Towards Accessible Interactive Visual Elements on the Web: An AI-Based Descriptive and Auditory Input/Output Approach for Screen Reader Users

Toni Barth
Anhalt University of Applied Sciences
Köthen, Germany
e-mail: toni.barth@hs-anhalt.de

Leonie Weinert
Anhalt University of Applied Sciences
Köthen, Germany
e-mail: leonie.weinert@student.hs-anhalt.de

Prof. Dr. Frank Heckel
Anhalt University of Applied Sciences
Köthen, Germany
e-mail: frank.heckel@hs-anhalt.de

***Abstract*— Interactive visual elements, such as digital whiteboards, have become essential components of modern collaboration, yet they remain largely inaccessible to screen reader users. While static images can be described using alternative text, dynamic and highly interactive visual content is particularly challenging due to rapid updates, spatial complexity, and the limitations of linear auditory output. This paper outlines a conceptual framework that leverages AI-driven event analysis, spatialized auditory feedback, and natural-language interaction to enable screen reader users to both perceive and manipulate interactive visual workspaces in real time. The proposed idea aims to close the accessibility gap by providing a bidirectional interaction loop that supports independent contribution in collaborative environments. Remaining challenges include efficient real-time change detection, minimizing cognitive load, and enabling privacy-preserving local computation. This work sets the foundation for future research toward more inclusive interaction paradigms for dynamic visual collaboration tools.**

Keywords—Audio-visual web applications; Web accessibility; Web-based collaboration; online personal assistants.

I. INTRODUCTION

The modern web increasingly favors images and other visual elements over plain text for communication. For visually impaired users who rely on screen readers, however, such visual content poses major accessibility barriers. Static images can be made accessible through alternative text descriptions, but dynamic or interactive content presents a much greater challenge: a constant stream of updates would be needed to reflect changes in real time. This challenge is particularly evident in collaborative workflows, where users need to understand and interact with visual information, as well as contribute to it. Currently, visually impaired users often depend on others to translate their ideas into visual form — a process that is time-consuming and prone to errors. Existing accessibility tools and standards (e.g. Accessible Rich Internet Applications (ARIA), Web Content Accessibility Guidelines (WCAG)) provide limited support for interactive visual elements, leaving a gap in enabling real-time, self-directed interaction. Visual elements, such as whiteboards, have become key components of processeslike project management, professional training, or online teaching, which makes it harder for visually impaired users to follow along. This paper presents an idea for bridging that gap: a method to describe interactive visual elements to screen reader users in real time and to enable direct interaction without external assistance. Addressing both aspects forms the foundation for making interactive visual content truly inclusive.

The remainder of this paper is structured as follows. In Section II, we review relevant background and related work on accessibility of dynamic and collaborative visual interfaces. Section III analyzes the key challenges involved in communicating visual changes and enabling interaction for screen reader users. In Section IV, we present our proposed AI-based descriptive and auditory input/output approach for accessible interactive visual elements. Finally, Section V concludes the paper and outlines directions for future work.

II. BACKGROUND AND RELATED WORK

Prior work in the field of accessibility highlights substantial barriers when visually impaired users interact with dynamic interfaces. Studies on auditory displays show that non-speech, spatialized audio can effectively communicate complex visual information and reduce cognitive load [1]. Research on screen reader accessibility further indicates that real-time updates in collaborative environments often remain inaccessible due to latency, verbosity, or pure lack of information [2]. Learnings can be taken from research on the effective description of visual material for people with visual impairments. However, most existing work has focused on scenarios that do not involve time-critical or real-time constraints [3].

While standards, such as ARIA and WCAG, describe approaches for static or moderately dynamic content, they provide little guidance for highly interactive visual tools, such as digital whiteboards. Building on these insights, this paper explores how AI-driven event analysis, spatialized auditory feedback, and natural-language interaction could be combined into a unified approach for making interactive visual content accessible in real time.

III. PROBLEM ANALYSIS

When analyzing the process of interacting with visual elements, two key challenges can be observed: communicating visual changes and enabling user interaction.

A. Communicating Visual Changes

Dynamic visual content often changes rapidly in both structure and semantics. The main challenge is to identify and compose these changes in a way that is informative but not overwhelming. Updates must be transmitted accurately and

concisely, including what changed, where the change occurred, and who made it. This requires detecting relevant changes and translating them into concise, real-time descriptions, which is particularly important as screen readers themselves are text-based and linear, which means that only one piece of information can be pronounced at a time. Visual components, however, allow multiple changes at the same time, especially when working in larger teams.

B. Enabling User Interaction

Users need alternative, non-visual ways to specify objects and locations. Traditional input methods, such as mouse or touch interactions, rely heavily on spatial awareness and are therefore inaccessible to many visually impaired users. While keyboard navigation offers some control, it remains limited for spatial tasks such as positioning or drawing. Users who cannot employ graphical input devices require alternative, non-visual mechanisms to specify objects, their properties, and their locations. Such mechanisms require a consistent machine-backed mental model of the workspace.

IV. PROPOSAL: AI-BASED DESCRIPTIVE AND AUDITORY INPUT/OUTPUT SYSTEM

Our proposal for solving these issues combines three main components: (1) AI-driven analysis of visual changes, (2) auditory representation of these changes, and (3) AI-based interpretation of user input to manipulate the visual content. Together, these components form a feedback loop that allows users to both understand and contribute to interactive visual environments independently.

