Find and Seek: Assessing the Impact of Table Navigation on Information Look-up with a Screen Reader

11

KRISTIN WILLIAMS, Human Computer Interaction Institute, Carnegie Mellon University
TAYLOR CLARKE, Software Engineering Institute, Carnegie Mellon University
STEVE GARDINER, Google
JOHN ZIMMERMAN, Human Computer Interaction Institute, Carnegie Mellon University
ANTHONY TOMASIC, Language Technologies Institute, Carnegie Mellon University

Web designers use visual cues such as layout and typography to make pages easier to navigate and understand. Yet, screen readers generally ignore these features and present page information in a linear audio stream. We investigate whether transcoding the visual semantics of grid-based layouts to tables supports better navigation. In a controlled experiment, participants navigated re-written pages significantly faster when doing data synthesis tasks and more accurately when looking up information meeting multiple criteria. Participants rated their table navigation experience better in terms of effort, memorization, ease of navigation, understanding of page information, and confidence in submitted answers. Participants attributed these gains to the table structure's support for (1) predictable audio presentation, (2) adopting an appropriate search strategy, and (3) making sense of page content. Contrary to the established belief that tables are inaccessible, our results show that tables can facilitate navigation when users need to synthesize across page content.

CCS Concepts: • **Human-centered computing → Empirical studies in accessibility**; *Auditory feedback*; Accessibility systems and tools;

Additional Key Words and Phrases: Table navigation, screen reader, spatial layout, web accessibility, audio interfaces, cognitive load

ACM Reference format:
Kristin Williams, Taylor Clarke, Steve Gardiner, John Zimmerman, and Anthony Tomasic. 2019. Find and Seek: Assessing the Impact of Table Navigation on Information Look-up with a Screen Reader. *ACM Trans. Access. Comput.* 12, 3, Article 11 (August 2019), 23 pages.
https://doi.org/10.1145/3342282

The contents of this paper were developed under a grant from the National Institute on Disability, Independent Living, and Rehabilitation Research (NIDILRR grant number NIDILRR 90DP0061). NIDILRR is a Center within the Administration for Community Living (ACL), Department of Health and Human Services (HHS).
Authors' addresses: K. Williams, Human Computer Interaction Institute, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213; email: krismawil@cs.cmu.edu; T. Clarke, Software Engineering Institute, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213; email: tclarke@alumni.cmu.edu; S. Gardiner, Google, 6425 Penn Avenue, Pittsburgh, PA 15206; email: steven.h.gardiner@gmail.com; J. Zimmerman, Human Computer Interaction Institute, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213; email: johnz@cs.cmu.edu; A. Tomasic, Language Technologies Institute, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213; email: tomasic@cs.cmu.edu.

1936-7228/2019/08-ART11 $15.00
https://doi.org/10.1145/3342282

1 INTRODUCTION

Modern web pages contain many semantic relationships that depend on visual perception for comprehension. Web designers use a combination of spatial location and typographic cues (e.g., hierarchy, font face, color) to create relationships between different pieces of information on a page. With the advent of responsive design, designers have developed even more complex visual semantics by establishing different versions of a page's layout based on the size of the users' screen. For example, a design that shows a grid of shopping choices on a laptop might render the items as a list when viewed on a mobile phone. These techniques are so widely used that templated code bases incorporate grid-based page flows (e.g., Bootstrap or Cascade) to facilitate adapting a page to different devices [3, 4].

Despite the important role these design techniques have in making a page easier to navigate, today's screen readers generally ignore much of a page's visual semantics when serializing a page's content to be read aloud. Instead, screen readers focus on a hierarchical representation found in the page's document object model (DOM). They prioritize features such as the page's title, headings, and other HTML landmarks. However, the DOM's hierarchy often differs drastically from a page's visual layout. When navigating some pages with screen readers, the reader can jump from the top of the page to the bottom, skipping over critical information that is intentionally placed in the middle of the page [37, 38]. Complicating this, different web browsers adapt grid layouts differently across devices and screen sizes, and this adaption can lead to screen readers reading aloud items that may not even be rendered on the device's screen [16, 26]. In effect, screen reader users are unable to benefit from design choices that help to simplify and quickly characterize a web page.

We investigate whether adapting the visual semantics of grid-based layouts to screen readers using HTML <table> tags supports better page navigation for screen reader users. We conducted a controlled experiment with 21 screen reader users employing participants' own preferred screen reader setup. We asked participants to complete data look-up tasks using either pages with visual semantics or those re-written by EnTable with table tags [20]. For each task, participants needed to use relationships among page elements to answer questions related to common activities such as visiting a restaurant, shopping, or going to a concert. We compared performance using speed, accuracy, and perceived task demands. We also elicited feedback from participants on their experiences using pages with either visual semantics or table tags to navigate the web. In addition, we interviewed participants about their typical web navigation experiences, and what kinds of alternative designs and data structures could better support screen reader navigation.

Our results show that pages re-written with table tags positively impacted participant performance and experience. In general, participants were faster and more accurate when using the re-written pages to complete data synthesis tasks. Participants also rated their navigational experience better when using the re-written pages in terms of perceived effort, memorization, ease of navigation, confidence in submitted answers, and understanding of page information. During interviews, participants described how table mark-up made cross-comparison of page information easier. Table mark-up supported additional screen reading functions (e.g., reading aloud metadata associated with an element) and navigation strategies that coupled well with participants' problem-solving approach to the data lookup question. Our results show that contrary to the widely held belief that tables are inaccessible, tables can facilitate page navigation when needing to synthesize information. While our experiment and study specifically focused on the role of tables in supporting screen reader users in making cross-comparisons of page information, our work informs (i) efforts to make web pages more easily navigable and (ii) approaches to identifying implicit page semantics for existing screen readers.

2 RELATED WORK

We review related research on navigating web page hierarchies using a screen reader and transcoding approaches that try to give navigational control to the user. This review includes prior work on the added costs of navigating a page using audio interfaces, automated approaches to model and predict these additional costs at the web design stage, and the accessibility of tables.

2.1 The Impact of Audio Interfaces on Web Design and Navigation Efficiency

Interpreting tabular data with a screen reader demands well-designed navigation support. Many daily web activities such as checking public transportation time tables or financial reports require navigation of tabular information [14, 27, 38]. When the amount of information exceeds the device's presentation capabilities, navigation becomes particularly important [13]. This occurs when a web page is bigger than what can fit on a screen, or when the output mode of presentation cannot handle the simultaneous display of information anticipated by the designer as with screen readers [13]. Listening to and comprehending speech introduces a time lag into navigation. Current cognitive task models of page navigation such as keystroke level models do not consider listening time, and so, screen reader navigation time becomes difficult to predict *a priori* [38]. Researchers estimate that listening time is likely to take 41% of the task's time in addition to other navigation subtasks such as selecting and traversing portions of the page [38]. Even when pages might be technically accessible to a screen reader, their design and layout can introduce additional time costs [26]. For example, some designs can force screen readers to go through the content line by line or refocus the virtual cursor away from the currently browsed content and so force the user to traverse previously covered material [26]. These temporal costs associated with using a screen reader to navigate a page emphasize a need for page designs that reduce audio navigation time [37, 38].

2.2 Transcoding

Transcoding is an approach to accessibility that transforms page content into an alternative format on the fly to make it more accessible [11]. Several transcoding systems have been developed to encapsulate web accessibility standards in the set of transformation rules applied by the system [30, 31]. Finding an error-proof method to test for whether a web page complies with accessibility guidelines remains an ongoing challenge [29], and even when the page complies with these guidelines, the experience of accessing that page may still be frustrating [26, 31]. Further, many guidelines do not translate neatly to transformation rules such as those that require an understanding of context [31]. Even when many accessibility standards can be successfully transcoded, accessing pages may still be frustrating due to problems with navigability [31]. A page's hierarchy can greatly impact the time it takes for a screen reader user to navigate a page. Inner-page links and web elements that work with a screen reader's shortcut keys reduce the path a screen reader would have to take to reach page content in what is known as the page's reaching time [36]. Transcoding approaches that have tried to address issues with reaching time typically rely on human-generated annotations of pages to ensure that transcoding is accurate and usable for end users by making page semantics explicit for the transformation rules [11]. For example, DANTE is a system that adopts a travel analysis framework for annotating the structure and role of page elements so that navigational features can be more easily accessed using a screen reader [44]. To do this, volunteer annotators are needed to identify a page's navigational cues according to a controlled vocabulary developed by DANTE's authors [44].

