UI Error Reduction for High Spatial Visualization Users when Using Adaptive Software to Verify Addresses

Thitivatr PatanasakPinyo[1] and Les Miller[2]

[1] Faculty of Information and Communication Technology, Mahidol University
Salaya, Nakhon Pathom, 73170, Thailand
`thitivatr.pat@mahidol.edu`

[2] Department of Computer Science, Iowa State University
Ames, Iowa, 50012, USA
`lmiller@iastate.edu`

Abstract

An adaptive software system is known as an application that can adapt itself based on different conditions of users. There are multiple conditions/criteria that can be used to direct how an application would adapt. Spatial visualization (VZ) is one of several human spatial abilities that is used to predict human's performance when using a computer application. Therefore, a difference in VZ level is a suitable choice as an adapting indicator, i.e., high VZ and low VZ users should get different features on a user interface (UI) to complete the same task. In this paper, we look at three studies where we asked participants to verify a set of housing addresses using a location-based application on an Android tablet with different versions of the application, especially, an adaptive version of the application was involved in the last study. We found that, for high VZ participants, the number of UI errors that participants created was significantly smaller when they were equipped with the adaptive software. We refer to a UI error (User Interface Error) as an error where a user tapped on a non-sensitive region of the screen. The results of the three studies and hypothesis tests for significance are reported.

1 Introduction

An interaction between a computer and a user becomes an essential topic in the area of computer science since the world has been in the era of ubiquitous computing. Most business transactions, education activities, and recreation activities of humans are done via interaction with a computer. However, most application software still comes with a single user interface (UI) for everybody, whereas there are multiple studies [4, 7] said that a single UI has never been a solution for usability problems. This statement holds in an address verification task using location-based software as a tool. The address verification task is a task done by US Census Bureau crews to verify correctness of a housing address, such that, both an actual real world location and the location of the house on the map correspond. Researchers [13, 15] found that users with different levels of spatial visualization (VZ) might require different UI features to work with when they used the location-based software to verify addresses because

their performance was better when they got the appropriate UI for their level of VZ. VZ is an individual spatial ability that tells how good we are when we have to do mental manipulation of spatial information/representation [18]. A user's level of VZ and performance when using a computer application are highly correlated [9, 11, 19, 22]. PatanasakPinyo [12] discovered that users, regardless of their level of VZ, could complete the task of address verification significantly faster when they used a VZ-based adaptive version of the location-based software.

We did three studies to observe different aspects of user performance. All three were address verification tasks using location-based software on an Android tablet.

In the three studies, we logged every action that a participant did on the UI such as pan or zoom. One variable that we logged was a total number of UI errors. A UI error, as defined by Taylor [21], is a tap/click that a user does on a non-sensitive region. A non-sensitive region is a region on a UI that contains no component nor control, i.e., a region without button, slider, text area, etc. In earlier work Taylor, et al. [21] showed that these errors result from users' failure to accurately use the sensitive parts of the UI. The results from the analysis in Section 4 illustrates that the errors are also tied to the usability of the UI features available to the high VZ users. In particular it points out the importance of the choice of UI features in adaptive systems based on spatial visualization.

We organized this paper as follows: Section 2 walks through related literature. Section 3 describes the details of the three studies, which include both designs and procedures. Section 4 shows the results of three studies, research questions, and hypothesis tests for significance. Finally, Section 5 summarizes the contribution described in this paper.

2   Related Work

Spatial visualization (VZ) is a component of spatial ability. VZ was defined by Salthouse et al. [18] as "mental manipulation of spatial information to determine how a given spatial configuration would appear if portions of that configuration were to be rotated, folded, repositioned or otherwise transformed" [Salthouse, Timothy A., et al. "Sources of individual differences in spatial visualization ability." Intelligence 14.2 (1990): 187-230]. Norman [9] stated that VZ was able to tell how good an individual was when using a computer system. His original message was "The primary cognitive factor driving differences in performance using computer-based technology is spatial visualization ability" [Norman, Kent L. "Spatial visualization - A gateway to computer-based technology." Journal of Special Education Technology 12.3 (1994): 195-206]. Multiple studies also supported this claim. Batinov et al. [1] found that users with different VZ level also had a different methodology when doing an address verification task. PatanasakPinyo et al. [14] developed prediction models to predict a user's level of VZ based on a set of actions that a user did on a UI when doing an address verification task.

