ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 2, Issue 2, pp: (224-237), Month: April-June 2014, Available at: www.researchpublish.com

Enhancement Of Learning Using Speech Recognition And Lecture Transcription

[1]Ashwini B V, [2]Laxmi B Rananavare

[1]PG Student, [2]Associate Professor
Department of Computer Science and Engineering
Reva University, Bangalore, INDIA

***Abstract*: Speech recognition (SR) technology has a burgeoning range of applications in education system .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 post-lecture transcription (PLT), has been developed to increase the word recognition accuracy. Both methods has been 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 and instructor audio for students to access online after class. It has been learnt that PLT provides more word recognition accuracy than RTC. The potential benefits of SR-mLA for students who have difficulty taking notes accurately and independently were discussed, particularly for non-native English speakers and students with disabilities.**

***Keywords*: Educational Technology, Electronic Learning, Multimedia Systems, Notetaking, Speech Recognition.**

I. INTRODUCTION

Speech recognition (SR) technology has a burgeoning range of applications in education system from captioning video, television for the hearing-impaired, voice controlled computer operation, till dictation. Some of the most popular commercially available 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% accuracy but for the spontaneous speech the accuracy cannot be achieved in the same way due to number of reasons [1]. Liberated Learning Consortium (LL) (www.liberatedlearning.com) and its members have done research for the past decade and demonstrated that standard commercially available SR software was unsuitable for real-time captioning or transcription of speech [1]. Transcripts produced from a continuous unbroken stream of text are additionally difficult to read and interpret without punctuation or formatting [2], [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 re-create 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 in order to retrieve specific text or video lecture content [7], [8].

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 2, Issue 2, pp: (224-237), Month: April-June 2014, Available at: www.researchpublish.com

In this study, comparison has been done on the classroom implementation, reliability, and academic performance impact of two different methods of SR-mediated lecture acquisition (SR-mLA). Both SR-mLA techniques were employed using conventional educational technology .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, post-lecture transcription (PLT), employed a digital audio recording of the instructor's lecture to provide transcripts, which were synchronized with the audio recording for students to view online or download after class. 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. The relevance of SR-mLA tools for instructors and students, particularly those with special needs, were discussed in depth for future study.

II. A LITERATURE SURVEY ON EXISTING SYSTEM

A. Liberated Learning Consortium (LL)

The LIBERATED LEARNING PROJECT (LLP) is an applied research project studying two core questions: 1) Can speech recognition (SR) technology successfully digitize lectures to display spoken words as text in the classrooms? 2) Can speech recognition technology be used successfully as an alternative to traditional classroom notetaking for persons with disabilities? This paper addresses these intriguing questions and explores the underlying complex relationship between speech recognition technology, university educational environments, and disability issues.

MERITS:

1] It can achieve 98% accuracy.

DEMERITS:

1] The accuracy of spontaneous speech cannot be assessed.

2] Standard commercially available SR software was unsuitable for real-time captioning or transcription of speech.

B. Liberated Learning Initiative (LLI)

Creative uses for both mainstream and adaptive technologies are proliferating in higher education. Accordingly, there are continued opportunities to evaluate how technology may be used to enhance accessibility, facilitate inclusion, offer innovative learning alternatives, and ultimately promote academic success for students with disabilities (SWD). Historically, students with disabilities have experienced inadequate access to lecture material in the classroom, and insufficient access to the academic resources necessary to sustain their progress. The Liberated Learning Initiative (LLI) has been committed to understanding how speech recognition technology (SR) may contribute to creating a barrier-free learning environment for students.

MERITS:

1] The benefits of producing lecture transcripts have shown to enhance both learning and teaching.

2] Coupled with a recorded audio/video lecture track and copies of the lecture slides, students could re-create the lecture material for replicating the lecture at their own learning pace.

DEMERITS:

1] Transcripts produced from a continuous unbroken stream of text are additionally difficult to read and interpret without punctuation or formatting.

C. High School Pilot Project

This research concentrated on valuating of how the technology might enhance accessibility, facilitate inclusion and promote academic success for students with disabilities. The Liberated Learning Project, used a collaborative model with IBM and other telecommunication company to explore the use of SR as an accessibility technology in the classroom. Their concept was relatively simple; a professor's lectures would be digitized and displayed on a screen in real time and the digitized lectures would be made available for students after the lecture on-line. It was believed from the start that this combination of access would be helpful for students who are deaf, hard of hearing, learning disabled or mobility impaired. Furthermore, the provision of post-lecture edited online notes would be superior to the conventional dependence on other persons to take notes for those who could not take their own lecture notes.

