Evaluating the Benefit of Highlighting Key Words in Captions for People who are Deaf or Hard of Hearing

Sushant Kafle, Peter Yeung, Matt Huenerfauth
Golisano College of Computing and Information Sciences
Rochester Institute of Technology (RIT), Rochester, NY, USA
sxk5664@rit.edu, pxy9548@rit.edu, matt.huenerfauth@rit.edu

ABSTRACT
Recent research has investigated automatic methods for identifying how important each word in a text is for the overall message, in the context of people who are Deaf and Hard of Hearing (DHH) viewing video with captions. We examine whether DHH users report benefits from visual highlighting of important words in video captions. In formative interview and prototype studies, users indicated a preference for underlining of 5%–15% of words in a caption text to indicate that they are important, and they expressed an interest for such text markup in the context of educational lecture videos. In a subsequent user study, 30 DHH participants viewed lecture videos in two forms: with and without such visual markup. Users indicated that the videos with captions containing highlighted words were easier to read and follow, with lower perceived task-load ratings, compared to the videos without highlighting. This study motivates future research on caption highlighting in online educational videos, and it provides a foundation for how to evaluate the efficacy of such systems with users.

AUTHOR KEYWORDS
Captioning System; Deaf and Hard of Hearing; Caption Highlighting; Text Highlighting; User Study; Feedback.

CCS CLASSIFICATION
• Human-centered computing~Empirical studies in accessibility

INTRODUCTION
People who are Deaf and Hard of Hearing (DHH) often use captioning services to access audio or audiovisual information, with captions conveying aural information in the form of text. Usually a human transcriptionist transcribes the audio information to digital-text using a keyboard, with captions being displayed on a screen, usually below or alongside the visual source. In real-time contexts, such as live-events or meetings, a well-trained transcriptionist can produce accurate real-time captions with a speed of over 200 words per minute [41]. Other low-resource and cost-effective caption services are also becoming popular for providing real-time captioning by automating (e.g. automatic speech recognition systems) or semi-automating (e.g., human supervised automation) the speech-to-text transcription.

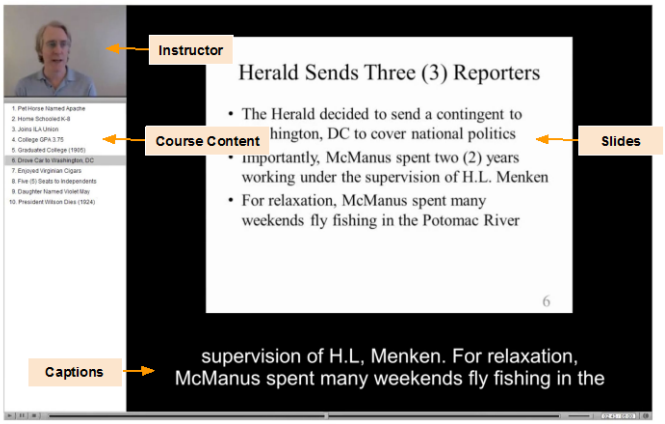

Figure 1. A typical arrangement of elements in an online educational video: with instructor, slides, and captions [18].

Although captions can make audiovisual content more accessible, there are still limitations. One challenge is the demand required on users' visual-attention, as they must access the visual information from the video and read the caption text simultaneously. This task is especially challenging when there are multiple visual references in the video or concurrent visual channels (e.g., multi-speaker settings). For instance, in an educational lecture setting, there may be multiple sources of simultaneous information content, e.g. the instructor, the slides, the captions, the sign language interpreters, etc. Figure 1 displays a screen configuration, typical of an online education platform [18], with a video of an instructor, the slides, and the captions displayed underneath. While hearing students may be able to process the audio speech and visual information concurrently, DHH students must switch their attention between the visual streams, which may result in students lagging behind or missing information [10].

Prior reading comprehension research has found that readers are often able to skim a text quickly for relevant information, rather than fixating on a portion for deep

ASSETS '19, October 28–30, 2019, Pittsburgh, PA, USA

ACM ISBN 978-1-4503-6676-2/19/10…$15.00
https://doi.org/10.1145/3308561.3353781

meaning [12, 14]. Consequently, researchers have investigated methods for highlighting important words or phrases in text to enable faster browsing and to support the reading task [22, 28, 30, 31, 36]; that work did not examine captioning contexts. Although recent computing research has investigated automatic methods for identifying important words in a caption text [23, 24], there has been a lack of research on the usability of highlighting key words for users. Such highlighting *in captions* may require special consideration: Unlike text documents, captions are dynamic, with shorter text segments, which are usually shown in 1 or 2 lines, for 2 to 4 seconds [28]. Moreover, users are known to be sensitive to caption display parameters such as speed, font size, or decorations: Several researchers have measured the influence of such visual parameters of caption appearance on the readability of captions for DHH users [6, 28, 43].

Thus, our work investigates text-highlighting in captions for videos viewed by DHH users, and our contribution is threefold: First, we examine DHH users' preferences for various visual parameters of highlighting important words in captions: In two rounds of in-person interview and prototype usability studies, DHH participants indicated when they would prefer to see word-highlighting, which style of visual markup for this highlighting they preferred, and what is the threshold percentage of words that should be highlighted. Secondly, to investigate the efficacy of text highlighting in captions, we present results from a larger study, in which DHH users responded to questions after viewing videos of two forms: with and without highlighting of important words in the captions, when viewing educational lecture videos. Thirdly, the question-types and empirical results in our larger study could be beneficial for future researchers when evaluating automatic methods for identifying important words for users in this educational-video context, with our results as a potential baseline.

BACKGROUND AND PRIOR WORK

Although services exist to provide access to spoken content for DHH users, e.g. sign language interpreters or captioning services, users face challenges in attending to multiple streams of visual information. Text captioning of video content is increasingly common, e.g. enabling educational institutions to satisfy legal requirements for making content accessible for DHH students [3, 16]. Yet, traditional text captions are not a complete solution to providing full access to video content for DHH users, especially when there are multiple concurrent visuals and/or visual-references within the captions [26, 29, 33]. DHH users who rely on visual information sources must strategically switch between the captions and other visual information in video content. Since human cognition is a limited resource, with bounds on processing concurrent visual information sources, there can be a loss of information, even when high-quality accessibility services are provided. Consequently, research has found that DHH users typically get less out of even *accessible* mainstream classroom lectures than their hearing peers do [33].

Many interventions (discussed below) have been proposed for enhancing visual browsing of text through highlighting important words [26, 29], but this work has looked at text (e.g. textbooks or webpages). As online video has become an increasingly popular source of news, education, or entertainment among the general population [13, 34], research is needed on whether highlighting important words would also be beneficial for caption text displayed during video, and whether this would benefit DHH users.

There is reason to believe why research is needed that specifically focuses on DHH users in this context:

Peripheral visual attention. Research has found that DHH users, especially those who have used sign-language since an early age, have greater peripheral vision skills than hearing users [8, 37]. Eye-tracking studies to understand DHH users' strategies when viewing captions have revealed that although DHH users spend a smaller amount of time reading captions (compared to hearing users viewing captions), the amount of time DHH users spend watching captions depends on the rate of change of captions and amount of motion in the images [11]. These findings suggest differences in how DHH users visually process a video caption text, as compared to general readers of text.

