2020
THE JOURNAL ON TECHNOLOGY AND PERSONS WITH DISABILITIES

Gaze Guidance for Captioned Videos for DHH Users

Jessica Li

Vassar College - Poughkeepsie, NY

jyli@vassar.edu

Matt Luettgen

Indiana University Bloomington - Bloomington, IN

mrluettg@iu.edu

Sedeeq Al-khazraji

Rochester Institute of Technology - Rochester, NY, University of Mosul, Mosul, Iraq

sha6709@rit.edu

Matt Huenerfauth, Reynold Bailey, Cecilia O. Alm

Rochester Institute of Technology - Rochester, NY

matt.huenerfauth@rit.edu; rjbvcs@rit.edu; coagla@rit.edu

Abstract

We evaluated whether DHH individuals benefit from the addition of subtle visual effects to captioned educational videos to guide their gaze toward potentially informative content.

Keywords

Deaf and Hard of Hearing, Emerging Assistive Technologies, Research & Development

Introduction

Automatic Speech Recognition (ASR) technology can generate real-time captions during classroom lectures for Deaf or Hard-of-Hearing (DHH) individuals, but the resulting experience may not be fully accessible due to errors in captions (Berke et al. 2018) or the challenges faced when users must split their attention between the caption and other visual information, e.g. slides displayed (Kushalnagar et al. 2010). This study focuses on students viewing captioned videos of lectures containing an instructor and other visual content. We investigate whether the addition of *gaze guidance* (brief subtle blinking elements added to the video to draw someone's gaze) could be used to guide the visual attention of DHH individuals toward regions of the video where key information may be displayed. To avoid the time-consuming work of manually identifying specific times and locations in the video when we should guide the DHH users' gaze away from captions, we sought to automate this process by analyzing where hearing individuals (people who learned English as a second language) directed their gaze when they viewed these videos. The main contribution of this work is empirical: We have explored whether gaze guidance added to educational lecture videos lead to differences in DHH user's looking at the non-caption region of the video or in their comprehension of the content.

Related Work

DHH students in classroom settings face challenges in splitting their visual attention among multiple information sources, e.g. captions, instructor, and slides (Kushalnagar et al. 2010). Some work has investigated user interfaces for DHH students with multiple video windows (one for each information source), with visual indicators to bring the user's attention to specific windows (Kushalnagar et al. 2010). Other work has considered pausing captions when users look away from the text, so that they do not miss material, but such solutions could lead students to lagging behind during a live lecture (Lasecki et al. 2014). Prior eye-tracking research found that DHH users often fixate on the caption region of a video (Rathbun et al. 2017) or rapidly switch their gaze between captions and other video regions (Szarkowska et al. 2011).

As an alternative method of visually guiding the eye-gaze of DHH users, we investigate *subtle gaze guidance,* a technique in which a semi-transparent flickering animation is placed on an image or video, typically in the user's peripheral vision. This technique is effective at directing users' gaze to other areas on screen (Bailey et al. 2009) and has been shown to improve task performance, e.g. guiding gaze toward obstacles to reduce reaction times in a driving simulator (Pomarjanschi et al. 2012). However, it has not previously been employed among DHH users in a captioned video context, which is a novel contribution of our study.

A limitation of any technique for guiding the visual gaze of DHH viewers is that target locations, at specific times in the video, must be determined. Asking the creator of an educational video to manually determine such targets is time-consuming; so, researchers have considered simple automatic rubrics, e.g. bring the user's gaze toward the slides whenever a slide-change occurs (Kushalnagar et al. 2010). In this work, we instead consider whether we can learn from where hearing individuals look during a video, to automatically identify where to encourage the DHH viewer to look. Some hearing individuals many never look at a captions in a video; so, we investigate whether the eye-gaze patterns of students who learned English as a second language can be analyzed automatically to identify where a DHH viewer should direct their gaze at the most informative regions of a video. Prior work has examined the use of captioning in videos for students of English as a second language (Etemadi 2012; Hayati & Mohmedi 2011); work has also examined online educational videos with captioning for such students (Van der Zee et al. 2017).

