Towards a Rich Format for Closed Captioning

Lloyd May
lloyd@ccrma.stanford.edu
Stanford University
Stanford, California, USA

Alex Williams
a.williams@snhu.edu
Southern New Hampshire University
Manchester, New Hampshire, USA

Saad Hassan
saadhassan@tulane.edu
Tulane University
New Orleans, Louisiana, USA

Mark Cartwright
mc232@njit.edu
New Jersey Institute of Technology
Newark, New Jersey, USA

Sooyeon Lee
sooyeon.lee@njit.edu
New Jersey Institute of Technology
Newark, New Jersey, USA

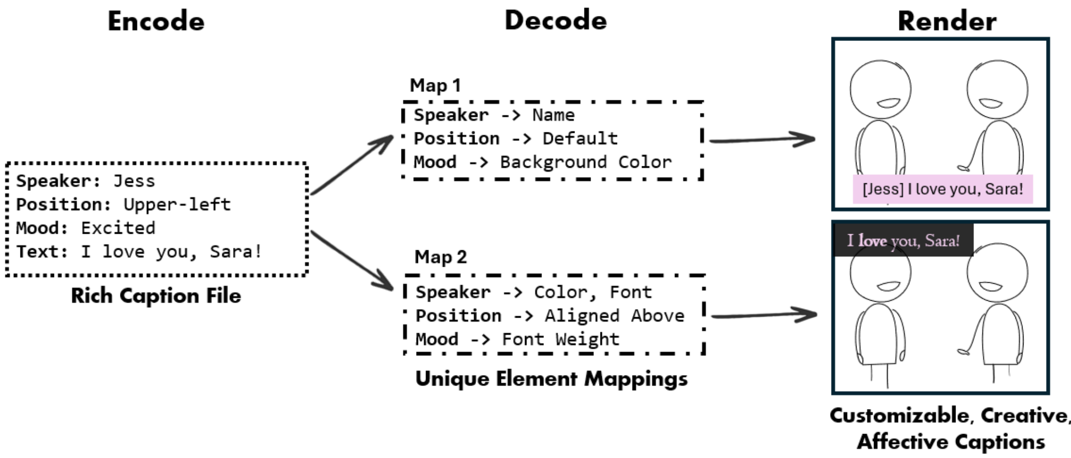


Figure 1: The proposed Rich Caption system that allows users to (1) create a single caption file with additional customizable parameters, (2) create multiple custom mappings between these parameters and characteristics of the caption, and (3) render the resultant captions onto the source video.

Abstract

Closed-captioning is an essential part of viewing audio-visual content for many people, including those who are D/deaf and Hard-of-Hearing. Traditional closed-captioning systems generally consist of a single track of timed text that offers limited options for personalization. Research into extending the capabilities of captioning, such as affective, poetic, and customizable captions has shown a desire among a subset of users for these features, but only in specific contexts. However, due to the difficulty in creating custom stimuli videos utilizing the custom captioning system, comparisons between systems and longitudinal studies have not been pursued. This demo paper introduces *Rich Captions*, a structured system that allows for a single closed-caption file to be tagged with additional information that can then be flexibly leveraged to render different customizable, creative, and poetic captions from the same file. Additionally, we introduce the *Rich Caption Editor* [1], a free, open-source software system designed to author, edit, and render rich captions. The system design was informed by a formative design workshop with closed-captioning researchers and advocates. The current design allows researchers to generate reproducible stimuli for closed-captioning studies. Once the design space and user preferences are better understood, the rich captioning framework could be refined to serve a general audience.