Reading literacy skills. Further, in standardized testing in the U.S., English literary rates have been measured to be lower among adults who are deaf [21, 32] – a result that may be due to reduced language exposure or other educational experiences during childhood. Lower reading literacy skills among DHH users could affect the usage of captions among these users [9, 43]. Moreover, users with lower literacy can find it especially challenging to follow fast-moving captions [43]. This further suggests that research on the benefits of text highlighting is needed specifically among this group.

Errors and omissions in the caption text. Even when captions are produced by a human transcriptionist or captioning service, there can be errors in the text that is provided, or delays in when it is presented. Automatically produced captions, e.g. produced through automatic speech recognition (ASR), may have an even greater percentage of errors. Moreover, unlike human-generated errors, errors produced by automatic systems have been found to be even more cognitively demanding for users [4, 28]. In addition, caption texts (especially if produced through an automatic method) customarily do not convey speaker traits like accents, vocal emphasis on words, or emotional subtext of speech, which could be useful for DHH users. Thus, while there has been prior work (discussed below) on highlighting words in static text, research is needed on captions.

Prior Work on Enhancing Caption Accessibility

Multiple researchers have investigated methods of increasing the accuracy of (semi-)automated methods of

generating captions for DHH users [4, 28]. In addition, researchers have focused on the usability of alternative methods of presenting or displaying captions:

Some researchers addressed the problem of visual dispersion in classroom settings for DHH students through a system that reduces the burden on students to attend to changes across multiple channels of visual information [10]. The system integrated multiple video streams into a single display, with notifications when changes occur on various streams, e.g. change of slides. Integrating multiple video streams onto one screen has also been studied [26]. While reducing the distance between information streams and notifying users of onscreen changes may benefit students, they must still integrate multiple visual sources of information, which may tax their working memory and impact their learning [2].

Other work has focused on problems that arise from fast-moving captions. Researchers have investigated whether context-switching (moving gaze between captions and other regions of the video) may result in DHH users losing track of the captions or the visual content. Researchers in [29] proposed a caption-display interface for viewers to pause the captions, to prevent them from falling behind in classroom settings. Pausing the caption allowed students to follow the content at their own pace and better associate visual content with references to it in the text. However, in a real-time classroom setting, pausing a caption could lead students to fall behind what the instructor speaks during that interval.

Notably, prior work on alternative user-interface designs for captioning of educational video-content for DHH students has not examined the use of important-word highlighting. This gap in the literature is not surprising, since technology for automatically predicting important words in a caption text is only a recent focus of computing research [23, 24].

Importance-based Highlighting in Text

Text highlighting provides a natural way of conveying important information in text, and research has found that readers are able to make use of this emphasis information, without special training as to its meaning [15]. In the education context, highlighting is a common strategy used by textbook authors, teachers, and students to indicate important concepts in a text, and this has been shown to enable faster browsing and recall of information by students [25, 45]. In general, researchers have also argued that strategic marking of words or phrases can enhance the reading experience [17, 25]. Readers use highlighting as a way of functional coding, which helps with retention and faster browsing [17]. Similarly, readers perform better on comprehension tests when reading text with highlights [25]. Other work has focused on the effect of text highlighting on word-retention and learning [22, 31, 36]. When comparing keyword-highlighting and non-highlighting conditions, researchers found that students performed better on a cloze task when a text-passage was highlighted [30].

With the increasing availability of digital text content, several highlighting interventions have been applied to such texts, including rendering text in a different color, with color backgrounds, or changing font decoration. Such special rendering enables readers to attend to the most important segments of the text or focus on relevant information quickly [12, 36, 25]. In addition, computing accessibility research has investigated text transformations to promote comprehension among struggling readers [39]. For instance, highlighting important words has been found to improve reading rate and comprehension among people with dyslexia [38]. In an eye-tracking study, researchers found that keyword highlighting improved comprehensibility and readability of onscreen texts for people with dyslexia [39].

Strategies for Text Highlighting

The effects of different highlighting strategies have also been explored. Table 1 presents some of the popular highlighting strategies that have been considered in prior research. For instance, in empirical studies, readers often interpret **bold** or UPPERCASE text as being spoken loudly, and they interpret *italicized* text as having softer emphasis [1]. In other work, researchers compared how well various highlighting strategies captured people's attention [42]: Among the alternatives in Table 1, yellow background and bold were the most attention-getting, while italics had the least effect.

Highlighting Strategy	Prior Research
Font Color (`color_r`)	Ponce and Mayer [36];
Bold Face (`bold`)	Rello *et al.* [39];
Background (`color_bg`)	Chi *et al.* [12];
UPPERCASE (`uc`)	Berger *et al.* [5]
Underlining (`ul`)	Vertanen *et al.* [44]
Italicized (`it`)	Berger *et al.* [5]
Font Size (`size`)	Wang *et al.* [46]

Table 1. Highlighting Strategies from Prior Work

Research has found that the ideal text-highlighting strategy largely depends upon the application and user group [1, 20, 42]. For instance, although *italics* has been a preferred strategy for text highlighting for children, researchers found that children with dyslexia had difficulty discriminating between italics and regular text [20]. This suggests that specific research is required with DHH users in the context of text-highlighting in video captions, to determine the best choice of highlighting style for these users and application.

Visual Markup of Text in Captions

Highlighting text in captions can be challenging, especially when considering the dynamic nature of the captions compared to static text. Users are engaged in an attention-demanding task of viewing a video with multiple sources of visual information in parallel to the caption-text stream, and

if the choice of highlighting style or the frequency with which words are highlighted is suboptimal, then such visual decoration of the text could be distracting. This speculation is supported by prior research that has investigated the effect of different visual markup of caption texts in another context: to convey the *confidence scores* of a captioning service (e.g., ASR system) as to the accuracy of its caption output [6]. In a study with 107 DHH users, Berke *et al.* discovered that although participants were receptive to the idea of having visual indicators of the confidence of an automatic caption system, they were concerned about distraction from changes in text appearance. DHH users viewing video are sensitive to text-appearance changes in captions, and there is risk that highlighting text could actually be *detrimental*. Thus, empirical research is needed to determine the best choice of highlighting styles. Then, based on this set of preferred design parameters, we can then determine whether there are indeed benefits for DHH users from text highlighting.

RESEARCH QUESTIONS

To understand the preferences of DHH users who are viewing videos, in regard to highlighting important words in the caption text, we investigate the following questions:

RQ1. Are DHH users receptive to the premise of importance-based word highlighting in captions, and what are their highlighting preferences?
- a) For what types of videos or tasks would DHH users find caption text highlighting most useful?
- b) What visual markup strategy would DHH users favor for highlighting important words in captions?
- c) What percentage of words would DHH users prefer for highlighting in captions?

RQ2. Given DHH users' preferences from RQ1, when viewing captions of online lecture video, do DHH users subjectively prefer highlighting in captions?