MERITS:

1] It would be helpful for students who are deaf, hard of hearing, learning disabled or mobility impaired.

DEMERITS:

1] There was inherent need and obligation of SRT on non- disabled students and on pedagogy and teachers.

D. SR-mLA technique using IBM tools

1) Two Approaches for SR-mLA:

In this study, based on comparison done on the classroom implementation, reliability, and academic performance impact two different methods of SR-mediated lecture acquisition (SR-mLA) was developed .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, post-lecture transcription (PLT), employed a digital audio recording of the instructor's lecture to provide transcripts, which were synchronized with the audio recording for students to view online or download after class.

Instructors teaching large lecture classes were used to wearing wireless microphones for voice amplification. The generated lecture transcripts were made available to student to access online later. The most time consuming task of 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 [23]. Compared to RTC with ViaScribe, PLT using HTS was found easier to implement, requiring no prior speaker training, and resulted in more than a 6% improvement in word recognition accuracy . Under favorable conditions for continuous SR applications, such as reading selected materials, trained users could achieve very high word recognition accuracy [1]. ViaScribe resulted in decreased word recognition accuracy than PLT for the same instructor in the same course. However, ViaScribe had problem running consistently during class due to operating system compatibility issues.

[23]: The learning within lectures of hearing-impaired students could be reduces by errors in captions generated by speech recognition. This problem was addressed by investigating ways of correcting these captions. Finally, correction framework was developed to permit quantitative and qualitative testing of correction methods.

A table 1 show summarizes and compares major technical functionalities between real-time captioning using ViaScribe and post-lecture transcription using HTS.

Table 1. Comparsion of Major Functionalities between ViaScribe and HTS System

Features	ViaScribe (RTC)	HTS (PLT)
Recording method	ViaScribe has inbuilt capability of recording through classroom PC that receives audio from the wireless headset microphone worn by the lectures.	This system runs software on classroom PC to record lecture audio in SR-compatible format.
User dependent / independent SR engines	Initial user-specific profile training is necessary for the lectures to accustom SR-software to speakers voice to maximum word recognition accuracy.	No user training is required. The speaker wears a high-quality microphone to record the lecture audio file.
Error correction	After the lecture, the generated raw transcript can be corrected in order to update user's voice profile continuously.	The recorded audio file is uploaded to the HTS website via user interface for transcription and error correction.
Display method	Classroom PC's runs server software to transmit the transcripts to students laptop PC's connected as clients and also on classroom projection screen.	Lecture audio is synchronised with the PPT's and diagrams which are provided as multimedia class notes.

a)IBM® ViaScribe: IBM® ViaScribe utilized a SR engine capable of transcribing live or pre-recorded speech developed collaboratively by IBM® and the LL Consortium. During class lectures displayed or captioned what the instructor uttered into text as it is being spoken. ViaScribe 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], [21].

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 2500 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 .

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. 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 .

b) *IBM® Hosted Transcription Service (HTS):* IBM® HTS was selected for post-lecture transcription primarily because of its higher word recognition accuracy rates compared to other systems, hence was found suitable for recognizing extemporaneous speech [22]. 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' SR engine employs a double-pass decoding technique, which dynamically adjusts to the speaker's voice, without requiring voice profile training or enrollment [22].

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 post hoc correction method similar to ViaScribe was performed. Afterwards, 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 post-production 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.

MERITS:

1] Improve the word recognition accuracy through speaker-dependent and speaker-independent SR systems developed by IBM Research.

DEMERITS:

2] These SR-mLA techniques developed by IBM Research group are too costly to afford and use.

E. Implications for Student Learning:

Past studies have demonstrated that acquiring and studying lecture notes resulted in a greater learning experience and higher overall academic performance for students [17], [18], [24]. Students had stated that 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.

SR-mLA was believed that it would be especially advantageous for students with special needs, such as non-native 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], [25]. 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 [16].

[24]: Researchers were interested in academic self- regulated learning have begun to study processes that students use to initiate and direct their efforts to acquire knowledge and skill. The social cognitive conception of self-regulated learning presented here

involves a triadic analysis of component processes and an assumption of reciprocal causality among personal, behavioral, and environmental triadic influences.