Benyon et al. and Benyon [2, 3] gave a definition of an adaptive system as "Adaptive systems are systems which can alter aspects of their structure, functionality or interface in order to accommodate the differing needs of individuals or groups of users and the changing needs of users over time" [Benyon, D., et al. "Special issue on intelligent interface technology: editor's introduction" Interact with Computers 12.4 (1987): 315-322]. The motivation for developing an adaptive system to bridge the gap of the usability problem when users have different levels of spatial ability were examined by several researchers [8, 16, 17].

3   Methodology

Each of the three studies performed for this work consist of two parts. The first part was an evaluation of a participant's VZ level. The second part was a field study where we asked every participant to do the address verification task using a location-based application on Android tablet. All three studies were approved by the IRB Committee of Iowa State University (IRB IDs are 14-526, 15-494, and 16-288). Any identifiable data of a participant was not recorded. Every participant was given a compensation. Every participant was a resident of Ames that was interested in participating and made a contact with the PI via an email address that was posted on community boards all around the town.

3.1   Evaluation of Spatial Visualization

For the VZ evaluation, we used the Paper Folding Test (VZ-2) [5] for all three studies. We asked each participant to complete the VZ-2 test, which consisted of two problem sets. Each set has ten problems with a time limit of three minutes. For each problem, a participant had to mentally fold and punch a hole on a square paper according to the instructions. A participant received 1 point for each correct answer. He/she received 0 point if they provided no answer. However, if an answer was incorrect, he/she received -0.33 point. The participant was classified as high VZ-2 person if he/she scored 13 or more. Scoring 8 or less was classified as a low VZ-2 person. Figure 1 shows an example of a problem in a VZ-2 test.
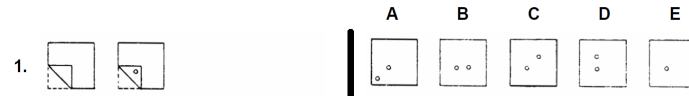

Figure 1: An Example of Problem in VZ-2 Test.

3.2   Field Studies

For all three studies, after a participant finished the VZ-2 test, we escorted him/her to the neighborhood located at the south of Iowa State University, main campus, to continue with a field study (Figure 2). Although the task that we asked participants to do was the address verification task using a location-based application on an Android tablet for all three studies, there were some differences in the detail of each study. This subsection provides a brief explanation of the studies that includes a design, a procedure, and a location-based application that were used.

3.2.1   Study 1

We had thirty-one participants in **Study 1**. After finishing the VZ-2 test, the principle investigator (PI) drove the participant to the neighborhood. The PI handed the participant a Google Nexus tablet that had the location-based application installed and launched. The PI also gave instructions of how to use it to verify all ten addresses. The participant started the address verification task when he/she was ready. The PI walked after the participant to observe and to be close in case that he/she had any difficulty using the application.
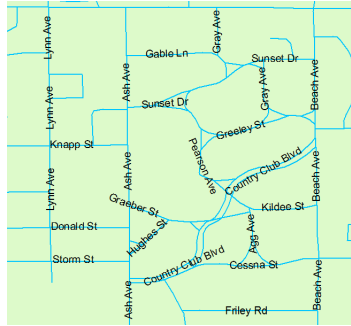
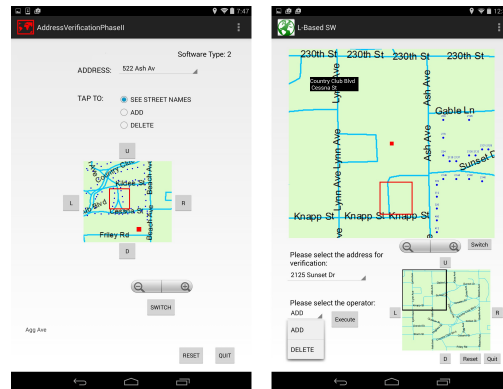Figure 2: The Neighborhood for the Field Study.

Figure 3: (a) Application UI of **Study 1**. (b) Application UI of **Study 2**.

