom the SR software to the speaker's voice to maximize word recognition accuracy.	No user training is required. The speaker wears a high-quality microphone to record a lecture audio file.
Error correction	The generated raw transcripts can be corrected offline after the lecture in order to continually update the user's voice profile.	Recorded audio file is uploaded to a HTS website for transcription and correcting errors via user interface.
Display method	Classroom PC runs server software during class to transmit captioning in real-time to student laptops PCs as clients or to a classroom projection screen.	Lecture transcripts can be synchronized with audio and PowerPoint slides and put online as multimedia notes.

4 SR Methods and Tools

Two different SR approaches for SR-mLA were used; 1) the first approach was RTC using IBM ViaScribe and 2) the second was PLT through IBM Hosted Transcription Service (HTS). Both software systems were available through Purdue's institutional membership in the LL Consortium, which also provided technical support. The selection of SR tools was based on the results of previous studies [4], [5], [6]. The general steps for performing both methods of SR-mLA are outlined (Fig. 1). In the following sections, the various SR software applications and utilities employed by each SR-mLA approach are described in detail to facilitate replication and testing by others.

Table 1 summarized and compared the major technical functionalities between real-time captioning using ViaScribe and postlecture transcription using HTS. Functionalities for comparison were divided according to:

1. process of recording instructor's speech,
2. speaker-dependent or -independent SR engines used,
3. error correction methods, and
4. display options.

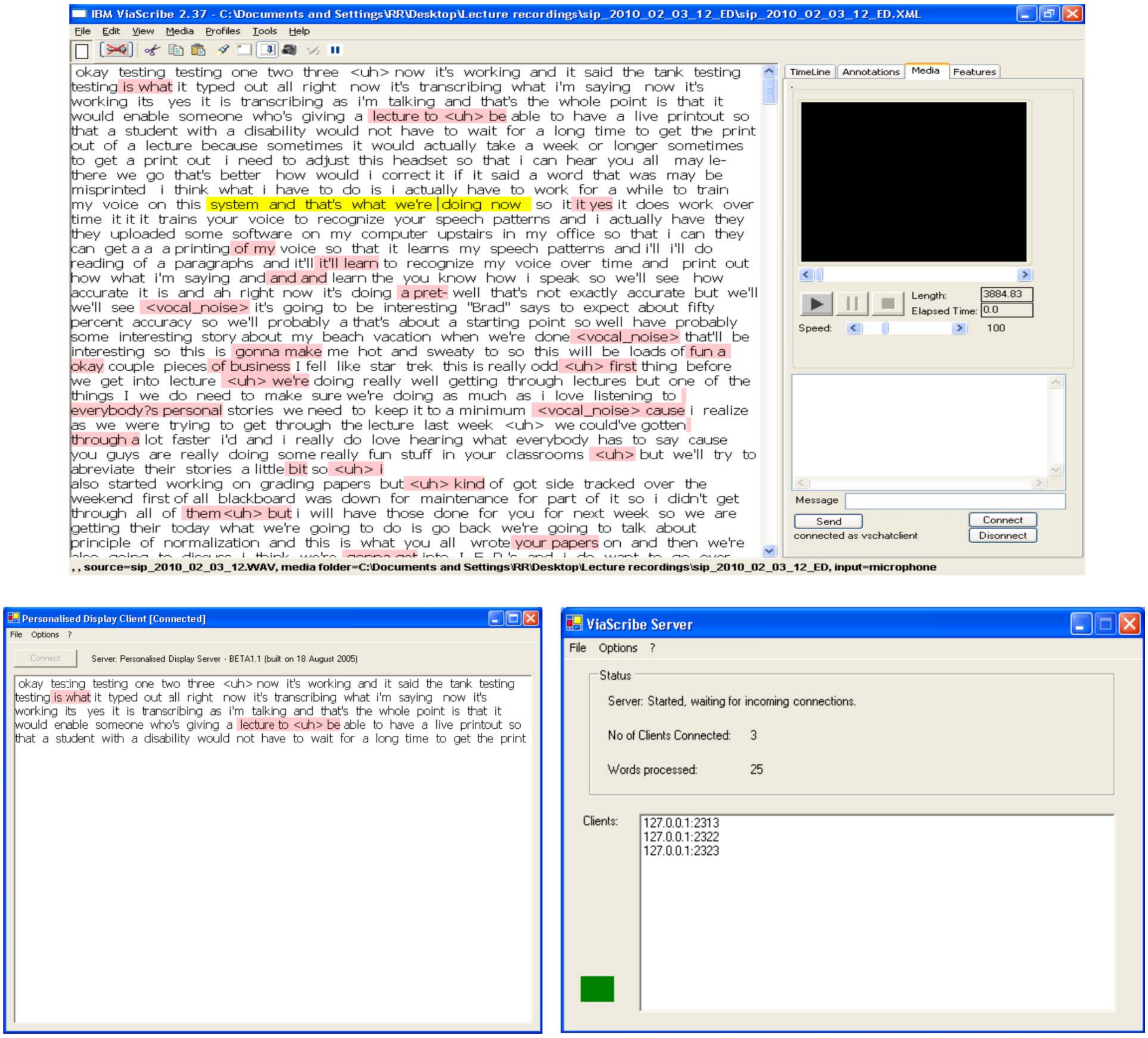


Fig. 2. a) ViaScribe window displaying real-time captioning with audio playback in synchrony. b) Client window connected to the instructor's server for in-class streaming. c) Client-server monitoring window showing connected students during class.

4.1 IBM ViaScribe

IBM ViaScribe utilized a SR engine capable of transcribing live or prerecorded speech developed collaboratively by IBM and the LL Consortium. During class lectures ViaScribe displayed or captioned what the instructor uttered into text as it is being spoken. ViaScribe (version 2.3.7) was chosen for real-time captioning, because it had a proven track record by LL members for reliable captioning and had a client-server platform for streaming live transcription to students' laptop PCs during lectures [3], [4], [36].

As natural spoken language does not explicitly state grammar and punctuation; ViaScribe transcription provided readability by introducing a paragraph break or other markers whenever the speaker paused for a breath. These pauses could be customized according to the speaker's individual speech characteristics.

To improve word recognition accuracy, users performed voice profile training. The commercial ViaVoice (version 10) application was used to create the initial voice profile and subsequently updated to the ViaScribe application. Profile training involved recording a minimum of 500 words of dialogue and vocabulary for proper speech recognition. Once the initial voice profile was performed, it could be updated by adding lecture transcripts that had been recorded and corrected by us. As more words were inputted, word recognition accuracy improved. However, inputting more than 2,500 words for profile training does not significantly improve word accuracy (LL, unpublished results).

ViaScribe can save the audio and PowerPoint slide images displayed during class for creation of multimedia files (SMIL, XML, WAV, RT, RTF) that could be published online to permit students to view and download lecture information in a format according to their individual learning preferences. ViaScribe offers the ability to listen to the audio lecture track and read and search the raw textual transcription immediately following the presentation. The transcript can also be edited to correct recognition errors and posted online (Fig. 2a).