We present the evaluation of our RQs in **two phases**: First, we conducted some studies in which we gathered subjective preferences from a small number of DHH users about various display options for highlighting in captions. These smaller preliminary studies were *not sufficiently powered* to enable us to observe statistically significant differences in user preferences. Instead, the goal of these *formative* studies was to provide some preliminary answers to RQ1, so that we were not making arbitrary choices about our design. Later, we conducted a larger user study to compare the experience of DHH users when viewing online educational lecture videos under two conditions: with and without text-highlighting. This final *summative* study utilized the best configuration settings for caption-highlighting found in our initial formative studies, which had identified the most preferred application use-case (educational online-lecture videos), choice of highlighting style (underlining), and percentage of words to highlight (at most 15%). To answer RQ2 in our final study, we compared the two video conditions: with and without text highlighting.

FORMATIVE STUDIES: METHOD AND RESULTS

The goal of our formative studies was to understand DHH users' interest in important-word highlighting in video captions, and their preferences among various visual markup strategies. Two studies, with 6 DHH participants each, helped us select display options for text-highlighting for our subsequent larger study. Rather than selecting design settings arbitrarily, we used this multi-study design to identify parameters for that final study.

Highlighting Configurations for Formative Studies

During this in-person interview and prototype-evaluation study, which was conducted in two rounds, we presented users with videos with different highlighting configurations. We were interested in two main design factors: the highlighting **markup style** and the **percentage of words** to highlight. One option for investigating these two factors would be conduct a single study with a large number of users to investigate various possible combinations of both factors, within a single study. Because these initial studies were planned as preliminary formative studies, in support of our larger final evaluation study, we chose to instead devote more personnel and time resources toward conducting the final study with as many participants as possible. Thus, we decided to investigate these two design factors in a cascaded manner, through a two-round formative study design, with each factor investigated independently in each round:

In round 1, to compare visual markup strategies, we conducted a within-subject study with **7 markup conditions** previously shown in Table 1. In all of the video stimuli shown in this round-1 study, the percentage of words highlighted was kept constant at around 20%.

In round 2, the stimuli videos included variations in the **percentage of words** highlighted in each caption. We investigated **4 conditions**: low percentage (5%), medium (15%), high (25%) and very high (35%). At the end of the round-1 study, we had determined that underlining was the preferred method of visual highlighting of important words. Thus, all of the stimuli video in this round-2 study used underlining as the method of visual highlighting.

As discussed in our Conclusion and Future Work, this choice to cascade the two small studies, each investigating a single factor, did not enable us to investigate interaction effects among variables. However, this tradeoff allowed us to devote more resources toward the larger final study.

Stimuli Preparation for Formative Studies

As discussed below, in some open-ended interview questions conducted during this formative study, we asked our DHH participants about the types of videos for which they may be interested in text-highlighting. In responses to those questions, users expressed interest in highlighting for online educational lecture videos, but we had not

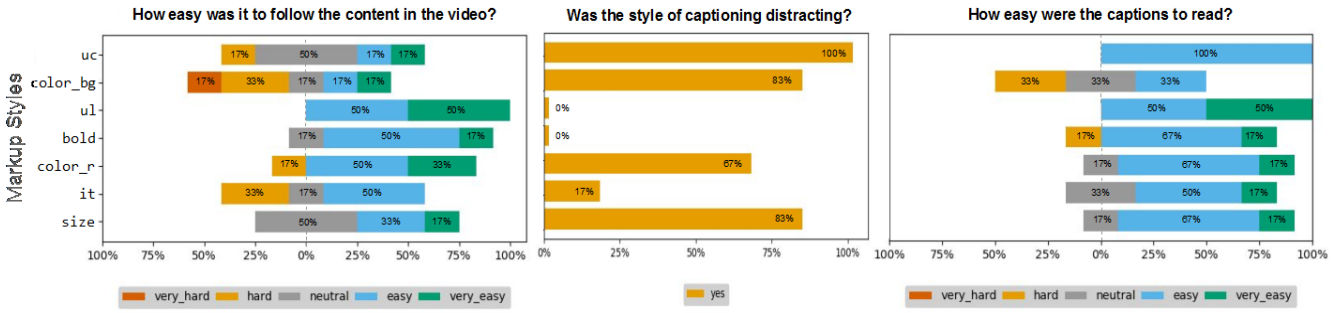

Figure 2. Round-1 Formative Study: Comparison of different visual markup-styles (list in Table 1) for highlighting in captions.

anticipated that finding when we had launched this round-1 formative study. Thus, the stimuli video used to display various text markup styles and highlighted-word percentage in the round-1 study was from a non-education genre: Specifically, as stimuli for this formative study, we used videos of a **fake business meeting** which had previously been used in research with DHH participants [23]. The meeting video was first chopped in 12 smaller videos with an average duration of 30 seconds. Each participant was shown the 12 videos in order, each with a different display configuration-setup arranged in pseudo-randomized order (Latin-Square) for each condition.

In order determine which words should be highlighted in the caption text for these videos, we utilized an **automatic word-importance prediction** system for scoring each word in the spoken dialogue according to its importance [24]. The word-importance scoring system computed a score for each word based on its influence on the meaning of the utterance. Compared to other word-importance prediction systems [23], this system had been specifically trained to operate over spoken dialogue, which differs from written language. For example, in English, spoken dialogue contains more filler words (e.g., *uh*, *um* etc.), ungrammatical sentences, and higher frequency of non-dictionary words. This system can provide a numerical score for each word in a text as to its importance. In order to produce each video stimulus with a particular percentage of words highlighted as important, we ranked words according to this score, and we highlighted a portion of the top-ranked words, to control the percentage of words with visual highlighting in the final caption text.

Recruitment and Participants for Formative Studies

For both rounds of formative studies, participants were recruited by e-mail and flyers at the Rochester Institute of Technology. Participants were eligible if they answered "yes" to both: *Are you Deaf or Hard-of-Hearing? Do you use captions when viewing television?* Participants met a DHH researcher fluent in both English and ASL in a private office to ensure a distraction-free environment. Participants were paid $40 for the 60-minute study.

Questionnaires for Smaller Studies

The questions asked and overall sequence of activities were identical for the round-1 and round-2 studies; the only difference was the video stimuli (either focusing on markup-style or percentage of highlighted words). At the beginning of each session, participants were informed that they would see captioned videos with some words shown differently, as a word-importance highlighting strategy.

Before viewing video stimuli, participants answered open-ended items, on a **pre-study questionnaire**, regarding the usefulness of word-importance highlighting for captions.

Next, participants viewed stimuli videos; after each video they answered three questions, which had been used by prior researchers to successfully gather subjective responses from DHH participants about caption quality [27]:

Q1: How easy were the captions to read? (Five-point scale from *very hard* to *very easy)*

Q2: How easy was it to follow the content in the video? (Five-point scale from *very hard* to *very easy*)

Q3: Did you find the caption distracting? (Yes/No)

Towards the end of the study, the participants were asked some **post-study questions** to gauge their interest in various use-case scenarios in which word-highlighting of captions may be preferred. The rationale for asking these questions at the end of the study was so that the users would have had some initial experience at viewing captions with various types of highlighting. For these questions, participants were encouraged to propose any situations and/or genres of videos where captioning highlighting might be beneficial.

