Breadth vs. Depth in Menu Structures for Screen Reader Users

Carolyn Bussey
DePaul University
243 S Wabash
Chicago, IL 60614 USA
buss0139@gmail.com

Yu (Jackson) Cui
DePaul University
243 S Wabash
Chicago, IL 60614 USA
cuiyu126@hotmail.com

Akash Singh
DePaul University
243 S Wabash
Chicago, IL 60614 USA
singh.akash5553@gmail.com

Somer Weich
DePaul University
243 S Wabash
Chicago, IL 60614 USA
somerweich@outlook.com

November 20, 2018

Abstract
Despite decades of study, there are still mixed recommendations on what hierarchical menu structures work best for users with screen readers. The organization and structure of these menus can make the experience of using a website enjoyable and efficient or frustrating and confounding. We replicated a study from 2010 examining the breadth vs depth argument in menu structures. Consistent with some previous studies, we found that broad, shallow hierarchical menu structures outperformed deep, narrow hierarchical menu structures for screen reader users.

Author Keywords
Accessibility; Menu structure; Screen readers; Blind; Usability; Navigation

ACM Classification Keywords
K.4.2 Computers and Society: Social Issues. Handicapped persons/special.

Introduction
The usability of large menus has been studied with sighted users over the past 30 years. In the mid-1980s researchers began exploring whether a menu with many options and few levels ("broad-shallow") (Figure 1) would be easier to use than a menu with few options

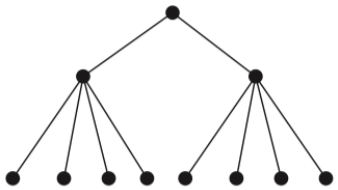

Figure 1: A broad-shallow hierarchy with many options and few levels.

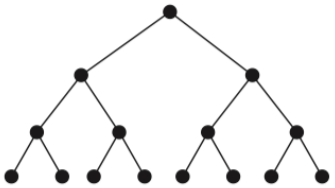

Figure 2: A narrow-deep menu with few options and many levels.

and many levels ("narrow-deep") (Figure 2). Researchers have evaluated an array of different types of broad and narrow menus, from 2^6 (six levels with two options at each level) to 64^1 (one level with sixty-four options). In most studies, broad-shallow structures tested better than narrow-deep ones [2,4,5,7,10]. An important consideration of much of the early work was that the menus typically contained less than 100 bottom level items. Typical menus found in content-heavy websites and applications, such as e-commerce, contain more than 100 bottom level items.

In 1998, Czerwinski and Larson [6] selected three menu structures to evaluate what they argued were representative of common menus; the menus they evaluated had 512 bottom level items. Their dependent variables included task time and a measure of hypertext lostness to evaluate navigational efficiency. They also collected data on their participants (n= 19) subjective preference. Their results were consistent with earlier research. Broad-shallow menus out-performed narrow-deep menus which were typically perceived as more complicated. Participants used a non-linear search pattern when scanning through options in a menu and the contents of the multiple branches of the narrow-deep menus were sometimes difficult to recall. The researchers concluded that selecting and remembering multiple categories was more cognitively taxing than scanning large amounts of information for a target item [6]. In summation, multiple researcher projects with sighted participants have come to similar conclusions; however, little research has considered users with screen readers.

Barnicle [1] studied the usability of websites for users with screen readers in 2000. The author examined common GUI elements, including menus, and recommended that narrow navigation structures were better for screen readers. Screen readers present links in a linear list, and this author hypothesized that shorter lists at each level would help "funnel" users into the right categories more quickly [1]. However, this recommendation conflicts with later findings from Hochheiser and Lazar [3] in 2010.

Hochheiser and Lazar [3] duplicated Czerwinski and Larson's study using three menu hierarchies (8x8x8, 16x32 and 32x16) with screen reader users (n=19). They found that users who were blind performed better with moderately broad-shallow menu hierarchies like their sighted counterparts. We argue that these conflicting recommendations suggest a need for further inquiry into web menu hierarchies for people who use screen readers.

In addition to the conflicting recommendations, a recent and popular web design trend is the mega-menu. Mega-menus are an extreme form of a broad-shallow hierarchy. These typically include visual aspects that group navigation options through layout, typography, and (sometimes) imagery or icons. According to the Nielsen Norman group, mega-menus are helpful because they allow users to see more options; more visual options decreases cognitive load by making all the various "levels" of menu hierarchy always visible [8]. Obviously, the visual organization and imagery of a typical mega-menu is not readily accessible to users who are blind. In this project, we aimed to examine how the design of menu structures impacts user-experience for people who use screen readers. We compared three designs (1) extreme broad-shallow menu (a mega-menu), (2) a moderate

broad-shallow menu, and (3) a narrow-deep menu by revisiting the previous comparison work from Czerwinski and Larson [6] and Hochheiser and Lazar [3].

