Hard Lessons: Effort-Inducing Interfaces Benefit Spatial Learning

Andy Cockburn[†], Per Ola Kristensson[‡], Jason Alexander[†], Shumin Zhai[*]

[†]Computer Science, University of Canterbury, Christchurch, New Zealand, {andy, Jason}@cosc.canterbury.ac.nz

[‡]Computer & Information Science, Linköpings Universitet, 581 83 Linköping, Sweden, perkr@ida.liu.se

[*]IBM Almaden Research Center, 650 Harry Road, NWE-B2, San Jose, CA 95120, USA, zhai@us.ibm.com

ABSTRACT

Interface designers normally strive for a design that minimises the user's effort. However, when the design's objective is to train users to interact with interfaces that are highly dependent on spatial properties (e.g. keypad layout or gesture shapes) we contend that designers should consider explicitly increasing the mental effort of interaction. To test the hypothesis that effort aids spatial memory, we designed a "frost-brushing" interface that forces the user to mentally retrieve spatial information, or to physically brush away the frost to obtain visual guidance. We report results from two experiments using virtual keypad interfaces – the first concerns spatial location learning of buttons on the keypad, and the second concerns both location and trajectory learning of gesture shape. The results support our hypothesis, showing that the frost-brushing design improved spatial learning. The participants' subjective responses emphasised the connections between effort, engagement, boredom, frustration, and enjoyment, suggesting that effort requires careful parameterisation to maximise its effectiveness.

Author Keywords

Skill acquisition, education, training, gesture stroke, pen input, text entry, spatial memory, learning.

ACM Classification Keywords

H5.2 [User Interfaces]: Interaction styles.

INTRODUCTION

Discovering mechanisms and methods that help users to acquire skills in using computer applications has long been an important research topic in HCI. An early example is Carroll's seminal work on "Training Wheels" [2, 3]. There is also extensive research into computer-based training for tasks that are dangerous or costly to practice in the real world (e.g. [1, 15]). The primary objectives of these training systems are to familiarise users with their task environment, to aid task planning and stress management, and to tune visual attention skills.

In this paper we study the hypothesis that "hard" interfaces that promote mental effort will be more beneficial than "easy" ones for the purpose of learning, particularly when acquiring skills concerning spatial memory. Although this hypothesis is supported both by intuition (we remember a driving route better if we spend effort navigating through it, rather than being a leisurely passenger) and by insights in the general psychology literature (e.g. [7, 8, 23, 34]), there has been little empirical work testing it with user interfaces. If it is true that effort facilitates learning, then the designers of training systems should consider making them "harder".

The current work explores this topic through theoretical analysis, interface design, and empirical investigation. We first review related work, and then we present the rationale of a frost-brushing interface design aimed at inducing greater effort on the user's part. The main body of the paper presents two formal experiments: one involved learning spatial locations of interface widgets, and the other involved learning both locations and spatial trajectories. Although spatial memory and learning affect human-computer interaction quite broadly [4, 5, 11, 12, 14, 28, 33, 35], two mobile text entry methods (virtual keyboarding and shape writing [18, 36]) were chosen as experimental tasks for the following reasons. First, mobile interaction is increasingly important and developing effective UI and UI training methods is a timely topic. Second, text entry is a skill that requires a great deal of learning or training. Third, spatial memory is intensely involved in virtual keyboarding (memorizing the key locations) and shape writing (memorizing gesture trajectories). We also discuss how the results can be applied in desktop user interfaces. Our studies show that the effortful frost-brushing design does improve spatial memory, but that level of effort needs careful consideration to balance the user's subjective experience of factors including boredom, frustration, and enjoyment.

RELATED WORK

Psychology research has long demonstrated that deeper, effortful, or intentional cognition improves users'

CHI 2007, April 28–May 3, 2007, San Jose, California, USA.

understanding of, and memory for, entities they interact with. Craig and Lockhart [7] proposed a "depth of processing" framework for human memory in which the strength of episodic memory is a positive function of the levels of semantic involvement in processing stimuli. Subsequent experiments supported the theory, showing that deeper encodings took longer to process, but resulted in higher performance in word memory tests [8]. Similar positive results for deep encodings have also been produced in spatial memory tests [23, 34]. A more direct basis of our hypothesis that a more effortful interface may facilitate better learning is Schmidt and Bjork's survey of skill acquisition research in two separate fields of human motor control and memory [29]. They found that a common effective mechanism for skill acquisition across different domains is to encourage active retrieval of information from memory. They also observe that "manipulations that degrade the speed of acquisition can support the long term goals of training" (p207).

The idea that soliciting greater effort may improve learning can also be found in the HCI literature. One example is Ehret's study [13]. Although Ehret's primary research objective was to predict performance using ACT-R cognitive architecture, his empirical findings of training "where to look" in graphical user interfaces is the closest prior work to our current investigation. His experimental interface displayed a circle of buttons representing different colours arranged around a central "cue" button. Trials involved clicking one of the circle buttons to match the colour of the cue. Participants trained to learn the colour/location associations using four conditions that varied the labelling of the circle buttons: "colour-match" labelled buttons with a colour dot; "meaningful" labelled them with the name of the colour; "arbitrary" displayed an arbitrary icon; and "no-label" used blank buttons. All conditions allowed users to consult a tooltip showing the button's colour after a 1-sec delay. These conditions were intended to influence the "evaluation cost" of determining the target location – from the low cost "colour match" to the demanding "no-label" condition. Ehret showed that participants could better re-construct the location of the buttons after using the high cost conditions. Perhaps due to the modelling objective of the research, the same number of trials was used in all four experimental conditions. This created a confound to the demonstrated learning effect since the participants on average spent more than twice the time training with the highest cost interface than with the lowest cost one. It is therefore unclear whether the learning advantage was due to higher "cost" or to longer practice time. Similarly, it remains unclear whether the benefit of the high cost interface could be obtained if the same amount of practice time was given to all conditions.

