In Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society, 2004

The Effect of Structure on Object-Location Memory

Carsten Winkelholz (winkelholz@fgan.de)
Christopher Schlick (schlick@fgan.de)
Mark Brütting (bruetting@fgan.de)

Research Establishment for Applied Science (FGAN)
Research Institute for Communication, Information Processing and Ergonomics
Neuenahrer Strasse 20
53343 Wachtberg, GERMANY

Abstract

This paper aims to study the effect of structure in a graphical layout on object-location memory. In two experiments several structures have been examined in respect to the performance of object-location retrieval. The results show that beside simple object-to-object spatial relations also the spatial relation of three objects is encoded in human spatial memory as a noisy distance-angular pair. Further the results show that noise in spatial memory is not symmetric, but seems to be distorted towards a higher accuracy to the horizontal directions.

Introduction

One aspect of human spatial memory is the usage of allocentric frames of references to encode and retrieve the location of an object. This aspect of human spatial memory implicates that the structure of a graphical layout might affect the performance of object-location encoding and retrieval.

Basically the study presented in this paper is motivated by some experiments performed recently in the community of information visualization. One experiment of Travanti & Lind (2001) investigated object location memory in hierarchical information structures across different instances of 2D and 3D displays. The results of their tests show, that the 3D display improves performance in the spatial memory task they designed. They were aware that their result does not prove their hypothesis that the natural appearance of the 3D display used in the test actually affected the improved performance. They speculated that possibly other visual properties of an item in the 3D display were used as a reminder for the memory task. Cockburn (2004) showed that neither the natural appearance nor the different sizes of the items in the 3D display affected the performance of object-location retrieval. In both studies the memory task was to associate alphanumerical letters to the items. Therefore Cockburn suspected that the vertical orientation of Travanti & Lind's 2D display made the formation of effective letter mnemonics more difficult than the horizontal 3D layout, because words and word combinations normally run horizontally left to right. By analyzing these studies we came to the conclusion that one major factor had not been considered - the factor of the object-to-object spatial relations (the structure of the graphical layout respectively).

The effect of layout structure on object-location encoding and retrieval could best be investigated if a computational model of human spatial cognition is considered. Recently some compelling works toward this goal has been published (Wang et al 2002; Johnson et al 2002). This paper shows one application area for computational models of human spatial memory, but also sheds some new light on the requirements of such a model.

Design of the Experiments

The papers cited above inspired the design of the experiments in this study. There were two phases in the cited experiments. In the encoding phase the subjects had to learn associations of alphanumerical letters to one object in the structure. During the encoding phase a click on one of the objects in the display highlighted the object and revealed a letter at the top of the display, which had to be associated to the position of the object. In the retrieval phase the subjects had to find all of the letters, one at a time. A randomly selected letter had been shown at the top of the display area, and the subject had to click the object associated with it.

This design of the experiment has two drawbacks. First the subjects are free to choose the objects in the encoding phase and second that alphanumerical letters are used as retrieval cues. The first point gives subjects the opportunity to develop strategies for learning the object-letter associations. In combination with the usage of alphanumerical letters this increases the probability that subjects create mnemonics through possible abbreviations of words that can be read from a row.

In respect to a cognitive model these are task specific aspects. The study of this paper was interested in more general mechanisms of object-location encoding/retrieval. To meet this goal the design of the experiment had to prevent subjects from further processing object-locations in the encoding phase. This suggested the task of memorizing a randomly created sequence of highlighted objects from the structure. The number of correct repeated sequences is used as a measure of performance. Furthermore allows this kind of memory task an effective analysis of the errors subjects make.

Two experiments were performed. The first experiment investigated the factor horizontal vs. vertical orientated layout structure and the factor of the existence vs. non-existence of symmetric features in the layout structure. The second experiment focused on the investigation of noise in the encoding of spatial object-to-object relations.

Subjects and Apparatus

30 volunteer subjects (only male, average age 35) were recruited from the staff of our institute to perform both experiments. All subjects had normal or corrected-to-normal vision. Three sets of different structures have been created. Each structure consisted of red spheres of equal size. The layout structures were presented against a black background on a 21'' VGA monitor with a resolution of 1280x1024 pixels. The monitor was in front of the subjects within 2 feet. Subjects were asked to respond by clicking with a mouse. Subjects wore a head-mounted eye-tracking device while they were conducting the experiments.

