New Orleans JAZZED about Engineering Education

Paper ID #16847

Inclusive Learning through Real-time Tracking Display of Captions

Dr. Raja S. Kushalnagar, Rochester Institute of Technology

Raja Kushalnagar is an Assistant Professor in the Information and Computing Studies Department at the National Technical Institute for the Deaf at the Rochester Institute of Technology in Rochester, NY. He teaches information and/or computing courses, and tutors deaf and hard of hearing students in computer science/information technology courses. His research interests focus on the intersection of disability law, accessible and educational technology, and human-computer interaction. He is focused on enhancing educational access for deaf and hard of hearing students in mainstreamed classrooms. He worked in industry for over five years before returning to academia and disability law policy. Towards that end, he completed a J.D. and LL.M. in disability law, and an M.S. and Ph.D. in Computer Science.

Mr. Gary W. Behm, Rochester Institute of Technology

Gary W. Behm, Assistant Professor of Engineering Studies Department, and Director of NTID Center on Access Technology Innovation Laboratory, National Technical Institute for the Deaf, Rochester Institute of Technology.

Gary has been teaching and directing the Center on Access Technology Innovation Laboratory at NTID for five years. He is a deaf engineer who retired from IBM after serving for 30 years. He is a development engineering and manufacturing content expert. He develops and teaches all related engineering courses. His responsibility as a director of Center on Access Technology Innovation Laboratory include the planning, implementation and dissemination of research projects that are related to the need of accessibility. He received his BS from RIT and his MS from Lehigh University. His last assignment with IBM was an Advanced Process Control project manager. He managed team members in delivering the next generation Advanced Process Control solution which replaced the legacy APC system in the 300 mm semiconductor fabricator. Behm has fifteen patents and has presented over 30 scientific and technical papers at various professional conferences worldwide.

Dr. Aaron Weir Kelstone

EDUCATION Ed.D in Education, Northeastern University, Boston, Massachusetts, 2013 M.A. in English Literature Cleveland State University, Cleveland, Ohio, 2001 B.A. in English Literature Cleveland State University, Cleveland, Ohio, 1994 PROFESSIONAL EXPERIENCE Senior Lecturer, 2010 & Program Director of Performing Arts, NTID ,2011 RECENT PUBLICATIONS American Deaf Prose: 1980-2010: Gallaudet Deaf Literature Series, Vol. 1, "Homecoming," Gallaudet UP, April, 2012 Wordgathering: A Journal of Disability Poetry, "Ruminations of a Cyborg," (WWW.wordgathering.com) March, 2010 Vignettes of a Deaf Character: Foreword, Gallaudet UP, November 2010 Tactile Mind Press,"25-Cents," Minneapolis, MN, 2001 RECENT GRANTS AND FOUNDATION FUNDING NSF-funded dance production that interprets scientific principles for a general audience. Astrophysics and Dance: Engaging Deaf, Hard-of-Hearing, and Hearing Individuals in Science Education (NSF Award No. DRL-1136221) culminated in a dance performance that toured the country. It used a multimedia theatrical production to communicate information about gravitational astrophysics to members of the general public, with a special emphasis on deaf and hard-of-hearing individuals OTHER RELEVANT EXPERIENCE Theatre background provides insight related to the use and implementation of access technology in performance and presentation environments to support development of prototypes for use in performances.

Mr. Brian Trager, Rochester Institute of Technology
Miss Mary Rose Weber, Rochester Institute of Technology

I am currently a fifth year student at Rochester Institute of Technology (RIT) and studying Information Technology for B.S. My concentrations are web development and mobile development. My abilities are working as a team member solving problems. I am strongly motivated by new challenges.

Mr. Shareef Sayel Ali, Rochester Institute of Technology NTID and VTCSecure's ACE Innovation Lab

Shareef wrote and designed the RTTD software. He works for VTCSecure and NTID on the FCC's Accessible Communication for Everyone (ACE) application. ACE is an open source platform that allows video calls and so much more. Shareef is pursuing his BS degree in Computer Science at RIT.

Mr. Jason Dominick Lee, Rochester Institute of Technology, Center on Access Technology

I am fifth year Electrical and Mechanical Engineering Technology in the College of Applied Science & Technology at RIT. For over two years, I have worked as a hardware engineer under Center on Access Technology (CAT) department. During that period, I have developed first generation Real-Time Tracking Display (RTTD). I currently work on data collection for second generation RTTD in classroom purpose. I also work on third generation RTTD development for theater purpose.

