Springer Link

SubtitleFormatter: Making Subtitles Easier to Read for Deaf and Hard of Hearing Viewers on Personal Devices

International Conference on Computers Helping People with Special Needs

ICCHP 2018: Computers Helping People with Special Needs pp 211-219 | Cite as

Part of the Lecture Notes in Computer Science book series (LNCS, volume 10896)

Abstract

For deaf or hard of hearing (DHH) viewers who cannot understand speech, many countries require video producers/distributors to provide speech-to-text over the video, also called subtitles that can be turned on or off by the viewer. These subtitles must comply with national subtitle quality standards. The growth in video capable personal devices has shifted viewers away from watching broadcast video on a standardized television display and towards watching video on interactive personal devices. However, personal devices range widely from tiny watch displays to enormous television displays, with different proportions which impact subtitle readability. SubtitleFormatter automatically formats subtitles according to a display's screen size and minimum font size for reading. A user study of subtitle formatting evaluates subtitle readability, and finds that viewers preferred SubtitleFormatted-segmented subtitles over wrap around (arbitrarily-formatted) subtitles.

Keywords

Subtitles   Speech-to-text   Deaf or hard of hearing
[Download](#) conference paper PDF

1 Introduction

Hearing loss is an invisible but significant barrier in daily life, including education and streaming television. In addition to approximately 2% of people born deaf or hard of hearing, approximately 31% of people over 65 have significant hearing loss [1]. Many rely on subtitles to access and enjoy videos.

Videos are a vital part of our shared cultural experience and shapes our identity as citizens. Subtitles provides accessibility to these individuals such that they are not shut out of society and culture. Subtitling laws and policies in many countries, including the United States, United Kingdom, Brazil and India, guarantee some access to video programming [2, 3].

These acts mandate aural-to-visual accommodations such as subtitles as shown in Fig. 1 which aids both people with disabilities, and people with situational accessibility needs. For instance, subtitles have been shown to be useful across a wide range of situations, such as bars, restaurants, airports and to improve literacy skills in children and people learning English as a Second Language.
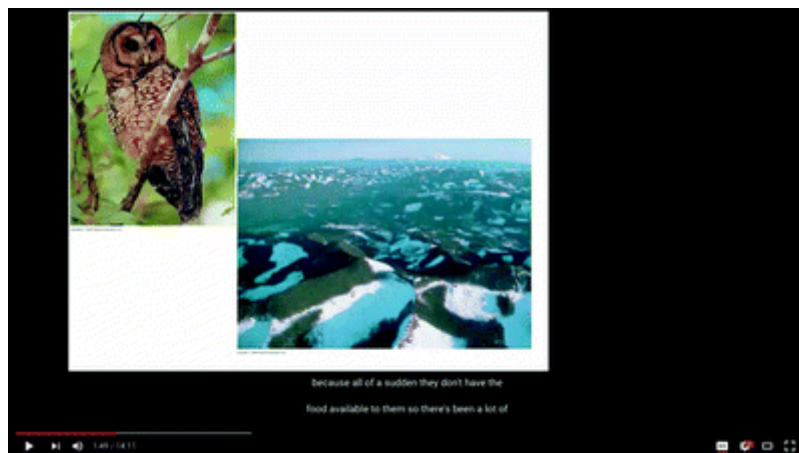

Fig. 1.

Video with subtitles. The subtitles are the white letters at the bottom.

2 Background

In the United States, pre-prepared subtitles were included in national TV shows from 1973 [4], and real-time subtitles were included in television shows from 1982 [5]. The timeline for introducing subtitles in television was similar in most developed countries. For over 30 years, standards for subtitling [6] were developed and standardized for an average DHH viewer who watched analog broadcasts on non-interactive, fixed format television displays. The television screen resolution and proportions were set for standard resolution (e.g., NTSC: 720 by 480 pixels or PAL: 720 by 576 pixels with an aspect ratio of 4:3). This resolution and aspect ratio remained unchanged till the advent of digital broadcasts in the 2000s.

