

2Across: A Comparison of Audio-Tactile and Screen-Reader based Representations of a Crossword Puzzle

Hrishikesh V. Rao

University of Michigan, School of Information
Ann Arbor, Michigan
hrishir@umich.edu

Sile O'Modhrain

University of Michigan, School of Information
Ann Arbor, Michigan
sileo@umich.edu

ABSTRACT

Crosswords are a popular recreational game that relies on the spatial relationship between words. As a player answers clues, they begin to organize words to form an intersecting grid. A good non-visual representation should convey the interrelation of words and support the user in building a practical spatial image of the crossword grid. This paper looks at two approaches to representing a crossword puzzle for visually impaired users: a screen reader based crossword, and an audio-tactile crossword puzzle. We evaluate the designs in a study with 10 visually impaired participants. The audio-tactile representation was found to support the practical use of the crossword's spatial structure while the screen reader based puzzle leveraged participant's prior experience in navigating websites. The paper discusses critical aspects of our study and presents a perspective on the use of multimodal interfaces for such spatial applications.

Author Keywords

Accessibility; Audio-Tactile; HCI; Refreshable Braille Display.

CCS Concepts

•Human-centered computing → Accessibility design and evaluation methods; Usability testing;

INTRODUCTION

This paper contrasts the use of two unique systems for reading and manipulating spatial information on computers by people with visual impairments. The first system involves a combination of a full-page refreshable braille display and text-to-speech, we call it the audio-tactile system (AT). The second, system is a commercially available product that uses a more traditional screen reader technology and we refer to it as the screen reader system (SR). Our studies use the crossword puzzle as an example of an application that involves reading and manipulating spatial information. We recruited 10 visually impaired participants (5 participants for each of the crossword representation conditions) for our study in order to answer the

following research questions concerning the design of spatial applications that use multimodal representations:

- How well do a screen-reader and an audio-tactile system each assist in solving a crossword puzzle?
- How might we design accessible spatial applications for people with visual impairments that can convey the spatial properties of the puzzle, and that can support interactions that use spatial information?

In this paper, we contribute: (1) An evaluation of the usability of an audio-tactile crossword puzzle and a screen reader compatible crossword puzzle. (2) Design recommendations for developing accessible applications that use multiple modalities for representing spatial information.

RELATED WORK

Previous work on the representation of spatial information has covered applications such as static [2], [7] and dynamic [21], [22], [23] maps, tactile graphics [14] and E-readers [12] among others. Some of the relatively complex spatial applications discuss input-output interactions for website layout tools [5], GUIs [13], crossword puzzles [10], [15] and sudoku [3]. These applications require participants to construct and regularly update their remembered spatial image of the representation for performing a task on the interface. For the blind users, spatial information in such applications is often presented in non-visual ways, for example through electronically synthesized speech, sound, static and dynamic tactile graphics, and at times in a combination of these modalities. We draw upon the advantage of using non-visual modalities for presenting spatial information as illustrated in some of user studies conducted by the authors cited above.

Full-page refreshable braille displays are ideal for presenting large amounts of braille text and tactile graphics. Despite the limited real estate on the display, interactions like scrolling and panning have been proposed as ways of accommodating more information [13], [18]. In such cases, the tactile representation of spatial information comes with a few inherent challenges. Users often indicate that it requires more effort to read and integrate new information with existing memory for a tactile representation than for a visual representation [20]. When working with information involving a lot of text, announcing the content using synthesized speech can be more convenient than presenting it in braille. Speech, however, is limited due to the need to process information in time [8]. This makes the process of assimilating large quantities of information a lot

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than the author(s) must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

CHI '20, April 25–30, 2020, Honolulu, HI, USA.

Copyright is held by the owner/author(s). Publication rights licensed to ACM.

ACM ISBN 978-1-4503-6708-0/20/04 ...\$15.00.

<http://dx.doi.org/10.1145/3313831.3376207>

more demanding, creating a bottleneck for the complexity of information that can be conveyed. For complex applications such as crosswords, excel sheets, and Sudoku puzzles, using speech alone for supporting users in constructing a spatial image may, therefore, be problematic. Use of tactile overlays on touch sensitive surface by [4] and [7] explore audio-tactile interfaces, limited to querying interactions. Due to the static nature of these overlays, new overlays are to be manufactured when the underlying information is updated.

With regard to spatial perception, prior work supports the idea of an amodal representation of the spatial image in humans, irrespective of the modality of the source of information [6]. Essentially, an effective stimulus in the audio, the tactile or the audio-tactile modality should be equally successful in updating a spatial image. In [11], the authors discuss the potential use of combined "audio-haptic" modalities to present spatial information, such that the information presented via speech complements the information in the haptic modality.

With respect to the temporal nature of speech versus the spatial nature of a tactile representation, it is interesting to consider the multiple ways in which spatial images can be constructed. The literature on navigation suggests that people use one of two strategies to develop spatial representations, a route-based, or a survey-based perspective. Route based perspective is an egocentric representation of the space, where the information is mapped with the user as a reference. A survey based perspective is allocentric, so the information about an object is mapped using another object as the point of reference. Previous studies show that survey knowledge produces better performance in survey-based tasks (involving spatial relations and estimation of the distance between objects), while route knowledge produces better performance in navigational tasks [19]. Some studies also show better performance from using route based descriptions, regardless of the nature of the task [9]. This body of work while being relevant to navigation based tasks has not yet been considered with respect to the different ways in which verbal and spatial information is conveyed to blind users interacting with access technologies and to how it impacts the nature of their interaction with speech versus tactile information.

