Co11ab: Augmenting Accessibility in Synchronous Collaborative Writing for People with Vision Impairments

Maitraye Das
maitraye@u.northwestern.edu
Northwestern University
Evanston, IL, USA

Thomas B. McHugh
mchugh@u.northwestern.edu
Northwestern University
Evanston, IL, USA

Anne Marie Piper
ampiper@uci.edu
University of California, Irvine
Irvine, CA, USA

Darren Gergle
dgergle@northwestern.edu
Northwestern University
Evanston, IL, USA

ABSTRACT

Collaborative writing is an integral part of academic and professional work. Although some prior research has focused on accessibility in collaborative writing, we know little about how visually impaired writers work in *real-time* with sighted collaborators or how online editing tools could better support their work. Grounded in formative interviews and observations with eight screen reader users, we built *Co11ab*, a Google Docs extension that provides configurable audio cues to facilitate understanding who is editing (or edited) what and where in a shared document. Results from a design exploration with fifteen screen reader users, including three naturalistic sessions of use with sighted colleagues, reveal how screen reader users understand various auditory representations and use them to coordinate real-time collaborative writing. We revisit what collaboration awareness means for screen reader users and discuss design considerations for future systems.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in collaborative and social computing**; **Empirical studies in accessibility**.

KEYWORDS

Collaborative writing, vision impairments, accessibility, screen readers, ability-diverse collaboration, collaboration awareness

ACM Reference Format:
Maitraye Das, Thomas B. McHugh, Anne Marie Piper, and Darren Gergle. 2022. Co11ab: Augmenting Accessibility in Synchronous Collaborative Writing for People with Vision Impairments. In *CHI Conference on Human Factors in Computing Systems (CHI '22), April 29-May 5, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 18 pages. https://doi.org/10.1145/3491102.3501918

CHI '22, April 29-May 5, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-9157-3/22/04…$15.00
https://doi.org/10.1145/3491102.3501918

1 INTRODUCTION

Online, distributed work has become the norm for many people in academic and industry settings. Yet, key technologies that support such collaboration (e.g., Google Docs, Microsoft Word) contribute to ongoing issues of inequity in professional and educational contexts. Consider the following vignette from Neil[1], a blind IT professional and proficient screen reader user who regularly performs collaborative writing with his sighted colleagues and friends:

> *"It's easy for your voice to become the small voice...when you're collaborating, because in a situation like this one (using Google Docs), I couldn't necessarily provide any constructive feedback on what change someone has made. It would take me kind of memorizing the document several times... I feel like that's a big responsibility to put on people [using] screen readers...rather than helping them get that information from the tool."*

Over 2 billion people worldwide use commercial collaborative writing tools on a monthly basis [1]. Still, these tools and their collaborative features (e.g., track changes, comments) offer only basic levels of screen reader access and do not afford people with vision impairments the same level of usability and efficiency their sighted peers experience [6, 20, 26, 27, 66]. Given the proliferation of remote work and collaboration in today's world, it is not only essential that we understand how ability-diverse teams—those involving people with and without vision impairments—engage in collaborative writing, we must design new technologies that support more accessible ways of working at a distance.

Research on accessible collaborative writing is nascent (e.g., [6, 26, 27]) compared to more than 30 years of research within HCI and CSCW on understanding collaborative writing practices [11, 13, 44, 55, 57, 79], developing theoretical frameworks [38, 47, 60, 73], and building new technologies (e.g., [5, 56, 78, 84]) to support collaborative work for sighted people. In their seminal 1992 paper [30], Dourish and Bellotti state that "...awareness of the content of others' actions allows fine-grained shared working and synergistic group behaviour which needs to be supported by collaborative applications." Myriads of collaborative writing tools have since applied these theoretical concepts and introduced efficient and effective ways for sighted people to develop collaboration awareness and coordinate actions through features such as comments,

[1] All names are pseudonyms.

suggested edits, real-time edit notifications, revision history, task assignment, and so on [11, 55, 57]. While not to the same extent, researchers have started investigating how visually impaired writers negotiate collaboration tools and practices with their sighted collaborators [26] and developing technologies to improve their collaborative writing experience [20, 27, 66, 80]. Yet, much of this work has focused on accessibility in *asynchronous* collaborative writing, leaving open questions around how visually impaired writers engage in *synchronous* collaboration (i.e., multiple authors working on a shared document in real-time) in ability-diverse teams.

To address this gap, the present paper analyzes how visually impaired writers interact with collaborative features to create, review, and revise documents with others in real-time and introduces new techniques to support their work. As a first step, we conducted formative interviews and observations with eight screen reader users who regularly perform real-time collaborative writing. Our analysis reveals three tasks that are essential for performing synchronous collaboration but difficult to accomplish using screen readers: 1) understanding who is doing what and where in real-time, 2) avoiding concurrent edits, and 3) developing a high-level overview of collaborative actions within a shared document. Drawing on these insights, we built *Co11ab*, a Google Docs extension that incorporates spoken and non-speech audio feedback and interaction techniques to enhance synchronous collaborative writing for screen reader users. We report results from exploratory design evaluations with fifteen screen reader users, including three naturalistic sessions of use with sighted colleagues, which detail how participants understood and reacted to audio cues in Co11ab as well as how they used the system to perform real-time collaborative writing.

This paper makes three core contributions to HCI and accessible computing. First, we contribute new empirically-based understandings of the complexities associated with synchronous collaborative writing for screen reader users, which complements prior work on accessible collaboration in asynchronous settings [20, 26, 27, 66, 80]. Second, we introduce novel auditory techniques to support screen reader access in collaborative editing environments, revealing new insights about how to design audio cues for related applications. Third, we revisit and extend theories of collaboration awareness in the context of our findings to better understand how to augment ability-diverse collaboration, particularly involving non-visual access, and the design of accessible collaborative systems.

2 RELATED WORK

Our work is informed by research on accessibility in collaborative work, collaborative writing tools and practices, as well as theoretical frameworks of collaborative writing and awareness.

2.1 Accessibility in Collaborative Work

Within HCI and accessible computing, a large and growing body of literature investigates collaborative practices of ability-diverse teams in professional, academic, and personal settings. Prior work has called attention to inequities faced by people with disabilities in collaborative environments [69], such as the 'invisible work' that blind employees must perform while working in predominantly sighted workplaces [16]. Researchers have also identified relationship maintenance as a critical component of creating accessibility in shared spaces and activities [15, 63, 71] and highlighted the ways in which disabled people negotiate co-creative practices with their able-bodied collaborators [25, 28, 48]. Others have studied how visually impaired people form a shared understanding with sighted collaborators while shopping together [83], performing programming tasks [58], receiving remote assistance [45], and seeking information online [3].

Complementing this work on understanding collaboration practices, researchers have developed new technologies to better support accessible collaborative work (e.g., [23, 33, 53, 74]). While a range of interaction modalities have been used in these technologies, auditory representations remain a dominant approach for enhancing the ways visually impaired people perceive textual, visual, and graphical information [2, 32]. As an example, Mendes et al. used spatial audio and multiple text-to-speech voices to facilitate workspace awareness of blind people on large multi-user interactive tabletops [50]. Shi et al. identified opportunities and challenges in using auditory feedback to help blind people detect visual cues while video calling [68]. Additionally, Metatla et al. developed a hierarchical auditory view of graphical diagramming software by combining speech and non-speech cues that helped blind and sighted co-workers explore diagrams together [52]. Other studies have examined the use of non-speech auditory feedback for improving the accessibility of collaboration in education and schooling [49, 75] and accessible craftwork [14].

Closely related to our study, researchers have identified inaccessible interfaces with modern screen readers that are related to understanding and referencing collaboratively written content [6, 20, 26, 66], and even performing basic writing tasks like formatting and resizing documents [22, 54]. Still others have designed auditory representations for asynchronous collaborative writing features such as comments and tracked changes [27]. Our work extends this prior research by studying and developing technologies to enhance accessibility in *synchronous* collaborative writing.

2.2 Collaborative Writing Tools and Practices

Over the years, HCI and CSCW researchers have investigated how people produce shared documents, exchange feedback, and interact with each other using collaborative writing tools and how to design systems to support people's collaborative writing practices. As an example, Wang, Olson, and colleagues highlight the many techniques that students use in groups to write collaboratively, which range from creating templates during the document creation stage to organizing contributions through a divide-and-conquer method [57, 78, 82]. Identifying where collaborators are and responding to them are also common practices, as Birnholtz and colleagues note with their observation of "group maintenance" behaviors [11, 12]. Similarly, Wang et al. [79] found that collaborators did not want to edit in close spatial proximity to others in a document in order to avoid exposing details of their own writing practices.

In addition to studying the writing process itself, researchers have also designed tools to enhance collaborative writing environments that range from extensions to commercially available systems such as Microsoft Word and Google Docs (e.g., [78, 79]) to new experimental systems (e.g., [84]). Many tools focus on visualizing collaborative actions and contributions to a written document over time. Perez-Messina et al. [59] designed a data structure to create interactive visualizations that show the organization and structure of collaborative text and historical edits. To better interpret asynchronous document changes, Wang et al. developed DocuViz [78] and AuthorViz [79], which aggregate and visualize revision history on Google Docs. Others have focused on visualizing collaborator contributions [43] and the location of collaborator's gaze within a real-time editor [42]. Although this body of research highlights the information people need to successfully collaborate in shared documents and approaches for presenting that information, we know considerably less about the experience of visually impaired writers and how to support their informational needs.

2.3 Theories and Frameworks of Collaborative Writing and Collaboration Awareness

Over the years, researchers have synthesized studies on collaborative tools and practices to develop theoretical frameworks and taxonomies to understand collaborative writing. Posner and Baecker [60] proposed a framework to describe key components of collaborative writing including roles, activities, document control, and writing strategies (e.g., joint writing, parallel writing, etc.). Lowry et al. [47] extended Posner and Baecker's framework by including work modes in terms of proximity and synchronicity, i.e., whether the collaborators will perform writing in the same or different places and same or different times. Relatedly, Dourish and Bellotti [30] defined the notion of collaboration awareness as an "understanding of the activities of others, which provides a context for your own activity" within a shared document. Depending on the degree of engagement and planning, Baecker et al. [5] defined two levels of awareness: focused collaboration (when people work together in a tightly-coupled manner) and peripheral awareness (when people have rough ideas about their collaborators' activities).

Building on these early ideas, Gutwin and Greenberg [38] developed a conceptual framework for workspace awareness in real-time small-group collaboration. They posited that collaborators need to be aware of *who* is doing *what* and *where* in a shared workspace. Within each of these broad categorical questions, they articulated specific informational elements that help people keep track of real-time activities occurring in the workspace. For example, 'who' category includes presence, identity, and authorship; 'what' category includes action, intention, and artifact; and 'where' category includes location, gaze, view, and reach. This framework also detailed the ways in which awareness information is used in various collaborative activities [38]. In particular, they discussed how peripheral awareness of others' actions help people coordinate actions, determine opportunities to assist one another, anticipate conflicts, and manage coupling [39] (i.e., recognizing when it might be appropriate to switch to more tightly-coupled work from loosely-coupled individual work). Additionally, awareness information also makes it easier to develop common ground [21] and communicate about tasks. To date, however, these theories of collaborative work focus on sighted people and neglect the ways in which ability-diverse teams interact and make use of collaborative tools.

3 FORMATIVE STUDY: METHOD

To understand accessibility in synchronous collaborative writing, we conducted remote interview and observation sessions with eight screen reader users. This study was approved by institutional review board of Northwestern University.

3.1 Participants

We recruited eight participants (aged 20-44; 4 identified as female, 3 as male, 1 as non-binary; 5 identified as White and 3 Hispanic). See Table 1 for participants' self-reported visual ability, occupation, how frequently they perform collaborative writing, and how frequently they use different collaborative writing tools. All participants were advanced or expert users of at least two screen readers among JAWS, NVDA, and VoiceOver, while some had experience with Narrator, TalkBack, and ChromeVox. All participants were residents of the United States.

3.2 Procedure

The first author conducted the session over Zoom between December 2020 to February 2021. We first asked questions about participants' usage of assistive technologies and collaborative writing tools, probing for challenges they encounter and strategies they develop when writing with others. Next, we remotely observed participants as they worked on a sample Google Doc in real-time with three coauthors. The researcher played the role of the coauthors (one using their own name and two anonymous profiles). The sample document was a one-page article about an animal pre-populated with several comments and suggested edits. Participants used their preferred screen reader to work on the document; four used JAWS, three NVDA, and one VoiceOver. We asked participants to share their screen via Zoom (including computer sound) and verified that participants had the recommended Google Docs and screen reader settings enabled (e.g., turning on collaborator announcements).

