

# CollabAlly: Accessible Collaboration Awareness in Document Editing

Cheuk Yin Phipson Lee\*  
Carnegie Mellon University  
Pittsburgh, PA, USA  
cheukyil@andrew.cmu.edu

Zhuohao Zhang\*  
University of Illinois at  
Urbana-Champaign  
Champaign, IL, USA  
zhuohao4@illinois.edu

Jaylin Herskovitz  
University of Michigan  
Ann Arbor, MI, USA  
jayhersh@umich.edu

JooYoung Seo  
University of Illinois at  
Urbana-Champaign  
Champaign, IL, USA  
jseo1005@illinois.edu

Anhong Guo  
University of Michigan  
Ann Arbor, MI, USA  
anhong@umich.edu

## ABSTRACT

Collaborative document editing tools are widely used in both professional and academic workplaces. While these tools provide some accessibility features, it is still challenging for blind users to gain collaboration awareness that sighted people can easily obtain using visual cues (e.g., who edited or commented where and what in the document). To address this gap, we present CollabAlly, a browser extension that makes extractable collaborative and contextual information in document editing accessible for blind users. With CollabAlly, blind users can easily access collaborators' information, track real-time or asynchronous content and comment changes, and navigate through these elements. In order to convey this complex information through audio, CollabAlly uses voice fonts and spatial audio to enhance users' collaboration awareness in shared documents. Through a series of pilot studies with a coauthor who is blind, CollabAlly's design was refined to include more information and to be more compatible with existing screen readers.

## CCS CONCEPTS

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

### ACM Reference Format:

Cheuk Yin Phipson Lee, Zhuohao Zhang, Jaylin Herskovitz, JooYoung Seo, and Anhong Guo. 2021. CollabAlly: Accessible Collaboration Awareness in Document Editing. In *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '21)*, October 18–22, 2021, Virtual Event, USA. ACM, New York, NY, USA, 4 pages. <https://doi.org/10.1145/3441852.3476562>

\*Both authors contributed equally to this work.

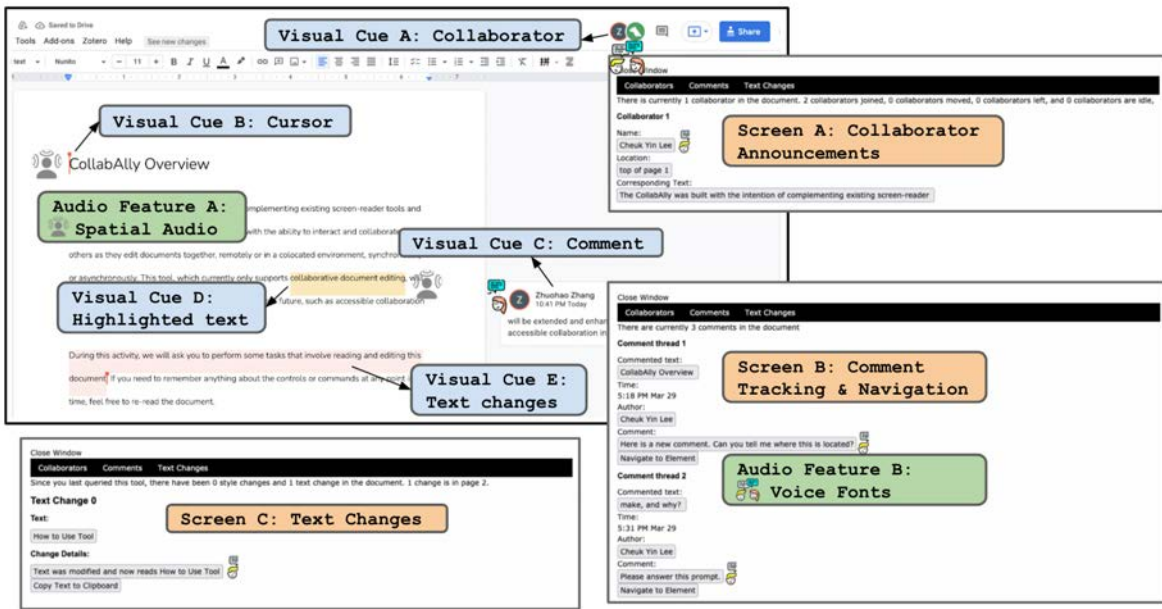
Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).  
*ASSETS '21, October 18–22, 2021, Virtual Event, USA*  
© 2021 Copyright held by the owner/author(s).  
ACM ISBN 978-1-4503-8306-6/21/10.  
<https://doi.org/10.1145/3441852.3476562>