In addition, students could voluntarily install client software on their own personal laptops during class receiving text as it is being streamed by the ViaScribe server (Fig. 2b). There is an inherent delay between when a word is spoken and when it is transcribed (regardless of whether SR is used or human captionists are employed). However, ViaScribe used a single pass decoding technique, which generated very little display lag compared to other SR systems that use different decoding techniques. A client-server monitoring application on the instructor's machine showed the current client connections, which could be deactivated, and the rate of streaming words from the server (Fig. 2c).

4.2 IBM Hosted Transcription Service

IBM HTS was selected for postlecture transcription primarily because of its higher word recognition accuracy rates compared to other systems [37]. HTS is a speaker-independent SR system developed by IBM Research that automatically transcribes a variety of standard audio or video file formats through a cloud service. HTS uses statistically derived acoustic and language models to convert speech to text. As opposed to statistical models designed for creating written language, which would not be ideally suited for recognizing extemporaneous speech, HTS used United States English Broadcast News models built from acoustic data from spoken language. HTS' SR engine employs a double-pass decoding technique, which dynamically adjusts to the speaker's voice, without requiring voice profile training or enrollment [37].

For HTS transcription, authenticated users had to visit the HTS service portal, log into their secure accounts, and then upload a media file for automatic transcription. Once HTS has processed the recorded lecture, the transcribed text could be viewed and edited for error corrections online employing a Flash-based interface. A posthoc correction method similar to ViaScribe was performed. Afterward, the audio recording in synchrony with the transcript could be downloaded. This multimedia content could be viewed in different predefined layouts and adjusted temporally by authors using postproduction tools provided by HTS. The presentation package was downloaded from HTS, which consisted of an XML file with timing data, audio WAV file, and lecture transcript to prepare the multimedia transcript.

4.3 Utilities and Tools

The SR-mLA system required the use of software utilities to record lecture audio, generate lecture transcripts, and host them online as multimedia class notes. The following sections describe such tools.

4.3.1 Recording Lecture Audio

The instructors' oral lectures were recorded during class using an Audio-Technica (Ohio, USA) 700 Series Freeway 8-channels UHF wireless microphone system. The receiver was connected to a notebook Windows PC via the mic-in jack. Instructors wore an Audio-Technica Pro 8HE hypercardioid dynamic microphone headset with wireless transmitter.

Audacity (audacity.sourceforge.net), open source software, was used for recording lecture audio for postlecture transcription during Phase 2 studies. Audacity provided various configuration options for audio recording in SR-compatible format settings (ADPCM (or PCM) WAV, 22,050 Hz or higher, 4-bit or higher, mono).

Microsoft PowerPoint 2007 was used as an alternate method for recording lecture audio and slide images for subsequent postlecture transcription during Phases 3 and 4. Instructors would deliver their PowerPoint slideshow presentation during class with their speech saved as WAV files along with each slide image and the slide timings. PPT to XML converter software converted PowerPoint 1997-2003 format (ppt) files and generated a set of images for the slides, an audio file, and a XML file that contained timings for synchronization. The timings were required to synchronize audio, text, slides, and for generating synchronized multimedia class notes [9].

4.3.2 Display of Multimedia Lecture Notes

Synote (www.synote.org/synote), a web-based application, created synchronized bookmarks or "Synmarks" that contain notes and tags to synchronize audio or video recordings, SR transcripts and PowerPoint slide images [10]. These individual files are compiled into a synchronized multimedia web resource for easy access, searching, and management by students and teachers. Synote displayed these synchronized, multimedia class notes online through a web interface for easy access (Fig. 3). Registered users can save their own remarks and notes for each set of class notes.

The links to access online multimedia transcripts were posted on Blackboard. Blackboard is a commercial web-based course management system accessible to students for taking quizzes, accessing lecture notes, and viewing their class grades online.

4.3.3 Word Error Rate Tool

A Word Error Rate (WER) evaluation tool developed by IBM is a command-line utility that compared a raw "decoded" SR text transcript to an edited "reference" text transcript. The input transcripts were plain text files with ANSI encoding. The WER tool computed a range of error results, including substitutions, deletions, and omissions for calculating accuracy. WER is defined as the percentage of total errors for the reference transcript or total number of words spoken. Accuracy is the percentage of correct words to total words transcribed [36]. WER percentages for each instructor were averaged from at least two lecture hours to a maximum of five lecture hours. Total words referenced for this study was 117,786 or approximately 19 hours of lectures.

4.4 Statistical Analysis

IBM SPSS (Statistical Product and Service Solutions version 17.0) software was used to compare WER and other metrics and student class performance. derived Paired Student's t-tests compared class performance during access to multimedia class notes (experimental phase) and without (control phase). Paired samples correlations were also calculated using t-tests and ANOVA for individual student's scores and WER analysis. Statistical significance was determined at $p = 0.05$ and $p = 0.001$ for extremely significant differences.

5 Classroom Evaluation Procedures

Both SR-mLA systems were evaluated during four phases of in-class testing during both social science and life science lecture courses at a public university to explore diversity in lecture material and teaching styles. We used a case study methodology incorporating cognitive walk-throughs with instructors, instructor interviews, and investigator observation to evaluate technical implementation of SR-mLA, or

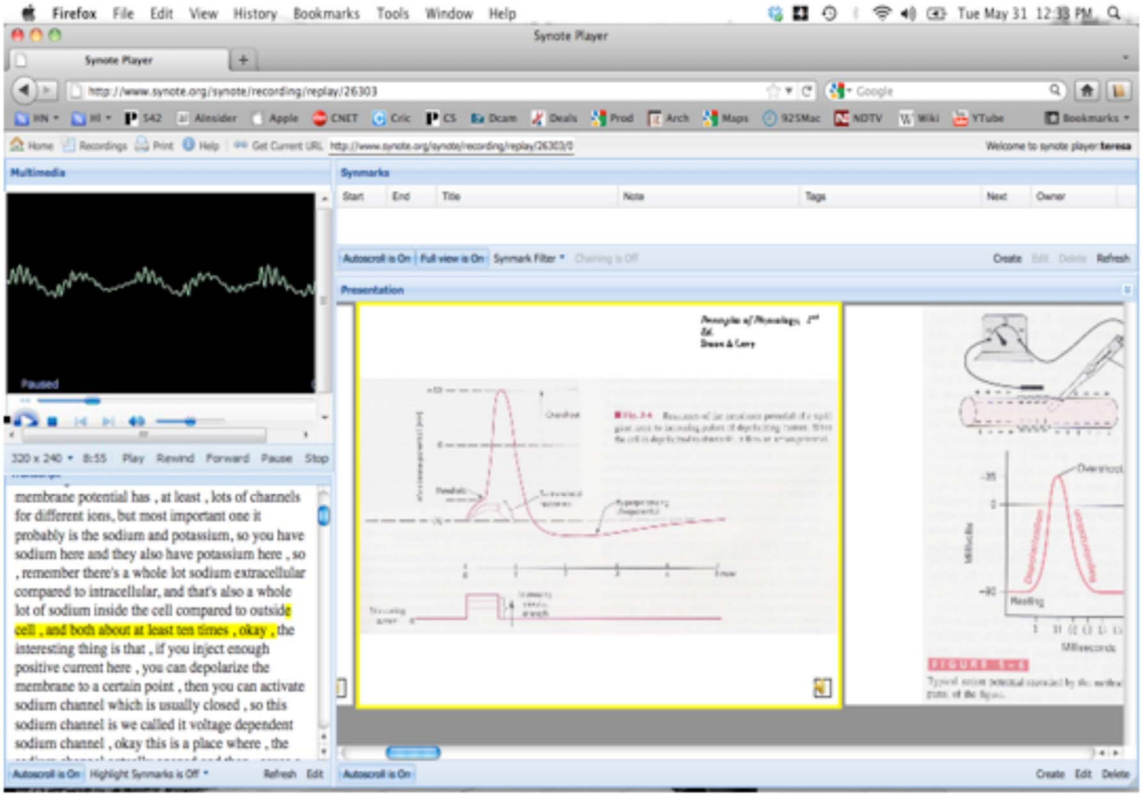


Fig. 3. Multimedia class notes in Synote.

word recognition accuracies. The lecture notes were corrected by graduate students not taking the classes.