In this paper, we use the crossword puzzle as an example of a spatial application. Crosswords have two high-level interactions: querying and editing. Solving a crossword puzzle requires users to understand the spatial structure of the grid and the interrelation between words. They also require users to perform a combination of route tasks, such as scanning the boxes that make up a word and survey tasks, such as locating an intersection between words. In [10], researchers discuss that an ideal crossword should be able to facilitate "users in understanding how words are interrelated" and subsequently in creating an accurate mental image. In addition to this requirement, we propose that the crossword representation should also be able to support both route and survey based interactions.

In our study we evaluate a screen reader compatible crossword website (SR Crossword) developed by the American Printing House, and an audio-tactile crossword system (AT

Crossword) developed by the authors of this paper. The navigation process in the screen reader based crossword requires users to hop over the grid, one box at a time, using the arrow keys on the keyboard. The system announces location information, stored character and the type of box as the cursor moves from one box to the next. In comparison, navigating across the grid in the audio-tactile crossword puzzle involves users feeling the tactile representation of the grid, followed by pointing at a specific location and querying for information at that pin as shown in Figure 2. On a high level, the two systems are designed to support manipulation of spatial information, however, there are significant differences in the types of navigation they support. The SR Crossword has elements of egocentric navigation, where information acquired is about the position of the cursor that the user is currently moving inside the grid. On the other hand, the AT Crossword's tactile representation provides survey like information and the users start their interaction from the overview of the crossword it provides. We observe the differences in participant's approach because of this distinction and present our evaluation over the next sections of the paper.

SCREEN READER COMPATIBLE CROSSWORD PUZZLE

The Screen reader (SR) representation of the crossword is a web application (<https://crossword.aphtech.org/>) developed by the American Printing House. This website is compatible with several screen readers across multiple browsers and operating systems. It also allows users to load custom crossword puzzles. We used the NVDA screen reader on a Windows system at default settings for the SR crossword. The crossword interactions used browser interactions to explore the crossword website. These interactions are common between JAWS and NVDA.

Crossword Design

The crossword website has two lists of numbered clues, one for the clues which run in the across direction and the other for the clues which run in the down direction. The grid consists of empty boxes - to store the characters, and closed boxes - which are blocked off. The starting character of a word is always a empty box with a number that corresponds to a clue. According to the user guide of the crossword website, this number can refer to either an across clue, a down clue, or both.

Keyboard Interactions & Screen Reader Announcements. The website is designed with simple and straightforward interactions. Users "tab" and "reverse tab" (shift +tab) between the list of across clues the crossword grid and the list of down clues in the order shown in the Figure 1. The arrow keys let you navigate over the list of clues, and pressing enter on a highlighted clue drops the cursor on the starting character of the word in the grid. The system announces row & column information, Empty/Closed boxes and filled in characters (if any) when the cursor moves onto any box inside the grid. Users can move around the grid using the arrow keys similar to the interactions in a spreadsheet. The system uses sound cues as notifications for events such as when the cursor hits the edge of the grid, and when the last character of a word is typed in.

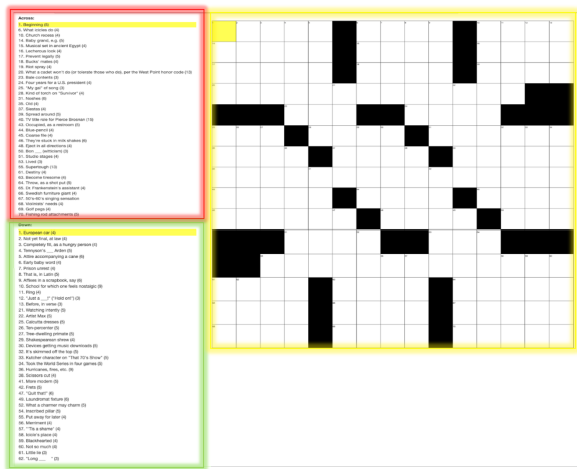


Figure 1. Screenshot of the SR crossword website (<https://crossword.aphtech.org>) with a red box highlighting the list of across clues, a yellow box highlighting the crossword grid and a green box highlighting the list of down clues

AUDIO TACTILE CROSSWORD PUZZLE

We have developed the audio-tactile (AT) crossword as a multi-modal interface on a full page refreshable braille display called the Graphiti [1]. The Graphiti communicates with a computer over a serial port and is controlled using the Graphiti APIs. The program also uses the computer’s built in speakers for making announcements and sound notifications, since the Graphiti does not have an onboard speaker system. Further details concerning the development of the AT crossword are reported elsewhere.

Full Page Tactile Display. Our AT crossword uses a full page refreshable braille display, developed by Orbit Research and American Printing House, called the Graphiti. The 2400 pins(60x40 matrix) can be set to one of 4 different heights, making it ideal for displaying tactile graphics. The crossword grid is represented as a matrix of raised pins, where each pin represents a box in the crossword grid. The touch sensitive surface on the refreshable pin matrix on the Graphiti allows co-located input-output interactions. For example, the haptic blink motion is a popular feedback interaction in the genre of refreshable braille displays. In the AT crossword, it serves to confirm if the system has read the user touch input accurately (the pin jumps under your finger) or if there was an error (the pin jumps elsewhere on the screen; the system then announces error message). The dynamic capabilities of the tactile output also serves as a landmark to connect information announced through speech to a physical location on the display.