Methodology

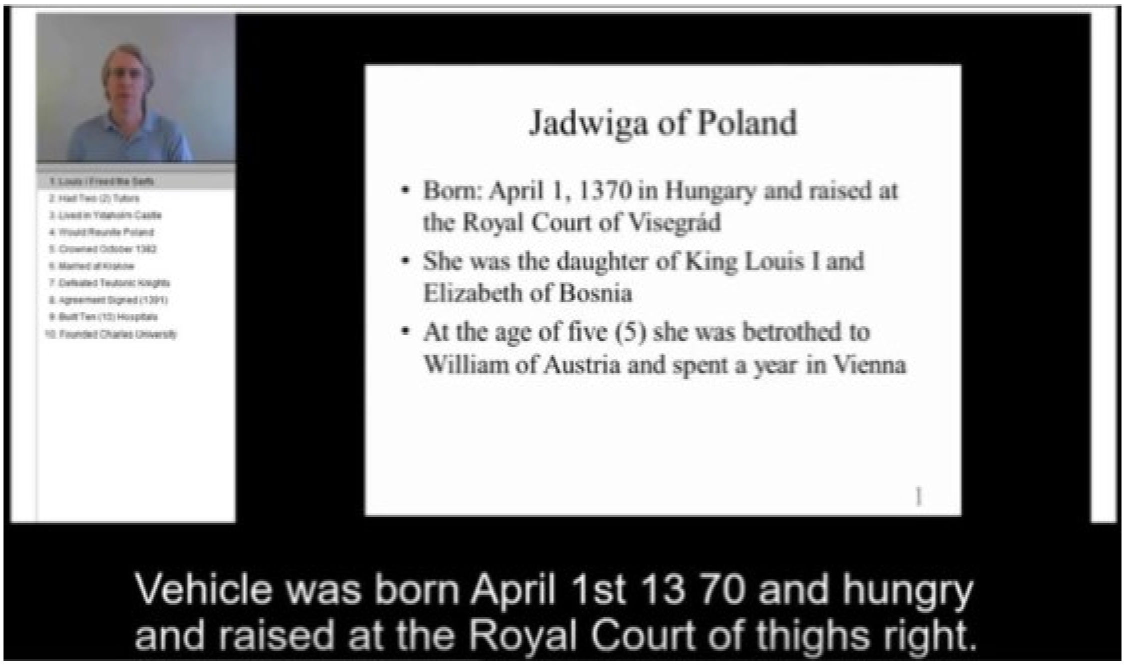
This study addresses three research questions:

RQ1: Do gaze patterns of hearing native English and nonnative English participants differ?

RQ2: Do gaze patterns of DHH individuals differ for videos with gaze guidance vs. videos without gaze guidance?

RQ3: Does gaze guidance improve comprehension for DHH individuals?

In this study, an SMI RED250 eye tracker (60 Hz), the SMI iViewRED software, and the iMotions Platform were used to collect eye data as participants viewed four educational lectures (2.5 minutes each) and answered comprehension questions; these videos and questions were produced by Harper (2015). We created new captions for these videos automatically using the Sonix automatic speech recognition (ASR) API, yielding a word error rate of 17%. We used imperfect captions to simulate the output of automatic captions produced by some online video platforms. Two studies were conducted: 13 native English users and 7 non-native English users (who learned English after age 5) participated in Experiment 1, and 11 individuals who identified as d/Deaf or Hard-of-Hearing (DHH) participated in Experiment 2.



1. Louis I Freed the Serfs
2. Had Two (2) Tutors
3. Lived in Ytaholm Castle
4. Would Reunite Poland
5. Crowned October 1382
6. Married at Krakow
7. Defeated Teutonic Knights
8. Agreement Signed (1391)
9. Built Ten (10) Hospitals
10. Founded Charles University

Jadwiga of Poland

- Born: April 1, 1370 in Hungary and raised at the Royal Court of Visegrád
- She was the daughter of King Louis I and Elizabeth of Bosnia
- At the age of five (5) she was betrothed to William of Austria and spent a year in Vienna

1

Vehicle was born April 1st 13 70 and hungry and raised at the Royal Court of thighs right.

Fig. 1. Example video lecture containing the lecturer, video content, and ASR-generated subtitles.