Figure 3 (a) shows the UI of the location-based application that was used in **Study 1**. The drop-down list on top contains all ten addresses that a participant must verify. The radio button list contains options how the UI responds when a participant taps on the primary map. There is one primary map located at the center. The primary map shows street names and address numbers of all houses inside that map area. The four pan buttons are located at the corresponding edges of the primary map. The zoom control and the switch button are located at the bottom-right of the primary map. A participant can tap on the switch button to set the map type of the primary map between a Tiger map and a street map that uses map images from Google maps. The reset button is used when a participant wants to start over. A participant taps on the quit button when he/she completed all ten addresses. Every action that a participant did on the screen was automatically logged by the software.

In this study, every participant was randomly assigned two out of three features (GPS, Object-Indexing, Mini Map). The participant who got the UI in Figure 3 (a) had GPS and Object-Indexing. GPS displayed a red dot on the primary map representing the actual location that the participant was at. Object-Indexing was activated when the first option in the radio button list was checked and the participant tapped on the primary map. The application listed all names of streets located in the tapped area denoted by a red rectangle. The list of street names was displayed at the bottom-left of the UI. Mini Map, which was an overview of the

whole neighborhood, was a feature that this participant did not have.

3.2.2   Study 2

For **Study 2**, we had another thirty-two participants. The procedure of **Study 2** was similar to **Study 1**. Every participant in **Study 2** took the VZ-2 Test and then was escorted to the neighborhood. Each participant had to verify ten addresses (different addresses from **Study 1**) using another version of the location-based application.

Figure 3 (b) shows the UI of the location-based application that a participant used in **Study 2**. The UI used in **Study 2** was a universal UI where each participant had the same components. The UI in Figure 3 (b) shows the components a participant used to verify an address such as pan buttons or a zoom control. The primary map is 4-time larger than the one in **Study 1** (Figure 3 (a)). Furthermore, all three features (GPS, Object-Indexing, Mini Map) are available all the time. GPS displays a red dot on the primary map representing the location of the participant. For Object-Indexing, a participant can tap a street or intersection on the primary map to see the corresponding street names. The street names are displayed at the top-left corner of the primary map. The Mini Map is located at the bottom-right of the UI. All ten addresses are contained in the drop-down list. Another drop-down list contains add and delete commands that a participant used to correct any housing address on the primary map.

3.2.3   Study 3

We divided the field study of **Study 3** into three phases. A participant had to verify four addresses, four addresses, and two addresses in Phase 1, Phase 2, and Phase 3, respectively. A participant had to do every phase in chronological order starting at Phase 1. We used the same neighborhood as the previous two studies with 10 new addresses. Each participant was randomly assigned one of two treatments: Treatment 0 and Treatment 1. If a participant was assigned Treatment 0, the application was launched with the Traditional UI for him/her to work with in Phase 1. After finishing Phase 1, the application switched to the appropriate Adaptive UI for working in Phase 2. The application worked in the opposite way if Treatment 1 was assigned. It gave a participant Adaptive UI for Phase 1 and Traditional UI for Phase 2. Since both the purpose and detail of Phase 3 were to develop a prediction model to predict a level of VZ in real time, which is not related to this paper, we omit an explanation of Phase 3.

The Traditional UI (Figure 4 (a)) is a UI that contains only the necessary components needed to do the address verification task plus some features that were proved to be beneficial for every user regardless of their VZ levels [4, 13]: GPS and Mini Map. There are two versions of the Adaptive UI: Adaptive UI for low VZ participants (Figure 4 (b)) and Adaptive UI for high VZ participants (Figure 4 (c)). The Adaptive UI for low VZ participants (Figure 4 (b)) also included Object-Indexing for Low VZ and landmarks. The Object-Indexing for low VZ works the same way as it did in previous studies except that it displays street names right above the region a participant taps, which allows low VZ participants to retrieve information easily without moving their focus out of the primary map. The inclusion of landmark is another feature that was added to the Adaptive UI for low VZ to represent a landmark such as a treehouse, a playground, or a stadium. The use of landmarks proved to be useful for users with a low level of VZ [6, 10, 20]. Since most high VZ users prefer a UI with minimal design, i.e., a UI that contains only necessary components required to get the task done, we added only Object-Indexing to it. The only difference is that Object-Indexing for high VZ participants

Figure 4: (a) Traditional UI. (b) Adaptive UI for Low VZ. (c) Adaptive UI for High VZ.

displays street names out of the primary map, in particular, it displays them at the bottom-left of UI as shown in Figure 4 (c). The reason is that we would like to have the primary map "clean" in the viewpoint of high VZ participants.

