Effect of Layout on User Performance and Subjective Evaluation in an Augmented-Reality Environment

Xin Lei, Yueh-Lin Tsai, and Pei-Luen Patrick Rau(✉)

Department of Industrial Engineering, Tsinghua University, Beijing, China
rpl@mail.tsinghua.edu.cn

Abstract. In this study, the effect of layout on user performance and subjective evaluation in an augmented-reality (AR) environment was investigated. A scenario where participants had to work on three windows simultaneously was used. Three basic layouts of these windows have been examined, i.e., a horizontal layout, a vertical layout, and a diagonal layout. Additionally, two experimental tasks had to be completed; one was a reading comprehension task requiring a low switching frequency (LSF), and the other was a classification task requiring a high switching frequency (HSF). The results revealed that first, participants performed best in the diagonal layout in the LSF task, whereas they performed best in the vertical layout in the HSF task. Second, no significant differences were found in the disorientation between different layouts. Third, participants were significantly less satisfied with the diagonal layout in the HSF task. In conclusion, a horizontal layout is first recommended for general tasks and a vertical layout is recommended for HSF tasks. The switching distance and switching path are two important factors to be considered in the layout design in an AR environment.

Keywords: Multi-window · Layout · Augmented reality · Switching

1 Introduction

People often work with several computer applications simultaneously; however, the size of the computer screen limits the display of multiple windows. Most windowing systems follow the independent overlapping windows approach; thus, the oldest window is always overlapped by the new active one. Owing to the current manner of interaction, frequent switching can cause users to easily lose location awareness and operation awareness, thus leading to a limited sense of perceived control [1]. Finally, the performance and work satisfaction of users can be negatively affected.

The advent of augmented reality (AR) has broken the limitations of current information display modes. For example, the AR head-mounted display (HMD), such as HoloLens, frees users from the computer screen and makes all environment space into users' "desktops". Users can divide the environment space into several regions; then, they may combine relevant windows and separate irrelevant windows. This enables users to configure a certain group of windows into a specific spatial region.

P.-L. P. Rau (Ed.): HCII 2019, LNCS 11576, pp. 376–385, 2019.
https://doi.org/10.1007/978-3-030-22577-3_27

Moreover, this changes the manner of interaction when users switch between multiple windows. Therefore, users may easily and conveniently access a window just by turning their heads.

The multiple-window layout displayed in one environment space has an important influence on the information processing of users. Although there are few studies that are focused on augmented-reality-based layouts, layouts have been investigated and they have been proved to be significant in traditional visual information presentation. For example, the Web page layout has been considered to be a major influencing factor on performance, orientation, and subjective satisfaction [2–5]. Thus, we suspect that the layout of multiple windows is highly important in the augmented-reality environment as well. The aim of this study is to examine the influence of several basic layouts on the user performance and the subjective evaluation, and then to identify the appropriate layouts for different tasks.

2 Related Work

Layouts have been proved to be significant in visual information presentation in numerous fields. First, layouts are important in website designs; placing web objects at expected locations and designing their appearance according to user expectations facilitates orientation; therefore, users can perform faster searches and remember more easily [6]. Second, layouts have influence on graph readability. Three layouts (i.e., force-directed, hierarchical, and orthogonal layout) have different levels of readability in different tasks; the force-directed layout outperformed the other layouts in certain tasks; however, all three layouts performed equally well in certain other tasks [7]. Moreover, in several research works, it has been shown that the graph layout affects the readability as well as the understanding of the underlying data [8]. Third, layouts are important in tag clouds as well, which have become a popular visualization and navigation interface on the Web. The layout of a tag cloud influences its perception; tags in the upper left quadrant are better recalled and can be noticed more quickly, whereas tags in the middle of the cloud attract more user attention than tags near the borders [9].

2.1 Performance

The overall screen layout is considered to have a major impact on task performance [5]. Horizontal menus (left and right) cause a significantly quicker reaction time than vertical menus (top and bottom) for both hits and correct rejections for the visual search task [10]. Moreover, a different study shows the same effect of layout both on accuracy and speed measures, with frames located at the top or left of the screen leading to better performance. Furthermore, layouts and tasks have an interaction effect on performance. For example, the force-directed layout outperformed other layouts in certain tasks; however, all layouts performed equally well in certain other tasks [7].