During the observation period, we asked participants to go through the sample document and talk aloud what information they were trying to access, what information they received, and how. Next, we enacted four different scenarios where the researcher: 1) placed their cursor in the paragraph where the participant had their cursor, 2) moved their cursor away from the participant's paragraph, 3) placed their cursor at exactly the participant's cursor location, and 4) typed a sentence in the participant's active paragraph. During each scenario, we asked participants to share what they thought was happening based on the collaboration notification they received (if any). We also asked them to find where the coauthors were editing and who was doing what, along with follow-up questions regarding whether they pay attention to such collaboration information in their regular work, in what contexts, and how. We spent approximately 10 minutes on each scenario. We encouraged participants to use any features available on Google Docs to access collaboration information (e.g., 'live edits'). Participants who had experience performing real-time editing on Microsoft Word also shared how that matched or differed from working on Google Docs. We concluded with a brainstorming period, where we asked participants to share their thoughts on how collaboration information can be presented to better support their real-time writing using screen readers. Each session lasted for approximately 90-100

Table 1: Details of participants (all names are pseudonyms). Participants were recruited for the formative study (F) and design evaluation conducted in two phases: researcher-facilitated sessions (RF) and naturalistic collaborative writing sessions (NW).

Name and participation phase	Self-reported visual ability	Occupation	Collaborative writing frequency	Collaborative tools used
Alex (F, RF)	Blind due to Retinopathy of Prematurity, some light perception	Technical writer	Monthly (real-time: less than once a month)	Docs: monthly, Word: less than once a month
Cory (RF)	Blind since birth from Pseudo-optic Dysplasia, has light perception	Assistive tech instructor	Daily (real-time: less than once a month)	Docs: weekly, Word: daily
Dan (RF)	Nearly totally blind since 8-9 years old, some light perception	IT professional	Monthly (real-time: monthly)	Docs: weekly, Word: weekly
Gina (F, RF)	Totally blind since birth due to glaucoma	Consultant	Weekly (real-time: weekly)	Docs: monthly, Word: daily, Office 365: weekly
Ian (RF, NW)	Blind since birth due to retinal detachment, has light perception	Student	Weekly (real-time: weekly)	Docs: monthly, Word: daily
Jim (F, RF, NW)	Totally blind due to Leber Congenital Amaurosis	Web developer	Weekly (real-time: monthly)	Docs: monthly, Word: daily
Joy (RF)	Legally blind, nearly functional print vision in one eye, born with cataract, developed glaucoma	Assistive tech specialist	Weekly (real-time: monthly)	Docs: monthly, Word: weekly, Office 365: once or twice
Kate (RF)	Nearly totally blind since birth, some light perception	Access tech director	Daily (real-time: weekly)	Docs: weekly, Word: daily
Kirk (RF)	Totally blind since birth, some light perception	Programmer	Weekly (real-time: weekly)	Docs: weekly, Word: weekly, Office 365: monthly
Leah (F, RF)	Totally blind due to Leber Congenital Amaurosis, some light perception	Assistive tech director	Weekly (real-time: weekly)	Docs: weekly, Word: Daily, Office 365: weekly
Mia (F, RF, NW)	Totally blind due to Retinopathy of Prematurity	Academic researcher	Weekly (real-time: monthly)	Docs: weekly, Word: daily
Neil (F, RF)	Totally blind due to Retinopathy of Prematurity, some light perception	IT professional	Weekly (real-time: monthly)	Docs: monthly, Word: weekly, Office 365: less than once a month
Ron (RF)	Totally blind due to Retinitis Pigmentosa since 8 months old, has light perception	Accessibility software architect	Daily (real-time: weekly)	Docs: daily, Word: daily
Ryan (RF)	Totally blind since birth due to glaucoma and cataracts	Accessibility analyst	Daily (real-time: daily)	Docs: less than once a month, Word: daily, Office 365: daily
Sean (F, RF)	Blind due to Optic Nerve Atrophy since 2 years old, can read large print (35pt+)	Access tech specialist	Weekly (real-time: monthly)	Docs: weekly, Word: daily
Vera (F)	Legally blind since birth, gradual vision loss	Program director	Weekly	Docs, Word, Pages

minutes. Participants were compensated with US$60 gift cards. All the sessions were recorded and transcribed for analysis.

3.3 Data Analysis

We analyzed the data following reflexive thematic analysis [17, 18]. The first author open-coded the transcripts and reviewed screen recordings with a particular focus on the way participants acted and reacted upon receiving collaboration notifications in different scenarios. Next, through iterative comparison between data to data and data to emerging themes and regular discussions as a group, we developed three themes that capture key challenges screen reader users face in real-time collaborative writing.

Although six participants explored the sample document with braille displays along with screen readers during the observation session, none of them received any collaboration notifications on their braille displays. In this paper, we focus our analysis on the auditory speech output of the screen reader but acknowledge that the lack of support for collaborative writing through braille displays needs to be addressed in future work.

4 FORMATIVE STUDY: FINDINGS

Our analysis revealed three challenging but essential tasks in collaborative writing for screen reader users: understanding who is doing what and where in real-time, avoiding concurrent edits, and getting a high-level overview of collaboration information.

4.1 Understanding Who is Doing What and Where in Real-Time

An important aspect of maintaining awareness in remote collaborative work is to understand who is doing what and where in the workspace [5, 30]. Access to this information shapes the way coauthors coordinate actions with each other, maintain attention, and fluidly transition between individual and shared tasks [38]. In existing collaborative writing tools (e.g., Google Docs, Microsoft Word), sighted people can visually follow color-coded cursors that move and type letters in real-time to understand who is editing what and where. They can also bring into focus (i.e., the visible portion of a page) the region where a coauthor has their cursor at a particular instance by clicking on the coauthor's avatar at the top of the window and thus monitor what the coauthor is doing. Screen reader users, however, only receive spoken announcements when they (or their coauthors) place their cursor on (or move away from) a paragraph where collaborators have their cursors positioned (e.g., 'X, Y, and Z are editing this paragraph,' 'Y has left this paragraph,' etc.). Thus, our participants shared that the only way for them to identify the paragraphs where others are working, especially on Google Docs, is to manually scroll through the document until they land on each coauthor's paragraph. Neil said, *"[If] I wanted to see if someone's currently changing something, I would literally be having to continuously scroll this to see... because it won't tell you that... unless you yourself move to that paragraph."*

Moreover, the spoken notifications screen reader users receive, both on Google Docs and Microsoft Word, only informs them of the presence of others in the particular paragraph but does not provide details about what or exactly where others are editing. Mia explained, *"The notifications... they're not helpful because you don't know where [exactly] the editing is happening, or you don't know the context of what's being edited...that you're supposed to be seeing on the screen at the same time."* During the observation session, some participants tried to find the exact location of the researcher by parsing ambiguous and inconsistent announcements they received through their screen reader, but often their guesses were incorrect.

Even when participants could identify the paragraph where someone was editing, they had to speculate the content of the new edits *"based on memory"* by reading the entire paragraph multiple times (see also [26]). To gather more information about the coauthor's real-time activities during the observation session, Jim, Leah, and Neil tried to use Google Docs' 'live edits' feature, which provides a periodically updated list of recent edits that screen reader users can traverse through. Most of the real-time edits did not appear at all in the 'live edits' region, potentially due to latency issues, and the ones that did appear lacked contextual details such that participants could not develop a coherent understanding of the location and content of the edits.

Given the complexities associated with accessing real-time edit notifications through screen readers, participants, in most cases, resort to *"communicating among ourselves more so than communicating through the document"* (Vera). Often they maintain an external communication channel (e.g., audio-video calling or chat applications) while working together on a shared document. Gina said that she *"would ask the person, 'Where exactly are you?', so I could at least find that paragraph and then go from there."* Nevertheless, this workaround was not ideal; participants talked about the 'extra work' [16, 26] their sighted collaborators needed to perform for sharing low-level collaboration details with them. Mia said, *"I'm just sitting there like, 'What are you actually doing?' She's describing what she's doing as best she can... I can understand how hard that is, because you're editing, you're making the changes at the same time too. So then, hav[ing] to describe that to somebody else can be hard."*

In addition to describing the challenges with understanding real-time collaborative actions, participants brainstormed about potential approaches to address this issue. Mia thought that a useful technique could be *"keeping track of the cursors... so that you can follow along and see what edits they're making."* Jim also suggested a similar approach: *"If you could switch to a collaborator's POV, point of view... [if] I could hear their words being placed in a document as if it was me editing it, that would be nice."* However, participants also cautioned against potential information overload in that *"what was being entered by someone else would override what you were entering"* (Alex). Gina, Alex, and Vera, instead, wanted a way to directly jump to a coauthor's cursor location and manually review their edits by moving the cursor.

4.2 Avoiding Concurrent Edits

Not only do existing collaboration notifications limit screen reader users' understanding of the real-time activities of their couthors, they also fall short in making users aware of the possibility of concurrent edits. As discussed before, the visual design of real-time edits makes it easier and intuitive for a sighted person to avoid typing at the same spot where their coauthors' cursors are. The moving cursors and real-time synced letters also allow sighted people to seamlessly edit on their own even when their cursors are close to their coauthors' cursors but not at the exact same spot. For instance, two sighted persons can edit at the beginning and ending portions of a paragraph without causing any friction. Screen reader users, however, have no way to gauge the proximity of their coauthors' cursor positions relative to their own, since they get the same notification (e.g., 'X, Y, and Z are editing this paragraph') when their coauthors' cursors are located anywhere on the paragraph they are editing. Thus, real-time collaborative writing tools (e.g., Google Docs, Microsoft Word) attempt to control concurrent edits by keeping screen reader users at least an entire paragraph away from their coauthors. This significantly limits the extent to which screen reader users can engage in close co-editing with others and increases chances of conflicting edits. Leah said, *"If I am editing the same thing you're editing, for example, I don't know how it shows you on the screen, but the screen reader doesn't really tell you. It only says that you're in that section (paragraph). It would be great to have that information because that way we wouldn't be working on the very same spot... we're not changing the same thing."*

Furthermore, the spoken notifications for coauthors' movements cause *"speech clutter,"* and disrupt the screen reader user's own reading or writing [26]. Neil said, *"Some of the speech feedback that you receive while you're doing collaborative writing, it gets very chattery. Specifically, as people move around the document… it gets very overwhelming very quickly especially with multiple people."* Participants shared that they prefer to keep the collaborator announcements off entirely, as they *"find them more distracting than helpful"* (Alex). Instead, they maintain a strict spatial separation between each other's editing zones, often enforced through external communication. Jim explained, *"If we're editing in real time, I might say, 'Okay, are you doing this? Or am I doing this?' Just so that we're not stepping on each other's toes."* This way, participants' real-time collaborative writing mostly takes a divide-and-conquer route, where they must coordinate with their coauthors to review their edits after the fact. Eventually, finding workarounds to deal with the constraints of collaborative tools leads to a suboptimal collaboration experience for our participants. Leah lamented, *"I was able to collaborate, but not really in real time."*

Going back to the root cause of these complexities, participants thought that having *"less cumbersome"* (Gina) ways to learn where coauthors are editing could help them avoid concurrent edits but still perform editing safely in close vicinity. Most participants said they would prefer some form of non-speech audio cues such as earcons with changing pitch and text-to-speech voices with varying pitch or speech rate. For example, Jim said, *"If the [screen reader] voice could switch when you're on a collaborator's line... then you can see like- oh, this is a line that we're both working on together."* In summary, the current lack of information regarding collaborator proximity makes it challenging for screen reader users to maintain distance from others and avoid concurrent edits.

4.3 Getting a High-Level Overview of Collaboration Information

In addition to attending to who is doing what and where in real-time, writers also need to develop a high-level understanding of how collaboration information is distributed in a document [73]. Such a high-level overview can help writers plan and prioritize their next course of actions [38]. Current collaborative writing tools (e.g., Google Docs, Microsoft Word), however, do not provide screen reader users with a straightforward way to develop a high-level overview. Instead, they must traverse the list of comments and suggested edits one by one and jump back and forth between the main text content and comment/edit panes. Neil described his mental process for building an overview:

> *"I imagine the full page of text kind of in front of me, maybe not with the contents of the text, but general idea of the layout of the paragraphs, the lines, as I kind of went through it, and kind of mapping that out after I read it for the first time. And then from that point kind of having just a representation or idea of where others have edited or made changes for myself... I visualize it as almost like a drafting paper grid, where the text is kind of represented with each square. And then people's presence and changes are marked by various boundaries around them like grid. To me, I feel like that just kind of helps orient my own position."*

The above process of 'visualizing' high-level information is cognitively intensive and requires Neil to navigate the entire document multiple times, by line and by paragraph, *"with the [spoken notifications for] changes made and without, and really understand that in and out."* Unlike Neil, others said that they rely on external communication with their coauthors *"to call attention to a specific thing. They'll tell me like, 'Hey, please check out page three. There's a ton of stuff in there. I want you to look at'"* (Gina).