Prior to evaluation, instructor testers were trained to operate the necessary software and provided initial best practices for improved SR accuracy before starting the real-time captioning or postlecture transcription.

5.1 Phase 1: Initial Evaluation with RTC

Phase 1 evaluation occurred during regular courses offered in two graduate-level special education classes taught by a female instructor in the College of Education. The goal of Phase 1 testing during these semester-long courses was to successfully install and run the ViaScribe system, assess this technology with feedback from the instructor and students, and collect data on the accuracy and feasibility of real-time captioning. Prior to use, the instructor underwent initial voice training to develop a voice profile for the ViaScribe system.

In Fig. 4, a schema of the RTC system used during classes is given. Briefly, the instructor used a portable, laptop PC running the ViaScribe server program during class. The laptop was then connected to the wireless microphone receiver and digital classroom projector. During class, the ViaScribe server program streamed textual captions, as it was being spoken and processed by the SR engine, to the classroom projection system and screen or to students' laptops running client software during class to serve as a closed captioning window (Fig. 2b). As clients, students acquired raw, unedited transcripts. Afterward students were provided with corrected transcripts. The instructor could provide additional information and create multimedia presentations to post on the web with keyword-searchable transcripts synchronized with the digital lecture audio (Fig. 2).

5.2 Phase 2: Initial Evaluation with PLT

PLT was assessed during the same education courses from Phase 1 with the same instructor. As shown in Fig. 5, PLT was deployed by digitally recording the lecture audio with software installed on the instructor's laptop. The lecture audio was recorded in a SR-compatible format using a wireless microphone system. After class, the lecture audio file was uploaded to the online HTS system to be automatically transcribed through the IBM SR engine. Once the HTS system finished processing, a notification e-mail was sent to the HTS website account holder. This process could take hours to a full day depending on the index of submission in the queue of jobs submitted by other users. The transcribed text was then corrected for errors by a trained graduate student using the HTS Flash interface. The editor would play the lecture audio while editing the transcribed text automatically advancing in synchrony.

The audio recording and generated transcripts were automatically synchronized through HTS; however, this content could be disaggregated to text only. The multimedia transcripts (synchronized text and audio) were uploaded to the university Blackboard system for students to download, search, print, or playback the audio using third-party applications if desired.

5.3 Phase 3: PLT Evaluation during a Science Course

PLT using HTS was alternately evaluated during the lecture portion of a graduate neuroanatomy course in the College of Veterinary Medicine rather than a social science course.

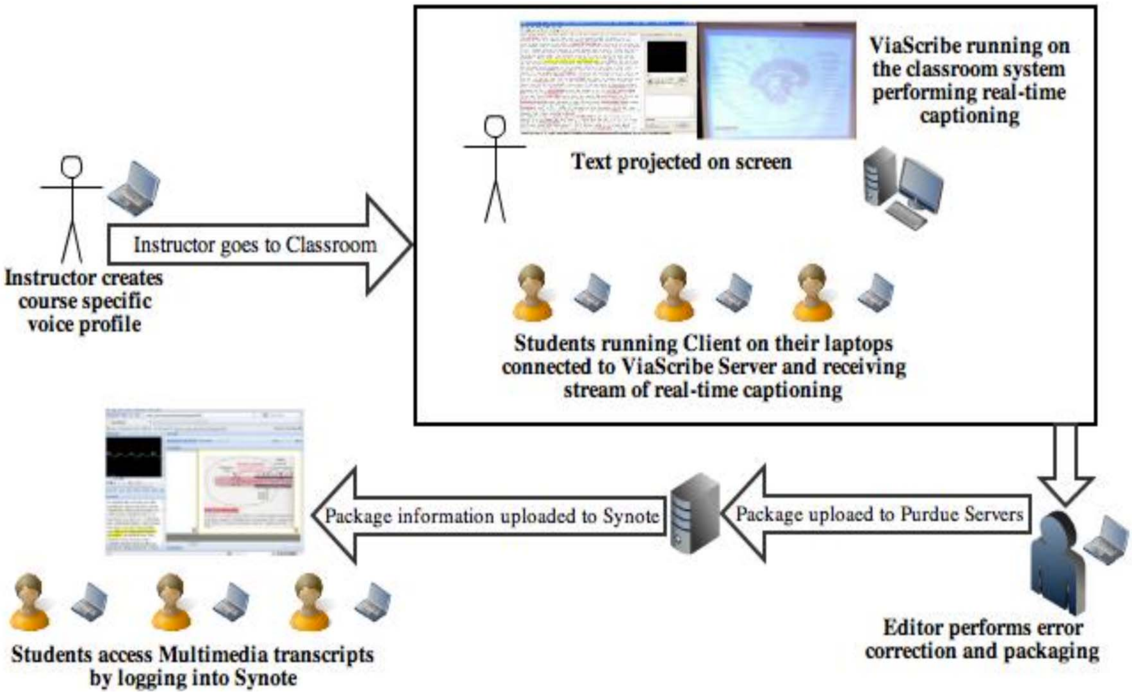


Fig. 4. High-level overview of the real-time captioning method using ViaScribe.

Likewise, this course was taught by a male instructor rather than a female instructor, which was performed previously (Fig. 5). The objectives of the Phase 3 evaluation were to 1) compare SR-mLA implementation in a science versus social science course, 2) compare instructors of different genders, and 3) assess a new method of providing students SR class transcripts by synchronizing them with the lecture audio and class PowerPoint slides to generate comprehensive multimedia class notes.

However, Microsoft PowerPoint was used to record the lecture audio and class slides simultaneously during class. The speech was automatically saved with each slide according to the time instructor spent on a particular slide. The PowerPoint file was separated to generate a set of slide images, an audio file, and XML file, which contained data on slide timings for synchronization. The audio file was subsequently uploaded to the HTS system for transcription. Transcript error correction was performed through the HTS website.