CCS Concepts

• **Human-centered computing → Accessibility systems and tools**; *Accessibility technologies.*

Keywords

Closed-Captioning, Affective Captioning, Creative Captioning, Subtitles

ACM Reference Format:
Lloyd May, Alex Williams, Saad Hassan, Mark Cartwright, and Sooyeon Lee. 2024. Towards a Rich Format for Closed Captioning. In *The 26th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '24), October 27–30, 2024, St. John's, NL, Canada.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3663548.3688504

ASSETS '24, October 27–30, 2024, St. John's, NL, Canada

ACM ISBN 979-8-4007-0677-6/24/10
https://doi.org/10.1145/3663548.3688504

[1]https://github.com/disjoylab/Rich-Caption-Editor

1 Introduction

Closed captioning is crucial for many, especially for D/deaf and Hard-of-Hearing (DHH) individuals, providing necessary access to audio information. However, captions are often viewed as a "one-size-fits-all" technology that affords little customization beyond changing basic static visual characteristics like size and color. Historically, closed captioning has also prioritized speech over non-speech information (NSI), such as background music or environmental sounds. NSI is often poorly captioned or omitted altogether in favor of nearly inaudible background speech [8, 16], despite clear guidelines [3]. NSI captions can deepen emotional connection, provide additional content, and communicate information critical for understanding video content [24]. A specific type of NSI, extra-speech information (ESI), indicates how speech is delivered (e.g., yelling, singing) and speaker identity, which is particularly important in multi-speaker scenes or where a speaker is off-screen or obscured. Previous research has examined customizable speaker identification in fast-paced settings where only captioning specific speakers may be advantageous [9]. However, effective communication of ESI, such as using italics for inflections, remains under-explored [2]. Communicating NSI through captioning has been explored by several artists and researchers. Previous work has highlighted the nuance present in NSI captioning by proposing a three-part model to understand NSI communication preferences [17] the selection (what NSI should be conveyed), curation (what information about the NSI should be conveyed), and communication (how should the curated information be conveyed). Sound artist Christine Sun Kim highlights the limitations and creative possibilities of captioning NSI, using descriptive and poetic language to enhance immersion [20, 21]. While Kim's poetic captions interrogated the role of NSI captions, they also raised practical questions of reading rate and overstimulation.

Previous work explored experimental captioning techniques to enhance NSI communication through modified captions and other communication strategies. Alonzo et al. used graphic captions and icons for user-generated content and found strong interest among DHH users for more customization options in captioning [1], while other studies have used color, animation, and icons within captions to convey emotion, volume, or speaker identity [4, 6, 10, 12, 14, 22]. NSI has also been communicated through sound visualizers [7, 17, 19], haptics [13, 15, 23], and off-screen projection [11]. While many researchers have explored this design space and produced promising initial results, there has been little work comparing these techniques or studying their use over a longer time period. A factor contributing to this is likely the difficulty present in replicating stimuli—custom, single-purpose software is often utilized to create the stimuli for these studies. Viewers cannot currently opt-in to use more experimental captioning techniques due to constraints of the current one-size-fits-all approach to captioning, likely contributing to the reluctance in widespread adoption.

In this demo paper, we outline the initial development of a *Rich Captions* system to author, edit, and render captions in a flexible, replicable, and customizable way. As illustrated in Figure 1, the system uses an encode -> decode -> render framework where caption files are encoded with additional structured information through abstract tags (i.e. *elements*). These abstract elements are then decoded according to a customizable mapping file that governs how the information captured by the element tags should be interpreted and displayed. Finally, the resultant captions are then rendered onto the desired video. In addition to proposing the rich caption system, we present the *Rich Caption Editor* (RCE), a free, open-source system to easily author, edit, and render videos with rich captions. Target users for the RCE are researchers, including technologists and hobbyists, who wish to explore rich-captioning. While eventually, we aim to open rich-captioning to a general audience, we believe the design space is not yet well explored to propose a well-defined, rigorously validated framework for end-users. Therefore, we aim for the RCE to be a tool for researchers to better understand the design space that would form stepping-stones to making a rich-captioning system available to a general audience.