To make the process of perceiving the distribution of collaboration activities easier and efficient, participants emphasized that *"an overview summary... would be useful, especially for a very large document… to be able to not have to read the whole thing to just quickly jump to where the changes are made"* (Sean). In particular, Gina thought that receiving non-speech audio cues could be helpful in understanding overview information, *"like some kind of sound or... way of telling me- if I go here, this is where the edits are... something is clustered in a certain area, to be able to know that upfront."*

5 SYSTEM DESIGN AND DEVELOPMENT

To address the key challenges we identified in our formative work, we developed Co11ab, a Google Chrome based extension that uses a variety of spoken and non-speech audio cues to present synchronous collaboration information.[2] The extension works in tandem with a visually impaired writer's screen reader and Google Docs and maintains these applications' underlying functionalities, e.g., shortcuts for navigating text and reading comments or suggested edits. Co11ab enables screen reader users to selectively access collaboration information through five features, which we detail below.

5.1 Query Location

Since figuring out the location of active coauthors can be difficult in existing collaborative writing platforms (see Section 4.1), Co11ab enables users to hear a list of co-authors who are editing in a document and where they are editing by pressing a shortcut (Control+Shift+K). This provides the exact location of their cursor at a specific instance (e.g., 'Joe is editing in page one, paragraph five, and line three').

5.2 Follow Mode

To help understand who is editing what in real-time, Co11ab enables screen reader users to invoke a 'follow' mode (Control+Shift+H) that switches to a certain coauthor's *"point of view,"* as our participants recommended in the formative study (see Section 4.1). In this mode, if the coauthor being followed is actively typing in the document, the system continuously emits a typing sound. Once the coauthor completes a line, the screen reader reads what they have just typed. If the coauthor being followed is not actively typing but only moving cursors from line to line, the screen reader reads the content of the lines they are moving their cursors from but does not emit any typing sound. Thus, the presence or absence of the typing sound indicates whether or not the followed coauthor is writing at that instance. This way, the screen reader user can listen to the text content a coauthor is writing line by line or where the coauthor is moving in the document without losing their own cursor position. When the screen reader user turns the follow mode off by pressing Escape, their cursor remains where it was at the time of invoking the follow mode.

5.3 Jump to a Coauthor's Location

Informed by our participants' suggestions in the formative study (Section 4.1), we incorporated a feature in Co11ab that allows a screen reader user to jump to the cursor location of a coauthor being followed by pressing a shortcut (Control+Shift+One). Follow mode automatically turns off upon the user jumping to the coauthor's location. An important distinction between this feature and the current version of Google Docs or Microsoft Word is that when a sighted user clicks on a collaborator's avatar, the document view is adjusted to include the collaborator's cursor location within the visible portion of the page. Co11ab, in contrast, supports querying specific information about collaborator location and allows rapid cursor movement to their position so that the screen reader can read aloud surrounding text content.

[2] Co11ab extension and our code repository are publicly available.

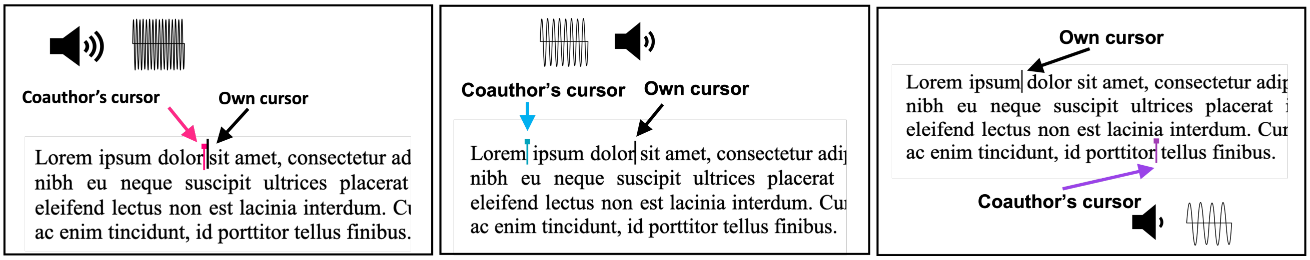


Figure 1: Relative proximity between a screen reader user's own cursor and a coauthor's cursor is indicated with earcons. Left: both cursors are at the same spot; a high-pitched or loud earcon is emitted. Middle: the cursors are in the same line; a medium-pitched or medium-loud earcon is emitted. Right: the cursors are within the same paragraph but far apart; a low-pitched or quieter earcon is emitted. Auditory icons with varied loudness work similarly.

5.4 Relative Proximity Notifications

To make screen reader users aware of the possibility of concurrent edits without causing *"speech clutter"* (see Section 4.2), Co11ab indicates relative proximity between cursors through non-speech audio cues [27]. It allows a screen reader user to choose from five different options of relative proximity information: earcon with varied pitch, earcon with varied loudness, auditory icon with varied loudness, spoken announcements, or no notifications. An earcon refers to an abstract auditory representation of an event whereas auditory icon resembles realistic sound [32]. The system emits a brief musical note as an earcon and the sound of typing on a keyboard as an auditory icon. We manipulated certain audio attributes (i.e, pitch and loudness for the earcon and only loudness for the auditory icon) to indicate how close or far away a coauthor is located from the screen reader user's own cursor. The audio attribute (pitch or loudness) gradually increases when a coauthor approaches the text portion where the screen reader user is editing, and decreases when the coauthor's cursor moves away from the user's cursor location (see Figure 1). We chose pitch and loudness as variables for relative proximity because these attributes can indicate similarity and difference between successive audio events [19, 29]. For the earcon, different coauthors' cursor movements were indicated by different musical instruments. In the current implementation of the system, these instruments were automatically assigned to different coauthors based on when they join the document (e.g., the first coauthor is encoded by banjo, the second one with violin, etc.). The auditory icon is the same for all coauthors.

In presenting relative proximity between coauthors' cursors, we follow the inverse relationship between collaborators' distance and the need for collaboration awareness described by the focus/nimbus model [62]. In other words, the audio variable (i.e., pitch or loudness) changes more when coauthors' cursors are closer to each other (e.g., within the same line), compared to when their cursors are far apart. We defined four semantic levels for relative proximity between coauthors' cursors: 1) at the exact same spot, 2) within the same line but not at the same spot, 3) within the same paragraph but not in the same line, and 4) not within the same paragraph. When a coauthor's cursor moves across these semantic levels, the audio variable changes accordingly. The loudness increases by 30% for each level, with 100% in the highest level (at the exact same spot) and 10% in the lowest level (not within the same paragraph). Similarly, the pitch increases by 5 full notes for each level, with notes B5 (996.75 Hz), D5 (592.67 Hz), F4 (352.40 Hz), and A3 (222.00 Hz) for the highest to the lowest level. Similar semantic categories for relative proximity are applied in preparing spoken announcements (e.g., 'Joe is not in your paragraph/ Joe is in your paragraph/ Joe is in the same line/ Joe is at the same spot'). To reduce information overload, the audio cues are played only when a coauthor's cursor moves across the semantic levels but not while a coauthor remains within a certain level.

5.5 Audio Scrollbar

Inspired by Hill et al.'s graphical presentation of 'edit wear' [40], we developed an audio scrollbar to provide a high-level overview of how coauthors' edits, cursors, or comments are distributed in a shared document. This information, although not readily available for screen reader users in existing collaborative writing tools (see Section 4.3), is essential for writers to understand which parts of a document have the most contributions from coauthors or which parts have remained comparatively unchanged [57]. A screen reader user can invoke the audio scrollbar by pressing a keystroke (Control+Shift+U). As they virtually move from paragraph to paragraph (or line to line) by repeatedly pressing the up/down arrow key, the audio scrollbar emits earcons with varying pitch (or repetitions) corresponding to the particular paragraph (or line). A short, quiet blip sound is played to denote a paragraph (or line) with no collaboration elements (i.e., edits, comments, or active cursors). An example of the audio scrollbar is shown in Figure 2. The system implements a hierarchical navigation approach so that a screen reader user can develop a high-level overview of collaboration information and also easily explore any particular sections they want in greater detail. For instance, the user can listen to meta-information about the number of collaboration elements in a paragraph or line (by pressing Control+M) or move directly to the paragraph or line to begin editing or read in detail (Control+Shift+One).

The user can choose from two different representations of the audio scrollbar. The first one consists of short-lived earcons (i.e., musical tones) with varying pitch relative to the number of collaboration elements per paragraph (or per line) in a document. The pitch of the earcons increases when there are more elements in a paragraph (or line) and vice versa. A screen reader user can choose

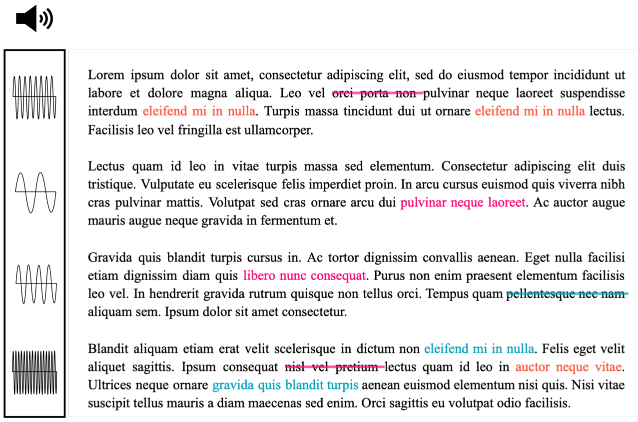


Figure 2: The audio scrollbar for a document with four paragraphs. When the audio scrollbar is invoked, earcons of different pitches play according to the number of edits in the paragraph. The second paragraph has the lowest pitch, since it has the lowest number of edits (one). Similarly, the fourth paragraph has the highest number of edits (four) and the corresponding earcon has the highest pitch.

whether the audio scrollbar indicates only comments, edits, or cursors, or all of those; and whether it will calculate the elements per line or per paragraph. To ensure that the earcons remain within the audible range and are euphonic and easily discernible, we used five short tones with different pitch levels generated by a musical instrument (piano) as earcons. In the second representation, instead of varying pitch, an earcon is played multiple times (2-5 times) according to the number of elements per paragraph (or line). In other words, an earcon is played five times for the paragraph with five or more edits and only once for the paragraph with only one edit, if edits are selected as the intended overview element. Note that our design of audio scrollbar is different from Yalla and Walker's auditory scrollbar [81], which uses audio cues for menu navigation.

5.6 Implementation details

Co11ab is implemented in JavaScript by using a combination of Web and Google Chrome extension-specific APIs. It consists of three components: a settings popup window, a persistent background storage and data transfer layer, and a collection of scripts implementing the functionalities of our system. The settings popup provides screen reader accessible lists of features that a user can toggle on and off. The background layer registers any updates in the feature settings and keypress events of feature-specific shortcuts to the Google Doc a user has opened through the Chrome runtime message passing APIs. It also contains logic to log a user's actions for our data analysis, as well as functionality for certain features that are not implementable through standard web APIs, such as automatic sound playing and text-to-speech conversion.

When a user loads a document, our extension injects the collection of scripts that implement the system's functionality into the web application's context using the Chrome Content Script feature. This provides Document Object Model (DOM) level access to the Google Doc, as well as the ability to make changes and execute JavaScript as if it was provided by the web application. The system's features generally follow a pattern of invoking the retrieval of some document information, such as text content, cursor location, or content placement, and then transforming that information into an accessible output. This invocation of information retrieval can be through the action of the user such as pressing a keyboard shortcut, by the action of a user's coauthors such as moving to a different location, or through a JavaScript mutation observer that fires when the content, styling, or position of a document element changes. The output generation can be as simple as taking a text input and announcing it through the user's screen reader using the ARIA-live API; or more complex, such as taking positional or other quantitative document information and passing it to the background layer's automatic sound playing implementation to play a specific pre-recorded sound or a tone with a specific frequency using the Tone.js JavaScript library.

6 DESIGN EVALUATION: METHOD

To understand how screen reader users understand and use the new techniques introduced in the Co11ab extension during synchronous writing, we conducted an exploratory study with fifteen screen reader users. Additionally, three screen reader users participated in a naturalistic writing session with a known sighted collaborator.

6.1 Participants

Participants were recruited from our research network and snowball sampling (aged 20-44, mean=32.9 years; 5 identified as female, 8 as male, 1 female/non-binary, 1 non-binary; 8 identified as White, 4 as Hispanic, 1 as African American, 1 as Middle Eastern, 1 as mixed race). Twelve participants were residents of the United States at the time of our study. Seven of them also took part in our formative interviews. Participants mostly perform collaborative writing with their colleagues and friends. All participants have both sighted and visually impaired collaborators, except Gina, Leah, and Kirk who collaborate with only sighted people and Cory whose collaborators are visually impaired. JAWS, NVDA, and VoiceOver were the most common screen readers among our participants, although some had experience with Narrator, TalkBack, and ChromeVox. While we focused on audio output for all participants, Cory used a braille display in tandem with their screen reader since they rely on both braille and audio output in their regular writing. See Table 1 for details of participants' self-reported visual ability, occupation, how frequently they perform collaborative writing, and how frequently they use different collaborative writing tools.

