Effects of Interaction on Human Memory

David Eigen

August 24, 2004

Introduction

Human memory performance is an important factor in visualization design. Often, most of the work scientific researchers spend on their problems is done while *not* using the visualization. Thus, if a researcher can remember a visualization better, he/she may be able to make better use of the visualization in their every-day work.

In addition, better memory performance in an online setting leads to more efficient use of the system. For example, if users can remember interesting settings or previous experiences better, they will be able to navigate and compare between them more easily.

This project's aim was to investigate whether there might be a link between how a user interacts with a dataset and the user's memory of the data. To test this, we performed a series of three experiments testing subjects' memory of one-dimensional lists of words. Subjects performed a matching task in which they moved a list of words to the left and right; immediately afterward, they were tested on their memory of the list structure with a forced-choice recall test. The first two experiments compared positional and velocity-based controls on a desktop computer. The third experiment compared three different interactions in a four-walled virtual environment (CAVE), each of which involved different types of body movements.

Prior Work

Usoh et al. [18] found that users felt greater presence when walking in real-world scenes than when flying. However, it is unclear whether greater presence results in a better understanding of scientific data presented in a virtual environment. It is also not clear whether real-world scenes are a special case, or if full-body interactions have more general effects that might apply in less real (even non-3D) scenes.

There has also been work contrasting of the effects of active and passive exploration in spatial integration of real-world scenes or objects. In active exploration, subjects control their own movement through a scene; in passive exploration, subjects view scenes without interacting with them, such as by watching a video of a walk-though. Gaunet et al [7] had subjects move through a driving simulation either by using a joystick to control their path according to spoken directions, or by watching an equivalent drive-through video. They found no performance differences between the two, in neither a recall task nor an orientation/pointing task. There was also no difference in a reconstruction task where the subject had to draw the path they took on a piece of paper.

However, others have found more subtle performance differences using virtual environments. Sun et al [17] found that subjects actively exploring scenes had orientation-free representations, while subjects passively exploring the same scenes had orientation-specific representations. James et al [8] found that active manipulation of objects lead to shorter reaction times in a recognition task when compared to passive exploration.

These experiments dealt with spatial integration of real-world scenes. In contrast, the stimuli in our experiments have a distinctly unreal but still spatial format. Instead of using "snapshot" memories of a scene, we examine whether the type of interaction may play a role in the memory encoding (or recall) of ordering relations between data items in the stimulus.

Memory of scenes is also dependent on context and viewpoint. Waller et al [19] found that the encoding of large, room-sized layouts is dependent on viewpoint (i.e. is orientation-specific). In addition, humans have greater ability to recall items when they are in (or even imagine) the same environment in which they learned them [16, 3]. People also remember contextual information, including the location where they learned something, or things they learned at around the same time. It is possible that full-body interaction might produce these sorts of contextual effects — for instance, users may encode the direction in which they had to walk to get from one point to another, and then use this information when recalling relationships in the stimulus.

Finally, cognitive load affects memory performance. Users who use more cognitive resources controlling a display have fewer resources left for attending to the content and encoding the information in long-term memory. It is possible that different types of interactions impose different cognitive loads, in which case memory performance will vary.

Experiment 1

Experiment 1 compared positional and velocity-based controls on a desktop computer. Subjects moved a list of words to the left and right using a mouse. The position of the mouse directly corresponded to either the position of the word list or the velocity (i.e. was either a zero- or first-order control). Each trial in this experiment consisted of two parts: a training phase, and a testing phase.

In each training phase, subjects were shown a list of 10 words. Only a subsection of the list fit in the screen width, so subjects had to interactively "scroll" the list in order to see all the words. Subjects performed a matching task in which they were shown a series of target words; subjects had to move the list so that the matching word on the list was aligned below the target. This forced subjects to actively interact with the display, instead of simply studying the list.

Immediately following each training phase, the subject's memory of the list was tested. Subjects were shown pairs of words displayed one on top of the other, and had to choose the word that appeared further to the left in the list. Subjects selected the word they thought was leftmost by typing either the '1' or '2' key.

We used words instead of graphical symbols because we felt that subjects could get symbols from different trials confused, and remember the locations of symbols from previous trials during the recall task. There is very little chance of this happening with words because of their vastly different meanings, and so the different trials are more independent. The drawback to words is that they are a cross-modal stimulus requiring use of both verbal and visuospatial components of working memory. However, because subjects knew of the recall test in advance, we felt they would encode the list using an explicit strategy anyway, making this difference not matter as much.

