# Modeling human spatial memory within a symbolic architecture of cognition

Carsten Winkelholz[1] and Christopher M. Schlick[2]

[1] Research Establishment for Applied Science (FGAN),
Neuenahrer Strasse 20, 53343 Wachtberg, Germany,
winkelholz@fgan.de
[2] Institute of Industrial Engineering and Ergonomics, RWTH Aachen University of Technology, Bergdriesch 27, 52062 Aachen, Germany
c.schlick@iaw.rwth-aachen.de

**Abstract.** This paper presents a study on the integration of spatial cognition into a symbolic theory. The concepts of encoding object-locations in local allocentric reference systems and noisy representations of locations have been integrated into the ACT-R architecture of cognition. The intrinsic reference axis of the local reference systems automatically result from the sequence of attended locations. The first part of the paper describes experiments we performed to test hypotheses on the usage of local allocentric reference systems in the context of object-location memory in graphical layout structures. The second part describes in more detail the theory and its integration into ACT-R. Based on the theory a model has been developed for the task in the experiments. The parameters for the noise in the representation of locations and the parameters for the recall of symbolic memory chunks were set to values in the magnitude quoted in literature. The model satisfyingly reproduces the data from user studies with 30 subjects.

## 1 Introduction

Symbolic theories of cognition are appealing for studies in the field of human-computer interaction. Symbolic theories allow the expression of the cognition process in relation to the task and the visual elements of the interface. One main issue for cognitive modeling in this field is the integration of visual information. Ehret [5] and Anderson et al. [2] describe symbolic models that reproduce learning curves for the location of information on a display. In these examples the underlying mechanism for the learning of locations is the same as for the learning of facts. After some practice the location of specific objects, such as menu buttons, can be retrieved without a time consuming random visual search and encoding of labels. In both studies the locations of visual objects are represented in absolute screen coordinates, and no noise in the representations of the scalar values has been assumed. Accordingly, the success of a retrieval only depends on the number of repetitions of the location of an item. However, the location of an object can only be identified within a frame of reference. In experimental

psychology it is well accepted to divide the frames of references into the following two categories: an egocentric reference system, which specifies the location of an object with respect to the observer, and an environmental (allocentric) reference system, which specifies the location of an object with respect to elements and features of the environment. A good review of the experimental evidence on the usage of these different reference systems in human spatial memory is given by McNamara [9]. This aspect of human spatial memory implicates that the structure of a graphical layout might affect the performance of object-location memory. Object-location memory in the context of graphical layout-structures has already been investigated in the field of information visualization by Tavanti & Lind [13] and Cockburn [4]. These studies showed that different kinds of displays influence performance in object-location retrieval from memory. In both studies the memory task was to associate alphanumerical letters to object-locations. Cockburn suspected that a horizontal oriented layout facilitates the formation of effective letter mnemonics, whereas Tavanti & Lind speculated that a more 'natural' appearance of a visualization enhances object-location memory. Both studies did not consider spatial relations of objects to each other within the structure as a factor. Therefore, we performed our own experiments in which the structure of the object-configurations was varied. In the first part this paper reports the design and results of these experiments. A detailed analysis of the user traces within this experiment suggests that users choose the last two attended locations as a reference axis to which they encode the currently attended location. The second part of the paper describes how we integrated this fact in combination with the concept of noisy location dimensions into the visual module of the symbolic ACT-R [2] architecture of cognition. Based on this extended visual module a symbolic model for the memorizing-task has been formulated. The parameters for noise in the location representation and activation decay of memory chunks have been set to fit the data and are compared to values quoted in literature.

## 2 Experiments

The task of our experiments was to memorize a randomly created sequence of highlighted objects from different structures. The number of correct repeated sequences is used as a measure of performance. This kind of memory task allows an effective analysis of the errors made by subjects. Two experiments were performed. The first experiment investigated the factor horizontal vs. vertical oriented 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. In the following the procedure of the experiment is only sketched. More details can be found in [18]. Thirty 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 were 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.