Beyond location learning, experiments by O'Hara and Payne [25, 26] showed that users learn to solve puzzles such as the "eight-puzzle" (which involves sliding eight numbered tiles in a 3×3 matrix to reach a target configuration) more effectively when the interface promotes "planning" by increasing the per-operation cost to the user. Alternative implementations of operation cost included slowing the rate of system feedback and increasing the complexity of action specification (e.g. typing "A1" to specify the tile to slide, rather than clicking on it directly). Again, the argument is that increasing interface effort has a positive learning outcome due to the promotion of deeper cognitive processing. Shedigan and Klawe [31] describe a similar interface strategy for promoting "reflective cognition" in math learning, but we are unaware of their empirical results.

Marking menus [19] provide an example of a design exhibiting an interface "cost" that might facilitate the transition to recall based expert behaviour. They delay the display of the pie menu content, forcing the user to either recall the direction of the desired menu item or to wait for the pie menu to reveal itself.

Finally, there is extensive research demonstrating the important role that spatial memory, which is the main dependent variable of our investigation, plays in computer use. Several studies agree that measures of spatial aptitude correlate well with efficiency when using text editors [12], computer games [14], and file managers [35]. Several studies also show that interfaces that lack spatial constancy harm performance [5, 33], while those that maintain spatial constancy aid it [4, 11, 28].

In summary there are compelling theoretical reasons and a few user interface examples that support the hypothesis that effortful interfaces result in better spatial learning. We set out to formally test the hypothesis in two experiments: the first involving spatial location learning of a keypad layout; the second involving both location and trajectory components in gestural shape-writing.

AN EFFORTFUL FROST-BRUSHING INTERFACE

User effort can be instantiated in many ways, as exemplified by the six measures of the NASA "Task Load Index" [17]: mental/physical/temporal demand, effort, performance and frustration. The goal of our design is to promote mental effort (active use of memory) by raising the physical and temporal effort of retrieving items from the display. Although we evaluate only one effort-inducing design, we later discuss other methods.

Our design overlays "frost" on graphical widgets, obscuring their labels (but leaving the widget outlines intact). To use a frosted object, one has to either mentally recall its content or "brush off the frost" by waving the mouse cursor over the object to reveal its label. If left alone, the object gradually fades back to its original frosted state (Figures 1 and 3).

This frost-brushing mechanism imposes a physical-cognitive effort trade-off on the user. To perform an interactive task or to activate a widget, the user has two options. The first is to spend time and physical effort to brush off the frost in order to follow the mentally "easy" path of reacting to the guidance of the visual display. The second is to take the more mentally challenging path of exerting cognitive effort to recall the label or function of the widget, which should foster better learning and greater memory strength. Gray and Fu have shown that people rely more on memory, even if imperfect, when the cost of accessing visual information is higher [16]. In our design, the amount of physical effort needed to reveal the visual information in exchange of mental effort is a continuum that can be adjusted by the properties of either the frost or the brush. For example, to make it physically easier, the frost can form slowly and thinly (so the labels of the widgets are not completely obscured). Conversely, the frost can be thick and fast forming. Similarly, the brush can be wide or narrow, more efficient (brushing away more frost with one stroke) or less efficient (repeated strokes are needed to remove the frost).

If designed and adjusted appropriately to the task, we believe a frost-brushing interface can make the learning experience more effortful, more planful, and more fun. We apply such a mechanism in two spatial learning tasks, one involving object locations only and the other involving both location and spatial trajectory.

EXPERIMENT 1: SPATIAL LOCATION LEARNING

This experiment tests the hypothesis that inducing greater effort improves learning the location of objects in graphical user interfaces. More specifically, it tests how well participants learn the location of buttons (keys) on graphical virtual keyboards when trained with a traditional visible interface or a more effortful brushing interface. Unlike Ehret's experiment, which held number of training trials constant, we held the training period constant for both experimental conditions.

Users often employ letter mnemonics to aid their recollection of key positions (e.g. "qwertyuiop" is a well known keyboard mnemonic). In order to avoid unwanted artifacts of letter mnemonics in our experiment, we used iconic symbols rather than letters on each button (see Lee and Zhai [21] for a study of keyboard mnemonics).

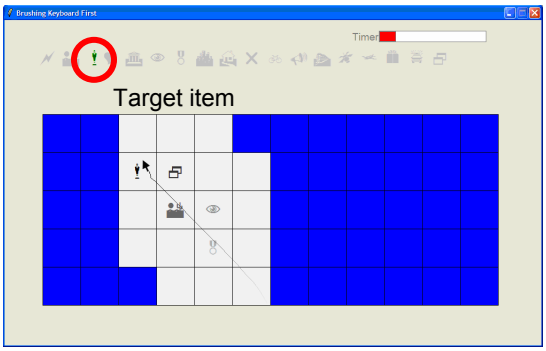


Figure 1. The brushing keyboard in Experiment 1. The faded line shows the trajectory of the cursor.

All participants trained for five minutes to learn the location of 18 iconic symbols using both the brushing interface and a visible interface as the control condition. In the visible condition the labels on the keys were displayed throughout training. Following training with each interface the participants' location recall was tested using a blank keyboard.

Apparatus

The entire experimental interface (Figure 1) ran in a window of fixed dimensions at 1000×600 pixels on a 15inch 1400×1050 pixel display. The eighteen active keys had a white background, and inactive keys were blue. A target-cuing region above the virtual keyboard showed the next target symbol, highlighted in green. It also contained a timer that showed the remaining training time.