Methods

In the following sections, we describe our participants, data collection and data analysis methods.

Participants

We recruited eleven participants with the help of the National Federation of the Blind (NFB) and Chicago Lighthouse. All participants were blind and used a screen reader to access websites. Two participants lived in the Chicago metropolitan area, six participants lived in Minneapolis, Minnesota, and three participants lived in Boise, Idaho (Table 1).

	Age	Gender	Level of Education	Screen Reader Years
P1	40-50	F	Master's	20+
P2	60-70	M	Master's	20+
P3	50-60	M	High School	20+
P4	70-80	F	Bachelor's	30+
P5	60-70	F	High School	< 1
P6	20-30	F	Bachelor's	10+
P7	50-60	F	Master's	5+
P8	50-60	F	Master's	5+
P9	60-70	F	Master's	20+
P10	40-50	F	Master's	20+
P11	50-60	F	Associate's	5+

Table 1: Participant demographics.

Data Collection

We used data and hierarchical menus from the original experiment by Czerwinski and Larson [6] with some modifications.

The original study used three hierarchical menus with 512 options: (1) three levels with eight choices at each level (8×8×8); (2) two levels with 16 choices followed by 32 choices (16×32); (3) two levels with 32 choices followed by 16 choices (32×16).

The first modification was the removal of the 32x16 hierarchy. This hierarchy is similar to the 16x32 hierarchy; both hierarchies are two levels, both have a broad-shallow structure, and the 16x32 hierarchy performed better than the 32x16 in both studies from Czerwinski and Larson [7] and Hochheiser and Lazar [4]. The second modification was the addition of an extremely broad-shallow hierarchy, a mega-menu, two levels with 8 choices followed by 64 choices (8x64).

We used Treejack, a menu structure validation tool. We first confirmed that the tool was accessible to participants using screen readers prior to running the sessions. Participants started the study with a warm-up task using a limited menu structure (2 choices followed by 8 choices, 2x8) to gain familiarity with navigating and using the tool.

We conducted the in-person sessions in differing locations, to suit the participants' comfort. After the warm-up task, participants completed three randomly selected tasks for each structure, 10 tasks in total. The tasks all involved asking the participant to find a specific option (e.g. Bluegrass Music) within a given menu. To prevent participants from learning the

hierarchies, we had them switch after each task was complete. For example, the first task was completed using the 8x8x8 hierarchy, the second task was completed using the 16x32 hierarchy, and so on. To compensate for ordering effects, we used different orders for the hierarchies; for example, one researcher started with the 8x8x8 hierarchy for the first task while another researcher started with the 8x64 hierarchy.

Data Analysis

We collected three dependent variables: (1) correctness; (2) task completion times; and (3) lostness. These three metrics were chosen as a way to measure the accuracy and efficiency of each hierarchy.

To determine correctness differences, we tracked whether the tasks were completed successfully, completed unsuccessfully, timed out after 4 minutes, or not completed at all. We used SPSS to run a Chi-squared test to determine if there were any significant differences in the success rate of tasks between these three hierarchies.

To determine task completion times, we tracked how long it took the participants to complete the tasks. We used SPSS to run an ANOVA test to determine if there were any significant differences in the task completion times between these three hierarchical menus.

To determine differences in lostness we tracked the menu options the participants clicked on while completing the task. Lostness is a calculation developed by Smith [9] that compares the number of clicks by the user to the number of clicks required for a task to be completed successfully, and returns a score from 0 to 1. A user that takes an optimal path will score 0. A score above .50 is considered lost. For example, if a task can be completed successfully with only three clicks, but a user clicks 10 times trying to locate their target item, then we would consider them to be lost within the hierarchical menu. We used SPSS to run a Friedman test to determine if there were any significant differences in the level of lostness between these three hierarchical menus.

To determine if there was a significant relationship between task completion times and lostness we used SPSS to run a correlation test.

Findings

We found that screen reader users who are blind performed better with a moderately broad-shallow menu hierarchical menu, mimicking the results of the previous study. We also found that while these users perform well with an extremely broad-shallow menu (8x64), like a mega-menu, they still perform better with a moderately broad-shallow menu (16x32). Findings from our study include: (1) correctness; (2) task completion times; (3) lostness; and (4) observations.