A new multimedia package synchronizing transcribed text, audio, and individual slide images was created and uploaded to Synote for students to view at their convenience. Students had to logon to Synote to view these multimedia class notes. Through Synote, students could

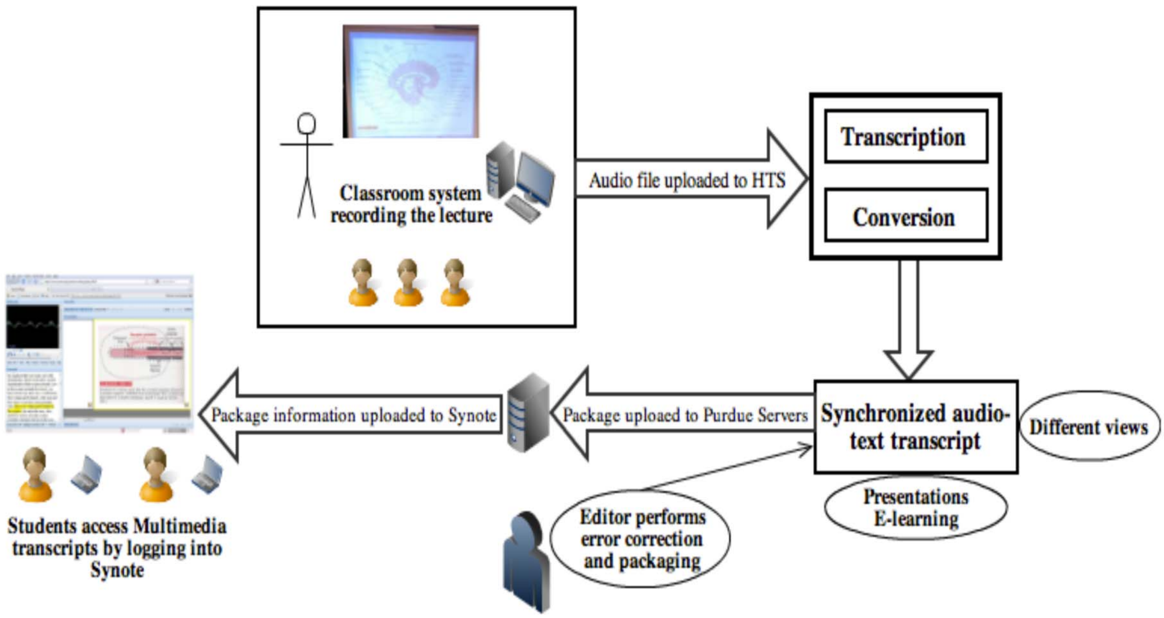


Fig. 5. High-level overview of the postlecture transcription method using HTS.

select, modify, annotate, or search the multimedia class notes according to their individual preferences (Fig. 3).

5.4 Phase 4: Evaluation of Student Class Performance in a STEM Course

Outcomes for quiz scores, exam grades, and student satisfaction of PLT were assessed in a preliminary study of the small cohort of nine students in a team-taught graduate level course in systemic mammalian physiology at the College of Veterinary Medicine. A nonmandatory online five-question quiz was posted each week for 12 weeks to course's Blackboard website covering the previous week's lecture material. Students were invited to voluntarily perform these quizzes and were informed that these quiz scores did not affect their course grades. Multimedia class notes with lecture transcripts were available to students to view on Synote for the first six weeks of the course or 16 lectures (experimental period). The multimedia class notes were unavailable for the next six weeks (control period). The most obvious incentive for students to voluntarily take online quizzes was to ensure their understanding of the lecture information. Student questionnaires regarding use of multimedia class notes and notetaking were completed by students at the end of the course.

The systemic physiology course was divided into separate topics covering blood, muscles, and the nervous and digestive systems. Throughout the course students were provided hardcopies of the lecture PowerPoint slides and lecture notes that the instructors would normally hand out when teaching this course. During the time period, students had access to multimedia class notes, they had two class exams totaling 160 points. When multimedia class notes were unavailable, students had three exams worth 170 points in total. The total exam points, quiz scores, and combined scores were compared between time periods to assess changes in class performance.

6 Observations

6.1 Technical Issues Performing RTC and PLT

The first issue concerning implementing either RTC or PLT in the classroom was to determine how best to integrate the system with the university-managed classroom technologies. The university's information technology department had to give permission in installation of ViaScribe and sound recording software on the classroom PC. Thus, PLT employing PowerPoint for lecture audio and slide recording proved easiest to set up.

During Phase 1 testing, the instructor experienced frequent freezes or crashes while recording and transcribing lectures on the classroom PC. We found that ViaScribe (developed in 2000) was unreliable with Windows 7, resulting in unexpected stalls and missed portions of lecture transcription. A laptop PC installed with Windows XP was used by the instructor instead of the classroom PC to run ViaScribe. ViaScribe's incompatibility with newer OS also limited its use by some students interested in running the ViaScribe client software on their own laptop PC during class. The majority of students in the class did not use a laptop PC.

Another initial logistical problem was inadequate audio quality from resident lecture podcasting offered by the university through the in-classroom audio recording system. Thus, we decided to use a wireless microphone system connected to the instructor's laptop to receive the lecture audio to the SR software for both RTC and PLT approaches. Instructors wore a wireless headset microphone during lecturing.

During Phase 3 testing, the instructor used the same PowerPoint presentation that he has in the past while saving his audio recording during lecture. One limitation, however, of recording while presenting the slides through PowerPoint during class was that the instructor could not go backwards or stop the slideshow presentation or the slide audio would be overwritten. Thus, instructors had to proceed sequentially through their lecture slide set for complete recording and transcription. A more robust software would rectify this limitation and will allow for a more natural style of presentation.

6.1.1 Best Practices for SR-mLA

The best practices for optimal word recognition accuracy for both SR-mLA methods were developed based on prior experiences of expert users and modified by instructors' responses and observations during the study [1], [2], [3], [4]. These generalizable best practices were shown to be important for achieving high levels of recognition accuracy. Past evidence shows that audio quality is directly correlated with recognition accuracy; thus, incorrect microphone positioning leads to poor quality audio and speaker training improved recognition accuracy [1], [2], [3]. These best practices for classroom use are as follows:

- The words should be clearly articulated by the speaker, especially word endings.
- Correct microphone positioning is important to avoid "breathiness." This is even more important when transcribing a live lecture or doing profile training.
- The speaker should not speak too fast or too slow but keep a relaxed, natural rate.
- If profile training is to be performed, commonly used course-specific keywords that would not be in dictionaries should be trained.
- Speakers should periodically break during long lectures to check the reliability of the SR system.
- Speakers should try not to look at the transcribed text while recording.
- Only the speaker's voice is reliably recorded. For example, if responding to student questions, the instructor should either repeat the question and then respond or pause recording.
- When gesturing or making other nonverbal cues to demonstrate a point; make sure to describe what is being done during transcription.

6.1.2 Word Error Correction

Correcting word recognition errors was the most time-consuming task for both real-time captioning and postlecture transcription. Error correction was performed using editing tools available with ViaScribe and HTS, respectively. Essentially, both approaches required an editor to listen to the lecture audio recording and change the transcribed text by correcting misrecognized words, inserting missed words, or deleting superfluous wording. Correcting transcripts with high error rates required considerable human effort. Graduate students unfamiliar with the course content did the editing. For a typical 1-hour

TABLE 2
Average Word Error Rates and Word Recognition Accuracies

Trial & SR-mLA Method	Word Error Rate (before, after) training	Recognition Accuracy
Phase-1 (RTC)	45%, 22.2%	78.8% (after training)
Phase-2 (PLT)	14.7%	85.3%
Phase-3 (PLT)	9.1%	90.9%
Phase-4 Instructor 1 (PLT)	18.0%	82.0%
Phase-4 Instructor 2 (PLT)	10.8%	89.2%

lecture with an error rate of over 20 percent, approximately 4 hours of editing for a single editor was required depending on the number of errors and whether the editor corrected for misrecognized words as well as for punctuation and formatting. Editing can be accomplished by teaching assistants or collectively by students in the class through the online correction application. The latter method shares the workload among several people [18].