2.1 Subtitled Videos on Personal Devices

The advent of digital television on personal displays led to far more diversity in resolution, size and proportions. Viewers consume video programming personal devices with varying resolutions and aspect ratios. This diversity of personal display characteristics can influence a viewer's preferred size and number of caption lines. Although viewers can view videos anywhere, any time, and no longer be tethered to their couches, it becomes harder to fit subtitles on the widely varying sizes and proportions of personal devices. Viewing devices range from tiny smart watch displays to enormous television displays. While there has been a substantial body of research focused on a standardized speech-to-text presentation for all users watching television programs built over 30 years, there is scant research focused on adapting subtitles according to the display characteristics.

Currently, most video providers follow television captioning standards. When they do add features in their speech-text caption interfaces, these features create additional problems. For example, if the interface offers resizable fonts, and the font is made bigger, the caption lines will become too big for the video display and wrap around, which disrupts the reading process. To the best of our knowledge, no video platform provides a feature to reformat subtitles depending on screen size or user preferences.

3 Related Work

We focus on improving two parts of the closed caption reading process – the cognitive process of reading subtitles, and the process of segmenting and formatting subtitles to fit the video display.

3.1 Cognitive Process of Subtitle Reading

Prior research has shown that the cognitive process of reading subtitles that are a form of real-time speech-to-text is different from reading static text or print. Speech-to-text is short and regularly changes, while print is long, formatted and does not change [7]. Reading subtitles often takes relatively more time and energy than it does to listen to spoken or signed languages, and those watching subtitles must often split their attention between the subtitles and other visual information, such as whatever is happening on the TV screen [8].

This study focuses on how to divide up lines on subtitles to maximize readability and comprehension. Professionals format the subtitles manually to ensure that the number of words per line to a standard television display. When these subtitles are viewed on a smaller screen than the standard television screen, the subtitles should be made larger to maintain readability and the line width and count should be adjusted downwards. Then there is not enough screen real-estate to accommodate the subtitles, and the caption lines will become too big for the video display and wrap around, which disrupts the viewer's reading process as shown in Fig. 2. Conversely, on large displays, the font size does not have to grow as much as the video, and it may work to increase the number of simultaneously displayed lines.

Fig. 2.

Picture on left – captions on small personal device display. The viewer increases subtitle size to comfortably read them. The size increase makes the line too long for the screen, and it wraps around, which impacts caption readability. Picture on right – captions on television display. The subtitles are

formatted for this resolution and screen by default. The subtitles normally follow caption guidelines to maximize readability for most viewers of the video.

Research on automatic caption segmentation seems to suggest it can have a positive impact, but is not conclusive. Perego et al. [9] found that segmenting subtitles in inappropriate places had no impact on sentence recall or eye movement. However, Rajendran et al. [10] found a significant difference in eye movement for different kinds of subtitle segmentation. Waller and Kushalnagar [11] suggest that segmentation may have an impact on our memory of the text.

4 SubtitleFormatter Design and Evaluation

We created a linguistically aware automatic formatting system, called SubtitleFormatter. The system automatically formats subtitles by parsing the text to break the text at linguistically appropriate places.

We conducted a user study to verify the utility of the system's parsing and breaking of subtitles. We compared the readability of unparsed subtitles versus human parsed subtitles that were generated by professional closed caption stenographers on both regular television screens and for small phone screens. We also compared the readability of unparsed subtitles versus automatic subtitles on both regular and small phone screens, to investigate whether the readability can reach the level and quality of human-parsed subtitles generated by professional closed caption stenographers.

4.1 Design

The SubtitleFormatter system analyzes the subtitles and the display so that it can format the subtitles according to the display. It has two parts – a linguistic analyzer and a display analyzer.