The principle challenge for annotation-based transcoding approaches is generating the annotations themselves either through volunteers, labor-intensive work by the developer, or through mixed-initiative interfaces [11]. Some transcoding work has tried to leverage the relationships between HTML attributes and CSS style sheets as an annotation framework to enable annotation-based approaches without requiring the labor of generating the annotation files [11]. In general, this work often adopts one of two approaches: concern with reading order or adaptation to small devices [11, 31, 45]. However, while transcoding for small devices using a fragmented representation of the page shares similar techniques as those adopted for web accessibility, the requirements for small devices and voice output differ [10]. Specifically, when empirical studies were run with a reasonable amount of participants (e.g., >10), researchers found that screen reader users still need (1) a way to identify the role of the fragment and (2) to be able to execute a global search using keyboard shortcuts (a feature that was inhibited by fragmenting the page) [10, 44]. One experiment, run with 20 blind participants and 20 sighted matched controls, found that transcoding for page hierarchy on small devices significantly improved search time for both groups [45]. However, this same experiment found that when techniques were personalized to the user, significant gains in search time for sighted participants were not experienced by screen reader users [45]. These studies emphasize a need for empirical studies of transcoding techniques that include a reasonable number of participants to verify whether the promise of these systems are experienced, in fact, by screen reader users.

Transcoding shows promise for reducing a page's reaching time by better adapting page content to navigational strategies while still positioning page elements among their peers [13]. Yet, navigation time can be reduced or eliminated altogether, if a screen reader user has a way of transforming visual knowledge to a structure that supports the ability to ask arbitrary questions of it. This transformation approach, known as the "Raman Principle," approaches screen reader accessibility by likening the design goal to querying a database with the user's own questions [28, 34]. In this spirit, our approach emphasizes that with tabular data, no single navigational path is obviously more important than another. In particular, tabular data is matrix-like and supports several different equally valid entry-points or indexing methods [13]. A navigational grid supporting many degrees of navigational freedom with a screen reader may be far more important for accessing a page than a verbal description of spatial layout by providing greater navigational flexibility [17, 45].

2.3 Table Accessibility

Contrary to the "Raman Principle," some scholars and professionals have argued that complex data structures such as tables or frames make page navigation problematic with a screen reader [43]. They argue that the multidimensionality of tables essentially makes non-visual understanding of the content difficult, because the screen reader's aural linearization cannot be considered equivalent to the non-linearity of the table's content [32, 33, 43]. However, screen reader users also have the ability to use screen reader commands—or, keyboard shortcuts—to explore and interrogate a page's structure. This functionality allows a screen reader user to direct aural feedback to report on specific areas of interest and to explore different aspects of the page using the tactile feedback of keyboard keys. When a screen reader's keyboard shortcuts are mapped in such a way to consider the tactile feedback of the keys as a spatial cue—as has been done by using the keyboard's arrow keys for table commands in modern, commercially available screen readers—the keyboard itself can provide two-dimensional cues for spatial layout [9, 11, 25].

There is some evidence that tables can be inaccessible, yet these findings are inconclusive and coincide with research advances in the design of table navigation. An early study of web accessibility using a screen reader found that when four participants were matched with sighted web users, data look up with a table "was trivial for sighted subjects, but almost impossible for the visually impaired subjects" [22]. However, the study's use of only a few participants and omission of the task conditions limit the generalizability of these findings. Around this time period, vertical navigation techniques were introduced to address limitations on horizontal only table navigation [9]. At the same time, transcoding techniques were developed to transform ungridded tables (that is, a table with a cell or cells that span more than one column creating irregularity in the table grid) to gridded tables so that vertical and horizontal navigation steps mapped directly onto the traversed table cells [9]. Ungridded tables proved particularly problematic for screen reader users when horizontal navigation was the only supported way to traverse a table (see Reference [43] for motivating examples). In a study conducted with three participants, the researchers found that tables could be more easily navigated using a screen reader when two degrees of navigational freedom were supported [9]. Another system allowed screen reader users to navigate tables in spreadsheets using two degrees of navigational freedom and provided contextual audio information such as the cell's category as determined by column headers [15]. This system was evaluated using eight sighted users and five blind users, and the researchers found that a query mechanism was needed to support users in interrogating the data with their own questions [15].

Early work on the IBM Home Page Reader analyzed HTML pages with complicated tables to infer table headers so that users could use keyboard shortcuts to access those headers directly and skip through parts of the table [11]. This system required users to switch to table navigation mode when they encountered a table [9]. Once in table navigation mode, users could move horizontally through the table by using the leftmost (number 4) and rightmost (number 6) keys on the keyboard's number pad, move vertically by using the middle top key (number 8) and the bottom middle key (number 2), read out a cell's context (like the column header) using a key combination, and jump to the next set of cells using another key combination [9]. The IBM Home Page reader was offered as a stand-alone browser product. However, the product has been discontinued. While the browser is no longer available, its compass-like navigation design has been incorporated into many modern screen readers' support for table navigation [2].

To address accessibility issues with complex data structures, ARIA is an annotation language for making interactive web page elements apparent to a screen reader user such as widgets and page information frequently updated with asynchronous JavaScript and XML (AJAX) [6]. ARIA supports a grid pattern for annotating pages using interactive tabular data and layout grids [6]. By assigning a grid pattern, the developer enables a screen reader user to navigate the grid using their arrow keys much like a table. However, unlike tables, the developer must provide code to manage the screen reader's focus on grid elements [6]. So, even when using ARIA, web pages with grid-based design must still be annotated by the developer to be more easily navigable with a screen reader, and the developer must write further code to support auditory feedback on inner-page relationships such as those conveyed by a page's visual semantics. To address the need for developer intervention, VoiceOver's web spots feature automatically analyzes the visual design of a page and groups visual sections of a page into areas supporting quick scanning with a screen reader [5]. Web spots provides auditory feedback to each region by reading out the beginning of that area's page content. This feature reduces the need for manual intervention to annotate page elements to support screen reader navigation. However, the implicit visual semantics of inner page relationships are presented as if they were the beginnings of different paragraphs and so, this solution neglects navigational support of the relationships themselves as supported with table tags. Both ARIA Grids and Voiceover's web spots allow a screen reader to more easily navigate a page's elements, but neither improve access to the page's visual semantics without requiring custom code to manage focus on those relationships.

We extend prior research on using crowd-sourcing approaches to address this human annotation problem. Crowd powered assistive technology has been shown to be a viable way to solve the human annotation problem by using crowd workers [12]. Building on this prior work, a browser extension was created to support crowd workers in rewriting web pages with grid-based designs as gridded tables using HTML table markup to facilitate navigation with a screen reader [20]. In an exploratory study, screen reader users described being able to navigate the resulting pages faster and more accurately than pages that had not been rewritten [20, 21]. To investigate the potential for using crowdsourced transcoded tables to support better navigation, we conducted a controlled experiment to assess whether the modified pages improve performance on data lookup tasks for screen reader users.

3 TRANSFORMING VISUAL SEMANTICS TO SUPPORT TABLE NAVIGATION

We used the system EnTable [20] to transcode the visual semantics of grid-based web page design into HTML tables [20]. Briefly, transcoding of a page can be requested via a button embedded in a browser. The page is then sent to a crowd worker. Using a spreadsheet-like labeling tool, the worker identifies the tables in the page by drag-and-drop of elements from the page into an associated spreadsheet. The labeling tool focuses on repetitive types of information in the page and leaves the larger page structure unchanged. The worker labels the first two rows of the page and EnTable derives the rest of the rows of the table using heuristics. When the worker is satisfied with the result, perhaps on multiple tables on the page, the system generates a transcoding program "wrapper" for the page. When the page is later accessed, the wrapper rewrites the page on-the-fly to transcode the labeled visual grid into HTML tables, thus supporting screen reader users with two-dimensional navigation to access to the page's information.

4 PILOT STUDY

We conducted a pilot study to examine (1) whether screen reader users benefit from EnTable without any further modifications, characterize the demographic assumptions of our target users, and (2) anticipate the likely technical infrastructure screen reader users would be adopting when accessing EnTable. We recruited screen reader users through local organizations, word of mouth, and listservs. Study sessions lasted 60–90min, took place at a location of the participant's choice, and used the participant's preferred access technology (requiring only a stable Internet connection and web browser). Participants were paid $30 for their time. Participants completed a background survey, and then did 12 data look-up tasks, similar to the tasks in the experiment described below. These tasks consisted in two practice tasks followed by two task blocks of five tasks each, counterbalanced for condition across participants. Finally, participants completed a semi-structured interview. The data look-up tasks consisted in using a web page—either a page found on the web or a page rewritten with EnTable—to answer a factual question using data from the page.

To assess whether pages rewritten with table tags would be likely to support faster and more accurate navigation (as reported in prior work [21]), we collected data on participant's speed, accuracy, and their described experience doing a data look-up task using a screen reader. For our speed measure, we measured the time it took participants to click on the provided page until they submitted an answer to our study's question using a web form. For our accuracy measure, we graded the submitted answers according to whether they identified the target answer. Last, we transcribed the interviews and coded them for themes of interest including interactions with audio feedback from the page, two-dimensional navigation, and experience getting factual information from the web.

While our pilot study helped us refine our protocol and research questions, it also substantially altered some of our hypotheses and motivated entirely new ones (discussed in Section 4.3). We found that participants struggled with the complexity of some of our tested pages and participants varied substantially in their familiarity with screen reader commands. We detail below what we learned and its impact on our study.