Method

Each subject in the experiment completed eight successive trials. The positional control was used during the training phase for half the trials. The velocity control was used for the other half. Subjects alternated between the two controls. Half the subjects began with the positional control, and half with the velocity.

Subjects were given a practice session before the first real trial began. In this practice, subjects performed the training task with each type of control, followed by a short testing phase. Both training phases in the practice session used the same list of six words, and all subjects saw this list in the same order. Twelve targets were used for each interaction type. Ten pairs of words were used in the practice testing phase. Subjects were given feedback to ensure that they understood the instructions: depending on whether they correctly indicated the leftmost word, a blue "CORRECT" or red "WRONG" was printed on the screen.

So that subjects did not use the start of each training phase to readjust to the interaction, each trial's training session was preceded by a warmup. The warmup was exactly like the training task, except that the subject matched numbers instead of words. The list used was the numbers 1 through 10, in increasing order. The same pattern of 14 targets was used for each warmup.

Each training task consisted of 15 successive target words. The order of the words in the word list was randomized. The sequence of target words was chosen by concatenating random permutations of the middle 8 words in the list. No two successive targets were the same. Although the two words on the ends were not used as targets during real training phases, they were used as targets during the warmups. Subjects were given unlimited time for the training task, but were instructed to complete it as quickly as possible.

The testing phase consisted of 30 word pairs. Half the word pairs were selected to correspond to (previous target, current target) pairs used in the training phase. Half the pairs were selected uniformly at random. The first and last words in the word list were not used. The order of each pair was determined at random, but was controlled so that the top word was the word on the left for exactly half the pairs. Because of this, expected chance performance was always 50% correct.

Subjects were given up to four seconds to respond to each pair. If a subject did not respond within the four seconds, the words "TIME UP" came on the screen. The subject was then presented with the next word pair after a one second rest time.

Each trial used a different word list. Each word list consisted of a set of 10 nouns obtained from the MRC psycholinguistic database [12, 22]. The concreteness, familiarity and imagability ratings of the words lied in the ranges 504-604, 452-618 and 501-600, respectively. All words consisted of either one or two syllables, and contained between four and six letters. Each list had at least one and at most three two-syllable words, and either one or two pairs of words starting with the same letter. No list had a triple of words starting with the same letter. To combat use of mnemonics involving the meanings of the words, words with similar meanings were not put into the same list. We also tried to ensure that no two words in the same list fell into a readily recognizable category (for example, "soda" and "vase" would not have gone into the same list, since they both have to do with liquids).

The order in which the word lists were presented to subjects was varied systematically. Four permutations of the eight lists were used, counterbalancing the word lists and interaction types. Two subjects did the experiment with each list order: one with position first and one with velocity first. In all, there were eight conditions, each of which was completed by one subject.

After subjects finished the last round of testing, they were briefly interviewed on the strategies they used to remember the words, as well as which interaction type they preferred.

While running test subjects, we found that a few people created stories or sentences that linked together the words in the list. Subjects who used this strategy would then recite their story to themselves during the testing phase. This strategy was highly effective for those who used it; most of these subjects would consistently perform at or very close to 100% correct. In addition, their performance was probably mostly dependent on how far into the sentence they could recite before timing out; their actual memory of the list was almost always perfect due to the mnemonic. For these reasons, we threw away data from subjects that reported creating stories or phrases, and reran the conditions.

Ten subjects participated in the experiment. We threw away data for two of the subjects and reran their conditions; both these subjects were consistently around or below chance. This left a total of eight subjects, one in each condition. No subjects reported making stories or using other questionable mnemonics. All subjects were right handed except for one. He controlled the mouse with his right hand, and said that he normally controls a computer mouse with his right hand.

Results

Memory performance during the testing task was measured as $\frac{\text{\# pairs with correct response}}{\text{\# pairs subject responded to}}$. Subjects always responded to either all or all but one of the pairs for each testing session, except for one subject, who responded to between 23 and 28 pairs for each trial. Below are the means and standard deviations for the percent correct scores and times subjects took to complete the training task.

Interaction	% correct mean , stdev	training time mean , stdev
Position	0.73 , 0.05	$38.8s$, 4.6s
Velocity	0.78 , 0.04	$75.6s$, 15.6s

An ANOVA analysis on the percent correct scores shows that this result is marginally statistically significant ($F = 5.24,\ p = 0.056$). Eliminating the subject that responded to fewer pairs, we get a more significant result ($F = 7.92,\ p = 0.031$), with percent correct means and standard deviations almost unchanged at 0.72 ± 0.04 and 0.78 ± 0.05.

