Evaluating Alternatives for Better Deaf Accessibility to Selected Web-Based Multimedia

Brent Shiver
IBM Accessibility
IBM Research - Austin
Austin, Texas, USA
bnshiver@us.ibm.com

Rosalee Wolfe
College of Computing and Digital Media
DePaul University
Chicago, Illinois, USA
wolfe@cs.depaul.edu

ABSTRACT

The proliferation of video and audio media on the Internet has created a distinct disadvantage for deaf Internet users. Despite technological and legislative milestones in recent decades in making television and movies more accessible, there has been less progress with online access. A major obstacle to providing captions for Internet media is the high cost of captioning and transcribing services.

This paper reports on two studies that focused on multimedia accessibility for Internet users who were born deaf or became deaf at an early age. An initial study attempted to identify priorities for deaf accessibility improvement. A total of 20 deaf and hard-of-hearing participants were interviewed via videophone about their Internet usage and the issues that were the most frustrating. The most common theme was concern over a lack of accessibility for online news. In the second study, a total of 95 deaf and hard-of-hearing participants evaluated different caption styles, some of which were generated through automatic speech recognition.

Results from the second study confirm that captioning online videos makes the Internet more accessible to the deaf users, even when the captions are automatically generated. However color-coded captions used to highlight confidence levels were found neither to be beneficial nor detrimental; yet when asked directly about the benefit of color-coding, participants strongly favored the concept.

Categories and Subject Descriptors

K.4.2 **[Computers and Society]**: Social Issues – *Assistive technologies for persons with disabilities.*

General Terms

Design, Experimentation, Human Factors, Economics.

Keywords

Multimedia accessibility, web accessibility, captioning, deaf, automatic speech recognition, speech-to-text.

1. INTRODUCTION

In recent years, researchers have investigated the accessibility of online resources to support deaf students [1] [2] including Web-based materials [3] [4] and social communities [5]. However, less discussion has occurred about the accessibility of the Internet to deaf adults who were born deaf or became deaf at an early age. This population typically differs in its language preference from those who became deaf later in life. In the United States, they use American Sign Language as their preferred language and view English as a second language.

According to a recent survey [6], adults commonly use the Internet for reading email, searching for information, finding answers to health questions, and reading the news. In recent years, these activities increasingly involve more than the printed word -- they involve multimedia [7]. Videos often accompany written text, or even replace it. These make the Internet increasingly less accessible to the deaf community.

To address this issue, the W3C Web Accessibility Initiative developed the Web Content Accessibility Guidelines (WCAG) for accessibility solutions [8]. Further, Section 508 of the Rehabilitation Act requires that U.S. Federal agencies' electronic and information technology be accessible to people with disabilities [9]. A recent study reconfirms that these guidelines and regulations are well warranted. Researchers [10] found that the barrier of "multimedia content without text alternative" to be the most critical for users who are deaf and hard-of-hearing.

The year 2014 witnessed modest gains in addressing this issue. The FCC recently finalized new quality standards for closed captioning of broadcast television [11] and now requires that any program televised after January 1, 2016 with captions must retain its captions when shown online [12]. However the new regulations do not address any media that was previously broadcast, nor does it cover videos that were published directly to the Web. Thus the majority of multimedia on the Internet remains inaccessible to those who are deaf or hard-of-hearing.

Even with assistance from legislative statute, it is virtually impossible to manually caption every single video or audio clip on the Internet due to the staggering cost. Captioning a video manually costs approximately $9 - $30 per minute [13] [14]. The costs cover skilled captionists who not only transcribe the audio content, but also keep video's text and audio in sync.

A transcription simply contains the text corresponding to the spoken words in a video. Transcriptions have several advantages over captions. When people begin reading a transcript, they have immediate access to the entire transcript and are limited only by the speed of their reading. In contrast, captioned text is synchronized with the video and the text is only revealed as the person in the video speaks. Viewers have to divide their attention between the captions and the visual activity in the video [15]. Captioned text is limited by the speed of produced speech (180-220 words per minute), which on average is slower than the speed of reading (250-300 word per minute) [16] [17]. Transcripts cost less than captioning, but they remain expensive, usually costing $1.50 - $3.00 per minute [18]. A lower-cost alternative is automatic speech recognition (ASR) [19] [20] [21], but it does not afford the accuracy of a trained captionist or transcriptionist [22].

There is a need to develop a rationale for leveraging scarce resources to do the most good. High-quality, manual captioning is prohibitively expensive. Transcripts offer a potential advantage due to their lower cost. Captioning by ASR is cheap, but inaccurate. In order to properly inform decisions about allocating resources, we asked two groups of related questions:

1. What aspects of using the Internet prove most difficult? For multimedia content, would transcripts be an acceptable alternative to captions? What is the current user experience with automatic captioning?
2. If there were no possibility of hiring a skilled captionist or transcriptionist, would lower-quality text, generated by ASR, be better than nothing? Would a visualization of the text's quality be useful?