4.2 Linguistic Analyzer

The system incorporates the Stanford Parser which is an open-source statistical-based natural language processing processor [12]. The parser can produce a syntax tree for any given sentence or sentences in a text. For example, for

the sentence "When it rains, the children like to play outside", the parser can display a syntax tree for the sentence as shown in Fig. 3.
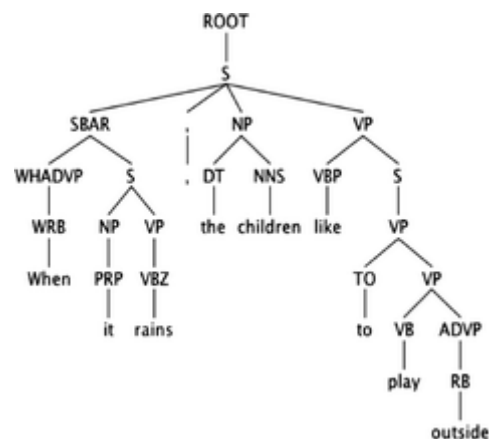

Fig. 3.

The parse tree generated by the Stanford Parser for the sentence: "When it rains, the children like to play outside."

It breaks down the sentence into phrases, which in turn are broken down into smaller and smaller phrases down to the level of words. The SubtitleFormatter system uses this information to identify optimal point breaks.

5 Evaluation

To evaluate caption segmentation and formatting, we evaluated the difference in readability between unparsed subtitles (A), human-parsed (H), and SubtitleFormatter-parsed (P) subtitles on a regular television display and on a personal phone.

For A, we counted the characters per line limit and split subtitles only after the line width was over the maximum in subtitling guidelines, as shown in Fig. 4, on right. For P, the program broke up the lines using as shown in Fig. 4, on left.

Fig. 4.

Left: snapshot of SubtitleFormatter, where a break is inserted when it encounters a logical break according to the parse tree, and still less than the maximum length. Right: A snapshot of unparsed subtitles that was segmented and formatted automatically by inserting a break when the line has exceeded the maximum length.

5.1 Participants

We recruited 34 deaf and hard of hearing participants. All participants regularly use captions when watching online videos, TV, and other audio-video content. The participants ranged in age from 20 to 48 years old: 20 men, and 14 women. By ethnicity, 18 participants identified as white, 5 identified as black, 5 identified as Asian or Asian-American, 4 identified as Hispanic and 2 as multiracial.

5.2 Study

Each participant watched six 4-minute videos with A, H or P subtitles, with breaks in between. The survey and videos were shown either on a 40-inch television set or on iPhone 5 s. The entire experiment took 30–45 min. Half viewed the first three videos on a 40-inch high-definition (1920 × 1080) television display, and the next three videos on a 4-inch iPhone 5 (1136 × 640) display, and the other half watched in the reverse order. Each viewer watched all six videos in a balanced, randomized order. We gathered data from all participants through three parts – a Likert rating questionnaire, a comprehension questionnaire, a sentence completion task, and an eye-tracking data gathering part. After each video, the participants were given a sentence completion task in which they were presented with the beginning part of a sentence from the text. Afterwards the researcher explained the purpose of the study, and the participant was invited to add a comment on either the study or subtitles.

6 Results

The results from the following evaluations were grouped by Likert ratings, comprehension scores including sentence completion scores. The Shapiro-Wilk test indicated the observed values were not normally distributed, so non-parametric testing was done. The Wilcoxon Signed Rank tests was used to perform post-hoc comparisons between pairs of samples, with Bonferroni corrections to address the multiple comparisons.

6.1 Likert Ratings

Television Display

For the subjective responses, no significant differences were found: A vs. H: Wilcox = 54, $p > 0.05$; A vs. P: Wilcox = 45.5, $p > 0.10$; H vs. P: Wilcox = 48.5, $p > 0.05$.

For satisfaction, none of the three conditions had a significant impact on the ratings relative to each other: A vs. H: Wilcox = 48.5, $p > 0.05$; A Vs. P: Wilcox = 51, $p > 0.05$; H vs. P: Wilcox = 75.5, $p > 0.05$.

