

Screen or Tabletop: An Eye-Tracking Study of the Effect of Representation Location in a Tangible User Interface System

Mina Shirvani Boroujeni^(✉), Sébastien Cuendet, Lorenzo Lucignano, Beat Adrian Schwendimann, and Pierre Dillenbourg

CHILI, École polytechnique fédérale de Lausanne, Lausanne, Switzerland
{mina.shirvaniboroujeni,lorenzo.lucignano,beat.schwendimann,
pierre.dillenbourg}@epfl.ch, sebastien.cuendet@gmail.com

Abstract. This article presents a comparison of the effects of input-output location (co-located versus discrete) on user performance in a tangible user interface (TUI) system. We conducted a mobile eye-tracking study with two different versions of a TUI system and, despite similar performances in both conditions, our findings revealed differences in the users gaze patterns, shedding new light on the underlying cognitive processes.

Keywords: Tangible interface · Eye-tracking · Representation location

1 Introduction

The potential benefits of tangible user interfaces (TUIs) and the related new possibilities for the design of learning experiences drew the attention of HCI and education research communities in recent years [12]. Training spatial skills is one of the areas in which TUIs are shown to be effective and this could be explained by the physicality of the tools towards the development of mental models [2]. TapaCarp Fig. 1a is a recent augmented tabletop system developed by Cuendet for improving spatial skills [1]. However, the high hardware and maintenance costs of TapaCarp prevent widespread application by educational institutes. To reduce the costs, the author developed a web-based version, named eTapaCarp, which does not rely on any dedicated hardware components but uses off-the-shelf web-cams. As a consequence, in eTapaCarp the digital feedback is not co-located within the manipulation space, but it is represented on a discrete plane (i.e. a monitor, Fig. 1a and b).

The goal of this user study is to discern the differences in user performances and gaze patterns brought by the design of the two systems, identifying to what extent it affects the users experience with TUI systems.

2 Related Work

Price and colleagues [9] provided a comprehensive framework to classify TUIs, which features the location of input and output as a discriminative parameter.

It refers to the relative spatial positioning of physical and digital components of the system. Three modalities have been proposed for location: discrete, co-located, or embedded. Location can impact attention demands [5], potential for representing multiple levels of abstraction [8], as well as the collaborative group working [10]. The investigation of those aspects of designing TUIs has rarely exploited the eye-tracking methodology, although it has been successfully employed in interaction and usability research [7]. Recently, authors in [4] have used eye-tracking data to compare the user experience in a TUI environment and GUI counterpart. Results of this study suggest some cognitive advantages in using TUIs as the gaze values in the tangible condition indicated less demanding mental efforts. In [11], Schneider and colleagues reported how the use of eye-tracker data could reveal some information regarding the underlying mental processes of the user while performing a task. Those works encouraged us to investigate the effect of location from the novel perspective of gaze behaviors.

3 Experimental Setup

The experimental setup distinguished two conditions: discrete (eTapaCarp) and co-located (TapaCarp) displays. We will refer to these two conditions by the system titles. TapaCarp (Fig. 1a) is a tabletop user interface system. Users interact with the system through manipulating small wooden blocks equipped with fiducial markers. The system dynamically projects the three orthographic projections (*top*, *front* and *side* views) of the blocks on the tabletop. In eTapaCarp (Fig. 1b), users manipulate the objects in the horizontal workspace and see the feedback on an external vertical screen. In eTapaCarp (Fig. 1b), users manipulate the objects in the horizontal workspace and see the feedback on an external vertical screen.