When training using the visible interface, the symbols were always visible on the keyboard. When using the brushing interface at most five symbols in the brushed area were visible, with the most recently brushed key displayed in black, and the others progressively fading to invisible (Figure 1). When users stopped brushing, the keys faded to invisible over a one second period.

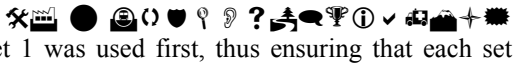
The experiment ran on a Compaq nx9010 laptop (2.66GHz Pentium 4) running Microsoft Windows XP. Input was received through a high quality optical mouse. The software controlled the participants' exposure to the experimental conditions and logged all user actions.

Participants

The 14 volunteer participants in this within-subjects experiment were all post-graduate computer science students or staff, two female. The experiment lasted approximately 20 minutes for each participant.

Procedure

All participants completed familiarisation, training and testing with their first interface, and then repeated the process with the second. Half of the participants used the brushing interface first; the other half used the visible interface first. Two sets of 18 iconic symbols were used, all from the Microsoft "Webdings" font:

- Set 1: 
- Set 2: 

Symbol set 1 was used first, thus ensuring that each set was balanced across training conditions. Each symbol was shown in the same keyboard location for all participants.

Familiarisation. The familiarisation process consisted of a brief explanation of their next training interface (visible or brushing), followed by 50 seconds practice using a keyboard containing the digits 0 to 9 randomly arranged.

Training. During the training period, the participants used their assigned interface (visible or brushing) for five minutes. They were instructed that the objective was to become as efficient with the keypad as possible, and that memorising item locations would help them achieve this. The 18 symbols were displayed in the target-cueing region, with the next target item highlighted green. Tasks involved finding and selecting the target on the keyboard (either by brushing or by visually searching for it if using the visible interface). Each successful acquisition caused a confirmation beep, and the next randomly selected symbol was highlighted green. An incorrect selection caused an error tone to be played. Participants continued to search for the same symbol until correctly selected.

To promote equitable training opportunity across the 18 symbols, each target item was randomly selected from the set of unused symbols until the set was empty, at which point it was refilled with all 18 items.

When the five-minute training period expired, the software presented three Likert-scale questions (1 disagree, 5 agree). All questions began with "Training with this interface was" and finished with either "boring", "fun", or "challenging".

Testing. The testing period consisted of two tasks: a *tapping task* and a *memory recall task*. The tapping task involved tapping out the 18 symbols on the visible virtual keyboard as quickly as possible, which enabled us to observe how those trained in the brushing interface adapt when switching to a visible interface, as compared to those trained in the visible condition. Each target symbol was randomly selected and highlighted green in the target-cuing region. Correct and incorrect selections were confirmed with an audible beep or error tone. The target symbol only advanced following a correct selection.

The memory recall task followed the tapping task. All of the symbols on the virtual keyboard were blank. Participants were asked to select the location of the 18 symbols, presented in a random order, without any feedback for correct or incorrect selections. An audible "click" confirmed that a selection had been made, and the next target in the cueing window was immediately identified by green highlighting. After selecting locations for all 18 items, the participants were asked to estimate how many items they had correctly selected.

The primary hypothesis is that the effortful frost-brushing interface will allow better spatial location learning than the visible interface. The following two dependent measures can support or reject this hypothesis:

- The number of correct selections during the memory recall task. Improved spatial location memory should result in a higher number of correctly selected targets.
- The mean Euclidian distance between the selected key and the correct key during the memory recall task: improved spatial location memory should result in a smaller distance between the actual target location and the location of the selected key.

Results

Performance improvement during training

The participants made fewer iterations through the 18 symbols when using the brushing interface because each selection took more time. Also, the participants varied in their speed of selections, meaning that some made several more iterations through the sets than others. To accommodate this variance, only the first n iterations are included in the analysis, where n is the number of iterations made by at least ten of the fourteen participants: brushing interface $n=5$; visible $n=10$.

Regression analysis of the mean selection times across iterations closely adhere to the power law of practice [9, 24] for both training conditions. This law states that task completion time improves across trials according to the formula $T_n=C\cdot n^{\alpha}$, where T_n is the time to complete trial n, C is the time on the first trial, and α is the steepness of the learning curve. Best fit models with the brushing and visible interfaces are $T_n=4.53\cdot n^{-0.466}$ and $T_n=2.38\cdot n^{-0.227}$, (both R^2=.99). The training data and the power-law models are shown in Figure 2.

For the brushing interface we also analysed the mean distance brushed per symbol in each iteration, which decreases from 1377 pixels in the first iteration to 395 in the fifth. These values correlate well with the mean time per iteration (linear R^2=.99). This suggests that participants refined their spatial memory across iterations, consequently reducing the search space for each symbol.

Tapping speed during the testing period

Although the participants completed fewer practice trials in the brushing condition, when they switched to the visible interface for the testing period, their average

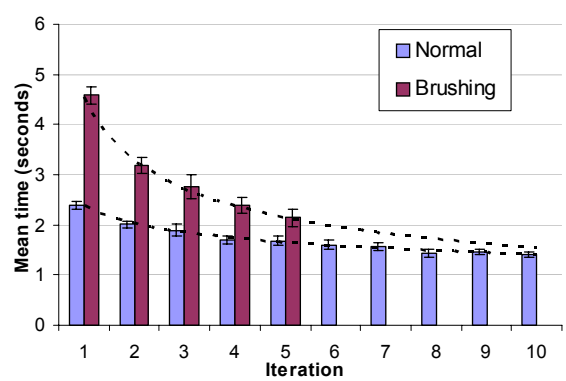


Figure 2. Mean time per iteration for the brushing and visible interfaces in Experiment 1 training. Dotted lines are power-law regression for both conditions.

tapping time was not significantly different from those who trained in the visible condition. The mean time was 27.8s (sd 5.4) in the brushing condition and 26.8s (sd 5.6) in the visible condition: paired T_{13}=1.5, p=.16.

