Visionary Caption: Improving the Accessibility of Presentation Slides Through Highlighting Visualization

Carmen Yip
carmen.yip.2019@sis.smu.edu.sg
Singapore Management University
Singapore

Jie Mi Chong
jiemi.chong.2019@sis.smu.edu.sg
Singapore Management University
Singapore

Sin Yee Kwek
sinyee.kwek.2019@sis.smu.edu.sg
Singapore Management University
Singapore

Wang Yong
yongwang@smu.edu.sg
Singapore Management University
Singapore

Kotaro Hara
kotarohara@smu.edu.sg
Singapore Management University
Singapore

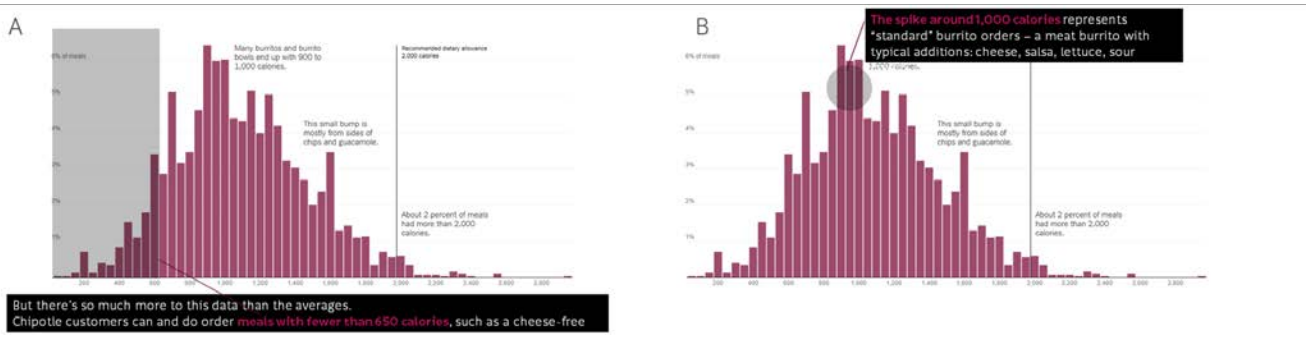

Figure 1: Screenshots of low-fidelity video prototypes. (a) *Annotation*. The prototype highlights a part of the visualization with a shape (a translucent rectangle or circle) and draws a connecting segment between the shape and the caption. (b) *Moving Box*. In addition to presenting a shape and a segment, the caption box moves to the proximity of the highlighted part of the visualization.

ABSTRACT

Presentation slides are widely used in occasions such as academic talks and business meetings. Captions placed on slides support deaf and hard of hearing (DHH) people to understand spoken contents, but simultaneously comprehending and associating visual contents on slides and caption text could be challenging. In this paper, we design and develop a visualization technique to highlight and associate chart on a slide and numerical data in caption. We first conduct a small formative study with people with and without hearing impairments to assess the value of the visualization technique using a lo-fidelity video prototype. We then develop *Visionary Caption*, a visualization technique that uses natural language processing to automatically highlight visual content and numerical phrases, and show the association between them. We present a scenario and personas to showcase the potential utility of *Visionary Caption* and guide its future development.

ASSETS '21, October 18–22, 2021, Virtual Event, USA

ACM ISBN 978-1-4503-8306-6/21/10.
https://doi.org/10.1145/3441852.3476539

CCS CONCEPTS

• **Human-centered computing** → **Human computer interaction (HCI)**; **Accessibility technologies**.

KEYWORDS

Accessibility, deaf and hard of hearing, information visualization

ACM Reference Format:
Carmen Yip, Jie Mi Chong, Sin Yee Kwek, Wang Yong, and Kotaro Hara. 2021. *Visionary Caption*: Improving the Accessibility of Presentation Slides Through Highlighting Visualization. In *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '21), October 18–22, 2021, Virtual Event, USA.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3441852.3476539

1 INTRODUCTION

Use of slides in presentation tools (*e.g.*, PowerPoint, Reveal.js) is prevalent in occasions like business meetings and academic talks. Such presentations tools, however, could pose a challenge for deaf and hard of hearing (DHH) people. DHH people use visual speech to comprehend spoken language [2, 9, 12–14], but visually attending to both slides and the speaker could tax their comprehension capacity due to visual dispersion [3]. Increasing use of slide-only presentations over video conferencing tools aggravate the problem by making lip-reading hard to impossible. Showing captions help

DHH people comprehend spoken contents. But simultaneously focusing on both captions and visual information presented on a slide and associating the two pieces of information could be a challenge.