Inclusive Learning through Real-Time Display of Captions

Abstract

Deaf and Hard of Hearing (DHH) students cannot follow classroom lectures without accommodations such as real-time speech-to-text transcription. Current classroom transcription systems, such as C-Print improve access to classroom lectures, but still do not provide equivalent access to spoken information. These transcription systems require the DHH students to watch the transcription on a personal laptop screen, which is suitable for speeches, but not engineering lectures. Unlike speeches, most engineering lectures include use of detailed visuals such as slides or diagrams, and sequential procedures. DHH students constantly look away from their laptop display to search and study the visuals. As a result, they spend less time watching lecture visuals and gain less information than their hearing peers. However, the need to process simultaneous aural and visual information can also be taxing for hearing students, and previous studies have shown that they also benefit from real-time speech-to-text transcription.

We evaluated the real-time display of captions (RTD) usability by both deaf and hearing students in an engineering class. It further examined the factors that influence hearing students' use of RTD as an alternative source of information to help with their learning process in the classroom, and the factors that influence deaf students' use of RTD.

Our evaluation showed that DHH students prefer a continuously moving RTD with three lines, and that is as close as possible to the teacher. On the other hand, hearing students prefer a RTD that has 6 lines that at a fixed location.

Challenges

Historically, DHH students are an under-represented and under-served minority in higher education because they do not receive adequate information in class. As a result these students are often unprepared for traditional STEM classrooms. Most DHH students cannot understand spoken lectures without the aid of aural-to-visual access. Prior to the introduction of the earliest accessibility laws in 1974, less than 4% of DHH individuals completed college in the 1960s. Although substantial progress has been made in the past 40-50 years in terms of accessibility laws and technology, the disparity in graduation persists: the graduation rate is 16% of DHH as compared to 30% for their hearing peers[1].

Currently there are over 31,000 DHH students enrolled in college and this enrollment number is up 15,000 over the past 10 years[2]. Deaf and hard of hearing students, and students with visual learning preferences are underrepresented in engineering, in part because it is difficult for visual learners to sustain attention on more than one visual simultaneously.Their attention is severely taxed when they have to switch attention between detail rich slide or demonstration visuals and the interpreter or teacher[3–5]. The underlying reason is that the classrooms are not designed to utilize students' visual skills and are not fully accessible by DHH students, including engineering classrooms[6]. When teachers maximize the benefits of visual learning, the barriers in regular lectures for DHH students, such as using spoken English is partially ameliorated[7].

Figure 1: The spread of visual information sources (interpreter, whiteboard, slides, teacher) in a classroom

Hearing students are able to simultaneously watch the visuals and listen to the spoken explanation, while DHH students have to choose either the visuals or captions. Typically, they constantly look away from the real-time transcript to search and observe details in the classroom visuals. As a result, they spend less time on the visuals and gain less than their hearing peers. They can also fall behind in reading the real-time transcript compared with their hearing peers.

Our paper compares viewing preferences between DHH and hearing students when they viewed two versions of real-time transcript display interfaces with C-Print, which had a 4-5 second delay. Our previous research has shown that DHH students preferred 3 lines, while hearing students preferred 6 lines[8]. The feedback indicated that DHH students reported that captions was their only access to spoken information, so the 4-5 second delay between the speech and captions was not an issue. On the other hand, hearing students reported using the captions as backup, for example, to review information if they misheard or did not understand a specific word or sentence.

Background

Real-time text is usually delivered through a display system. One popular integrated captioning and display system is C-Print, which was developed at the National Technical Institute for the Deaf, a federally funded institution for deaf and hard of hearing students. The system has significantly improved access to lectures for DHH individuals in many programs around the country[9,10]. It also benefits individuals with other disabilities, such as those with a visual impairment or a learning disability[a].

[a] (https://www.rit.edu/ntid/cprint/)

C-Print displays the printed text of spoken English on a laptop/tablet in real time, which is a proven and appropriate means of acquiring information for some individuals who are DHH. A trained operator, called a C-Print captionist, produces a text display of the spoken information in classroom or other settings. At the same time, one or more students read the display to access the information.

The main advantage of the C-Print system as compared with professional stenographers output is cost and availability. In terms of cost, C-Print typists need far less training and time as compared with stenographers who and are cheaper than professional stenographers, who type on a specialized keyboard using short-hand notation [10]. Although stenographers are more accurate, studies have shown that C-Print typists' accuracy is high enough that studies have not found any significant difference between C-Print and sign language interpreters in terms of student learning in the classroom [19, 4]. The main disadvantage of C-Print, like CART, is that the "flow" of text is not smooth or consistent, unlike LegionScribe [10].