4   Results and Discussion

This section reports the results and a discussion of our analysis of data that we gathered from the three studies. This also includes a research question and corresponding hypothesis tests.

4.1   Results

Although we logged every action (i.e., every tap on UI) of every participant from all three studies, we report only one variable due to space limitation, which is a total number of error taps, i.e., a tap on a non-sensitive region. Box plots (Figure 5) show the distribution of the number of error taps for both low VZ participants and high VZ participants in each study.
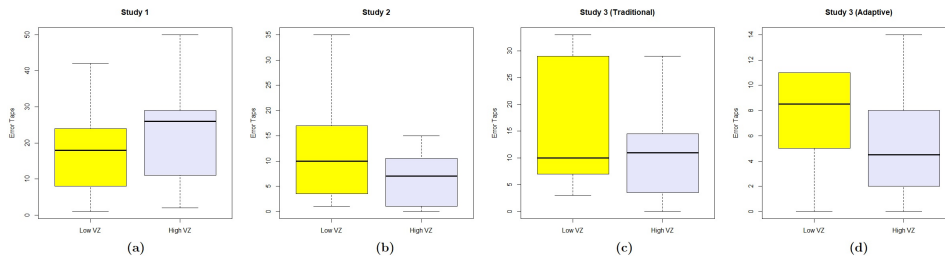

Figure 5: Box Plots for(a) **Study 1**. (b) **Study 2**. (c) **Study 3 (Traditional)**. (d) **Study 3 (Adaptive)**.

4.1.1   Study 1

From thirty-one participants, seventeen participants tested low VZ and thirteen tested high VZ (one participant tested mid-VZ which was not included in the process of hypothesis test). Figure 5 (a) shows the distribution of the number of error taps done by low VZ participants (yellow) and high VZ participants (lavender).

4.1.2   Study 2

From thirty-two participants, fifteen participants tested low VZ and seven participants tested high VZ (ten participants tested mid-VZ which were not included in the process of hypothesis test). Figure 5 (b) shows the distribution of the number of error taps done by low VZ participants (yellow) and high VZ participants (lavender).

4.1.3   Study 3 (Traditional)

From thirty participants, ten participants tested low VZ and twenty participants tested high VZ. Figure 5 (c) shows the distribution of the number of error taps done by low VZ participants (yellow) and high VZ participants (lavender).

4.1.4   Study 3 (Adaptive)

Figure 5 (d) shows the distribution of the number of error taps done by low VZ participants (yellow) and high VZ participants (lavender) when they used the Adaptive UI in **Study 3** regardless of the assigned treatment.

4.2   Analysis

According to the box plots in Figure 5, we can observe a rough pattern that both low VZ participants and high VZ participants produced less error taps when they used the UI that contained features that were useful for their group (Figures 5 (b) & (d)) than another type of UI (Figures 5 (a) & (c)). We raised a research question to see whether or not the result we had is statistically useful.

4.2.1   Normalization of Number of Error Taps

Since the total number of addresses to be verified in each study was not equal, particularly, **Studies 1** and **2** had ten addresses, but the first two phases of **Study 3** had four addresses each, this incident might intuitively cause a biased situation such that a participant probably produced more error taps when he/she had more addresses to verify. Hence, we normalize a number of error taps of each participant in **Studies 1** and **2** before making any comparison or statistical process with those in **Study 3** using the following formula:

$$e' = \frac{n_3 e}{n}$$

From the above formula, e' is the normalized number of error taps of the participant of either **Studies 1** or **2**. e is an original number of error taps. n and n_3 are a number of addresses to be verified in **Study 1** (or **Study 2**) and **Study 3** (either phase), respectively.

4.2.2   Research Questions and Hypothesis Tests

A research question was set up as "Is there a difference in a number of error taps that a participant of same VZ level produced in two distinct studies?". With this research question, two hypotheses are proposed:

H_0 : There is no difference in numbers of error taps of participants in **Study** x and **Study** y.

H_a : There is a difference in numbers of error taps of participants in **Study** x and **Study** y.