Round-1 Results: Comparing Markup-Styles

A total of 6 DHH individuals participated in the round-1 study, with 3 males and 3 females, and self-identified hearing-status of 4 Deaf and 2 Hard of Hearing. Participants were shown captions with different visual markup (shown in Table 1). Figure 2 summarizes the responses of the participants for the different caption markup strategies. In particular, the graphs on the left and right of Figure 2 present participants' responses to Likert questions; diverging stacked bar graphs like these are recommended for presentation of Likert response data [40]. The segments of each bar indicate the percentage of responses for each Likert option, with the conditions of the study along the Y-axis. The neutral response is centered horizontally, with

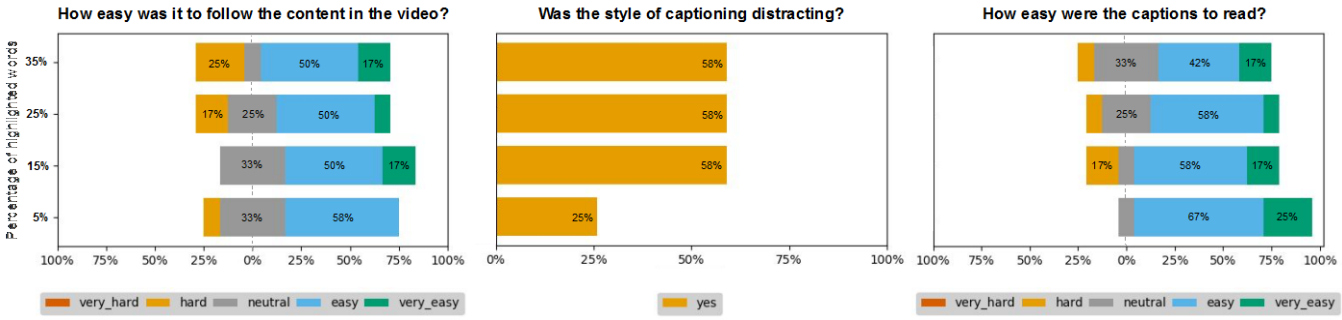

Figure 3. Round-2 Study: Comparison of the percentage of words highlighted in captions.

negative responses to the left and positive responses to the right. Participants preferred the underlining (`ul`) strategy for highlighting words in captions, and the bold strategy was a close second. Although the italics (`it`) markup was recognized as one of the least distracting strategies, it was harder to follow compared to other more distracting strategies like font color (`color_red`). Participants further reported that italicizing was harder to read. Notably, strategies like uppercasing (`uc`) and font size (`size`) changes were indicated as one of most distracting markups. The small sample size of this formative study did not support statistical significance testing.

Round-2 Results: Comparing Highlight Percentage

Based on the results of the round-1 study above, the video stimuli shown in the round-2 study used the underlining method of highlighting. In this round-2 study, a total of 6 newly recruited DHH individuals participated, with 3 males and 3 females, and self-identified hearing-status of 4 Deaf and 2 Hard of Hearing. Participants in this round viewed videos using underlining as a highlighting strategy, with the video stimuli differing as to what percentage of words were highlighted. Participants preferred videos with 5% to 15% of the words highlighted. As shown in Figure 3, participants indicated that captions were most readable when 5% of the words were highlighted, and when 15% of the words were highlighted, participants found it easier to follow along with the captions. Participants found the 5% highlight condition the least distracting. The small sample size of this formative study did not support statistical significance testing.

Round-1 and Round-2 Results: Interest in Highlighting

Across both rounds, participants were asked identical questions about their interest in highlighting. Since we did not observe a difference in feedback comments across the two rounds, for brevity, responses from all 12 participants across both rounds are presented together below.

On the pre-survey questionnaires, participants shared their initial thoughts about importance-based highlighting for video captions. Participants were fairly open to the premise: 8 participants out of 12 welcoming the idea. When asked to elaborate, one participant responded as follows:

> *I think important words being highlighted in captions should be worthwhile because it helps to get my attention in any matter. In my experience, sometimes I am too lazy to read all the captions, but I will be more attentive if there is something important to know. (P9)*

Two participants expressed concerns about using such a feature, especially given its novelty and their lack of familiarity with this new form of text appearance. They were also concerned that visual distraction due to highlighting would decrease the readability of the captions:

> *It might be hard to read; I haven't seen that before. It could be useful, but I would have to see to make judgements. (P6)*

A few participants indicated doubts about the usefulness of this feature. Although they saw some potential, they were not sure if it was a *silver-bullet* solution:

> *It depends, if it's in a classroom setting then, yeah, it sounds helpful but if it's based on the persons voice, etc. Then, no! If it's actually an important vocabulary, then yeah it would be nice. (P7)*

Furthermore, in responses to our post-study questionnaire, many participants indicated that highlighting would be beneficial for online lecture videos, with some saying:

> *Highlighting helps me to keep track of the online materials and video content. (P9)*

> *When the teacher talks too long, the deaf people have a hard time catching up. That's why the students need to know which words are important. (P7)*

Other contexts in which participants indicated that text-highlighting may be useful were: meetings with hearing peers (mentioned by 5 of 12), classroom lectures (by 4 of 12) and news/political video announcements (by 3 of 12).

Discussion of Results from Round-1 and Round-2

The results from these round-1 and round-2 formative studies began to address research question RQ1. However, these small formative studies were conducted as a preliminary exploration of this design space, with a goal of informing the design of our final larger study below. Given the small number of participants in these formative studies, they were too underpowered to enable statistical significance testing. While not yet providing a conclusive

answer to the various design questions raised by RQ1, these formative studies did enable us to avoid making arbitrary design choices as to the appearance of word highlighting for our final study.

LARGER STUDY: METHODS

The goal of our final study was to understand whether DHH users **subjectively prefer** word importance highlighting in captions (RQ2). We utilized the results from our two rounds of formative studies, as summarized in Table 2.

Parameters	Value
Markup Strategy	underline (`ul`)
Highlight Percent	5 – 15%
Video Genre	Online-lecture videos

Table 2. Results of preliminary studies used in final study.

Preparation of the Stimuli Video

Since online-lecture video was the most popular scenario suggested by participants for word-importance highlighting, for our final study, we needed to produce new video stimuli to match this context. When generating stimuli, we used "underline" style highlighting, with 5-15% of words highlighted. As discussed earlier, many online education platforms use a screen arrangement with multiple concurrent visual streams, including an image of the instructor, of slides, of captions, etc. As a basis for our stimuli, we made use of a public dataset of educational video stimuli that had been produced in [18], which possessed several desirable characteristics for use in experimental studies:

- **Content Obscurity:** The content in the videos had been engineered such that it was obscure, to remove content-bias from a participant having prior knowledge of a topic. The content was partially fictionalized, using fake names and other historical details wherever necessary.
- **Content Homogeneity:** As discussed in [18], a similar density of information types (names, dates, etc.) was distributed throughout all the slides of the lecture and the script of the instructor spoke, to enable the videos to be partitioned into segments for experimental studies.
- **Visual Homogeneity:** As discussed in [18], the videos have an onscreen layout with multiple regions typical of lecture videos on many education platforms. As shown in Figure 1, each contains: the instructor, the slides, a topic list for the lecture, and captions. The videos had been designed such that they promote visual homogeneity: with limited and consistent color use across the visual streams over time, and with limited upper body gestures by the instructor. This homogeneity enables the videos to be partitioned into segments with similar appearance.