Design Decisions

We followed [11]’s insights on multimodal representation of graphics in our design decisions when distributing information between the audio and tactile modalities.

- Insight 1: *"Adopting a less is more design philosophy; this is the best way to ensure that a tangible display will be both meaningful and usable."*

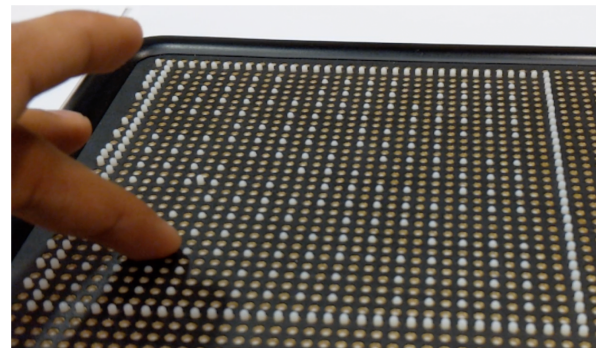


Figure 2. A.T. Crossword represented on the Graphiti. The tallest pins form borders and enclose the crossword grid. An empty box is represented by a raised dot, a filled box by a slightly lowered pin and closed box by completely lowered pin.

The AT representation uses the tactile display strictly for non-textual information that is crucial to build a spatial image of the crossword puzzle. Figure 2 shows the rows and columns in the crossword are spaced out by one line of lowered pins on either side to assist in accurate touch input. We vary the heights of the pins to signify empty, filled, and closed box. An initial pilot study with 3 participants to decide the combination of heights revealed that during exploration, taller pins "stood out" more. Therefore for levels ranging from 0 to 4, where 0 is a completely lowered pin and 4 is a completely raised pin, we assigned height level 2 for an empty box, height level 1 for filled and height level 0 for a closed box. Border pins stand at height level 4.

- Insight 2: *"Taking into consideration the capabilities of the perceptual systems for which the media will be designed; this requires consideration of how information should be distributed across modalities and how efficiently it can be picked up even within a given modality."*

Audio announcements present rich information about the status of the puzzle and add context during the tasks. For example, the system announces the structure of a partially filled word(such as "H, blank ,L ,L ,blank" when the word associated with a clue is partially filled- H_LL_); As the user fills in the missing letters, the system echos the characters that a user has filled in for the solution before they move on to the next clue, serving as a quick confirmation. Contextualized announcements refer to the system information announced during interactions. For example, the system announces the structure of a partially filled word when a user enters the edit mode; and summarizes the solution upon exiting the edit mode. The system also announces characters that were stored at a location during the edit mode so participants can accordingly hop over the existing characters if they do not want to change them.

- Insight 3: *"Supporting multimodal information in the interface; this is the best way to reach more users, via the most intuitive information delivery, for supporting the widest array of tasks."*

We operationalize this insight through our design of a multi-modal representation of the crossword. The system output

Interaction	Control	Description
Querying location	Point + space	Announces the location and currently stored character; Cursor appears on the pin that is pointed at to highlight the location
Requesting Clue	Point + Direction Key (Right or Down)	Announces the clue; The Cursor appears on the pin that is pointed at to highlight the location
Enter Character	Point + Select	Announces the location, and alerts the user to type. Cursor appears on the starting pin to highlight the location
Typing Characters	Braille Keyboard	Character announced, and cursor proceeds along the direction until interrupted or word is complete
Editing a character	Point + select	The Cursor appears at the pin that is pointed at to highlight the location. The System announces the location and alerts the user to type.

Table 1. Commands for Audio - Tactile Crossword puzzle.

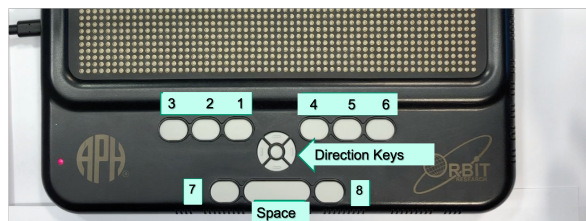


Figure 3. Layout of the braille keyboard on the Graphiti

can take advantage of a large information bandwidth by distributing information over the speech and tactile modalities. Speech is ideal for textually rich information such as text for the clues, repeating the solution that was entered, etc. In contrast, the tactile modality is suitable for conveying spatial information such as the shape of the grid, the location of a word within the grid and the intersection between words that constitute the grid. By carefully picking critical information for a set of tasks involving the crossword puzzle, we divide the information between the two modalities.

AT Interactions and walk through

The interactions for the AT crossword representation were designed to combine use of the Graphiti's touch input with its Perkins style braille input keyboard (Figure 3). Table 1 shows a list of possible interactions with the Graphiti and the controls.