4.1 Participant Backgrounds

We successfully recruited 22 screen reader users (10 male, 12 female). When asked about their proficiency using a screen reader, 18.2% described themselves as advanced, 68.2% as intermediate, and 13.6% as beginners. When asked about their proficiency using the Internet, 22.7% described themselves as advanced, 68.2% as intermediate, and 9.1% as beginners. 64% of respondents reported that they customize their screen reader, while 36% reported that they do not customize their screen reader. When asked how customized their screen reader is, 0% reported very customized, 40.9% reported that their screen reader was somewhat customized, 40.9% reported slightly customized, and 18.2% reported that their screen reader was not customized.

Four of our participants were unable to complete the tasks, because they were not familiar enough with using the Internet to navigate our web form on their own, or they did not know enough of the screen reader's shortcuts to complete the practice task without substantial frustration. Participant times varied between 1 and 8min per task. While able to complete the study, one of our participants required twice the calibrated time of 90min to do so. This variance is in line with previous literature, but suggests that participant background may substantially impact whether a person will benefit from web accessibility improvements [27].

4.2 Participant Feedback on Pilot Tasks

Participant feedback on our pilot study highlighted how sizeable page information with little markup can negatively impact the navigation experience.

4.2.1 Memory. Participants encountered substantial demand on their memory to successfully complete the task. Thirteen participants described memory as posing problems when navigating a page, as P19 stated: *"What doesn't work well [is] when I'm relying upon my memory."* The auditory presentation required participants to synthesize across a substantial amount of information building a representation of the page. *"It's almost like you're memorizing the table as you go. You're only seeing one piece of it at a time"* (P2). The serial audio presentation of the screen reader interface required participants to build up a gradual understanding of the page but also retain information that had ceased being read aloud.

4.2.2 Understanding Repetitive Page Structure. Participant responses suggested that pages with table tags structure the audio presentation of page information in a way that enhances understanding the page's contents. Fourteen participants described how table headers supported fluid navigation once they entered a table. P13 described the tedium of going line by line when a page's contents were not in a table: *"It was just a paragraph I had to go by words to find the telephone number."* In contrast, pages with table tags provided guidance in context. P18 described what this audio feedback was like: *"The title of the row or the column was repeated over and over, and that's one of the things that makes it really easy to just navigate through and not lose my place."* Tables facilitated understanding and navigation by creating an association between descriptors and content.

4.2.3 Uncertainty Arising from Significant Task Load. Participants described having to track many different pieces of information and adopted alternative methods to track information they were certain of. Sorting through relevant and irrelevant page information led to participant uncertainty and the need to verify the task's question and answer. P8 described having *"to go back and make sure [they were] really listening specifically to what information was asked for and then going back and then searching out again what it was."* Participants used coping strategies independent of the study's technology to help with attending closely to the presented information. P6 made quick notes of an answer he thought would help with a task question, *"I sort of jotted down the name of the city, figuring that that would be the name, but it wasn't."* We observed participants develop many notetaking strategies including copy/paste, manual braille punch, and speech to text both in an alternative application and on an additional device to cope with task load.

4.3 Pilot's Impact on Controlled Experiment

The high number of participants who struggled with our practice tasks led us to modify our experiment in three ways. First, we implemented a more stringent set of screening questions asking potential participants more about how they use the Web to ensure that they actively go to web pages, navigate forward and back in the browser, and traverse web page content as opposed to simply following e-mailed links or confining their screen reader usage to only their operating system. Second, we introduced a training period into our experiment to actively acquaint our participants with a set of screen reader shortcuts they could use to navigate the experiment's web pages. Last, based off participant feedback, we developed a set of hypotheses that tables lessened task load and demands on memory. We describe these hypotheses in more detail below.

5 CONTROLLED FIELD EXPERIMENT

We conducted a controlled experiment to assess whether screen reader navigation of web pages using grid-based design is better when EnTable semi-automatically rewrote the pages to use table tags. We evaluated the navigational experience in terms of speed, accuracy, perceived task demands, and participants' feedback. We describe our experiment setup in more detail below.

5.1 Method

We asked participants to use their preferred screen reader setup to complete several data lookup tasks using pages found on the web. This approach allowed us to increase ecological validity in our experimental design by testing whether transforming pages with tables could enable table navigation across variations in browser version, screen reader software and version, and hardware setup. For our tested tasks, participants were asked to use pages with either grid-based visual design or our modified version using table tags. We adopted a human-in-the-loop approach to modify pages to simulate the role of crowdsourcing accessibility fixes to pages (described more below in the Workflow section). We then evaluated differences between the two conditions using a combination of subjective and objective measures of navigational experience as well as elicited participant feedback on how the experiment went.

5.1.1 Participants. We recruited participants through local NGOs, word of mouth, and listservs. We then screened participants for regular use of a screen reader with the Internet. Our screening criteria required participants to regularly visit web pages through a web browser (this last condition served to differentiate participants who used e-mail, but did not navigate web sites, as discovered during our pilot). We paid participants $30 in cash for participating.

5.1.2 Page Conditions. We identified five pages from the Internet that used grid-based design (visual semantics, our baseline condition) and provide information that would be needed to do daily tasks. These included sites to lookup the location of an event, shop online, choose a restaurant, determine the weather forecast, and purchase a plane ticket. To create a comparable set of five tasks, we rewrote the content of the pages to cover different information, but retained the HTML and CSS formatting for the pages, including linked images. The new content included pet adoption, grocery items, book reviews, shoe shopping, and ordering take-out food. We then reviewed audio renderings of these pages using aDesigner [35] to catch any residual content that we may have missed in our rewriting that would be noticeable when accessing the page with a screen reader. We made further adjustments to the rewritten pages accordingly. Finally, we mixed the original pages with the rewritten pages to create two task sets of five pages each. Each set consisted in both original and rewritten content, and contained only one instance of five different page structures (the HTML and CSS formatting found in the original pages). This process gave us a total of 10 pages.

5.1.3 Task Types. For each task block, we created five questions, each of which, could be answered using only one of the pages from the task set. The questions increased in degree of difficulty according to how much of the page content needed to be covered to answer it, as given below:

- Find an item on the page,
- Find an item on the page (but, in fact, the item is missing from the page),
- Find an item that meets a minimum criteria,
- Find an item from a set that has the lowest value relative to preset criteria, and
- Find an item that meets two criteria.

For all 10 pages, we created 10 equivalent pages rewritten with table tags so that we had a total of 20 pages to make up four task sets (two sets of the visual semantics pages and two sets of the pages with table tags). For each of the 20 pages all interactive elements including forms, drop down menus, widgets, active web-elements, links, and ARIA-live regions (elements of a page that update with real-time information and interrupt with an audio notification of new content) were disabled. We did not test the pages for accessibility, and instead, relied on aDesigner to give a sense of how the screen reader would convey the page to ensure residual content from the original page did not intrude on the experience of the newly created page (see References [29, 35] for discussions of challenges anticipating the screen reader experience). Prior research has pointed out that overemphasis on automated accessibility checks and strict adherence to accessibility standards can obscure whether the experience of navigating the page is frustrating and instead, an audio rendering and navigation of the page is needed to anticipate what a screen reader user's experience will be like [26, 29, 31, 38]. As recommended by this prior work, we used a combination of aDesigner and testing the pages with JAWS to ensure our pages supported grid-based or table tag navigation without biasing our experiment to pages formatted to ideal accessibility standards. This ensured our experiment pages had higher ecological validity for our use case where screen reader users could submit navigationally difficult pages for automatic transformation.

5.1.4 Workflow for Rewriting Pages with Tables. We used a workflow for wrapping webpages with table tags using a human-in-the-loop approach found in prior work [12, 20]. A member of the research team, who was an expert user of the EnTable plugin and not involved in user testing, wrapped the task pages with table tags using the plug-in. Future designs envision a service that would support a screen reader user being able to submit for revision navigationally difficult pages to a crowdworker, and automatically receive a modified page that would be more amenable to the range of shortcuts available through a screen reader (see Reference [20] for more detail). This wrapping activity consisted in first selecting pages from the web that met the experiment design criteria of representing a day-to-day task and using grid-based design. Then, we submitted selected pages to EnTable with information detailing how to correctly wrap the page with table tags. These pages were then incorporated into our web app for the experiment.