[25]: In class, note-taking is a vital learning activity in secondary and post-secondary classrooms. The process of note-taking helps students stay focused on the instruction, forces them to cognitively process what is being presented, and better retain what has been taught, even if they never refer to their notes after the class.

F. Existing SR-Tools

1) Dragon NaturallySpeaking (also known as Dragon for PC, or DNS), is a speech recognition software package developed by Dragon Systems . It runs on Windows computers. and the most recent package is version 12.0, which supports 32-bit and 64-bit editions of Windows XP, Vista, 7 and 8.The Mac OS version is called DragonDictate or Dragon for Mac.

2) IBM ViaVoice is a range of language-specific continuous speech recognition software products offered by IBM. The current version is designed for use in embedded devices and other latest stable version of IBM ViaVoice to improve decoding accuracy, by reading prepared texts of a few hundred sentences. The acoustic models were tuned by recorded text to that specific user. In addition, user specific text files could be parsed to tune the language model. Correction of misrecognised words was also used to improve subsequent decode accuracy.

[2]: Automatic Speech Recognition (ASR) can support universal access to communication and learning through the cost-effective production of text synchronized with speech and describes achievements and planned developments of the Liberated Learning Consortium to support preferred learning and teaching styles; assist those who had difficulty in notetaking; assisting learners to manage and search online digital multimedia resources; provide automatic captioning of speech for deaf learners or when speech is not available or suitable; assist blind, visually impaired or dyslexic people to read and search material; and, assist speakers to improve their communication skills.

[3]: Accessibility in the workplace and in academic settings has increased dramatically for users with disabilities. For persons who are deaf and hard of hearing full participation requires complete access to audio materials, both for live settings and for prerecorded audio and visual information. Even for users with adequate hearing, captioned or transcribed materials offer another modality for information access in certain situations, such as listening in noisy environments, interpreting speakers with strong accents, or searching audio media for specific information

[5]: In this study, examination of continuous automated speech recognition was done in the classroom. The participants were both native speakers of English (L1) and English as a second language students (L2) enrolled in an information systems course (Total N=160). After an initial training period, it is observed that an L2 lecturer in information systems delivered three 2-hour lectures over a three-week period to the participants and other students. Compared with L1 students, a significantly greater number of L2 students and special needs students reported that the system had potential as an instructional support mechanism. However, a greater accuracy in the system's recognition of lecture text vocabulary needs to be achieved.

[7]: The number of digital lecture video recordings has increased dramatically since recording technology became easier to use and searching within this large archive are limited and difficult. Additionally, detailed browsing in videos is not supported due to the lack of an explicit annotation. Manual annotation and segmentation is time-consuming and therefore useless. In this paper , an indexing method for computer science courses based on their existing recorded videos are discussed.

[8]: Software for automatic capture and recording of live presentations has become popular especially at universities. Many research projects have experimented with automatic lecture recording and making the resulting documents available for access to the students over the Web.

- The system is used on a regular basis for lecture recording at university as well as by many other institutions. It has not only attracted a broad variety of users but meanwhile served as a basis for a commercial product, as well.
- With AOF, recording takes place automatically in the background while the presenter gives his lecture in a traditional way. The result of the recording process is a multimedia document containing the audio (i.e., voice of the presenter), the slides of the presentation together with annotations on these slides, as well as an optional video image of the lecturer.

G. Utilities and Tools

The existing 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.

1) Recording Lecture Audio:

The instructors' oral lectures were recorded during class using an Audio-Technica Freeway 8-channel 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®, open source software, was used for recording lecture audio for PLT. Audacity provided various configuration options for audio recording in SR-compatible format settings. Instructors delivered their PowerPoint™ slideshow presentation during class with their speech saved as WAV files along with each slide image. To generate a synchronized multimedia notes and to synchronize audio, text, slides certain timing were required[9].

2) Display of Multimedia Lecture Notes:

To synchronize audio or video recordings, SR transcripts and PowerPoint slide images Synote (www.synote.org/synote), a web-based application was used to create a synchronized bookmarks or 'Synmarks' that contained notes and tags [10]. Synote displayed these synchronized, multimedia class notes online through a web interface for easy access . Registered users can save their own remarks and notes for each set of class notes. Later students could access multimedia transcripts through links posted on Blackboard® which was a commercial web-based course management system accessible to students for taking quizzes, accessing lecture notes, and viewing their class grades online.