Correctness

We had 11 participants attempt three tasks each for all three menu hierarchical menus, for a total of 99 possible tasks. There were 74 correct responses, nine incorrect responses, 10 timeouts, and six tasks where no activity was recorded (Figure 3, Table 2). There were no significant differences between the hierarchies in the number of correctly completed tasks ($\chi^2(6) = 3.83$, $p = 0.73$).

	8x8x8	16x32	8x64
No activity	4	1	1
Timed Out	3	3	3
Incorrect	2	4	4
Correct	24	25	25

Table 2: This table shows the values for Figure 3.

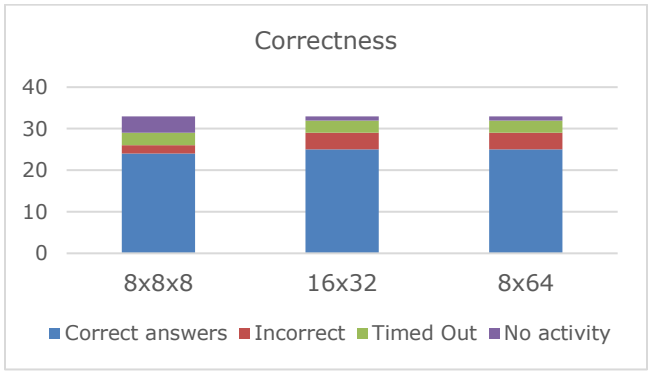

Figure 3: In this graph, the correctness of the three hierarchical menus are compared.

To maintain the assumption of normality, subsequent analyses exclude the six tasks with no activity, one set of tasks from one participant who completed all of the tasks twice with two different devices, and the nine timed out tasks that were outliers as a result of users being confused during the tasks. Leaving 78 tasks for analysis.

Task Completion Time

Task completion times were fastest for the 16x32 hierarchy (M = 61.11, SD = 13.70), followed by the 8x64 hierarchy (M = 61.64, SD = 18.14), and 8x8x8 hierarchy (M = 108.10, SD = 62.08) (Figure 4). A Kolmogorov-Smirnov test indicates that the task completion times for all three hierarchies do follow a normal distribution: (1) 8x8x8 hierarchy, D(26) = 0.11, p=0.88; (2) 16x32 hierarchy, D(26) = 0.19, p=0.30; (3) 8x64 hierarchy, D(26) = 0.26, p=0.05. A one-way repeated-measures ANOVA revealed significant main effects of hierarchy (F(2, 50) = 7.20, p=0.002), with post hoc Bonferroni indicating that the 8x8x8 hierarchy was significantly slower than the 16x32 hierarchy (p=0.006) and the 8x64 hierarchy (p=0.023). To be safe, a non-parametric Friedman test of differences among repeated measures confirmed significant differences in the three hierarchies (χ^2(2) = 8.31, p = 0.016).

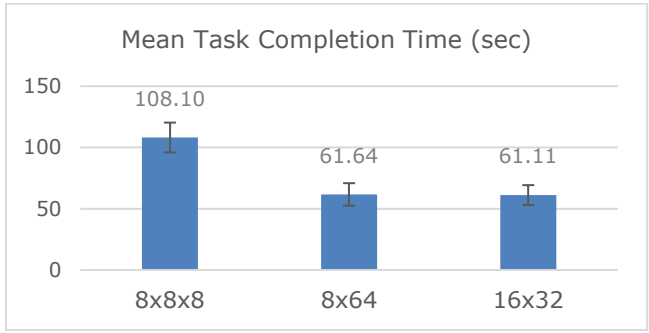

Figure 4: In this graph, the mean task completion times of the three hierarchical menus are compared.

Lostness

Participants were lost four times in the 8x8x8 hierarchy (*lostness*=0.52, 0.52, 0.52, 0.80), one time in the 16x32 hierarchy (*lostness*=0.68), and one time in the 8x64 hierarchy (*lostness*=0.96). Participants were most lost with 8x8x8 (M = 0.24, SD = 0.18), followed by the 8x64 (M = 0.06, SD = 0.14) and 16x32 (M = 0.06, SD = 0.14) hierarchies (Figure 5). A Kolmogorov-Smirnov test indicates that the lostness scores for: (1) the 8x8x8 hierarchy do follow a normal distribution, D(26) = 0.23, p=0.10; (2) the 16x32 hierarchy do not follow a normal distribution, D(26) = 0.51, p=0.00001; and (3) the 8x64 hierarchy do not follow a normal distribution, D(26) = 0.50, p=0.00003. A non-parametric Friedman test of differences among repeated measures was conducted and rendered a Chi-square value of 20.96 which was significant (p<.01).