5.1.5 Data Lookup Task. For each task, participants answered the task's question by completing our custom created web form. The question was labeled with the heading "Question" followed by the text. For example, for Figure 1's task using the original page, the participant was prompted with the question, *"What day had the lowest chance of precipitation?"* The subsequent text then instructed the participant to answer the question by following the provided link. After the question and the link, a text entry box was provided for participants to write in their answer and submit it using a button. Participants were given the options to answer the question, write in "no answer

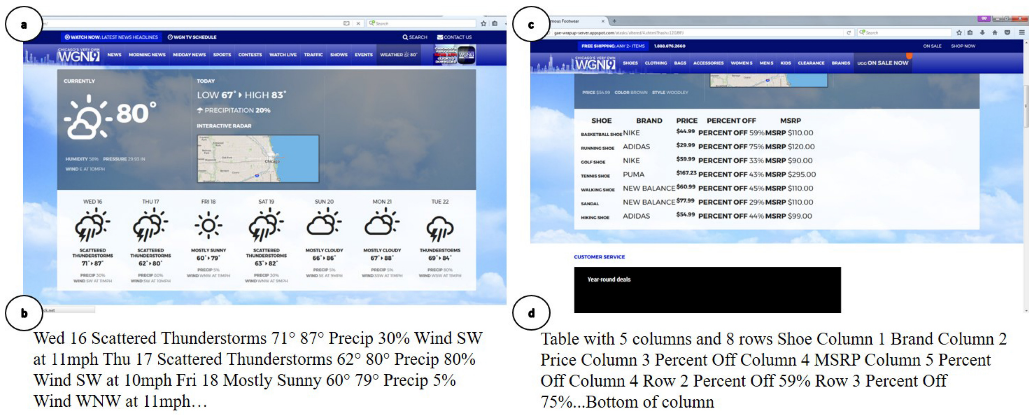

Fig. 1. The experiment's task pages are shown above for the fourth task type—find an item from a set that has the lowest value relative to preset criteria—both in the visual semantics condition (a) and in the table tag condition (c). The page's audio feedback when using a screen reader to navigate is shown below each page (b and d). In this task set, the weather page is the original page as found on the web and the shoe page is the page rewritten (to control for learning effects) and wrapped with table tags. Task sets were counterbalanced across conditions so that half of the participants used the weather page in the visual semantics condition and half used it in the table tag condition.

found" for cases where they thought the answer was missing, and "pass" where they had become so frustrated with navigating the page that they did not wish to continue with the task.

5.1.6 Measures. We measured speed and accuracy for each task. For speed, we measured the start time as the time the participant clicked on the provided linked page and the end time as when they pressed the submit button. For accuracy, we graded whether or not the participant provided the answer to the question as given by the intended web page. So, for instance, while the brand of an item might be casually accepted as an answer to a question asking about a shopping item, we graded this as incorrect (we carefully worded our questions with descriptions to make it clear what answer was accepted, i.e., using the name of the content's label). Whenever a participant answered "pass," we graded this task as incorrect. After each task set, we asked participants to rate their experience doing the tasks on 1–7 semantically anchored scales in terms of effort to complete tasks, memorization of page layout, ease of navigation, confidence in correctness of submitted answers, and their understanding of relationships between the page elements. These measures were designed to adapt the NASA Task Load Index to our experiment's research questions based off the findings from our pilot and to support our oral protocol [23]. This instrument allowed us to assess the impact of page layout on task demands when the page's content is presented in serial order as found in audio interfaces.

5.1.7 Apparatus. Participants completed our experiment using the browser and screen reader setup of their choice. Participants could complete the experiment at their home, at a local library for the blind/non-governmental organization, at a university, or at their workplace. We encountered two computing setups—one at a participant's home and one at the participant's place of work—where the experiment could not be completed on the participant's setup. In the first case, the participant's screen reader and browser conflicted with one another for navigational control over the page. In the second, the participant's web browser was so outdated that it would not render a web form. In these two cases the participant was given the option to use a Lenovo Thinkpad T450 laptop running Windows 7 with JAWS version 18.0, and web browsers Internet Explorer, Mozilla

Firefox, and Google Chrome. This environment was also used when participants chose to come to the lab.

Participants who used our provided laptop were given the option to plug in a mouse, keyboard, and speakers as they wished. They could either use their own peripheral devices as in the cases for the above two participant (one setup occurred at home and the other at work). Or, if they requested them for the lab setting, then we provided one. When the participant loaded our experiment's page, the researcher assigned the participant's condition and then opened the page with our experiment's web form.

5.1.8 Procedure. Like the pilot, the procedure was calibrated to last 60–90min, took place at a location of the participant's choice, and used the participant's preferred access technology (requiring only a stable Internet connection and web browser). When we first recruited participants, we confirmed what screen reader they would be using for the experiment and directed them to the screen reader appropriate WebAIM link to review their screen reader commands for finding information on a page and navigating a table [2]. In cases where the screen reader commands for the participant's screen reader were not hosted on WebAIM (this occurred only once), the researcher located and shared the equivalent commands for the participant's screen reader. When the experiment began, participants first completed a background survey (administered orally) covering their demographics, screen reader usage and expertise, and both work and education history.

Then participants completed four practice tasks. During the first practice task, the researcher instructed the participant to search the entire page using the find command, locate a table, read across a table's row, read down a table's column, and read aloud the table headers. For the second practice task, the researcher reviewed strategies for interpreting the task question and locating an answer on the page. The participant then completed two additional practice tasks on their own. The practice tasks used a simple table layout with column headers (see Reference [8] for impacts of table complexity). They also exposed participants to cases when spelling and content distinctions were important to find the correct answer, when empty table cells, when tables were used for page layout, and when there were multiple tables on a page. All of the practice tasks consisted of pages found in their original form from the web so as not to bias participants to the visual semantics or table tag condition. In short, the practice tasks were designed to be somewhat more difficult than the test tasks so that participants could ask any questions if they were uncertain what to do—such as remembering the screen reader commands or what they should answer if they did not think the answer was on the page. Then, participants completed two task blocks and a semi-structured interview as in the pilot while being observed by the researcher. The researcher ensured that the participant was not distracted during participation in the experiment.

5.1.9 Hypotheses. Based on our pilot study and prior research [20, 21], we hypothesize that screen reader users will be faster, more accurate, and have a better navigational experience using web pages rewritten with table tags. Our hypotheses are motivated by the fact that table tags make the visual relationships implicit in web pages explicit to the navigational experience of a screen reader. We detail our hypotheses more below.

H1. Speed: *Participants will be faster when using pages rewritten with table tags than when using pages with grid-based design.* Pages with table tags support two degrees of navigational freedom with a screen reader. So, they will require traversal of fewer page elements and thus, take less time to navigate.

H2. Accuracy: *Participants will be more accurate when using pages rewritten with table tags than when using pages with grid-based design.* Tables support screen reader feedback on relationships

between page elements such as reading out column headers before cell content, and so participants will be better able to infer relationships across page content.
H3. Perceived Experience: *Participants will rate tasks in the visual semantics condition as having greater demand than the table tag condition for all perceived task demand measures given below.*

(1) **Effort.** *Participants will rate the visual semantic condition as requiring greater effort than the table tag condition.* Pages rewritten with table tags will lower the effort required to find needed information, because row and column organization will support identification of related information.
(2) **Memory.** *Participants will rate the visual semantic condition as requiring more memorization than the table tag condition.* Pages rewritten with table tags will lessen demand on working memory to synthesize content by making repeated patterns of content explicit in auditory feedback by reading out content metadata.
(3) **Ease of Navigation.** *Participants will rate the visual semantic condition as more difficult to navigate than the table tag condition.* The navigational freedom of tables will be perceived as easing search demands by providing multiple routes to the same piece of information.
(4) **Confidence.** *Participants will rate their confidence in their submitted answers for the visual semantics condition lower than in the table tag condition.* Pages rewritten with table tags will lessen uncertainty in identifying page information by relating content with descriptors through linking cell data with column and row headers. So participants will have higher confidence in the information identified as a candidate answer.
(5) **Understanding.** *Participants will rate their understanding of the page relationships in the visual semantic condition lower than in the table tag condition.* Because tables arrange similar information nearby on the page, participants will understand page relationships better on pages rewritten with table tags.

5.1.10 Exploratory Questions. When we piloted our study, participant performance suggested that pages requiring more traversal of page elements contributed to fatigue and frustration. To ensure participants completed the full set of tasks, we presented task types in order of task type difficulty so that participants benefited from learning gains across task types. Thus, we did not randomize presentation of task type, and include task type as an exploratory question only (note, we did counterbalance task sets across participants, see below for experimental design).

Task Type Speed: *Participants will be faster at tasks requiring greater comparisons across sub-elements of a page in the table tag condition than the visual semantics condition.* Content in pages rewritten with table tags can be skipped, because table-supported, two-dimensional navigation provides a navigational route directly between sub-elements. So, the more these routes are used, the greater the speed benefit.
Task Type Accuracy: *Participants will be more accurate at tasks requiring greater comparisons across sub-elements of a page in the table tag condition than the visual semantics condition.* Tables support screen reader feedback on the page structure, and so, the more tasks rely on inferring relationships across page content, the greater the accuracy benefit.

5.1.11 Design & Analysis. We used a within subjects design with the page condition (visual semantics pages or pages with table tags) as our independent variable. For each condition, all five task types were completed in succession (see above for discussion). The order of the task sets and order of condition were fully counterbalanced across participants, and participants were randomly assigned to an order. Participants were asked to rate task demands on 7-point scales following the pattern *"On a scale of 1 to 7, with 1 being LOW/LITTLE X and 7 being HIGH/MUCH X, please rate* *how X"* (see Reference [39] for discussion of oral administration of Likert scale measures to those with visual impairment).