People often falsely assume that the use of captions in the classroom enables full access to real-time text for DHH people. This assumption minimizes other information accessibility issues, such as simultaneous visual streams or content complexity. For example, DHH students have to constantly switch gaze between the C-Print display and the lecture slides. So, for current classrooms, captions should be presented in a way that reduces the effects of visual dispersion and cognitive overload[11].

Our RTD system addresses the challenges described in the previous section and in previous work[8]. The goal of our RTD system is to minimize visual dispersion and maximize use of captions for the caption viewers. The system involves three components: C-Print, Kinect 2, and the RTD system. The system can be easily set up and turned on within a few minutes by the captionist.

For capturing and displaying the captions, the RTD system uses a laptop and application to integrate the C-print application with a Kinect 2 device, so that the display of real-time text by the C-Print captionish is displayed over the head of the person. The RTD system is mounted on a cart so that it can be moved and quickly deployed. The RTD real-time transcript is produced by C-Print above the presenter, which is all within the viewer's peripheral vision as shown in Figure 2.

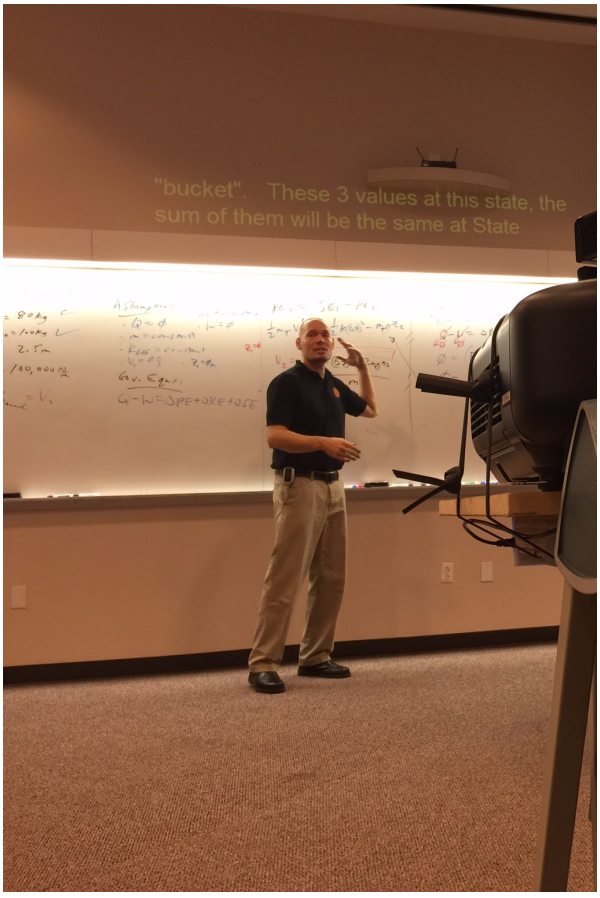

Figure 2: RTD System with captions

The RTD system uses the Kinect 2 system to track and record the current location of the presenter, as shown in Figure 5.

The system records the location of the presenter, and displays the projected captions at a fixed distance from the presenter's body, usually over their head. This projection scheme ensures that the captions and presenter are always close to each other at a fixed distance. The fixed distance and location predictability enables the viewer to easily see both the transcript and presenter's information simultaneously. They can detect changes, easily follow and read the real-time transcript without missing the presenter's information.

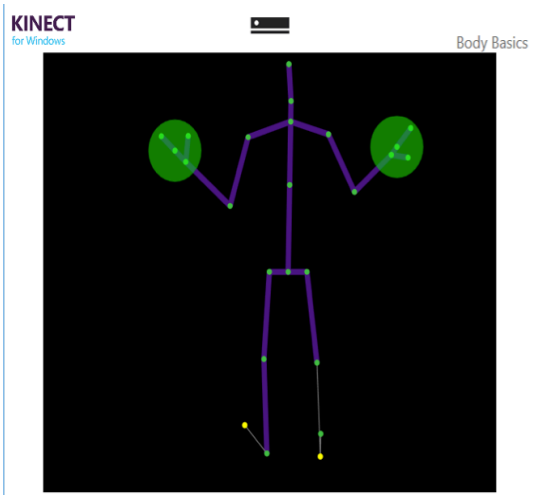

Figure 3: Kinect 2 3D view of presenter

The startup time is less than a minute and mostly the time for the projector to start up, which is about 30 seconds. The cost of the entire system currently ranges from about $700 to $1000. The bulk of the cost is to cover a laptop, such as a Surface Pro for $500. The other items are somewhat less expensive, ranging from $200 for the Kinect 2 for Windows system at the low end. The final components of the system include the provision of a cart to provide easy transportation of the system from one place to another and vice versa.