Experiment 1

The first experiment aimed at showing if the performance of recalling objects is still improved in the horizontal oriented structures, even if in the experimental design no semantic content is used. Further one horizontal structure was added that contains not the symmetric features of the horizontal structure used by Travanti & Lind and Cockburn. Another purpose of this experiment was to show if there is any learning progress in the performance of object-locations encoding/retrieval. It might be possible, that subjects become more familiar with a structure the longer they are exposed to them. In combination with the factor of symmetric features in the structure it might be speculated, that in the presence of symmetric features a subject needs less time to become familiar with the structure.

Materials

Figure 1 shows the three structures that were used in the first experiment. The first two structures are similar to the structures used by Travanti & Lind. Each structure consists of 25 spherical items. The first structure represents a 2D display of a tree-structure, like it is used in most common graphical user interfaces. The second structure represents the structure of the 3D display, where any perspective clues have been removed. The third structure is equivalent to the first one except that it is rotated by 90^0 counterclockwise.

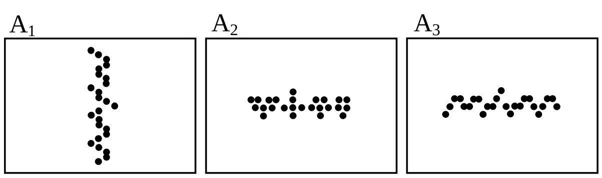


Figure 1: Set of structures used in experiment 1.

Design and Procedure

In each encoding retrieval trial, the subject was presented one structure. After an acoustical signal the computer started to highlight objects of one randomly created sequence. Only one object of the sequence was highlighted at once. The sequences were five items long. The highlighted object differed from the not highlighted objects by color (blue instead of red), increased size and a cross that appeared within its circle shape. The end of a sequence was indicated by an second acoustical signal. Subjects were instructed to repeat the highlighted objects in correct order, by clicking them with the mouse. After five objects had been clicked, another acoustical signal rang out and a short online questionnaire with a subjective rating occurred. Subjects had to rate how confidant they were about their answer and the degree of difficulty to memorize the sequence. The questionnaire was inserted between the tests of two sequences to reduce stress by diversion. Each subject was tested on all structures. The experiments consisted of three blocks. In each block the same structure was tested four times in succession. Between each block there was a break of one minute. Subjects were randomly divided into six groups with five persons, where in each group the order of the three blocks belongs to one of the six possible permutations.

Before the main experiment started, each subject passed through a training run, consisting of two blocks of four sequences. The structures presented in the training run consisted of 16 objects randomly located on the display. The length of the sequences subjects had to learn varied between four and six items.

All sequences for the training run and for the actual test were created randomly only with the property that not the same item occurred in the sequence one behind another. For each subject new random sequences were created. This was done to avoid that for one structure an easy sequence would have been created by chance (e.g. the items of a sequence are only in one row). In general for each structure there might be sequences that are easy to learn, but for some structures these are more likely than for others. And clearly this is a property of a structure that one likes to deduce from its spatial layout. To fix the sequence across subjects would mean that two different factors are controlled. Creating random sequences for each subject means to balance the factor of the sequence among subjects. To fix a sequence across subjects would be interesting to study one specific factor in detail. This was done in parts in the second experiment that is reported in the next section.

Results and Discussion

Accuracy data The number of correct and incorrect repeated sequences for each structure is shown in Table 1.

Table 1: Contingency table (2x3) of correct and erroneous sequences

Structure	A_1	A_2	A_3
Correct seqs.	46	61	63
Erroneous seqs.	74	59	57

The effect of structure approaches significance (2x3 contingency table p = 0.056, χ^2=5.77). When comparing the numbers of correct repeated sequences between each pair of structures with a one-sided analysis of the corresponding 2×2 contingency tables, the exact Fisher test yields that performance in the horizontal oriented structures are significantly higher ($p<0.05$), whereas the symmetric features in the structure did not show any significant effect.

Learning progress Figure 2 shows the development of the performance by each trial in the same structure.

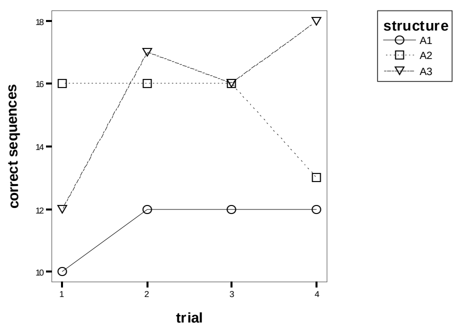


Figure 2: Development of performance in dependence of number of trials.