Our dependent variables were task time, task accuracy, and subjective rating of experience for each condition. We use paired t-tests to determine whether table tags had a significant effect on task time analyzed according to task type. We do not compare across task types as we did not randomize task order to encourage full experiment completion. Instead, we provide descriptive statistics across task types to address our exploratory questions. For error counts, we use a Matched Pairs Sign Test [40]. For Likert data, we used Wilcoxon Signed Rank tests. To determine the strength of our results, we analyzed the effect sizes using Cohen's d and Pearson's r (see References [18, 41]).

Last, we coded our transcripts from our semi-structured interviews with deductive codes derived from previous work and found in the literature [21], as well as inductive codes derived from our pilot study interviews. To validate our code set, a person independent of the research team coded two randomly chosen transcripts independently, and we calculated interrater-reliability. We used Cohen's Kappa as our measure, and attained $\kappa = 0.71$. We report on these findings below.

5.2 Controlled Comparison Findings

From our measured observations, we found that pages with table tags had a significant effect on speed, accuracy, and perceived navigation experience. For our exploratory questions, we also found preliminary evidence that pages with table tags had a greater impact on the speed of participants when the task type required greater synthesis of the page's contents. We present these results in more detail below.

5.2.1 Demographics. We successfully recruited 21 participants (8 female, 13 male) ranging in age from 26 to 75 years of age. Participants identified with the following race or ethnicities: 4.8% Asian, 4.8% Hispanic, 85.7% White, and 4.8% Two or more races. Participants reported the highest level of education attained as follows: 9.5% high school degree, 14.3% associate's degree, 47.6% college degree, 23.8% master's degree, and 4.8% doctorate degree. 81% of participants reported being born blind or becoming blind at <5 years of age. Of the other participants, 4.8% became blind between 6 and 10 years, 9.5% between 18 and 25 years, and 4.8% between 36–45 years of age. When asked to rate their level of expertise using a screen reader on a 1–7 point scale with one being novice and seven being expert, participants rated their expertise as the following: 19.0% answered 3, 9.5% answered 4, 47.6% answered 5, 19.0% answered 6, and 4.8% answered 7. Of the most commonly used Internet browsers, 71.4% used Internet Explorer, 14.3% Firefox, 4.8% Chrome, and 9.5% Safari. When asked to describe which screen reader they most regularly used, 85.7% said they used JAWS, 9.5% used VoiceOver, and 4.8% used System Access. When asked to characterize how many hours they spent using the Internet the week before the experiment, participants responded the following: 9.5% spent <2h, 23.8% spent 3–10h, 28.6% spent 11–20h, 19.0% spent 21–30h, 9.5% spent 31–40h, and 9.5% spent >40h.

These demographics characterize our participant pool as largely made up of college educated, white adults who became blind before the age of 5 years. The majority of them rated their screen reading expertise as greater than a 5 on a 7-point scale and they primarily used JAWS with Internet Explorer to go online for >10h a week.

5.2.2 Task Speed. In general, participants were faster when completing tasks using pages that had been rewritten with table tags (M = 126.4s, SD = 72.0s, range: 36–463s) than in the visual semantics condition (M = 183.5s, SD = 120.7s, range: 40–709s). Paired t-tests revealed this difference was significant for the minimum criteria tasks, least value tasks, and two criteria tasks ($p < 0.01$; see Table 1 and Figure 2).

Table 1. Tasks Using Table Tags were Significantly Faster When the Lookup Task Required the User to Synthesize Across Page Elements such as Finding a Value that Meets a Minimum Criteria, that is the Smallest of a Set, and that Meets Two Criteria

Task	Time (s) Visual Semantics *M*	Visual Semantics *SD*	Table Tags *M*	Table Tags *SD*	*t-value*	*p-value*[a]	*Cohen's d*	Accuracy Visual Semantics *Hits*	Visual Semantics *Misses*	Table Tags *Hits*	Table Tags *Misses*	*M*	*p-value*
Find Value	119.2	75.2	114.6	83.8	0.18	0.857	0.04	18	3	21	0	1.5	0.13
Missing Value	170.6	105.8	129.8	98.6	1.26	0.214	0.39	19	2	15	6	-2.0	0.96
Minimum Criteria	222.9	118.7	118.0	49.3	3.65	0.001**	0.99	17	4	19	2	1.0	0.31
Minimum Value	192.6	103.5	122.8	46.4	2.75	0.009**	0.66	7	14	10	11	1.5	0.25
Two Criteria	265.9	158.9	146.9	62.7	3.12	0.003**	1.13	8	13	15	6	3.5	0.033*

Speed and Accuracy for Each Data Look-up Task

Table tag tasks were significantly more accurate when the lookup task required a value meeting two criteria. Each task's descriptive statistics, significance test results and effect sizes are given above.
[a]Tasks with * show significant differences between the table tag task and the visual semantics task at p < 0.05, and tasks with ** at p < 0.01.

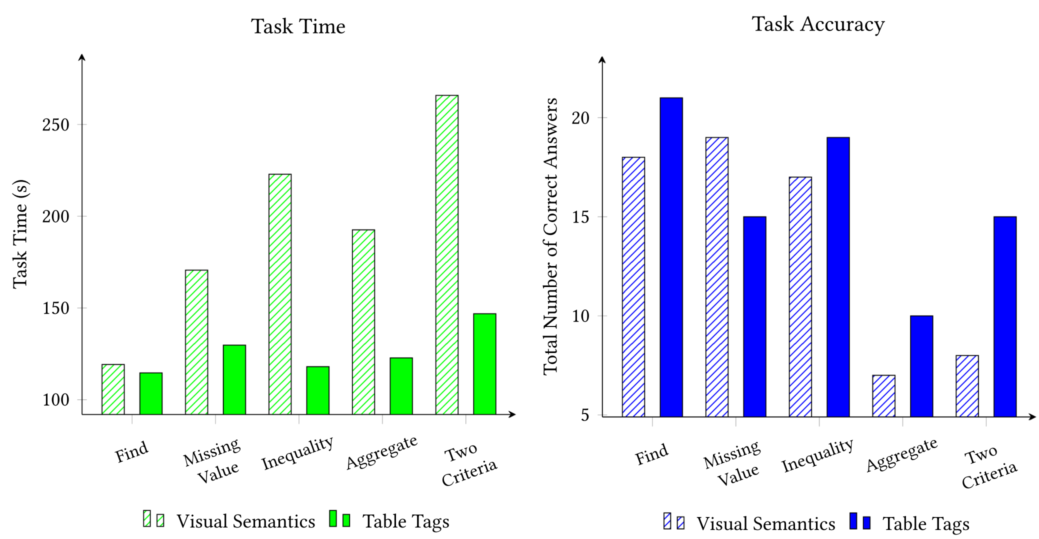

Fig. 2. Average task times across all participants were lower for each task in the table tag condition than in visual semantics condition. In total, participants were more accurate on all tasks in the table tag condition than in visual semantics condition, except for the second task (when the answer was missing from the page). The differences in time were significant for the inequality, aggregate, and two criteria and in accuracy for the two-criteria task.

5.2.3 *Task Type Speed.* When using pages from the visual semantics condition, participant task times showed an upward trend as the task required greater synthesis of page content (average visual semantics times: $t_1 = 119.2$s, $t_2 = 170.6$s, $t_3 = 214.5$s, $t_4 = 184.8$s, $t_5 = 228.2$s). This trend was not present for task times using pages modified with table tags (average table tag times: $t_1 = 114.6$s, $t_2 = 129.7$s, $t_3 = 118$s, $t_4 = 122.8$s, $t_5 = 146.9$s).

Table 2. Participants Rated Their Experience Doing Data Look-up Tasks Using Pages with Table Tags Significantly Better than When Using the Visual Semantics Pages on All of Our Collected Subjective Measures

Experience	Perceived Experience of Data Look-up Tasks				p-value	Pearson's r
	Visual Semantics		Table Tags			
	Mdn	*Range*	*Mdn*	*Range*		
Effort	3	1–5	6	2–7	<0.0001	0.96
Memorization	4	1–7	6	2–7	<0.0001	0.98
Ease of Navigation	3	1–5	6	3–7	0.002	0.67
Confidence	5	1–7	6	2–7	0.015	0.53
Understanding	5	2–7	6	2–7	0.003	0.66

Summary statistics, significance results, and effect sizes are shown above.

5.2.4 Task Accuracy. Participants were more accurate when completing tasks using pages that had been rewritten with table tags ($M = 3.81$, $SD = 0.96$, range: 2–5) than when using the visual semantics pages ($M = 3.29$, $SD = 1.24$, range: 1–5). A Matched Pairs Sign test revealed the page condition had a significant effect on answer accuracy when the task consisted in finding a value meeting two criteria ($p < 0.05$; see Table 1). However, while participants were more accurate when using table tag pages (error rate: 24% table tag condition, 34% visual semantics condition), they also gave a wrong answer more frequently when using table tag pages (wrong answer counts: 24 table tag condition, 11 visual semantics condition). We revisit this discrepancy in the discussion section below.