A user starts with an empty crossword grid. To query for information about a location, the user points at the corresponding pin on the grid and presses the space bar on the Graphiti to have the information announced (supported with a blinking motion of the pin). Similarly, to ask for the clue, the user points to the starting character of the word and presses the right (for across clue) or down (for down clues) buttons of the Graphiti to have the clue announced. The bi-manual interaction helps to keep one hand of the user on the tactile representation at all times to read haptic feedback through the blinking pin.

The next step is to press select to enter the edit mode, and a cursor appears at the starting character of the clue. The user

can fill in the solution using the braille keyboard, and upon receiving a character, the cursor moves to the next spot in the word. The system announces if there are characters stored at a location when the cursor moves there. To exit the edit mode, the user presses the select button once again and the system repeats the characters that were entered for confirmation and the pins representing the solved word are lowered.

STUDY DESIGN

The objective of our study was to collect interview data, observational notes and participant comments to assess the two representations of the crossword puzzle. From this data, we learn about the strengths and the pitfalls of the two representations and their contribution in building a practical spatial image of the crossword.

We assigned participants to either of the representations at random, and used the same crossword puzzle across all participants for consistency. Although we designed the AT crossword to be functionally similar to the SR crossword in terms of possible interactions, there were certain biases that could not be prevented. For example, it is important to recognize that the SR crossword is a website purposed to solve a crossword puzzle. The interaction are similar to navigating a website. Participants benefit from their prior experience with screen readers in the SR condition, meanwhile none of the participants using the AT crossword had used a full page braille displays, as these devices are not yet available outside laboratories and research institutions. The AT crossword lacked consistency of touch input due to technical limitations. This meant, most often the errors occurred because of system limitations (and not that of the users), thus affecting the overall performance of the participants. Our study presents a contrast between the experiences of using the two systems, bearing in mind the advantages and the limitations of the two access technologies.

This study explores the usability of the two representations in an effort to improve the design of spatial applications using the comments from the users and our observations. Our research questions focus on the strategies used by participants in solving the crossword puzzle. We do not compare the two sys-

tems based on the performance of participants. Performance based comparison using metrics such as error rate, time for completion etc would be affected by the inherent advantages and disadvantages of the two systems. However, future studies can be set up to compare parameters that are quantitative and more descriptive of the performance.

Participants

We recruited 15 participants, age 18 and above, who were assigned one of the two systems at random. All participants were legally blind. All participants had prior experience working with a computer with a screen-reader (4 JAWS, 1 NVDA, 5 Multiple). Participants also had experience with a Perkins style braille keyboard (such as that incorporated with the Graphiti). After eliminating no-shows (2), poor data(2), and incorrect study(1) we used the data from 10 participants for our analysis. Data excluded for being poor included participants that did not complete the task, or appeared disengaged with the task. 6 participants (3 in SR group, 3 in AT group) of these 10 participants had prior experience with a crossword puzzle in the form of a static tactile representation, or with a visual puzzle - solved along with a sighted partner. It was, therefore, essential to go over the rules to ensure that participants understood the objective of the game during the training session. Each session lasted approximately an hour and was facilitated by a researcher. Participants were compensated \$20 in cash for their time.

Procedure

The participants were briefed about the objective of this study at the beginning of their session. Each participant answered a short verbal questionnaire about their demographic information, occupation, prior knowledge and self rated their skills which are relevant to this study. They completed a short training session in which the researcher introduced the participants to the system. This was followed by a practice session where the participants were encouraged to solve a few clues while the researcher was available to assist them through the issues they might encounter.

Once a participant reported that they were confident enough to proceed, the researcher set a timer for 20 minutes during which participants were asked to solve the puzzle. During the testing period the participants were encouraged to think out loud as they navigated through the interface to solve clues. The researcher who was observing participants during this period would briefly interrupt the participant between their attempts, to ask questions about their strategies and probe further based on their comments. This method of conducting brief semi structured interviews between two clue attempts is known as prompted retrospective think-aloud. It minimizes disturbance to the flow of a user solving the puzzle due to interruption and prevents the researcher from speaking over the voice of the screen-reader or speech output of the system.

Participants would not be interrupted if they guessed a wrong answer, and would be encouraged to try it out instead. This was a more natural way of solving a crossword, and provided an a chance to observe when participants could or couldn't catch and correct their mistakes. Participants were encouraged

to try out multiple strategies for solving the puzzle, such as: cycling through the list of clues to find a relatively easy clue, word and solving it. Participants were also asked about their interpretation of the structure of the crossword puzzle, including questions about their understanding of the direction of a word, their recollection of the location of a word within the crossword grid, and about their strategy for solving the puzzle. The session ended with a debriefing post-test interview to inquire about any challenges and frustrations they experienced. Participants also answered a verbal post-test questionnaire to provide feedback on their overall experience of solving the puzzle.

Evaluation

We transcribed the interviews and voice recordings, and combined them with the researcher's observation notes. The combined data was open coded [16] independently for the two group (based on the type of crossword puzzle) and then clustered within the groups. We reconstructed the crossword puzzle for all participants using the recorded data, and overlaid it with markers on location where they queried for a clues. The markers were color coded according to the direction of the clue for every query. We call these diagrams as *heat-maps*. The heat-maps of participants of a group were referred to during the clustering of the open codes.

RESULTS

S.R. Crossword Puzzle

1. Acquiring the unavailable information using familiar methods.

Participants solving the SR crossword lacked an easy access to some of the critical information of the crossword puzzle. We observed participants using methods from their experience in navigating websites to gather the missing information.