Memory recall

The primary hypothesis – that the "effortful" brushing interface will improve spatial location learning – is supported by the *memory recall* performance. Both the number of correct selections and the mean miss distance show reliably better location memory when trained using the brushing condition. Participants made more correct selections when trained with the brushing interface (mean 13.4, sd 3.8) than with the visible interface (mean 11.9, sd 4.7): paired T_{13}=2.5, p<.05. Similarly, the miss-distances were much smaller with the brushing interface: brushing mean 498 pixels (sd 480), visible mean 754 pixels (sd 637), paired T_{13}=3.1, p<.01. The relatively large difference between the mean miss distances across the two training conditions (498 versus 754 pixels) suggests that the participants were more likely to guess key locations following training with the visible interface. This interpretation is supported by our observation that participants frequently made comments of the nature "I've no idea" following visible training.

The number of correct selections was not significantly influenced by the difference between symbol Set 1 (mean 12.2, sd 4.5) and Set 2 (mean 13.0, sd 4.1), as tested by an unpaired T-Test: T_{26}=0.5, p=.6.

Subjective responses

In practice any improvement in location learning provided by an effortful interface would be of little value if the interface is perceived to be more boring or less fun – users would simply abandon it earlier. Analysis of the subjective data suggests that the brushing interface made its training condition less boring and more challenging. Mean responses to the five-point Likert-scale questions (1-disagree, 5-agree) showed that the brushing interface was less boring than the visible one (paired T_{13}=2.33, p<.05), with means of 2.7 (sd 1.2) and 3.9 (sd 1.1) respectively. Similarly, the brushing interface was more challenging than the visible interface (T_{13}=4.6, p<.01): brushing mean 3.9 (sd 1.1), visible mean 2.2 (sd 1.3). There was no significant difference between the perceived fun of the two training interfaces (T_{13}=1.5, p=0.16), with brushing and visible means of 3.0 and 2.4 respectively.

After their testing period with each training condition, the participants were asked to estimate how many items they had selected correctly. There was a marginal difference between these estimates for the two training conditions, suggesting the participants were aware of their better performance with brushing. The mean estimates with the brushing and visible interfaces were 10.9 (sd 5.3) and 9.4 (sd 5.1) respectively: T_{13}=2.1, p=.05.

Finally, as seen in previous experiments [6] the participants underestimated the acuity of their spatial memory, indicated by a significant difference between their estimation of the number of items correctly selected (mean 10.1, sd 5.1) and the actual number of correct selections (mean 12.6, sd 4.2): T_{27}=4.7, p<.01.

The main conclusions of Experiment 1 are as follows:

1. During training with the brushing interface users quickly rely on their spatial memory to reduce the amount of brushing required to find objects.
2. An effortful interface improves spatial memory when the amount of training time (rather than the number of training trials) is experimentally controlled.
3. Participants found training with the brushing interface less boring and more challenging than a constantly visible interface.

EXPERIMENT 2: SPATIAL TRAJECTORY LEARNING

To further test the hypothesis that inducing effort improves spatial learning, we decided to replicate and extend the findings of Experiment 1. While Experiment 1 focused purely on spatial *location* learning, Experiment 2 inspects the impact of effort and planning on skill acquisition in a task involving both location and trajectory. We chose shape writing, a word level text entry method [18, 36] as the experimental task. Shape writing uses pen strokes on a graphical keyboard to enter words. Each word is defined by its "sokgraph" – the path connecting the characters in the word on a graphical keyboard layout (Figure 3 top). Figure 3 (bottom) shows the user entering the word "good" using the frost-brushing version of ShapeWriter. The sokgraph for "good" is roughly a "V" shape, from "g" in the middle of the top row of keys to "o" vertically below, then diagonally up and left to "d" on the top row. The sokgraph can be drawn with pauses separating movements from key to key (while the user visually searches for the next character) or using a fluid gesture for the entire word.

Depending on the level of experience, the user can either draw the sokgraph with a series of discrete stylus movements, pausing to visually search for each subsequent key, or the whole sokgraph can be entered in a fluid gesture if they can retrieve the word shape from their memory. We apply the brushing/frosty technique to training with shape writing in this experiment. The user has to brush away the frost on the keyboard to determine the location of the required characters to draw the sokgraphs, unless they can recall their shape and location from past memory. In addition to active recall, the brushing interface should also promote planning because users must brush to reveal not only individual letters (like in Experiment 1), but also enough letters to visualize the trajectory that links a series of letters to make a word. As one can imagine, this experiment is much more demanding than the previous one, which gave us an

Figure 3. ShapeWriter (top) and the 'brushing' training game (bottom). The ideal trace of the recognized sokgraph is displayed on the keyboard.

opportunity to observe whether there is an appropriate level of effort beyond which the brushing interface ceases to be effective.