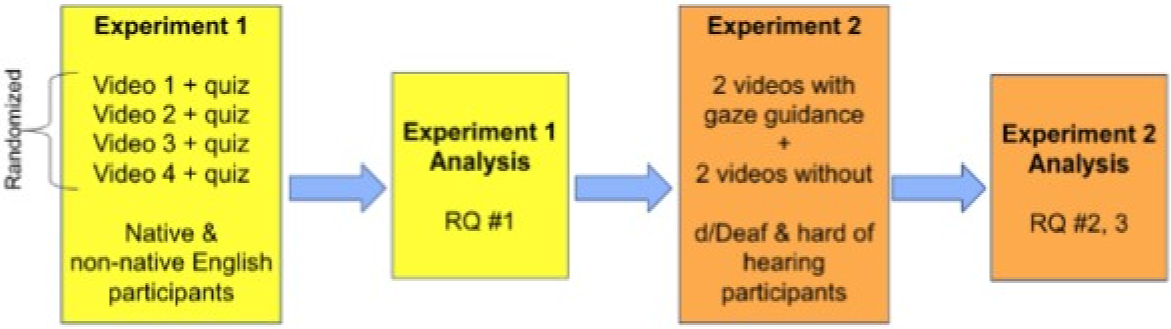


Fig. 2. Overview of experiments, which address research questions.

Experiment 1

In Experiment 1, participants viewed four captioned video lectures. Each video was followed by 12 to 13 comprehension questions and 7 subjective questions (closed-ended Likert-scale ordinal questions) about the video-viewing experience. Four subjective questions were from a prior study (Berke et al. 2018), and 3 were a subset of NASA Task Load Index questions used in (Kafle et al. 2019).

A two-sample t-test comparing the percent fixation time of native English (M = 27.89, SD = 16.33) and nonnative English (M = 39.57, SD = 20.49) participants showed that percent fixation time in the captions was significantly greater in the nonnative group (Figure 3); Experiment 1 results supported that non-native speakers make more use of captions and look at captions more (RQ1).

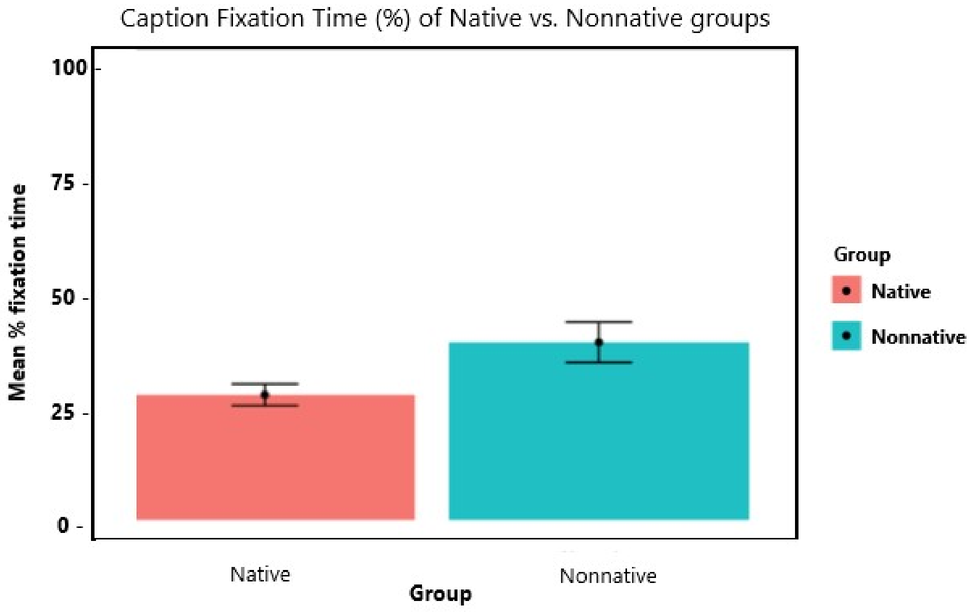


Fig. 3. Percent fixation times by language.

Experiment 2

As the eye movements of native and non-native speakers were significantly different, we only used the eye data of non-native English users to select our gaze guidance locations for DHH users. To identify consensus locations/times when hearing individuals looked at locations in the video, we analyzed the eye data using mean shift clustering[1]. As illustrated in Figure 4, the top 30 clusters (from each 2.5-minute video) were used to select the location and timing for inserting gaze guidance modulations into each video. Modulations consisted of brief light/dark flickers on

[1] From https://github.com/mattnedrich/MeanShiftypy with a bandwidth value of 125

a region of the video, with 0.5 seconds duration; prior work found these settings to be 75% successful at drawing viewers' attention (Bailey et al. 2009).