For example, P1, P5, P8 & P10 cycled over the list of clues and answered them in an order. They used the clue from their previous attempt as a landmark when querying for the next clue in the list (instead of the clue #). This way they answered clues without ever moving to or requiring spatial knowledge of the grid.

"I don't have a good idea what's going on the actual crossword puzzle. For me right now it's just me answering questions" - P1, when asked about their strategy solving the puzzle.

When in doubt, participants (P5, P10, P8) toggled between clues using the arrow keys to have the previous clue repeated. We speculate that the participants may have struggled to gather the necessary information from the first announcement because of the amount of information announced at one time. The clue announcement illustrated in Figure 4 shows the breakdown of the announcement text in the SR crossword, and the announcement text in the AT crossword. It is to be noted that clues in both representations may be longer and at times contain nuanced hints that require attention, to come up with the solution.

[After asking for the clue the first time] " did it say how many characters it had?" [repeats the previous clue to confirm the length of the clue] - P5

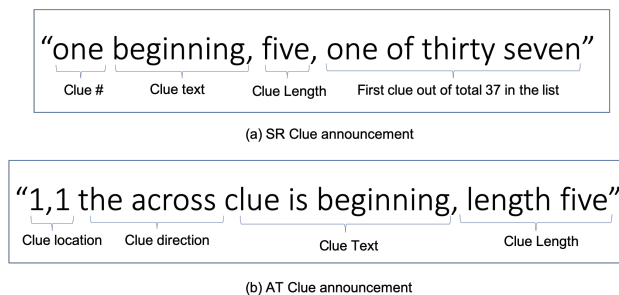


Figure 4. A breakdown of the clue announcements made in (a) SR crossword and (b) AT clue announcements.

2. Missing context, missing feedback and desired customizability

In the S.R. condition, we observed a need to customize the system for the purpose of a spatial application on two levels: (1) having interactions that are specific to the application and provide context when necessary, and (2) allowing participants to interact with the system using their custom screen reader and other assistive devices.

In the SR crossword, the system would announce the direction from the title of the list when navigating to it from other parts of the puzzle. This was also the only time the system announced any information about the direction of a clue. All participants expressed trouble remembering the direction over the duration of their attempt.

"I'm having trouble remembering what set (across or down) of clues I am in" -P1

"I wish it would say on the grid if this word runs down or across..." [After losing the location of a word which they were solving while navigating the grid] -P7

Few of the common errors among the participants originated because of the system was not able to provide feedback to the users about the current status of the crossword puzzle. For example P1, P8 & P10 would sometimes type in the answer without pressing the enter key so the website would not register their solution into the grid. The system did not provide a feedback in such situations. Instead, it announced the characters as they were typed in due to keyboard echo settings of the screen reader. This rather served as a confirmation from the system instead of notifying the user about their mistake.

P8 types in all characters of a word after forgetting to press enter on a clue. They immediately realize they are not on the grid, and go back to correct their mistake and enter the word.

Participant (P5, P8) unknowingly changes the direction of a word before they began typing the solution, by pressing the arrow keys while on the grid. The system doesn't notify them about this change in direction and registers their answer in the wrong direction.

Information such as whether a word is partially filled due to intersecting words, or if changing a character disturbs the intersecting word is difficult to gather without extensive explo-

ration of the grid, in the midst of solving a clue. It is however extremely important to have this information when solving the crossword. Not having an easy way to learn this kind of information could "undo" the user's progress in the crossword, despite having a website that is screen reader compliant.

"I wish it would give more like and audio clue. . . Let's say in a word and I already had a letter in there and I didn't have the right answer, and I type in a different letter, I don't know if it'll tell me that it replaced the letter. . .because I am undoing the work I have already done..." - P10

"Does it announce if I get the word wrong (after solving an intersecting word)? Because I wont be going over all of this to be able to double check..."-P1

The system did not notify participants P5, P7 and P8 when they overwrote a character that was already present in a grid. The new character modified the original answer for the intersecting word and users weren't made aware of this change.

Unfamiliarity with the system's voice was an issue for some of the participants. They also expressed a need to be able to use custom screen reader shortcuts for exploring the website. Accessing the table properties and shortcuts on the grid (P1, P8, P10) using the home button on the list of clues (P5,P8), and accessing the spelling of the clue to be able to read nuances (P8, P10) were some of the features that participants wished to use with the crossword application. A limitation of the design of our study forced the participants to only use an NVDA screen reader with default shortcuts which could not provide access to the custom settings participants used on their personal devices.

3. Difficulties in accessing spatial information from the grid.

We observed that exploring the area around a landmark on the SR crossword grid helped in building a spatial image in the vicinity of the landmark. However, on a larger scale, the system was unable to provide a practical spatial structure for the crossword that the participants could use to orient themselves. We speculate this may be because processing the keystrokes, announced coordinates and announced characters became more demanding as the size of the exploration increased.

Participants P1, P7, P8, & P10 were able to learn the structure of a partially filled word by "arrowing" over the word. This interaction was also used by P5, P7, P8, & P10 for reviewing a word it after filled in. In these cases, the word was used as a landmark to explore around it and build a local spatial image.

"... it's just more memory involved and more time spent kinda arrowing through things just so I know where I am ..." - P5.