Addressing these questions required two studies. The first study assessed the patterns of Internet usage among the adult deaf population to understand if these patterns were similar to those in the general population. It also investigated which types of multimedia content posed the most critical barriers, and gathered preference opinion on the viability of using transcripts instead of captioning. The results from this study could potentially lend insight into how to set priorities for accessibility efforts.

In addition, the results from the first study shaped the context for the second study. This was a follow up effort that considered the user experience when viewing text that was generated by ASR as contrasted with viewing text that was created manually.

2. EXPLORATORY STUDY

One of the goals of the first study was to identify multimedia content that was considered a high priority in the deaf community that would then serve as a case study for applying and evaluating speech visualization technology. Since the current literature did not provide a clear indication of a compelling choice, this exploratory study was necessary. The study proposal was evaluated and approved by DePaul Institutional Review Board (IRB) as noted by DePaul IRB #JR052311NUR.

2.1 Participants

A total of 20 deaf and hard-of-hearing adults from various parts of the United States participated in the study. They were recruited primarily through email invitation and some others were contacted through social media outlets such as Facebook. Nineteen of 20 participants were profoundly deaf, and one participant was hard-of-hearing. Fourteen participants stated that they were born deaf while five reported becoming deaf younger than five years old. Just three became deaf at the age of five years or older. Eleven of them were aged 30-39, while nine were 40 years or older. Thirteen were identified as male and seven female.

2.2 Procedure

All interviews were conducted by a deaf facilitator via video phone. The first step of the interview was to gain informed consent from the participants. The facility provided them with an information sheet explaining the study and notifying the participant that the interview would be recorded. The participants had an opportunity to ask questions or raise concerns prior to the interview. Each participant answered several basic demographic questions and a set of questions about Internet usage. Each then responded to 13 open-ended questions regarding their experience in using multimedia over the Internet.

2.3 Interview Results

Table 1 summarizes the Internet activities selected by the participants, rank ordered by frequency of use. Email was the most frequent activity, followed by getting the news, and accessing social media sites such as Facebook or Twitter. Watching YouTube videos occurred less frequently, as did shopping online. Online auctions and education/training were the least-frequently occurring activities.

Table 1: Patterns of Internet usage by study participants

Activity	Many times a day	Once a day	Once a week	Rarely	Never
Email	19	1			
News	10	6	3	1	
Social media sites	8	7	3	2	
YouTube	2	6	7	5	
Online shopping			10	10	
Online auctions			3	14	4
Education / training			2	16	2

Responses from the 13 open-ended questions were transcribed from sign language to English. A card sort analysis [23] was conducted on the responses to each individual question to identify patterns of commonality in the responses. The following is a listing of each question and a summary of the responses, based on the card-sorting results.

1. *When you read news articles on the Internet, do you ever watch the videos? Why or why not?*

When asked whether they watch news-related videos on the Internet, all twenty participants agreed that news videos are not useful without captions. Some of them pointed out that they tend to read news articles as opposed to videos online. One participant emphasized that if there are no captions, it isn't worth watching. Another said, "It's not worth my time."

2. *Have you watched YouTube?*

All 20 users have visited YouTube at least once when using the Internet to watch videos. Out of 20, 15 watch it "sometimes." Five pointed out that videos are often not captioned. Six users preferred watching Deaf-oriented or ASL-signed videos and two mentioned teaching ASL using YouTube.

3. *Have you ever found yourself needing the information on a video?*

Everyone responded at least “sometimes” but 16 of them pointed to lack of captions. Five complained of being stuck because videos lacked captions or were not accessible. Four users mentioned resorting to searching via Google or other search engine for texts related to the video.

4. *Can you describe the type of video it was?*

Fourteen respondents reported that the video they wanted was news-related and five mentioned CNN specifically. Knowledge about what is going on in the world was emphasized by four participants. Finally, four users mentioned interest in using training videos to increase knowledge and skills and keep up with current trends. Five users complained of being referred to a page with video, but after clicking on the link they realized that it was a video without captions and thus the information was inaccessible.

5. *What do you do to obtain the information or contents from videos?*

All participants would resort to reading related text or articles when available. Half of them would use Google or search online for related text, articles, and/or posts such as Facebook to learn more about contents from videos. Three users would attempt to contact the source or author to request transcripts of the video. Three participants would resort to asking an interpreter or hearing person to help translate selected videos.

6. *How often are you frustrated about inaccessibility when you use the Web?*

Sixteen users experienced frustration while four have either developed a tolerance or have given up altogether. Eight reported frustration every day and seven reported being frustrated sometimes. A participant succinctly compared the unavailability of captions to the stirred-up feelings of frustration that occur when the Internet is down.