The visual and content homogeneity of these videos allowed us to create a controlled setting for understanding the effects of text highlighting in captions. In total, this dataset contains **four lessons** [18], with each lesson being five minutes in duration, and containing exactly 10 presentation slides, each having a time duration of 30 seconds. Each of the lessons was originally of duration 5-minutes. To generate our stimuli, we split each into sub-lessons of 1.5-minute duration each (discarding the final 0.5 minutes of each lesson). This yielded a set of 12 short videos, each of which was rendered in two conditions: with and without highlighting.

While we had used an automatic system [24] for identifying important words during our initial formative studies, that system had been designed to operate on conversational-style speech, rather than formal academic lectures. While it would be possible for researchers to re-train a word-importance system for this new genre of speech, building an automatic system for this task was beyond the scope of addressing our research questions in this study. Thus, instead of using an automatic system, we manually identified important words in the text that should be highlighted. To reduce individual bias, the words were identified based on a consensus labelling by a group of 3 researchers. We used the threshold criterion in Table 2 for selecting the number of words to be highlighted such that 15% of the words were highlighted in each stimulus video. Given the short video duration and our methodology of asking multiple researchers to agree upon the importance-labeling of words to be highlighted, practically it was easier to achieve a consensus at the 15% level.

Study Setup and Questionnaires

In pilot testing, participants indicated that watching four 5-minute videos (in their original duration) was too tiring; so, each lesson was split into three shorter segments. Each participant saw 9 videos (segments of three lessons) during the study. However, each sub-lesson video within a lesson was always presented in sequence, to preserve the original temporal flow of each lesson. The highlighting and non-highlighting conditions were presented in an alternate order in the videos, with the assignment of conditions to each stimulus video counterbalanced across participants.

Similar to the preliminary study, participants were asked **pre-study** and **post-study** questionnaires. During the study, participants viewed videos with and without highlighting and answered questions about the readability of the captions. In addition to the questions asked in the preliminary study, we also included questions about the user-perceived workload of the comprehension task: We asked about three dimensions of the NASA Task Load Index (NASA-TLX) [19]:

- **Mental Load**: This scale measures how much mental and perceptual activity was required for the task, e.g. thinking, decoding, remembering, looking, searching etc.
- **Temporal Demand**: This scale measures how much time pressure the user felt due to the pace of the task.
- **Effort**: This scale measures how hard the user had to work to accomplish their level of performance.

Questions	Scale
*Q1. It was **easy to follow** the content of the video and captions.* *Q2. It was **easy to read** the caption.* *Q3. I was **able to identify** the important words and concepts.* *Q4. I **understood** all of the content of the video and captions.*	5-point Likert Scale from *Strongly Agree* to *Strongly Disagree*.
*Q5. How **mentally demanding** was the task (reading and understanding the captions in the video)?* *Q6. How **hurried or rushed** was the task (reading and understanding the captions in the video)?* *Q7. How **hard** did you have to work to read and understand the captions in the video?*	21-point NASA-TLX scale from *Very Low* to *Very High*.

Table 2. List of questions used in the final study.

The wording of the TLX items was modified slightly to include the phrase "reading and understanding the captions in the video" after any mention of "the task." The final set of questions as shown in this study appear in Table 3.

To summarize, each participant was shown videos of 1.5-minute duration each. Half of the videos contained captions with words highlighted in them, and half of the videos, without highlighting. For each video, our participants answered 7 questions in total (Table 3). Q1-Q4 were questions that were inspired from our earlier preliminary study, and Q5-Q7 were adopted from the NASA-TLX.

Recruitment and Participants

We recruited participants for this study using similar methods as in our formative studies. Participants were paid $40 for the 60-minute study. A total of 30 DHH individuals (age distribution with mean = 25 and standard deviation = 6.02) participated, with 15 males and 15 females, and self-identified hearing-status of 17 Deaf and 13 Hard of Hearing. To evaluate the English literacy skill of participants, we used Wide Range Achievement Test 4th edition (WRAT4)[1], which had been previously validated with DHH users [35]. Our participants reported an average WRAT score of 82.6, which is one standard deviation below the standard score (100) among adults in the U.S.

RESULTS

For each question in Table 3, we collected 270 responses from the 30 DHH participants on the stimuli videos. This section presents a comparison of participants' subjective preference of each of the two conditions: videos containing captions with highlighting (*highlight*) and videos containing captions without any highlighting (*no_highlight*).

[1]https://www.pearsonclinical.ca/en/products/product-master/item-11.html

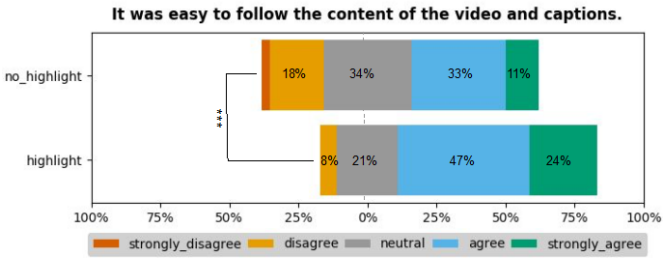

Figure 4. Percentage distribution of participants' responses on the ease of following the content of the video and the caption.

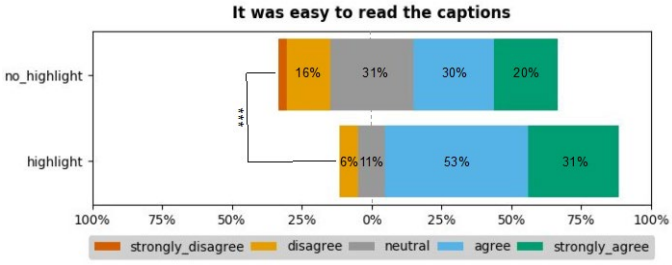

Figure 5. Percentage distribution of participants' responses on the readability of the caption.

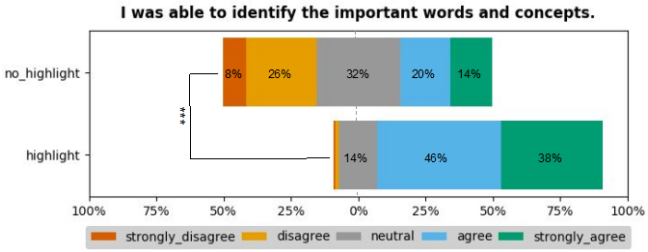

Figure 6. Percentage distribution of participants' responses on being able to identify the important words and concepts.

Figure 4 compares responses when asked about the ease of following the content of the video under our two highlighting conditions, represented on the Y-axis of the chart. After conversion of scalar Likert responses to integer (e.g., "*Strongly Disagree*"=1, "*Disagree*"=2, etc.), a Wilcoxon rank-sum test indicated a significant difference for this question [W=576, p<0.0001] with the average rating-score at [μ_1=3.32 and μ_2=3.9] for the *no_highlight* and *highlight* conditions respectively. In Figures 4 through 9, brackets indicate significant differences as follows: *** p<0.0001, ** p<0.001, * p<0.01, or N.S. not significant.

Similarly, a Wilcoxon rank-sum test indicated significant differences on the general ease of reading the captions in videos [W=631, p<0.0001] with the two means at [μ_1=3.53 and μ_2=4.09] for *no_highlight* and *highlight* conditions respectively. Figure 5 summarizes the responses for the participants for this measure.