### 2.1 Experiment 1

The first experiment aimed at showing whether the performance of recalling object-locations is still improved in the horizontal oriented structures, even if no alphanumerical letters are used as retrieval keys.

**Materials** Fig. 1 shows the three structures that were used in the first experiment. 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 is horizontal oriented and exhibits some symmetrical features. The third structure is equivalent to the first one except that it is rotated counterclockwise by 90 degrees.

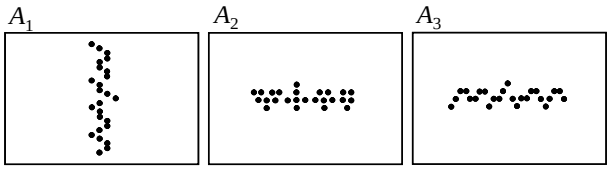



**Fig. 1.** Set of object configurations used in Experiment 1.

**Design and Procedure** In each encoding retrieval trial, the subject was presented with one structure. After an acoustical signal, the computer started to highlight objects of one randomly created sequence. Only one object of the sequence was immediately highlighted. The sequences were five items long. The highlighted object differed from the unhighlighted 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 a 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. After a short break the next sequence was presented to the subject. All sequences were created randomly with the property that no object is highlighted twice in succession. New random sequences were created for each subject in order to avoid the event that an easy sequence was created by chance for any structure (e.g., all objects of a sequence are only in one row). By creating random sequences for each subject the factor of the sequence itself is balanced among subjects. In order to examine a specific factor in detail the establishment of a sequence for all test subjects is of interest. This was done in parts in the second experiment reported below.

**Results** The number of correct and erroneous repeated sequences for each structure is shown in Table 1.

**Table 1.** Contingency table ($2 \times 3$) of correct and erroneous sequences in set $A$ of Experiment 1.

| | $A_1$ | $A_2$ | $A_3$ |
|---|---|---|---|
| Correct seqs | 46 | 61 | 63 |
| Erroneous seqs | 74 | 59 | 57 |

The effect of structure approaches significance ($2 \times 3$ contingency table $p = 0.056$, $\chi^2 = 5.77$). When comparing the numbers of correct repeated sequences between each pair of structures with a one-sided analysis of the corresponding $2 \times 2$ contingency tables, the exact Fisher test yields that performance in the horizontal oriented structures is significantly higher ($p < 0.05$), whereas the symmetric features in the structure did not show any significant effect.

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.

### 2.2 Experiment 2

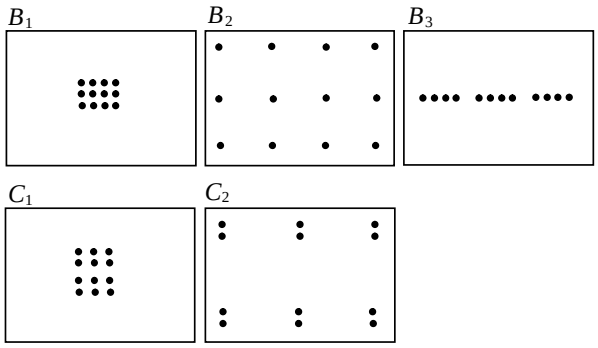



**Fig. 2.** Set of object configurations used in Experiment 2

The second experiment aimed at showing how the usage of local frame of references in human spatial memory, as they are discussed by McNamara [9], affects the performance of object-location encoding/retrieval in dependence on different graphical layout structures.