6.2 Researcher-Facilitated Sessions

The first author conducted one-on-one sessions with participants via Zoom between July and August 2021. Each session lasted for approximately 90-115 minutes. We collected consent from the participants and provided detailed instructions for installing Co11ab and required setup steps. To help participants explore Co11ab, we shared a sample one-page Google Doc with them. The document had six paragraphs and several comments and suggested edits from two collaborators (the first author and one anonymous profile). The text content was copied from a blog post on a pet animal. We used distinct copies of the same document for each participant.

At the beginning of the sessions, participants shared their screen and computer sound through Zoom and read the content of the sample document shared with the researcher. Next, we guided participants through the steps for selecting various audio representations (e.g., earcon, auditory icon, or spoken announcements for relative proximity) and invoking different features available in the extension (e.g., follow mode and audio scrollbar). Participants first explored the features related to real-time activities followed by the relative proximity notifications and finally the audio scrollbar feature. The researcher played the role of collaborators through two author profiles (one in their own name and one anonymous). For example, when a participant pressed the shortcut to follow the anonymous collaborator, the researcher performed typing or cursor movement in the document from the anonymous profile so that the participant could experience how the follow mode works. After each interaction, we asked participants to share their thoughts on what happened and why, prior to explaining what the feature does. We concluded the sessions with an overall debrief on the entire system, probing participants for their thoughts on potential use cases, trade-offs, and challenges that might arise while using the system. Participants were compensated with US$60 gift cards.

6.3 Naturalistic Collaborative Writing Sessions

We conducted follow-up naturalistic writing sessions with three participants (Ian, Mia, and Jim), each of whom invited an existing sighted collaborator to join them for the session. Ian (student) and Lily (student) are siblings who perform real-time collaborative writing using Google Docs on a monthly basis to prepare homework and essays. While not regularly, Mia (postdoctoral associate) and Ivy (undergrad) have worked together a few times using Google Docs for data entry and bibliography management. Jim (web developer) and Ava (college student) had previously collaborated on developing ideas for a story Ava was planning to write, but our study was the first time they wrote together on a shared document in real-time. All dyads had collaborative writing experience using Google Docs and often used external tools such as phone calls or chat applications to communicate with each other while writing.

We began each session by reminding the visually impaired participant about the features available in Co11ab and asked them to configure the system based on the representations they preferred in the researcher-facilitated session. The visually impaired participant shared their screen through Zoom (without computer sound) and recorded screen reader utterances locally on their computer (except Ian who was unable to record screen reader audio due to technical issues). Participants' main task was to use Google Docs with Co11ab to prepare a blog post on "top things to do when moving to a new home." We chose this prompt as a topic on which most participants would have some experience or understanding. We gave participants a starter Google Doc with a couple paragraphs. The paragraphs were seeded with comments, suggested edits, and a few grammatical errors. We asked participants to generate 3-4 additional points with a short description (2-3 sentences) for each point, rearrange the points in decreasing order of importance, and add a brief introduction and a conclusion summarizing and highlighting the main ideas. We provided these prompts as pointers to get them started to generate content together and explore how they might use different features of Co11ab in their work. As such, we did not measure participants' success or failure in accomplishing these tasks and did not enforce the use of any features. We also did not require participants to memorize keystrokes and offered reminders as needed throughout the study.

The sessions ended with a one-on-one debrief interview with the visually impaired participant in a Zoom breakout room, where we asked them to share their thoughts on using Co11ab to write with their sighted collaborator. We probed for any changes and challenges during the writing session compared to their regular collaboration practices and how they might customize the auditory features moving forward. We collected the sighted participants' feedback through a short survey. The sessions lasted for 100-120 minutes. All participants received US$60 gift cards each.

6.4 Data Analysis

All sessions were recorded via audio and screen capture and transcribed for analysis. We analyzed observational and interview data collected from the researcher-facilitated sessions alongside the naturalistic sessions of use. Our approach to data analysis follows reflexive thematic analysis [17, 18], which involved a process of open and selective coding, memoing, and iterative comparison of data to emerging themes. We first analyzed our data to understand participants' reactions to the specific design features and audio cues within Co11ab. Next, we analyzed the ways in which participants used Co11ab during real-time collaborative writing. In particular, we performed micro-analysis of video and screen recordings [10] attending to participants' moment-to-moment interactions and how they established shared meaning with each other [36, 72]. While our analysis was informed by theories of collaboration awareness (e.g., [5, 30, 38, 73]), we maintained analytic flexibility to understand the unique experiences of screen reader users and the work of collaborative writing. Finally, our approach to evaluation and data analysis is informed by the political/relational model of disability [41] and other feminist disability scholarship [8, 31, 46]. Drawing on this theory, we focus on how our participants make sense of new technologies as part of creating access and how such technology may reshape their collaborative work and the labor of access rather than focusing on task performance data [9].

7 DESIGN EVALUATION: FINDINGS

Below we present findings from the researcher-facilitated and naturalistic writing sessions. We first discuss how participants understood and reacted to the design of audio cues and representations available in Co11ab. Second, we detail the ways in which participants used Co11ab to support real-time collaborative writing.

7.1 Design of Audio Cues and Representations

One of our primary goals was to understand whether or not the non-speech auditory representations introduced in Co11ab are intuitive and easy to understand for screen reader users. Below we detail how participants understood and reacted to Co11ab's features and suggested areas of improvement.

7.1.1 Query Location, Follow Mode, and Jump-to-Location. On the whole, participants appreciated the spoken feedback announcing the page, paragraph, and line number when they queried a collaborator's position or jumped to their location. Three participants (Dan, Kate, Kirk), however, felt that this announcement lacked important contextual information. Kirk explained, *"Paragraph four line four tells me absolutely nothing... Particularly because Google Docs doesn't have line numbers for screen reader users... So, if you just add a couple of words from that line into that announcement, I think it's a lot more clear where in the document you were just jumped to."* Based on this suggestion, we updated the extension before conducting the naturalistic writing sessions so that the screen reader would read the content of the new line after a cursor jump.

Participants also wanted the follow mode to have 'verbosity' settings in which they could specify how and with what level of detail collaborators' actions are announced. For example, some felt that listening to what a coauthor is adding in the document line by line (i.e., the current implementation) was helpful because *"I don't want to hear as you are typing what is happening, because you could be deleting and adding things and it will be confusing"* (Ron). Others (Alex, Dan, Gina, Kirk, Ryan), however, felt that they might need to hear what their coauthors are typing by word or by character depending on specific close co-editing scenarios.

The current implementation of the follow mode also does not allow screen reader users to actively type or move their cursor in the document unless they intentionally turn this mode off or take their cursor to the coauthor's location. This implementation separates 'following' from 'editing' in that a user cannot do both at the same time. Four participants (Cory, Kirk, Ian, Kate) mentioned that this was useful *"because it's going to be very confusing [if] I'm writing and I hear the words that I type, and then, all of a sudden I hear other types of word, but it's from the person that I was following"* (Ian). However, Dan and Jim shared that they might need to listen to what others are typing and simultaneously edit on their own, especially if they want to *"wordsmith the sentence as they're (coauthors) typing it... I could on the fly more easily correct them, or give them hints or guide them in the right direction."* Jim, Kate, and Sean suggested using different voices to differentiate a coauthor's edits from their own screen reader feedback (i.e., voice coding [27]), although spatial audio [50, 76] and secondary whisper [65] might be useful as well.

7.1.2 Relative Proximity. While sighted writers often have access to visual markers of where others are editing within a shared document, Co11ab provides such support through relative proximity notifications that use loudness and pitch change with earcons (i.e., musical notes), an auditory icon (i.e., typing sound), and spoken announcements. Overall, participants (except Ryan and Alex) preferred non-speech audio for relative proximity over spoken announcements because of conciseness and reduced verbiage. Joy explained that spoken feedback can *"cause me very quickly to lose my train of thought and be kind of annoyed at the whole situation."* Importantly, some participants who preferred non-speech audio (Joy, Kate, Neil) still felt that they would need a more *"discordant"* or *"jarring"* notification specifically when coauthors' cursors collide, such as *"two notes that are not in harmony"* (Joy).

Among participants who preferred non-speech audio cues, everyone except Ron preferred earcons. Six (Joy, Kate, Cory, Neil, Kirk, Leah) reported that the earcons were easier to differentiate while performing collaborative writing than the auditory icon because the typing sound was *"too similar to other noises in my environment"* (Joy) and that *"I almost could mistake it for you pushing keys on your keyboard"* (Cory). While earcons were preferred to the auditory icon, multiple people mentioned issues with the *"learning curve"* (Jim) and importance of being able to assign certain instruments to frequent collaborators so that they could learn them over time.

In terms of mapping proximity, everyone except four participants (Ryan, Mia, Alex, Ron) understood without prior explanation that the loudness and pitch increased when coauthors came closer to their cursor location and decreased when they moved farther away. Four participants (Kate, Jim, Mia, Leah) indicated that the inverse relationship with distance was more *"realistic"* for the change in loudness than pitch. However, the change in pitch was preferred by everyone except Jim and Sean (and Mia who wanted a combination of both), because participants felt that they had to concentrate harder to *"pick up on the subtlety of the volume changes"* (Cory) than the changes in pitch. Others said that volume changes will be more difficult to detect with other audio (e.g., music) in the background (Dan) and that it will depend on the global volume settings of their computer (Kate). Additionally, three participants (Alex, Ron, and Kirk) pointed out that the current representation of relative proximity in Co11ab lacked directional information such that *"I don't necessarily know...if you're moving down towards where I am or you're moving up towards where I am"* (Alex). While some participants wanted Co11ab to encode direction of collaborator movement (e.g., using pitch, loudness, or spatial audio), Dan and Kate cautioned against creating *"information overload"* with extraneous audio effects. Relatedly, to minimize auditory overload, some participants felt that relative proximity notifications should be automatically paused while the follow mode or audio scrollbar is on, since writers would not need to be aware of both at the same time. We updated Co11ab based on this feedback prior to the naturalistic writing sessions.

7.1.3 Audio Scrollbar. Overall, the audio scrollbar aligned with participants' mental models in that most participants could anticipate that higher pitch corresponded to higher amount of comments/edits and that the repetition of earcons indicated the number of comments/edits. Six participants (Jim, Cory, Ryan, Leah, Kirk, Sean) preferred the use of repeated earcons, because it gave *"precise information"* whereas pitch change provided relative information about which paragraph (or line) has more or less comments/edits. Participants felt that with pitch change, *"You might just get confused... Your estimates wouldn't be accurate depending on the range"* (Mia). Nevertheless, six participants (Kate, Gina, Dan, Ron, Neil, Mia) preferred earcons with varied pitch, because they were *"concise"* and better for *"at a glance quickness"* than repeated earcons which could get *"very noisy"* in a document with extensive comments/edits. Several participants who are proficient in music and programming (Dan, Kirk, Ron, Neil) thought that a wider range of pitch with a *"sliding scale"* (e.g., incremental change by a half or full note) can be applied to denote the specific number of comments/edits. Some others (Kirk, Cory, Joy, Neil), however, suggested an encoding approach that would combine pitch variation with earcon repetition.

Despite appreciating the concept of getting an auditory overview of how comments, edits, and cursors are distributed in the document, several participants (Joy, Jim, Dan, Kate, Kirk) shared that they expected a different navigational structure for the audio scrollbar. They envisioned the audio scrollbar to be a *"skim tool"* such that they could hear audio cues for the number of comments, edits, or cursors in a paragraph upon invoking the default navigational commands they use for skim reading in their regular work (e.g, Control + Down arrow to move to the next paragraph). This way, they could hear the "context" (i.e., first few words) of a paragraph or a line while also getting an idea of the magnitude of collaborative activities within that region. While the current hierarchical implementation of the audio scrollbar allows a screen reader user to take their cursor to a paragraph they want to explore in more detail, participants felt that combining the scrollbar audio cues with default navigation would further streamline their work processes.

7.2 Coordinating Activity during Real-Time Collaborative Writing

Our analysis of researcher-facilitated and naturalistic writing sessions revealed five key ways participants used Co11ab to coordinate their actions and collaboratively edit a shared document: (1) monitoring a coauthor's edits in real-time, (2) directing a coauthor's attention to specific edits, (3) avoiding cursor collisions and conflicting edits, (4) switching between individual work and monitoring edits, and (5) quickly reviewing document edits and comments.

7.2.1 Monitoring a Coauthor's Actions in Real-Time. Analysis of observational and interview data reveals that the follow mode and jump-to-location feature were useful in monitoring and understanding a collaborator's actions while working synchronously. Participants felt that they could track their coauthors' edits in real-time using follow mode and jumping to their location when necessary without needing to *"wait until they were done and reread the whole thing"* (Leah). We observed an example of this during the naturalistic writing session involving Jim (screen reader user) and Ava (sighted). Midway through the session, they brainstormed about a new point and planned how they would add it to the document. Jim suggested, *"Do you want to write this one, and I'll kind of expand it a little bit?"* As Ava started writing, Jim turned on the follow mode to monitor Ava's new edits in real-time. While doing so, he simultaneously provided verbal feedback to Ava. He said, *"I'm reading what you have. That's really good."* Jim later compared this experience with his regular work on Google Docs: *"I was...able to follow her typing when she was typing points. That was more feedback than I had ever received with other- I mean, using Google Docs normally, cause I wouldn't be able to actually see where she's editing."*

During the writing session with Ava, we also observed that Jim developed a routine of manual tracking instead of entirely relying on the follow mode. This was because during the follow mode, the screen reader remained silent if Ava's cursor was static at a certain point, causing confusion about her actions, which needed verbal clarification (e.g., Jim once said, *"Are you done typing, Ava? I see it's not [reading]."*) In such cases, Jim sometimes jumped to Ava's location and manually monitored her edits by moving his cursor back and forth or up and down. This manual following was particularly useful when Jim made minor revisions on-the-fly as Ava was typing instead of waiting for her to complete adding content and reviewing it later.