6.2 Comparison of Word Recognition Accuracies

In Table 2, the average word recognition accuracy was compared using ViaScribe before voice profile training (45 percent) to after training (22.2 percent) for the same instructor. Such profile training was essential to improving word recognition accuracy.

Despite prior voice profile training, the SR accuracy rate for RTC (78.8 percent) was lower than speaker-independent PLT (85.3 percent) by the same instructor (Table 2; Phases 1 and 2). This was due in part by the immense processing load required for real-time captioning during class. The HTS engine used for PLT performed a double-pass recognition process that improved word recognition accuracy significantly.

Though PLT was performed by a female instructor in a social science class (Phase 2) and by a male instructor in a life science course (Phase 3), their total word recognition accuracies 85.3 and 90.9 percent, respectively, were somewhat different. During Phase 4, Instructor 1 was a nonnative English speaker with a recognition accuracy of 82.0 percent. Instructor 2 was a native American English speaker with an 89.2 percent word recognition accuracy (Table 2).

We also wanted to determine if other factors affected word recognition accuracy, such as fatigue. We divided lecture transcripts generated by PLT into eight samples consisting of the raw number of words and the edited number of words and calculated WER. When comparing the first, third, fourth, and seventh WER samples over five lectures, we found no significant changes over time ($r = -0.0753$, $p = 0.75$).

6.3 Classroom Display Methods

For RTC during lecture, one projection screen could be used to display the instructor's PowerPoint class slides and another to show the lecture captioning in real time (Fig. 4). However, during Phase 1, the classroom used only had a single projection screen, which was dedicated for displaying class slides. Hence, we had to suspend displaying RTC during class.

For PLT lecture, transcripts were combined with the lecture audio and the PowerPoint class slides to generate comprehensive multimedia class notes during Phase 3, Adobe Presenter was tested to demonstrate slides during class; however, HTS did not support Adobe's Flash format and would not convert the slide images for producing multimedia class notes. Therefore, only the lecture audio and HTS-generated text transcripts could be synchronized for students.

During Phase 4, the multimedia files combining audio, transcripts, and PowerPoint slide images were viewable through a website, called Synote, to benefit from additional features such as making comments and remarks, having content searchable by keyword, slide, or class time displayed in synchrony (Fig. 3).

6.4 Student Performance Using Multimedia Class Notes

6.4.1 Student Class Performance

When students had multimedia class notes (with transcripts included), their mean quiz score was 38.8 percent. This was statistically significantly higher than the average score of 23.7 percent when multimedia class notes were unavailable ($n = 9, p = 0.032$). With multimedia class notes, the percentage of total scores for exams 1 and 2 averaged 91.5 percent, which was extremely significantly higher than total exam percentages for exams 3, 4, and 5 (81.3 percent) without multimedia class notes ($p = 0.00032$). In addition, the combined quiz and exam scores averaged 83.2 percent when multimedia class notes were available and was significantly higher than without, which averaged 72.7 percent ($p = 0.00023$) (Fig. 6).

Eight out of nine students voluntarily participated in taking the online quizzes, which tested students' recall of the previous week's lecture content. Quiz taking varied among students between time periods. Students performed an average of 55.6 percent of the quizzes when multimedia transcripts were accessible, while students took an average of 39.0 percent of quizzes when these transcripts were not provided. However, since three students performed more than 80 percent of all quizzes during both phases, no statistical significant difference was determined for class quiz participation ($n = 8, p > 0.05$).

Paired samples correlations were calculated for individual student's quiz participation and quiz scores to the presence of the multimedia class notes. There were statistically significant relationships between greater quiz participation ($n = 9, r = 0.728, p = 0.026$) and higher total quiz scores among the cohort of students when multimedia class notes were available ($r = 0.790, p = 0.011$).

6.4.2 Student Experiences with Multimedia Class Notes

Eight students returned our questionnaire. All respondents felt notetaking was important when taking this class. Seven students stated using the synchronized multimedia class notes. The features of the multimedia class notes that they found especially beneficial were the ability to repeat the

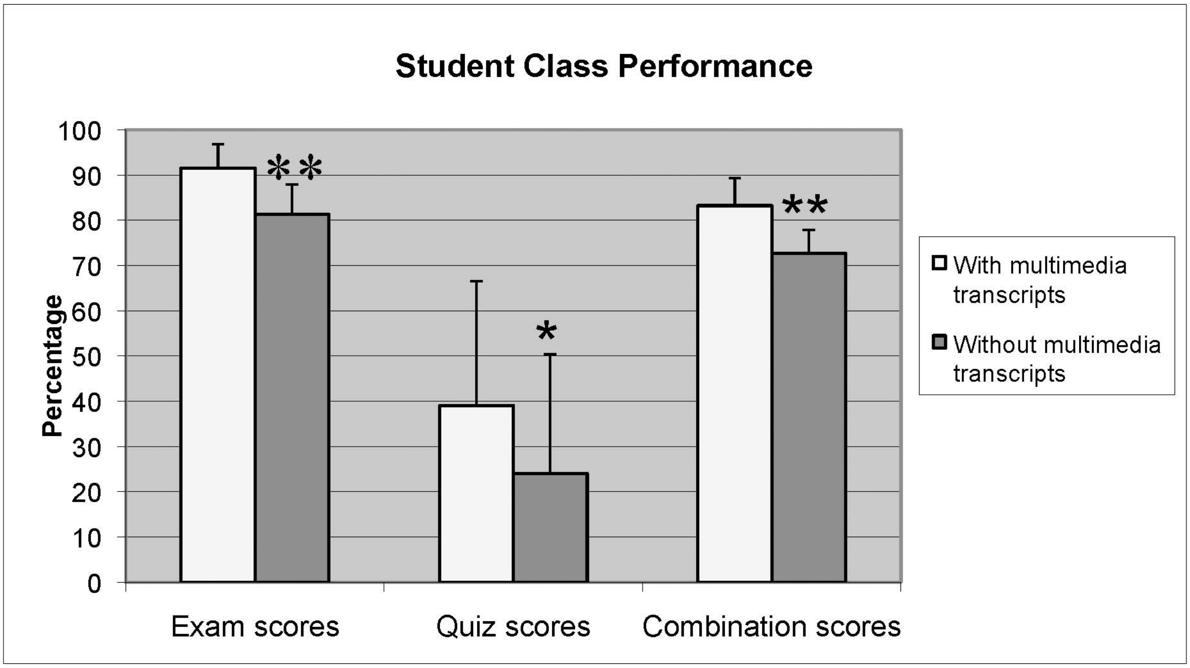


Fig. 6. Differences in class performance with multimedia class notes. Error bars indicate standard deviation. * indicates $p < 0.05$, ** indicates $p < 0.001$.

lecture content as many times as needed, the synchronization of the lecture audio and text with the slides, and being able to dynamically search for specific lecture topics. The aspects that students found inconvenient were inaccuracies in speech-to-text transcription and having to register and log into a website to access the multimedia notes.