Discussion

While the velocity control yielded better memory performance, subjects took about twice as long to complete the training task with it. Thus, the improved memory performance might not be due to the interaction at all, but from the fact subjects spent more time exposed to the stimulus. This is backed up by the fact that all subjects reported having an easier time using the positional control. In the second experiment, we controled for this by fixing the amount of time subjects spent in the training task.

Experiment 2

In experiment 2, subjects were given a fixed amount of time in each training session. Two different times were used, 35s and 75s, roughly corresponding to the amount of time subjects took to complete the 15 targets in experiment 1. Each subject alternated between position and velocity control every trial, and alternated between the two times every other trial.

12 right-handed subjects participated in this experiment. Half the subjects started with the positional control, and half with velocity. Within each of these groups, half the subjects started with the 75s condition, and half with the 35s condition. Four word list orderings were used — each ordering was used by three subjects.

Note that because the time was fixed, the number of target words in each trial was variable. Subjects matched as many target words as they could in the given time. Thus, the distribution of word pairs from experiment 1, where half the pairs came from successive targets, was slightly modified. Instead of using the list of all targets, half the pairs were chosen from the first 15 targets. The target words were pre-generated; if a subject failed to get through 15 targets (which was commonly the case), then the first 15 targets that they would have received were used for this purpose. As in experiment 1, the other half of the pairs were generated uniformly at random. This ensured that the distribution of word pairs in experiments 1 and 2 was identical.

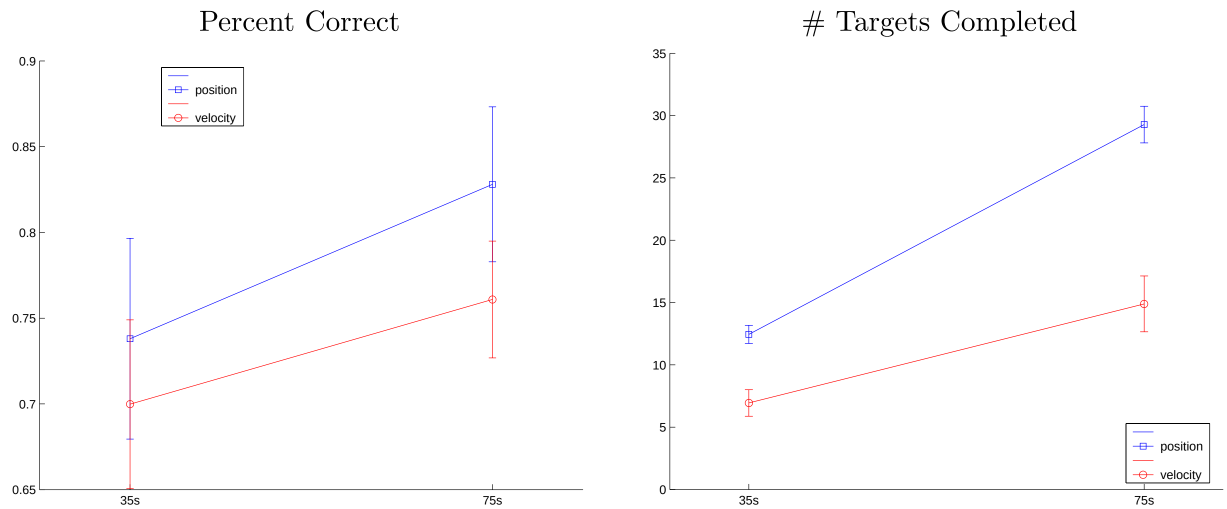

Figure 1: Percent correct and number of targets completed for experiment 2. The position control yielded a higher percent correct score for the 75s case, but almost no improvement for the 35s case. Subjects were able to complete around twice as many targets using the positional control as with the velocity control. Error bars indicate one half standard deviation in each direction.

Results

Two subjects formed sentences out of the list words as a mnemonic device; their data was thrown away. A third subject was also not used — he took much longer to understand the instructions than most participants; he also seemed a little odd performing the tasks, but did not do anything that overtly merited disqualification. Including his data makes the results statistically insignificant ($p > 0.38$ comparing percent correct in the 75s case), but eliminating it produces highly significant result, as shown below. We therefore believe he is an outlier.

Memory performance with the positional control was significantly better than performance with the velocity control. The overall means and standard error for the percent correct scores were 0.784 ± 0.033 for position and 0.733 ± 0.026 for velocity, with $p = 0.02$ in an analysis of variance. Positional control was better in both the 35s and 75s case, though the difference was more pronounced for 75s, and not so distinguishable in the 35s condition (see Figure 1).