3) Word Error Rate Tool:

A Word Error Rate (WER) was a evaluation tool developed by IBM® 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 [21].

III. BACKGROUND

A. 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 was vitally important. The addition of text transcripts versus purely audio enhances accessibility [9], [10]. 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]. For students with significant hearing loss, captionists are traditionally employed to interpret and transcribe what the instructor is lecturing [11]. 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 [13], [14]. SR-mLA has been shown to help international students who are non-native speakers with regard to the language of instruction (in this case English) to acquire accurate lecture notes [5], [16].

[9]: Multimedia has become technically easier to create (e.g. recording lectures) but while users can easily bookmark, search, link to, or tag the WHOLE of a podcast or video recording available on the web they cannot easily find, or associate their notes or resources with, PART of that recording.

[10]: Although manual transcription and captioning can increase the accessibility of multimedia for deaf students it is rarely provided in educational contexts due to the cost and shortage of highly skilled and trained stenographers. Speech recognition has the potential to reduce the cost and increase the availability of captioning if it could satisfy accuracy and readability requirements.

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 2, Issue 2, pp: (224-237), Month: April-June 2014, Available at: www.researchpublish.com

[11]: Deaf and hard of hearing people can find it difficult to follow speech through hearing alone or to take notes when lip-reading or watching a sign-language interpreter. Lectures can be digitally recorded and replayed to provide multimedia revision material for students who attended the class and a substitute learning experience for students unable to attend.

[13]: Federal legislation requires that supplementary aids and services should be provided to students with diagnosed disabilities to assist them in accessing the general-education curriculum. Such modifications to the curriculum may include assistive technology (AT) devices and services. AT can be any item, piece of equipment, or teacher-made product that is designed to improve a student's functional capability or help a student succeed in accessing the general education curriculum.

[15]: Listening to a speech recording was found more difficult than visually scanning a document because of the transient and temporal nature of audio. Audio recordings capture the richness of speech, yet it was difficult to directly browse the stored information. This article describes techniques for structuring, filtering, and presenting recorded speech, allowing a user to navigate and interactively find information in the audio domain. This article describes the Speech Skimmer system for interactively skimming speech recordings. Speech Skimmer uses speech-processing techniques to allow a user to hear recorded sounds quickly, and at several levels of detail.

[16]: Innovative speech recognition (SR) technology was utilized to assist students acquiring more complete and accurate lecture notes. SR-assisted note taking can be accomplished in different ways to allow students to devote more attention to understanding course material than manually recording the instructor's lecture content. Lecture notetaking was physically challenging for many students with disabilities who cannot take notes for the bulk of lecture notes themselves and must rely upon hired notetakers for class notes, which may not readily available or feasible. They are also dependent on the skills and knowledge of the notetaker for the quality of their notes [1].

B. Notetaking in STEM subjects

Notetaking was a fundamental and ubiquitous learning activity that students were expected to perform and master during their educational development [17], [18]. 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 [19].

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 . Science courses, such as histology and biochemistry were selected by students as classes they would most want assistance with notetaking [20].

[20]: This report describes the effects of making audio podcasts of all lectures in a large, basic biochemistry course promptly available to students. The audio podcasts complement a previously described approach in which a tablet PC is used to annotate PowerPoint slides with digital ink to produce electronic notes that can be archived. The survey data suggest that the students have a positive attitude toward the combination of tablet-based instruction and audio podcasting.

[18]: Previous research had shown that providing written organizational lecture cues boosts notetaking and that boosting notetaking raises achievement. To find out this, participants were made to listen to a lecture that contained or did not contain spoken organizational lecture cues. Participants either recorded lecture notes or refrained from notetaking while listening to the cued or uncued lecture. Results showed that spoken organizational lecture cues boosted the number of noted organizational points and details by 39 and 35%, respectively.

IV. PROPOSED SYSTEM

Despite of several documented hurdles of implementing new technologies in classroom settings, it was assumed that it could be 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. These methods are assumed to be inexpensive and relatively easy to set up prior to class. Both the approaches of SR-mLA would require the following core components: recording hardware and software, SR software, error correction methodology for transcription accuracy, strategy for transcription display, and a usable distribution/publishing platform shown in figure. Both SR-mLA technologies did not interfere with typical lecture teaching activities, only requiring the instructor to wear a wireless microphone.