5.2.5 Task Type Accuracy. As tasks required greater synthesis of the page's content, participants passed on the task more during the visual semantics condition (number of passes: $t_1 = 0$, $t_2 = 0$, $t_3 = 1$, $t_4 = 2$, $t_5 = 4$). In contrast, participants did not pass on any of the tasks during the table tag condition. Similarly, participants answered that they did not think the answer was on the page the more the task required synthesis across page content during the visual semantics condition (number of "no answer": $t_1 = 2$, $t_2 = 0$, $t_3 = 3$, $t_4 = 8$, $t_5 = 5$). In contrast, "no answer" was given only once during the table tag condition (t3). Participants also provided incorrect answers more frequently as tasks required greater synthesis during the baseline condition (number of wrong answers: $t_1 = 1$, $t_2 = 2$, $t_3 = 0$, $t_4 = 4$, $t_5 = 4$). This trend was not present during the table tag condition (number of wrong answers: $t_1 = 0$, $t_2 = 6$, $t_3 = 1$, $t_4 = 11$, $t_5 = 6$).

5.2.6 Subjective Ratings. Participants rated the task demands of the table tag condition lower than those of the visual semantics pages on all dependent measures, and these results were significant for all dependent measures at $p < 0.05$. We summarize these findings below and report the results of inferential statistics in Table 2.

Effort. Participants rated the visual semantics pages as requiring more effort to use (Mdn = 3, range: 1–5) than the pages with table tags (Mdn = 6, range: 2–7).
Memorization. Participants rated the visual semantics pages as requiring greater need to memorize the page's contents (*Mdn* = 4, range: 1–7) than the pages with table tags (*Mdn* = 6, range: 2–7) to successfully complete the task.
Ease of Navigation. Participants rated the ease of navigating visual semantics pages (*Mdn* = 3, range: 1–5) lower than that of navigating pages with table tags (*Mdn* = 6, range: 3–7).
Confidence. Participants rated their confidence in their answers taken from pages using visual semantics (*Mdn* = 5, range: 1–7) lower than their confidence in answers taken from pages with table tags (*Mdn* = 6, range: 2–7).

Understanding. Participants rated their understanding of relationships between page content of pages using visual semantics (*Mdn* = 5, range: 2–7) lower than their understanding of relationships between page content using table tags (*Mdn* = 6, range: 2–7).

In summary, pages using table tags facilitated faster task completion times for tasks requiring participants to synthesize across page data such as meeting one criterion, finding the least value of a set, or meeting two criteria. On this last task, participants were also more accurate. Pages using table tags also had a significant effect on perceived task demands. Participants rated pages using table tags as requiring less effort, easing memorization and navigation, and providing greater understanding of relationships between page content. Participants also rated their confidence in their submitted answers using pages with table tags higher than when using the visual semantics pages even though they were more frequently wrong when using the pages with table tags. Our exploratory results further suggest that when pages using visual semantics are unmodified, participants will more frequently give up on the task or conclude that the answer is not on the page the more synthesis the task requires.

5.3 Interview Findings

During our semi-structured interviews, we found that tables provided a predictable and consistent auditory presentation of page information. Table keyboard shortcuts to control audio interaction coupled well with participant approaches to answering the task questions. Participants reported being slowed down when needing to consider more and more information (either required by task demands or page layout) and when they encountered pages that violated their expectations for reasonable page layout. We review these results in more detail below.

5.3.1 Fluid Navigation. The consistent audio cues of column and row headers in table markup supported quick adoption of a task strategy. Nineteen participants described ways in which navigation proceeded quickly and fluidly in the desired way. Of these 19, 13 thought that web pages formatted using table mark-up supported better navigation. These 13 participants described how they frequently became uncertain whether the screen reader had fully presented the page's information, and they thought the table format provided more certainty. Specifically, tables standardize audio and sequential cues and so, stabilized expectations about what information would be presented next. P8 highlighted how the table's standardization helped with the task to find a flight: "*They were consistent, so if you know that a row would have flight times, then it was always flight times.*" These cues included reading out metadata in the forms of column and row headers, or those for the current cell, characterizing sequential relationships, and accurately describing portions of the table that were irrelevant and so could be skipped. The table format enabled participants to predict what the subsequent page structure would be like and so adopt *"a kind of process of elimination"* (P13). Confirming our hypotheses that many templated pages would benefit from table markup, P10 described how implicit patterns in the page can be made explicit with tables: *"Everything that is a repetitious item, all the items make sense as, you know, being part of a table."* Once these items are wrapped with a table, the screen reader can make these patterns explicit with repetitive audio cues.

5.3.2 Page Structure. Page navigation strategies depended on whether the page structure and content would be amenable to using certain screen reading commands. Twenty participants detailed the strategies they adopt to get an overview of the page. Common strategies included going through the headers, skimming quickly past ads, searching for keywords found in the prompting question, and creating a list of links. However, 13 participants highlighted how they adopt these strategies given their expectations for how the page would be laid out or how the information would be organized. P16 described a mismatch between page structure and its information in the visual semantics condition: *"The columns in the second [set of tasks] were not in line."* While P16 recognized a repeating pattern in the page's information, it was not presented in a data structure that allowed easy cross-comparison using screen reader commands. P10 expressed frustration at this: *"That kind of always irritates me, because then you can't use table commands. You know, they look like they're in columns and you should be able to do them with table commands, but you can't."* Mismatches between the page information and its format in the visual semantics condition misled participants in planning a navigation strategy.

The implicit structure of the page's content cost participants time as they had to figure out what pattern they could best leverage in their navigation strategy. Eleven participants characterized ways that they tried to infer a pattern from the page's content. However, this added an additional task step: *"You sort of had to get an initial feel for how the data was laid out . . . These last tasks were not standardized, so you kind of had to figure out what their method was"* (P18). Inferring the page content's pattern was also prone to false navigation starts as participants tried to capitalize on their expectations. P2 detailed how this cost compounded during the task search for an item meeting multiple criteria. *"My first assumption was that they would be listing the departure times in order, but they weren't, so after going down the list of departure times, I had to navigate over a few columns to find departure and arrival times . . . and I guess the first thing I expected was for them to be listed in order . . . but they may have been ordered more by price."* Although P2 sought out an implicit pattern to hasten the search, the page's ordering by price mismatched P2's expectation that the results would be ordered temporally. As P2 searched for a flight during a particular time frame, the time to navigate over unrelated information quickly added up as P2 eliminated flights one by one.

5.3.3 Task Load. Attending to the page's information while remembering the task's question became increasingly difficult if the page required inferring relationships either to answer the task's question or because of the page structure. Sixteen participants described ways in which experiment tasks compounded to quickly become formidable. As expected, tasks requiring more synthesis of page information were more difficult. P13 attributed this difficulty to demands on short term memory: *"You had to look for two or three pieces of information, and you had to compare two or three things in your head at the same time."* 13 participants thought that balancing attention to search criteria with making data comparisons impacted their ability to remember all of the relevant information. P12 explained how managing attention, the search criteria, and the data slowed the task down, *"remembering what I was looking for . . . [and] listening to that information, so I was going back . . . that's when I went back and forth on where—because I thought there was something else."* Tasks that asked participants to synthesize across information presented by the screen reader substantially increased demands on attention and memory, because attention was already being paid to search criteria and page structure.

Tables lessened task demands by supporting navigation strategies that helped manage demands on memory. Six participants used a process of elimination strategy to lessen the amount of information they had to keep track of. P7 characterized how this strategy helped manage active listening to presented information alongside attending to the search criteria: *"What I would do is go along, and whenever I saw one that was lower than what I had seen as the previous low one, I would think that would be the best guess unless it's something below that."* The table format supported the elimination strategy, by enabling rapid comparison of related page data. P19 chose *"to read down the column until the first number [they] heard being more than [the search criteria]."* P4 highlighted the way headers would be read out as he passed through each row in a column helping him keep track of where he was at. In contrast, the visual semantics condition increased demands on synthesis and memory. Participants felt they had to *"remember where [they're] at, or count how many, like* *of this number, or the percentage is three cells away from the date"* (P4). Although the table format facilitated scanning to eliminate possible answers, participants still needed to consider a substantial amount of information to be effective. P13 reported having *"to go through the table two or three times trying to weed out in [his] head which one, you know, which one [he was] going to choose."* While tables supported direct comparison of page data when adopting an elimination strategy, this did not prevent participants from having to traverse the same information several times.

5.3.4 Ecological Validity. Participants largely reported that the experiment represented their typical experience, but highlighted current limitations where table tags may help. Although 6 participants admitted that they rarely use tables or avoid large datasets altogether, 13 participants thought the search tasks and pages were representative of their web browsing. These activities included shopping, looking up time tables for a bus or train, reviewing financial information, and looking up baseball statistics for favorite players. A few participants thought there was a role for using tables to make highly specialized tasks encountered at work accessible. P8 thought that tables could make article metadata available specifically for *"language in translation kind of things . . . if I'm looking for a particular manuscript it might have a table with the provenance and the different date and things like that."* P14 described how mandatory work trainings often fail to consider the fact that his screen reader needs to interpret the materials, which often come bundled in proprietary presentation formats and software that are inaccessible to his screen reader.