Most students stated having access to multimedia class notes did not change their typical manual notetaking habits during class. However, three students responded that they could pay more attention to the instructor than worry about capturing all the lecture information.

The nonnative English speaking student was the most positive about having access to the multimedia class notes. She stated that with the notes she "can learn many times as long as you want and could follow notes which not followed during class." She also claimed that she "can pay more attention listening, not have to write down notes during class."

7 Discussion

Current applications for SR technology have focused primarily on document dictation; which requires punctuation, capitalization, and syntax to be specified, discrete word entry, and voice user interface control (e.g., call systems, home automation, driver control of vehicle features). SR-mLA provides an ideal model for studying continuous SR whereby extemporaneous speech by a single speaker (lecturer) is transcribed for student use in a controlled, noise-limiting environment.

By investigating the technical challenges and merits of implementing SR-mLA in typical university classes, we can evaluate its current feasibility and develop practical strategies for incorporating this technology in other educational environments. In this pilot study, we chose two different SR-mLA methods, real-time captioning and postlecture transcription, using SR engines we knew to be reliable and accurate. Other SR engines could be substituted within these RTC or PLT frameworks.

7.1 Two Approaches for SR-mLA

We defined the technical specifications of testing SR-mLA within the context of a public university system using contemporary classroom educational technology. Because of its unique nature, it is challenging to define a universal approach that would work in all settings. However, any specification would require the following core components (Fig. 1): recording hardware and software, SR software, error correction methodology for transcription accuracy, strategy for transcription display, and a usable distribution/publishing platform. Each should be evaluated to find the right set of components, which can be fine-tuned to work best together in a given environment. We asserted that college lectures would be a suitable environment for evaluating innovative learning technology.

Despite documented hurdles of implementing new technologies in classroom settings [38], it was feasible to set up and execute RTC and PLT in concert with existing classroom audio-visual equipment and a high-quality microphone during single-instructor and team-taught undergraduate and graduate courses. These methods were inexpensive and relatively easy to set up prior to class. Both SR-mLA technologies did not interfere with typical lecture teaching activities, only requiring the instructor to wear a wireless microphone. Instructors teaching large lecture classes were used to wearing wireless microphones for voice amplification.

During our case studies, instructors were able to successfully integrate lecture transcripts in the different educational environments of social and life science courses for students to access online. The most time-consuming task of furnishing lecture transcripts was editing for errors. Error correction serves two purposes: 1) to improve the readability of the lecture transcripts for student use and 2) to enhance the accuracy of future SR for speaker voice profile improvements and building better acoustic and language models to be shared with other LL Consortium members [39]. For RTC, speaker-dependent profile training was necessary to increase word accuracy (by an additional 20 percent in our case. Profile training required approximately one hour of training prior to teaching, then future lecture transcription files could be used to improve the instructor's profile. Some instructors may find this step burdensome.

Compared to RTC with ViaScribe, PLT using HTS was easier to implement, requiring no prior speaker training, and resulted in more than a 6 percent improvement in word recognition accuracy (Table 2). SR accuracy varied between ViaScribe and HTS due to differences in SR engines and decoding techniques. HTS performed a double-pass decoding routine on the lecture audio, which dynamically adjusts to the speaker's speech without requiring voice profile training, while ViaScribe must decode speech instantaneously.

Under favorable conditions for continuous SR applications, such as reading selected materials, trained users could achieve very high word recognition accuracy [1]. Expert speakers who devoted considerable time to voice profile training were reported to achieve up to 90 percent word recognition accuracy using ViaScribe in a university setting [4]. We assessed different strategies of improving word recognition accuracy. ViaScribe resulted in decreased word recognition accuracy than PLT for the same instructor in the same course. However, we had problems running ViaScribe consistently during class due to operating system compatibility issues. In a recent study, several factors were identified as significantly affecting WER. For instance, native English speakers had higher recognition accuracy than nonnative English speakers and males were more accurate than females. Also, decreased WER was evident in STEM than social science courses, but speaking rate was not a factor [37]. These results were consistent with our findings for the nonnative English speaking Phase 4 instructor and between Phase 2 (female) and Phase 3 (male) instructors. These findings are not surprising because the HTS' SR engine utilized US Broadcast acoustic models that would likely be predominantly male voice recordings. It underscores the importance to employ more diverse acoustic models or models specific to the discipline being lectured and transcribed.

We also sampled the word recognition accuracy along the time course of several select one-hour lectures. We found no significant changes in WER among different instructors over time. Therefore, the impact of lecture fatigue or practice effect can be discounted.

The SR best practices we developed during this case study offers strategies to improve word recognition accuracy. Instructors usually became more proficient once they became more familiar with using the SR-mLA technology, were prepared for teaching, and maintained proper speaking techniques. We believe the SR best practices could be universally applied to either RTC or PLT approaches as well as with other SR engines.

7.2 Implications for Student Learning

During this pilot study, there was a positive correlation among individual students in voluntarily taking the noncompulsory online quizzes more frequently when multimedia class notes were available than when they were not provided. Additionally, students received higher quiz scores when SR multimedia transcripts were accessible. Voluntary quiz taking and quiz performance rose most noticeably among those students who stated they were interested in studying material offered outside of class. We believe that having access to multimedia class notes was an added incentive for students to take the optional online quizzes. Once the multimedia class notes were unavailable, students took the online quizzes much less.

Past studies have demonstrated that acquiring and studying lecture notes result in a greater learning experience and higher overall academic performance for students [27], [28], [30], [40]. In this study, greater class grade performance was observed when synchronized multimedia class notes were available during the course. Students scored 10.2 percent higher on exams, 15.0 percent higher on noncompulsory online quizzes, and 10.5 percent on total scores. However, there are many factors involved in students getting high grades. Therefore, it is difficult to determine a direct correlative effect between having access to multimedia class notes and class performance for individual students. More study is needed to discern the full impact of SR multimedia class notes on student learning.

Students stated they benefitted the most from having multimedia class notes by

1. being able to pay more attention to the instructor instead of focusing on recording complete class notes,
2. the ability to review the lecture material multiple times,
3. synchronization of the instructor's lecture audio, transcripts, and slides, and
4. ability to make notes, comments, remarks and dynamically search for specific lecture keywords, time periods, or slides in these multimedia notes.

We believe SR-mLA would be especially advantageous for students with special needs, such as nonnative English-speaking students and students with disabilities, to obtain class notes without having to rely upon classmates or paid notetakers or captionists [4], [5], [6], [41]. Students incapable of or not confident in their own notetaking are able to acquire through PLT accurate and comprehensive multimedia class notes, which they could review at their own convenience and pace [25]. Various formats of these multimedia class notes can enable greater access for all—the visually impaired, deaf or hard of hearing, mobility impaired, learning disabled, nonnative speakers, distance learners, and any student who has a need of synchronized searching of the lecture material.

