Co-Designing Systems to Support Blind and Low Vision Audio Description Writers

Lucy Jiang
lucjia@cs.washington.edu
Paul G. Allen School of Computer Science & Engineering,
University of Washington
Seattle, WA, USA

Richard Ladner
ladner@cs.washington.edu
Paul G. Allen School of Computer Science & Engineering,
University of Washington
Seattle, WA, USA

ABSTRACT

Audio description (AD), an additional narration track that conveys visual information in media, improves video accessibility for blind or low vision (BLV) viewers. Despite being the primary beneficiaries of AD, BLV audiences are limited in how they can contribute to the AD writing process due to technology inaccessibility and societal biases. In this poster, we (1) prototype and test AccessibleAD, an accessible AD writing system, (2) analyze what context and features are valued by BLV description writers, and (3) explore nonvisual involvement in AD creation. This work expands on existing literature regarding audio description and explores best practices for expanding access to AD writing.

CCS CONCEPTS

• **Human-centered computing** → **Accessibility**.

KEYWORDS

audio description, video accessibility

ACM Reference Format:
Lucy Jiang and Richard Ladner. 2022. Co-Designing Systems to Support Blind and Low Vision Audio Description Writers. In *The 24th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '22), October 23–26, 2022, Athens, Greece.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3517428.3550394

1 INTRODUCTION

For over 300 million blind and low vision (BLV) people across the world [2], audio descriptions are a critical extension to visual storytelling. Audio description (AD) is "the descriptive narration of key visual elements of live theater, television, movies, and other media" [9], but only a small fraction of all digital content is described [8]. AD is traditionally written by sighted professionals. However, their priorities may not always align with those of BLV audiences, who wish to hear more about characters' races, costumes, and disabilities [12]. Disabled people are also frustratingly subjected to sanitized experiences by sighted writers, which can take the form of PG-rated descriptions for R-rated scenes [13].

ASSETS '22, October 23–26, 2022, Athens, Greece

ACM ISBN 978-1-4503-9258-7/22/10.
https://doi.org/10.1145/3517428.3550394

Although AD informed by BLV perspectives is of a higher quality and better matches the needs of BLV audiences [4], BLV people are often denied equal access to AD employment opportunities [5]. This industry must be accessible so that BLV individuals can have the same opportunities as sighted people to participate in creating AD. Access can be achieved through creating technology that supports the needs of BLV writers and eliminating societal stigmas. To explore alternate methods to creating AD, we: (1) designed an accessible AD writing prototype and (2) evaluated the prototype with six BLV individuals interested in writing AD.

2 RELATED WORK

Audio description is an emerging research area within HCI and accessibility. Yuksel et al. reported that a human-in-the-loop machine learning approach reduced barriers for creating AD [14]. ViScene, a collaborative AD writing tool used by BLV and sighted reviewers, decreased costs for creating non-professional AD [7]. However, these works targeted sighted writers of AD scripts.

In moving towards automated AD, Campos et al. utilized pre-existing movie scripts to improve computer-generated AD [1]. Another line of work involved providing BLV people with additional information about a video to augment existing AD. Ihorn et al. developed two AI-driven tools, NarrationBot and InfoBot, to generate baseline and on-demand AD. For on-demand AD, the bot used visual question answering (VQA) systems with and without humans-in-the-loop [3]. However, this research did not aim to help BLV users produce the AD themselves.

This poster makes two primary contributions: (1) a prototype of a system, AccessibleAD, that improves access to AD writing and (2) insights regarding key elements of BLV-written descriptions.

3 METHODOLOGY

3.1 Prototype Design

We began our research process by interviewing two experienced BLV audio describers to inform the design of an accessible platform. We determined that any AD system must: (1) be accessible to BLV users, (2) support targeted navigation throughout a video, and (3) provide context regarding a video's visuals, as AD writing is grounded by semi-objective visual observations.