One or two participants thought that the datasets were smaller than what they typically encounter. Another two participants recognized and were impressed with the transformation of templated pages to tables. P4 remarked, *"The ones where it was setup like headings, that was really different. I've never seen it like that before. Those were the ones I think I got through the fastest."* In general, 19 participants thought the experiment setup fairly represented their typical experience using a screen reader to look for information on the web.

5.3.5 Summary. Participants described ways that tables support consistent audio presentation of page information, enabled quick adoption of page navigation strategies, and supported ready cross-comparison of page content. Participants reported that tables made page patterns explicit that would otherwise require making inferences about page content. Thus, tables can lessen task load especially in cases where multiple pieces of information must be attended to at once. While participants envisioned a role for pages with table tags in their daily tasks, they also thought it could be used to transform highly specialized materials encountered at work or could be used to transform a page when a pattern in the page's information was available.

6 DISCUSSION

While Gardiner et al.'s exploratory study found that participants reported tables enabled faster and more successful completion of navigation tasks, our findings validate these initial reports and show when these findings hold and when they do not [21]. Pages rewritten with table tags facilitated navigation when search criteria required participants to synthesize across page content by providing a stable audio presentation of the page information. However, this only addressed part of the task demands. When tasks required participants to keep track of a substantial amount of information including search criteria, page information, and the need to synthesize results, participants reported being overloaded. We consider how some of our observations may address these limitations below but also outline further questions and future work that these findings give rise to.

6.1 Confident Attempts

Although participants had a lower error rate in the table tag condition, their total number of incorrect answers was also higher than in the visual semantics condition. This finding can be partly explained by our accuracy measure. We penalized cases when participants passed on the task or concluded that the answer was not on the page. Because participants did not pass on tasks and concluded the answer was missing only once during the table tag condition, their error rates benefited.

At the same time, participants indicated that they were more confident in their answers in the table tag condition and thought they had a greater understanding of the relationships between the page's content. This finding suggests that the table structure supported participants in feeling they better grasped the page's material and so increased the number of attempts they made at answering the task's question during the table tag condition. Previous work found that when websites induce uncertainty due to users' unmet expectations or accessibility barriers, this can impact the confidence a user has in their own skills [7]. Our findings suggest that by lowering uncertainty over page structure, table tags support greater user confidence that they found the needed information. However, because participants were aware of which condition they were in, their frequent attempts could also be an artifact of satisficing and so should be interpreted with caution.

6.2 Whitespace

Notably, the most frequent errors in the table tag condition occurred when the answer was missing from the page. Despite the table formats' support for reading column and row headers for a table cell with missing information, and reading the cell information as blank, this formatting was not a sufficient audio cue for making missing information explicit. This finding may be because the screen reader announces the same "blank" audio cue for the cell as it does for web page white space such as breaks in paragraphs. Thus, the "blank" audio cue conflates white space with missing information. Further study is needed to examine what audio cues are effective for conveying missing information.

6.3 Influence of Screen Reader Setup on Navigational Experience

While participants rated their experience in the table tag condition significantly better than the visual semantics condition, the mobility of the device used appeared to negatively influence their ratings of their navigational experience overall. Although all of participants' tested screen reader setups shared the minimum set of navigation commands required for our experiment (abilities to find keywords, vertically and horizontally navigate a table, and identify table headers), we observed cases where participants were not able to access the full set of keyboard shortcuts reported by software support documentation or WebAIM during the training portion of our experiment. This limitation was largely due to variation in hardware in participants' preferred screen reader setups. To survey the extent of this loss, we subdivided participants into groups based off the degree of mobility their screen reader device supported and compared their average experience ratings. When participants were grouped accordingly, we observed that their ratings on perceived navigation experience decreased by a full point overall (see Figure 3). We do not present these results as part of the findings, because the differences in number of participants across categories do not support drawing strong conclusions. Rather, we mention them here to discuss ways our experiment could have been impacted by the participant's screen reader setup.

Making a web page accessible to a screen reader may not be the most desirable form of assistive technology as the device design becomes more mobile. We observed that mobile devices lose the full screen reader command set as the device is more mobile. For example, even in a relatively small mobility change—like shifting from a desktop setup to a laptop—participants lose the screen reader command to invoke a cell's column and row headings, because most laptop keyboards eliminate the numeric keypad needed for the shortcut. As our experiment did not balance for mobility of device used, our experiment is heavily represented by participants who used desktop

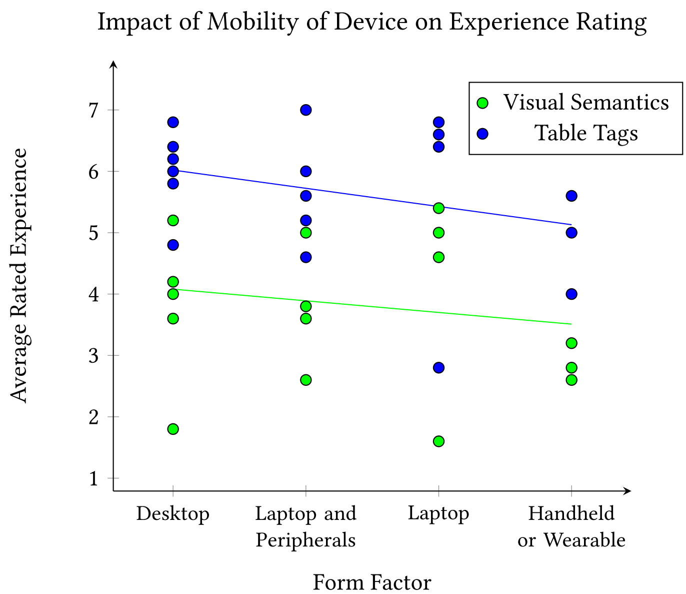

Fig. 3. Although participants rated their experience using pages with table tags significantly higher than the baseline pages, the mobility of the device may have impacted participants' experience overall. Ratings dropped by one point overall as the device became more mobile irrespective of the condition.

and laptop setups with only a few choosing to use a device more mobile than a laptop. Future research is needed to investigate interactions between mobile device design and support for a robust set of commands/navigational routes to access the same piece of information. While prior work suggests that mobile devices introduce accessibility problems not present in desktop experiences [24, 42], research on the screen size's impact on the usability of adaptive layouts for sighted users suggests that the mobility of the device is likely to interact with the perceived experience of the information's presentation [19]. The predictability of presented information may be more important for desktop setups than for mobile setups [19], and so, the predictability of screen reader feedback afforded by table tags could have less of an important role for mobile devices. A further study is needed to determine if there our limitations to our findings for mobile devices where presumably adaptive layouts have a more important role for adjusting page content to the device. However, whether the significance of device form factor for sighted users will hold for screen reader users is likely to differ [45].

6.4 Table Accessibility

Our findings contradict long held views that tables—as an example of a complex data structure—are inaccessible to screen reader users. Prior work reports that providing a table user with additional audio cues for the data structure worsens performance [33]. However, this earlier work used test conditions that simulated screen reader navigation for sighted users. Our findings uphold consensus within the research community that users independent of the target group are not good proxies for testing how assistive technology will be adopted and used by the target group [1, 27]. In our pilot with 22 screen reader users and field experiment with 21 screen reader users, we found that for data synthesis tasks, tables provide a stable audio presentation of the page, lessen task load, facilitate cross-comparison, and couple well with problem solving strategies. This suggests that the task type, in part, determines when tables will be perceived as helpful. Our findings suggest that audio cues that assist screen reader users in shifting their strategies from complex questions to those used to answer simple questions when using tables function analogously to the visual cues that assist sighted users in their strategies [33]. These results encourage future work investigating interactions between task complexity and audio cuing for chunking and grouping parts of a page to lessen task demands.

7 CONCLUSION

In summary, we conducted a controlled experiment with 21 screen reader users comparing whether rewriting pages using visual semantics with tables supported a better navigational experience for screen reader users. We found that pages rewritten with table tags often facilitated significantly faster and more accurate data lookup. Overall, table tags also had a significant effect on perceived task demands including effort, memory, navigational experience, understanding of page relationships, and confidence in submitted answers. Participants described ways that tables support consistent audio presentation of page information, quickly adopting page navigation strategies, and ready cross-comparison of page content. While we found that table tags had a positive effect on screen reader navigation of pages adopting grid-based design, we also observed cases in which making a web page technically accessible was not sufficient to address the limitations of screen reader audio cuing or the limited navigation shortcuts available on mobile devices. Contrary to the established belief that tables are inaccessible, our results show that tables can facilitate navigation when users need to synthesize across page content and provide a promising approach for automatically adapting existing web pages to screen readers.

ACKNOWLEDGMENTS

We would like to thank Blind and Vision Rehabilitation Services of Pittsburgh, Golden Triangle Council for the Blind, the Pittsburgh Cultural Trust, Three Rivers Rowing, and the Carnegie Library for the Blind and Physically Handicapped in addition to the many of the members of these organizations for their help with recruiting for this study.

REFERENCES

[1] Cynthia Ann Bowman and Paul T. Jaeger. 2005. *Understanding Disability: Inclusion, Access, Diversity, and Civil Rights*. Greenwood Publishing Group, Westport, CT.