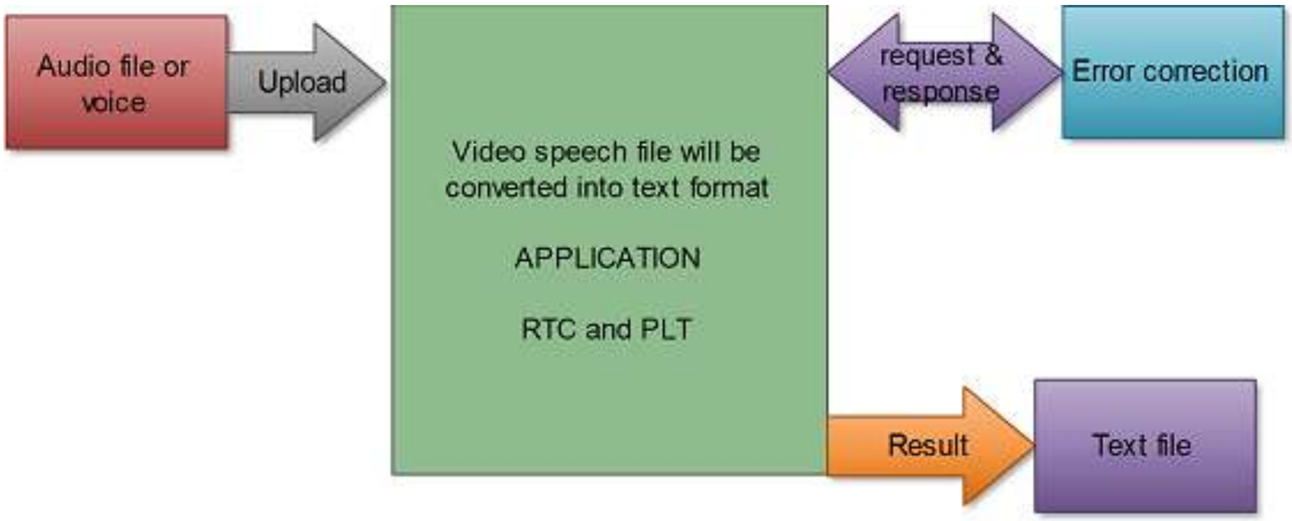

Fig: Overview of SR-mLA methodology

In this study, we compared the classroom implementation, reliability, and academic performance impact of two different methods of SR-mediated lecture acquisition (SR-mLA). 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, post-lecture transcription (PLT), employed a digital audio recording of the instructor's lecture to provide transcripts, which were synchronized with the audio recording for students to view online or download later after the class.

A. Module Specification

Modules

- Speech acquisition
- Request for the RTC or PLT function
- Speech pre-processing
- Text storage

Speech Acquisition

The data acquisition process has been designed for the minimization of the error and acquisition process the will recover all types of the error of the word.

During speech acquisition, speech samples are obtained from the speaker in real time and stored in memory for preprocessing. Speech acquisition requires a microphone coupled with an analog-to-digital converter (ADC) that has the proper amplification to receive the voice speech signal, sample it, and convert it into digital speech. The system sends the analog speech through a transducer, amplifies it, sends it through an ADC, and transmits it to the processor in the system through the communications interface. The system needs a parallel/serial interface to the processor and an application running on the processor that acquires and stores data in memory. The microphone input port with the audio codec receives the signal, amplifies it, and converts it into 16-bit PCM digital samples at a sampling rate of 8 KHz.

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 2, Issue 2, pp: (224-237), Month: April-June 2014, Available at: www.researchpublish.com

Request for the RTC or PLT function:

The project converts speech into text using an application in the system. Run-time speech data is acquired from a transducer through analog-to-digital (A/D) conversion and is sent to the system, which provides speech preprocessing and recognition using the processor. The system sends text data to a PC for storage. Functionally, the project has the following blocks:

Speech Preprocessing

The speech signal consists of the uttered digit along with a pause period and background noise. Preprocessing reduces the amount of processing required in later stages. Generally, preprocessing involves taking the speech samples as input, blocking the samples into frames, and returning a unique pattern for each sample, as described in the following steps.

1. The system must identify useful or significant samples from the speech signal. To accomplish this goal, the system divides the speech samples into overlapped frames.