In this project, we explored design of visualization techniques to alleviate the difficulty of associating visual information and caption text through a user-centered design process. We created three low fidelity video prototypes of visualization techniques that highlight a part of a chart on a slide and link it with a phrase in a caption box (Fig. 1). The design of the technique is inspired by research on video caption placement [1, 4, 6–8], visualization verbalization [5, 10, 11, 15] — a technique to convert data into natural language for easy comprehension of data — and a presentation practice where people point to a part of a slide using a presentation pointer. In the initial formative study, we showed low fidelity prototypes to four participants, of which two had hearing loss. We asked them to perform comprehension tasks and collected their informal feedback on the prototypes.

Convinced by the result of the formative study, we moved on to developing a high fidelity prototype of one of the visualization techniques. We used JavaScript (D3.js), HTML, and Python to develop a visualization technique — *Visionary Caption* — that dynamically shows the association between captions and a histogram visualization. Based on the formative study results and the experience of one of the authors with hearing impairment, we designed preliminary personas and a scenario that demonstrates the use of the system and drives the future development of *Visionary Caption*.

2 FORMATIVE STUDY

Method. We recruited $N = 4$ participants through personal connections. Of the four participants, two had severe hearing loss. All participants were students. We asked the participants to watch three video prototypes — *Default Caption* (Video Figure 1a), *Annotation* (Fig. 1a; Video Figure 1b), and Moving Box (Fig. 1b; Video Figure 1c). Each video prototype showed a histogram that visualizes a distribution of meals' calories ordered at Chipotle. The histogram was taken from a New York Times (NYT) article [1]. All the video prototypes showed text from a paragraph from the same NYT article as open captions. But their presentation differed. *Default Caption* showed captions at the bottom of the slide, just like typical subtitles. *Annotation* showed a circle or a rectangle on the histogram to highlight which part of the visualization a presenter was referring to. Numerical phrases in the caption that described the highlighted part of the visualization were colored magenta and a black segment connected the shape and the caption. *Moving Box* prototype also used shapes to highlight parts of the visualization, but the caption box moved around on the video frame such that it was placed in proximity to the shape. One of the authors read out the paragraph verbatim and the audio clip was used to create the three video prototypes. We uploaded videos to YouTube for ease of access. This had an added benefit of automatically generating timestamped caption files (.srt file).

After a participant finished watching each video, we asked him/her a comprehension question (*e.g.,* "*What is the proportion of meals that have more than 1,600 calories?*") and further clarified if the captions and visualization helped them understand which part of the histogram the speaker was referring to. We prepared a different comprehension question for each prototype. At the end of the study session, we also asked which prototype they preferred, what they thought were the pros and cons of each prototype, whether these kinds of visualization helped them understand or present visual data, what other things they would like to see in the visualization, and what methods they currently use to comprehend complicated visualizations in presentations.

Result. The two hard of hearing participants answered 2/3 and 1/3 comprehension questions correctly, while the participants without hearing impairments answered all the questions correctly. This may suggest that being able to hear helped them remember verbally described visual information more accurately. And for the participants with hearing impairments, focusing on both captions and visual information may have been taxing. All participants preferred the *Annotation* prototype (Fig. 1a) among the three prototypes. Participants did not have any comments about the design of *Default Caption*, because it looked like a conventional way of caption presentation in a video. All the participants disliked the *Moving Box* prototype, because they could not anticipate where the caption box would appear, disorientating them. Referring to the *Annotation* prototype, a participant with hearing loss noted that *"some parts of the graph were blocked off [by shapes], but not important details."* While the shapes and line segments did not severely occlude the visual contents, the comment suggested it is important to take a good care to not occlude visualization too much. A participant without hearing impairment suggested another prototype design where the user interface can zoom in to the part of the histogram instead of showing highlight annotations.

3 VISIONARY CAPTION

Based on the findings from the formative study, we decided to design and develop *Visionary Caption* — a technique that presents a histogram, captions, and shape annotations that highlights parts of the histogram that correspond to what are referred in the caption text (Video Figure 2). *Visionary Caption* is a web-based application. *Visionary Caption* had two components: (i) a back-end component that uses Python with spaCy NLP library and Flask web framework to process and serve caption text; and (ii) a front-end component that uses HTML and JavaScript (D3.js) to visualize the histogram, caption, and annotations. We used the same text and meals' calories data that were used in the formative study to develop the high fidelity prototype.