2 System Design

2.1 Formative Study

To gain insights into designing a tool for authoring and rendering rich captions, we conducted preliminary design probe workshops. The core target user group of this tool is closed-captioning researchers and advocates. We recruited 9 total participants from the team's research network through email, including 1 Deaf technologist, 1 Deaf captioning advocate/evangelist, and 7 human-computer interaction (HCI) researchers from academic institutions. 4 participants self-identified as female and 5 as male.

The online workshop was held on three separate occasions to accommodate participants' schedules. Each session followed the same structure, lasted approximately 55 minutes, and had 2-4 participants present, with an American sign language interpreter present when requested. The session began with a brief explanation of the scope of the project and the current target group for the tool (see Appendix A.1), followed by introductions and a short ice-breaker activity. To establish a shared baseline of examples and prompt discussion, participants were then given access to a playlist of seven 20–35-second video clips with open captions from various *Netflix* shows (see Appendix A.2). The clips were selected to highlight scenarios that might require additional nuance in closed-captioning such as an off-screen speaker, a speaker imitating a celebrity, or multiple speakers wearing surgical masks speaking quickly. Participants were asked to take 5 minutes to view any number of the clips that might interest them and reflect on ways they might improve the captioning in the clip, assuming that no financial, technical, or time constraints are present. Participants then shared these reflections and entered a short facilitated discussion. We then conducted a design probe study where participants were asked to watch a 4-minute video overview of a possible system for authoring and rendering rich captions. Notably, this was framed as an example of a system rather than a prototype to encourage a wider scoping of participant feedback. The example system is described in detail in Appendix A.3. Participants were asked to reflect on the example system, share these reflections with the group, and then enter a longer facilitated discussion. To close each session, the main themes from the session were highlighted, with participants invited to correct or add to the highlighted themes, and participants completed a short exit survey.

After reflecting on the challenging captioning scenarios, participants indicated several contextual factors that informed their recommendations, such as the genre of the video, the aesthetic intention of the production team, as well as the cultural literacy assumed. The design probe outlined an initial system of adding fixed features taken from previous literature [1, 4, 16], such as valence, arousal, position, and model confidence, to traditional caption files (see Appendix A.3). Participants found the idea of rich captions interesting and exciting, particularly the encoder-decoder structure of the system. They noted that this structure would allow researchers to flexibly design, render, and compare creative and affective captioning solutions and that a more well-defined, distilled version of this system could be useful to a more general audience. Certain features that could be used to personalize what NSI is captioned, such as narrative importance and on-screen redundancy were noted as particularly interesting and under-studied. The option to provide features that could alter the lexical content of the captions, such as offering *simplified* and *full* options for dialogue or *poetic* and *sonic* options for NSI captioning, were also noted as particularly interesting. Additionally, two researchers were intrigued by the possibility of rich captions to include APIs that could be used to overtly control and cue external technologies such as a haptic feedback device or external lighting.

While acknowledging the opportunities presented by the example system, participants raised several important critiques of the example design. Most notably, participants raised many concerns over how the features included in the system were selected and implemented. For example, was the valence-arousal model of emotion, widely used in prior affective captioning work [1, 5] truly the most appropriate for all use cases, and, if so, should this feature be encoded on sentence, word, or phoneme level? It was clear from this feedback that participants desired increased flexibility to create their own features and specify the data structure of each feature, and that current literature does not currently provide sufficient justification for these decisions.

2.2 Rich Caption Editor

We gained several design insights from the formative study workshop and exploration of previous literature, namely: (1) affording feature and element customization, (2) allowing for the creation of text variations such as "simplified English", (3) offering support for elements to alter the lexical content of the captions. These findings were incorporated into the development of the *Rich Caption Editor* (RCE): an open-source tool to author and render rich captions for audio-visual media. Rich captions utilize an encoder-decoder framework where additional information is embedded within the caption file itself. This additional information is then combined with a user-definable mapping or preference file that is then interpreted by a decoder which renders the captions. This framework allows for a single rich-caption file that enables captioning researchers to more easily create and replicate stimuli, technologists, and creatives to explore more experimental approaches to captioning, and potentially one day allow end-users increased control over their captioning experience.