2. The system checks the frames for voice activity using endpoint detection and energy threshold Calculations.

3. The speech samples are passed through a pre-emphasis filter.

4. The frames with voice activity are passed through a Hamming window. HMM Training Hidden Markov Models HMM-Based Recognition Speech Preprocessing Speech Text Storage Acquisition C Program on PC External Hardware SOPC VB Program on PC SOPC-Based Speech-to-Text Conversion 87

5. The system performs autocorrelation analysis on each frame.

6. The system finds linear predictive coding (LPC) coefficients using the Levinson and Durbin algorithm.

7. From the LPC coefficients, the system determines the cepstral coefficients and weighs them using a tapered window. The cepstral coefficients serve as feature vectors.

Text Storage

This speech-to-text conversion system can send the recognized digit to a PC via the serial, USB, or Ethernet interface for backup or archiving. For our testing, we used a serial cable to connect the PC and RS-232 port on the DE2 board. Processor on the DE2 board sends the digital speech data to a PC; a target program running on the PC receives the text and writes it to the disk. PC program is written using Visual Basic 6 (VB) using a Microsoft serial port control. The VB program must be run in the background for the PC to receive the data and write it to the hard disk. The Windows HyperTerminal software or any other RS-232 serial communication receiver could also be used to receive and view the data. The serial port communication runs at 115,200 bits per second (bps) with 8 data bits, 1 stop bit, and no parity. Handshaking signals are not required. The speech-to-text conversion system can also operate as a standalone network device using an Ethernet interface for PC communication and appropriate speech recognition server software designed for the processor.

HMM-Based Recognition

Recognition or pattern classification is the process of comparing the unknown test pattern with each sound class reference pattern and computing a measure of similarity (distance) between the test pattern and each reference pattern. The digit is recognized using a maximum likelihood estimate, such as the Viterbi decoding algorithm, which implies that the digit whose model has the maximum probability is the spoken digit Preprocessing, feature vector extraction, and codebook generation are same as in HMM training. The input speech sample is preprocessed and the feature vector is extracted. Then, the index of the nearest codebook vector for each frame is sent to all digit models. The model with the maximum probability is chosen as the recognized digit. After preprocessing in the processor, the required data is passed to the hardware for Viterbi decoding. Viterbi decoding is computationally intensive so we implemented it in the system for better execution speed, taking advantage of hardware/software co-design. The Viterbi decoder has been written in Verilog HDL and included it as a custom instruction

ISSN 2348-1196 (print)
International Journal of Computer Science and Information Technology Research ISSN 2348-120X (online)
Vol. 2, Issue 2, pp: (224-237), Month: April-June 2014, Available at: www.researchpublish.com

in the processor. The custom instruction copies or adds two floating-point numbers from data a and data b , depending on the prefix input. The output (result) is sent back to the processor for further maximum likelihood estimation.

V. RESULTS

The Proposed SR-mLA technique has resulted in the following snapshots of two methods: Real time captioning and Post lecture transcription.

Fig A shows the Home Page of the SR-mLA technique where the new user can register and the existing user can login using login option providing the username and password to login to his/her secured account. After the user has logged in, the next page (Fig B) contains options select the lecture and course to get the required notes for the following lecture/course. User could either select RTC for Real time captioning or Generate for uploading video/Audio file for PLT operation to generate Post lecture transcription

Fig B shows the Page that has converted the speech to text using RTC; this text could be edited and saved. Fig C is the PLT showing the translated text of the uploaded lecture video which could be edited and hosted online for further access by student and lectures.

Fig A: Home Page with login option

Fig B: RTC showing speech transcription generated

Fig C: PLT showing translated text of uploaded video file

VI. CONCLUSION AND FUTURE WORKS

In this proposed paper we have seen how SR-mLA would help students to improve their academic performance. 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 could be done to provide the facility for PowerPointTM slideshows, related diagrams with PLT and also for 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. These SR systems can be implemented locally or virtually as a service via cloud computing environment. 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 post-lecture transcription from recording the lecture to delivery of multimedia class notes.

ACKNOWLEDGMENTS

We would like to thank faculties of CSE Dept and Principal of Reva ITM for their continuous support in preparation of this paper.

REFERENCES

[1] K. Bain, S. Basson, and M. Wald, "Speech recognition in university classrooms: liberated learning project," Proc. The Fifth International ACM SIGCAPH Conference on Assistive Technologies (ASSETS), pp. 192-196, 2002.