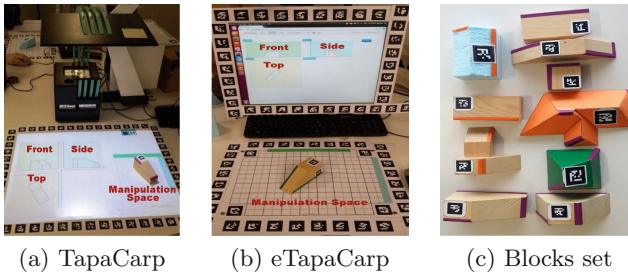


Fig. 1. Experimental setup

Research Questions: There are two main research questions we address in this paper. Q_1 : “Does the separation of input and output spaces in eTapaCarp affect the user performance?” Our initial hypothesis is that this design choice in eTapaCarp could lead to a split of attention, which in turn might affect the task performance, for example by lowering the accuracy or interaction speed. Q_2 : “Does the eye gaze data provide insight into differences of representational

location and user interaction?” Previous works such as [2,3] often came short explaining the observed differences, whereas the use of eye-tracker data might enable us to go a step further into understanding how users are thinking.

Task: The experimental task was a set of 10 edge finding questions. In an edge finding question, three edges are highlighted using colored tapes on the physical block (Fig. 1c). The user is given the task of identifying and selecting the corresponding edges on each of the three orthographic projections. The edge selection is performed by a mouse click.

Participants: A total of 18 university students (16 men, 2 women) from the 2nd and 4th academic year in Mechanical Engineering and Microtechnique sections participated in the experiment. Their average age was 21.4 years and 9 participants were randomly assigned to each of the two conditions.

Procedure: The experiment was composed of two parts: (1) A pretest on spatial skills, consisting of a mental rotation and a paper folding task [6] with a time limit of 3 min for each test, (2) the set of 10 edge finding questions, performed on TapaCarp or eTapaCarp, depending on the condition. There was no time limit for completing this part and participants gaze data was collected employing SMI Eye-Tracking Glasses with binocular pupil tracking at 30 Hz.

4 Results

The statistical tests were made using ANOVAs on linear models. Repetitions were taken into account using mixed effect models when needed. For the eye-tracking data analysis, the workspace is divided into four areas of interest (AOIs) corresponding to the physical block (*object*) and the three orthographic views. The unit for the eye tracking analysis is the dwell, defined as one visit in an AOI from entry to exit.

Pretest Scores: The 18 participants got an average score of 73.61% (SD: 24.63%) in the mental rotation test and 85.55% (SD: 12.47%) in the paper folding test. According to the literature [1], these scores indicate well-developed spatial reasoning skills.

Activity Duration: The average time to complete all the edge finding questions was 9.5 min (SD: 3.12 min). Comparing the two conditions, in eTapaCarp completion of all questions took slightly less time than in TapaCarp (8.6 min vs. 10.3). However this difference was not statistically significant ($F[1,16]=1.40$, $p > .1$).

Performance Accuracy: Participants on average completed 8.11 (SD: 1.97) out of 10 questions with no mistake in any of the views. Analysis of the number of correctly answered questions in the two conditions reveals no significant difference (8.33 in TapaCarp vs. 7.89 in eTapaCarp, $F[1,16]=0.22$, $p=.65$). The total number of mistakes (missing or excess edges) made by each user was also not significantly different between the two conditions (around 7.15 mistakes, over

90 edges to be found in total, in both conditions). In general, most mistakes occurred in the *side* view (4.3 in TapaCarp and 5.11 in eTapaCarp) and the number of mistakes in *front* and especially *top* view were relatively low.

Strategy: All participants in the experiment preferred to find all edges in one view before switching to next view, which we call a “group-by-view” strategy. Most of the participants adopted this strategy starting from the first question, but 4 in TapaCarp and 3 in eTapaCarp started with a “group-by-edge” strategy and after one or two questions, they also switched to group-by-view since they found their initial approach quite confusing. The most common order for completing the views was *front*, *side* and *top* which can be related to the positioning of the views in the output plane, where the *front* view is in the top-left area.