The difference in performance between the 35s and 75s times was also significant, with the overall mean percent correct and standard errors 0.717 ± 0.035 for the 35s condition and 0.800 ± 0.027 in the 75s condition, with $p = 0.017$ in an ANOVA (Figure 1).

Subjects were able to complete around twice as many targets during the matching task using the positional control as compared to the velocity control. The overall means and standard errors were 21.031 ± 0.533 versus 11.375 ± 1.073, with $p < 0.001$ (see Figure 1). This is consistent with experiment 1, where subjects took about half as long to complete the matching task for a fixed number of targets using the positional control.

Discussion

This experiment clearly demonstrates that subjects had higher memory performance when using the positional control. In addition, subjects were much faster with this control, as they were in experiment 1.

Interestingly, subjects sometimes moved the words so fast with the positional control that they would be unreadable. This could not happen with the velocity control, since the maximum speed was fixed so that the words were always just readable. Thus, subjects performed better with the positional control even though the words might have been readable for slightly less time than with the velocity.

An argument that could be made is that the positional control was better because subjects were able to complete more targets, and thus saw different parts of the list more frequently. However, in experiment 3, the body control was just as good (and often better) than the hand or head control, even though subjects completed significantly fewer targets with it.

The results of this experiment are compatible with current cognitive load theories. According to [21], positional control imposes a lower cognitive load than velocity when implemented with a mouse[1]. Because of this, users are able to allocate more resources to memory encoding, and thus have higher memory performance.

Users said that they much preferred the positional control, and that the velocity control was harder to operate. Some even said that they had to "concentrate" too much on controlling the list when using the velocity control. This is consistent with the cognitive load explanation.

Experiment 3

In experiment 3, we used three different interaction modes in a virtual environment: a hand-based interaction, a head-based interaction, and a full-body interaction.

With the hand interaction, subjects interacted with the word list using a tracker held in the hand, by pointing at a position on the projection wall. The word list moved according to the horizontal position of this point.

With the head interaction, subjects did the same thing by moving their head. Thus, the gaze direction was (approximately, given that there was no eye-tracking) the "pointing" direction.

With the full-body interaction, subjects had to walk in a 8ft space to control the list.

[1] However, when implemented with a joystick, velocity control is expected to have lower cognitive load. This is because it has a natural zero position and offers resistance when being pushed in a direction. [21]

	Interaction Order					
	123	213	312	132	231	321
List Order						
8693ab75c142	A	B				
a327b58194c6		C	D			
982a6c4b3175			E	F		
25cb3461798a				G	H	
7c93a5142b68					I	J
c1268479a5b3	L					K

Figure 2: The twelve conditions in experiment 3. Word lists are identified by a string of hex digits. Each hex digit is the number of a word list 1-12 (1-c in hexadecimal). The order of digits in the string is the order they were used the experiment. The order of the interactions is given by a permutation of the numbers "123" — 1 is hand, 2 is head and 3 is body. The conditions used in the experiment are labeled A-L.

Method

Each subject performed twelve trials, each of which included the matching task and the testing task. The eight word lists from experiments 1 and 2 were used, plus four additional lists. Subjects rotated through the three interaction types round-robin, performing four trials with each interaction. One third of the subjects started with each interaction technique. Half cycled in the direction hand → head → body, and half did body → head → hand.

Six word list orders were used. Each order was used in two experimental conditions, set up as shown in Figure 2. The word list orders were generated at random, and fed into a randomized algorithm that mutated them to fit the following constraints: (1) Given the conditions in Figure 2, each word list is used with each interaction type an equal number of times; (2) Each word list appears in the last four trials exactly twice; and (3) Each word list appears in the first trial at most once.

Subjects were given 65 seconds in each training phase (the matching task), and 34 pairs in each testing phase. The 34 pairs included four distractor pairs containing the words at either end of the list (described below). The end words were not included in any of the 30 remaining pairs.

We used 65s as the time for the training phase instead of 35s or 75s as in experiment 2 because 75s was too long and produced a ceiling effect (note all the control types were positional). 35s was too short a time for our test subjects — subjects were slower using all three cave controls than with the positional desktop control (see comparison section).

In addition, some subjects in experiment 2 noticed that the words on either end of the list were not used, and some did not. We eliminated this discrepancy for the current experiment by adding distractor pairs and targets. Each end word appeared exactly once in the first 10 targets of the matching task. Subjects were always able to match more than 10 targets, so they always saw the end words. Each testing phase contained four pairs with an end word. The other word in each of these pairs was chosen uniformly at random from the other 9 words in the list. These four pairs were not included in the analysis, since most subjects easily remembered the end words due to primacy/recency effects.