In the creation of the rich caption file structure, we adapted the WebVTT format [2] in an effort to make potential future integration easier. The software was developed in *Unity* and *C#*, and is publicly available here: https://github.com/disjoylab/Rich-Caption-Editor The file structure consists of five basic components, namely: (1) Element: an abstract information indicator such as mood, speaker, or emphasis. These do not directly affect how the captions are rendered but rather indicate information that may be useful when rendering the captions. (2) Setting: A concrete parameter that directly impacts the visual rendering of the caption, such as bold, font family, size, color, etc. (3) Feature: A collection of settings that can be mapped to an element. For example, we could define *valence* as an element that takes values between -1 and 1, and then define a feature *color gradient and bold* and would alter the *color* and *bold* settings of the text of the caption according to the value of valence. (4) Cue: Contains a single captioning event, such as a line of dialogue or a sound effect. Each cue must contain a start time and end time. The data structure of a cue is illustrated in 2b. (5) Cue-group: A higher-level collection of cues, such as "Music" or "Speech", allowing for features to be easily applied to a collection of cues.

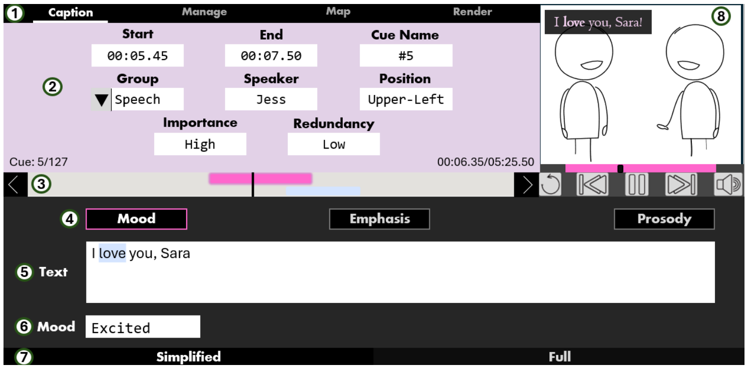
Elements are further subdivided by hierarchical position within the captioning system, with the most precise element taking rendering priority over more elements. Elements can apply to: (1) an entire cue-group, such as "Music" or "Sound Effects", (2) an entire cue such as "Speaker", or (3) to a single word, phoneme, or letter such as "Prosody". Finally, *Global Defaults* are defined as the default visual style of the captions. This results in a hierarchy of elements, namely: Character-Level Elements > Cue-Level Elements > Cue-Group-Level Elements > Global Defaults.

The RCE was designed with a flexible feature system that allows users to create and modify their own elements and features. The RCE, as illustrated in Figure 2a, is divided into four tabs by functionality, namely: The *Caption* editor screen allows users to edit the content and elements of cues, as well as preview a rendering of the current cue over the relevant video segment. The *Manage* tab is used to select which elements and features are available in the current project. This was implemented to reduce visual clutter and streamline the user experience. The *Map* tab is where users can construct features by creating a mapping of settings (visual display parameters) to specific elements (abstract tags in the caption). A suite of features can then be selected, grouped, and saved in a standardized format as a *Feature Map*. The *Render* is a full-screen view of the associated video file with the captions rendered according to the currently selected Feature Map. This can be used for screen-recording stimuli or enjoying the end result of the captioning process with no visual clutter.