With RTC student subjects could view extemporaneous speech from the instructor to actively participate in class discussions during lectures. For instance, students with hearing loss can be engaged in lectures and respond during class with timely questions [11], [12]. In one study, students felt that RTC improved teaching and learning in class as long as word recognition was greater than 85 percent and the transcription and display lag was negligible [42].

SR-mLA is not synonymous with notetaking. Notetaking practices can vary depending on the student, who may choose to include or omit any lecture content or record this information in a way that helps their understanding.

However, verbatim notetaking was reported as the goal of half of college students, but it was estimated that less than 40 percent of lecture information was actually recorded [28]. SR-mLA enables the students to capture all lecture information. During this study, we achieved an average word recognition accuracy of 87 percent for PLT, which is more than twice as effective as verbatim notetaking by manual notetakers.

8 Future Plans

The next step in evaluating SR-mLA will be to test how students with disabilities can best utilize this technology to achieve particular learning outcomes. Although the initial findings regarding student performance are very encouraging, further research with larger class numbers and multiple courses is required to fully understand the impact of SR-mLA on academic performance, particularly for students with special needs. We also plan on evaluating the impact of SR transcripts on class performance in greater detail such as the effect and perception of raw unedited transcripts in comparison to the edited transcripts. Noted pioneers in SR applications, IBM Research and Nuance, are continually developing new SR engines and different platforms to convert speech to text. These SR systems can be implemented locally or virtually as a service via cloud computing environment. Future studies will take advantage of cloud technology. Access to SR transcription services would be more efficient through local hosting of a SR instance as a cloud service. The cloud computing model "Software as a Service" would allow users to remotely access SR-mLA using internet web browsers. The main advantages of this model are: on demand availability of SR-mLA without any software installation on user systems, access to greater processing power than on local PCs, and automating the whole process of postlecture transcription from recording the lecture to delivery of multimedia class notes.

Acknowledgments

This project was possible through financial support from the Indiana Family and Social Services Administration through a grant by the Regenstrief Center for Healthcare Engineering and Center for Assistive Technology, Center for Paralysis Research at Purdue University, and NIH Director's Pathfinder Award to promote diversity in the scientific workforce funded by the National Institute of General Medical Sciences (1DP4GM096842-01). Technical assistance was provided by the LL Consortium and their partnership with IBM Corporation. The authors appreciate assistance from Steve Witz, Bart Collins, Bharat Bhargava, Mike Wald, Riyi Shi, Andrew Marquis, James Walker, Li Hwa Chong, and Eli Asem. Special thanks to the journal's editors and reviewers for help in revision of this manuscript.

References

[1] K. Bain, S. Basson, and M. Wald, "Speech Recognition in University Classrooms: Liberated Learning Project," *Proc. Fifth Int'l ACM SIGCAPH Conf. Assistive Technologies (ASSETS '02)*, pp. 192-196, 2002.

[2] M. Wald and K. Bain, "Universal Access to Communication and Learning: the Role of Automatic Speech Recognition," *Universal Access in the Information Soc.*, vol. 6, no. 4, pp. 435-447, 2008.

[3] K. Bain, S. Basson, A. Faisman, and D. Kanevsky, "Accessibility, Transcription, and Access Everywhere," *IBM Systems J.*, vol. 44, no. 3, pp. 589-604, 2005.

[4] D. Leitch and T. MacMillan, "Liberated Learning Initiative Innovative Technology and Inclusion: Current Issues and Future Directions for Liberated Learning Research," Year III Report, Saint Mary's Univ., Nova Scotia, Canada, 2003.

[5] K. Ryba, T. Mcivor, M. Shakir, and D. Paez, "Liberated Learning: Analysis of University Students' Perceptions and Experiences with Continuous Automated Speech Recognition," *E-J. Instructional Science and Technology*, vol. 9, no. 1, Mar. 2006.

[6] D. Leitch, "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, Saint Mary's Univ., 2008.

[7] S. Repp, A. Grob, and C. Meinel, "Browsing within Lecture Videos Based on the Chain Index of Speech Transcription," *IEEE Trans. Learning Technologies*, vol. 1, no. 3, pp. 145-156, July-Sept. 2008.

[8] W. Hürst, T. Kreuzer, and M. Wiesenhutter, "A Qualitative Study Towards Using Large Vocabulary Automatic Speech Recognition to Index Recorded Presentations for Search and Access over the Web," *Proc. IADIS Int'l Conf. WWW/Internet (ICWI '02)*, pp. 135-143, 2002.

[9] M. Wald, G. Wills, D. Millard, L. Gilbert, S. Khoja, J. Kajaba, and Y. Li, "Synchronised Annotation of Multimedia," *Proc. IEEE Ninth Int'l Conf. Advanced Learning Technologies*, pp. 594-596, July 2009.

[10] M. Wald, "Synote: Accessible and Assistive Technology Enhancing Learning for All Students," *Proc. 12th Int'l Conf. Computers Helping People with Special Needs (ICCHP '10)*, pp. 177-184, 2010.

[11] L. Elliot, S. Foster, and M. Stinson, "Student Study Habits Using Notes from a Speech-to-Text Support Service," *J. Exceptional Children*, vol. 69, no. 1, pp. 25-40, 2002.

[12] D. Stover and N. Pendegraft, "Revisiting Computer-Aided Notetaking Technological Assistive Devices for Hearing-Impaired Students," *Clearing House: A J. Educational Strategies, Issues and Ideas*, vol. 79, no. 2, pp. 94-97, 2005.

[13] M. Wald, "Captioning for Deaf and Hard of Hearing People by Editing Automatic Speech Recognition in Real Time," *Proc. 10th Int'l Conf. Computers Helping People with Special Needs (ICCHP '06)*, pp. 683-690, 2006.

[14] M.S. Stinson, S. Eisenberg, C. Horn, J. Larson, H. Levitt, and R. Stuckless, "Real-Time Speech-to-Text Services," http://www.ntid.rit.edu/research/cprint/pdf/TFR_RTCaptioning_99.pdf, 1999.

[15] US Dept. of Education, "An Institutional Perspective on Students with Disabilities in Postsecondary Education," Nat'l Center for Educational Statistics, Postsecondary Education Quick Information System, Aug. 1999.

[16] J. Robison and C. Jensema, "Computer Speech Recognition as an Assistive Device for Deaf and Hard of Hearing People, Challenge of Change: Beyond the Horizon," *Proc. Seventh Biennial Conf. Postsecondary Education Persons Who Are Deaf or Hard of Hearing*, Apr. 1996.

[17] S. Watson and L. Johnston, "Assistive Technology in the Inclusive Science Classroom," *J. Science Teacher*, vol. 74, no. 3, pp. 34-38, 2007.

[18] M. Wald and K. Bain, "Enhancing the Usability of Real-Time Speech Recognition Captioning through Personalised Displays and Real-Time Multiple Speaker Editing and Annotation," *Proc. HCI Int'l Conf.*, vol. 7, pp. 446-452, July 2007.

[19] J. Boyle, "The Process of Note Taking: Implications for Students with Mild Disabilities," *Clearing House: A J. Educational Strategies, Issues and Ideas*, vol. 80, no. 5, pp. 227-230, 2007.