Structure A_1 with no symmetric features exhibits an increasing performance with each trial, whereas the horizontal oriented structure with symmetric features even shows a decrease in the last trial. The effect of trials on performance is not significant in any structure (2x4 contingency table $\chi 2$ statistical test). Hence, this effect may result from noise in the data.

The most important result of experiment 1 is that it shows that the horizontal oriented structures do improve performance, even if no alphanumerical letters are used as retrieval keys. However, this result may be culture dependent. For example people, who are used to read in columns instead of rows, might be more familiar with horizontal oriented structures.

One culture independent reason for this result might be that the human field of view is more extended into the horizontal direction. This increases noise of allocentric and egocentric memory chunks in vertical directions. If this hypothesis was right, people used to read in columns would profit from a horizontal oriented tree view in two ways: First the horizontal structure would increase performance of object-location retrieval and secondly, inscriptions could be written in columns instead of rows.

Experiment 2

The second experiment aimed at showing how the aspects of human spatial memory, like they are discussed in Wang et al (2001, 2002), affect the performance of object-location encoding/retrieval in dependence on different graphical layout structures. Another purpose of experiment 2 was to collect eye movement data for a more detailed analysis of how subjects encode object-locations.

Materials

The structures used in the second experiment are shown in Figure 3. They are divided into two subsets, because the limited pool of subjects didn't allow testing all permutations needed to prevent order effects.

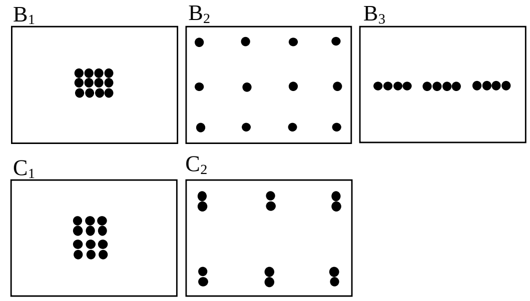


Figure 3: Set of structures used in experiment 2.

Justification The structures in set B and C were created to test some factors assumed to play an important role in the process of object-location encoding/retrieval in structures. The motivation to choose these structures is founded in the assumptions and expectations before the experiments were performed. Mainly the following factors were expected to contribute to the overall performance:

1. Hierarchical features.
2. Noise in the location of an allocentric memory chunk.
3. Noise in the location of an egocentric memory chunk
4. Higher activation of allocentric memory chunks if objects are in spatial vicinity.

The last factor seems plausible, because the effort to assess spatial object-to-object relations is smaller if objects are close together; possibly no eye movement is needed. This last factor would give the spatial narrow matrix B_1 an advantage over the spatial wide matrix B_2 in respect to performance of object-location encoding/retrieval. But also the other factors listed above may contribute. In the linear structure the noise in the memory chunks are more grievous than in the matrices, because there is only one dimension that contributes information, whereas in the case of the matrices also the direction contributes. The tables below show which structure profits by which factor compared to another structure in its set. A + sign in one cell means, that the structure of the row takes an advantage over the structure in the column in respect to the factor of the table, whereas a – sign indicates the opposite The factor of hierarchical features is balanced within each set, so this factor is not included in the tables. (For this purpose the linear structure has been separated into three groups with four objects).

Table 2: Which structure profits by which factor in set B.

Less noise in allocentric memory chunks

	B_1	B_2	B_3
B_1		0	++
B_2	0		++
B_3	--	--	

Less Noise in egocentric memory chunks

	B_1	B_2	B_3
B_1		--	-
B_2	++		+
B_3	+	-	

Higher activation of allocentric memory chunks

	B_1	B_2	B_3
B_1		++	+
B_2	--		-
B_3	-	+	

Table 3: Which structure profits by which factor in set C.

Less noise in allocentric memory chunks

	C_1	C_2
C_1		+
C_2	-	

Less noise in egocentric memory chunks

	C_1	C_2
C_1		-
C_2	+	

Higher activation caused by spatial vicinity

	C_1	C_2
C_1		+
C_2	-	

To estimate the overall performance, the tendencies shown in the tables have to be quantified. Furthermore, not any factor might contribute equally to the overall performance. Without any computational model it can only be speculated about these questions. However, in the setup used in the experiment, it can be assumed that the differences in the noise of the egocentric memory chunks are nearly negligible, because the changes in the average visual angles between the different objects in the scene are small compared to the human field of view. Whereas the directional angular of the allocentric memory chunks possibly covers the whole range. The effect of noise in the allocentric memory chunks in the structure B_1 and B_2 are expected to have an equal effect, because all relative distances are equal. It was expected, that the effect of decrease in performance in the linear structure would be very distinct.