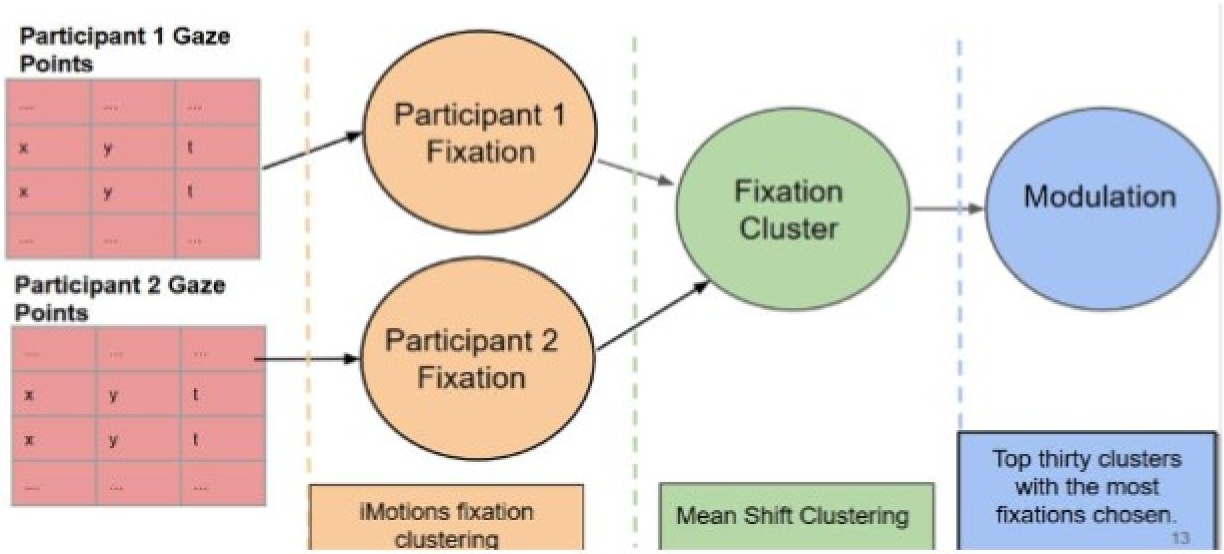


Fig. 4. Algorithm to select gaze guidance modulations.

Experiment 2 was identical in design with Experiment 1, except the participants were DHH individuals and half of the videos included gaze guidance (Latin squares assignment of the guidance or no-guidance condition for each participant).

A paired t-test did not reveal any significant difference in users' percent fixation time looking at the caption text region of the video, when comparing guidance and no-guidance conditions (Figure 5); t(10) = -1.2779, p = 0.1151. Thus, we did not observe that gaze guidance drew DHH users' gaze away from the caption region of the video significantly.

A paired t-test did not reveal any significant difference in comprehension-question accuracy scores, when comparing the guidance or no-guidance videos (Figure 6); t(10) = -1.2593, p = 0.8817. Thus, we did not observe any improvement in comprehension-question success from adding gaze guidance to videos (RQ3).

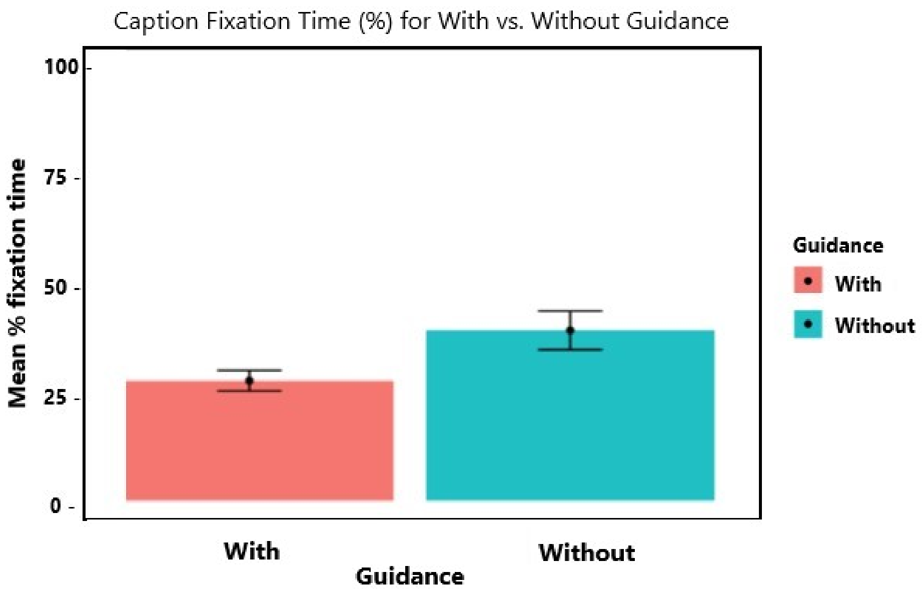


Fig. 5. Percent fixation times by guidance.