"I can understand... hard if (I'm) in the grid... hard to tell where one word ends and other starts" - P7, when asked about their understanding of the structure of the crossword puzzle.

"I am having a little trouble remembering and then also sorta build a map in my head of what this puzzle looks like..." - P10

In their session of solving the crossword puzzle, P3 did not appear to notice the blink when querying. In one of the attempts at querying, the researcher observed the system announced clue for a different location than where P3 was pointing. The participant went on to solve the clue, without realizing that it was in a different location. The disconnect between the information made them inherently more confused about their location within the puzzle (also mentioned in P3's quote in the next section).

1.2 Distributing information between audio and tactile modalities.

Participants used the tactile features of the crossword as a reference when reviewing the status of the puzzle, such as for finding empty and filled out words by feeling the heights of dots. Participants also used this information to gather the overview of the puzzle. However, as the density of pins containing characters increased, a large neighborhood of pins would be lowered, rendering them similar to the closed boxes in the crossword puzzles. Participants (P4, P9, P2) complained about the confusion between the closed box and filled box, and at times mistakenly assumed one for the other. Following quote from P9 is an example of such issue, when they confused a cluster of closed boxes next to filled characters as part of the words they have solved, however, they did not remember filling them out :

"Alright lets see ... this has been filled out ... so now ... any other clues we have ... has this been filled out? [pointing to the closed box spaces]" - P9

A concern that was common across all participant was to have braille representation of characters of crossword on the tactile representation. Being able to read braille characters would allow participants to orient themselves in the grid using previously solved words as landmark.

"It's hard to get a picture of the whole thing since you can't see the real letters" - P3

"I can't tell you which words are where, I can only tell where the words are" - P4, when asked if they can locate a word within the grid.

To be noted, our decision to not use braille representation in the AT crossword was due to the large dot size and inter-dot spacing on the Graphiti compared to the standard. For a 15x15 crossword, a standard braille representation would occupy 47x62 which is larger than the dimensions of the display.

To access details like coordinates, or letters stored at particular grid points required participants to repeatedly query the grid through the braille/speech interface. The poor touch sensitivity of the Graphiti made it hard for the system to accurately locate a user's touch on the first attempt. This may be one of the reasons why participants preferred asking the researcher to read clues.

"Oh we almost got this section... [pointing to the lowered pins in the top left corner of the grid] That's the 'NISI', I think ... [pointing to 'KATE'. 'NISI' was in a different part of the puzzle]... are you allowed to tell me which one this is?" - P3

2. Spatial knowledge through tactile representation.

Participants P2, P3, P4 and P9 would review the lowered pins on the grid upon filling in the solution to a clue and exiting edit mode, while the system repeated the characters entered. This review gesture was observed several times over the study. When P9 typed in the solution DRIP which partially filled ('D ___'), they overwrote the first character with the remaining characters instead of hopping over the first spot. The word was stored as 'RIP_' upon exiting the edit mode. P9 noticed the last character was empty when they went back to feel the lowered pins and promptly used the querying interaction to ask for the character stored in the last spot. The system announced that it was 'blank':

"Blank?... it should be P from DRIP... let me see.. this word I thought it was easy... lets see... [asks for the clue associated with that word]" - P9.

The system announces clues when participants query for it at the starting character of the word. All participants struggled to identify the starting character of a word at some point during the study as the tactile representation does not have a clear landmark for starting characters and the direction in which the word ran.

"...forgetting where to go... I should have had more focused... instead of randomly just checking spots...I should have had more focused trip around the crossword..." - P4, during post test interview

Participants demonstrated a work around for this problem, utilizing the structure of the crossword puzzle to locate the starting character of the puzzle. P3 demonstrated awareness that words on the grid have either a border pin or a closed box on either ends, no matter which direction. Meaning that no two words are organized back to back. Thus the pin to the right or below a closed box or a border pin must be the starting character. It is to be noted that P3 had prior experience working with a crossword puzzle in tactile form.

"[After the researcher explains that the system announces clues only when asked from the starting character of each word] ...Oh... so that is the pins after the blank ones right (referring to the closed box represented by completely lowered pins)" - P3, during training.

In some cases, the puzzle had more than one word in a row/column. Here, users demonstrated different strategies for querying clues. P4 attempted words in an order from left to right for each row. As they filled in a word and those pins were lowered, it was easier to continue on that row to find the next raised pin which is the starting location of the next word in the row.

We also observed that not all participants used the blinking pin appearing as a cursor during edit mode. We speculate that unlike conventional braille displays that employ similar blinking cursors, the A.T. crossword does not display the characters near the blinking cursor so this information is not as useful for participants. Participants (P3,P6), who tried to look for the cursor during training expressed that it was hard to locate it on the grid. We believe this may also discourage

Participants expressed the desire to read braille characters on the tactile representation as an intuitive way to review characters and to orient using words as landmarks. However, using a braille characters on the grid would increase the size of the representation, and take away the high level spatial information offered in a non-braille representation. We believe that designing to allow switching between the two representations is a plausible compromise. Creating multiple layouts of representations based on the information they convey has been demonstrated on tactile representations before [17]. Having multiple layouts is especially beneficial for layered information. For example, in the case of a crossword puzzle, layouts could be used to distribute layers of properties of a box such as - its location in the grid, the clues that it is a part of, whether it is a starting character of a word, if it is empty or filled or closed off, the character it holds if it is filled.