The experiment was implemented in a 4-walled CAVE virtual environment. The stimulus appeared directly on the front wall, with words visible on the entire width of the wall. Nothing was projected onto the right or left walls. The floor was used only to show a mark between trials indicating where to stand. Since the stimulus appeared directly on the front wall, the 3D stereo capability was not used. Subjects did not wear 3D glasses.

Each interaction technique was implemented by intersecting a ray with the front projection wall. The horizontal (x) coordinate of this intersection corresponded to the position of the word list.

The ray used depended on the interaction technique. For the hand interaction, the ray came from a hand-held tracker, and was controlled like a pointing device (e.g. laser pointer). For the head interaction, the ray emanated from a tracker attached to a headband, perpendicular to the subject's forehead. For the body interaction, the ray came from the head tracker, and was perpendicular to the front projection wall — for this interaction, subjects had to use large translational motions, i.e. steps to the left or right. Note that although the implementation of each interaction was done using these rays, the user did not actually see the rays drawn — the only visible stimulus was the word list and target word.

Because the subject's gaze direction had to change while using the the head interaction, the target word moved horizontally for all interactions. The center of the target word was the same as the horizontal projection of the intersection of the ray with the wall.

Before the experiment began, subjects calibrated the head tracker by fixating on a stationary cross, and adjusting the headband so that the cross appeared in the middle of a red circle. The circle's horizontal position was the ray intersection for the head interaction described above. This ensured that the ray from the head tracker was exactly in the gaze direction, and also controlled for eye dominance.

We ran one subject in each condition. 17 total subjects participated, all right-handed. Three subjects used the story mnemonic described in experiment 1, and one subject performed consistently at or below chance. Their data was thrown away and conditions rerun. In addition, the subject for condition E claimed to have a bit of trouble remembering which button to press for each word in the recall task; nothing in the data indicated he was having trouble other than somewhat lower percent correct scores (in particular, he had no more timeouts during testing than normal). This condition was rerun, although the data from the original subject were not thrown away.

Results

There was a fair amount of within-subject variance and outlying data. A small part of this was most likely due to chance orders of the words. For example, one subject reported seeing a word list in almost alphabetical order. To help tighten the data, we removed all trials whose percent correct score was more than 2.5 standard deviations away from the mean for that subject. Standard

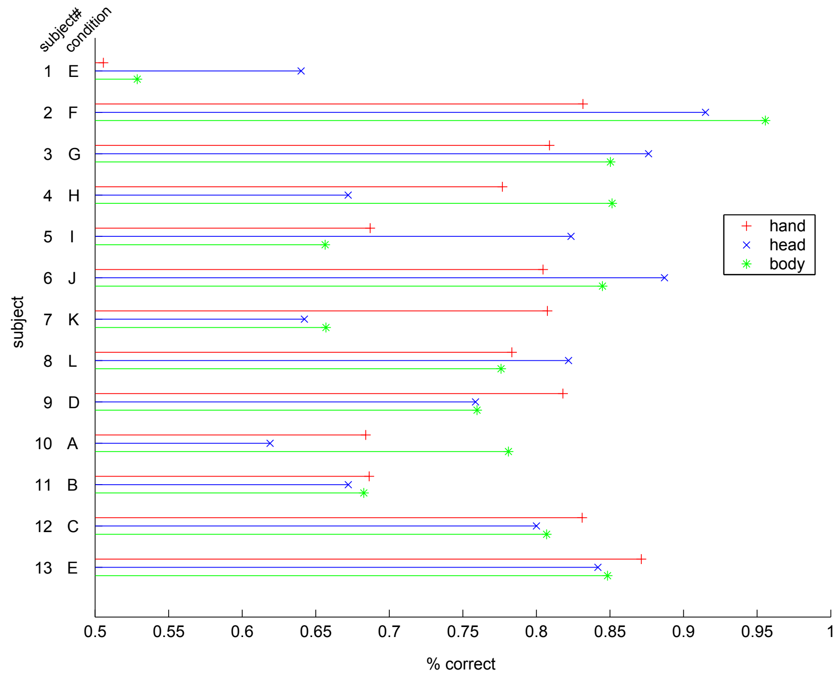

Figure 3: Mean percent correct in recall test for each interaction type, for each subject. Although there was no single interaction that was clearly best, some subjects performed much better with some interactions than others.

deviation and mean for this purpose were calculated using the subject's 10 middle scores — that is, all but the maximum and minimum. This amounted to eliminating at most two trials per subject, for a total of 11 trials (7.6% of data). We also threw away data from all trials where subjects responded to less than 25 of the 30 word pairs — there were three of these. In all, 14 trials were thrown away (9.72% of data); no more than 2 trials were thrown away for any individual subject.