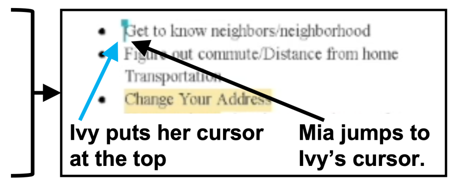
Overall, participants felt that the ability to monitor a collaborator's edits in real-time through Co11ab would make their collaborative work processes *"easier and efficient"* (Ron). They emphasized that being aware of coauthors' real-time activities would simplify their current work routine, which generally follows an asynchronous hands-off style. Joy said, *"It means that we don't have to keep this paper trail of passing the document back and forth... or an absolute nightmare, we'll have 10 or 20 emails in a chain about a two-page document."* While working at the same time on a shared document, participants felt that querying a coauthor's location and jumping to their location could also reduce their reliance on external verbal communication for low-level details [38] and thus, help them achieve common ground with the least collaborative effort [21]. Neil said, *"I would constantly have to ask people, oh, who added what, where, and then I would be like, find it with my screen reader and then manually track it down and see the exact lines that got changed. So it was a very tedious process. And I think this takes that whole tedious concept out of that workflow."* Jim felt that being aware of coauthors' real-time activities without extensive verbal communication could be especially useful in time-sensitive projects or in situations where simultaneous external communication is not possible. Others also shared that the follow mode would be particularly useful in *"brainstorming meetings... [where] someone's taking notes and everyone can see them"* (Kate) or in *"pair writing where someone dictates and the other person writes or someone doodles and the other person sort of gives an indication of what they think of that"* (Kirk). Neil further explained how the follow mode and jump-to-location feature could support collaborative programming: *"If I was working with a programmer... I would want to have that more granular information, just to kind of understand what exactly in the code they're changing in detail."* For screen reader users, engaging in such pair writing, programming, or meeting scribing activities is typically difficult, since they cannot monitor others' edits in real-time [58].

7.2.2 Directing a Collaborator's Attention to Specific Edits. In addition to actively monitoring a collaborator's actions in real-time, writers may periodically need to call a collaborator's attention to specific sections of a document. Participants explained that the follow mode and the jump-to-location feature could improve their coordination dynamics with their sighted colleagues, who might provide ambiguous verbal guidance to a location of interest or become potentially "annoyed" with the tedious workarounds. Mia explained: *"Blind people... they might be more specific with you about, 'Oh, find this exact text that I'm at', whereas a sighted person, they don't usually do that. They tell you where to go more based off of distance... They're using more spatial language... might use colors, like 'Ohh, just I'm in the paragraph in blue'... So maybe it'd be even more helpful with sighted [people] than with blind people."*

Indeed, we observed multiple instances of participants using the follow mode and jump-to location features in tandem to direct a collaborator's attention to an area of interest. For example, Mia (visually impaired) and Ivy (sighted) demonstrated the process of coordination and directing attention through cursor placement during the naturalistic writing session. They employed a reactive (i.e., joint) writing style throughout the session where they brainstormed together about the content to be added and closely read, reviewed, and revised each other's edits [47]. In this vignette (Table

Table 2: Ivy (sighted) directs Mia's (screen reader user) attention to specific edits. SR: Screen reader speech

1		*(Ivy and Mia are talking about reordering points)*
2	Ivy:	I just moved like the neighborhood one first. Then the commute one second above, like the ones that are already written.
3	Mia:	... Wait, where are you putting this?
4	Ivy:	I put them above like-
5	Mia:	The paragraphs?
6	Ivy:	The paragraphs. Yeah.
7		*(Mia moves her cursor to find where Ivy has put the list of points)*
8	Mia:	Wait, but oh- so confusing. But the first thing in here is her paragraph. Not the list of stuff-
9	Ivy:	The first comment is on change your address, which is the first line of the blog post that was already here.
10	Mia:	Wait, let me go to where your cursor is.
11	Ivy:	My cursor. Yeah, I'll put it at the top. *(Ivy puts her cursor at the top)*
12		*(Mia jumps to Ivy's location by pressing shortcut)*
13	SR:	`Moved cursor to Ivy's location on page one paragraph one and line one. List bullet. Get to know neighbors slash neighborhood..`
14	Mia:	Ohh, yeah, I see what you did.



2), Ivy and Mia are talking about reordering points in a list. Mia becomes confused about where Ivy added one section of text (line 3). Ivy begins to explain verbally but then senses Mia's confusion and places her cursor at the point of interest (line 11) so that Mia knows the exact location where the change occurred. Co11ab then announces the interaction, which specifies that Mia moved to Ivy's location, the precise page/paragraph/line location, and reads part of the content at that location. Mia then confirms that she understands the change (line 14). The success of this interaction depends on both the system's features and their use by the dyad. Rather than relying on a verbal description of the cursor location and the change, Ivy can put her cursor to a point and rapidly direct Mia's attention to the section of interest. What's more, Co11ab provides not only the specific location of the change but contextual information about what's at that location. Thus, the collaborative processes of grounding [21] and joint attention [61, 70] are enacted through both the availability of technological features and their use during interaction. For a screen reader user, a collaborator's cursor, which was previously ambiguous and difficult to detect, now becomes a powerful mechanism for pointing, directing attention, and establishing joint understanding.

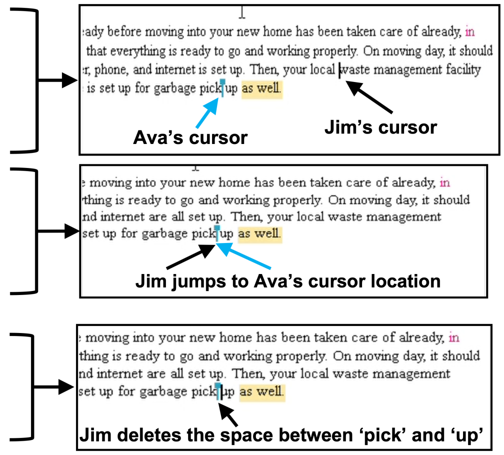
To further understand how participants used the jump-to-location feature to direct attention to specific edits, we turn to another vignette from Jim (visually impaired) and Ava (sighted). In this example, when Jim heard an earcon with medium loudness (for paragraph level), he became aware that Ava had joined his paragraph and said *"I see you"* (Table 3, line 5). The earcon prompted Jim to explore exactly where Ava was in the paragraph and what she was doing. He jumped to Ava's cursor location and said *"Okay, so..."* (Table 3, line 9). Prompted by this verbal signal along with visually monitoring Jim's cursor movement, Ava drew Jim's attention to a specific phrase where she had her cursor. Then Ava said, *"'Pick up' could be one word"* (Table 3, line 10). Jim explored the particular phrase by moving his cursor back and forth by a few words and deleted the space between 'pick' and 'up' (Table 3, lines 11-15). This example illustrates how Ava was able to call Jim's attention to a particular word and Jim could quickly navigate to the place in the document where Ava was attending. What's more, Ava was able to verbally suggest the edit without making the edit for Jim. Thus, the auditory cues and navigation features introduced in Co11ab helped Jim and Ava to swiftly initiate shared attention [70] and achieve common ground [21].

After the session, Jim emphasized that he never needed to ask Ava for guidance on her location. He explained, *"It was kind of very, very cohesive and very, very, very smooth, which is, pre-collaborative writing extension wasn't so much the case with Google docs. We got to focus a little bit more on bouncing ideas than '...Where are you now? Okay. Here I am' kind of thing... So it left a lot more room for actual discussion, brainstorming, and writing."* In this example, we see that Co11ab not only provides a mechanism for grounding and coordinating joint attention, it fundamentally reshapes interaction. Instead of spending extensive time on verbally sharing low-level details required for reaching a shared understanding of the area of interest, the dyad is able to spend more time on higher-level activities that are important for effective collaborative writing, such as discussion and brainstorming [47, 60].

7.2.3 Avoiding Cursor Collisions and Conflicting Edits. As our formative findings show, screen reader users currently have a limited sense of how close they are to their coauthors in a document. Without understanding their position relative to others, they may inadvertently collide and make conflicting edits to the same portion of text. Joy, Kate, and Sean explained that cursor collisions are more likely to occur when they work with other blind coauthors, since screen reader users must move their cursor to read the content unlike sighted people: *"I feel like it's less likely that I'm going to bump into a sighted collaborator because they don't have to move their cursor to read, where I assume I'm more likely to run into another blind user"* (Joy). Participants reported that relative proximity notifications would help them anticipate situations when their cursor might collide with their coauthors' cursors so that they can continue *"working on my things without having to worry*

Table 3: Jim (screen reader user) jumps to Ava's (sighted) location to review a particular text

1		*(Jim's cursor is on line 3 of the paragraph)*
3		*(Ava places cursor on the phrase 'pick up' in line 4 of Jim's paragraph.)*
4		*(SR emits an earcon with paragraph level loudness; then continues reading line 3)*
5	Jim:	I see you.
6	Jim:	Okay, let's see-
		(Jim presses shortcut to jump to Ava's location)
7		*(Jim quickly moves his cursor up to line 3 and down to line 4)*
8		*(SR is reading line 4)*
9	Jim:	Okay, so-
10	Ava:	Pick up could be one word.
11		*(Jim moves cursor back and forth by word on line 4)*
12		*(SR reads content of line 4 word by word)*
13		*(Jim's cursor is just at the beginning of the word 'up')*
14	Jim:	I think- let me see-
15		*(Jim deletes the space between 'pick' and 'up')*



about stepping on their toes" (Joy). Gina added, *"It would help avoid that kind of conflict where maybe I'm erasing something and you're adding something and I'm just erasing what you're adding."*

We observed Mia (visually impaired) do this during the naturalistic writing session with Ivy (sighted). Mia had her relative proximity notifications off at the beginning. Later, when they started reorganizing the list items and making direct edits in the document, both attempted to fix the same grammar error at one point. Although they coordinated their actions and avoided cursor collisions through verbal discussion, this prompted Mia to turn on the relative proximity notifications so that she could anticipate potential conflicting edits moving forward. Similarly, after using Co11ab with his sighted colleague (Ava), Jim said the cues he received about Ava's location helped him *"to safely work on different stuff... because I knew exactly where she was editing... I feel like, okay, she's working on this. I can work on this above it, below it, or I can work on a conclusion and I can make... what normally sighted collaborators would see as snap decisions to kind of speed up the work."*

Given the lack of awareness cues for screen reader users in current collaborative writing tools, many participants described adopting a divide-and-conquer writing style in their regular work (i.e., splitting up aspects of writing or areas of the document). Ian (visually impaired) and Lily (sighted) also used a divide-and-conquer approach during their naturalistic writing session, where they discussed and identified which paragraphs each would edit at the beginning of the session. For most of the writing session, Ian and Lily worked independently in different paragraphs and maintained spatial separation between their work areas. Nevertheless, Ian explained that he wanted to remain aware of where Lily was working: *"I was typing my part and she was typing hers, but I wanted to know how close she was or how far she was... And the (musical) notes I noticed when she was getting closer... that was helpful."* We observed multiple instances where their cursors collided, at which one or both participants quickly moved their cursor to maintain spatial separation. What's interesting here is that cursor collisions are not inherently problematic. Indeed, cursor collisions are key to the ways in which participants referenced areas of interest (Section 7.2.2). Instead, cursor collisions and the associated relative proximity notifications help build peripheral awareness [5, 35] about who is working in which sections of the document, which is critical information for avoiding unintentional concurrent or conflicting edits. What's more, the peripheral awareness gained through Co11ab has important implications for the screen reader user as well as the sighted collaborator and the overall group dynamic. Lily noted that Co11ab features helped Ian to focus on his own writing: *"He (Ian) was able to type faster and keep focused on his part of the document [compared to their regular work]."* Thus, having a better understanding of where collaborators are working may not only help to prevent conflicting edits, it may also enhance screen reader users' ability to contribute during collaborative writing.