Structure C_1 and C_2 differ only by the distances between the six pairs of objects; the distances between the two objects within a pair are equal. The hypothesis for this structure is that for sequence containing transitions between objects of two far distant pairs it will become more difficult for the subject to encode the location of the object within a pair. This results from a higher noise in the spatial object-to-object relation. To show this effect one predefined sequence was used. This allows analyzing behavior of subjects more efficient. Data from experiments can be used for the parameterization of stochastic models. The regularities found by the algorithms can be analyzed and interpreted (Winkelholz et al., 2003).

Design and Procedure

The experimental design was similar to experiment 1. This time the sequences were six items long. Furthermore, the experiment consisted of two blocks instead of three and in one block each structure from each set was presented once. The first three structures in each block were chosen from set B ordered by one of the possible six permutations. The last two structures in each block were C_1 and C_2, which order again was balanced within groups of subjects.

Except for one sequence in each block all sequences were created randomly for each subject. One sequence for the structures of set C was predefined. Like mentioned above, this was done to be able to analyze experimental tracing data effectively. The sequence was predefined for the structures C_1 and C_2 respectively. The predefined sequence is shown in Figure 4 on the left. It was used in the first block for structure C_1 and in the second block for C_2 or vice versa. By alternating, which structure in the first block starts with the predefined sequence, the effect of remembering the sequence in the second block had been balanced between the structures C_1 and C_2.

Results and Discussion

Accuracy data The numbers of correct repeated sequences are shown in the contingency tables 4 and 5.

Table 4: Contingency table (2x3) of correct and erroneous sequences in set B.

	B_1	B_2	B_3
Correct seqs.	38	34	16
Erroneous seqs.	22	26	44

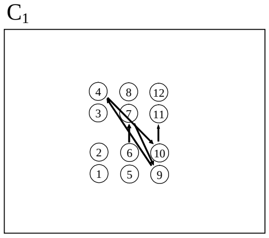
Table 5: Contingency table (2x2) of correct and erroneous sequences in set C.

	C_1	C_2
Correct seqs.	35	25
Erroneous seqs.	25	35

The performance in the linear structure is significantly lower than in the structures of the matrices (exact Fisher-test $p<0.001$). Although the number of correct sequences in the narrow structure is a little bit higher than in the wide matrix, this difference is not significant. In table 5 the number of correct and incorrect sequences from the randomly created sequences and the predefined sequence are combined.

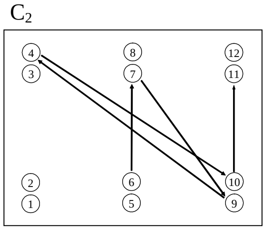
Analysis of errors A look at the errors subjects made in their answers gives more insight into the underlying cognitive processes. To analyze the answer sequences for the predefined sequence in set C we used a modified algorithm for variable length markov chains (VLMC) (Ron et al 1996, Bühlmann & Wyner 1998) to parameterize a stochastic model by the answer sequences. Roughly speaking this algorithm can be seen as a filter for subsequences (called contexts) from the data that contain predictive information. We modified this algorithm in a way that only contexts that contain significant predictive information in a statistical sense are included into the model (Winkelholz et al 2004). To apply this algorithm to the answer sequences the objects in the structure has to be assigned to symbols. The contexts of erroneous behavior found by this method in the answer sequences of the structures C_1 and C_2 are shown in Figure 4. In the first column of the table the contexts found by the algorithm are shown in parenthesis, followed by an arrow, and the most probable next symbols that occur in the answer sequences, if this context is given. E.g. "(7,10)->3", means: If subjects had clicked on object 7 followed by object 10, the most probable object they will click next is object 3. If on the right side of an arrow, more than one symbol/number is listed, they are ordered by their probabilities, with the most probable next symbol first. On the right of an arrow possible next symbols are listed, as far as their frequencies for the given context meet one of the two conditions: First, the frequency is significantly higher than for the symbols with lower frequencies. Second, the frequency does not differ significantly from the frequency of the symbol with the next higher probability.

C_1



Context -> Next Symbol	Interpretation
(4)->10, 11	Omission of object 10.
(6)->7,9	Omission of object 7.
(10) ->11, e	Surprised by the end.
(b)->6,7 (7,6) -> 9 (b, 7) -> 6	Changed order of the first two events in the sequence
(4,9)->11 (9)->4,11	The spatial vicinity, caused subject to mistake 9 for 10.