The back-end component identified phrases to be highlighted in a four-step process. First, it read the caption text from .srt file, performed tokenization, dependency parsing, and parts-of-speech tagging on the text. Second, the component extracted phrases containing numerical operations by finding 'nummod' token (*e.g.,* "1000" in "1000 calories") and extracted the phrase containing this token from the parse tree. Third, it classified the extracted phrases into three types of operations — 'mode', 'more than', and 'less than' — using prefixed rules. Finally, using the classification result and the numerical information, the component decided to show a shape as either a circle or a rectangle, and identify the coordinates and size of shapes. The data was served to the front-end with the timestamp

[1] At Chipotle, How Many Calories Do People Really Eat? https://www.nytimes.com/interactive/2015/02/17/upshot/what-do-people-actually-order-at-chipotle.html

information that it retained from the .srt file. See the supplemental material for the code.

The front-end visualization component used the data passed from the back-end to render shapes to highlight parts of the histogram and emphasize phrases (Video Figure 2). It rendered a circle to highlight 'the spike' since the presenter is bringing attention to that small area of the visualization. It draws a rectangle to annotate 'more than' or 'less than' area as the presenter is bringing attention to the area of the graph up to or more than the relevant number. Timing to strengthen text phrases and render shapes is synchronized using the timestamps taken from the original .srt file.

4 PERSONA AND SCENARIO

We describe one scenario with two personas to contextualize how *Visionary Caption* could be used in the future. We anticipate them to drive the future design and development iterations of *Visionary Caption*, too. The personas and scenarios are based on both our interaction with the formative study participants and the experience of one of the authors who has hard of hearing.

Abbie. She is an undergraduate student who is hard of hearing. She attends remote and on-site meetings for her course work, research, and internship. She reads captions during meetings when it is available (*e.g.,* meeting over Microsoft Teams where automated captioning is available by default). In a meeting where a speaker uses visual media heavily, she occasionally has a challenge in simultaneously comprehending visual media and looking at a speaker and/or caption.

Bob. Bob is a university professor. When giving presentations in the lectures or research meetings, he often uses charts (*e.g.,* a histogram) in his lecture materials and meeting slides.

Scenario. Abbie attends Bob's online lecture over Microsoft Teams where he presents consumers' behavior patterns in ordering fast food. Abbie turns on closed captions while listening to Bob's presentation. In one of his presentation slides, Bob shows a histogram in which the x-axis represents total calories of set menus and y-axis represents the frequency of orders. Abbie turns on *Visionary Caption*, then it dynamically highlights a part of the histogram that Bob is referring to with a circular or rectangular annotation and connects the annotation with the caption using a line segment. Abbie finds it easier to understand which part of the visualization Bob is referring to and it makes her understand the topic better.

5 DISCUSSION AND CONCLUSION

We conducted a formative study with four participants (two with hearing impairments) to design the tool to support them in watching a presentation with a visualization and caption. Based on the formative design step, we developed *Visionary Caption*, a technique that presents a histogram, caption, and shape annotation that highlights a part of the visualization which corresponds to what is referred in the caption text. We created a scenario and two personas to demonstrate how *Visionary Caption* could support presentation audience with hearing impairments.

Our participants with hearing impairments noted that visualization techniques like *Visionary Caption* could help them while watching a visual-heavy presentation. While the feedback helped us in directing our design and development effort, the data is not strong enough to suggest the utility of the visualization technique. Toward showing the usefulness of the visualization technique, we plan to evaluate the effectiveness of the *Visionary Caption* with DHH people with a larger sample size in the future. In the future work, we also plan to extend the prototype to support various types of visualizations other than histograms. We plan to improve the NLP component so that it can convert natural language into visualization more robustly. Another key technical goal is to supporting conversion of speech-to-visualization, which will use automated speech recognition technology to generate captions and converts them into visualization in real-time.

ACKNOWLEDGMENTS

This research was supported by the Singapore Ministry of Education (MOE) Academic Research Fund (AcRF) Tier 1 grant and a Lee Kong Chian Fellowship.