2.2 Disorientation

Disorientation has been defined as the tendency to lose one's sense of location and direction in a nonlinear document [11, 12]. It can cause users to become frustrated, lose interest, and experience a measurable decline in efficiency [13]; however, it can be minimized via the improvement of navigation design [14]. Therefore, disorientation can serve as an additional tool to evaluate information technology. However, it is not easy to measure disorientation; two fundamental approaches exist regarding disorientation: one that claims a link between the actions of the users and disorientation, and another that claims that user disorientation can only be measured by asking users about their perceptions [11, 15]. Furthermore, it has been investigated that perceived disorientation is predictive of task performance in an interactive search task; however, the actions of the users are not [11]. Moreover, there are differences in the perceived disorientation of the same system that are linked to the sex of the user because certain sex differences have been identified in spatial abilities, including spatial navigation, object location, and spatial rotation [16].

2.3 Perceived Satisfaction

User satisfaction has been recognized as the most dominant criterion of website success. Muylle et al. [17] empirically validated a standard instrument for measuring website user satisfaction that consisted of three components, i.e., information, connection, and layout. Similarly, a different website quality assessment considers the usefulness and the layout as the two most important criteria [18]. Furthermore, it has been found that layouts have great effect on the perceived satisfaction of users [17, 19–21].

3 Design and Evaluation of Different Layouts

3.1 Design of Multi-window Layouts

In this study, a scenario is considered in which users work using three windows simultaneously. Users have to collect and process information from these windows, and then respond accordingly. Multitasking in user behavior can be represented along a continuum in terms of the time spent on one task before switching to another [22]. Thus, three types of windows are defined in this study owing to the total time allocated on a window while multitasking. Window A represents the core task, which consumes most of the user time, window B represents the secondary task, and window C represents the auxiliary task, which consumes the least time. According to previous studies, people tend to concentrate more on the center and upper-left areas; thus, important windows (namely A and B) should be located in the two most appropriate areas. Therefore, we propose three common layouts for the three windows, i.e., the horizontal layout (H), the vertical layout (V), and the diagonal layout (D) (Fig. 1).

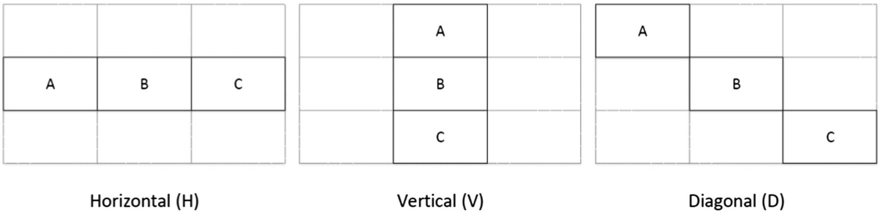

Fig. 1. Three layouts examined in this study

3.2 Evaluation Experiment

Experimental Scene. We used Unity 3D to design our experimental scenes and conducted the experiments on the AR HMD HoloLens. Each layout consisted of three windows, which was set 100 cm in front of users. Three windows were in the equal sizes of 50 $\times$ 30 cm, and the font size of the information shown in each window was 10 pt.

Experimental Task. There were two types of tasks, namely one with a low switching frequency (LSF) and the other with a high switching frequency (HSF).

The LSF task was reading comprehension. Participants had to read an article and complete three multiple-choice questions in 5 min. The articles and questions used in this study were all extracted from the College English Test-6 in China, and they are in the same difficulty level. The LSF task involved three windows. A window presented the English article. A window presented three multiple choice questions. The last window presented the Chinese meaning of the English words in the article that users might not understand, but users may not frequently look at this window. According to the time length that users spent on each window, the article window was type-A window, the question window was type-B window, and the dictionary window was type-C window. In the LSF task, participants had five minutes to finish the questions. When the time was up, the correct answer would appear automatically. At this time, participants could not continue answering the questions and we counted the number of their correct answers.

HSF task was data classification. Participants had to classify twelve events into four categories according to their degree of importance and degree of urgency. A window was the description window that presented the degree of importance and urgency of all events, and users had to get necessary information from this window. A window was the working window where participants labeled events in each category. The last window presented the classification rule, but participants may not frequently look at this window because they should be able to remember the classification rule. They might only occasionally refer to this window to check the rule. According to the time length that users spent on each window, the description window was type-A window, the working window was the type-B window, and the rule window was the type-C window. In the HSF task, the timer would automatically stop when the participant would correctly complete the task and the completion time was recorded. However, if the participant would not succeed in completing the task in one trial, the data were not used for analysis because this completion time would be significantly longer than that of the one-time success.