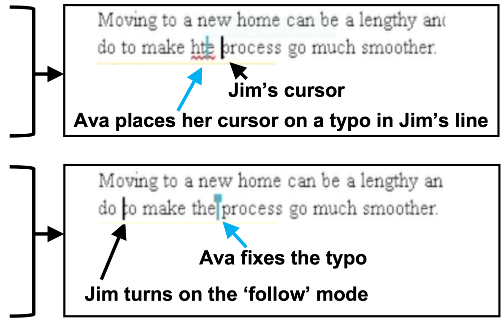
7.2.4 Switching between Individual Work and Monitoring Edits. Participants explained that the relative proximity cues provided a signal for knowing when to pause individual work and attend to a collaborator's actions [35, 39]. Dan said, *"When I heard that higher tone, that definitely caught my attention to- okay, wait a minute. What are you doing here? So I think as soon as I heard that, I would stop and then I would probably turn on the 'follow' to hear what you're trying to do... So it's kind of more like the intro to let me know that someone's in the area and then I can investigate further."*

We observed an instantiation of this during a naturalistic writing scenario involving Jim (visually impaired) and Ava (sighted). Midway through the session, Ava entered Jim's line to fix a typo (Table 4, line 3), triggering a high volume earcon for Jim. Realizing that Ava moved into the immediate space where he was working, Jim turned on the follow mode to observe Ava's edits and initiated a verbal conversation to clarify the type of actions Ava was performing. Once Ava clarified that she was only proofreading what Jim added (Table 4, line 9), Jim switched back to his own work and continued adding content to the document. Here we see that the proximity notifications provided peripheral awareness for deciding when to break focus on individual writing and more fully attend to a collaborator's actions [35, 39].

7.2.5 Quickly Reviewing Document Edits and Comments. Participants felt that the audio scrollbar would be most helpful for planning and prioritizing collaborative writing tasks in both synchronous and asynchronous scenarios. They shared that they could use the audio scrollbar to *"streamline my work"* (Ron), especially if *"I*

Table 4: Jim (screen reader user) turns on the follow mode to review Ava's (sighted) edits.

1		*(Jim's cursor is on line 2 of the paragraph)*
2		*(Ava places her cursor on a typo 'hte' in Jim's line)*
3		*(SR emits an earcon with line level loudness; then continues reading line 2)*
4	Ava:	Got it!
5		*(Ava changes 'hte' to 'the')*
6	Jim:	Got it? Ohh, great, great!
7		*(Jim turns on the 'follow' mode and promptly silences SR announcement)*
8	Jim:	Are you typing? Yeah, I think. Are you taking it from here?
9	Ava:	Ahh, I was just proofreading while you were-
10	Jim:	Ow, gotcha, gotcha.
		(Jim turns off the 'follow' mode)



didn't open the document in a while and then... [I] want to know exactly what happened since I left" (Ian). Participants said that the audio scrollbar would help them identify *"parts of the document that were sort of contested or under review"* (Alex) and allocate their time on tasks accordingly. Leah said, *"It's sort of like when you open a document and you can visually see what's going on and say, oh my gosh, I'm going to do this later, or I got time for this now... I can just quickly go through all the paragraphs and see, okay... how do I need to program my time for this?"* Furthermore, participants added that the audio scrollbar feature would be most useful while working on long, complex documents with a lot of coauthors and documents that people created on their own. Gina said, *"Particularly if I'm super familiar with a document and don't want to keep having to read and reread everything all the time... that will help me prioritize what I hit first."* Cory further explained how the audio scrollbar could make the process of "scanning" a document for collaboration information easier and efficient for screen reader users:

> *"Visually people could scan for different color changes [for comments or edits] or people could basically just move their mouse to whatever comment, whereas with a screen reader, you've got to do a little bit more work, whether it's going to the comments list or moving to that section and then looking for the comments. And it's still a little bit more tedious than if someone's visually scanning. And I feel like with the scrollbar, you get a little better feel for the overall audio picture without having to go and examine each part."*

During the naturalistic writing sessions, Jim and Ian each invoked the audio scrollbar at the very beginning to find the comments and edits in the sample document. Mia, on the other hand, did not use the scrollbar at all, since she and Ivy were discussing edits primarily through verbal exchange. Mia later explained, *"She (Ivy) was already giving me enough verbal info that I didn't need to have the meta information about the number of changes... but I feel like if I was working by myself, I might've used that feature (audio scrollbar) more."* In summary, participants felt that the audio scrollbar would be most useful for planning and time allocation as well as when writers are working on their own and may not have (or want) a sighted collaborator to provide a verbal overview.

8 DISCUSSION

Multiple decades of research in academia and industry have led to a proliferation of tools that support real-time distributed collaborative work. This body of work is rooted in empirical studies and theories that make implicit ability assumptions and neglect the experiences of ability-diverse teams. Designing more equitable collaborative technologies requires rethinking how collaboration awareness works and designing systems that support associated collaborative processes and informational needs. We now revisit our findings to discuss theoretical and practical future directions.

8.1 Revisiting Collaboration Awareness for Screen Reader Users

Creating more equitable and inclusive collaboration technologies requires revisiting the notion of collaboration awareness and how it has been conceived over the years. For example, Gutwin and Greenberg's collaboration awareness framework [38] asserts that people need to understand *who* is doing *what* and *where* to perform collaborative work in real-time. On the surface, these broad categories are still relevant to the experience of visually impaired writers. Indeed, our data show many examples of blind writers understanding each of these elements. Yet, the ways in which these understandings are achieved are fundamentally different due to both the experience of blindness and how the tools (e.g., screen readers, accessibility features) reconfigure collaborative meaning making. In Gutwin and Greenberg's framework [38], key informational elements for developing peripheral awareness and joint attention in real-time collaboration include collaborators' location (where are they working?), gaze (where are they looking?), and view (where can they see?). Our analysis, however, brings forth questions about the meaning of these core constructs of collaboration awareness: what does gaze or view (either their own or that of their collaborators) mean for a screen reader user? And, how do asymmetries in collaborators' access to and understanding of this information shape their work?

To date, the theories and frameworks of collaboration awareness have assumed symmetry in collaborators' ability to access the shared workspace in that they make use of the same (or similar) tools and process information in the same modalities while working together. Yet, the interfaces through which sighted and visually impaired collaborators generate and consume information on collaborative platforms (e.g., Google Docs, Microsoft Word) are fundamentally different; see Section 4. While sighted people make use of visual-spatial representations to navigate and perceive information in a document, visually impaired writers rely on screen reader speech output and keystroke-based navigation that are ephemeral, sequential, and linear [7] (although some also use braille). The notion of gaze and view depend on principles of visual gestalt perception [77], the ability to glance or scan, and persistent visual representations. For screen reader users, understanding where a collaborator is working or looking requires the system to provide precise and meaningful location information, sufficient context to understand that location, and then understanding or remembering the overall structure of the document—all of which are readily apparent for a sighted user looking at a graphical display. Similarly, understanding what a collaborator did in the document requires mental effort to revisit and remember changes or get an auditory overview of the document, whereas visual markup, color coding, and transparent overlays (i.e., comment spans) make this information immediately available and persistent for sighted users. Thus, more inclusive notions of collaboration awareness must attend to not only the availability of information but the effort and time required to access that information.

Further, collaboration awareness is not solely about the availability of information; it is also about how people understand and make use of that information during interaction. In addition to differences in information modalities, people have different mental models and ways of relating to artifacts that are shaped by disability experience. As our analysis reveals, there are fundamental differences in the way a sighted writer and a screen reader user understand the structure of a document and collaborators' locations. While sighted collaborators may use spatial (e.g., "bottom of the page") or color-coded (e.g., "paragraph in blue") descriptions to refer to an area of interest, these visual-spatial references do not always align with a screen reader user's understanding of a document. Even the same collaborative action may have different meanings to a screen reader user. As an example, for sighted people, jumping to a coauthor's location means bringing the coauthor's editing area (where they have their cursor positioned) into the field of view. For screen reader users, however, this means putting their own cursor at the coauthor's cursor location, either through verbal guidance, manually moving from paragraph to paragraph to find the coauthor's position, or in the case of Co11ab, using the jump-to-location shortcut. Collaborative tools and frameworks that do not understand such asymmetries will fall short in helping collaborators achieve common ground [21] and coordinate joint attention [70]. At best, these failures require additional time and effort by both collaborators to resolve, and at worst, they lead to systemic exclusion and a blind writer feeling their voice has *"become the small voice"* (Neil).

A deeper understanding of inclusive collaboration requires recognizing the implicit ability assumptions and asymmetries in collaboration awareness. At its core, Co11ab aims to help shift this imbalance, and our analysis shows signs of that happening. Consider the example where Mia directed her attention to a specific text that Ivy was referring to by jumping to Ivy's cursor location. The jump-to-location feature enabled them to initiate joint attention to a point of interest similar to the way sighted dyads can use gaze visualizations to attend to the same object [24, 42, 61]. In another instance, we observed how relative proximity alerts made Jim aware of Ava's presence nearby, which further prompted him to momentarily pause his own writing and follow what Ava was doing. Thus, the relative proximity cues along with the follow mode helped Jim manage coupling [64] in that he could determine an appropriate moment to transition from an individual task to a tightly-coupled shared activity (i.e., reviewing collaborator's edits) [35, 39]. Furthermore, the audio scrollbar introduced a way for participants to quickly 'scan' documents for highly contested areas and develop an overview of the document state so that they can plan and prioritize their course of actions accordingly [38]. In each of these examples, interaction depended on both the availability of technology resources and human collaborators who enact more accessible practices together. Thus, rethinking collaboration awareness requires attending to how able-bodied collaborators make use of accessibility features and bring them into interaction as well as how new technologies make new ways of working possible.

8.2 Design Considerations for Accessible Collaborative Writing

Through the design and iterative user exploration of Co11ab, our study illustrates how auditory representations shape screen reader users' workflow and ability to make sense of collaborative information in real-time. Below we discuss the trade-offs and considerations of various auditory designs for collaboration information.

8.2.1 Compromising learnability with discernibility. In designing several auditory features of Co11ab, we incorporated a skeuomorphic design approach where an action or interface is mapped on to its real-world counterpart [67]. These features include the auditory icon that imitated the real-world action of typing and increasing loudness of earcon or auditory icon as coauthors' cursors come closer. Indeed, our participants found these designs to be realistic representations of collaborators' movements and more intuitively understandable and learnable than abstract representations such as earcons (i.e., musical notes) and changing pitch [4, 34]. However, participants also pointed out that the auditory icon (i.e., typing sound) and change in loudness are less discernible amidst other audio streams in the surroundings and contingent upon one's workspace setup (e.g., global volume settings or background music). The typing sound, in particular, can even be confused with typing activities of one's own or others. In contrast, while earcons and pitch change need explicit learning and memorizing [32, 34], most participants preferred them as they found it easier to distinguish and thus more likely to provide them with the peripheral awareness of when coauthors are working close to their space.

8.2.2 Needing disruption to avoid conflicts. Echoing prior work [27], our analysis showed that participants preferred non-speech audio cues over spoken announcements for relative proximity since the latter was more disruptive to their own reading, writing, and *"train of thoughts."* Interestingly though, in real-time collaborative writing scenarios, there are certain instances when disruption to individual work is desired. For example, participants shared that they would need more *"discordant"* alerts when coauthors' cursors are very close to their cursor so as to warn them against potential conflicting edits. Cursor collision is unlikely to happen in asynchronous collaboration but is an important consideration for developing peripheral awareness in synchronous collaboration. Furthermore, screen reader users need to be able to customize auditory cues based on the task at hand at a particular instance, such as switching to a more disruptive representation of proximity cues when they are working on a heavily edited area with a lot of collaborators and turning off alerts when working in a region on their own or not making active edits.

8.2.3 Balancing efficiency with specificity. Our analysis revealed that while processing collaboration information, screen reader users need to balance efficiency with specificity. As an example, participants shared mixed reactions towards the representations of the audio scrollbar. While some preferred repeated earcons because it provided specific information about the number of comments, edits, and active coauthors within a certain threshold, others preferred pitch change because of conciseness. The same reasoning holds for spoken announcements and non-speech audio cues. Although spoken announcements are more specific than non-speech audio cues, all but Alex and Ryan preferred the latter because of the *"economy of sound that you don't necessarily get with words"* (Kate). Importantly though, the efficiency gained from earcons and pitch change comes with the compromise in learnability, as discussed above.

8.2.4 Managing cognitive overload across simultaneous activities. Multitasking or paying attention to multiple streams of auditory information can be complex and cognitively overwhelming for screen reader users [76] due to the ephemeral and linear nature of screen reader speech [7]. Our design exploration with Co11ab illustrated that minimizing extraneous information that is not relevant to the particular task at hand is important to control such information overload [51]. For example, participants felt that relative proximity alerts should be automatically turned off while the follow mode or the audio scrollbar is on, because proximity information is not needed when monitoring a coauthor's edits or getting high-level overview. Similarly, some participants thought that following and editing should be toggleable such that they do not hear what others are typing when they are editing on their own. However, certain styles of real-time writing (e.g., reactive writing [47]) requires writers to quickly switch back and forth between reviewing and active revising (e.g., fixing typos). This means that screen reader users may want to 'follow' a coauthor while simultaneously editing on their own, as we observed Jim doing during the naturalistic writing session. In such cases, collaborative systems can use audio manipulation techniques such as voice coding [27], concurrent speech [37], spatial audio [50, 76], and secondary whisper [65] to distinguish between the coauthor's edits and own screen reader typing echo.