The experiment used an order-balanced within-participant design to compare training effects in brushing and visible conditions: each participant, in either A-B or B-A order, performed in both conditions A and B. Although more sensitive than a between-participant design, a within-participant design runs the risk of asymmetrical skill transfer [27]. While we did not find asymmetrical skill transfer in Experiment 1, which employed different symbol sets between the two conditions, it did occur in Experiment 2 for some dependent variables, as indicated by their statistically significant Order × Condition interaction effects. We therefore rely on a between-subject analysis based only on data collected in the first condition for each participant (dropping half of the data, collected as the second condition of each participant). One exception is the subjective ratings. Since there was no performance involved in these measures and participants' subjective ratings could only be more informed with a relative comparison, we used all within-participants data for the subjective measure analyses. Note that for the objective performance measures none of the conclusions based on the between-participant analysis was changed qualitatively when a within-analysis was performed.

Participants

The 22 volunteer participants were all university students, six female. None had previous experience with shape writing. Six had previously used stylus input devices, and five were familiar with gesture-based systems having used systems such as marking menus and Graffiti. The female participants, and those experienced with stylus-based input systems, were balanced between the conditions. Each participant's involvement in the experiment lasted approximately 60 minutes.

Apparatus

The experiment ran on a Pentium 4 computer running Microsoft Windows XP. The experimental interface (Figure 3) ran in a 340×450 pixel window within a 1400×1050 pixel display. Input was received through a Wacom Graphire2 tablet and stylus.

A customized version of the ShapeWriter software controlled the participants' exposure to the experimental conditions, and it logged all user actions.

Procedure and Results

This was a multi-phase experiment. Phase 1 involved training participants to enter words using either the brushing interface or the visible one. Phase 2 tested the accuracy in reproducing those trained words on a visible keyboard. Phase 3 assessed participants' memory recall of the words trained during Phase 1 without the benefit of seeing the keyboard at all.

For ease of comprehension, results from each phase of the experiment are presented immediately after the phase's rationale, procedure and measurement are described.

Phase 1 — Training with the visible or brushing interface

At the start of the experiment, the participants were given a two-minute demonstration of ShapeWriter, with the experimenter showing how to enter the text "Shape writing is fast" both by moving discretely from character to character and by gesturing shapes for words. The alphabetical tendency in the ATOMIK layout (Figure 3 top) [32] was explicitly illustrated by the experimenter at the launch of the ShapeWriter software. Participants were shown how to delete words by striking through the text and how to correct words by selecting from the probable alternatives menu associated with each word. They were also shown the training game, including the frosting behaviour of the brushing interface when using it (Figure 3, bottom). They were then allowed two minutes of free practice with the interface, playing the balloon game with words randomly selected from the phrase "Soon good at shape text".

Table 1. The words used in Experiment Two.

Length			
2	**3**	**4**	**5**
of	and	that	which
in	was	have	their
is	you	they	could
he	not	were	other

The participants then trained with shape writing for fifteen

minutes either in the brushing or visible condition. The training period involved repeatedly entering sixteen English words (Table 1), which consisted of the four most frequently occurring English words of lengths 2, 3, 4, and 5 characters (http://www.wordcount.org.)

During the fifteen minute training period, each of the sixteen words in the word set was presented in a random order. All of the words were presented *n* times before any word was presented *n+1* times. Each randomly selected word was presented within a balloon in the training game (Figure 3). Each balloon "floated" to the ceiling and stayed there until "popped" by correctly entering its sokgraph, which caused the next target word to appear. This process continued until the training period expired.

Phase 1 Results

Since one has to sweep away the frost to reveal letters in the brushing-condition, it took markedly longer to enter each word. The mean word-entry time was 4.5s (sd 2.0) for the visible condition and 8.4s (sd 4.7) for the brushing condition. Since we held the same training period (15 minutes) for both conditions, the brushing condition had fewer practice repetitions with each word. The mean repetition counts with the visible and brushing conditions were 12.0 (sd 4.1) and 6.1 (sd 3.4).

Several participants had substantial difficulty seeking and remembering the characters in five-letter words when using the brushing interface. Participant 15, for example, spent in excess of one minute entering each of the three five-letter words on their first presentation. The maximum times spent on a five letter word were 81s with brushing and 35s with visible.

Figure 4 shows the mean time taken to enter words with the two training conditions across repetition, based on the number of blocks completed by at least nine of the eleven participants in each condition. Regression analysis shows good fits with traditional power-law of practice models: brushing R^2=0.99, visible R^2=0.94.

Phase 2 — Testing with visible keyboard

After the 15 minutes training, all participants did testing with the visible ShapeWriter interface — importantly the brushing mode was not used regardless of training condition. The same balloon game was used in this test, with one randomly selected word appearing in each balloon, and new balloons appearing until all sixteen words had been correctly entered three times. Like in Phase 1, all words from the set were presented *n* times before any word was shown for an $n+1^{th}$ time. The dependent variables were the time taken to enter each word, measured from the time the balloon first appeared to successfully "popping" it by correctly entering the text, and the percentage of trials containing an error.

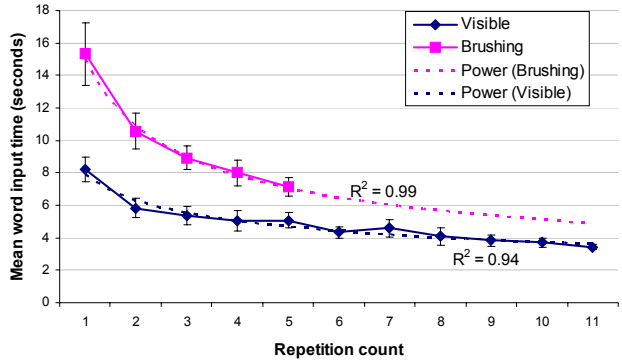


Figure 4. Mean time to input words during Phase 1 practice with the visible and brushing interfaces.

Having completed the testing period, the participants were asked to rate the training interface using measures from the NASA-TLX (Task Load Index) worksheets.