We developed AccessibleAD, a semi-functional web-based platform designed to streamline the AD writing process. All features were created by the first author to facilitate Wizard of Oz experiments. The first author wrote baseline AD (objective, minimally detailed information about key visual content) for each clip, and

descriptions were recorded with IBM's Watson Text to Speech engine to simulate the experience of receiving computer-generated AD. In accordance with industry conventions, we only wrote AD for areas in the video without dialogue.

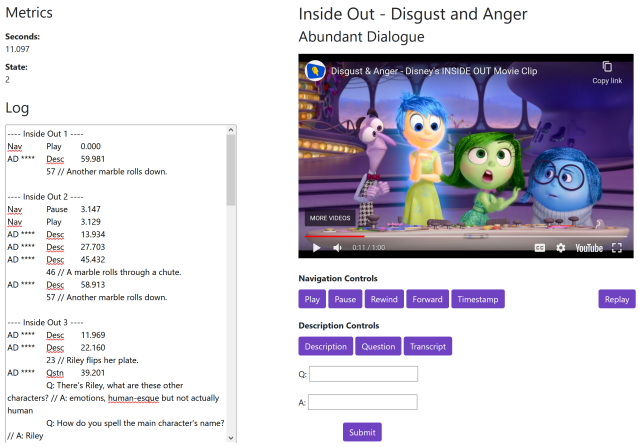

Figure 1: Screenshot of the AccessibleAD prototype. Features include navigation controls (play, pause, rewind, forward, timestamp, and replay) and description controls (description, question, and transcript). The log panel on the left side of the screen tracked all user actions during the study.

3.2 Prototype Evaluation

We evaluated our prototype with six BLV people interested in writing AD, including the two AD writers from the preliminary portion of the study. We recruited participants through the Audio Description Twitter community and snowball sampling. All participants were over 18 years of age and were compensated for their time.

The remote study consisted of three parts. First, we asked participants about their AD preferences and writing experience. Second, we shared our screen and asked participants to describe two videos using AccessibleAD. Participants controlled the prototype either through voice commands or keyboard commands sent through the "chat" feature on the meeting platform, and were given 15 minutes to describe each clip. We emphasized that they did not need to finish writing their AD nor fix typos or grammar. For participants who were unfamiliar with the clips, defined as rating their familiarity at 5/10 or lower, we provided the movie's IMDb synopsis for context. We observed and logged participants' interactions and experiences with the prototype, and evaluated the efficacy of three key features: (1) closed captions, (2) baseline descriptions (main objects, people, spatial relations, interactions, movement, on-screen text, etc.), and (3) on-demand descriptions (accessed by asking quantifiable or yes/no questions at any point in the video). Third, the interview closed with questions about participants' thoughts on their AD writing process, the usability of AccessibleAD, and combating stigmas against BLV description writers.

In each study, we used two Disney Pixar clips: Inside Out - Disgust and Anger [11] and Ratatouille - Remy Fixes the Soup [10]. Due to disparities in the amount of dialogue in the videos, we wrote four baseline descriptions for Inside Out and ten for Ratatouille.

4 FINDINGS

4.1 Context Required for AD Writing

Four of the six participants sought more detail about specific elements in each video. Half of the participants (N = 3) asked for more information about character identities, such as a character's race, age, expressions, and body language. Participants were also interested in knowing more about character actions. In Ratatouille, a majority of participants (N = 4) asked for further details regarding the character's actions. Some participants (N = 2) asked about actions that they missed during a lengthy description gap, while others sought to gather more context about actions hinted at in the pre-generated AD.

Two participants mentioned that information about the setting or location was important to them as well. After one pass of the pre-generated AD for Ratatouille, which ends with a mention of the lights turning on, both P3 and P1 inquired about the setting of the scene. Lastly, participants asked questions to clarify audio cues, building on sound effects and tone of speech to understand the full meaning of a scene. In the Inside Out clip, notable sounds included marbles clinking and Riley's temper tantrum.

