VerbalEyes: A Large-Scale Inquiry into the State of Audio Description

Lucy Jiang*
University of Washington
Seattle, United States
lucjia@cs.washington.edu

Daniel Zhu*
University of Washington
Seattle, United States
dzhu99@cs.washington.edu

ABSTRACT

Audio description (AD), the spoken narration of a video's key visual elements, improves video accessibility for blind or visually impaired viewers. Current processes for incorporating AD are manual and expensive, preventing the widespread adoption of audio description in mainstream media. We conducted user research on preferences for AD within the blind and visually impaired (BVI) community, surveying 107 BVI individuals and interviewing 43 subject-matter experts. We looked for (1) *when* they use AD, (2) *how* they use it, and (3) *on which platforms* they use it. Our main focus was to uncover *what they value in a high quality audio description experience*, including a range of user preferences for brevity, voice, and audio mixing. Our findings show that the most prominent challenge is the lack of available AD. To advance toward ubiquitous AD, we tested the usability of a tool that we developed to automatically describe videos, which we call VerbalEyes. Through VerbalEyes, we expand on knowledge of AD preferences and propose a solution to provide automatic audio descriptions based on novel user insights.

CCS CONCEPTS

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

KEYWORDS

Accessibility; blind and low vision; video; audio descriptions

ACM Reference Format:
Lucy Jiang and Daniel Zhu. 2021. VerbalEyes: A Large-Scale Inquiry into the State of Audio Description. In *Tapia 2021: CMD-IT/ACM Richard Tapia Celebration of Diversity in Computing Conference.* ACM, New York, NY, USA, 2 pages.

1 INTRODUCTION

Without adequate audio cues, digital video content is inaccessible for the 285 million people across the world who are blind or visually impaired[1]. Audio description, a secondary narration track describing essential visual content in videos, empowers people with visual impairments to access and understand videos. However, readily accessible, low-cost audio description is not available where people with visual impairments need it most – in learning environments, on user-generated content, and in on-demand video platforms.

We conducted a large-scale inquiry into the state of audio description, combining the perspectives of AD users and entities seeking to provide AD. Among the 111 BVI individuals who we surveyed and interviewed, there was universal frustration with the inaccessibility of digital video content. 92 survey respondents emphasized that they have a good experience with AD, but that it is not offered widely enough. Traditional AD vendors charge an expensive $9-15+ per video minute, with a slow turnaround time of 4-9 business days, preventing audio description from becoming as ubiquitous as closed captioning. The current AD creation process is in desperate need for innovation to match the growing scale of digital video content.

Our work establishes users' preferences for AD to inform the design of automatic methods for creating high quality audio description. Through two rounds of interviews involving contextual inquiries, usability tests, and general questions about participants' lived experiences, we follow the principles of inclusive design to determine preferences of end users within the BVI community. We evaluate their satisfaction with AI-generated audio descriptions created with VerbalEyes, a tool we developed for automatically creating AD. Our contributions in this poster are threefold:

(1) insights on user preferences for audio description,
(2) an analysis of AD in industry and higher education, and
(3) the automation of audio description technology.

2 BACKGROUND AND RELATED WORK

Prior work has attempted to improve the accessibility of AD by developing easier audio description writing interfaces [2], creating semi-automatically generated AD for movies [1], and incorporating human-in-the-loop approaches for writing AD [3]. Pavel et al. [2] introduced Rescribe, a tool to help authors create and refine audio descriptions, and proposed the idea of extended-inline AD. Campos et al. [1] developed a system to automatically generate audio descriptions for movies based on the original script and subtitles, and Yuksel et al. [3] reported that a human-in-the-loop machine learning approach was effective at reducing barriers for creating AD. These works focus on the partial automation of AD creation and perform limited user research regarding AD. There is still significant ambiguity about BVI individuals' specific preferences for different AD attributes; we contribute a deeper qualitative analysis of what users value in a high quality AD experience. Our work is the first large-scale inquiry into the audio description landscape by synthesizing the perspectives of BVI individuals, industry accessibility experts, and higher educational institutions.

* Equal contribution.

Tapia 2021, September 15 - 17, 2021, Virtual

[1] https://www.who.int/blindness/publications/globaldata/en/

3 METHODOLOGY

To understand the current state of audio description, we surveyed and interviewed 3 main groups of experts: (1) individuals from the BVI community, (2) industry accessibility experts, and (3) accessibility groups at higher educational institutions.

3.1 Surveys

We began by releasing a survey on various online platforms, including a forum on AppleVis (an online resource for BVI users of Apple products) and the Audio Description Discussion Facebook Group. We explored questions on internet usage, challenges with browsing the internet, screen reader usage, audio description usage, experiences with AD and suggested improvements, and envisioned platforms on which users wanted to access automatic AD. In total, we received 119 responses, including 107 from BVI individuals.