C_2



Context -> Next Symbol	Interpretation
(9)-> 4,3 (3)->10	The spatial vicinity, caused subject to mistake 3 for 4.
(7) -> 9,10	The spatial vicinity, caused subject to mistake 9 for 10.
(7,10) -> 3	If subject mistake 10 for 9 he also mistake 3 for 4.
(4)->10,11 (4,11)->e	Omission of object 10.

Figure 4: Contexts of erroneous behavior found by the parameterization of a stochastic model. Left: The structure with symbols assigned to the objects and the predefined test sequence. Right: Table with contexts and possible interpretation.

In structure C_2 with the more distant pairs there are more contexts concerning with the confusion of the objects within the pairs of the upper left, and down right corners, whereas for structure C_1 there are more contexts concerning the omission of an object. The most notable context for structure C_2 is "(7,10)->3". The angular between the line from 7 to 10 and the line from 10 to 3 is nearly similar to the angular between the lines 7 to 9 and 9 to 4. Therefore this context indicates that subjects used the relative change in angular direction of two transitions as a reminder.

Eye movement data Currently only the eye movement data of the structures C2 and B2 have been analyzed. Only these two structures exhibit spatial distinct features that allow a reliable assignment of fixations to attended features in the structure. In the structure C2 the fixations were only assigned to one pair. The resolution of the eye tracking device was not sufficient to distinguish between fixations within each pair. For the analysis of the eye movement data in the encoding phase of the predefined test sequence the same method as in the analysis of the errors in the answer sequences was used. The pictures obtained from this procedure are shown in Figure 5. Each picture shows the transitions in the eye movement between the pairs of objects, when the object shown as a filled circle is highlighted. The most probable pairs of objects that will be fixated next if one fixation and the highlighted object is given are presented as arrows starting at the currently fixated pair of objects.

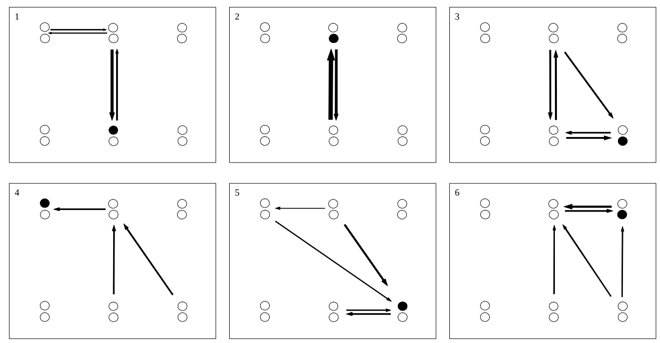


Figure 5: Eye movement data during the encoding phase for the predefined sequence (see figure 4).

The sizes of the arrows are scaled by the frequency of this transition. Although the predefined test sequence does contain two transitions that connect the objects from the upper left corner to the down right corner, there is only one transition in eye movement that connects these pairs directly. Even in the case of a transition from the down right to upper left corner in the test sequence subjects first fixated the group more near to the currently fixated pair of objects (picture 3-4). It was expected that after these transitions in the test sequence occurred, subjects would tend to repeat these transitions by eye movement to create memory chunks for this spatial relation. Instead subjects seem to create spatial relations to the pairs in the middle column. This result becomes more affirmed by taking a look at the eye movement data of the randomized sequences of the structures C_2 and B_2. An overlay of the transitions in the randomized test sequences and the corresponding transitions in the eye movement data are shown in Figure 6.

Although the transitions in the test sequences contain equally transitions between distant objects, these transitions are merely absent in the eye movement data. In both structures most transitions in eye movement are transitions between locations in the vicinity of the two objects in the middle of the screen. In the case of the matrix, movements of the fixation toward objects at the border are very sparse, whereas in the structure with the pairs of objects there are noticeable more fixation movements toward each pair of objects. This also explains the not expected result**,** that there is no significant difference in the performance of the wide and the narrow matrix structure. Possibly, it is sufficient to fixate a location in the middle of the screen to asses most of the spatial object-to-object relations. Moving attention in the visual buffer to repeat transitions is possible without moving fixation. Therefore the effort to repeat transitions of the test sequences in structure B_1 and B_2 are similar. Different in structure C_2; here subjects needed to move fixation to resolve which object within a pair had been highlighted.