Evaluation of RTD

We ran a study in a class that was about 2 hours long for each lecture. We asked the students to evaluate the readability and usability of the RTD captions, with varying amount of lines and movement. Prior to the experiments, we displayed RTD in a static location in two class lectures to let students become comfortable with the technology.

The study was conducted over four lectures. On the first day, the captions were displayed at fixed location with three lines (STD3). On the second day, the captions were displayed above the teacher with three lines (RTD3). On the third day, the captions were displayed at a fixed location with six lines (STD6). On the fourth day, the captions were displayed above the teacher with six lines (RTD6).

At the start of each class, we announced that we were evaluating speech-to-text accommodations, and that we would distribute a survey about the speech-to-text display at the end of class, and that it would take about a minute to complete. We distributed the surveys and collected them at the end of class, after they were completed.Our survey consisted of a demographic question on hearing status, four Likert questions on user preferences, and two open-ended questions to solicit feedback on their experience with the system in action. All of the enrolled students in attendance for the course lectures had experience with C-Print, by virtue of having watched it in the class lectures prior to the evaluation period.

- D1: 'Are you deaf or hearing?'

- Q1: ‘How easy is it to follow [TRTD3/TRTD6/SRTD3/SRTD6]?’, with a Likert scale that ranged from 1 through 5, with 1 being ‘Not at all easy’ to 5 being ‘very easy’
- Q2: ‘How easy was it to see the teacher, teacher’s writing and [TRTD3/TRTD6/SRTD3/SRTD6]?’, with the same Likert scale response as before.
- Q3: ‘How helpful is it to see the teacher, teacher’s writing and [TRTD3/TRTD6/SRTD3/SRTD6]?’, with a Likert scale that ranged from 1 through 5, with 1 being ‘Not at all helpful’ to 5 being ‘very helpful’
- Q4: ‘Would you recommend [TRTD3/TRTD6/SRTD3/SRTD6] to others?’, with a Likert scale that ranged from 1 through 5, with 1 being ‘Not at all recommend’ to 5 being ‘very much recommend’.
- O1: ‘What are the strengths of [TRTD3/TRTD6/SRTD3/SRTD6]?’
- O2: ‘What are the weaknesses of [TRTD3/TRTD6/SRTD3/SRTD6]?’

Results

Based on the answers to the demographic question D1, a total of 30 hearing students and 1 deaf student from the section participated in the study to compare the four conditions. Since there was only one deaf student, we only report the feedback from the deaf student. The hearing students strongly preferred the static display of real-time captions over tracked display of real-time captions, so we compared only these two conditions: SRTD3 and SRTD6. Since the Likert scores were skewed and not normal, we analyzed the responses using the non-parametric Wilcoxon Signed-Rank test to evaluate whether there was a statistically significant difference.

The Q1 responses (How easy is it to read [SRTD3/SRTD6]?) indicated that the hearing students significantly preferred 6 lines over 3 lines in a static location($Z = 1.371$, $p < 0.01$).

The Q2 responses (How easy was it to see the teacher, teacher’s writing and [SRTD3/SRTD6]?) indicated that the hearing students had no statistically significant preference between 6 lines or 3 lines in a static location ($Z = 3.522$, $p = 0.372$).

The Q3 responses (How helpful is it to see the teacher, teacher’s writing and [SRTD3/SRTD6]?) indicated that the hearing students had no statistically significant preference between 6 lines and 3 lines in a static location ($Z = 3.771$, $p = 0.379$).

The Q4 responses (Would you recommend [SRTD3/SRTD6] to others?) showed that the hearing students had a statistically significant preference for recommending 6 lines over 3 lines in a static location to other students ($Z = 1.261$, $p < 0.01$).

Participant comments

Several common themes emerged on analyzing the feedback. The deaf student reported that they did not notice a delay in the captions. They also noted that regardless of lines shown, it was easier for the student to see the teacher, whiteboard and the captions when the text followed the teacher than when it did not. The student also noted it was easier to follow the teacher’s expressions and body language more clearly, and felt more ‘connected’ with the teacher with tracked rather than static display.

By contrast, the hearing students reported that it was very distracting to deal with the delay between the actual speech and the transcribed text, and that it was harder to read the text when it was moving. Regarding the delay between the speech and text, they found it distracting to listen to the speech and to read the captions. They also wanted to leave the text in a more static location to minimize their search process as they listened to the speech and looked up to search for the key words in the text display immediately after they either misheard or missed it completely. Both the deaf student and hearing students wanted to see different amounts of transcript lines for different reasons.