From both hypotheses, $x, y \in \{1, 2, 3(\text{Traditional}), 3(\text{Adaptive})\}$. Since each study was independent and had different number of participants, we used Mann-Whitney as a hypothesis test to test for significance of each pair of studies. However, we report only test results of high VZ participants because there was no test on low VZ participants of any pair of studies that yielded significant difference. Table 1 shows results of hypothesis test. The first column indicates which two studies that we ran the test on their data (numbers of error taps produced by high VZ participants). The second column indicates the p-value we got from the test. The last column indicates whether or not we rejected H_0 according to the corresponding p-value.

Table 1: Results of hypothesis test on error taps by high VZ participants in each pair of studies.

Studies (only high VZ)	p-value	Test Result
Study 1 and Study 2	0.008769	Reject null hypothesis
Study 1 and Study 3 (Traditional)	0.74	Fail to reject hypothesis
Study 1 and Study 3 (Adaptive)	0.01419	Reject null hypothesis
Study 2 and Study 3 (Traditional)	0.01991	Reject null hypothesis
Study 2 and Study 3 (Adaptive)	0.1019	Fail to reject hypothesis

From Table 1, the three tests show that a difference in numbers of error taps is statistically significant. The first test that confirmed this significance is a test between **Studies 1** (a UI with necessary components and two random features) and **2** (a UI with necessary components and complete three features regardless of users' preference), which returned a p-value of 0.008769. The test between **Study 1** and **Study 3 (Adaptive)** also confirmed the significance with a p-value of 0.01419. Recall that **Study 3 (Adaptive)** came with a UI that had necessary components and a feature (Object-Indexing for high VZ) that was proved to be useful for high VZ participants. The last test that showed a significance is a test between **Study 2** and **Study 3 (Traditional)** with a p-value of 0.01991. A UI of **Study 3 (Traditional)** had only necessary components without any extra feature.

In an analysis view, we can categorize three studies into two groups. Group 1, which contains **Study 1** and **Study 3 (Traditional)**, comes with a concept that a UI has a limited set of tools that a participant can use to verify addresses. Some tools or features that they had, if they belonged to either of these two studies, might be seen as non-suitable or non-interesting by such participants. On the other hand, Group 2, which contains **Study 2** and **Study 3 (Adaptive)**, comes with a concept that a UI had extra tools/features that were really useful

to participants and they used them to enhance their task of address verification. From Table 1, it is obvious that every test between a study in Group 1 and a study in Group 2 resulted that high VZ participants from a study in Group 1 produced more error taps than participants from a study in Group 2 significantly. The other two tests, which were a test between two studies from the same group, resulted that numbers of error taps were not significantly different as we expected.

5   Conclusion

We did three studies. All of them were to invite participants to do the address verification task using a location-based application on Android tablet. Every participant took VZ-2 test to determine a level of spatial visualization (high/low) before going to the neighborhood to verify addresses. Each study had a different version of UI. Every action that a participant did was automatically logged by the application. After all three studies were done, we extracted, from a log file, the number of error taps that each participant made to analyze. Although we cannot see any significance from low VZ participants, we found that high VZ participants who were in a study that used a UI with extra features that proved to suit users with high level of VZ produced less error taps than high VZ participants who were in a study that used a UI without extra features or a UI that had irrelevant features. The difference that we found was tested significant.

Future work in this area will look at option for low VZ users. Of particular interest is whether the design of features for low VZ UIs need to take into consideration the techniques used by Taylor, et al. [21] to adjust the sign of action components on the screen such as buttons, radio buttons, text boxes, etc.

6   Acknowledgements

We would like to express our special thanks to Georgi Batinov, Kofi Whitney, and Adel Sulaiman for their insights during group discussion.

References

[1] Georgi Batinov, Michelle Rusch, Tianyu Meng, Kofi Whitney, Thitivatr Patanasakpinyo, Les Miller, and Sarah Nusser. Understanding map operations in location-based surveys. In *Eighth International Conference on Advances in Computer-Human Interactions (ACHI 2015)*, pages 144–149, Lisbon, Portugal, 2015.

[2] D Benyon, PR Innocent, and D Murray. System adaptivity and the modeling of stereotypes, interact'87. In *Second IFIP Conference on Human-Computer Interaction, Stuttgart, Elsevier Science Publishers*, 1987.

[3] David Benyon. Accommodating individual differences through an adaptive user interface. *Human Factors in Information Technology*, 10:149–149, 1993.