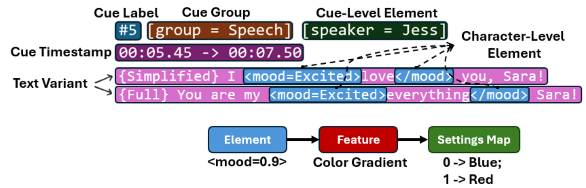
3 Discussion & Future Work

Prior literature [1, 4, 10] and the formative study indicate a desire among users and researchers for additional customization and expressive capabilities in captioning systems. Currently, creating stimuli for these studies is time-intensive, leading to difficulties in replication, comparison, and longitudinal data collection in this research space. Exploring the design space can prove challenging, as the number of possible display features, genres, and viewing settings leads to a combinatorial explosion of possibilities [4]. Therefore, a tool like the RCE may also be useful to create participatory design studies that allow participants to quickly iterate through

[2] https://www.w3.org/TR/webvtt1/



(a) The user interface of the Rich Caption Editor.



(b) Top: An illustrative example of the modified WebVTT formatting used to save the rich caption files. Bottom: The relationship between abstracted elements in the text, which connect to a collection of visual display settings grouped together under a single feature.

Figure 2: (A)The user interface of the Rich Caption Editor consisting of: (1) Tab selection to navigate between various features of the RCE, (2) the cue-level element editing environment, (3) scroll-able medium-scale timeline with cue timing and duration indicated by colored rectangles, (4) the element tool selector, allowing users to edit character-level elements, (5) the lexical content of the cue, (6) the value of the selected character-level element, in this case "love" has a "mood" value of Excited, (7) text variation selection, as different text variations are able to have unique character-level elements, and (8) the preview window that displays the current cue according to the currently selected mapping. The preview window displays the section of the video currently being captioned and can be looped, paused, or used to navigate to the next caption chronologically.

design possibilities. The majority of prior work has focused solely on one facet of captioning in isolation, such as affective captions or NSI, and the RCE may be useful for combining these areas of inquiry. Additionally, the genre and viewing context, such as action movies on a large TV versus short-form social media on a smartphone, introduce additional unique considerations [18], and the RCE could help explore this space. Given the diverse needs of researchers exploring this large design space, the need to prioritize flexibility in defining elements, mappings, and features over premature standardization was clear. While this does introduce additional system complexity, the ability to explore these varied research questions using an easily replicable system appeared to be vital in the pursuit of supporting and unifying efforts to design and evaluate expressive, customizable captioning systems.

Future work will include an evaluation study and further refinement of the RCE to ensure it is meeting the needs of researchers. Following this, we aim to create stimuli for a variety of longitudinal and comparative studies of creative captioning techniques, as well as use the RCE as a pedagogical tool in accessible technology design classes. Additionally, given the advances in machine learning models for automatic captioning, sound event classification, and multi-modal video understanding, future work could explore how the rich captioning format could be used as a framework for models to structure their outputs and provide additional outputs such as model confidence or nearest-neighbor results. Ultimately, the end goal of this work is to refine and standardize the rich-captioning system and make a system for expressive, customizable captions available to a general audience.

References

[1] Oliver Alonzo, Hijung Valentina Shin, and Dingzeyu Li. 2022. Beyond Subtitles: Captioning and Visualizing Non-speech Sounds to Improve Accessibility of User-Generated Videos. In *Proceedings of the 24th International ACM SIGACCESS Conference on Computers and Accessibility*. 1–12.

[2] Akhter Al Amin, Abraham Glasser, Raja Kushalnagar, Christian Vogler, and Matt Huenerfauth. 2021. Preferences of deaf or hard of hearing users for live-TV caption appearance. In *Universal Access in Human-Computer Interaction. Access to Media, Learning and Assistive Environments: 15th International Conference, UAHCI 2021, Held as Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings, Part II*. Springer, 189–201.

[3] Ben Caldwell, Michael Cooper, Loretta Guarino Reid, Gregg Vanderheiden, Wendy Chisholm, John Slatin, and Jason White. 2008. Web content accessibility guidelines (WCAG) 2.0. *WWW Consortium (W3C)* 290 (2008), 1–34.

[4] Caluã de Lacerda Pataca, Saad Hassan, Nathan Tinker, Roshan Lalintha Peiris, and Matt Huenerfauth. 2024. Caption Royale: Exploring the Design Space of Affective Captions from the Perspective of Deaf and Hard-of-Hearing Individuals. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–17.