Conclusion

The deaf student reported that the transcript was the only access to spoken information, so the 4-5 second delay between the speech and transcribed text was not an issue. One way to interpret the student's preference for tracking can be that the student found it quick and easy to switch views and find the relevant information to read immediately and do less searching. When six lines were displayed, the student found it difficult to scan through the lines to locate the relevant part of the transcript.

By contrast, the hearing students significantly preferred more lines at a static location, because they used the transcript display as a backup as they listened to the teacher. They would review information on the transcript if they misheard or did not understand a specific word or sentence. So their preference for 6 lines at a static location can be interpreted to mean that they wanted enough lines to review missed information merely by looking at the upper lines of the display. The responses suggested that students use transcript in different ways, and that it is important to offer an user adjustable number of lines or to switch between tracked and static location.

The hearing students were annoyed by the 4-5 second delay between the speech and transcript. Most hearing participants noted that three lines were not enough because each line displays around one second worth of speech. Six lines of text would be roughly about six seconds worth of speech, which allows hearing students to read the transcript if they misheard, provided that the the delay is less than six seconds. This suggests that the number of caption lines displayed should be adjustable, according to the teacher's speaking rate and delay of the captioner, in order to better serve the needs of students. Our results show that providing additional captions history may be worth the attention switching overhead for looking away and resuming reading. The additional history enabled the viewers to review words that they have not heard or seen.

The survey results indicate that both the deaf student and hearing students benefited from the captions and would recommend it to their peers. For the deaf student, the tracked display of captions reduces visual dispersion, and creates a more inclusive and versatile classroom environment. For the hearing student, the The findings suggest that the number of caption lines should be adjustable, so that they can keep up with the text. The feedback from hearing and deaf students indicate that latency is a serious annoyance for all students, more so for hearing students than for deaf students. Overall, the responses suggested that deaf and hearing students benefit from transcripts, but use them in different ways, which influences their preferences.

References

1. Erickson W, Lee C, Von Schrader S. Disability Statistics from the 2011 American Community Survey (ACS). 2013.

2. Aud S, Hussar W, Kena G, Bianco K, Frohlich L, Kemp J, Tahan K. The Condition of Education 2011. NCES 2011-033. National Center for Education Statistics. 2011.

3. Marschark M, Pelz JB, Convertino C, Sapere P, Arndt ME, Seewagen R. Classroom Interpreting and Visual Information Processing in Mainstream Education for Deaf Students: Live or Memorex(R)? American Educational Research Journal. 2005 [accessed 2010 Sep 7];42(4):727–761. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=1440927

4. Cavender AC, Bigham JP, Ladner RE. ClassInFocus: Enabling improved visual attention strategies for deaf and hard of hearing students. In: Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS '09. New York, New York, USA: ACM Press; 2009. p. 67–74. http://portal.acm.org/citation.cfm?doid=1639642.1639656

5. Kushalnagar RS, Lasecki WS, Bigham JP. Accessibility Evaluation of Classroom Captions. ACM Transactions on Accessible Computing. 2014;5(3):1–24.

6. Behm GW, Mondragon AF. A Teaching Model for Teaching Deaf/Hard-of-Hearing and Hearing Students with Course Accessibility and Real World Product Design. In: 2014 ASEE Annual Conference. ; 2014. p. 1–13.

7. Marschark M, Sapere P, Convertino C, Pelz J. Learning via direct and mediated instruction by deaf students. Journal of Deaf Studies and Deaf Education. 2008 [accessed 2010 Aug 22];13(4):546–561. http://www.ncbi.nlm.nih.gov/pubmed/18453639

8. Kushalnagar RS, Behm GW, Kelstone AW, Ali S. Tracked Speech-To-Text Display: Enhancing Accessibility and Readability of Speech-To-Text. In: ACM, editor. Proceedings of the 17th International ACM SIGACCESS Conference on Computers and Accessibility. Lisbon, Portugal: ACM; 2015. p. 223–230.

9. Elliot LB, Stinson MS, Easton D, Bourgeois J. College Students Learning With C-Print's Education Software and Automatic Speech Recognition. In: American Educational Research Association Annual Meeting. New York, NY: AERA; 2008.

10. Elliot L, Stinson M, Coyne G. Student learning with C-Print's educational software and automatic speech recognition. In: American Educational Research Association Annual Meeting, San Francisco, CA. ; 2006. p. 1–22. http://www.ntid.rit.edu/research/cprint/pdf/AERA 2006.pdf

11. Mayer RE, Moreno R. A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. Journal of Educational Psychology. 1998;90(2):312–320.