Figure 6 shows the distribution of Likert-responses when participants indicated if they could identify the important words and concepts in the video. A Wilcoxon rank-sum test indicated significant differences for this question [W=246,

p<0.0001] with means at [μ_1=3.05 and μ_2=4.18] for *no_highlight* and *highlight* conditions respectively.

When asked to indicate the overall understandability of the captions, participants subjectively preferred the *highlight* condition over the *no_highlight* condition, with means at [μ_1=3.9 and μ_2=3.5] respectively. A Wilcoxon rank-sum test indicated significant difference for this measure [W=759, p<0.001]. Figure 7 summarizes the percentage distribution of responses for this question.

Participants also reported differences in the mental demand required to read and the understand the captions in the video under the two highlighting conditions, as shown in Figure 8. The box plot reveals that the median score for the *highlight* condition was lower than that of the *no_highlight* condition. A Wilcoxon rank-sum test indicated a significant difference for this measure [W=2185, p<0.01].

Figure 9 shows the box-and-wisker diagram summarizing the responses of the participants when asked about the temporal demand of the task. While the box plot may appear to show that the median score for the *highlight* condition was slightly lower than *no_highlight* condition, a Wilcoxon rank-sum test indicated no significant difference between the *highlight* and *no_highlight* conditions for this measure.

Lastly, there was a significant difference in responses to the question about the effort required to read and understand the captions in the video, under the two conditions, revealed through a Wilcoxon rank-sum test [W=1743, p<0.001]. Figure 10 shows the results of this analysis.

DISCUSSION

Our results indicate that participants are open to the idea the highlighting in captions, and we measured statistically significant differences when comparing participants' responses after they viewed captions with each highlighting condition, for online lecture videos.

In particular, participants indicated that lecture videos containing highlighted words in captions were easier to read and follow, as compared to videos without any highlighting. This was an important finding because prior research on the incorporation of visual markup in captions (for conveying confidence of words in captions generated automatically) had revealed that participants had concerns about being distracted by text decoration. On the contrary, our results indicate that our markup strategy for highlighting in captions was preferred by DHH participants on this task, who reported a significantly higher (p<0.0001) readability score on videos with captions containing highlighting. Although not distracting, the participants found the highlighting in captions to be noticeable enough to be able to identify important words and concepts in the video. Participants also reported an overall increase in the understandability of the content of the video and captions under highlighting.

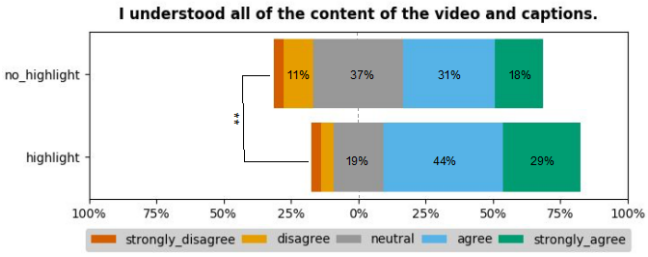

Figure 7. Percentage distribution of participants' responses on the understandability of the content of the video and the captions.

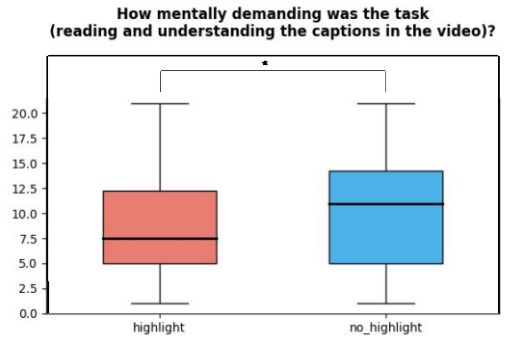

Figure 8. Summary of participant's rating on the mental demand when reading and understanding the captions in the video.

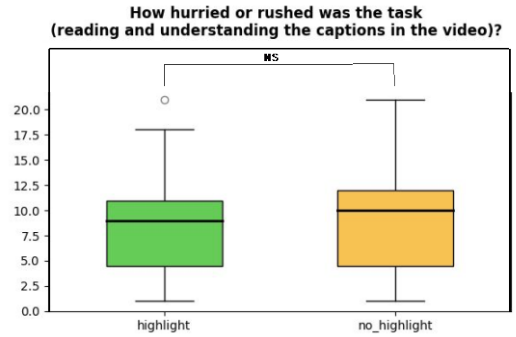

Figure 9. Summary of participants' rating on the temporal demand of reading and understanding the captions in the video.

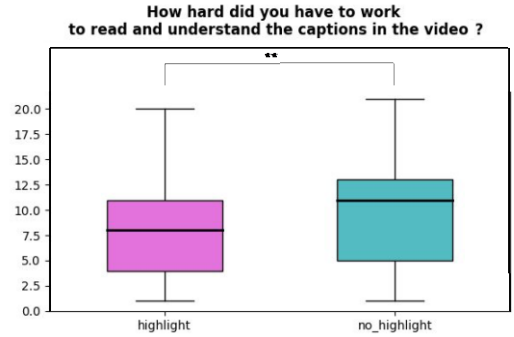

Figure 10. Summary of participants' rating on the difficultly of reading and understanding the captions in the video.

Similarly, we observed a significant difference ($p<0.01$) in participants' *mental and perceptual load* required to read and understand the captions, with and without highlighting. Overall, they reported less mental load when viewing videos with captions that contained highlighting. In addition, participants indicated that it required *less effort* to read and understand captions under the highlighting condition.

However, we did not observe a difference in participants' rating of the *temporal demand,* under our two highlighting conditions. This result suggests that although the participants found highlighted captions easier to read, this did not influence their perception of time pressure from the pace of the video. It is important to note that in both highlighting conditions, participants indicated a relatively low degree of temporal demand (*highlight*: 8.46, *no_highlight*: 8.90), suggesting that participants were rather comfortable with the pace of the task in either case.

Although our investigation of the various design issues from RQ1 was only formative in nature, those preliminary studies had suggested that we should focus on the context of educational online lecture videos, with underlining markup, with 5%-15% of words highlighted. While we did not provide any statistically significant empirical evidence that these are the optimal settings of these design variables through the findings presented in this paper, our final study was able to confirm that for this specific combination of these variables, users did prefer captions with highlighting.

CONCLUSION AND FUTURE WORK

Although captioning of videos is essential for making content more accessible for DHH users, prior research has found that there is room for improvement, especially for videos with multiple channels of visual information. The additional demands of attention management and visual processing for such videos may detract from an individual's ability to comprehend the content. This effect would have particular significance in educational contexts, including for videos of lectures in which students must keep track of the instructor, slides, and other visual information sources.

In this work, we have investigated whether highlighting important words in captions leads to improvements in DHH viewers' subjective rating of their experience. Although the benefits of highlighting words in a non-dynamic text have been studied in prior work, highlighting of important words in video captions had been relatively under-studied. We therefore conducted formative and summative studies to investigate whether there are benefits to highlighting important words in captions for people who are DHH.