7. *Describe the top three frustrations that you've experienced. What happened?*

The top frustration was lack of captions on new video clips. Seventeen participants mentioned this. Five users pointed out non-captioned self-tutorials and e-learning videos as a problem. Five viewers complained about lack of captions on YouTube. Finally, three users expressed disgust when clicking on a link only to be redirected to a video without captions

8. *Are you familiar with Google automatic captions? If yes, please tell me about your experience.*

Seventeen respondents had some experience while the other three didn't have any. Twelve pointed out that the captions had accuracy issues or too many errors, and four thought they were useless or too hard to follow. However, three users felt it was a good start and a step in right direction. In addition, two said it was better than no captions at all. Two participants thought it might be useful for hearing people who spoke another language.

9. *Have you used any other automatic speech recognition technology? If so, what was it? How did it work for you?*

Fourteen respondents stated they have not used ASR technology, five have used it, and one did not respond. Three users mentioned either Dragon Naturally Speaking [24] or iPhone's Siri feature [25] but they did not elaborate on how they used it.

10. *Now I need to review a couple of items of terminology with you. Captioning is the process of displaying text on a television, video screen or other visual display. Captions typically show a transcription of the audio portion of a program as it occurs. A transcript is a document containing a complete written or printed version of content originally presented as a video or recording. Which approach do you prefer? What are the advantages and disadvantages of each?*

Captions were preferred by 19 participants because the captions were usually better than transcripts. However one participant said it is fine either way. Nine described reading transcripts as being harder to use and requiring too much effort. Eleven pointed out that captions made “logical sense” because they’re always in sync with the video.

11. *Are there any situations where you prefer captions over transcripts and vice versa?*

Seven participants favored captions in all situations while five preferred captions and mentioned that transcripts were only acceptable as a backup. Four pointed out that transcripts could be useful for reference, information verification, and research. Six would rather have captions because they are in sync with videos and are easier to follow.

12. *We are currently investigating technologies that may improve accessibility. Which situations that you mentioned earlier do you feel this would benefit the most?*

News-related videos were mentioned 15 times while four participants demanded all videos to be captioned. Four believed that all television shows should be accessible and three felt that investment or financial clips should be covered. Three emphasized any television shows or movies that are already captioned should also be captioned online.

13. *Do you have any advice or suggestions in regards to improving accessibility on the Internet?*

Five participants suggested additional government involvement such as legislation, lawsuits, and FCC, and three recommended educating people about the need because there may be a lack of awareness. Four demanded everything to be captioned. Three people suggested that any videos that have already been captioned should also be captioned online. Lastly, three thought speech recognition might be useful.

2.4 Analysis

This study addressed the first group of questions mentioned in the introduction, which were:

1. What aspects of using the Internet prove most difficult?
2. For multimedia content, would transcripts be an acceptable alternative to captions?
3. What is the user experience with automatic captioning?

In regards to the question, “What aspects of using the Internet prove most difficult?” the most common theme was the lack of accessibility to news online videos. All of the participants mentioned that they have used the Internet to access the news and most of them (16/20) did so at least once a day. Some participants were especially disappointed with well-known news outlets because they made little effort to make their videos accessible.

In regards to the second question about the acceptability of transcripts as an alternative to captions, the responses indicated strongly that captions were preferable to transcripts for news videos. Of the 20 participants, 19 indicated a preference for captions, and the remaining participant said that either format was equally preferable. Further, nine of the participants described reading transcripts as being harder than reading captions and required too much effort. More than half of the participants pointed out that captions made "logical sense" because they're always in sync with the video.

Further, when asked if there were any circumstances where transcripts would be preferable to transcripts, 16 of the 20 participants responded in the negative. Five participants mentioned that the only use for transcripts were as a backup to captions.

The consensus is that captions are designed to be synchronous with the videos which make them easier to follow. There was also consensus that it takes extra work to go back-and-forth between viewing the video and reading a separate transcript. Captions shown on the video itself were perceived as easier to follow.

In regards to the final question, "What is the user experience with automatic captioning?" participant responses were mixed. Of the 17 participants who had tried automatic captioning, 12 found that the results had too many errors to be useful. However, three of the participants thought that the technology was promising and two of them stated that automatic captions were better than no captions at all.

From this first study we understood that making Internet news videos more accessible was a high priority for members of the deaf community. Further, the data indicated that captions were preferable to transcripts for the news videos, so we decided to focus our efforts on captioning. Even though transcripts are more economical to produce, the test participants strongly indicated that captioning was preferable.

Based on this information, we focused our second study on evaluating access to Internet news using captions generated through ASR. We hoped that having the opportunity to explore different styles of automatic captioning might encourage interest in utilizing this technology.

3. VISUALIZATION STUDY

This study investigated the second set of questions (listed in the introduction) that concern the acceptability of automatically-generated captions. The study addressed these questions in the context of accessing news videos. We chose news videos because they were the type of video mentioned most often in the first study. The second study also considered the utility of indicating possible errors in the captions through a visualization of the word confidence level. The study was reviewed and approved by the DePaul University Institutional Review Board (BS031313CDM).