Phone Display

For the subjective and satisfaction responses, there was a significant difference between A vs. H, and A vs. P, but not H vs. P.

For satisfaction, there was no significant difference between A vs. H and A vs. P, but there was a significant difference between P and H: H vs. P: Wilcox = 14.8, $p < 0.05$.

6.2 Comprehension Scores

Television Display

For general comprehension questions, there was no significant difference: A vs. H: $t = 0.583$, $p > 0.05$; A vs. P: $t = 0.432$, $p > 0.05$; H vs. P: $t = -0.24$, $p > 0.05$.

For sentence completion, A was significantly different than either H or P: A vs. H: $t = -1.048$, $p < 0.05$; A vs. P: $t = 1.052$, $p < 0.05$, but not H vs. P: $t = 0.414$, $p > 0.05$.

However, participants got significantly more sentence completion questions correct for P than A: $t = 2.169$, $p < 0.05$, but not for the other pairwise comparisons.

Phone Display

For general comprehension, none of these means differed significantly: A vs. H: $t = 0.541$, $p > 0.05$; A vs. P: $t = 0.471$, $p > 0.05$; P vs. H: $t = 0.349$, $p > 0.05$.

For sentence completion, H was not significantly different than either of the other conditions: A vs. H: $t = -1.276$, $p > 0.05$; H vs. P: $t = 1.874$, $p > 0.05$.

However, participants got significantly more sentence completion questions for P vs. A: $t = 2.3224$, $p < 0.05$, but not for the other pairwise comparisons.

6.3 Comments

The participant comments were generally negative about unsegmented subtitles. They were generally positive about grammatically segmented subtitles, generated either by professionals, or SubtitleFormatter program, on both on small and large displays. When the participants had different comments between large and small displays, they said that the lines or words were too hard to see on small displays. On large displays, they said the lines were not complete, or that they did not like it and could not explain why. For human-segmented subtitles, 28 participants

wrote comments, and 15 wrote down identical comments for both small and large displays. When they wrote the same comment for both, they said that it was easy to read, or that they could understand each line. When the participants had different comments between large and small displays, they generally said that the lines or words were too fast on small displays, and that on large displays, they said the lines were not complete.

7 Results

Participants significantly preferred H or P segmented subtitles for smaller screens, but not for bigger screens. They performed significantly better on sentence completion tasks for either P or H over A. However, they did not perform significantly better on either H or P over A for either large or small screens. The general lack of significant differences between H and A for bigger screens agrees with the assertion by Perego et al. that segmentation in captioning has little or no impact on readability. It is possible that human or SubtitleFormatter subtitles were easier to remember, but the difference from unsegmented subtitles did not rise to the level of significance.

The preference for human or SubtitleFormatter subtitles on smaller screens could be that when lines are shorter, the sentence concepts are more likely to be distributed on multiple lines and that breaking outside of phrases is likely to be confusing. The fact that segmentation has an impact on user preferences and sentence recall for smaller screens has important implications for captioning, as currently caption guidelines all encourage proper segmentation

The SubtitleFormatter supports viewer preferences for proportionately larger text on small screens by automatically adjusting caption line width according to screen size. It can be viewed as an automatic enhancement of accessibility for viewers who use captions, like how people with diverse magnification needs can benefit from automatic magnification. We present a novel approach to automatic subtitle segmentation which generates and selects optimal segmentation points according to SubtitleFormatter. SubtitleFormatter segmentation can be an inclusive approach for viewers who wish to follow best practice in segmentation guidelines and rules, or one that fits their own needs.

8 Future Work

Automatically formatting subtitles by display can also benefit people with limited English proficiency, or viewers with situational auditory barriers, e.g., quiet public spaces.