Participants in our formative studies expressed an interest in highlighting of important words in the context of online education lecture videos, and they indicated a preference for the use of underlining to mark important words (**bold** strategy was a close second), with 5%-15% of words highlighted as important. In an experimental study with 30 DHH participants who viewed lecture videos with and without highlighting, participants reported higher readability and understandability scores and lower task-load scores when viewing captions with highlighting. These findings suggest the efficacy of highlighting of important words in captions for this online lecture video setting. As a final contribution of this work, we have demonstrated a set of question items and an experimental methodology that can be employed by future researchers who wish to investigate the efficacy of word-highlighting in additional contexts, or with alternative appearance or implementation settings.

There were several limitations of our study that we would like to address in future work: Our smaller formative studies were too underpowered (i.e. with too few participants) to enable us to investigate the design options in research question RQ1 conclusively. Although those studies served their formative purpose for this paper, we believe that future researchers and designers would benefit from a more conclusive investigation of those design options in a larger study. Relatedly, there was a limitation in the cascaded nature of our two rounds of formative studies, where we collected preference scores on 1) the markup strategy for highlighting and then 2) the threshold percentage of words to be highlighted, independently. While it would have required a larger number of participants, a two-factor study would have been more ideal to enable us to investigate interaction between factors.

In future work, we would like to investigate the generalizability of our results across different tasks and application contexts; we could investigate if highlighting would benefit other groups of users or would be useful for other video genres or other communication scenarios, such as live captioning in multi-party meetings. In prior methodological research [7], we found that in experiments with DHH participants evaluating captions, comprehension-question probes were less discriminative than other question-types. Based on that finding, we had not included comprehension-question probes in this current study, but we could include comprehension probes in future work.

In conclusion, our findings demonstrate the potential benefit of caption highlighting in online educational videos for DHH users, which motivates additional research in this area, and we have provided a methodological foundation for the evaluation of such systems with DHH users.

ACKNOWLEDGMENTS

We thank Allen V. R. Harper who designed the educational video stimuli during his dissertation research. This material is based upon work supported by the National Science Foundation under Award Numbers 1462280 and 1822747; by the Department of Health and Human Services under Award No. 90DPCP0002-01-00; by a Microsoft AI for Accessibility (AI4A) Award; by a Google Faculty Research Award; and by funding from the National Technical Institute of the Deaf (NTID).

REFERENCES

[1] Cynthia Acrey, Christopher Johnstone, and Carolyn Milligan. 2005. Using universal design to unlock the potential for academic achievement of at-risk learners. *Teaching Exceptional Children* 38, 2 (2005), 22–31.

[2] Paul Ayres and John Sweller. 2005. The split-attention principle in multimedia learning. *The Cambridge handbook of multimedia learning* 2 (2005), 135–146.

[3] Keith Bain, Sara Basson, Alexander Faisman, and Dimitri Kanevsky. 2005. Accessibility, transcription, and access everywhere. *IBM systems journal* 44, 3 (2005), 589–603.

[4] Keith Bain, Sara H. Basson, and Mike Wald. 2002. Speech Recognition in University Classrooms: Liberated Learning Project. In *Proceedings of the Fifth International ACM Conference on Assistive Technologies* (ASSETS '02). ACM, New York, NY, USA, 192–196. DOI: http://dx.doi.org/10.1145/638249.638284

[5] Stephanie Berger, Oliver Niebuhr, and Kerstin Fischer. 2018. Eliciting extra prominence in read-speech tasks: The effects of different text-highlighting methods on acoustic cues to perceived prominence. In *Proc. 9th International Conference on Speech Prosody* 2018. 75–79.

[6] Larwan Berke, Christopher Caulfield, and Matt Huenerfauth. 2017. Deaf and Hard-of-Hearing Perspectives on Imperfect Automatic Speech Recognition for Captioning One-on-One Meetings. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility* (ASSETS '17). ACM, New York, NY, USA, 155–164. DOI: http://dx.doi.org/10.1145/3132525.3132541

[7] Larwan Berke, Sushant Kafle, and Matt Huenerfauth. 2018. Methods for Evaluation of Imperfect Captioning Tools by Deaf or Hard-of-Hearing Users at Different Reading Literacy Levels. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18). ACM, New York, NY, USA, Paper 91, 12 pages. DOI: https://doi.org/10.1145/3173574.3173665

[8] Rain Bosworth G. and Karen Dobkins R. 2002. The Effects of Spatial Attention on Motion Processing, in Deaf Signers, Hearing Signers, and Hearing Nonsigners. *Brain and Cognition* 49, 1 (2002), 152 – 169. DOI: http://dx.doi.org/10.1006/brcg.2001.1497

[9] Barbara B Braverman and Melody Hertzog. 1980. The effects of caption rate and language level on comprehension of a captioned video presentation. *American Annals of the Deaf* (1980), 943–948.

[10] Anna C. Cavender, Jeffrey P. Bigham, and Richard E. Ladner. 2009. 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). ACM, New York, NY, USA, 67–74. DOI: http://dx.doi.org/10.1145/1639642.1639656

[11] Claude Chapdelaine, Valérie Gouaillier, Mario Beaulieu, and Langis Gagnon. 2007. Improving video captioning for deaf and hearing-impaired people based on eye movement and attention overload. In *Human Vision and Electronic Imaging XII*, Vol. 6492. International Society for Optics and Photonics, 64921K. DOI: http://dx.doi.org/10.1117/12.703344

[12] Ed H Chi, Lichan Hong, Michelle Gumbrecht, and Stuart K Card. 2005. ScentHighlights: highlighting conceptually-related sentences during reading. In *Proceedings of the 10th international conference on Intelligent user interfaces*. ACM, 272–274.

[13] P. Diwanji, B. P. Simon, M. MÃd'rki, S. Korkut, and R. Dornberger. 2014. Success factors of online learning videos. In 2014 International Conference on Interactive Mobile Communication Technologies and Learning (IMCL2014). 125–132. DOI: http://dx.doi.org/10.1109/IMCTL.2014.7011119

[14] Ana-Belén Domínguez, María-Soledad Carrillo, María del Mar Pérez, and Jesús Alegría. 2014. Analysis of reading strategies in deaf adults as a function of their language and meta-phonological skills. *Research in developmental disabilities* 35, 7 (2014), 1439–1456.

[15] John Dunlosky, Katherine A Rawson, Elizabeth J Marsh, Mitchell J Nathan, and Daniel T Willingham. 2013. Improving students' learning with effective learning techniques: Promising directions from cognitive and educational psychology. *Psychological Science in the Public Interest* 14, 1 (2013), 4–58.

[16] Maria Federico and Marco Furini. 2012. Enhancing Learning Accessibility Through Fully Automatic Captioning. In *Proceedings of the International Cross-Disciplinary Conf. on Web Accessibility*. ACM, New York, NY, USA, Article 40, 4 pages. DOI: http://dx.doi.org/10.1145/2207016.2207053

[17] Shawn M Glynn. 1978. Capturing readers' attention by means of typographical cuing strategies. *Educational Technology* 18, 11 (1978), 7–12.

[18] Allen VR Harper. 2015. Eye Tracking and Performance Evaluation: Automatic Detection of User Outcomes. (2015). *CUNY Academic Works.*

[19] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in psychology*. Vol. 52. Elsevier, 139–183.

[20] Rozita Ismail and Azizah Jaafar. 2014. Important Features in Text Presentation for Children with

Dyslexia. *Journal of Theoretical & Applied Information Technology* 63, 3 (2014).