3.1 Participants

Participants were volunteers recruited through Deaf mailing lists. Additional participants were discovered through forwarding of the solicitation email. All were 18 years or older, were deaf or hard-of-hearing and had at least some college education and had watched captioned videos. A total of 95 people participated in the study.

3.2 Stimuli

The stimuli were four videos simulating news stories, each captioned in a different style:

1. Captions created through ASR. Words recognized at a higher confidence level were displayed in a more prominent color.
2. Captions created through ASR but without the visualization technique.
3. No captions.
4. Manually-created captions.

The role of stimuli 3 (no captions) and 4 (manually-created captions) were to act as a worst case and a best case respectively. Stimulus three would simulate the barriers described in [10] and stimulus four would simulate the use of a trained captionist.

There is no perfect solution for creating stimuli for this type of test. For full control, each stimulus would have to present the same news article. However, this would create a large transfer-of-learning bias since the same story would be repeated with each treatment, and the participants would be answering the exact same questions about content on four separate occasions. Thus it was necessary to identify four different news stories.

Using actual news stories as test stimuli posed an additional problem because viewers may have seen the story previously and have prior knowledge of its content. The question became, "Where can we find stories that are not actual news stories, and how do we control for the level of difficulty of stories and the content questions?"

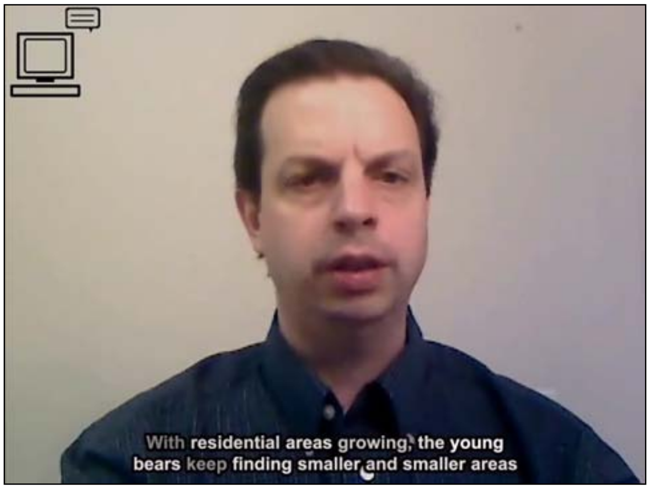
To control for this possibility, this study utilized four simulated news stories by selecting material from standardized 8th-grade reading tests. The material needed to be believable as a news story, but also be from a reading test that had been previously validated for level of difficulty. Four reading passages chosen from the standardized tests were produced as news stories. [26] [27]. The questions from the original standardized reading tests served as the basis for the performance metrics in this study.

Videos were created from the stories by recording a single speaker who read the stories from a software teleprompter in an environment consisting of a neutral background, a table and a chair. To control for nonverbal cues, the news reader kept his arms on the table and used a neutral tone and facial expression through the reading. We created custom software utilizing the Microsoft Speech SDK to perform the recognition and to produce the timing and color-coding for captions in the SubStation Alpha format, which we then combined with the video via Virtual Dub. We then verified that the captions were temporally aligned to the audio track of the original video before stripping the sound from the final form of the video. To emphasize, the choice of speech recognition tool is not the point of the study – at present no speech recognition tool performs perfectly. What we wanted to evaluate was whether ASR-generated captions are a viable alternative when there are no manually-created captions available.

This resulted in simulated news videos that were controlled for speaker, speaker environment, nonverbal communication as well as reading difficulty. When reformatted as Internet media, stimuli

1, 2 and 4 received captions, and stimulus 3 received no captions. The manually-captioned stimulus had no errors, and the WER did not apply to the video since it did not have captions. Because the stories were different, the performance of the speech recognizer differed as well - giving a WER of 20% in the first instance and 12% in the second. However, according to [28] the word error rates for the first two stimuli were sufficiently low that the captions retained their utility.

Figure 1 is a screen shot from stimulus one. The videos captioned through ASR also carried an icon in the upper left corner that showed a voice bubble emanating from a computer. The intent of the icon was to indicate that the captions were generated through software that might not produce results that were as accurate as those created by professional captionists.