Dwell Percentage, Duration and Revisit Time: As depicted in Fig. 2, in both systems the largest proportion of dwells was related to the physical block (42% in TapaCarp and 34% in eTapaCarp), and the smallest to the *top* view (12% in TapaCarp and 19% in eTapaCarp). In TapaCarp the percentage of dwells on the physical block was higher than eTapaCarp (marginally significant $F[1,16]=4.04$, $p < .06$), while in TapaCarp dwells on the *top* view were significantly lower in proportion ($F[1,16]=38.41$, $p < .0001$). For the other two AOIs (*front* and *side* views), participants in both conditions had quite equal dwell ratios. Moreover, TapaCarp participants showed shorter dwell duration ($F[1,16]=14.41$, $p=.001$), especially on the *side* view for which the average dwell duration was significantly smaller than eTapaCarp ($F[1,16]=21.68$, $p < .001$). The revisit times for each AOI, namely the time between two consecutive entries in an AOI, were shorter in TapaCarp condition ($F[1,16]=13.21$, $p=.002$). Moreover, there was a stronger significant difference between the revisit times on the *object* in the two conditions: the revisit time for the *object* in TapaCarp was shorter ($F[1,16]=15.11$, $p=.001$).

Gaze Transitions Among AOIs: According to Fig. 2, transitions among views and *object* were generally more prevalent in TapaCarp (except for *top* view) whereas transitions within the orthographic views were more frequent in eTapaCarp. However transitions in both graphs exhibit an *object* centric distribution. In TapaCarp transitions between *object* and 2D views account for around 85% of total transitions and this value in eTapaCarp is around 70% which are both considerably large proportions. Furthermore, transitions back and forth between *side* view and *object* were significantly higher in TapaCarp ($F[1,16]=4.96$, $p < 0.5$), whereas the same value for the *top* view was significantly higher in eTapaCarp ($F[1,16]=10.36$, $p < .001$).

5 Discussion and Conclusion

Regarding our first research question (user performance), participants in both conditions performed equally well. Moreover, the absence of a significant difference in the task duration implies that there was no considerable time overhead imposed by the separation of input and output planes in eTapaCarp. Considering the task completion strategies, the group-by-view approach was preferred

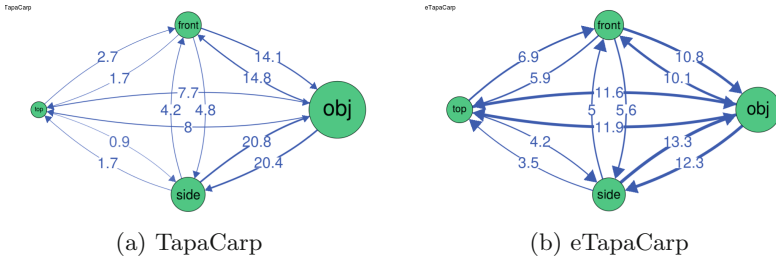


Fig. 2. Transition between AOIs: Nodes size are proportional to the average percentage of dwells on the AOIs, and edge labels indicate the average percentage of transitions among AOIs.

since the cognitive cost imposed by switching the views in the group-by-edge approach is higher than switching the edges in group-by-view strategy.

Regarding our second research question (can eye-tracking tell us more?), the eye-tracking data provided more insights about the completion strategies between the two conditions. In both conditions, *object* absorbed most of the dwells and the transition graphs showed the centrality of this area. This indicates that relating views with the physical blocks was preferred by the participants probably due to the ease of extracting spatial information from the 3D block as compared to the 2D representations. However, two main results characterized the two conditions: in TapaCarp, the short revisit time and the higher dwell proportion on the *object* suggest that the users referred more often to the physical block, frequently exploiting the scaffold provided by the external representation rather than referring to their mental model; in eTapaCarp, the lower percentage of transitions among views and block could be resulting from the separation of input and output planes which increases the cost of switching between the two parts of the interface. This resulted in mapping the orthographic view of interest to another orthographic projection in which the participant had already identified the edge and implies the use of a partially internalized mental model. Extracting information in this way requires more effort and this is probably the reason why participants in eTapaCarp reported longer dwell duration. Another point coming from the transition analysis is the higher connectivity level between the *side* view and the *object* in TapaCarp, and between the *top* view and the block in eTapaCarp. A possible explanation for this can be the proximity effect. In TapaCarp the spatially closest view to the block zone is the *side* view, while in eTapaCarp it is the *front* view. According to the high ratio of references to the *object* and the relatively higher difficulty level of the *side* view, placing this view next to the block zone is a rational choice in the interface design of TapaCarp.