[20] T. Mortimore and W. Crozier, "Dyslexia and Difficulties with Study Skills in Higher Education," *J. Studies Higher Education*, vol. 31, no. 2, pp. 235-251, 2006.

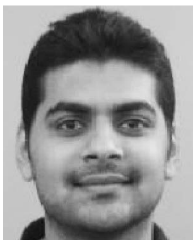
[21] C. Hughes and S. Suritsky, "Notetaking Skills and Strategies for Students with Learning Disabilities," *J. Preventing School Failure*, vol. 38, no. 1, pp. 7-12, 1993.

[22] B. Arons, "SpeechSkimmer: A System for Interactively Skimming Recorded Speech," *ACM Trans. Computer-Human Interaction*, vol. 4, no. 1, pp. 3-38, 1997.

[23] J. Huang, "The Effects of Academic Skills on Chinese ESL Students' Lecture Comprehension," *College Student J.*, vol. 40, no. 2, pp. 385-392, 2006.

[24] K. Hadjikakou, V. Polycarpou, and A. Hadjilia, "The Experiences of Students with Mobility Disabilities in Cypriot Higher Education Institutions: Listening to Their Voices," *Int'l J. Disability, Development and Education*, vol. 57, no. 4, pp. 403-426, Dec. 2010.

[25] B. Duerstock, R. Ranchal, Y. Guo, T. Doughty, J. Robinson, and K. Bain, "Assistive Notetaking Using Speech Recognition Technology," *Proc. Festival Int'l Conf. Caregiving, Disability, Aging and Technology (FICCDAT '11)*, 2011.
[26] J. Bonner and W. Holliday, "How College Science Students Engage in Notetaking Strategies," *J. Research Science Teaching*, vol. 43, no. 8, pp. 786-818, 2006.
[27] S. Peverly, V. Ramaswamy, C. Brown, J. Sumowski, M. Alidoost, and J. Garner, "What Predicts Skill in Lecture Note Taking?" *J. Educational Psychology*, vol. 99, no. 1, pp. 167-180, 2007.
[28] B. Titsworth and K. Kiewra, "Spoken Organizational Lecture Cues and Student Notetaking as Facilitators of Student Learning," *J. Contemporary Educational Psychology*, vol. 29, no. 4, pp. 447-461, Oct. 2004.
[29] F. DiVesta and G. Gray, "Listening and Note Taking: II. Immediate and Delayed Recall as Functions of Variations in Thematic Continuity, Note Taking, and Length of Listening-Review Intervals," *J. Educational Psychology*, vol. 64, no. 3, pp. 278-287, June 1973.
[30] K. Kiewra, "Investigating Notetaking and Review: A Depth of Processing Alternative," *J. Educational Psychologist*, vol. 20, pp. 23-32, 1985.
[31] P. Van Meter, L. Yokoi, and M. Pressley, "College Students' Theory of Note-Taking Derived from Their Perceptions of Note-Taking," *J. Educational Psychology*, vol. 86, no. 3, pp. 323-338, Sept. 1994.
[32] N.H. Van Matre and J. Carter, "The Effects of Note-Taking and Review on Retention of Information," *Proc. Ann. Meeting Am. Educational Research Assoc.*, 1975.
[33] H. Lyles, B. Robertson, M. Mangino, and J.R. Cox, "Audio Podcasting in a Tablet PC-Enhanced Biochemistry Course," *Biochemistry and Molecular Biology Education*, vol. 35, no. 6, pp. 456-461, 2007.
[34] K.E. Brobst, "The Process of Integrating Information from Two Sources: Lecture and Text," Doctoral dissertation, Columbia Univ., Teachers College, Dissertation Abstracts Int'l, vol. 57, p. 217, 1996.
[35] R.A. Palmatier and J.M. Bennett, "Notetaking Habits of College Students," *J. Reading*, vol. 18, pp. 215-218, 1974.
[36] IBM, "IBM ViaScribe," http://www-03.ibm.com/able/accessibility_services/ViaScribe-accessible.pdf, 2011.
[37] Liberated Learning, "Hosted Transcription Service," http://www.transcribeyourclass.ca/hts.html, 2012.
[38] D.M. Carey and P. Sale, "Practical Considerations in the Use of Technology to Facilitate the Inclusion of Students with Severe Disabilities," *Technology and Disability*, vol. 3, pp. 77-86, 1994.
[39] J. Bell, "Enhancing Accessibility through Correction of Speech Recognition Errors," *SIGACCESS Newsletter*, no. 89, pp. 22-27, Sept. 2007.
[40] B.J. Zimmerman, "A Social Cognitive View of Self-Regulated Academic Learning," *J. Educational Psychology*, vol. 81, no. 3, pp. 329-339, 1989.
[41] D. Hayden, D. Colbry, J.A. Black Jr., and S. Panchanathan, "Note-Taker: Enabling Students Who Are Legally Blind to Take Notes in Class," *Proc. 10th Int'l ACM SIGACCESS Conf. Computers and Accessibility (ASSETS '08)*, pp. 81-88, 2008.
[42] I. Weiss, "Report of the 2000 National Survey of Science and Mathematics Education," technical report, Horizon Research, 2001.



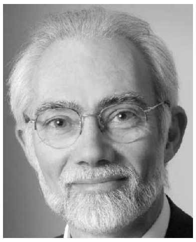
Rohit Ranchal received the MS degree in computer science from Purdue University in 2011 and the BTech degree in information technology from DAV Institute of Engineering and Technology, Jalandhar, India, in 2009. He is currently working toward the PhD degree in computer science at Purdue University, West Lafayette, Indiana. He is a member of the IEEE.

Teresa Taber-Doughty received the PhD degree in special education from Georgia State University, Atlanta, and has served as an educator for 27 years. She is an associate professor of special education at Purdue University. She is currently the president of the Division on Autism and Developmental Disabilities for the Council for Exceptional Children.

Yiren Guo received the BS degree in psychology from Shanghai Normal University, China, in 2010 and the MS degree in educational psychology from Purdue University in 2012.

Keith Bain is the lead of the International Liberated Learning Consortium and an adjunct professor in the Department of Finance, Management Science, and Information Systems at the Sobey School of Business. He cofounded an international research group dedicated to improving accessibility through speech recognition based captioning and transcription systems.

Heather Martin received the BA degree from Mount Allison University, the BEd degree from Saint Thomas University in New Brunswick, and the BSc degree from Saint Mary's University in Halifax, Nova Scotia. She is with Liberated Learning in the capacity of training, technical support, and software development.



J. Paul Robinson received the PhD degree in immunopathology from the University of New South Wales, Australia. He is the SVM professor of cytomics in the School of Veterinary Medicine and a professor in the Weldon School of Biomedical Engineering, Purdue University. He is a fellow of the American Institute for Medical and Biological Engineering and a member of both the IEEE and the IEEE Computer Society.

Bradley S. Duerstock received the BSE degree in biomedical engineering and the PhD degree in neuroscience from Purdue University. He is an associate professor of engineering practice in the Weldon School of Biomedical Engineering and School of Industrial Engineering and holds a courtesy appointment in the College of Veterinary Medicine at Purdue University. He is a member of the Center for Paralysis Research and director of the Institute for Accessible Science to promote the inclusion of persons with disabilities in the sciences.