## 1 INTRODUCTION

Collaborative document editing tools such as Google Docs, Microsoft Word, and Overleaf have become ubiquitous in today's professional and academic workplaces. Using shared documents, collaborators can synchronously or asynchronously access and modify document content, manage different versions, and track feedback and comments. Prior work has studied collaborative writing practices and found that they may offer productivity gains by allowing for a mix of synchronous and asynchronous writing [2, 9]. The collaboration awareness afforded by these tools allows users to decide what parts of a document to work on and when based on role structures and privacy needs [12].

However, collaborative document editing is challenging for blind users as tools are often not completely accessible [1, 3, 10]. Even when prior work has explored a number of ways to improve collaborative editing accessibility [6, 7, 11] and the fact that existing tools have accessibility support, they often do not provide the same collaboration awareness that visual versions have. While sighted users can take advantage of visual markup alongside the document text (i.e., highlights, nested comments, color-coded cursors), these benefits are not replicated non-visually [3]. Thus, existing accessibility support lacks the continuous and contextualized awareness of collaborator actions that visual cues provide, making the collaboration process more cognitively demanding for blind users. For example, while Google Docs provides verbal collaborator announcements when someone else enters or leaves the document, and Microsoft Word provides screen reader support for reading and replying to comments, it is still challenging for blind users to know who commented or edited the document, the location of the comment or changed text, and the content of the comment or change. Additionally when collaborating asynchronously, it is difficult to get contextual changes since they last accessed the file.

To assist blind users in collaborative document editing, we present a prototype system CollabAlly that makes extractable collaborative and contextual information accessible. CollabAlly consists of a backend engine that reverse engineers the structures and changes of a Google Doc from its DOM trees in real-time, and a browser extension that presents information including collaborators, comments, and changes in a 2D document canvas, allowing them to be laid out and vocalized using spatial audio and voice fonts [13].



**Figure 1: CollabAlly system overview: (i) Extracts visual cue information from Google Doc’s HTML DOM tree, including collaborators, cursors, comments, highlighted text, and text differences; (ii) Parses this data into a readable format and display it in three pop-up dialog boxes on-demand: (a) enhanced collaborator announcements, (b) comment tracking and navigation, (c) real-time and asynchronous text changes; (iii) Conveys contextual information using audio features, including spatial audio and voice fonts.**

CollabAlly also allows easy navigation for blind users leveraging a familiar browser-injected dialog popup and voice-guided interactions within the 2D document canvas.

## 2 COLLABALLY SYSTEM

CollabAlly aims to make collaboration awareness in document editing accessible to screen reader users. The CollabAlly system consists of a backend engine and a browser extension, that provides screen reader users access to information about the current document state and recent changes using a dialog box, accessed using keyboard shortcuts. The dialog box presents easy to navigate information access to collaborator locations, comments, and recent changes. CollabAlly also presents a summary of past asynchronous changes when users log into the document after being away.

### 2.1 Fetching and Modeling Document Changes

CollabAlly first reverse engineers the structures and changes of a Google Doc from its DOM trees in real-time. We identified several technical challenges based on how Google Docs are structured: (i) document and real-time collaborator information is only available by accessing the HTML DOM tree and parsing desired elements, as the Google Docs API is limited to basic reads and writes, (ii) only visible content on the browser screen can be fetched in real-time, (iii) HTML elements must be parsed, organized and transformed into a readable format, and (iv) 2D spatial locations of these elements must be communicated audibly to help guide blind users as they navigate the document.

To address these challenges, we identified different class names of desired HTML elements (i.e., collaborator cursors and comments)

to locate them in the document, and added a loading screen to scroll over the entire document so that non-visual elements can also be fetched. To parse and generate a readable format of all information, we store and track different states of collaborators, comments, and document text. Using this data, we can not only provide real-time support, but can also store all document states and provide updated document histories when users reopen the document.

CollabAlly intermittently queries the shared document to identify the cursor positions of collaborators, comment boxes and their corresponding highlighted texts, as well as text changes. Then, CollabAlly maintains an updated state of these elements and compares it with previously queried states. Specifically, CollabAlly uses Google’s “Diff-Match-Patch” library [4] to analyze text changes. The resulting changes are then reformatted using a library we developed that parses style and content changes, to be communicated to users in a digestible format. Each change is mapped to the page number and location in the document, to help users locate and navigate to it.