9 CONCLUSION

This study set out to investigate the ways in which visually impaired writers perform synchronous writing with sighted collaborators and develop novel auditory technologies to enhance their collaborative writing experience. Grounded in our formative interviews and remote observations with eight screen reader users, we built Co11ab, a Google Docs extension that uses various spoken and non-speech audio cues to support screen reader users in monitoring others' real-time activities, avoiding concurrent edits, and developing a high-level overview of collaboration information in a shared document. Our design evaluation study involving screen reader users, in fifteen researcher-facilitated and three naturalistic writing sessions with known sighted collaborators, illustrated how participants used Co11ab to coordinate joint attention, maintain peripheral awareness, and transition fluidly between individual and shared tasks with their collaborator. These insights also encourage thinking differently about collaboration awareness, particularly considering the asymmetries in collaboration experience among ability-diverse team members and bring forth practical considerations and tradeoffs for the design of accessible collaborative systems.

ACKNOWLEDGMENTS

This work was supported by NSF grant IIS-1901456 and a graduate research grant from Northwestern University. We thank our participants for their contributions to the study. We also thank Bryan Pardo, Cynthia Bennett, Marcelo Worsley, Abir Saha, and members of the Inclusive Technology Lab and CollabLab for their feedback at various points of this research.

REFERENCES

[1] 2020. Google quietly reached a major milestone in its battle with Microsoft to rule the workplace. https://www.businessinsider.com/google-g-suite-gmail-2-billion-vs-microsoft-office-2020-3 Retrieved September 6, 2021.

[2] Ádám Csapó and György Wersényi. 2013. Overview of Auditory Representations in Human-machine Interfaces. *ACM Computing Surveys (CSUR)* 46, 2, Article 19 (December 2013), 23 pages. https://doi.org/10.1145/2543581.2543586

[3] Dena Al-Thani, Tony Stockman, and Anastasios Tombros. 2015. The Effects of Cross-modal Collaboration on the Stages of Information Seeking. In *Proceedings of the XVI International Conference on Human Computer Interaction* (Vilanova i la Geltru, Spain) *(Interaccion '15)*. ACM, New York, NY, USA, Article 54, 8 pages. https://doi.org/10.1145/2829875.2829925

[4] T. S. Amer and Todd L. Johnson. 2018. Earcons Versus Auditory Icons in Communicating Computing Events: Learning and User Preference. *International Journal of Technology and Human Interaction* 14, 4 (2018). https://doi.org/10.4018/IJTHI.2018100106

[5] Ronald M. Baecker, Dimitrios Nastos, Ilona R. Posner, and Kelly L. Mawby. 1993. The User-Centered Iterative Design of Collaborative Writing Software. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems* (Amsterdam, The Netherlands) *(CHI '93)*. ACM, New York, NY, USA, 399–405. https://doi.org/10.1145/169059.169312

[6] Matthew J. Baker, E. M. Nightingale, and Suzy Bills. 2021. An Editing Process for Blind or Visually Impaired Editors. *IEEE Transactions on Professional Communication* 64, 3 (2021), 275–287. https://doi.org/10.1109/TPC.2021.3090722

[7] Mark S. Baldwin, Jennifer Mankoff, Bonnie Nardi, and Gillian Hayes. 2020. An Activity Centered Approach to Nonvisual Computer Interaction. *ACM Transactions on Computer-Human Interaction (TOCHI)* 27, 2, Article 12 (March 2020), 27 pages. https://doi.org/10.1145/3374211

[8] Cynthia L. Bennett, Erin Brady, and Stacy M. Branham. 2018. Interdependence As a Frame for Assistive Technology Research and Design. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility* (Galway, Ireland) *(ASSETS '18)*. ACM, New York, NY, USA, 161–173. https://doi.org/10.1145/3234695.3236348

[9] Cynthia L. Bennett, Daniela K. Rosner, and Alex S. Taylor. 2020. The Care Work of Access. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376568

[10] Jeff Bezemer and Diane Mavers. 2011. Multimodal transcription as academic practice: a social semiotic perspective. *International Journal of Social Research Methodology* 14, 3 (2011), 191–206. https://doi.org/10.1080/13645579.2011.563616

[11] Jeremy Birnholtz and Steven Ibara. 2012. Tracking Changes in Collaborative Writing: Edits, Visibility and Group Maintenance. In *Proceedings of the ACM 2012 Conference on Computer Supported Cooperative Work* (Seattle, WA, USA) *(CSCW '12)*. ACM, New York, NY, USA, 809–818. https://doi.org/10.1145/2145204.2145325

[12] Jeremy Birnholtz, Stephanie Steinhardt, and Antonella Pavese. 2013. Write Here, Write Now!: An Experimental Study of Group Maintenance in Collaborative Writing. In *Proceedings of the SIGCHI Conference on Human Factors in Computing*

Systems (Paris, France) *(CHI '13)*. ACM, New York, NY, USA, 961–970. https://doi.org/10.1145/2470654.2466123

[13] Tom Boellstorff, Bonnie Nardi, Celia Pearce, and T. L. Taylor. 2013. Words with Friends: Writing Collaboratively Online. *Interactions* 20, 5 (September 2013), 58–61. https://doi.org/10.1145/2501987

[14] Katya Borgos-Rodriguez, Maitraye Das, and Anne Marie Piper. 2021. Melodie: A Design Inquiry into Accessible Crafting through Audio-Enhanced Weaving. *ACM Transactions on Accessible Computing (TACCESS)* 14, 1, Article 5 (March 2021), 30 pages. https://doi.org/10.1145/3444699

[15] Stacy M. Branham and Shaun K. Kane. 2015. Collaborative Accessibility: How Blind and Sighted Companions Co-Create Accessible Home Spaces. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. ACM, New York, NY, USA, 2373–2382. https://doi.org/10.1145/2702123.2702511

[16] Stacy M. Branham and Shaun K. Kane. 2015. The Invisible Work of Accessibility: How Blind Employees Manage Accessibility in Mixed-Ability Workplaces. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility* (Lisbon, Portugal) *(ASSETS '15)*. ACM, New York, NY, USA, 163–171. https://doi.org/10.1145/2700648.2809864

[17] Virginia Braun and Victoria Clarke. 2006. Using Thematic Analysis in Psychology. *Qualitative Research in Psychology* 3, 2 (2006), 77–101.

[18] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health* 11, 4 (2019), 589–597. https://doi.org/10.1080/2159676X.2019.1628806

[19] Albert Bregman. 1990. Auditory Scene Analysis: The Perceptual Organization of Sound. *Journal of the Acoustical Society of America* (1990). https://doi.org/10.1121/1.408434

[20] Maria Claudia Buzzi, Marina Buzzi, Barbara Leporini, Giulio Mori, and Victor M. Penichet. 2014. Collaborative Editing: Collaboration, Awareness and Accessibility Issues for the Blind. In *Proceedings of the Confederated International Workshops on On the Move to Meaningful Internet Systems: OTM 2014 Workshops - Volume 8842*. Springer-Verlag, New York, NY, USA, 567–573. https://doi.org/10.1007/978-3-662-45550-0_58

[21] Herbert H. Clark and Susan E. Brennan. 1992. Grounding in communication. In *Readings in Groupware and Computer Supported Cooperative Work: Assisting Human-Human Collaboration*, Ronald Baecker (Ed.). Morgan Kaufmann.

[22] Michael Connolly, Christof Lutteroth, and Beryl Plimmer. 2010. Document Resizing for Visually Impaired Students. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction* (Brisbane, Australia) *(OzCHI '10)*. ACM, New York, NY, USA, 128–135. https://doi.org/10.1145/1952222.1952248

[23] Clare Cullen and Oussama Metatla. 2019. Co-designing Inclusive Multisensory Story Mapping with Children with Mixed Visual Abilities. In *Proceedings of the 18th ACM International Conference on Interaction Design and Children* (Boise, ID, USA) *(IDC '19)*. ACM, New York, NY, USA, 361–373. https://doi.org/10.1145/3311927.3323146

[24] Sarah D'Angelo and Darren Gergle. 2018. An Eye For Design: Gaze Visualizations for Remote Collaborative Work. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–12. https://doi.org/10.1145/3173574.3173923

[25] Maitraye Das, Katya Borgos-Rodriguez, and Anne Marie Piper. 2020. Weaving by Touch: A Case Analysis of Accessible Making. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. ACM, New York, NY, USA, 1–15. https://doi.org/10.1145/3313831.3376477

[26] 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. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW, Article 191 (Nov. 2019), 26 pages. https://doi.org/10.1145/3359293

[27] Maitraye Das, Anne Marie Piper, and Darren Gergle. 2022. Design and Evaluation of Accessible Collaborative Writing Techniques for People with Vision Impairments. *ACM Transactions on Computer-Human Interaction (TOCHI)* 29, 2 (2022), 42 pages. https://doi.org/10.1145/3480169

[28] Maitraye Das, John Tang, Kathryn E. Ringland, and Anne Marie Piper. 2021. Towards Accessible Remote Work: Understanding Work-from-Home Practices of Neurodivergent Professionals. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1, Article 183 (apr 2021), 30 pages. https://doi.org/10.1145/3449282

[29] Susan L. Denham and István Winkler. 2014. Auditory perceptual organization. In *Oxford Handbook of Perceptual Organization*, J. Wagemans (Ed.). Oxford University Press, Chapter 29.

[30] Paul Dourish and Victoria Bellotti. 1992. Awareness and Coordination in Shared Workspaces. In *Proceedings of the 1992 ACM Conference on Computer-supported Cooperative Work* (Toronto, ON, Canada) *(CSCW '92)*. ACM, New York, NY, USA, 107–114. https://doi.org/10.1145/143457.143468

[31] Elizabeth Ellcessor. 2016. *Restricted Access: Media, Disability, and the Politics of Participation*. NYU Press. http://www.jstor.org/stable/j.ctt18040rg

[32] Euan Freeman, Graham Wilson, Dong-Bach Vo, Alex Ng, Ioannis Politis, and Stephen Brewster. 2017. Multimodal Feedback in HCI: Haptics, Non-Speech Audio, and Their Applications. In *The Handbook of Multimodal-Multisensor Interfaces: Foundations, User Modeling, and Common Modality Combinations - Volume 1*. ACM and Morgan & Claypool, 277–317. https://doi.org/10.1145/3015783.3015792

[33] Vinitha Gadiraju, Annika Muehlbradt, and Shaun K. Kane. 2020. BrailleBlocks: Computational Braille Toys for Collaborative Learning. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376295

[34] Stavros Garzonis, Simon Jones, Tim Jay, and Eamonn O'Neill. 2009. Auditory Icon and Earcon Mobile Service Notifications: Intuitiveness, Learnability, Memorability and Preference. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1513–1522. https://doi.org/10.1145/1518701.1518932

[35] William W. Gaver. 1991. Sound Support for Collaboration. In *Proceedings of the Second Conference on European Conference on Computer-Supported Cooperative Work* (Amsterdam, The Netherlands) *(ECSCW'91)*. Kluwer Academic Publishers, USA, 293–308.

[36] Charles Goodwin. 2006. Human Sociality as Mutual Orientation in a Rich Interactive Environment: Multimodal Utterances and Pointing in Aphasia. In *Roots of Human Sociality*, Nicholas J. Enfield and Stephen C. Levinson (Eds.). Berg, London, 96–125.