Subjects were not able to complete as many targets with the body interaction as with the hand or head interactions. This was expected — we thought the body interaction would be inherently slower, because subjects had to walk. The mean number of targets completed was 19.615 for hand, 19.474 for head and 16.058 for body; standard errors were 0.545, 0.599 and 0.494, respectively. An analysis of variance shows this is statistically significant ($F = 17.653, p < 0.0005$), with pairwise comparisons between the hand or head and the body control significant ($p < 0.0005$ and $p < 0.001$). The pairwise comparison between hand and head yielded no significant difference. In fact, this comparison suggests that that there is probably no difference between the two ($F = 0.399, p > 0.999$).

There were no significant differences in memory performance between any of the interactions, as measured by percent correct as explained in experiment 1. However, some subjects performed much better with some interaction types than with others. The percent correct scores for each subject are in Figure 3.

In Figure 3, subjects 1, 2, 3, 4, 5, 6 and 10 all did much better with either the body or head control than with the hand control. The body control was better than the hand control for subjects 2, 3,

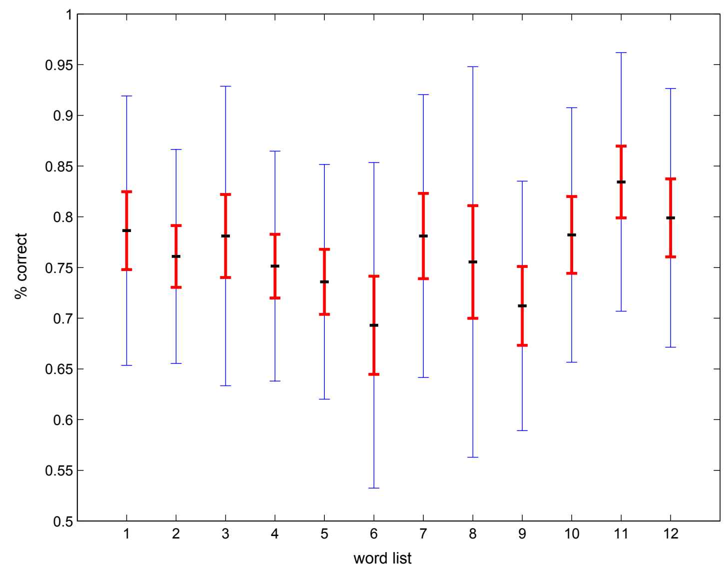

Figure 4: Memory performance for each word list. Subjects performed about equally well for each word list, except possibly lists 6, 9 and 11. Shown are means (black), means $\pm$ standard error (red), and means $\pm$ standard deviation (blue).

4, 6 and 10, and the head control was better for subjects 1, 2, 3, 5 and 6. The hand control seemed best only for subjects 7 and 9. There was not much difference between any of the controls for subjects 8, 11, 12 or 13. This suggests that different interactions may be best for different people, perhaps depending on their encoding strategies or familiarity with computer interaction. This is addressed further in the discussion section below.

Subjects performed about equally in the testing task for each of the 12 word lists (Figure 4). Subjects were slightly worse at lists 6 and 9, and slightly better at list 11. However, this was not statistically significant. We performed a t-test between the samples for each list and those for all other lists. The results for lists 6, 9 and 11 were $p = 0.2548$, 0.3537 and 0.1022, respectively.

In addition, there were no significant differences in the amount of time subjects took to complete the training task, nor in the number of times subjects timed out. There were also no significant differences in the reaction times to word pairs, calculated both including and excluding pairs that timed out.

Discussion

Although there was no statistically significant difference in memory performance for the whole data, many subjects did much better with some interactions than with others. This suggests that there may be a large amount of individual variation, and that different people may do better with different interaction types.

Anecdotal evidence supports this. In the interviews we conducted with subjects after the experiment was over, one subject (subject 3) said she thought she remembered the words better with the body interaction; indeed, this subject did do better with this control (Figure 3). Another subject (subject 11) said she felt “more aware” of were things were using the body control, though there was no noticeable difference in her recall performance.

