Preferences of Deaf or Hard of Hearing Users for Live-TV Caption Appearance

Akhter Al Amin[1], Abraham Glasser[1], Raja Kushalnagar[2], Christian Vogler[2], Matt Huenerfauth[1]

Rochester Institute of Technology, Rochester NY, USA
{aa7510,atg2036,mphics}@rit.edu
, Gallaudet University, Washington DC, USA
{raja.kushalnagar,christian.vogler}@gallaudet.edu

1 Abstract(800 Words)

There are a variety of appearance variations in television and streaming video captioning, e.g. text color, font typeface, caption background, number of lines, caption placement, etc. The effect of these variations in appearance properties on the TV-watching experience of Deaf and Hard of Hearing (DHH) users has not been specifically addressed in existing research-based guidelines nor incorporated into the design of state-of-the-art caption-evaluation metrics. Therefore, we have investigated what appearance attributes are preferred by DHH users while watching captioned videos.

To ensure DHH users' access to speech and non-speech onscreen information, some regulatory organizations require broadcasters to provide captions with specific characteristics, e.g. verbatim caption text, synchrony between caption text and the person speaking, or captions not occluding any salient onscreen content [8, 1, 11]. To support more frequent evaluation of the quality of captioning in broadcast television, a variety of automated caption-quality metrics have been introduced, yet most evaluate a caption based on the quality of transcription of the spoken content, e.g. [2–4, 9]. Despite this progress in guidelines and metrics, DHH viewers are not fully satisfied with the quality of captions provided by TV broadcasters [15, 10]. Additional research is needed on how to evaluate caption-quality in regard to appearance characteristics, e.g. the contrast between caption text and background color, font size, movement and placement style, and other issues, which have been shown to influence DHH viewers' perception of information during captioned video [13, 12, 1].

In this study, we held two 90-minute focus groups, with a total of 17 DHH participants aged 18 to 45 (median = 25). Participants watched videos with captions that had various properties, which we had engineered to reflect various common styles of text appearance. Participants answered questions and discussed their prior experience and preferences for the appearance of captions. We performed a thematic analysis with inductive open-coding of the text transcripts of the sessions [5]. Several major themes that emerged from this analysis are described below:

1.1 Preferred Caption Background

We displayed sample videos with three commonly used caption backgrounds: black, grey (semi-transparent), or transparent. Across ten participant comments, there was a preference for grey or transparent background, with participants indicating that it increased visibility of onscreen content:

> I agree, the grey or transparent one is the best. You can still see some visibility of what is behind it. You can still see some behind it. (P11)

This preference for content visibility aligns with prior work and guidelines, e.g. [6–8, 1].

1.2 Preferred Caption Font Size

After watching videos containing caption text of various font sizes as shown in Figure 1, there was consensus among participants that font size had a strong influence on the readability of the captions such as:

> Bigger is easier to read. (P9)

While this finding is consistent with [14], some said larger font might impede viewers' opportunity to watch other graphical contents.

> If the text is too big, then I have to look at it constantly, because I'm not getting enough information fast enough. (P11)

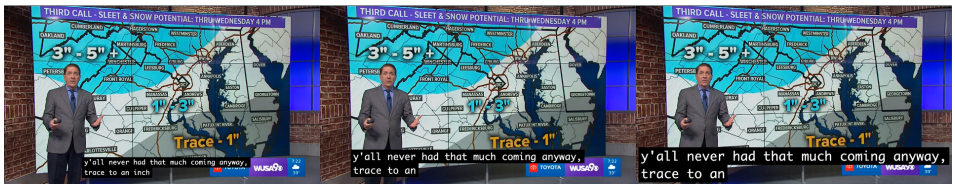

Fig. 1. Television weather news image with captions at small, medium, and large sizes.

1.3 Preferred Caption Text Color and Typeface

Participants' individual preferences for text color or typeface varied. Several discussed how the selection of these properties may depend upon the genre of video, e.g.

> I have never changed the font for captions, but it would be cool for fonts to match the genre of the show, e.g. Harry Potter (wizard style), or scary movie (spooky font). (P17)

1.4 Preferred Caption Movement Style

Participants' preferences for various forms of caption-movement also varied. Four of our 17 participants preferred a pop-up style (in which new text appeared as chunks), and six preferred roll-up style (in which text scrolls as new lines appear):

> If there is scrolling, then my eyes need to follow the moving words as I am reading words. I wish they would step up instead of roll up. (P1)

> Maybe a little faster, but stays longer onscreen, too. If that makes sense. (P12)

Two participants indicated that their preference for movement style was genre-specific, i.e. with roll-up for news and pop-up for movies:

> News or live program scrolling is fine but movies should have pop up. (P2)

1.5 Preferred Caption Placement

While some television programs position captions in various locations onscreen, e.g. to be closer to the individual who is speaking, several of our participants indicated a preference for captions to remain in a single location onscreen.