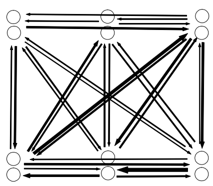
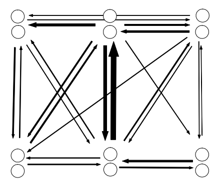
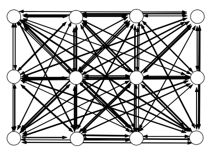
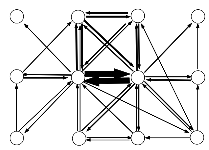
Overlay of all transitions in the randomized test sequences	Overlay of all transitions in eye movements.
	
	

Figure 6: Comparison of transitions in the randomized test sequences with transitions in the eye movement data.

Conclusions and Future Work

The experiments reported in this paper showed how single aspects of human spatial memory affect the overall performance in memorizing tasks of object-locations in layout structures. A computational model that quantifies the interaction of the different aspects of object-location memory is needed to get reliable predictions about the overall performance. The development of such a model within a general architecture of cognition like ACT-R, (Anderson & Lebiere, 1998) enables the implementation of meaningful cognitive models for the application field of information visualization.

The results of the two experiments make the following suggestions with regard to a computational model within the ACT-R architecture:

First, like Wang et al (2002) suggested the model should encode spatial object-to-object relations between the previously and currently attended objects as memory chunks.

Second, also the relation between three objects should be encoded in a memory chunk. In the same fashion as object-to-object relations are encoded this can be done by the visual module whenever attention shifts between three different objects. This memory chunk should be of the form of a noisy angular. Thus the model would show the systematic failures found in the analysis of the answer sequences.

Third, the results from the comparison of the horizontal and vertical oriented structures in the first experiment suggest that noise in the memory chunks of spatial memory is distorted towards a higher accuracy in the horizontal direction. This is a plausible assumption, because the human vision field of view is more extended into the horizontal direction and this should be true for coordinates in all frames of references.

Fourth, eye movement data showed, that subjects need not to gaze at objects they are attending to assess their locations in different frames of references. Therefore it may be disputed, if developers of cognitive models within ACT-R need to control fixation and attention independently. The noise in the assessed object locations should depend on the distance to the current location of fixation.

References

Anderson J. R.; Matessa M. & Lebiere Christian (1998): *The Atomic Components of Thought*. Lawrance Erlbaum Association, Inc. Publishers

Bühlmann, P. & Wyner A.J. (1999). Varaible Length Markov Chains. Annals of Statistics 27 (1), pp. 480-513, 1999.

Cockburn A (2004). Revisiting 2D vs 3D Implications on Spatial Memory. *Proceedings of the Fifth Australasian User Interface Conference* (AUIC2004). Dunedin, New Zealand. January 2004, pages 25-32.

Johnson, T. R., Wang, H., Zhang, J., & Wang, Y. (2002). A Model of Spatio-Temporal Coding of Memory for Multidimensional Stimuli. In W. Gray & C. Schunn (Eds.), *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society* (pp. 506-511). Mahweh, NJ: Lawrence Erlbaum Associates.

Ron, D., Singer, Y. & Tishby, N. (1996): The Power of Amnesia: Learning Probabilistic Automata with Variable Length. *Machine Learning* 25, 2/3, 1996, pp.117-149.

Travanti , M. & Lind M. (2001). 2D vs 3D, Implications on Spatial Memory. In proceedings of the IEEE Symposium on Information Visualization 2001.

Wang, H., Johnson, T. R., Zhang, J., & Wang, Y. (2002). A study of object-location memory. In W. Gray & C. Schunn (Eds.), *Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society* (pp. 920-925). Mahweh, NJ: Lawrence Erlbaum Associates.

Wang, H., Johnson, T. R., & Zhang, J. (2001). The mind's views of space. *Proceedings of the 3rd International Conference of Cognitive Science* (pp. 191-198).

Winkelholz C., Schlick C. & Motz, F. (2003): Validating Cognitive Human Models for Multiple Target Search Tasks with Variable Length Markov Chains. SAE-Paper 2003-01-2219. In: *Proceedings of the 6th SAE Digital Human Modeling Conference*, June 2003, Montreal, Canada.

Winkelholz C.; Schlick Ch. (2004, submitted): *Statistical Variable Length Markov Chains for the Parameterization of Stocastic User-Models from Sparse Data*. 2004 IEEE International Conference on Systems, Man and Cybernetics