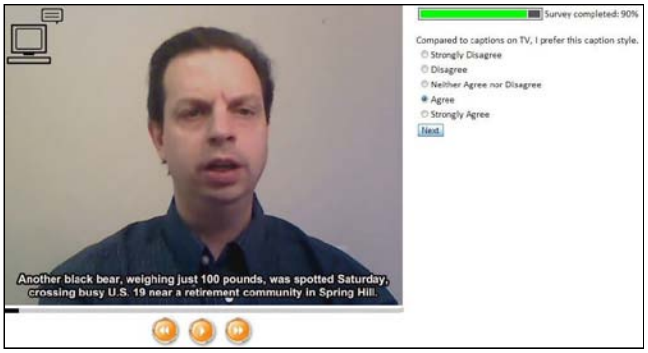
With residential areas growing, the young
bears keep finding smaller and smaller areas

Figure 1: Simulated news video with automatic captions visualized for word confidence level

A fully factorialized design using the four conditions was not possible due to the small number of available participants. Since we anticipated that the condition with a WER of zero would produce the best performance results, we presented this stimulus last for all of the participants. It was already anticipated that this stimulus would serve as the upper bound for participant performance and preference; placing it last would take advantage of any transfer of learning that did occur in the test and would only serve to enhance the upper bound. The remaining three stimuli were fully randomized.

3.3 Procedure

The tests were performed online. Participants read the informed consent and filled out a qualifying questionnaire to confirm their eligibility for the study. They then viewed the four simulated news videos, ranging from 2:19 to 4:12 minutes long. The first three videos were presented in randomized order to minimize order bias, while the one with perfect captions was presented last. At the end of each video, the participant answered questions that were extracted from the standardized reading tests about the video's content. While answering the questions, the participant could replay the video as often as desired. After answering the content-related questions, they rated their viewing experience. After the four videos, participants answered a post-test questionnaire, which asked the participant to rate and compare all four captioning styles. An honorarium of a $15 gift card was emailed to the participants at the conclusion of the test.



Survey completed: 90%

Compared to captions on TV, I prefer this caption style.

○ Strongly Disagree
○ Disagree
○ Neither Agree nor Disagree
● Agree
○ Strongly Agree

Next

Another black bear, weighing just 100 pounds, was spotted Saturday,
crossing busy U.S. 19 near a retirement community in Spring Hill.

Figure 2: Screen shot of the test instrument from the user preference portion of the evaluation.

3.4 Results and Analysis

Table 2 lists performance metrics as the mean number of correctly answered questions for each captioning style. A single-factor analysis executed on the means showed that higher scores for the captions generated by ASR were significant. A post hoc analysis employing Tukey simultaneous comparison t-values showed that there were only two pairs the reflected significant differences. These were visualized ASR captions compared to no captions and ASR captions compared to no captions. The rest of the pairs fell below the critical value of 3.18 (p = 0.01), including comparisons with the perfectly-captioned video.

The study asked two sets of preference questions. The first set measured the participant initial reactions with the captioning style. Participants were asked to rate several aspects of the captioning style on a 5-point Likert scale (0 = strongly disagree, 4 = strongly agree). A Kruskal-Wallis analysis converted the responses into rankings. See Table 4.

For the question "The captions were easy to read," a Mann-Whitney calculation with Bonferroni adjustment identified all pairs except for Caption Style 1 (ASR captions with visualization) compared to Caption Style 4 (manual captions) as being statistically significant. For question two, "The captioning made it easy to understand the story," each of three captioned styles compared to the non-captioned style were statistically significant, but there were no differences among the three styles with captioning. The analysis of responses to question three, "I have confidence in the accuracy of the captioning," yielded a similar result. The analysis of the rankings for question four, "I like this style of captioning," produced results that were similar to the results for question 1.

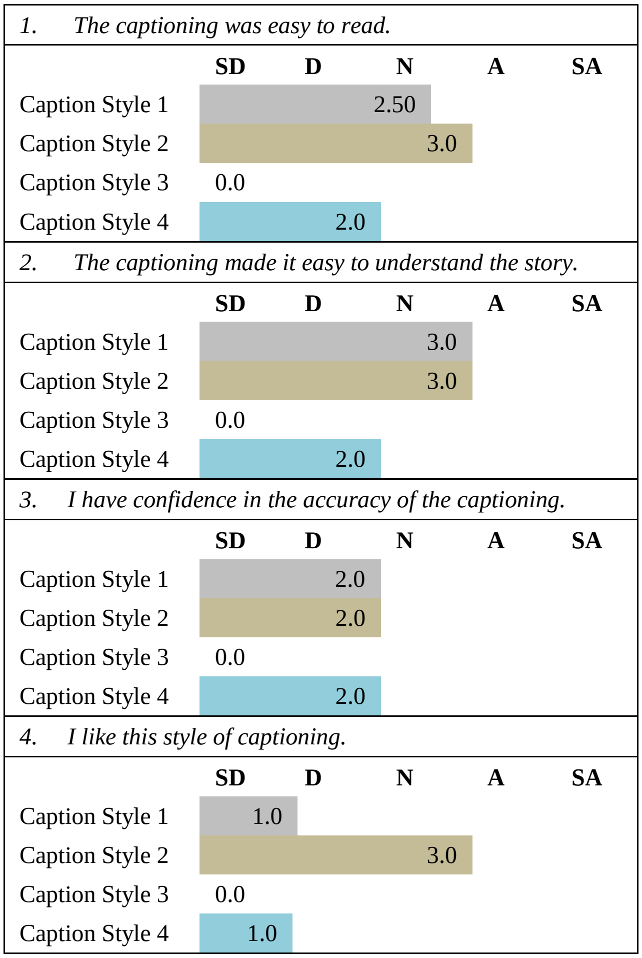
Table 2: Correctly-answered content questions as a function of caption style