Phase 2 Results

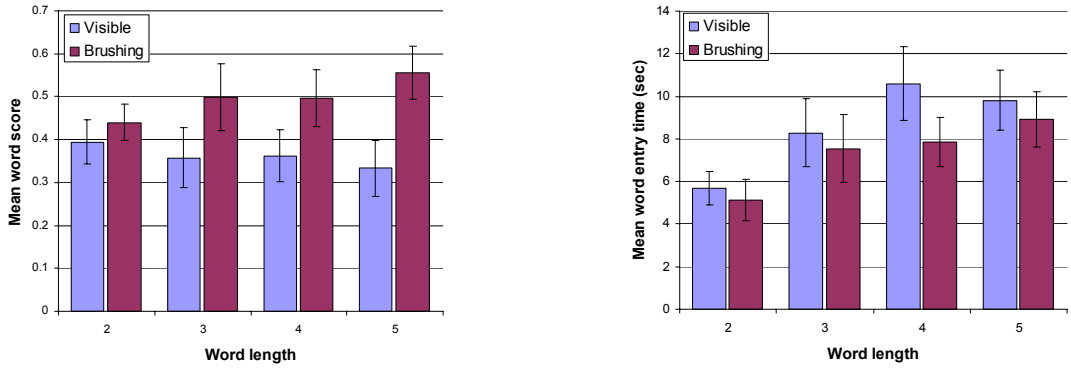
Time and error data are analysed using a 2×4 mixed-factors analysis of variance for factors *training condition* (visible or brushing, between-subjects) and *word length* (2, 3, 4, or 5 characters, within-subjects).

Results show that switching from the brushing interface to the visible one did not significantly damage performance, despite the very different behaviour required and the significantly fewer practice trials available during training. Training condition had no significant effect on error (visible mean 8%, brushing mean 12%, $F_{1,20}$=1.1, p=0.3) nor on word entry time (visible mean 3.6s, brushing mean 3.9s, $F_{1,20}$<1). There was a significant effect of word-length on word entry time ($F_{3,60}$=33.9, p<0.01), with means increasing from 2.2s (sd 0.5) for two-letter words to 4.8s (sd 1.9) for 5 letter words. There was no significant condition × length interaction for errors or word entry time ($F_{3,60}$<1 in both cases).

Table 2 summarises the subjective responses of the two training interfaces rated on a five point scale from 1 to 5 (low to high). As stated earlier, subjective analysis is within-participants. It shows that the participants found the brushing interface significantly less enjoyable, more

Table 2: Questionnaire responses from 1 (low) to 5 (high). Wilcoxon matched pairs significance tests.

Measure	Mean (sd)		Significant?		
	Visible	Brushing		z	p
Enjoyment	3.8 (1.1)	3.0 (1.4)	✓	2.2	0.01
Effective (poor-good)	4.0 (0.7)	4.0 (1.0)	×	0.1	0.5
Mental demand	2.6 (1.0)	3.2 (1.1)	✓	2.4	0.01
Physical demand	2.0 (1.0)	1.9 (1.1)	×	0.6	0.29
Temporal demand	2.3 (0.9)	2.4 (1.3)	×	0.2	0.4
Performance (poor-good)	3.5 (0.9)	3.0 (1.0)	✓	1.9	0.02
Frustration	2.0 (1.1)	2.4 (1.1)	✓	1.9	0.03
Overall effort	2.9 (1.0)	3.3 (1.1)	✓	2.4	0.01



(a) System word score function. **(b) Word entry time.**

Figure 5. Mean word score and entry time in Phase 3.

mentally demanding, more frustrating and more effortful than the visible one. These results are interesting as they confirm our impression that the brushing condition demanded *too much* effort and planning. In contrast to the first experiment, where the brushing interface had a positive effect on subjective measures, here we see the inverse, with effort reducing enjoyment and increasing frustration. This issue is further discussed below.

Phase 3 – Memory recall without the keyboard

Phase 3 answered the main question of this experiment – whether spatial learning in a trajectory-based task can be improved by the more effortful brushing interface. We tested participants' memory recall performance when the key labels on the keyboard are invisible. All sixteen words trained in Phase 1 were presented in a random order, one at a time, in a cuing window. The participants were asked to create the shape needed to enter each word on a *blank* keyboard, showing only the keypad grid. No feedback was presented.

Shape memory recall was measured by system-generated score function between 0 and 1 that reflects the distance between the stroke drawn and the ideal sokgraph trace of the target word. Such a score is also the basis of the recognition engine in ShapeWriter [18].

Phase 3 Results

Data was analysed using the same 2×4 RM-ANOVA as Phase 2. Participants trained in the brushing condition showed significantly higher recall scores (0.5) than those trained in the visible one (0.36): $F_{1,20}=4.3$, $p=0.05$. In other words, although subjectively the participants found the brushing interface less enjoyable and more frustrating, their memory recall performance was indeed better in the more effortful brushing condition. The mean time to draw the words was 7.4s following the brushing condition and 8.6s following visible, but not statistically significant ($F_{1,20}<1$). Figure 5 summarises these results.

Longer words again were slower to input than short ones ($F_{3,60}=15.4$, $p<0.01$), but the score function remained consistent across word length ($F_{3,60}<1$). There were no interactions between training condition and length for word entry time or score value ($F_{3,60}\sim1$ in both cases).

DISCUSSION AND FUTURE WORK

Greater effort indeed improves spatial learning

The results of the two experiments show that the users' spatial memory for both location and trajectory can indeed be improved by inducing greater effort during training through the frost-brushing interface. Although only practiced for short periods of time (5 and 15 minutes respectively), significant and marked differences in memory recall performance were measured between the frost-brushing and visible conditions. In both experiments the total amount of practice time was held constant between the two conditions, therefore exclusively attributing the performance difference to the mechanisms imposed in the frost-brushing interface. Importantly the findings were consistent in both location and trajectory spatial memory tasks in two separate experiments involving different degrees of effort and task complexity.