[2] 2017. WebAIM: Web Accessibility in Mind. Retrieved from http://webaim.org/articles/jaws/#tables.

[3] 2018. Bootstrap Toolkit. http://getbootstrap.com/.

[4] 2018. Cascade Framework. http://cascade-framework.com/.

[5] 2019. Apple Voiceover Documentation. Retrieved from https://www.apple.com/voiceover/info/guide/_1134.html#vo27959.

[6] 2019. WAI-ARIA. Retrieved from https://www.w3.org/WAI/standards-guidelines/aria/.

[7] Amaia Aizpurua, Myriam Arrue, and Markel Vigo. 2015. Prejudices, memories, expectations and confidence influence experienced accessibility on the Web. *Comput. Hum. Behav.* 51 (2015), 152–160.

[8] Dagmar Amtmann, Kurt Johnson, and Debbie Cook. 2002. Making Web-based tables accessible for users of screen readers. *Library Hi Tech* 20, 2 (2002), 221–231.

[9] Chieko Asakawa and Takashi Itoh. 1999. User interface of a nonvisual table navigation method. In *CHI'99 Extended Abstracts on Human Factors in Computing Systems*. ACM, 214–215.

[10] Cheiko Asakawa and Hironobu Takagi. 2000. Annotation-based transcoding for nonvisual web access. In *Proceedings of the 4th International ACM Conference on Assistive Technologies*. ACM, 172–179.

[11] Cheiko Asakawa and Hironobu Takagi. 2008. *Web Accessibility*. Springer-Verlog, London, UK. 231–260.

[12] Jeffrey P. Bigham, Richard E. Ladner, and Yevgen Borodin. 2011. The design of human-powered access technology. In *Proceedings of the 13th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, 3–10.

[13] Silas S. Brown and Peter Robinson. 2004. Transformation frameworks and their relevance in universal design. *Univers. Access Info. Soc.* 3, 3–4 (2004), 209–223.

[14] Jenny Craven and Peter Brophy. 2003. *Non-visual Access to the Digital Library: The Use of the Digital Library Interfaces by Blind and Visually Impaired People.* Manchester Metropolitan University, Centre for Research in Library and Information Management.

[15] Iyad Abu Doush and Enrico Pontelli. 2010. Non-visual navigation of spreadsheet tables. In *Proceedings of the International Conference on Computers for Handicapped Persons.* Springer, 108–115.

[16] Sylvia Egger. 2012. Flexbox Tests—Test Cases for CSS Accessibility Community Group. Retrieved from https://www.w3.org/community/cssacc/2012/06/12/flexbox-tests-test-cases-for-css-accessibility-community-group/.

[17] Ruslan Fayzrakhmanov, Max Göbel, Wolfgang Holzinger, Bernhard Krüpl, Andreas Mager, and Robert Baumgartner. 2010. Modelling web navigation with the user in mind. In *Proceedings of the International Cross Disciplinary Conference on Web Accessibility (W4A'10).* ACM, 14.

[18] Andy P. Field and Raphael Gillett. 2010. How to do a meta-analysis. *Brit. J. Math. Statist. Psych.* 63, 3 (2010), 665–694.

[19] Leah Findlater and Joanna McGrenere. 2008. Impact of screen size on performance, awareness, and user satisfaction with adaptive graphical user interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, 1247–1256.

[20] Steven Gardiner, Anthony Tomasic, and John Zimmerman. 2015. EnTable: Rewriting web data sets as accessible tables. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers and Accessibility.* ACM, 443–444.

[21] Steven Gardiner, Anthony Tomasic, and John Zimmerman. 2016. The utility of tables for screen reader users. In *Proceedings of the 13th IEEE Annual Consumer Communications & Networking Conference (CCNC'16).* IEEE, 1135–1140.

[22] Jon Gunderson and R. Mendelson. 1997. Usability of world wide web browsers by persons with visual impairments. In *RESNA Annual Conference.* 330–332.

[23] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of empirical and theoretical research. In *Advances in Psychology*, vol. 52. Elsevier, 139–183.

[24] Shaun K. Kane, Chandrika Jayant, Jacob O. Wobbrock, and Richard E. Ladner. 2009. Freedom to roam: A study of mobile device adoption and accessibility for people with visual and motor disabilities. In *Proceedings of the 11th International ACM SIGACCESS Conference on Computers and Accessibility.* ACM, 115–122.

[25] Rushil Khurana, Duncan McIsaac, Elliot Lockerman, and Jennifer Mankoff. 2018. Nonvisual interaction techniques at the keyboard surface. In *Proceedings of the CHI Conference on Human Factors in Computing Systems.* ACM, 11.

[26] Jonathan Lazar, Aaron Allen, Jason Kleinman, and Chris Malarkey. 2007. What frustrates screen reader users on the web: A study of 100 blind users. *Int. J. Hum.-Comput. Interaction* 22, 3 (2007), 247–269.

[27] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. 2010. *Research Methods in Human-Computer Interaction.* John Wiley & Sons.

[28] Clayton Lewis. 2013. Pushing the Raman principle. In *Proceedings of the 10th International Cross-Disciplinary Conference on Web Accessibility.* ACM, 18.

[29] Jennifer Mankoff, Holly Fait, and Tu Tran. 2005. Is your web page accessible?: A comparative study of methods for assessing web page accessibility for the blind. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* ACM, 41–50.

[30] Pavel Panchekha, Adam T. Geller, Michael D. Ernst, Zachary Tatlock, and Shoaib Kamil. 2018. Verifying that web pages have accessible layout. In *Proceedings of the 39th ACM SIGPLAN Conference on Programming Language Design and Implementation.* ACM, 1–14.

[31] Bambang Parmanto, Reza Ferrydiansyah, Xiaoming Zeng, Andi Saptono, and I. Wayan Sugiantara. 2005. Accessibility transformation gateway. In *null.* IEEE, 183a.

[32] Enrico Pontelli, D. Gillan, W. Xiong, Emad Saad, Gopal Gupta, and Arthur I. Karshmer. 2002. Navigation of HTML tables, frames, and XML fragments. In *Proceedings of the 5th International ACM Conference on Assistive Technologies.* ACM, 25–32.

[33] Enrico Pontelli and Tran Cao Son. 2002. Planning, reasoning, and agents for non-visual navigation of tables and frames. In *Proceedings of the 5th International ACM Conference on Assistive Technologies.* ACM, 73–80.

[34] T. V. Raman and David Gries. 1997. Documents mean more than just Paper! *Math. Comput. Model.* 26, 1 (1997), 45–54.

[35] Hironobu Takagi, Chieko Asakawa, Kentarou Fukuda, and Junji Maeda. 2004. Accessibility designer: Visualizing usability for the blind. In *Proceedings of the International ACM SIGACCESS Conference on Computers and Accessibility.* ACM, 177–184.

[36] Hironobu Takagi, Shin Saito, Kentarou Fukuda, and Chieko Asakawa. 2007. Analysis of navigability of Web applications for improving blind usability. *ACM Trans. Comput.-Hum. Interact.* 14, 3 (2007), 13.

[37] Shari Trewin, Rachel Bellamy, John Thomas, Jonathan Brezin, John Richards, Cal Swart, and Bonnie E. John. 2010. Designing for auditory web access: Accessibility and cellphone users. In *Proceedings of the International Cross Disciplinary Conference on Web Accessibility (W4A).* ACM, 4.

[38] Shari Trewin, Bonnie E. John, John Richards, Cal Swart, Jonathan Brezin, Rachel Bellamy, and John Thomas. 2010. Towards a tool for keystroke level modeling of skilled screen reading. In *Proceedings of the 12th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, 27–34.
[39] Shari Trewin, Diogo Marques, and Tiago Guerreiro. 2015. Usage of subjective scales in accessibility research. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, 59–67.
[40] Jacob O. Wobbrock and Matthew Kay. 2016. Nonparametric statistics in human–computer interaction. In *Modern Statistical Methods for HCI*. Springer, 135–170.
[41] Koji Yatani. 2016. Effect sizes and power analysis in HCI. In *Modern Statistical Methods for HCI*. Springer, 87–110.
[42] Hanlu Ye, Meethu Malu, Uran Oh, and Leah Findlater. 2014. Current and future mobile and wearable device use by people with visual impairments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 3123–3132.
[43] Yeliz Yesilada, Robert Stevens, Carole Goble, and Shazad Hussein. 2004. Rendering tables in audio: The interaction of structure and reading styles. In *Proceedings of the International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, 16–23.
[44] Yeliz Yesilada, Robert Stevens, Simon Harper, and Carole Goble. 2007. Evaluating DANTE: Semantic transcoding for visually disabled users. *ACM Trans. Comput.-Hum. Interact.* 14, 3 (2007), 14.
[45] Dongsong Zhang, Lina Zhou, Judith O. Uchidiuno, and Isil Y. Kilic. 2017. Personalized assistive web for improving mobile web browsing and accessibility for visually impaired users. *ACM Trans. Access. Comput.* 10, 2 (2017), 6.

Received July 2018; revised March 2019; accepted April 2019