Caption Style	Mean	n	Std. Dev.
1. ASR captions, visualized	0.7105	95	0.26376
2. ASR captions	0.6877	95	0.23349
3. No captions	0.5579	95	0.28818
4. Perfect captions	0.6158	95	0.21481

Table 3: Pairs with significant differences

Pair	Score
Caption Style 1 vs. Caption Style 3 (no captions)	3.56
Caption Style 2 vs. Caption Style 3 (no captions)	4.18

Table 4: Participants' initial reactions to captioning styles



There was one additional response gathered for Caption Style 1, which used highlighting to convey the word confidence levels. Over two thirds (68 of 95) of the participants disagreed or strongly disagreed with the statement, "The color coding was helpful for understanding the story."

After viewing all of the videos, the participants responded to a second set of questions. These measures were taken after the participant had viewed all four captioning styles. This gave participants the opportunity to reflect on the various styles and compare all of them. When asked which was the hardest to understand, most the participants (80 of 95) chose Caption Style 3, which had no captions. When asked which was the easiest to understand, the selections were Caption Style 2 (44), Caption Style 4 (32) and Caption Style 1 (19).

When asked "If you had a video with automatic captions containing errors, would you want color coding to indicate possible errors?" 73 of 95 responded, "Yes." When asked, "Did you see any indication on the screen that indicated that the captions were generated through automatic speech recognition?" 51 of the 95 participants responded "Yes." When asked, "Is there a better way to indicate that captions are created with automatic speech recognition?" 33 of 95 responded "Yes."

When asked, "If you had a choice of watching a video with no captions or watching a video with captions created through automatic speech recognition, which would you choose?" the majority (85 of 95) chose the latter.

3.5 Discussion

The increasing use of inaccessible multimedia on the Internet is obviously detrimental for the deaf community. Manual captioning, providing transcripts and using ASR to generate captions are some of the possible remedies, but each has its advantages and disadvantages.

Both performance and preference results of this study lend credence to the hypothesis, "If there were no possibility of hiring a skilled captionist to create high-quality captions, would low-quality captions, generated by ASR, be better than nothing?" However, the second hypothesis, "Would visualization of the caption's quality be useful?" was not strongly supported. While it seemed like a given that ASR-generated captions would be better than nothing, we were surprised that the visual indication of this type of captions was not strongly supported at the outset of the investigation.

In the performance portion of the test, participants scored significantly higher on the videos having captions generated by ASR, both with and without visualizations of the word confidence level. When rating the captioning styles, participants found that the captions generated by ASR where easier to read and made it easier to understand the story. Further, the style most often rated as easiest to read was captioning generated through ASR. When asked the direct question, "If you had a choice of watching a video with no captions or watching a video with captions created through automatic speech recognition, which would you choose?" the overwhelming majority (85 of 95) responded in the affirmative. We believe these results would be valuable to broadcast to decision makers who are choosing whether or not to caption Internet multimedia.

The icon indicating that the captions were generated via ASR had mixed success. A little over half (53%) of the participants noticed its appearance on the news videos, and a third of the participants felt that the icon did not do an effective job of communicating its intent.

The data collected to evaluate the second hypothesis yielded mixed results. When comparing captioning styles head-to-head, participants preferred the captioning style without visualizations. However, when asked the question whether they would want a visualization to show that the captions were created via ASR instead of being manually transcribed, over 75% of the participants indicated that they wanted visualization. More research appears to be necessary in order to determine the viability of the indicator icon. One interesting avenue to explore regarding this visualization is whether, after most videos are captioned with varying degrees of quality, people would be more likely to want the icon so they would know what quality level to expect.

4. RELATED WORK

These results reinforce the results of previous studies [10] that found the multimedia content without text alternative is a critical barrier for users who are deaf and hard-of-hearing. Using ASR to generate captions results in word error rates that are consistent with other strategies for producing captions at low cost [29] [30]. The results also suggest that closed captions generated by ASR [20], [21] may be viewed as a viable alternative by deaf users when it is not possible to hire a trained captionist.

The desirability of visualizations of word confidence levels remains an open question. Consistent with the hearing participants in a study by [31], participant preference ratings in this study showed no significant difference between the captions with and without visualizations. However, over three quarters of the participants stated that they wanted the visualizations, consistent with the findings in [32].