The principal behind the distribution of information between the two modalities in our audio-tactile crossword was to announce textually heavy information over the speech modality and to represent spatial data on the tactile modality. We saw that this division of information can be improved with careful consideration of the context and the tasks entailed by the application. It can be further refined by working with people with visual impairments to incorporate their feedback and comments during the interactions.

A critical aspect in our SR crossword study was that the participants could use their proffered screen reader and screen reader settings and shortcuts which are customized as they are comfortable with. The crossword website was screen reader compliant, however, it did not usable in certain scenarios. For example, in the case of the crossword grid, many participants wished to "treat it as a table", like any other grid like representation such as spreadsheets or HTML tables. The lack of these aspects closed doors for the strategies that involve the participant's prior experience with screen reader based exploration of grids and tables.

The website representation of the crossword puzzle conveys the visual features of the puzzle, such as - empty and closed boxes, coordinates, clue numbers etc, through announcements. Spatial properties of the grid, such as intersections of words, partially filled characters, location of the word within the grid are relatively harder to access. Visual and spatial properties may have a few common properties in their subset, however in their entirety, they convey two different meanings. Therefore, translating the visual representation to a screen reader compatible application may not do justice to the spatial tasks to be performed on that application.

Participants demonstrated strategies where they utilize the spatial organization of the crossword website as a whole, for example, repurposing the clues from the clue list as landmarks to jump to different regions of the puzzle. However, our study revealed that it was not easy for participants to learn the spatial structure of the crossword grid. Their prior knowledge of keyboard interactions and their familiarity with how websites are organized certainly contributed to their performance in solving the crossword puzzle. A direct comparison between the data from the participants cannot be made, as prior experience with

screen reader disturbs the level playing field for a new technology like the Graphiti. Our studies highlight the contrast in participant strategies at solving the crossword puzzle using the two representations. The AT representation inspires several strategies that involve using the spatial properties of the crossword grid. The unfamiliarity with the audio-tactile interface may discourage participants solving the AT crossword from performing complex interactions in the midst of the task. The tactile display, as we learned, was useful in providing a high-level overview of the puzzle. We observed our participants successfully moving hands between the pin matrix and the keyboard when performing bimanual interactions. On the other hand, the SR representation allows participants to perform the task of answering clues without a need to understand the spatial layout of the crossword grid. Participants' prior experience in navigating websites on a computer assists in developing strategies for solving the SR crossword puzzle. They are also able to demonstrate localized spatial awareness in some of these strategies.

In future, it would be useful to explore system designs that leverage the combine the advantage of user familiarity with a screen reader with the audio-tactile system's ability to provide an overview of the grid's structure and current status.

IMPLICATIONS FOR DESIGN

This study evaluated two representations of crossword puzzles with significant differences in the types of navigation they support. The insights of this study enable the following design recommendations for future work on similar applications.

- Spatial applications need to be designed to support both high-level spatial tasks and local navigational tasks.
- Integration with user's existing devices and settings should be accounted for in the design of any system developed to convey spatial information.
- Through the combination of audio and tactile modalities, designers can make use of the direct perception of spatial information through the tactile modality and large bandwidth for textually rich information through the speech modality. Multi-modal interfaces allow for synchronous presentation of information by careful multi-modal cuing, explained further by the authors of [11].
- The context of the interaction must be considered when deciding the distribution of information across the modalities. The relevance of information is tied to the stage of the task a user is currently at. Designers of multi-modal interfaces can design effective interactions by carefully considering this shifting relevance of information when distributing the information between modalities.

ACKNOWLEDGMENTS

This work is supported by an NSF grant (Award 1741435). We also acknowledge support from the American Printing House (APH), Orbit Research, the Smith Kettlewell Eye Research Institute and the National Federation for the Blind. We would also like to thank James Coughlan, Vaishnav Kameswaran, Brooke Sasias and Eytan Adar.