References

1. Pearson, J.D., Morrell, C.H., Gordon-Salant, S., Brant, L.J., Metter, E.J., Klein, L.L., Fozard, J.L.: Gender differences in a longitudinal study of age-associated hearing loss. J. Acoust. Soc. Am. **97**(2), 1196–1205 (1995)
   CrossRef (https://doi.org/10.1121/1.412231)
   Google Scholar (http://scholar.google.com/scholar_lookup?title=Gender%20differences%20in%20a%20longitudinal%20study%20of%20age-associated%20hearing%20loss&author=JD.%20Pearson&author=CH.%20Morrell&author=S.%20Gordon-Salant&author=LJ.%20Brant&author=EJ.%20Metter&author=LL.%20Klein&author=JL.%20Fozard&journal=J.%20Acoust.%20Soc.%20Am.&volume=97&issue=2&pages=1196-1205&publication_year=1995)

2. United States Congress, Americans with Disabilities Act, Pub. L. No. 101-336, 104 Stat. 328. United States of America (1990)
   Google Scholar (https://scholar.google.com/scholar?q=United%20States%20Congress%2C%20Americans%20with%20Disabilities%20Act%2C%20Pub.%20L.%20No.%20101-336%2C%20104%20Stat.%20328.%20United%20States%20of%20America%20%281990%29)

3. National Congress of Brazil, Ley de Igualdad de Oportunidades para las Personas con Discapacidad (Law of Equal Opportunities for Persons with Disabilities) (1992)
   Google Scholar (https://scholar.google.com/scholar?q=National%20Congress%20of%20Brazil%2C%20Ley%20de%20Igualdad%20de%20Oportunidades%20para%20las%20Personas%20con%20Discapacidad%20%28Law%20of%20Equal%20Opportunities%20for%20Persons%20with%20Disabilities%29%20%281992%29)

4. NIDCD: Captions For Deaf and Hard-of-Hearing Viewers, NIH Publ., no. 4834 (2017). https://www.nidcd.nih.gov/health/captions-deaf-and-hard-hearing-viewers (https://www.nidcd.nih.gov/health/captions-deaf-and-hard-hearing-viewers)

5. Block, M.H., Okrand, M.: Real-time closed captioned television as an educational tool. Am. Ann. Deaf **128**(5), 636–641 (1983)
   Google Scholar (https://scholar.google.com/scholar?q=Block%2C%20M.H.%2C%20Okrand%2C%20M.%3A%20Real-time%20closed%20captioned%20television%20as%20an%20educational%20tool.%20Am.%20Ann.%20Deaf%20128%285%29%2C%20636%E2%80%93641%20%281983%29)

6. United States General Publications Office, 47 CFR 79.A.1. (2015)
   Google Scholar (https://scholar.google.com/scholar?q=United%20States%20General%20Publications%20Office%2C%2047%20CFR%2079.A.1.%20%282015%29

)

7. Thorn, F., Thorn, S.: Television captions for hearing-impaired people: a study of key factors that affect
    reading performance. Hum. Factors **38**(3), 452–463 (1996)
    CrossRef (https://doi.org/10.1518/001872096778702006)
    Google Scholar (http://scholar.google.com/scholar_lookup?
    title=Television%20captions%20for%20hearing-
    impaired%20people%3A%20a%20study%20of%20key%20factors%20that%20affect%20reading%20perform
    ance&author=F.%20Thorn&author=S.%20Thorn&journal=Hum.%20Factors&volume=38&issue=3&pages=4
    52-463&publication_year=1996)

