Leveraging Spatial Interactions to Enhance BLV Users' Navigation of Virtual Environments

Project Progress Report

By: Anna-Samsara Daefler, Demetria Mack (Group name: SpaceInt)
avd2131 , dmm2285

1. Overview

   Our project aims to extend the technology developed by the Computer-Enabled Abilities Laboratory (CEAL) based at Columbia University which supplements desktop screen readers for blind and low-vision (BLV) users with spatial interactions - specifically, directional input and spatial sound output[1,2]. The code for CEAL's screen reader (WebNExt) is open source and available at https://github.com/ColumbiaCEAL/spatial-interactions-extension.

2. Research Questions

   We would like to build several extensions to the CEAL Lab's open-source WebNExt tool and evaluate how these extensions impact BLV user's experiences navigating the web.

   1. Can additional audio feedback around visual concepts provide a more seamless experience using WebNExt?
   1.1. How does building in additional audio feedback to WebNExt around visual concepts (such as when directional navigation hits the visual edge of the web page and navigating past large blank regions) affect BLV user's sense of orientation?
   1.2. Can meaningful image descriptions when alt text is missing be generated using lightweight AI models? How well does the GPT 4o vision model compare to existing alt text generation tools? And are the generated image descriptions helpful or distracting?

   We are planning to compare our work with existing tools and lead a structured evaluation session with ophthalmologist, Dr. Keale Cade at Howard University, to answer these questions.

   2. Can a speech-command interface that allows BLV users to navigate web pages through natural language reduce the cognitive load associated with keyboard-driven screen reader navigation? The goal is to simplify navigation by enabling commands like go to the first heading or find the search bar which can be more intuitive than memorizing keyboard shortcuts. Many accessibility tools focus on auditory output, but few address input flexibility, especially via speech. We'll enhance the existing WebNExt framework or create a standalone wrapper using Python's speech_recognition, pyaudio.
   2.1. Does speech-based navigation reduce task completion time? We will evaluate by

comparing the time, accuracy, and satisfaction of completing navigation tasks (e.g., time to find elements, number of interactions) using standard keyboard controls versus voice commands.
2.2. How do users perceive control and ease of use compared to keyboard-based methods? We will conduct a test session with the ophthalmologist to assess whether voice navigation feels more accessible or empowering in everyday scenarios.

3. Value to User Community

   The prospective users of our project are BLV people who use screen readers to navigate the web. Screen readers linearly parse the Document Object Model (DOM) tree structures of web pages. While this approach encompasses all the information on a web page, it omits spatial information such as the size, shape, and position of elements. The WebNExt study shows that BLV screen-reader users want to engage with visual semantics when navigating websites and find spatial navigation intuitive and immersive. A spatial understanding can improve collaboration with sighted web users by creating a shared frame of reference[3,4]. We have based our research questions on the feedback of BLV users given in the WebNExt study[2]. Our work will be open source and available on GitHub.

4. Demo: We expect to show the class a demonstration of our extensions to the WebNExt screen reader.

5. Delivery: Our code is public on github at https://github.com/demetriam/6156-Final-Project.

6. References

   1. Arnavi Chheda-Kothary, David A Rios, Kynnedy Simone Smith, Avery Reyna, Cecilia Zhang, and Brian A. Smith. 2023. Understanding Blind and Low Vision Users' Attitudes Towards Spatial Interactions in Desktop Screen Readers. In Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '23). Association for Computing Machinery, New York, NY, USA, Article 83, 1–5. https://doi.org/10.1145/3597638.3614490

   2. Arnavi Chheda-Kothary, Ather Sharif, David A. Rios, and Brian A. Smith. 2025. “It Brought Me Joy”: Opportunities for Spatial Browsing in Desktop Screen Readers. In CHI Conference on Human Factors in Computing Systems (CHI ’25), April 26-May 1, 2025, Yokohama, Japan. ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3706598.3714125

3. Juan Nino, Sherezada Ochoa, Jocelyne Kiss, Geoffreyjen Edwards, Ernesto Morales, James Hutson, Frederique Poncet, and Walter Wittich. 2024. Assistive Technologies for Internet Navigation: A Review of Screen Reader Solutions for the Blind and Visually Impaired. International Journal of Recent Engineering Science 11, (December 2024), 260–274. https://doi.org/10.14445/23497157/IJRES-V11I6P122

4. Venkatesh Potluri, Tadashi E Grindeland, Jon E. Froehlich, and Jennifer Mankoff. 2021. Examining Visual Semantic Understanding in Blind and Low-Vision Technology Users. In Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21). Association for Computing Machinery, New York, NY, USA, Article 35, 1–14. https://doi.org/10.1145/3411764.3445040

5. Michael F. Chiang, Roy G. Cole, Suhit Gupta, Gail E. Kaiser, and Justin B. Starren. 2005. Computer and World Wide Web Accessibility by Visually Disabled Patients: Problems and Solutions. Survey of Ophthalmology 50, 4 (2005), 394–405. https://doi.org/https://doi.org/10.1016/j.survophthal.2005.04.004