3.2 Semi-Structured Interviews

We interviewed 27 BVI individuals who were recruited from our survey respondents. All participants were volunteers between the ages of 18 and 65 and were from the United States, Denmark, Australia, and Estonia. Interviews were conducted virtually on Zoom and participants will be compensated for their time via a grant received from UW CREATE. So far, we have conducted two rounds of interviews with BVI participants:

(1) **First round (*N*=17)**: Our initial user interview strategy focused on learning about BVI users' lived experiences with AD and contextualizing their survey responses.
(2) **Second round (*N*=10)**: To build upon the previous round of interviews, we conducted a contextual inquiry with users to identify challenges in their video watching processes, and shared early demo videos[1] with participants to garner reactions, feedback, and areas for improvement.

We also interviewed 10 industry accessibility experts from Microsoft, Google, Facebook, and other companies. Our interviews aimed to understand the extent to which each company works on AD, how they create AD, and the impact of AD on their user base. We found that most companies have not prioritized providing widespread AD.

Lastly, we interviewed 7 representatives from accessibility groups at the University of Washington, Whatcom College, Bellevue College, and the Washington State School for the Blind. Our interviews uncovered issues with current AD processes and the pressing need for accessible and instant audio description in academic contexts.

4 RESULTS AND DISCUSSION

During our first-round interviews, we found that users have drastically different preferences on the level of detail of AD. Around 33% of participants preferred as little AD as possible to gain an objective understanding and prevent cognitive overload. Few wanted extremely detailed descriptions, and most favored something in the middle. A majority of participants preferred audio ducking (the practice of lowering background volume when AD is being read), and all did not want AD to interrupt or overlap with dialogue.

Participants emphasized the importance of having a low barrier to accessing and saving user settings. They also favored an AD tool that spanned across multiple forms of technology, such as websites, browser extensions, or apps, but acknowledged that they preferred using their mobile devices for watching videos.

Video Accessibility Pain Points	**Number of Respondents**
No AD available	28
Inaccessible video interface	25
Audio Description Experience	**Number of Respondents**
Not enough AD	92
Quality of AD is lacking	16
Bad volume mixing	14
Top Video Genres Needing AD	**Number of Respondent**s
How-to / Tutorial	35
Videos with no narration	22
Movies and TV shows	21

Table 1: Key insights from our survey on video accessibility; listing top responses where the number of respondents > 10.

In a second round of interviews, we identified that on popular video platforms such as YouTube and Instagram, advertisements can make the interface difficult to navigate. Not knowing which videos are described can also deter BVI individuals from using a platform. When we asked interviewees to rank their level of understanding and satisfaction upon their first exposure to a demo video, they gave a rating of 7.8/10 on average. Other key takeaways from user feedback involved lowering the original audio track for ease of comprehension and differentiating between AD and optical character recognition.

5 CONCLUSION

Our findings uncover the variety of AD preferences for description brevity, voice, and audio mixing, which prompt prominent design considerations for automatic audio description technology. To accommodate varying preferences for brevity, we propose presenting an accessible slider to allow users to adjust the level of description on a range spanning minimal, medium, and maximal, while setting the default at medium. We also suggest producing a separate audio track and accompanying script as opposed to integrating the AD into the original video, as this enables flexibility with volume settings.

Regarding next steps, we will continue to experiment with a two-pronged approach for evaluating AD serviceability, testing both an individual's overall understanding and their ability to recall specific video details. We will use these metrics and insights to calibrate our AD technology to better align with users' expectations for high quality audio description. We plan to extend our novel user research findings to the development of VerbalEyes, an AI-driven audio description technology that is accessible, scalable, and easy to use.

REFERENCES

[1] Virginia Campos, Tiago Araújo, Guido Souza Filho, and Luiz Gonçalves. 2020. CineAD: a system for automated audio description script generation for the visually impaired. *Universal Access in the Information Society* 19 (03 2020). https://doi.org/10.1007/s10209-018-0634-4

[2] Amy Pavel, Gabriel Reyes, and Jeffrey Bigham. 2020. Rescribe: Authoring and Automatically Editing Audio Descriptions. https://arxiv.org/pdf/2010.03667.pdf

[3] Beste F. Yuksel, Pooyan Fazli, Umang Mathur, Vaishali Bisht, Soo Jung Kim, Joshua Junhee Lee, Seung 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. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference (DIS '20)*. Association for Computing Machinery, New York, NY, USA, 47–60. https://doi.org/10.1145/3357236.3395433

[1] https://www.youtube.com/playlist?list=PLg-YuFNawnltIgIU68Y3LUIbhoE1D5lGP