Experimental Procedure. The experiment consisted of three phrases. In the first phase, the experimenters introduced the experimental tasks and rules to the participant. Then, the participant had to practice the data classification task on paper, no less than three times in order to become familiar with the task. In addition, the participant had to practice the "select" operation on the HoloLens because it would be heavily used in the formal experiment. The practice prior to the testing was aimed toward avoiding the effect of inexperience on the completion time in the formal experiment. In the second phase, participants had to complete LSF tasks and HSF tasks on the HoloLens, and they had to fill in a short questionnaire every time they completed a task. In the third phase, participants were interviewed about their preferences and they had to comment on the different layouts.

Measurements. Three dependent variables were measured in this study, i.e., the performance, the disorientation, and the perceived satisfaction. The performance consisted of accuracy and efficiency; the accuracy was measured in the LSF task, which was the number of correct answers, whereas the efficiency was measured in the HSF task, which was the completion time. The disorientation and the satisfaction were measured through a five-point Likert scale. A larger value of disorientation indicated that it was easier for participants to become confused in this layout. A higher value of satisfaction indicated that participants were more satisfied with this layout.

Participants. Twenty-four participants from Tsinghua University took part in the experiment. Their average age was 23.4 (SD = 1.39). All participants signed an informed consent agreement prior to testing.

Data Analysis. All dependent variables did not obey the normal distribution; thus, non-parametric analysis was used. The Friedman test was used to examine the main effect of the layout on the performance, the switching times, the disorientation, and the satisfaction. Then, the Wilcoxon signed rank test was used to perform the pairwise comparison between different layouts.

4 Results

4.1 Performance

Table 1 lists the descriptive statistics of the performance under different layouts as well as the results of the Friedman test and of the pairwise comparison. Regarding the LSF task, the performance is the number of correct answers; therefore, a higher value corresponds to a better performance. Regarding the HSF task, the performance is the completion time; therefore, a higher value corresponds to a worse performance.

In the LSF task, the statistical difference in accuracy was marginally significant, namely, $\chi_2^2 = 4.78$and $p = .091$. As the post-hoc analysis showed, the accuracy of the participants was significantly higher in the diagonal layout than that in the horizontal layout, and the effect size was considerable, namely, $V = 64.5, p = .042,$ and $r = .41$. Although the difference between the horizontal layout and the vertical layout was not statistically significant, the effect size was medium, namely, $V = 66.5, p = .124,$ and $r = .31$. Hence, the diagonal layout was the best and the vertical layout was the worst.

In the HSF task, the layout had a significant effect on the completion time, namely, $\chi_2^2 = 20.08 \, and \, p < .001$. The vertical layout led to the shortest completion time; the difference between the vertical layout and the remaining two layouts was large, both yielded $r > .55$. The diagonal layout resulted in the longest completion time and the difference between the diagonal layout and the horizontal layout was not significant; however, the effect size was medium, namely, $V = 205, p = .121,$ and $r = .32$. Therefore, the diagonal layout was the worst; however, the vertical layout was the best. This result contradicted the results of the LSF task.

Table 1. The main effect of layout on user performance

	LSF task: reading comprehension				HSF task: classification			
Layout	Mean	SD	χ_2^2	p-value	Mean	SD	χ_2^2	p-value
Horizontal	1.75	0.85	4.78	.091	92.13	33.47	20.08	<.001
Vertical	1.46	0.88			84.12	25.51		
Diagonal	1.92	0.72			96.27	25.98		
Post-hoc Analysis								
	V	p-value		r	V	p-value		r
H-V	66.5	.124		.31	242	.007		.55
D-H	50.5	.356		.19	205	.121		.32
D-V	64.5	.042		.41	300	<.001		.87

4.2 Disorientation

Table 2 lists the descriptive statistics of disorientation under different layouts as well as the results of the Friedman test and those of the pairwise comparison. A higher value means higher disorientation; therefore, a lower value is preferred instead of a higher one. However, no statistically significant differences were found in the disorientation for different layouts in both LSF and HSF tasks. In the HSF task, the perceived disorientation was greater in the vertical layout than in the horizontal layout and diagonal layout; although the p-values were higher than .05 and the effect sizes were medium, namely, they were both rs = .30.