A. AI-driven Analysis of Ongoing Changes

An artificial intelligence system continuously monitors the visual element to detect relevant changes such as the addition, movement, or modification of components. Conceptually, this involves (1) detecting low-level visual events, (2) aggregating them into higher-level interaction events, and (3) assessing their relevance based on semantic and collaborative context. Rather than relaying every minor update, the AI filters and classifies changes according to semantic relevance, ensuring that only meaningful updates are presented to the user. For instance, on a collaborative digital whiteboard, adding a new visual object or significantly revising an existing drawing constitutes a meaningful update, whereas fine-grained, continuous drawing motions that do not yet reveal a stable visual structure are filtered out. This assessment may rely on a combination of computer vision techniques for visual change detection and learned models that distinguish transient interaction traces from stable visual artifacts. Each change is described in terms of what occurred, where it occurred within a two-dimensional coordinate system, and who initiated it, if applicable. This analysis provides the foundation for efficient and context-aware auditory feedback.

B. Auditory Interpretation of Visual Changes

The identified changes are communicated to the user through non-speech audio cues mapped to the spatial structure of the visual element. For instance, changes on the left side of the workspace may be represented by sounds localized in the left audio channel, while shape-specific tones or textures can indicate object types or actions (e.g., drawing, resizing, moving). This parallel auditory representation allows users to perceive multiple simultaneous events without the limitations of linear speech output. For larger or more complex visual scenes, users can navigate the workspace using input devices such as a keyboard or touchpad. Only changes within the user's area of focus are announced, minimizing cognitive overload. When the user selects a specific component, the system can provide detailed spoken feedback and enable follow-up questions via natural-language interaction with the AI.

C. AI-based Interpretation of User Input

To allow users to contribute actively, the system includes a natural-language interface for generating visual content. Users can describe intended modifications verbally or via text — for example, "Draw a blue rectangle in the top right corner" or "Connect node A to node B." User utterances are processed by a language understanding component that extracts intents (e.g., create, modify, connect) and parameters such as object type, spatial reference, and attributes. The AI interprets these instructions and translates them into corresponding graphical operations. Ambiguous or underspecified commands can be resolved through follow-up questions or clarification prompts. This enables non-visual creation and manipulation of complex visual elements while maintaining synchronization with other collaborators working visually.

The proposed architecture closes the interaction loop between perception and action for visually impaired users. By combining semantic AI analysis with spatialized auditory feedback and natural-language input, it offers a promising approach for making visual collaboration tools, dashboards, and design interfaces accessible in real time. This system could significantly reduce dependence on intermediaries, increase independence, and broaden participation in visual collaboration environments.

V. CONCLUSION AND FUTURE WORK

The next step is to determine how to interface with existing solutions for digital collaboration. Open-source tools, such as Excalidraw, might allow for native integration. In contrast, closed-source applications, such as Zoom and Microsoft Teams, will require alternative strategies to enable whiteboard accessibility, such as web browser extensions and image analysis.

Extensive research is also required to identify effective methods for performing all processing locally, without relying on cloud-based AI providers, in order to ensure that confidential data remains within the user's own ecosystem. If fully local processing proves impractical — for example, due to hardware constraints or performance limitations — efforts will focus on

developing a deployable service that can be hosted on the user's own infrastructure. This approach will preserve data ownership while still enabling advanced functionality.

Further investigation is needed to determine appropriate intervals for analyzing visual elements and to establish robust criteria for when analysis should be triggered. Continuous reanalysis at very short intervals (e.g., every three seconds) is inefficient if no changes have occurred since the previous analysis. Consequently, adaptive strategies based on change detection and interaction patterns will be explored to balance responsiveness with computational efficiency.

Finally, substantial user research will be required to design an interface that is intuitive, efficient, and accessible, with particular emphasis on the needs of screen reader users. This includes iterative usability testing, accessibility evaluations, and close collaboration with users with disabilities to ensure that the final solution supports inclusive and accessible interaction.

Ultimately, this work points toward a more inclusive future for visual collaboration tools. By enabling equitable participation in environments such as digital whiteboards and design interfaces, the proposed system has the potential to reduce dependency on intermediaries and expand opportunities for visually impaired users. Advancing this concept can contribute to the development of new accessibility standards and inspire further research into bridging the gap between visual and non-visual interaction paradigms.

REFERENCES

[1] T. Hermann, A. Hunt, and J. G. Neuhoff, *The Sonification Handbook*. 2011, pp. 23–24. Accessed: Jan. 27, 2026. [Online]. Available: https://sonification.de/handbook/download/TheSonificationHandbook-HermannHuntNeuhoff-2011.pdf

[2] P. N. Vargas and D. G. Trevisan, "Seeing What We Can't See in the Non-textual Representation from the Educational Context," in *Human-Computer Interaction – INTERACT 2025*, C. Ardito, S. Diniz Junqueira Barbosa, T. Conte, A. Freire, I. Gasparini, P. Palanque, and R. Prates, Eds., Springer Nature Switzerland, Sept. 2025.

[3] K. Anderer, K. Müller, L. Strobel, M. Wölfel, J. Niehues, and K. Gerling, "Making Lecture Videos Accessible for Students who are Blind or have Low Vision through AI-Assisted Navigation and Visual Question Answering," in *ASSETS '25: Proceedings of the 27th International ACM SIGACCESS Conference on Computers and Accessibility*, S. Kane, K. Shinohara, C. Bennett, and M. Mott, Eds., Association for Computing Machinery, Oct. 2025. doi: 10.1145/3663547.3746349.