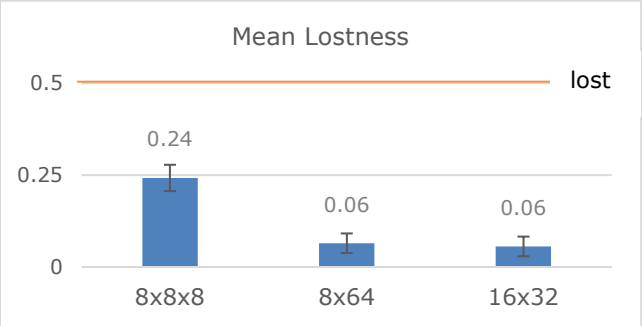

Figure 5: In this graph, the lostness scores of the three hierarchical menus are compared, lostness line shown.

We found a strong correlation between task completion time and lostness scores ($r(78) = 0.59$, $p < 0.01$).

Observations

We observed three distinct strategies from our participants while navigating through the menus; "careful listening", "auditory scanning", and "jumping to known links".

Most participants began the study by carefully listening to each link fully. Once participants were more familiar with Treejack and the content, they auditorily scanned the list by moving through the options before each one was fully read by the screen reader. This sometimes caused errors. Jumping to a known link was another common behavior we observed. Not all participants used a shortcut to jump to a known link, but those that began using these shortcuts continued to use them instead of listening to anything on the page. This occasionally confused participants; when used in an inappropriate submenu, they did not get any results or feedback.

Discussion

In this study we compared a new visual design trend for website menus, the mega-menu, to other commonly used structures to determine how it would impact people who use screen readers. It can be challenging for users who are blind to traverse menu structures that are designed for visual searching.

The task completion time results are consistent with those of the earlier studies by Czerwinski and Larson [6] and Hochheiser and Lazar [3]. The 16x32 hierarchy had the best performance, followed by the 8x64 hierarchy, and the 8x8x8 hierarchy. Lostness results were similar, with the 16x32 structure showing the lowest lostness values, followed by 8x64 and 8x8x8.

We found that the design of the menu structure does affect people who use screen readers, and the results are similar to previous research for sighted users.

Limitations and Future Work

Our sample size was small and did not represent a variety of screen reader users. To better inform our research, we would need to increase the sample size and add more diversity to the sample population.

We could not control the system setup for all participants. To better inform our research, we would need to control the system setup; including either the exact same setup for all participants or allow all participants to use their own device and settings.

We could not control the usability of the TreeJack system. To better inform our research, we would need to track metrics using a more stable, usable, and error free system.

References

1. Kitch Barnicle. 2000. Usability testing with screen reading technology in a windows environment. In *Proceedings of the 2000 Conference on Universal Usability* (CUU '00), 102-109. http://doi.acm.org/10.1145/355460.355543

2. Stewart Card. 1984. Visual search of computer command menus. *Attention and performance X: Control of language processes*, 97-108.

3. Harry Hochheiser and Jonathan Lazar. 2010. Revisiting breadth vs. depth in menu hierarchical menus for blind users of screen readers. *Interacting with Computers* 22: 389-398.

4. Julie A. Jacko, and Gavriel Salvendy. 1996. Hierarchical menu design: breadth, depth, and task complexity. *Perceptual and Motor Skills* 82: 1187–1201.

5. John Kiger. 1984. The depth/breadth tradeoff in the design of menu-driven interfaces. In *International Journal of Human-Machine Studies*. 201-213.

6. Keith Larson and Mary Czerwinski. 1998. Web page design: implications of memory, hierarchical menu and scent for information retrieval. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (CHI '98), 25-32. http://dx.doi.org/10.1145/274644.274649

7. Dwight P. Miller. 1981. The depth/breadth tradeoff in hierarchical computer menus. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 25 (1): 296–300. https://doi.org/10.1177/107118138102500179

8. Jakob Nielsen, Angie Li. 2014. Mega menus work well for site navigation. Retrieved October 15, 2018 from https://www.nngroup.com/articles/mega-menus-work-well/

9. Pauline A. Smith. 1996. Towards a practical measure of hypertext usability*. Interacting with Computers* 8 (4): 365–381.

10. Kathleen Snowberry, Stanley R. Parkinson, Norwood Sisson. 1983. Computer display menus. *Ergonomics* 26 (7): 699-712.