5. LIMITATIONS

The performance metrics showed that even when there were no captions available, participants answered more than 50% of the content questions correctly. On reexamination of the content questions, it may have been that several of the questions could have been answered using previous knowledge. Also, the performance on the perfectly-captioned video was lower than that on the two videos captioned through ASR. One might reason that having no errors in the captions would lead to a higher score in answering the content questions. One possible explanation might be the order of presentation. The goal of placing the manually-captioned video last in the presentation order was to give it the benefit of any transfer of learning that may have occurred. However, participants may have become fatigued from the lengthy study.

6. FUTURE WORK

The future work will involve exploring and identifying visualization techniques that help make ASR-generated captions more useful. Creating a better visual indicator that the captions are actually generated by the speech recognition engine may help serve as a reminder that they are not perfect and will include errors. This may foster wider acceptance of ASR-based captions as a possible alternative to non-captioned videos. Additional studies are necessary to develop appropriate color-coding options to help identify words having lower confidence levels. User preferences could be utilized to allow users to customize how the colors should be presented. For example, the user can tweak how transparency, colors, and styles (underline, strikethrough, or other patterns) appear when the words fall under certain confidence levels.

It should be pointed out that the captions created by a speech recognition engine still are not of the same quality as those created by a skilled captionist. Much work needs to be done before this cost-effective alternative could become a reality. It is possible that the alternative could become a "backup" solution when a deaf or hard of hearing person comes across a non-captioned video. However, deaf advocacy groups could be concerned that organizations may attempt to substitute automatic captions in order to meet legal obligations.

Alternative strategies involving crowdsourcing might help improve the quality of automatic captions through low-cost means [29]. When ASR software has access to a speaker's voice profile, the resulting recognized text has higher accuracy. One possibility would be to maintain voice profiles for speakers online. When users desiring automatic captions submit a video, they could supply the identity of the speaker. The result would be more accurate captions. Another intriguing alternative would be to explore the application of gaming techniques similar to the crowdsourcing for soliciting volunteers for manual captioning in [30].

The benefits of utilizing speech recognition to improve Internet accessibility for deaf users are endless. It would help narrow the accessibility gap that deaf Internet users experience daily and lead to leveling of the playing field. No one should be denied access to the abundance of information that the Internet has to offer. "Knowledge is power," is a well-known quote coined by Francis Bacon in 1597 in the *Meditationes Sacrae* [33] and it resonates well with the motivation behind this work.

7. ACKNOWLEDGEMENTS

This work was performed while Brent Shiver was a student at DePaul University. A special appreciation goes to John McDonald who graciously volunteered to serve as the news anchor for the news clips. We also thank the American Sign Language Synthesizer Project team at DePaul University for taking the time to review and share feedback on the clips and online survey. Finally, we wanted to recognize the deaf participants of both experiments for contributing their time and valuable input**.**

8. REFERENCES

[1] Bain, K., Basson, S., and Kanevsky, D. Accessibility, transcription, and access everywhere. *IBM Systems Journal* (2005), 589-603.

[2] Sheng, L. and Xu, J. Using social software to improve learning performance of deaf university learners. In *The 2nd IEEE International Conference on Information Management and Engineering (ICIME)* (Chengdu, China 2010), IEEE, 703-706.

[3] Sloan, D., Stratford, J., and Gregor, P. Using multimedia to enhance the accessibility of the learning environment for disabled students: reflections from the Skills for Access project. *ALT-J, Research in Learning Technology* (Mar 2006), 39-54.

[4] Freire, A. P., de Bettio, R. W., Frade, E. G., Ferrari, F. B, Monserrat Neto, J., and Libardi, H. Accessibility of web and multimedia content: techniques and examples from the educational context. In *Proceedings of the 19th Brazilian symposium on Multimedia and the web* (Salvador, Brazil December 2013), ACM, 7-8.

[5] Kožuh, I., Hintermair, M., Ivanišin, M., and Debevc, M. The concept of examining the experiences of deaf and hard of hearing online users. *Procedia Computer Science*, 27 (2014), 148-157.

[6] Zickuhr, Kathryn. 2010. Online Activities. Pew Research Internet Project. Retrieved January 12, 2015 from http://www.pewinternet.org/2010/12/16/online-activities/.

[7] Cheung, A. and Slavin, R. How features of educational technology applications affect student reading outcomes: A meta-analysis. *Educational Research Review* (2012), 198-215.

[8] Caldwell, B., Cooper, M., Reid, L., and Vanderheiden, G. 2008. Web Content Accessibility Guidelines (WCAG) 2.0. W3C. Retrieved January 12, 2015 from http://www.w3.org/TR/WCAG20/.

[9] GSA IT Accessibility and Workforce. 2015. Opening Doors to IT. Section 508.gov. Retrieved January 12, 2015 from http://www.section508.gov/.

[10] Pascual, A., Ribera, M., and Granollers, T. Impact of web accessibility barriers on users with hearing impairment. In *Proceedings of the XV International Conference on Human Computer Interaction.* (Puerto de la Cruz, Spain 2014), ACM, 8.