[5] Caluã de Lacerda Pataca, Matthew Watkins, Roshan Peiris, Sooyeon Lee, and Matt Huenerfauth. 2023. Visualization of Speech Prosody and Emotion in Captions: Accessibility for Deaf and Hard-of-Hearing Users. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–15.

[6] Caluã de Lacerda Pataca, Matthew Watkins, Roshan Peiris, Sooyeon Lee, and Matt Huenerfauth. 2023. Visualization of Speech Prosody and Emotion in Captions: Accessibility For Deaf And Hard-of-Hearing Users. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 831, 15 pages. https://doi.org/10.1145/3544548.3581511

[7] Jordan Aiko Deja, Alexczar Dela Torre, Hans Joshua Lee, Jose Florencio Ciriaco IV, and Carlo Miguel Eroles. 2020. Vitune: A visualizer tool to allow the deaf and hard of hearing to see music with their eyes. In *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–8.

[8] Gregory J Downey. 2008. *Closed captioning: Subtitling, stenography, and the digital convergence of text with television*. JHU Press.

[9] Benjamin M Gorman, Michael Crabb, and Michael Armstrong. 2021. Adaptive Subtitles: Preferences and Trade-Offs in Real-Time Media Adaption. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–11.

[10] Saad Hassan, Yao Ding, Agneya Abhimanyu Kerure, Christi Miller, John Burnett, Emily Biondo, and Brenden Gilbert. 2023. Exploring the Design Space of Automatically Generated Emotive Captions for Deaf or Hard of Hearing Users. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 125, 10 pages. https://doi.org/10.1145/3544549.3585880

[11] Brett R Jones, Hrvoje Benko, Eyal Ofek, and Andrew D Wilson. 2013. IllumiRoom: peripheral projected illusions for interactive experiences. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 869–878.

[12] JooYeong Kim, SooYeon Ahn, and Jin-Hyuk Hong. 2023. Visible Nuances: A Caption System to Visualize Paralinguistic Speech Cues for Deaf and Hard-of-Hearing Individuals. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing

Machinery, New York, NY, USA, Article 54, 15 pages. https://doi.org/10.1145/3544548.3581130

[13] Raja S Kushalnagar, Gary W Behm, Joseph S Stanislow, and Vasu Gupta. 2014. Enhancing caption accessibility through simultaneous multimodal information: visual-tactile captions. In *Proceedings of the 16th international ACM SIGACCESS conference on Computers & accessibility*. 185–192.

[14] Daniel G Lee, Deborah I Fels, and John Patrick Udo. 2007. Emotive captioning. *Computers in Entertainment (CIE)* 5, 2 (2007), 11.

[15] Lloyd May, Sarah Miller, Sehuam Bakri, Lorna C Quandt, and Melissa Malzkuhn. 2023. Designing Access in Sound Art Exhibitions: Centering Deaf Experiences in Musical Thinking. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 380, 8 pages. https://doi.org/10.1145/3544549.3573872

[16] Lloyd May, Keita Ohshiro, Khang Dang, Sripathi Sridhar, Jhanvi Pai, Magdalena Fuentes, Sooyeon Lee, and Mark Cartwright. 2024. Unspoken Sound: Identifying Trends in Non-Speech Audio Captioning on YouTube. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–19.

[17] Lloyd May, So Yeon Park, and Jonathan Berger. 2023. Enhancing Non-Speech Information Communicated in Closed Captioning Through Critical Design. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility* (New York, NY, USA) *(ASSETS '23)*. Association for Computing Machinery, New York, NY, USA, Article 16, 14 pages. https://doi.org/10.1145/3597638.3608398

[18] Emma J McDonnell, Tessa Eagle, Pitch Sinlapanuntakul, Soo Hyun Moon, Kathryn E Ringland, Jon E Froehlich, and Leah Findlater. 2024. "Caption It in an Accessible Way That Is Also Enjoyable": Characterizing User-Driven Captioning Practices on TikTok. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–16.