4.2 Accessible Features

Pre-generated baseline AD tracks are effective ways to provide BLV writers with additional context about a video's visuals. During the AD task, most interviewees (N = 5) requested to listen to all of the baseline AD to understand a video's context; the only participant who did not do so reported familiarity with both movies. However, baseline AD does not provide full access. When describing what information was important in an AD script, P5 stated: *"I just want full access to what someone… just because they have a functioning pair of eyeballs, has access to."* VQA support must be integrated into AD writing systems to build on context afforded by audio cues and baseline AD. Additionally, while two participants specifically liked the voice control interface, two others wished for additional input methods such as keyboard shortcuts. Having multiple input methods and control mechanisms enables participants to choose their preferred mode of interaction, which can improve both efficiency and accessibility.

4.3 Quantitative Feedback

Our findings signal how accessible AD writing systems can improve the experiences of BLV writers. Participants gave an average rating of 5.42/10 for their satisfaction with their descriptions for Inside Out. Two participants cited the abundant dialogue of the clip as a challenge. As Inside Out was presented first, the moderate learning curve of the system could have interfered with participants' writing efforts and reflected in their ratings. Participants rated their satisfaction with their Ratatouille descriptions at 6.92/10 on average, and all rated Ratatouille the same or higher than Inside Out. Participants gave an overall satisfaction rating of 6.58/10 on average. Two interviewees found the task to be difficult given the time constraints and their lack of experience with writing AD, but all (N = 6) liked the overall system design and its intuitive features.

4.4 Stigma

All participants (N = 6) also expressed their staunch support for increasing the involvement of BLV creatives in AD production pipelines. P3 thought writing AD would be a positive way for him to contribute to his community. Despite recognizing great value in technological augmentations of AD, P2 also remarked that changing the societal perception of BLV writers is just as critical. P5 advocated for the increased agency of BLV writers in AD workflows, and expressed her frustration with existing AD scripts. Despite the good intentions of sighted AD writers, she mentioned how *"there's a real, big kind of historical problem where… [disabled peoples'] experiences need to be sanitized"* (P5), leaving BLV audiences with unequal information about gory or otherwise explicit scenes that sighted viewers can access. P4, who has been using AD since the early 1990s, shared a similar viewpoint.

> *"Blind people can author audio description scripts. Adaptations are required, but it's no different than modifications which allow people with disabilities to accomplish all manner of tasks and jobs… we should be filling these roles given that we know best what blind people want in AD."* (P4)

5 DISCUSSION

5.1 Design Considerations

Through our research, we identified several insights to guide future work on AD creation tools. While pre-generated descriptions can be helpful, they may also contain incorrect information. Reporting a confidence level alongside automatically generated descriptions can signal their approximate accuracy. Additionally, studios which hope to increase AD quantity but not quality may misuse these technologies as "weapons for compliance" [3]. For example, BLV audiences believe using text-to-speech technology instead of human voice talents is jarring and unenjoyable for entertainment content [6]. While sighted writers may not need pre-generated AD or VQA systems, adaptations for accessing videos in an alternative way can be beneficial for any describer to identify key visual elements or split a video into more manageable segments. Introducing automation into the AD industry can be incredibly valuable as long as creators prioritize quality and the needs of the BLV community.

5.2 Limitations

We operated the prototype in a Wizard of Oz fashion based on participants' voice and keyboard commands. Participants did not have as much flexibility as they could have had when using the prototype on their own, which may have negatively impacted their satisfaction with the system. As both the baseline AD and the VQA were provided by a human rather than a computer, the detailed descriptions and answers to participants' questions are unrepresentative of what is currently possible by state-of-the-art VQA systems. Although the technical capacity to automatically generate a baseline description is not yet realized, this work begins to understand BLV users' preferences to inform future VQA development. Lastly, the small number of participants (N = 6) included in this study limits the generalizability of the results.