In conclusion, although this study presents some limitation due to the laboratory setting and a relatively small number of participants who belonged to a specific population, it contributes to the TUIs research providing design hints based on quantitative eye-tracking measurements, which provided a new insight on the problem and enabled us to go a step further in the explanation of the

differences caused by the variation of output location. Despite the absence of significant differences in the performance measures between the two systems, co-located representation could be more appropriate for training low skilled users, thanks to the facilitation in accessing the physical 3D model and in linking it to the digital information; Discrete location discourages the reference to the tangible model, in favor of relying on the mental model. Consequently, splitting the input-output is more suitable for users with better developed visual-spatial abilities. Future researches are encouraged to explore deeply the impact of design choices and extend design rules for tangible interface and their applications in learning activities.

References

1. Cuendet, S.: Tangible interfaces for learning: training spatial skills in vocational classrooms. Ph.D. thesis, EPFL (2013)
2. Cuendet, S., Bumbacher, E., Dillenbourg, P.: Tangible vs. virtual representations: when tangibles benefit the training of spatial skills. In: Proceedings of the 7th Nordic Conference on Human-Computer Interaction: Making Sense Through Design, pp. 99–108. ACM (2012)
3. Cuendet, S., Jermann, P., Dillenbourg, P.: Tangible interfaces: when physical-virtual coupling may be detrimental to learning. In: Proceedings of the 26th Annual BCS Interaction Specialist Group Conference on People and Computers, pp. 49–58. British Computer Society (2012)
4. Lucignano, L., Cuendet, S., Schwendimann, B.A., Shirvani Boroujeni, M., Dehler, J., Dillenbourg, P.: My hands or my mouse: comparing a tangible and graphical user interface using eye-tracking data. In: Fablearn 2014. No. EPFL-CONF-204226 (2014)
5. Patten, J., Ishii, H., Hines, J., Pangaro, G.: Sensetable: a wireless object tracking platform for tangible user interfaces. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 253–260. ACM (2001)
6. Peters, M., Laeng, B., Latham, K., Jackson, M., Zaiyouna, R., Richardson, C.: A redrawn vanderberg and kuse mental rotations test-different versions and factors that affect performance. *Brain Cogn.* **28**(1), 39–58 (1995)
7. Poole, A., Ball, L.J.: Eye tracking in HCI and usability research. *Encycl. Hum. Comput. Interact.* **1**, 211–219 (2006)
8. Price, S., Falcão, T.P., Sheridan, J.G., Roussos, G.: The effect of representation location on interaction in a tangible learning environment. In: Proceedings of the 3rd International Conference on Tangible and Embedded Interaction, pp. 85–92. ACM (2009)
9. Price, S., Sheridan, J.G., Falcão, T.P., Roussos, G.: Towards a framework for investigating tangible environments for learning. *Int. J. Arts Technol.* **1**(3–4), 351–368 (2008)
10. Rogers, Y., Lindley, S.: Collaborating around vertical and horizontal large interactive displays: which way is best? *Interact. Comput.* **16**(6), 1133–1152 (2004)
11. Schneider, B., Sharma, K., Cuendet, S., Zufferey, G., Dillenbourg, P., Pea, A.D.: 3D tangibles facilitate joint visual attention in dyads. In: International Conference on Computer Supported Collaborative Learning (CSCL), pp. 158–165 (2015)
12. Shaer, O., Hornecker, E.: Tangible user interfaces: past, present, and future directions. *Found. Trends Hum. Comput. Interact.* **3**(1–2), 1–137 (2010)