Subject 4 said she felt the hand control moved the words too fast and that it “hurt her eyes.” Indeed, her performance was better for the body control than for the hand or head controls, which are both faster. The mean number of targets she completed for head, hand and body controls were 18.75, 19.67 and 13.25, with standard deviations of 2.06, 2.52 and 0.96, respectively.

Another subject (subject 7) reported that the hand was “easiest by far,” and thought he had better memory performance with it. This was indeed the case, as shown in Figure 3.

In addition, subjects completed significantly fewer targets with the body control; yet many had better recall performance with it, and most others had around the same performance. This is in contrast to experiment 2, in which subjects performed worse at the recall task when using the slower velocity control.

One possible reason why the body interaction produced the same — or possibly better — performance is that subjects didn’t like to walk or make large movements. All but one subject liked the body control the least out of the three interactions. Common descriptions included the words “slow” or “clumsy”; many subjects said they disliked it because they were “lazy.” In addition, many subjects leaned to the left and right instead of stepping when possible. This suggests that most subjects were adverse to performing large movements, especially steps. Because of this, subjects had more incentive to remember the locations of the words: if they remembered the word locations, they could make fewer searches and thus move less. Thus, subjects that performed better with the body control might have done so *because* they didn’t like the interaction.

In addition, our results provide evidence that the three interaction types impose similar cognitive loads. If any imposed a higher load, subjects would not be able to concurrently encode the list as well, and would therefore perform significantly worse at the recall task. Note it is possible that each interaction imposed a different load for each subject; this is supported by the large amount of between-subject variation discussed above. However, the fact that almost all subjects rated the body control worst and hand best suggests that the degree of difficulty for each of the interactions was fairly consistent between subjects.

Comparison with Experiment 2

Because of changes between the two experiments, the results of experiments 2 and 3 are not readily comparable. The biggest change was the amount of time subjects had to complete the matching task (35s or 75s in experiment 2, but 65s in experiment 3). The results of experiments 1 and 2 indicate a strong speed-accuracy tradeoff, so this change greatly affects the comparability. We also added the end words as distractors in experiment 3; this change is important, but probably not as much as the difference in training time. Still, memory performance was obviously not drastically

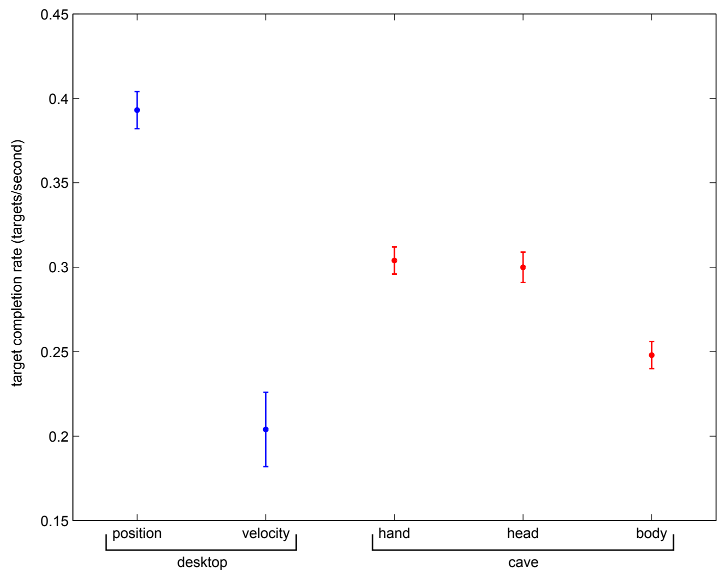

Figure 5: Rates at which subjects lined up targets in experiments 2 and 3. The desktop positional control was best, and velocity control worst. Cave performance was in the middle, with the body control slower than the hand and head. Error bars indicate standard error.

different — mean percent correct scores were around 75-80% in all cases.

However, it is still possible to compare the speed at which subjects were able to align target words. Figure 5 shows this rate for each interaction in the desktop and cave. Only the 75s conditions were used to calculate the rates for the desktop environment; all conditions were used for the cave. Rates were calculated by dividing the mean number of targets completed by the time for the matching task.

All CAVE interactions were significantly slower than the positional desktop control, but faster than the velocity control. As discussed above, the end words were used as targets in the CAVE experiment but not the desktop. However, only two of the targets were end words, out of a mean of 19.7, 19.5 or 16.1 targets for the hand, head or body controls, respectively. Thus, this difference probably did not affect the rates much. Note that if anything, the rates should be slightly faster than they would have been without the endpoint targets. This is because the word list was clamped at the endpoints, so subjects had an easier time keeping words aligned there.