REFERENCES

[1] Andy Brown, Rhia Jones, Mike Crabb, James Sandford, Matthew Brooks, Mike Armstrong, and Caroline Jay. 2015. Dynamic Subtitles: The User Experience. In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video* (Brussels, Belgium) *(TVX '15)*. Association for Computing Machinery, New York, NY, USA, 103–112. https://doi.org/10.1145/2745197.2745204

[2] Adam B Buchwald, Stephen J Winters, and David B Pisoni. 2009. Visual speech primes open-set recognition of spoken words. *Language and cognitive processes* 24, 4 (2009), 580–610. https://doi.org/10.1080/01690960802536357

[3] 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* (Pittsburgh, Pennsylvania, USA) *(Assets '09)*. Association for Computing Machinery, New York, NY, USA, 67–74. https://doi.org/10.1145/1639642.1639656

[4] Michael Crabb, Rhianne Jones, Mike Armstrong, and Chris J. Hughes. 2015. Online News Videos: The UX of Subtitle Position. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility* (Lisbon, Portugal) *(ASSETS '15)*. Association for Computing Machinery, New York, NY, USA, 215–222. https://doi.org/10.1145/2700648.2809866

[5] Fred Hohman, Arjun Srinivasan, and S. Drucker. 2019. T ELE G AM : Combining Visualization and Verbalization for Interpretable Machine Learning.

[6] Richang Hong, Meng Wang, Xiao-Tong Yuan, Mengdi Xu, Jianguo Jiang, Shuicheng Yan, and Tat-Seng Chua. 2011. Video Accessibility Enhancement for Hearing-Impaired Users. *ACM Trans. Multimedia Comput. Commun. Appl.* 7S, 1, Article 24 (Nov. 2011), 19 pages. https://doi.org/10.1145/2037676.2037681

[7] Yongtao Hu, Jan Kautz, Yizhou Yu, and Wenping Wang. 2015. Speaker-Following Video Subtitles. *ACM Trans. Multimedia Comput. Commun. Appl.* 11, 2, Article 32 (Jan. 2015), 17 pages. https://doi.org/10.1145/2632111

[8] Chris J. Hughes, Mike Armstrong, Rhianne Jones, and Michael Crabb. 2015. Responsive Design for Personalised Subtitles. In *Proceedings of the 12th International Web for All Conference* (Florence, Italy) *(W4A '15)*. Association for Computing Machinery, New York, NY, USA, Article 8, 4 pages. https://doi.org/10.1145/2745555.2746650

[9] Adam R Kaiser, Karen Iler Kirk, Lorin Lachs, and David B Pisoni. 2003. Talker and lexical effects on audiovisual word recognition by adults with cochlear implants. *Journal of speech, language, and hearing research : JSLHR* 46, 2 (apr 2003), 390–404. https://doi.org/10.1044/1092-4388(2003/032)

[10] Shahid Latif and Fabian Beck. 2019. VIS Author Profiles: Interactive Descriptions of Publication Records Combining Text and Visualization. *IEEE Transactions on Visualization and Computer Graphics* 25, 1 (Jan. 2019), 152–161. https://doi.org/10.1109/TVCG.2018.2865022

[11] Shahid Latif, Diao Liu, and Fabian Beck. 2018. Exploring Interactive Linking Between Text and Visualization. In *EuroVis*.

[12] Dominic W Massaro and Joanna Light. 2004. Using visible speech to train perception and production of speech for individuals with hearing loss. *Journal of speech, language, and hearing research : JSLHR* 47, 2 (apr 2004), 304–320. https://doi.org/10.1044/1092-4388(2004/025)

[13] Teresa V Mitchell and Melissa T Maslin. 2007. How vision matters for individuals with hearing loss. *International journal of audiology* 46, 9 (sep 2007), 500–511. https://doi.org/10.1080/14992020701383050

[14] Lawrence D Rosenblum. 2008. Primacy of multimodal speech perception. In *The handbook of speech perception*. Blackwell Publishing, Malden, 51–78.

[15] Stephanie Rosenthal, Sai P. Selvaraj, and Manuela Veloso. 2016. Verbalization: Narration of Autonomous Robot Experience. In *Proceedings of the Twenty-Fifth International Joint Conference on Artificial Intelligence* (New York, New York, USA) *(IJCAI'16)*. AAAI Press, 862–868.