[4] Thorsten Büring, Jens Gerken, and Harald Reiterer. Usability of overview-supported zooming on small screens with regard to individual differences in spatial ability. In *Proceedings of the working conference on Advanced visual interfaces*, pages 233–240. ACM, 2006.

[5] Ruth B Ekstrom, John W French, Harry H Harman, and Diran Dermen. Manual for kit of factor-referenced cognitive tests. *Princeton, NJ: Educational testing service*, 1976.

[6] Liisa AM Galea and Doreen Kimura. Sex differences in route-learning. *Personality and individual differences*, 14(1):53–65, 1993.

[7] Chris M Law, Ji Soo Yi, Young Sang Choi, and Julie A Jacko. Unresolved problems in accessibility and universal design guidelines. *Ergonomics in Design: The Quarterly of Human Factors Applications*, 15(3):7–11, 2007.

[8] Alan F Newell and Peter Gregor. "user sensitive inclusive design"—in search of a new paradigm. In *Proceedings on the 2000 conference on Universal Usability*, pages 39–44. ACM, 2000.

[9] Kent L Norman. Spatial visualization—a gateway to computer-based technology. *Journal of Special Education Technology*, 12(3):195–206, 1994.

[10] Elizabeth M O'Laughlin and Bradley S Brubaker. Use of landmarks in cognitive mapping: Gender differences in self report versus performance. *Personality and Individual Differences*, 24(5):595–601, 1998.

[11] Richard Pak, Wendy A Rogers, and Arthur D Fisk. Spatial ability subfactors and their influences on a computer-based information search task. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 48(1):154–165, 2006.

[12] Thitivatr PatanasakPinyo. *Flattening methods for adaptive location-based software to user abilities*. Graduate Theses and Dissertations, Iowa State University, 2017.

[13] Thitivatr PatanasakPinyo, Georgi Batinov, Kofi Whitney, and Les Miller. Methods that flatten the user space for individual differences in location-based surveys on portable devices. In *31st International Conference on Computers and Their Applications (CATA 2016)*, pages 65–70, Las Vegas, Nevada, 2016. International Society for Computers and their Applications (ISCA).

[14] Thitivatr Patanasakpinyo, Georgi Batinov, Kofi Whitney, Adel Sulaiman, and Les Miller. Enhanced prediction models for predicting spatial visualization (vz) in address verification task. In Gordon Lee and Ying Jin, editors, *Proceedings of 34th International Conference on Computers and Their Applications*, volume 58 of *EPiC Series in Computing*, pages 247–256. EasyChair, 2019.

[15] Thitivatr PatanasakPinyo, Georgi Batinov, Kofi Whitney, Adel Sulaiman, Les Miller, and Stephen Gilbert. Extracting useful features for users with different levels of spatial visualization. In *33rd International Conference on Computers and Their Applications (CATA 2018)*, pages 86–91, Las Vegas, Nevada, 2018. International Society for Computers and their Applications (ISCA).

[16] Krish Ramachandran. Adaptive user interfaces for health care applications. *IBM developerWorks*, 2009.

[17] Karthik Narayanan Ramalingam. Modeling error-based adaptive user interfaces. 2011.

[18] Timothy A Salthouse, Renee L Babcock, Debora RD Mitchell, Roni Palmon, and Eric Skovronek. Sources of individual differences in spatial visualization ability. *Intelligence*, 14(2):187–230, 1990.

[19] Maung K Sein, Lorne Olfman, Robert P Bostrom, and Sidney A Davis. Visualization ability as a predictor of user learning success. *International Journal of Man-Machine Studies*, 39(4):599–620, 1993.

[20] Lynn A Streeter and Diane Vitello. A profile of drivers' map-reading abilities. *Human factors*, 28(2):223–239, 1986.

[21] Alfred Taylor Sr, Les Miller, Sree Nilakanta, Jeffry Sander, Saayan Mitra, Anurag Sharda, and Bachar Chama. Using an error detection strategy for improving web accessibility for older adults. In *Advances in Computer-Human Interactions, 2009. ACHI'09. Second International Conferences on*, pages 375–380. IEEE, 2009.

[22] Kim J Vicente, Brian C Hayes, and Robert C Williges. Individual differences in computer-based information retrieval. In *Ergonomics and human factors*, pages 225–229. Springer, 1987.