8. Kushalnagar, R.S., Behm, G.W., Stanislow, J.S., Gupta, V.: Enhancing caption accessibility through
    simultaneous multimodal information: visual-tactile captions. In: ASSETS14 - Proceedings of the 16th
    International ACM SIGACCESS Conference on Computers and Accessibility (2014)
    Google Scholar (https://scholar.google.com/scholar?
    q=Kushalnagar%2C%20R.S.%2C%20Behm%2C%20G.W.%2C%20Stanislow%2C%20J.S.%2C%20Gupta%2C
    %20V.%3A%20Enhancing%20caption%20accessibility%20through%20simultaneous%20multimodal%20inf
    ormation%3A%20visual-tactile%20captions.%20In%3A%20ASSETS14%20-
    %20Proceedings%20of%20the%2016th%20International%20ACM%20SIGACCESS%20Conference%20on%2
    0Computers%20and%20Accessibility%20%282014%29)

9. Perego, E., Del Missier, F., Porta, M., Mosconi, M.: The cognitive effectiveness of subtitle processing. Media
    Psychol. **13**(3), 243–272 (2010)
    CrossRef (https://doi.org/10.1080/15213269.2010.502873)
    Google Scholar (http://scholar.google.com/scholar_lookup?
    title=The%20cognitive%20effectiveness%20of%20subtitle%20processing&author=E.%20Perego&author=F.
    %20Missier&author=M.%20Porta&author=M.%20Mosconi&journal=Media%20Psychol.&volume=13&issue
    =3&pages=243-272&publication_year=2010)

10. Rajendran, D.J., Duchowski, A.T., Orero, P., Martínez, J., Romero-Fresco, P.: Effects of text chunking on
    subtitling: a quantitative and qualitative examination. Perspectives (Montclair) **21**(1), 5–21 (2013)
    CrossRef (https://doi.org/10.1080/0907676X.2012.722651)
    Google Scholar (http://scholar.google.com/scholar_lookup?
    title=Effects%20of%20text%20chunking%20on%20subtitling%3A%20A%20quantitative%20and%20qualitat
    ive%20examination&author=Dhevi%20J..%20Rajendran&author=Andrew%20T..%20Duchowski&author=Pi
    lar.%20Orero&author=Juan.%20Mart%C3%ADnez&author=Pablo.%20Romero-
    Fresco&journal=Perspectives&volume=21&issue=1&pages=5-21&publication_year=2013)

11. Waller, J.M., Kushalnagar, R.S.: Evaluation of automatic caption segmentation. In: Proceedings of the 18th
    International ACM SIGACCESS Conference on Computers and Accessibility - ASSETS 2016, pp. 331–332

(2016)

[Google Scholar](https://scholar.google.com/scholar?q=Waller%2C%20J.M.%2C%20Kushalnagar%2C%20R.S.%3A%20Evaluation%20of%20automatic%20caption%20segmentation.%20In%3A%20Proceedings%20of%20the%2018th%20International%20ACM%20SIGACCESS%20Conference%20on%20Computers%20and%20Accessibility%20-%20ASSETS%202016%2C%20pp.%20331%E2%80%93332%20%282016%29)

12. Klein, D., Manning, C.D.: Accurate unlexicalized parsing. In: Proceedings of the 41st Annual Meeting on Association for Computational Linguistics - ACL 2003, vol. 1, pp. 423–430 (2003)
     [Google Scholar](https://scholar.google.com/scholar?q=Klein%2C%20D.%2C%20Manning%2C%20C.D.%3A%20Accurate%20unlexicalized%20parsing.%20In%3A%20Proceedings%20of%20the%2041st%20Annual%20Meeting%20on%20Association%20for%20Computational%20Linguistics%20-%20ACL%202003%2C%20vol.%201%2C%20pp.%20423%E2%80%93430%20%282003%29)

Copyright information

About this paper

Cite this paper as:

Kushalnagar R., Kushalnagar K. (2018) SubtitleFormatter: Making Subtitles Easier to Read for Deaf and Hard of Hearing Viewers on Personal Devices. In: Miesenberger K., Kouroupetroglou G. (eds) Computers Helping People with Special Needs. ICCHP 2018. Lecture Notes in Computer Science, vol 10896. Springer, Cham

- Buy this book on publisher's site
- Reprints and Permissions

Personalised recommendations

SPRINGER NATURE