[37] João Guerreiro and Daniel Gonçalves. 2015. Faster Text-to-Speeches: Enhancing Blind People's Information Scanning with Faster Concurrent Speech. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility* (Lisbon, Portugal) *(ASSETS '15)*. ACM, New York, NY, USA, 3–11. https://doi.org/10.1145/2700648.2809840

[38] Carl Gutwin and Saul Greenberg. 2002. A Descriptive Framework of Workspace Awareness for Real-Time Groupware. *Computer Supported Cooperative Work* 11 (2002), 411–446. https://doi.org/10.1023/A:1021271517844

[39] Christian Heath, Marina Jirotka, Paul Luff, and Jon Hindmarsh. 1995. Unpacking Collaboration: The Interactional Organisation of Trading in a City Dealing Room. *Computer Supported Cooperative Work* 3, 2 (1995), 147–165. https://doi.org/10.1007/BF00773445

[40] William C. Hill, James D. Hollan, Dave Wroblewski, and Tim McCandless. 1992. Edit Wear and Read Wear. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Monterey, CA, USA) *(CHI '92)*. ACM, New York, NY, USA, 3–9. https://doi.org/10.1145/142750.142751

[41] Alison Kafer. 2013. *Feminist, Queer, Crip*. Indiana University Press. http://www.jstor.org/stable/j.ctt16gz79x

[42] Grete Helena Kütt, Kevin Lee, Ethan Hardacre, and Alexandra Papoutsaki. 2019. Eye-Write: Gaze Sharing for Collaborative Writing. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland, UK) *(CHI '19)*. ACM, New York, NY, USA, Article 497, 12 pages. https://doi.org/10.1145/3290605.3300727

[43] Ida Larsen-Ledet and Henrik Korsgaard. 2019. Territorial Functioning in Collaborative Writing. *Computer Supported Cooperative Work* 28 (2019), 391–433. https://doi.org/10.1007/s10606-019-09359-8

[44] Ida Larsen-Ledet, Henrik Korsgaard, and Susanne Bødker. 2020. Collaborative Writing Across Multiple Artifact Ecologies. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. ACM, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376422

[45] Sooyeon Lee, Madison Reddie, Chun-Hua Tsai, Jordan Beck, Mary Beth Rosson, and John M. Carroll. 2020. The Emerging Professional Practice of Remote Sighted Assistance for People with Visual Impairments. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–12. https://doi.org/10.1145/3313831.3376591

[46] Simi Linton. 1998. *Claiming Disability: Knowledge and Identity*. NYU Press. http://www.jstor.org/stable/j.ctt9qfx5w

[47] Paul Benjamin Lowry, Aaron Curtis, and Michelle René Lowry. 2004. Building a Taxonomy and Nomenclature of Collaborative Writing to Improve Interdisciplinary Research and Practice. *The Journal of Business Communication (1973)* 41, 1 (2004), 66–99. https://doi.org/10.1177/0021943603259363

[48] Kelly Mack, Maitraye Das, Dhruv Jain, Danielle Bragg, John Tang, Andrew Begel, Erin Beneteau, Josh Urban Davis, Abraham Glasser, Joon Sung Park, and Venkatesh Potluri. 2021. Mixed Abilities and Varied Experiences: A Group Autoethnography of a Virtual Summer Internship. In *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, New York, NY, USA, Article 21, 13 pages. https://doi.org/10.1145/3441852.3471199

[49] Sergio Mascetti, Andrea Gerino, Cristian Bernareggi, and Lorenzo Picinali. 2017. On the Evaluation of Novel Sonification Techniques for Non-Visual Shape Exploration. *ACM Transactions on Accessible Computing (TACCESS)* 9, 4, Article 13 (April 2017), 28 pages. https://doi.org/10.1145/3046789

[50] Daniel Mendes, Sofia Reis, João Guerreiro, and Hugo Nicolau. 2020. Collaborative Tabletops for Blind People: The Effect of Auditory Design on Workspace Awareness. *Proceedings of the ACM on Human-Computer Interaction* 4, ISS, Article 197 (Nov. 2020), 19 pages. https://doi.org/10.1145/3427325

[51] Oussama Metatla, Nick Bryan-Kinns, and Tony Stockman. 2018. "I Hear You": Understanding Awareness Information Exchange in an Audio-Only Workspace.

In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–13. https://doi.org/10.1145/3173574.3174120

[52] Oussama Metatla, Nick Bryan-Kinns, Tony Stockman, and Fiore Martin. 2012. Cross-modal collaborative interaction between visually-impaired and sighted users in the workplace. In *Proceedings of the 18th International Conference on Auditory Display* (Atlanta, GA, USA) *(ICAD '12)*. 9 pages.

[53] 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, Article 14 (July 2013), 37 pages. https://doi.org/10.1145/2493171.2493172

[54] Lourdes Morales, Sonia M. Arteaga, and Sri Kurniawan. 2013. Design Guidelines of a Tool to Help Blind Authors Independently Format Their Word Documents. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems* (Paris, France) *(CHI EA '13)*. ACM, New York, NY, USA, 31–36. https://doi.org/10.1145/2468356.2468363

[55] Ricardo Olenewa, Gary M. Olson, Judith S. Olson, and Daniel M. Russell. 2017. Now That We Can Write Simultaneously, How Do We Use That to Our Advantage? *Communication of the ACM* 60, 8 (July 2017), 36–43. https://doi.org/10.1145/2983527

[56] Judith S. Olson, Gary M. Olson, Marianne Storrøsten, and Mark Carter. 1993. Groupwork Close up: A Comparison of the Group Design Process with and without a Simple Group Editor. *ACM Transactions on Information Systems (TOIS)* 11, 4 (Oct. 1993), 321–348. https://doi.org/10.1145/159764.159763

[57] Judith S. Olson, Dakuo Wang, Gary M. Olson, and Jingwen Zhang. 2017. How People Write Together Now: Beginning the Investigation with Advanced Undergraduates in a Project Course. *ACM Transactions on Computer-Human Interaction (TOCHI)* 24, 1, Article 4 (March 2017), 40 pages. https://doi.org/10.1145/3038919

[58] Maulishree Pandey, Vaishnav Kameswaran, Hrishikesh V Rao, Sile O'Modhrain, and Steve Oney. 2021. Understanding Accessibility and Collaboration in Programming for People with Visual Impairments. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1, Article 129 (April 2021), 30 pages. https://doi.org/10.1145/3449203

[59] Ignacio Perez-Messina, Claudio Gutierrez, and Eduardo Graells-Garrido. 2018. Organic Visualization of Document Evolution. In *23rd International Conference on Intelligent User Interfaces* (Tokyo, Japan) *(IUI '18)*. ACM, New York, NY, USA, 497–501. https://doi.org/10.1145/3172944.3173004

[60] Ilona R. Posner and Ronald M. Baecker. 1992. How People Write Together. In *Proceedings of the 25th Hawaii International Conference on System Sciences (HICSS '92)*. 127–138.

[61] Johanna Pöysä-Tarhonen, Nafisa Awwal, Päivi Häkkinen, and Suzanne Otieno. 2021. Joint attention behaviour in remote collaborative problem solving: exploring different attentional levels in dyadic interaction. *Research and Practice in Technology Enhanced Learning* 16, 11 (2021). https://doi.org/10.1186/s41039-021-00160-0

[62] Tom Rodden. 1996. Populating the Application: A Model of Awareness for Cooperative Applications. In *Proceedings of the 1996 ACM Conference on Computer Supported Cooperative Work (CSCW '96)*. ACM, New York, NY, USA, 87–96. https://doi.org/10.1145/240080.240200

[63] Abir Saha and Anne Marie Piper. 2020. Understanding Audio Production Practices of People with Vision Impairments. In *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility* (Virtual Event) *(ASSETS '20)*. ACM, New York, NY, USA, Article 36, 13 pages. https://doi.org/10.1145/3373625.3416993

[64] Tony Salvador, Jean Scholtz, and James Larson. 1996. The Denver Model for Groupware Design. *SIGCHI Bulletin* 28, 1 (Jan. 1996), 52–58. https://doi.org/10.1145/249170.249185

[65] Daisuke Sato, Shaojian Zhu, Masatomo Kobayashi, Hironobu Takagi, and Chieko Asakawa. 2011. Sasayaki: Augmented Voice Web Browsing Experience. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Vancouver, BC, Canada) *(CHI '11)*. ACM, New York, NY, USA, 2769–2778. https://doi.org/10.1145/1978942.1979353

[66] John G. Schoeberlein and Yuanqiong Wang. 2014. Usability Evaluation of an Accessible Collaborative Writing Prototype for Blind Users. *Journal of Usability Studies* 10, 1 (Nov. 2014), 26–45. https://dl.acm.org/doi/10.5555/2817310.2817313

[67] Suleman Shahid, Jip ter Voort, Maarten Somers, and Inti Mansour. 2016. Skeuomorphic, Flat or Material Design: Requirements for Designing Mobile Planning Applications for Students with Autism Spectrum Disorder. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct* (Florence, Italy) *(MobileHCI '16)*. ACM, New York, NY, USA, 738–745. https://doi.org/10.1145/2957265.2961866

[68] Lei Shi, Brianna J. Tomlinson, John Tang, Edward Cutrell, Daniel McDuff, Gina Venolia, Paul Johns, and Kael Rowan. 2019. Accessible Video Calling: Enabling Nonvisual Perception of Visual Conversation Cues. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW, Article 131 (Nov. 2019), 22 pages. https://doi.org/10.1145/3359233

[69] Kristen Shinohara, Michael McQuaid, and Nayeri Jacobo. 2020. Access Differential and Inequitable Access: Inaccessibility for Doctoral Students in Computing. In *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility* (Virtual Event) *(ASSETS '20)*. ACM, New York, NY, USA, Article 7, 12 pages. https://doi.org/10.1145/3373625.3416989

[70] Barbora Siposova and Malinda Carpenter. 2019. A new look at joint attention and common knowledge. *Cognition* 189 (2019), 260–274. https://doi.org/10.1016/j.cognition.2019.03.019

[71] Kevin M. Storer and Stacy M. Branham. 2019. "That's the Way Sighted People Do It": What Blind Parents Can Teach Technology Designers About Co-Reading with Children. In *Proceedings of the 2019 on Designing Interactive Systems Conference* (San Diego, CA, USA) *(DIS '19)*. ACM, New York, NY, USA, 385–398. https://doi.org/10.1145/3322276.3322374

[72] Jürgen Streeck, Charles Goodwin, and Curtis Lebaron. 2011. Embodied Interaction in the Material World: An Introduction. In *Embodied Interaction: Language and Body in the Material World*. Cambridge University Press, 1–26.

[73] James Tam and Saul Greenberg. 2006. A Framework for Asynchronous Change Awareness in Collaborative Documents and Workspaces. *International Journal of Human-Computer Studies* 64, 7 (July 2006), 583–598. https://doi.org/10.1016/j.ijhcs.2006.02.004

[74] Anja Thieme, Cecily Morrison, Nicolas Villar, Martin Grayson, and Siân Lindley. 2017. Enabling Collaboration in Learning Computer Programing Inclusive of Children with Vision Impairments. In *Proceedings of the 2017 Conference on Designing Interactive Systems* (Edinburgh, UK) *(DIS '17)*. ACM, New York, NY, USA, 739–752. https://doi.org/10.1145/3064663.3064689

[75] Shannon M. Tomlinson. 2016. Perceptions of Accessibility and Usability by Blind or Visually Impaired Persons: A Pilot Study. In *Proceedings of the 79th ASIST Annual Meeting: Creating Knowledge, Enhancing Lives Through Information & Technology* (Copenhagen, Denmark) *(ASIST '16)*. American Society for Information Science, Silver Springs, MD, USA, Article 120, 4 pages.

[76] Yolanda Vazquez-Alvarez and Stephen A. Brewster. 2011. Eyes-Free Multitasking: The Effect of Cognitive Load on Mobile Spatial Audio Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 2173–2176. https://doi.org/10.1145/1978942.1979258

[77] Johan Wagemans. 2014. Historical and conceptual background: Gestalt theory. In *Oxford Handbook of Perceptual Organization*, Johan Wagemans (Ed.). Oxford University Press, Chapter 1.

[78] Dakuo Wang, Judith S. Olson, Jingwen Zhang, Trung Nguyen, and Gary M. Olson. 2015. DocuViz: Visualizing Collaborative Writing. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems* (Seoul, Republic of Korea) *(CHI '15)*. ACM, New York, NY, USA, 1865–1874. https://doi.org/10.1145/2702123.2702517

[79] 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. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW, Article 107 (Dec. 2017), 18 pages. https://doi.org/10.1145/3134742

[80] M. M. Waqar, Nafees Ahmad, Muhammad Aslam, and Martinez-Enriquez A. M. 2020. Cloud based Co-authoring platform for visually impaired people. In *17th International Conference on Electrical Engineering, Computing Science and Automatic Control (CCE)*. 1–6. https://doi.org/10.1109/CCE50788.2020.9299126

[81] Pavani Yalla and Bruce N. Walker. 2008. Advanced Auditory Menus: Design and Evaluation of Auditory Scroll Bars. In *Proceedings of the 10th International ACM SIGACCESS Conference on Computers and Accessibility* (Halifax, Nova Scotia, Canada) *(Assets '08)*. ACM, New York, NY, USA, 105–112. https://doi.org/10.1145/1414471.1414492

[82] Soobin Yim, Dakuo Wang, Judith Olson, Viet Vu, and Mark Warschauer. 2017. Synchronous Collaborative Writing in the Classroom: Undergraduates' Collaboration Practices and Their Impact on Writing Style, Quality, and Quantity. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, OR, USA) *(CSCW '17)*. ACM, New York, NY, USA, 468–479. https://doi.org/10.1145/2998181.2998356

[83] Chien Wen Yuan, Benjamin V. Hanrahan, Sooyeon Lee, Mary Beth Rosson, and John M. Carroll. 2017. "I Didn't Know That You Knew I Knew": Collaborative Shopping Practices Between People with Visual Impairment and People with Vision. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW, Article 118 (December 2017), 18 pages. https://doi.org/10.1145/3134753

[84] Yeshuang Zhu, Shichao Yue, Chun Yu, and Yuanchun Shi. 2017. CEPT: Collaborative Editing Tool for Non-Native Authors. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, OR, USA) *(CSCW '17)*. ACM, New York, NY, USA, 273–285. https://doi.org/10.1145/2998181.2998306