[11] Phillips, A. 2014. FCC Issues New Rules to Improve Captioning Quality. National Association of the Deaf. Retrieved January 12, 2015 from http://nad.org/news/2014/3/fcc-issues-new-rules-improve-captioning-quality.

[12] National Association of the Deaf. 2014. FCC Requires Closed Captioning of Online Video Clips. National Association of the Deaf. Retrieved July 2, 2014 from http://nad.org/news/2014/7/fcc-requires-closed-captioning-online-video-clips.

[13] Automatic Sync Technologies, LLC. 2014. Closed Captioning Cost and Pricing. Caption Sync. Retrieved January 22, 2015 from http://www.automaticsync.com/captionsync/closed-captioning-cost/.

[14] CPC Computer Prompting & Captioning Co. 2015. CPC Closed Captioning Blog. CPC: Home of e-Captioning. Retrieved January 12, 2015 from http://cpcweb.com/blog/2012/09/the-top-5-closed-captioning-concerns-answered/.

[15] Lasecki, W. S., Kushalnagar, R., and Bigham, J. P. Helping students keep up with real-time captions by pausing and highlighting. In *Proceedings of the 11th Web for All Conference* (Seoul 2014), ACM, 39.

[16] Human Factors International. 2000. Human Interaction Speeds. Human Factors International. Retrieved January 12, 2015 from http://www.humanfactors.com/newsletters/human_interaction_speeds.asp.

[17] Smith, S. M. and Shaffer, D. R. Speed of speech and persuasion: Evidence for multiple effects. *Personality and Social Psychology Bulletin*, 21, 10 (1995), 1051-1060.

[18] CGBiz LLC. 2012. The Pre-Payment Requirement. Scribie: Audio Transcription, Perfected. Retrieved September 7, 2012 from https://scribie.com/blog/2012/09/the-pre-payment-requirement/.

[19] Wactlar, H. D., Kanade, T., Smith, M., and Stevens, S. M. Intelligent access to digital video: Informedia project. *Computer* (1996), 46-52.

[20] Jutla, D. N. and Kanevsky, D. wisePad services for vision-hearing-and speech-impaired users. *Communications of the ACM* (2009), 64-69.

[21] Google. 2009. Automatic captions in YouTube. Official Blog. Retrieved November 19, 2009 from http://googleblog.blogspot.com/2009/11/automatic-captions-in-youtube.html.

[22] National Institute of Standards and Technology. 2012. Speaker Recognition Evaluation. Information Technology Laboratory. Retrieved March 2, 2012 from http://nist.gov/itl/iad/mig/sre12.cfm.

[23] Fincher, S. and Tenenberg, J. Making sense of card sorting data. *Expert Systems*, 22, 3 (2005), 89-93.

[24] Nuance Communications, Inc. 2015. Dragon Speech Recognition Software. Dragon NaturallySpeaking. Retrieved January 17, 2015 from http://www.nuance.com/dragon/index.htm.

[25] Apple Inc. 2015. Siri: Your wish is its command. iOS 8. Retrieved January 17, 2015 from http://www.apple.com/ios/siri/.

[26] Public Schools of North Carolina. 2013. North Carolina End-of-Grade Tests -- Grade 8. Department of Public Instruction. Retrieved June 30, 2013 from http://www.ncpublicschools.org/docs/accountability/testing/releasedforms/grade8readingreleased.pdf.

[27] University of South Florida. 2013. FCAT Express Eighth Grade Reading. Florida Comprehensive Assessment Test. Retrieved June 30, 2013 from http://fcit.usf.edu/fcat8r/default.htm.

[28] Munteanu, C., Baecker, R., Penn, G., Thoms, E., and James, D. The Effect of Speech Recognition Accuracy Rates on the Usefulness and Usability of Webcast Archives. In *CHI Proceedings: Visualization and Search* (Montréal 2006), ACM, 493-502.

[29] Wald, M. Crowdsourcing correction of speech recognition. In *Proceedings of the 8th International Cross-Disciplinary Conference on Web Accessibility* (Hyderabad, India 2011), ACM, 22.

[30] Kacorri, H., Shinkawa, K., and Saito, S. Introducing game elements in crowsourced video captioning by non-experts. In *Proceedings of the 11th Web for All Conference* (Seoul 2014), 2014, 29.

[31] Vermuri, S., DeCamp, P., Bender, W., and Schmandt, C. Improving speech playback using time-compression and speech recognition. In *Proceedings of the SIGCHI conference on human factors in computing* (Vienna 2004), ACM, 295-302.

[32] Vertanen, K. and Kristensson, P. O. On the benefits of confidence visualization in speech recognition. In *Proceedings of the SIGCHI conference on human factors in computing systems* (Florence 2008), ACM, 1497-1500.

[33] Bartlett, J. 1919. Familiar Quotations, 10th ed. Bartleby.com. Retrieved August 20, 2013 from http://www.bartleby.com/100/139.39.html.