In addition to memory recall, we also measured participants' time and accuracy performance when switching from the brushing interface to a normal visible interface. Moving from a harder interface to an easier interface could mean slower or less accurate performance if the skills learned in the harder interface do not apply in a normal setting. The experimental results were not significantly different between training conditions even though the mechanisms and number of practice trials were quite different. The participants could switch to the normal visible interfaces without significant performance degradation. On the other hand, despite the improved spatial memory, participants were not significantly faster or more accurate in the presence of a visual display. This opens questions for further investigation. It is plausible that automaticity [22, 30] only develops after extensive learning (much more than the 5 or 15 minutes training given in our experiments), and that only then will the spatial memory advantage be strong enough to significantly improve performance beyond that available through visual reaction to graphical displays. Whether the

more effortful learning, as induced by frost-brushing or some other interface, would help the development of automaticity earlier than simple repetition remains to be discovered in future longitudinal studies.

Level of effort, engagement, and alternative designs

In Experiment 1, the participants rated the frost-brushing interface more positively than the visible interface, suggesting the levels of effort imposed were appropriate or acceptable. However, the level of brushing effort required in Experiment 2 was apparently beyond participants' tolerance level, resulting in more negative subjective ratings despite their improved memory.

Obviously the level of effort required should be adjusted to be challenging and engaging, above boredom (which can be the case of simple repetition practice with a visible interface) but below frustration (which borders with the case of Experiment 2). As Csikszentmihalyi [10] describes, engaging experience requires more challenge as skills increase. Fortunately for the frost-brushing interface, the level of effort required can be continuously adjusted by changing the quantitative properties of the brush (such as the brush width) or the frost (such as thickness and speed). Furthermore, it is also possible to change the type of effort required. For example, several participants criticised the nature of the brushing effort required in Experiment 2 because it prohibited partial task completion – with brushing the users had to accurately plan their entire shape trajectory prior to beginning the stroke, whereas with the visible interface they could stroke over the initial characters of a word before pausing to seek the latter ones. There are many alternate interface designs that would not prohibit partial task completion: for example, the display of all characters on the keyboard could "pulse", periodically fading into and out of view, or the pressure sensitive capabilities of the tablet could be used to reveal progressively more characters around the stylus when pressure is increased.

Applications

Findings of basic theoretical and empirical research can be applied in many ways, often beyond those imagined during the initial investigation. Our study shows that users of virtual keyboard and shape writing could improve their spatial memory of key locations and gesture trajectories if they practice these tasks with the frosting interface, even for a short period of time, although more practice may be needed for the memory strength to be great enough to actually speed up their text input speed. Our findings are also relevant to many other applications. For example it is possible to design e-learning software for learning a new set of command gestures or a foreign writing system (such as Chinese) in which a new character or gesture is frosted in order to increase the learner's effortful behavior.

For another example, there is evidence that graphical user interfaces poorly support the transition from novice to expert behaviour [20], with users failing to capitalize on shortcut techniques such as toolbars and keyboard shortcuts. Instead, users trap themselves in beginner mode with mouse-driven menu selections. An "effortful" interface could promote the development of efficient use by pausing temporarily while the toolbar item or keyboard shortcut is animated on the display. A small amount of short-term frustration might, therefore, yield substantial long-term productivity gain.

The understanding that effort or temporal cost improve spatial memory also lends support to the design of interfaces such as marking menus [19] which have a short temporal delay (increasing cost) prior to posting the visual menu, but this delay can be preempted by gesturing in the direction of the yet-to-be-displayed menu item. From our results, we suspect that users would learn marking menu gestures less well if the delay was eliminated.

With further research and development, effort inducing mechanisms such as these may also be beneficial in e-learning applications beyond spatial memory tasks.

CONCLUSIONS

We contend that "harder" interfaces, or interfaces that require greater effort from the user, benefit learning spatial tasks in graphical user interfaces. We propose a frost-brushing interface design, which can be varied in many ways, as one method of imposing greater effort on the users to mentally recall and plan their spatial action. Our two formal experiments, one involving location and the other location and trajectory, both show greater memory recall performance following training with the frost-brushing interface than with a visible interface.

Benefits in spatial memory are important because it has been shown to play a critical role in interacting with visual displays. Our further work will focus on the transition from conscious recall to expert automaticity.

ACKNOWLEDGEMENTS

Thanks to Colin Ouellette and Carl Gutwin (University of Saskatchewan) for assistance, advice and support. This project was started at, and supported by, IBM Almaden Research Center during Andy Cockburn's sabbatical visit.

REFERENCES

1. Baker, D., Prince, C., Sherestha, L., Oser, R. and Salas, E. Aviation computer games for crew resource management training. *International Journal of Aviation Psychology 3* (1993), 143-156.
2. Carroll, J. *The Nurnberg Funnel: Designing Minimalist Instruction for Practical Computer Skill*. MIT Press, Cambridge, MA, 1990.
3. Carroll, J. and Carrithers, C. Training Wheels in a User Interface. *CACM 27*, 8 (1984), 800-806.