Table 2. The main effect of layout on disorientation

	LSF task: reading comprehension				HSF task: classification			
Layout	Mean	SD	χ^2_2	p-value	Mean	SD	χ^2_2	p-value
Horizontal	2.75	0.99	0.24	.888	2.08	0.65	2.63	.268
Vertical	2.79	0.98			2.50	0.98		
Diagonal	2.83	0.76			2.46	1.06		
Post-hoc analysis								
H-V	76	1		0	40	.144		.30
D-H	32.5	.738		.10	42	.145		.30
D-V	56	.847		.04	73.5	.903		.02

4.3 Perceived Satisfaction

Table 3 lists the descriptive statistics of perceived satisfaction under different layouts as well as the results of the Friedman test and those of the pairwise comparison. A higher value indicates higher satisfaction; therefore, a higher value is preferred instead of a lower one. No statistically significant differences were found in satisfaction of different layouts in both LSF and HSF tasks. However, in the HSF task, participants were more satisfied with the horizontal and vertical layouts than they were with the diagonal layout, both $ps < .077$ *and* $rs > .36$. The differences were marginally significant and the effect sizes were large.

Table 3. The main effect of layout on user satisfaction

	LSF task: reading comprehension				HSF task: classification			
Layout	Mean	SD	χ^2_2	p-value	Mean	SD	χ^2_2	p-value
Horizontal	3.13	0.95	0.70	.703	3.25	1.22	3.09	.213
Vertical	3.29	1.04			3.25	1.11		
Diagonal	3.42	1.02			2.63	0.97		
Post-hoc analysis								
	V	p-value		r	V	p-value		r
H-V	62	.489		.14	123	.803		.05
D-H	87	.324		.20	72	.071		.37
D-V	87.5	.609		.10	58.5	.077		.36

5 Discussion

The experimental results showed that the layout of multiple windows has a significant effect on the user performance and the subjective evaluation in the AR environment. According to the short post-experiment interviews with participants, two major factors were identified as the ones that influenced user performance and evaluations toward different layouts: the switching distance and the switching path. The two factors can explain the experimental results at a certain extent.

The most important factor is the switching distance, namely, the distance that the heads of the participants moved when participants switched from one window to another. The diagonal layout presented the longest switching distance, followed by the horizontal layout, and then the vertical layout. This can explain why participants were significantly less satisfied with the diagonal layout in the HSF task but not in the LSF task. In the HSF task, users had to frequently switch between windows, thus the influence of the switching distance was more evident. In the LSF task, however, users did not need to frequently switch; therefore, the complaints of the users about the long switching distance was not particularly evident. This could also explain why users performed best in the vertical layout in the HSF task. The classification task was focused on the completion time of the users. The shortest switching distance was observed in the vertical layout, namely, the shortest switching time, where the shortest completion time was observed as well.

The second factor was the switching path, namely, the direction toward people had to turn their heads during switching. In the horizontal layout, users turned their heads left and right; in the vertical layout, users turned their heads up and down; in the diagonal layout, users turned their heads in two sets of directions, namely left–right and up–down. The left–right movement was the most natural for human users, followed by the vertical layout, which was followed by the diagonal layout. This can also explain why participants were significantly less satisfied with the diagonal layout in the HSF task but not in the LSF task. The influence of the head-moving direction was more evident when users had to frequently turn their heads.

6 Conclusion

In this study, the effect of the layout on the user performance and the subjective evaluation in the AR environment was examined. In the experiment, participants had to work on three windows simultaneously. Two experimental tasks were required to be completed; one was a reading comprehension task, which required a low switching frequency, and the other was a classification task, which required a high switching frequency. According to the experimental results, the average number of switching times in the LSF tasks was approximately 4 per minute, whereas the average number in the HSF tasks was approximately 16 per minute. In the LSF tasks, participants performed better in the diagonal layout and the horizontal layout than in the vertical layout. In addition, participants switched less frequently in the diagonal layout than in the horizontal layout, and they switched most frequently in the vertical layout. In the HSF tasks, participants performed significantly better in the vertical layout than in the horizontal layout. Moreover, participants performed worst in the diagonal layout. Additionally, participants were less satisfied with the diagonal layout than with the horizontal and vertical layouts. Two important factors that influenced user performance and evaluations toward different layouts were the switching distance and the switching path.

In conclusion, certain AR-based layout design suggestions should be proposed. First, the switching distance and the switching path should be strongly considered in AR-based multi-screen layout designs. Short switching distances and natural switching paths were preferred. Second, the horizontal layout would be recommended first. Its switching distance was moderate, and the switching path was left and right, which is natural and in line with the daily habits of people. User performance and perceived satisfaction were both acceptable for the horizontal layout. Third, the vertical layout is recommended for tasks requiring frequent switching and emphasizing efficiency. The switching distance of this layout is short; however, its switching path is up and down, which is not as comfortable as the horizontal layout. Finally, the designers should avoid locating a window that would require user operation in the left region.