6 CONCLUSION

Despite being critical to providing video access for BLV audiences, a vast majority of today's video content lacks audio description. BLV writers are limited in the ways they can participate and contribute to current AD pipelines due to harmful stigmas and inaccessible AD writing technology. This poster introduces the application of VQA in creating AD, uncovers perspectives from the BLV community regarding the context and features that are important to them when writing AD, and explores how advancing technology can reduce stigmas and further disability inclusion. Next steps include deploying a system that can be navigated by BLV users on their own devices and integrating automated VQA systems to fully explore independent writing possibilities. AD has proliferated and grown to become a major industry which must respect, empower, uplift, and employ BLV creatives in the AD creation process to achieve full parity, equality, and excellence. This work extends previous AD and accessibility research to provide new insights into co-designing audio description technology and to push for societal change.

ACKNOWLEDGMENTS

This research was supported by the University of Washington Center for Research and Education on Accessible Technology and Experiences (UW CREATE).

REFERENCES

[1] Virginia P. Campos, Tiago M. U. de Araújo, Guido L. de Souza Filho, and Luiz M. G. Gonçalves. 2020. CineAD: a system for automated audio description script generation for the visually impaired. 19 (2020), 99–111. https://doi.org/10.1007/s10209-018-0634-4

[2] The International Agency for the Prevention of Blindness (IAPB). 2022. *Magnitude and Projections.* https://www.iapb.org/learn/vision-atlas/magnitude-and-projections/

[3] Shasta Ilhorn, Yue-Ting Siu, Aditya Bodi, Lothar Narins, Jose M. Castanon, Yash Kant, Abhishek Das, Ilmi Yoon, and Pooyan Fazli. 2022. NarrationBot and InfoBot: A Hybrid System for Automated Video Description. (2022). https://doi.org/10.48550/arXiv.2111.03994

[4] Robert Kingett. 2021. *Adventure Beasts and Audio Description Verbs.* https://blindjournalist.wordpress.com/2021/10/22/adventure-beasts-and-audio-description-verbs/

[5] Ren Leach. 2022. *Post on Twitter.* https://twitter.com/renleach/status/1524379908373303296

[6] Byron Lee. 2020. *Post in the Audio Description Discussion Facebook Group.* https://www.facebook.com/groups/AudioDescriptionDiscussion/permalink/1435708533242756/

[7] Rosiana Natalie, Jolene Loh, Huei Suen Tan, Joshua Tseng, Ian L.Y. Chan, Ebrima H. Jarjue, Hernisa Kacorri, and Kotaro Hara. 2021. The Efficacy of Collaborative Authoring of Video Scene Descriptions. (2021). https://doi.org/doi/abs/10.1145/3441852.3471201

[8] The American Council of the Blind. 2022. *ADP Master List of Audio Described Videos.* https://adp.acb.org/masterad.html

[9] The American Council of the Blind. 2022. *The Audio Description Project.* https://adp.acb.org/

[10] Pixar Animation Studios. 2010. *Ratatouille Cooking Scene.* https://www.youtube.com/watch?v=jwLKPDJqldw

[11] Pixar Animation Studios. 2015. *Disgust & Anger - Disney's INSIDE OUT Movie Clip.* https://www.youtube.com/watch?v=AQ3hjymiCCg

[12] Salamishah Tillet. 2021. *'Bridgerton' Takes On Race. But Its Core Is Escapism.* https://www.nytimes.com/2021/01/05/arts/television/bridgerton-race-netflix.html

[13] Robbie Whelan. 2022. *'Bridgerton' Is About to Get Saucier.* https://www.wsj.com/articles/bridgerton-superfans-embrace-audio-option-that-narrates-steamy-on-screen-action-11648223396

[14] Beste F. Yuksel, Pooyan Fazli, Umang Mathur, Vaishali Bisht, Soo Jung Kim, Joshua Junhee Lee, Sueng Jung Jin, Yue-Ting Siu, Joshua A. Miele, and Ilmi Yoon. 2020. Human-in-the-Loop Machine Learning to Increase Video Accessibility for Visually Impaired and Blind Users. (2020), 47–60. https://doi.org/doi/abs/10.1145/3441852.3471201