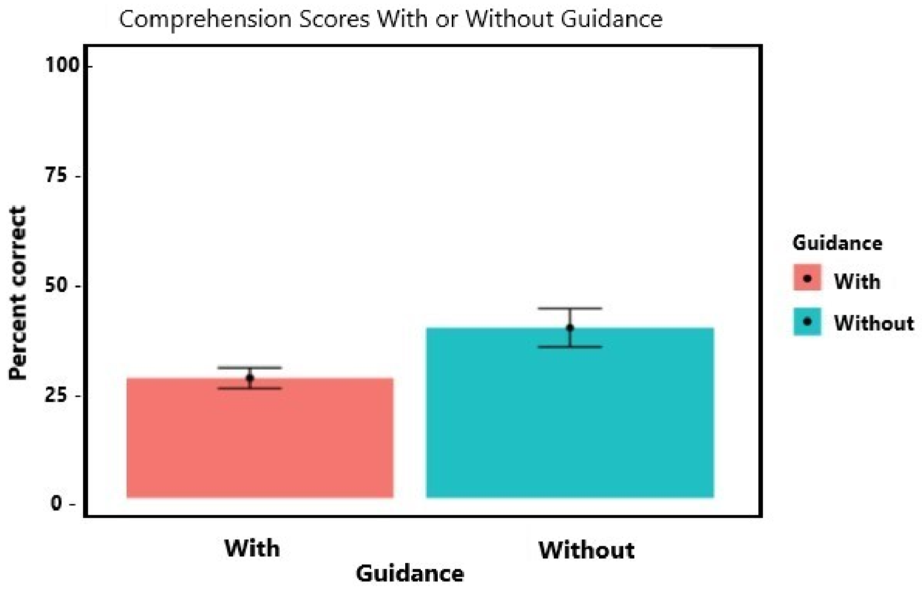


Fig. 6. Comprehension scores.

Limitations and Future Work

We identified four limitations of our study and propose future work to address this:

Placement of Modulations: Hearing participants did not always look where important information was; so, some of their eye data did not yield helpful target locations. Future work could improve the algorithm to help it aim to areas of the screen that likely have important information. *Style of Modulations*: The gaze guidance modulations in this study were set based on Bailey et al. (2009), but future work could investigate different gaze guidance modulation, tailored to DHH users. *Style of Video*: We only examined educational-lecture videos in this study, but future work could investigate different video genres. Small sample size: In future work, we will repeat this study with a larger number of participants.

Conclusion

We studied whether gaze guidance helps DHH individuals understand video content. We found that the eye-gaze patterns of non-native English users differ from native English users, and we demonstrated an algorithm for using the eye movements of non-native hearing individuals to predict targets where we could suggest a DHH individual to direct their gaze during videos. We applied gaze guidance to videos using this algorithm, and evaluated the resulting videos with DHH individuals. However, we did not observe a significant difference in the degree to which DHH individuals in our study looked away from the captions nor had success on comprehension questions, when comparing the guidance or no-guidance conditions. We speculate that further research is needed on identifying optimal gaze guidance appearance parameters for DHH users in this educational lecture video context.

Acknowledgements

This material is based upon work supported by the National Science Foundation under Award Nos. 1851591, 1462280, 1540396, 1763569, 1822747; by the Department of Health and Human Services under Award No. 90DPCP0002-01-00; and by a Microsoft AI for Accessibility Award. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of sponsors.

Works Cited

Bailey, Reynold, Ann McNamara, Nisha Sudarsanam, and Cindy Grimm. “Subtle gaze direction.” *ACM Transactions on Graphics (TOG)* 28.4 (2009): 100. Web.

Berke, Larwan, Christopher Caulfield, and Matt Huenerfauth. “Deaf and Hard-of-Hearing Perspectives on Imperfect Automatic Speech Recognition for Captioning One-on-One Meetings.” *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*. Baltimore, Maryland, USA: ACM, 2017. 155–164. Web. ASSETS ‘17.

Berke, Larwan, Sushant Kafle, and Matt Huenerfauth. “Methods for Evaluation of Imperfect Captioning Tools by Deaf or Hard-of-Hearing Users at Different Reading Literacy Levels.” *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. Montreal QC, Canada: ACM, 2018. 91:1–91:12. Web. CHI ‘18.