[21] Dorothy W Jackson, Peter V Paul, and Jonathan C Smith. 1997. Prior knowledge and reading comprehension ability of deaf adolescents. *Journal of Deaf Studies and Deaf Education* (1997), 172–184.

[22] Dale D Johnson and Bonnie von Hoff Johnson. 1986. Highlighting vocabulary in inferential comprehension instruction. *Journal of Reading* 29, 7 (1986), 622–625.

[23] Sushant Kafle and Matt Huenerfauth. 2017. Evaluating the usability of automatically generated captions for people who are deaf or hard of hearing. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, 165–174. DOI: http://dx.doi.org/10.1145/3132525.3132542

[24] Sushant Kafle and Matt Huenerfauth. 2018. A Corpus for Modeling Word Importance in Spoken Dialogue Transcripts. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation* (LREC-2018).

[25] Yuka Kawasaki, Hironori Sasaki, Haruhisa Yamaguchi, and Yumi Yamaguchi. 2008. Effectiveness of highlighting as a prompt in text reading on a computer monitor. In *Proceedings of the 8th WSEAS International Conference on Multimedia systems and signal processing*. World Scientific and Engineering Academy and Society (WSEAS), 311–315.

[26] Raja S. Kushalnagar, Anna C. Cavender, and Jehan-François Pâris. 2010. Multiple View Perspectives: Improving Inclusiveness and Video Compression in Mainstream Classroom Recordings. In *Proceedings of the 12th International ACM SIGACCESS Conference on Computers and Accessibility* (ASSETS '10). ACM, New York, NY, USA, 123–130. DOI: http://dx.doi.org/10.1145/1878803.1878827

[27] Raja S. Kushalnagar, Walter S. Lasecki, and Jeffrey P. Bigham. 2013. Captions Versus Transcripts for Online Video Content. In Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility (W4A '13). ACM, New York, NY, USA, Article 32, 4 pages. DOI: http://dx.doi.org/10.1145/2461121.2461142

[28] Raja S. Kushalnagar, Walter S. Lasecki, and Jeffrey P. Bigham. 2014. Accessibility Evaluation of Classroom Captions. *ACM Trans. Access. Comput.* 5, 3, Article 7 (Jan. 2014), 24 pages. DOI: http://dx.doi.org/10.1145/2543578

[29] Walter S. Lasecki, Raja Kushalnagar, and Jeffrey P. Bigham. 2014. Helping Students Keep Up with Real-time Captions by Pausing and Highlighting. In *Proceedings of the 11th Web for All Conference* (W4A '14). ACM, New York, NY, USA, Article 39, 8 pages. DOI: http://dx.doi.org/10.1145/2596695.2596701

[30] Detlev Leutner, Claudia Leopold, and Viola den Elzen-Rump. 2007. Self-regulated learning with a text-highlighting strategy. *Zeitschrift fur Psychologie* 215, 3 (2007), 174–182.

[31] Robert F Lorch. 1989. Text-signaling devices and their effects on reading and memory processes. *Educational psychology review* 1, 3 (1989), 209–234.

[32] John L Luckner and C Michele Handley. 2008. A summary of the reading comprehension research undertaken with students who are deaf or hard of hearing. *American Annals of the Deaf* 153, 1 (2008), 6–36.

[33] Marc Marschark, Jeff B Pelz, Carol Convertino, Patricia Sapere, Mary Ellen Arndt, and Rosemarie Seewagen. 2005. Classroom interpreting and visual information processing in mainstream education for deaf students: Live or Memorex®? *American Educational Research Journal* 42, 4 (2005), 727–761.

[34] Limor Peer and Thomas B. Ksiazek. 2011. YouTube and the Challenge to Journalism. *Journalism Studies* 12, 1 (2011), 45–63. DOI: http://dx.doi.org/10.1080/1461670X.2010.511951

[35] LeAdelle Phelps and Barbara Jane Branyan. 1990. Academic achievement and nonverbal intelligence in public school hearing-impaired children. *Psychology in the Schools* 27, 3 (1990), 210–217.

[36] Hector R. Ponce and Richard E. Mayer. 2014. An eye movement analysis of highlighting and graphic organizer study aids for learning from expository text. *Computers in Human Behavior* 41 (2014), 21 – 32. DOI: http://dx.doi.org/10.1016/j.chb.2014.09.010

[37] Jason Proksch and Daphne Bavelier. 2002. Changes in the spatial distribution of visual attention after early deafness. *Journal of cognitive neuroscience* 14, 5 (2002), 687–701.

[38] Luz Rello, Ricardo Baeza-Yates, Laura Dempere-Marco, and Horacio Saggion. 2013. Frequent words improve readability and short words improve understandability for people with dyslexia. In *IFIP Conference on Human-Computer Interaction*. Springer, 203–219.

[39] Luz Rello, Horacio Saggion, and Ricardo Baeza-Yates. 2014. Keyword highlighting improves comprehension for people with dyslexia. In *Proc. of the 3rd Workshop on Predicting and Improving Text Readability for Target Reader Populations* (PITR). 30–37.

[40] Robbins, Naomi B., and Richard M. Heiberger. 2011. Plotting Likert and other rating scales. In *Proc. of the 2011 Joint Statistical Meeting*, 1058-1066.

[41] Michael S. Stinson, Pamela Francis, Lisa B. Elliot, and Donna Easton. 2014. Real-time caption challenge: C-print. In Proceedings of the 16th International ACM SIGACCESS conference on Computers & Accessibility, ASSETS '14, Rochester, NY, USA, October 20-22, 2014. ACM, 317–318. DOI: http://dx.doi.org/10.1145/2661334.266133740

[42] Hendrik Strobelt, Daniela Oelke, Bum Chul Kwon, Tobias Schreck, and Hanspeter Pfister. 2016. Guidelines for effective usage of text highlighting techniques. *IEEE transactions on visualization and computer graphics* 22, 1 (2016), 489–498.

[43] Michael D. Tyler, Caroline Jones, Leonid Grebennikov, Greg Leigh, William Noble, and Denis Burnham. 2009. Effect of Caption Rate on the Comprehension of Educational Television Programmes by Deaf School Students. *Deafness & Education International* 11, 3 (2009), 152–162. DOI: http://dx.doi.org/10.1179/146431509790559606

[44] Keith Vertanen and Per Ola Kristensson. 2008. On the benefits of confidence visualization in speech recognition. In *Proceedings of the 2008 Conference on Human Factors in Computing Systems*, CHI 2008, 2008, Florence, Italy, April 5-10, 2008, Mary Czerwinski, Arnold M. Lund, and Desney S. Tan (Eds.). ACM, 1497–1500. DOI: http://dx.doi.org/10.1145/1357054.1357288

[45] Suzanne E Wade, Woodrow Trathen, and Gregory Schraw. 1990. An analysis of spontaneous study strategies. *Reading Research Quarterly* (1990), 147–166.

[46] Fangzhou Wang, Hidehisa Nagano, Kunio Kashino, and Takeo Igarashi. 2017. Visualizing Video Sounds with Sound Word Animation to Enrich User Experience. *IEEE Trans. Multimedia* 19, 2 (2017), 418–429. DOI: http://dx.doi.org/10.1109/TMM.2016.2613641