### 2.2 Presenting Contextual Information Using Audio Features

CollabAlly uses spatial audio and earcons to support document navigation by representing text locations as audio locations relative to the user’s cursor location. The location of the comment or collaborator in the document relative to the user’s location in the document is indicated by panning the sound source to the left or right of the user. We applied this feature to a set of audio beacons each associated with a different document change. When a collaborator enters or leaves the document, adds or deletes a comment,

or gets closer or further to the users cursor, corresponding earcons are played in the background using spatial audio.

All audio repositioning is implemented using the Resonance Audio SDK [5].

CollabAlly supports voice fonts. Each collaborator is assigned unique voice ID that can be pronounced through different Text-to-Speech (TTS) synthesizers. Currently, this feature is natively available for screen readers supporting Nuance Vocalizer engine [8]. For those using non-vocalizer-supported screen readers, CollabAlly also provides voice fonts through Google Cloud's TTS interface to generate server-side speech. Blind users, therefore, can easily distinguish what comments belong to whom using different voice colors. Information about a collaborator's location, comment text, or recent changes displayed in CollabAlly's dialog will be read using their voice font by the TTS synthesizer when the user presses a button associated with the change.

## 2.3 User Interactions

**2.3.1 Enhanced Collaborator Announcement.** CollabAlly provides an automatic event (in the background) and an active method (trigger by the user) to present information about collaborators' activities and locations. In the background, CollabAlly intermittently queries the shared document to identify the cursor positions of the collaborators relative to text elements in the document. Upon fetching that information, it uses sound notifications to communicate when a collaborator has moved their cursor within proximity of the user's cursor; when a collaborator's cursor has left the user's cursor proximity; when a collaborator has left or is idle; and when a collaborator has joined the document.

When users access the "collaborators" tab in the dialog box, CollabAlly provides a summary of how many collaborators recently joined or left the document and a list of current collaborators with their locations in the document. CollabAlly provides both a general estimate of the collaborator's location (i.e., "top of page one"), as well as the specific line of text that the collaborator's cursor is located in. Users can jump to any of these locations by pressing the toggle button. Additionally, they may press on the "Copy to Clipboard" button to copy the selected text that the collaborator is on, to later navigate to that element using Google Doc's built-in "search text" feature.

**2.3.2 Comment Sensemaking.** Using spatially positioned sound notifications, CollabAlly communicates whether a comment has been added or removed. Furthermore, upon opening the dialog box, users can also access details about who made the comment, when the comment was made, and the selected text that the comment is referring to. Comments are organized as threads, such that replies to a comment are nested chronologically within that comment thread. Voice fonts are also used to distinguish between collaborators and the authors of each comment.

CollabAlly also maintains an updated state of all the comments and compares it with the previously queried comments, so that users can track which comments have been added, removed, or updated. These changes are then formatted in the dialog box and also mapped to sound notifications using spatial audio.

**2.3.3 Real-time Text Changes.** Similarly, when collaborators are making text changes, blind users can learn about who made the change, what the change was and where it was located within one synthesized snippet. Furthermore, users can learn how the style of the text has been modified. By selecting the "Copy to Clipboard" feature, users can also navigate to any added or stylized text.

**2.3.4 Asynchronous Content Changes.** When users exit the document, all comment changes and text changes are stored in an external database. Once users re-enter the document, all asynchronous comments and text updates since the user last existed the document will be communicated to them. Users can easily access all asynchronous updates and changes in one centralized location, as opposed to manually searching through version histories of the document.

## 3 PILOT STUDY AND FUTURE WORK

CollabAlly's current design was refined through a series of pilot studies with a coauthor who is blind. We designed a series of reading and writing tasks in Google Doc and asked visually impaired participants to perform these tasks using Google Doc Accessibility Settings with CollabAlly installed. The tasks include (i) finding and locating collaborators in the document, (ii) locating comments and their corresponding text and resolving them, and (iii) locating real-time and asynchronous text changes.