REFERENCES

- [1] APH and Orbit Research. 2019. Graphiti® - a Breakthrough in Non-Visual Access to All Forms of Graphical Information. (2019). <https://www.orbitresearch.com/product/graphiti/>
- [2] Emeline Brule, Gilles Bailly, Anke Brock, Frederic Valentin, Grégoire Denis, and Christophe Jouffrais. 2016. MapSense: Multi-Sensory Interactive Maps for Children Living with Visual Impairments. (2016), 445–457. DOI: <http://dx.doi.org/10.1145/2858036.2858375>
- [3] René Gutschmidt, Maria Schiewe, Francis Zinke, and Helmut Jürgensen. 2010. Haptic Emulation of Games : Haptic Sudoku for the Blind. *PETRA '10 Proceedings of the 3rd International Conference on Pervasive Technologies Related to Assistive Environments* (2010), Article 2. DOI: <http://dx.doi.org/10.1145/1839294.1839297>
- [4] Shaun K. Kane, Meredith Ringel Morris, and Jacob O. Wobbrock. 2013. Touchplates: Low-cost Tactile Overlays for Visually Impaired Touch Screen Users. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '13)*. ACM, New York, NY, USA, Article 22, 8 pages. DOI: <http://dx.doi.org/10.1145/2513383.2513442>
- [5] Jingyi Li, Son Kim, Joshua A. Miele, Maneesh Agrawala, and Sean Follmer. 2019. Editing spatial layouts through tactile templates for people with visual impairments. *Conference on Human Factors in Computing Systems - Proceedings* (2019), 1–11. DOI: <http://dx.doi.org/10.1145/3290605.3300436>
- [6] Jack M. Loomis, Roberta L. Klatzky, and Nicholas A Giudice. 2014. Chapter 8 Representing 3D Space in Working Memory: Spatial Images from Vision, Hearing, Touch, and Language. In *Multisensory Imagery*. Chapter 8, 131–155. DOI: <http://dx.doi.org/10.1007/978-1-4614-5879-1>
- [7] Joshua A. Miele, Steven Landau, and Deborah Gilden. 2006. Talking TMAP: Automated generation of audio-tactile maps using Smith-Kettlewell's TMAP software. *British Journal of Visual Impairment* 24, 2 (2006), 93–100. DOI: <http://dx.doi.org/10.1177/0264619606064436>
- [8] EN Mitsopoulos and ADN Edwards. 1999. A principled design methodology for auditory interaction. *Human-Computer Interaction - INTERACT '99* (1999).
- [9] Matthijs L. Noordzij, Sander Zuidhoek, and Albert Postma. 2006. The influence of visual experience on the ability to form spatial mental models based on route and survey descriptions. *Cognition* 100, 2 (2006), 321–342. DOI: <http://dx.doi.org/10.1016/j.cognition.2005.05.006>
- [10] Stavroula Ntoa, Iliia Adami, Giannis Prokopiou, Margherita Antona, and Constantine Stephanidis. 2011. A-cross: An accessible crossword puzzle for visually impaired users. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 6766 LNCS, PART 2 (2011), 342–351. DOI: http://dx.doi.org/10.1007/978-3-642-21663-3_37
- [11] Sile O'Modhrain, Nicholas A. Giudice, John A. Gardner, and Gordon E. Legge. 2015. Designing media for visually-impaired users of refreshable touch displays: Possibilities and pitfalls. *IEEE Transactions on Haptics* 8, 3 (2015), 248–257. DOI: <http://dx.doi.org/10.1109/TOH.2015.2466231>
- [12] Denise Prescher, Oliver Nadig, and Gerhard Weber. 2010a. Reading Braille and tactile ink-print on a planar tactile display. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 6180 LNCS, PART 2 (2010), 482–489. DOI: http://dx.doi.org/10.1007/978-3-642-14100-3_72
- [13] Denise Prescher, Gerhard Weber, and Martin Spindler. 2010b. A tactile windowing system for blind users. *Proceedings of the 12th international ACM SIGACCESS conference on Computers and accessibility - ASSETS '10* (2010), 91. DOI: <http://dx.doi.org/10.1145/1878803.1878821>
- [14] Rameshsharma Ramloll and Stephen Brewster. 2002. A generic approach for augmenting tactile diagrams with spatial non-speech sounds. *Conference on Human Factors in Computing Systems - Proceedings* (2002), 770–771.
- [15] H. Rao and S. O'Modhrain. 2019. Multimodal representation of complex spatial data. In *Conference on Human Factors in Computing Systems - Proceedings*. DOI: <http://dx.doi.org/10.1145/3290607.3313249>
- [16] Johnny. Saldaña. 2009. *The coding manual for qualitative researchers*. Number London ; Thousand Oaks, Calif. :. Sage, London ; Thousand Oaks, Calif. :.
- [17] Maria Schiewe, Wiebke Köhlmann, Oliver Nadig, and Gerhard Weber. 2009. What you feel is what you get: Mapping GUIs on planar tactile displays. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 5615 LNCS, PART 2 (2009), 564–573. DOI: http://dx.doi.org/10.1007/978-3-642-02710-9_63
- [18] Michael Schmidt and Gerhard Weber. 2009. Multitouch Haptic Interaction. *Universal Access in Human-Computer Interaction. Intelligent and Ubiquitous Interaction Environments* 5615 (2009), 574–582.
- [19] Holly A. Taylor, Susan J. Naylor, and Nicholas A. Chechile. 1999. Goal-specific influences on the representation of spatial perspective. *Memory and Cognition* 27, 2 (1999), 309–319. DOI: <http://dx.doi.org/10.3758/BF03211414>
- [20] Maarten W A Wijntjes and Astrid M L Kappers. 2007. Angle discrimination in raised-line drawings. *Perception* 36, 6 (jan 2007), 865–879. DOI: <http://dx.doi.org/10.1068/p5603>

- [21] Limin Zeng, Mei Miao, and Gerhard Weber. 2012. Interactive Audio-haptic Map Explorer on a Tactile Display. *Interacting with Computers* 27, 4 (2012), 413–429. DOI:<http://dx.doi.org/10.1093/iwc/iwu006>
- [22] Limin Zeng and Gerhard Weber. 2011. Accessible maps for the visually Impaired. *IFIP INTERACT Workshop on Accessible Design in the Digital World* (2011), 1–12. <http://ceur-ws.org/Vol-792/Zeng.pdf>
- [23] Limin Zeng and Gerhard Weber. 2012. Building augmented you-are-here maps through collaborative annotations for the visually impaired. *CEUR Workshop Proceedings* 888 (2012), 7–12.