> My preference would be for the caption to stay in one place on the screen. I prefer captions to remain in one place so my eyes don't move all over. (P5)

This finding aligned with prior work [7].

1.6 Summary of Contribution

This study contributes empirical evidence about DHH viewer's preferences for caption appearance. The key findings have been briefly summarized in this abstract, and the full-length paper would provide additional methodological details, participant quotations, and discussion of these findings – especially in regard to how they may inform future guidelines for caption appearance or the creation of new metrics for evaluating the quality of captions for these users.

References

1. The described and captioned media program, 2010, captioning key for educational media, retrieved from: http://access-ed.r2d2.uwm.edu/resources/captioning-key.pdf.
2. Ali, Ahmed and Renals, Steve. (2018) Word Error Rate Estimation for Speech Recognition: e-WER.Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers), 20-24.

3. Romero-Fresco P., Pérez J.M. (2015) Accuracy Rate in Live Subtitling: The NER Model. In: Piñero R.B., Cintas J.D. (eds) Audiovisual Translation in a Global Context. Palgrave Studies in Translating and Interpreting. Palgrave Macmillan, London
4. Caption Accuracy Metrics Project Research into Automated Error Ranking of Real-time Captions in Live Television News Programs The Carl and Ruth Shapiro Family National Center for Accessible Media at WGBH (NCAM) By Tom Apone, Brad Botkin, Marcia Brooks and Larry Goldberg September 2011
5. Fatemeh Rabiee. Focus-group interview and data analysis. Proceedings of the Nutrition Society, 63(4):655–660, 2004. DOI: http://10.1079/PNS2004399
6. Bo Jiang, Sijiang Liu, Liping He, Weimin Wu, Hongli Chen, and Yunfei Shen. Subtitle positioning for e-learning videos based on rough gazeestimation and saliency detection. InSIGGRAPH Asia Posters. 15–16. 2017
7. Kuno Kurzhals, Fabian Göbel, Katrin Angerbauer, Michael Sedlmair, and Martin Raubal. A view on the viewer: Gaze-adaptive captions for videos. InProceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI '20, page 1–12, New York, NY, USA, 2020. Association for Computing Machinery.
8. Federal Communications Commission (2014). Closed Captioning Quality Report and Order, Declaratory Ruling, FNPRM. Online: https://www.fcc.gov/document/closed-captioning-quality-report-and-order-declaratory-ruling-fnprm
9. 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 (ASSETS '17). Association for Computing Machinery, New York, NY, USA, 165–174. DOI:https://doi.org/10.1145/3132525.3132542
10. Ofcom. Measuring live subtitling quality, uk., 2015. Online: https://www.ofcom.org.uk/research-and-data/tv-radio-and-on-demand/tv-research/live-subtitling
11. BBC.BBC Subtitle Guidelines. 2019. Online: https://bbc.github.io/subtitle-guidelines
12. Larwan Berke, Khaled Albusays, Matthew Seita, and Matt Huenerfauth. Preferred appearance of captions generated by automatic speech recognitionfor deaf and hard-of-hearing viewers. InExtended Abstracts of the 2019 CHI Conference on Human FaWERctors in Computing Systems, CHI EA '19,page 1–6, New York, NY, USA, 2019. Association for Computing Machinery.
13. Larwan Berke. Displaying confidence from imperfect automatic speech recognition for captioning.SIGACCESS Access. Comput., (117):14–18,February 2017.
14. Minjee Kim, Sung Hee Ahn, Youlbeen Kang, Soo-chan Jee, Sondo Kim, Ji Hwan Park, Hunsik Shin, and Myung-Hwan Yun. The effect of font anddisplay sizes on the readability for mobile devices. InProceedings of HCI Korea, HCIK '16, page 468–475, Seoul, KOR, 2016. Hanbit Media, Inc.
15. S. Nam, D. I. Fels, and M. H. Chignell. Modeling closed captioning subjective quality assessment by deaf and hard of hearing viewers.IEEETransactions on Computational Social Systems, 7(3):621–631, 2020.