Our pilot study informed two design improvements for CollabAlly. First, we modified the way users could access information within CollabAlly. We initially only used TTS to read out information about the collaborators, comments, and changes. However, we observed that TTS lacks user control and can thus overlap with existing screen reader announcements, and TTS also cannot be accessed such as using screen reader reading mode or refreshable Braille displays. Therefore, we updated our design to employ a dialog box to display and organize the information using a combination of easy to navigate earcons, texts, and buttons. Second, we modified how spatial audio was used in CollabAlly. We initially generated and positioned all audio snippets spatially, manipulating the volume and position properties to make the spatial layout more pronounced. We learned that the varying text volume of the actual document content was distracting and hard to perceive. Therefore, we updated our design to only play the sound icons spatially to deliver the spatial information, followed by the regular speech output of the document text content.

Although CollabAlly can accurately keep track of all useful HTML elements in a Google Doc, the system is still fragile as such reverse engineering could fail when Google Doc updates the way it constructs the HTML DOM tree. Additionally, CollabAlly also requires extra compatibility among the browser, Google Doc, and screen readers. The user experience could be much improved if these tools could adopt and integrate CollabAlly's features.

In future work, we plan to gather insights to further iterate and improve our design to make collaborative writing more accessible. We also plan to conduct evaluations to compare how users perform collaborative writing tasks and gain collaboration awareness using Google Doc's accessibility settings with and without CollabAlly. In the future, we plan to deploy and open source our tool to

the community, and to extend it for future research on accessible collaboration.

## REFERENCES

- [1] Khaled Albusays, Stephanie Ludi, and Matt Huenerfauth. 2017. Interviews and observation of blind software developers at work to understand code navigation challenges. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility*. 91–100.
- [2] Tom Boellstorff, B. Nardi, C. Pearce, and T. L. Taylor. 2013. Words with friends: writing collaboratively online. *Interactions* 20 (2013), 58–61.
- [3] Maitraye Das, Darren Gergle, and Anne Marie Piper. 2019. "It Doesn't Win You Friends": Understanding Accessibility in Collaborative Writing for People with Vision Impairments. *Proc. ACM Hum.-Comput. Interact.* 3, CSCW, Article 191 (Nov. 2019), 26 pages. <https://doi.org/10.1145/3359293>
- [4] Google. Software Retrieved 06/23/2021. Diff-match-patch: repackaging of google's diff match and patch libraries. <https://github.com/diff-match-patch-python/diff-match-patch>
- [5] Google. Software Retrieved 06/23/2021. Resonance audio. <https://resonance-audio.github.io/resonance-audio/>
- [6] Muhanad S Manshad, Enrico Pontelli, and Shakir J Manshad. 2013. Exploring tangible collaborative distance learning environments for the blind and visually impaired. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*. 55–60.
- [7] Jonas Moll and Eva-Lotta Sallnäs Pysander. 2013. A haptic tool for group work on geometrical concepts engaging blind and sighted pupils. *ACM Transactions on Accessible Computing (TACCESS)* 4, 4 (2013), 1–37.
- [8] Nuance. 2021. Nuance Vocalizer. <https://www.nuance.com/omni-channel-customer-engagement/voice-and-ivr/text-to-speech/vocalizer.html>
- [9] Ricardo Olenewa, Gary M. Olson, Judith S. Olson, and Daniel M. Russell. 2017. Now That We Can Write *<i>Simultaneously</i>*, How Do We Use That to Our Advantage? *Commun. ACM* 60, 8 (July 2017), 36–43. <https://doi.org/10.1145/2983527>
- [10] John G Schoeberlein and Yuanqiong Wang. 2012. Accessible collaborative writing for persons who are blind: a usability study. In *Proceedings of the 14th international ACM SIGACCESS conference on Computers and accessibility*. 267–268.
- [11] John G Schoeberlein and Yuanqiong Wang. 2013. Improving collaborative writing for the blind. *ACM SIGACCESS Accessibility and Computing* 107 (2013), 11–16.
- [12] Dakuo Wang, Haodan Tan, and Tun Lu. 2017. Why Users Do Not Want to Write Together When They Are Writing Together: Users' Rationales for Today's Collaborative Writing Practices. *Proc. ACM Hum.-Comput. Interact.* 1, CSCW, Article 107 (Dec. 2017), 18 pages. <https://doi.org/10.1145/3134742>
- [13] Wikipedia. 2020. Voice font. [https://en.wikipedia.org/w/index.php?title=Voice\\_font&oldid=954705999](https://en.wikipedia.org/w/index.php?title=Voice_font&oldid=954705999) Page Version ID: 954705999.