Cummins, Jim. “Bilingual Education and Special Education: Issues in Assessment and Pedagogy.” San Diego: College Hill, 1984.

Dorr, Michael, et al. “Eye movement modelling and gaze guidance.” *Fourth International Workshop on Human-Computer Conversation*. Citeseer. 2008. Web.

Etemadi, Aida. “Effects of bimodal subtitling of English movies on content comprehension and vocabulary recognition.” *International journal of English linguistics* 2.1 (2012): 239. Web.

Evans, Charlotte J. “Literacy development in deaf students: Case studies in bilingual teaching and learning.” *American Annals of the Deaf* 149.1 (2004): 17–27. Web.

Hayati, Abdolmajid and Firooz Mohmedi. “The effect of films with and without subtitles on listening comprehension of EFL learners.” *British Journal of Educational Technology* 42.1 (2011): 181–192. Web.

Harper, Allen V. R. “Eye Tracking and Performance Evaluation: Automatic Detection of User Outcomes.” (2015).

Kafle, Sushant, and Matt Huenerfauth. “Evaluating the Benefit of Highlighting Key Words in Captions for People who are Deaf or Hard of Hearing.” *Proceedings of the 21th International ACM SIGACCESS Conference on Computers & Accessibility*. Pittsburgh, Pennsylvania, USA: ACM, 2019. Web. ASSETS ‘19.

Koolstra, Cees M and Jonannes WJ Beentjes. “Children’s vocabulary acquisition in a foreign language through watching subtitled television programs at home.” *Educational Technology Research and Development* 47.1 (1999): 51–60. Web.

Kushalnagar, Raja S., et al. “Enhancing Caption Accessibility Through Simultaneous Multimodal Information: Visual-tactile Captions.” *Proceedings of the 16th International ACM SIGACCESS Conference on Computers & Accessibility*. Rochester, New York, USA: ACM, 2014. 185–192. Web. ASSETS ‘14.

Kushalnagar, Raja S., 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: https://doi.org/10.1145/1878803.1878827

Lasecki, Walter S., Raja Kushalnagar, and Jeffrey P. Bigham. "Helping students keep up with real-time captions by pausing and highlighting." Proceedings of the 11th Web for All Conference. ACM, 2014.

Pomarjanschi, Laura, Michael Dorr, and Erhardt Barth. “Gaze Guidance Reduces the Number of Collisions with Pedestrians in a Driving Simulator.” *ACM Trans. Interact. Intell. Syst*. 1.2 (Jan. 2012): 8:1–8:14. Web.

Rathbun, Kevin, Larwan Berke, Christopher Caulfield, Michael Stinson, and Matt Hunerfauth. “Eye movements of deaf and hard of hearing viewers of automatic captions.” *Journal on Technology & Persons with Disabilities 5* (2017): 130–140. Web.

Santella, Anthony and Doug DeCarlo. “Robust clustering of eye movement recordings for quantification of visual interest.” *Proceedings of the 2004 symposium on Eye tracking research & applications*. ACM. 2004. 27–34. Web.

Stewart, Melissa A and Inmaculada Pertusa. “Gains to language learners from viewing target language closed-captioned films.” *Foreign language annals* 37.3 (2004): 438–442. Web.

Szarkowska, Agnieszka, Izabela Krejtz, Zuzanna Klyszejko, and Anna Wieczorek. “Verbatim, Standard, or Edited?: Reading Patterns of Different Captioning Styles Among Deaf, Hard of Hearing, and Hearing Viewers.” *American Annals of the Deaf* 156.4 (2011): 363–378. Web.

Van der Zee, Tim, Wilfried Admiraal, Fred Pass, Nadira Saab, and Bas Giesbers. “Effects of subtitles, complexity, and language proficiency on learning from online education videos.” *Journal of Media Psychology* 29 (2017): 18–30. Web.

Waller, James M. and Raja S. Kushalnagar. “Evaluation of Automatic Caption Segmentation.” *Proceedings of the 18th International ACM SIGAC-CESS Conference on Computers and Accessibility*. Reno, Nevada, USA: ACM, 2016. 331–332. Web. ASSETS ‘16.

Winke, Paula, Susan Gass, and Tetyana Sydorenko. “Factors influencing the use of captions by foreign language learners: An eye-tracking study.” *The Modern Language Journal* 97.1 (2013): 254–275. Web.