[2] M. Wald, K. Bain, "Universal access to communication and learning: the role of automatic speech recognition," Universal Access in the Information Society, vol. 6, no. 4, pp. 435-447, 2008.

[3] K. Bain, S. Basson, A. Faisman, D. Kanevsky, "Accessibility, transcription, and access everywhere," IBM Systems Journal, 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 University, 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," EJournal of Instructional Science and Technology (e-JIST), 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 University, Nova Scotia, Canada, 2008.

[7] S. Repp, A. Grob, and C. Meinel, "Browsing within Lecture Videos Based on the Chain Index of Speech Transcription," IEEE Transactions on Learning Technologies, pp. 145-156, 2008.

[8] W. Hürst, T. Kreuzer, and M. Wiesenhüter, "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. 9th IEEE International Conference on Advanced Learning Technologies, pp. 594-596, Jul. 2009.

[10] M. Wald, "Synote: Accessible and Assistive Technology Enhancing Learning for All Students," Proc. ICCHP 2010 Part II LNCS 6180, pp. 177-184, 2010.

[11] M. Wald, "Captioning for Deaf and Hard of Hearing People by Editing Automatic Speech Recognition in Real Time," Proc. Tenth International Conference on Computers Helping People with Special Needs ICCHP 2006, LNCS 4061, pp. 683-690, 2006.

[12] M. S. Stinson, S. Eisenberg, C. Horn, J. Larson, H. Levitt and R. Stuckless, "Real-time speech-to-text services," Reports of the National Task Force on Quality Services in Postsecondary Education of Deaf and Hard of Hearing Students, Rochester, NY, 1999.

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

[14] 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 International Conference, vol. 7, pp. 446-452, Jul. 2007.

[15] B. Arons, "SpeechSkimmer: a system for interactively skimming recorded speech," ACM Transactions on Computer-Human Interaction (TOCHI), 4(1), 3-38, 1997.

[16] B. Duerstock, R. Ranchal, Y. Guo, T. Doughty, J. Robinson, and K. Bain, "Assistive Notetaking Using Speech Recognition Technology," Proc. Festival of International Conferences on Caregiving, Disability, Aging and Technology (FICCDAT): RESNA/ICTA3, Toronto, Canada, 2011.

[17] S. Peverly, V. Ramaswamy, C. Brown, J. Sumowski, M. Alidoost, J. Garner, "What Predicts Skill in Lecture Note Taking?" Journal of Educational Psychology, vol. 99, no. 1, pp. 167-180, 2007.

[18] B. Titsworth, and K. Kiewra, "Spoken Organizational Lecture Cues and Student Notetaking as Facilitators of Student Learning," Journal of Contemporary Educational Psychology, vol. 29, no. 4, pp. 447-461, Oct. 2004.

[19] N. H. Van Matre, J. Carter, "The Effects of Note-Taking and Review on Retention of Information," American Educational Research Association, Washington, D.C., 1975.

[20] 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(6), pp. 456-461, 2007.

[21] IBM®ViaScribe,http://www03.ibm.com/able/accessibility_services/ViaScribeaccessible.pdf, 2011.

[22] Liberated Learning Hosted Transcription Service (HTS), http://www.transcribeyourclass.ca/hts.html, 2012.

[23] J. Bell, "Enhancing accessibility through correction of speech recognition errors," SIGACCESS Newsletter, Issue 89, Sep. 2007.

[24] Self-Regulated Learning and Academic Achievements: An Overview, http://www.unco.edu/cebs/psychology/ kevinpugh/ motivation_project/resources/zimmerman90.pdf

[25] D. Hayden, D. Colbry, J.A. Black Jr. and S. Panchanathan, "Note-Taker: Enabling students who are legally blind to take notes in class," 10th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS 2008), Halifax, Nova Scotia, Canada, pp. 81-88, 2008.

[26] I. Weiss, "Report of the 2000 National Survey of Science and Mathematics Education," Technical Report, Chapel Hill, NC: Horizon Research, 2001.

[27] Rohit Ranchal, Teresa Taber-Doughty, Yiren Guo, Keith Bain, Heather Martin, J. Paul Robinson, and Bradley S. Duerstock,"Using Speech Recognition for Real-Time Captioning and Lecture Transcription in the Classroom", IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES,2013