4. Cockburn, A., Gutwin, C. and Alexander, J. Faster Document Navigation with Space-Filling Thumbnails. in *Proc. CHI'06*, ACM Press, (2006), 1-10.
5. Cockburn, A., Gutwin, C. and Greenberg, S. A Predictive Model of Menu Performance. in *Proc. CHI'07*, ACM Press, (2007), To appear.
6. Cockburn, A. and McKenzie, B. Evaluating the Effectiveness of Spatial Memory in 2D and 3D Physical and Virtual Environments. in *Proc. CHI'02*, ACM Press, (2002), 203-210.
7. Craik, F. and Lockhart, R. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior 11* (1972), 671-684.
8. Craik, F. and Tulving, E. Depth of processing and the retention of words in episodic memory. *J. of Experimental Psych.: General 103*, 3 (1975), 268-294.
9. Crossman, E. A theory of acquisition of speed-skill. *Ergonomics 2*, 2 (1959), 153-166.
10. Csikszentmihalyi, M. *Flow: The psychology of optimal experience*. Harper & Row, New York, 1991.
11. Czerwinski, M., van Dantzich, M., Robertson, G. and Hoffman, H. The Contribution of Thumbnail Image, Mouse-Over Text and Spatial Location Memory to Web Page Retrieval in 3D. in *Proc. INTERACT '99*, (1999), 163-170.
12. Egan, D. and Gomez, M. Assaying, Isolating, and Accommodating Individual Differences in Learning a Complex Skill. in Dillon, R. ed. *Individual Differences in Cognition*, 1985, 173-217.
13. Ehret, B. Learning Where to Look: Location Learning in Graphical User Interfaces. in *Proc. CHI'02*, ACM Press, (2002), 211-218.
14. Gagnon, D. Videogames and Spatial Skills: An Exploratory Study. *Educational Communication and Technology 33*, 4 (1985), 263-275.
15. Gopher, D., Weil, M. and Bareket, T. Transfer of skill from a computer game trainer to flight. *Human Factors 36*, 3 (1994), 387-405.
16. Gray, W. and Fu, W. Soft constraints in interactive behavior: the case of ignoring perfect knowledge in-the-world for imperfect knowledge in-the-head. *Cognitive Science 28* (2004), 359-382.
17. Hart, S. and Staveland, L. Development of NASA-TLX: Results of Empirical and Theoretical Research. in Hancock, P. and Meshkati, N. eds. *Human Mental Workload*, 1988, 139-183.
18. Kristensson, P. and Zhai, S. SHARK2: A large vocabulary shorthand writing system for pen-based computers. in *Proc. UIST'04*, ACM Press, (2004), 43-52.
19. Kurtenbach, G., Sellen, A. and Buxton, W. An Empirical Evaluation of Some Articulatory and Cognitive Aspects of Marking Menus. *Human-Computer Interaction 8*, 1 (1993), 1-23.
20. Lane, D., Napier, H., Peres, S. and Sandor, A. Hidden costs of graphical user interfaces: Failure to make the transition from menus and icon toolbars to keyboard shortcuts. *IJHCI 18*, 2 (2005), 133-144.
21. Lee, P. and Zhai, S. Top-down Larning Strategies: Can They Facilitate Stylus Keyboard Learning. *IJHCS 60*, 5-6 (2004), 585-598.
22. Logan, G. Attention and preattention in theories of automaticity. *American Journal of Psychology 105* (1992), 317-339.
23. Naveh-Benjamin, M. Recognition memory of spatial location information: another failure to support automaticity. *Memory and Cognition 16* (1987), 437-445.
24. Newell, A. and Rosenbloom, P. Mechanisms of Skill Acquisition and the Law of Practice. in Anderson, J. ed. *Cognitive Skills and their Acquisition*, Erlbaum, Hillsdale, NJ, 1981, 1-55.
25. O'Hara, K. and Payne, S. The effects of operator implementation cost on planfulness of problem solving & learning. *Cognitive Psychology 35* (1998), 34-70.
26. O'Hara, K. and Payne, S. Planning and the User Interface: The Effects of Lockout Time and Error Recovery Cost. *IJHCS 50* (1999), 41-59.
27. Poulton, E. Unwanted asymmetrical transfer effects with balanced experimental designs. *Psychological Bulletin 66*, 1 (1966), 1-8.
28. Robertson, G., Czerwinski, M., Larson, K., Robbins, D., Thiel, D. and van Dantzich, M. Data Mountain: Using Spatial Memory for Document Management. in *Proc. UIST'98*, ACM Press, (1998), 153-162.
29. Schmidt, R. and Bjork, R. The Conceptualizations of Practice: Common Principles in Three Paradigms Suggest New Concepts for Training. *Psychological Science 3*, 4 (1992), 207-217.
30. Schneider, W. and Shiffrin, R. Controlled & automatic human information processing: Detection, search, and attention. *Psychological Review 84*, 1 (1977), 1-66.
31. Sedighian, K. and Klawe, M. An Interface Strategy for Promoting Reflective Cognition in Children. in *Companion Proc. CHI'96*, (1996), 177-178.
32. Smith, B. and Zhai, S. Optimised Virtual Keyboards with and without Alphabetical Ordering - A Novice User Study. in *Proc. INTERACT'01*, (2001), 92-99.
33. Teitelbaum, R. and Granda, R. The Effects of Positional Constancy on Searching Menus for Information. in *Proc. CHI'83*, (1983), 150-153.
34. Van Asselen, M., Fritschy, E. and Postma, A. The influence of intentional and incidental learning on acquiring spatial knowledge during navigation. *Psychological Research 70*, 2 (2005), 151-156.
35. Vicente, K., Hayes, B. and Williges, R. Assaying and Isolating Individual Differences in Searching a Hierarchical File System. *Human Factors 29*, 3 (1987), 349-359.
36. Zhai, S. and Kristensson, P. Shorthand Writing on Stylus Keyboard. in *Proc. CHI'2003*, (2003), 97-104.