References

1. Zhang, Y., Mao, M., Rau, P.-L.P., Choe, P., Bela, L., Wang, F.: Exploring factors influencing multitasking interaction with multiple smart devices. Comput. Hum. Behav. **29**, 2579–2588 (2013)
2. Brinck, T., Gergle, D., Wood, S.D.: Usability for the Web: Designing Web Sites that Work. Elsevier, Amsterdam (2001)
3. Deaton, M.: The Elements of User Experience: User-Centered Design for the Web. Interactions **10**, 49–51 (2003)
4. Krug, S.: Don't Make Me Think! A Common Sense Approach to Web Usability. Pearson Education India (2000)
5. Parush, A., Shwarts, Y., Shtub, A., Chandra, M.J.: The impact of visual layout factors on performance in Web pages: a cross-language study. Hum. Factors **47**, 141–157 (2005)
6. Roth, S.P., Tuch, A.N., Mekler, E.D., Bargas-Avila, J.A., Opwis, K.: Location matters, especially for non-salient features–an eye-tracking study on the effects of web object placement on different types of websites. Int. J. Hum.-Comput. Stud. **71**, 228–235 (2013)
7. Pohl, M., Schmitt, M., Diehl, S.: Comparing the readability of graph layouts using eyetracking and task-oriented analysis. In: Computational Aesthetics, pp. 49–56 (2009)
8. Huang, W.: Using eye tracking to investigate graph layout effects. In: 2007 6th International Asia-Pacific Symposium on Visualization, APVIS 2007, pp. 97–100. IEEE (2007)
9. Lohmann, S., Ziegler, J., Tetzlaff, L.: Comparison of tag cloud layouts: task-related performance and visual exploration. In: Gross, T., Gulliksen, J., Kotzé, P., Oestreicher, L., Palanque, P., Prates, R.O., Winckler, M. (eds.) INTERACT 2009. LNCS, vol. 5726, pp. 392–404. Springer, Heidelberg (2009). https://doi.org/10.1007/978-3-642-03655-2_43
10. Pearson, R., van Schaik, P.: The effect of spatial layout of and link colour in web pages on performance in a visual search task and an interactive search task. Int. J. Hum.-Comput. Stud. **59**, 327–353 (2003)
11. Ahuja, J.S., Webster, J.: Perceived disorientation: an examination of a new measure to assess web design effectiveness. Interact. Comput. **14**, 15–29 (2001)
12. Head, M., Archer, N., Yuan, Y.: World wide web navigation aid. Int. J. Hum.-Comput. Stud. **53**, 301–330 (2000)
13. McDonald, S., Stevenson, R.J.: Effects of text structure and prior knowledge of the learner on navigation in hypertext. Hum. Factors **40**, 18–27 (1998)
14. Webster, J., Ahuja, J.S.: Enhancing the design of web navigation systems: the influence of user disorientation on engagement and performance. MIS Q. **30**, 661–678 (2006)

15. Santa-Maria, L., Dyson, M.C.: The effect of violating visual conventions of a website on user performance and disorientation: how bad can it be? In: Proceedings of the 26th Annual ACM International Conference on Design of Communication, pp. 47–54. ACM (2008)
16. Stenstrom, E., Stenstrom, P., Saad, G., Cheikhrouhou, S.: Online hunting and gathering: an evolutionary perspective on sex differences in website preferences and navigation. IEEE Trans. Prof. Commun. **51**, 155–168 (2008)
17. Muylle, S., Moenaert, R., Despontin, M.: The conceptualization and empirical validation of web site user satisfaction. Inf. Manag. **41**, 543–560 (2004)
18. Grigoroudis, E., Litos, C., Moustakis, V.A., Politis, Y., Tsironis, L.: The assessment of user-perceived web quality: application of a satisfaction benchmarking approach. Eur. J. Oper. Res. **187**, 1346–1357 (2008)
19. Cyr, D.: Modeling web site design across cultures: relationships to trust, satisfaction, and e-loyalty. J. Manag. Inf. Syst. **24**, 47–72 (2008)
20. Eighmey, J.: Profiling user responses to commercial web sites. J. Advertising Res. **37**, 59–67 (1997)
21. Eighmey, J., McCord, L.: Adding value in the information age: uses and gratifications of sites on the World Wide Web. J. Bus. Res. **41**, 187–194 (1998)
22. Salvucci, D.D., Taatgen, N.A., Borst, J.P.: Toward a unified theory of the multitasking continuum: from concurrent performance to task switching, interruption, and resumption. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 1819–1828. ACM (2009)