[19] Suranga Chandima Nanayakkara, Lonce Wyse, Sim Heng Ong, and Elizabeth A Taylor. 2013. Enhancing musical experience for the hearing-impaired using visual and haptic displays. *Human–Computer Interaction* 28, 2 (2013), 115–160.

[20] Christine Sun Kim. 2020. Artist Christine Sun Kim rewrites closed captions | pop-up Magazine. https://www.youtube.com/watch?v=tfe479qL8hg

[21] Niels Van Tomme and Christine Sun Kim. 2021. Activating Captions, ARGOS. https://www.argosarts.org/activatingcaptions/info

[22] Fangzhou Wang, Hidehisa Nagano, Kunio Kashino, and Takeo Igarashi. 2016. Visualizing video sounds with sound word animation to enrich user experience. *IEEE Transactions on Multimedia* 19, 2 (2016), 418–429.

[23] Yiwen Wang, Ziming Li, Pratheep Kumar Chelladurai, Wendy Dannels, Tae Oh, and Roshan L Peiris. 2023. Haptic-Captioning: Using Audio-Haptic Interfaces to Enhance Speaker Indication in Real-Time Captions for Deaf and Hard-of-Hearing Viewers. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–14.

[24] Sean Zdenek. 2015. Reading sounds. In *Reading Sounds*. University of Chicago Press.

A Workshop Materials

A.1 Workshop Introduction

The following was read to all participants at the beginning of each workshop:

"Welcome everyone, thank you for joining us today. Currently, captioning is one size fits all, and that does not appropriately meet the needs of many users. Closed-captioning can be extremely creative and expressive, yet captioners are forced into best practices because the current format does not allow users to opt-in to more poetic or expressive captions. In this workshop, we'll be exploring the topic of "Rich Captioning" technologies. Rich Captioning includes affective captioning, which includes things like modulated typography, where aspects of the text such as size, font weight, or vertical position, are changed in response to features like the emotion in the speaker's voice. Rich Captioning additionally includes customizable captions, where viewers can communicate their captioning preferences, such as which speakers should be captioned or color assigned to them. Currently, amazing affective and expressive captioning research has been done, but currently meaningfully comparing prototypes often means reconstructing previously published prototypes. This project aims to create an open-source, freely available rich captioning format and an authoring/editing tool that will make all of this easier. The tool is currently aimed at HCI researchers and power users, and future work will look more at the end-user experience. Note, that the current project uses a 'kitchen-sink' approach of including all features we might possibly want. Future work will whittle this down into a more mass-user-friendly state."

A.2 Short Clip Activity

The clips were taken from *Breaking Bad* season 5 episode 11 (3:00 onwards), *The Standups* season 3 episode 6 (1:30 onwards), and *New Amsterdam* season 1 episode 5 "Cavitation" (19:00 onwards). Note that the *New Amsterdam* clip was edited to remove graphic depictions of surgery that may have caused some participants unnecessary stress. The clips are available for reference at the time of publication here: https://www.youtube.com/playlist?list=PL1Sc0ESlWaVVLeBwAS7szUYs-6sD0Bpg7.

A.3 Design Probe Activity

The following was read to all participants at the beginning of the design probe activity: "This is an example design of a tool that can help author and edit customizable captions, and render them according to custom mappings and user preferences."

A subset of the following questions was asked during the following-up discussion, depending on themes that participants previously mentioned.

- What are some features you would want to see in a tool like this?
- What features/data should be encoded in the caption?
- What rendering techniques or tools should be included to interpret the data/features?
- What features of the software editing/authoring tool itself would you like to see?
- Would you use a tool like this?

The design probe video is available at the time of publication here: https://youtu.be/q4xzXsOGYVY