References

[1] Edward Awh, John Jonides, and Patricia A. Reuter-Lorenz. Rehearsal in spatial working memory. *Journal of Experimental Psychology: Human Perception and Performance*, 24(3):780–790, 1998.

[2] Raymond Bruyer and Jean-Christophe Scailquin. The visuospatial sketchpad for mental images: Testing the multicomponent model of working memory. *Acta Psychologica*, 98:17–36,

1998.

[3] Simon Chu, Victoria Handley, and Simon R. Cooper. Eliminating context-dependent forgetting: changing contexts can be as effective as reinstating them. *The Psochological Record*, 53:549–559, 2003.

[4] Ruud Custers and Henk Aarts. On the role of processing goals in evaluative judgements of environments: Effects on memory-judgment relations. *J Environ Psychol*, 23:289–299, 2003.

[5] Bruce Durding, Curtis Becker, and John Gould. Data organization. *Human Factors*, 19(1):1–15, 1977.

[6] Johannes Engelkamp. Visual imagery and enactment of actions in memory. *Br J Psychol*, 86:227–240, 1995.

[7] Florence Gaunet, Manuel Vidal, Andras Kemeny, and Alain Berthoz. Active, passive and snapshot exploration in a virtual environment: influence on scene memory, reorientation and path memory. *Cog Brain Res*, 11:409–420, 2001.

[8] K. James, G. Humphrey, T. Vilis, B. Corrie, R. Baddour, and M. Goodale. "active" and "passive" learning of three-dimensional object structure within an immersive virtual reality environment. *Behavior Research Methods, Instruments, & Computers*, 34(3):383–390, 2002.

[9] Peter Johnson. The functional equivalence of imagery and movement. *Quarterly Journal of Experimental Psychology*, 34A:349–365, 1982.

[10] Kayako Matsuo, Chikako Kato, Tomohisa Okada, Tetsuo Moriya, Gary Glover, and Toshiharu Nakai. Finger movements lighten neural loads in the recognition of ideographic characeters. *Cognitive Brain Research*, 17:263–272, 2003.

[11] George A. Miller. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2):81–97, 1956.

[12] University of Western Australia School of Psychology. Mrc psycholinguistic database (dict interface): `http://www.psy.uwa.edu.au/mrcdatabase/uwa_mrc.htm`, 2004.

[13] Amanda Parker, Angus Gellatly, and Mitchell Waterman. The effect of environmental context manipulation on memory: Dissociation between perceptual and conceptual implicit tests. *Eur J Cog Psychol*, 11(4):555–570, 1999.

[14] Daniel L. Schwartz and Tamara Black. Inferences through imagined actions: Knowing by simulated doing. *J Exp Psychol*, 25(1):116–136, 1999.

[15] R. D. Seidler, J. J. Bloomberg, and G. E. Stelmach. Context-dependent arm pointing adaptation. *Behavioural Brain Research*, 119:155–166, 2001.

[16] Steven M. Smith. Environmental context-dependent memory. In G. Davies and D. Thomson, editors, *Memory in Context: Context in Memory*, chapter 2. John Wiley, 1988.

[17] Hong-Jin Sun, George S. W. Chan, and Jennifer L. Campos. Active navigation and orientation-free spatial representations. *Memory & Cognition*, 32(1):51–71, 2004.

[18] Martin Usoh, Kevin Arthur, Mary C. Whitton, Rui Bastos, Anthony Steed, Mel Slater, Frederick P. Brooks, and Jr. Walking > walking-in-place > flying, in virtual environments. 1999.

[19] David Waller, Daniel R. Montello, Anthony E. Richardson, and Mary Hegarty. Orientation specificity and spatial updating of memories for layouts. *J Exp Psychol: Learning, Memory, and Cognition*, 28(6):1051–1063, 2002.

[20] Richard Wesp and Jennifer Hesse adn Donna Keutmann. Gestures maintain spatial imagery. *Am J Psychol*, 114(4):591–600, 2001.

[21] Christopher D. Wickens and Justin G. Hollands. *Engineering Psychology and Human Performance*. Upper Saddle River, NJ: Prentice Hall, 2000.

[22] Michael Wilson. Mrc machine usable dictionary. version 2.00. Technical report, Oxford Text Archive, Oxford University Computing Service, 1987.

[23] Paul Wilson, Nigel Foreman, and Michael Tlauka. Transfer of spatial information from a virtual to a real environment. *Human Factors*, 39(4):526–531, 1997.

[24] Hubert Zimmer and Johannes Engelkamp. Does motor encoding enhance relational information? *Psychological Research*, 51:158–167, 1989.