**Materials** The structures used in the second experiment are shown in Fig. 2. They are divided into two subsets because the limited pool of subjects did not allow the testing of all permutations needed to prevent order effects. 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 from before the experiments were performed. Mainly the following factors were expected to contribute to the overall performance:

1. Hierarchical features,
2. Noise in the allocentric location-representation,
3. Noise in the egocentric location-representation,
4. Higher availability from memory of locations represented in allocentric frames of references 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 narrow matrix $B_1$ an advantage over the wide matrix $B_2$ in respect to performance of object-location encoding/retrieval. The other factors listed above may also contribute. The noise in the location-representation is more grievous in the linear structure than in the matrices since there is only one dimension that contributes information. In the case of the matrices though, direction also contributes. Table 2 shows 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; 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 $B_3$ has been separated into three groups with four objects). To estimate the

**Table 2.** The factors that the structures profit from in the structures of B and C (FOR - frame of reference)

Less noise in allocentric FOR

| | $B_1$ | $B_2$ | $B_3$ |
|---|---|---|---|
| $B_1$ | | 0 | ++ |
| $B_2$ | 0 | | ++ |
| $B_3$ | – – | – – | |

| | $C_1$ | $C_2$ |
|---|---|---|
| $C_1$ | | + |
| $C_2$ | – | |

Less noise in egocentric FOR

| | $B_1$ | $B_2$ | $B_3$ |
|---|---|---|---|
| $B_1$ | | – – | – |
| $B_2$ | ++ | | + |
| $B_3$ | + | – | |

| | $C_1$ | $C_2$ |
|---|---|---|
| $C_1$ | | – |
| $C_2$ | + | |

Higher availability of allocentric FOR

| | $B_1$ | $B_2$ | $B_3$ |
|---|---|---|---|
| $B_1$ | | ++ | + |
| $B_2$ | – – | | – |
| $B_3$ | – | + | |

| | $C_1$ | $C_2$ |
|---|---|---|
| $C_1$ | | + |
| $C_2$ | – | |

overall performance, the tendencies shown in the tables must be quantified. Furthermore, not every factor might contribute equally to the overall performance. Without any computational model as described in Sect. 3 and 4 only speculation can occur about these questions. However, in the setup used in the experiment, it can be assumed that the differences in the noise of the egocentric location-representation 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. This is in contrast to an allocentric location-representation, where the angles take values on the whole range. The effect of noise in the allocentric location-representation in the structure $B_1$ and $B_2$ is 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 transitions within a sequence 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 because the directions do not significantly differ. One predefined sequence was used to show this effect. User traces can be used for the parameterization of stochastic models. The regularities found by the algorithms can be analyzed and interpreted [16].

**Design and Procedure** The experimental design was similar to Experiment 1. This time the sequences were six items long. With one exception all sequences were created randomly for each subject. One sequence for the structures of set C was predefined. As mentioned above, this was done so experimental tracing data could be effectively analyzed. The sequence was predefined for the structures $C_1$ and $C_2$, respectively. The predefined sequence is shown in Fig. 3 on the left. Its usage within the experiments was such that the probability that subjects remembered the sequence from a previous presentation was low. Furthermore, this effect had been balanced between the structures $C_1$ and $C_2$.

**Results - Performance** The numbers of correct repeated sequences are shown in the contingency Table 3. The performance in the linear structure is signifi-

**Table 3.** Contingency table ($2 \times 3$) of correct and erroneous repeated sequences in set $B$ and $C$.

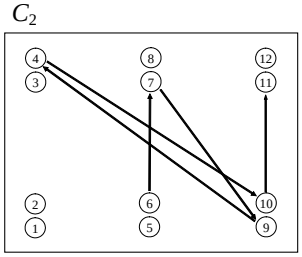
| | $B_1$ | $B_2$ | $B_3$ |
|---|---|---|---|
| Correct seqs | 38 | 34 | 16 |
| Erroneous seqs | 22 | 26 | 44 |

| | $C_1$ | $C_2$ |
|---|---|---|
| Correct seqs | 35 | 25 |
| Erroneous seqs | 25 | 35 |

cantly 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 slightly higher than in the wide matrix, this difference is not significant. For $C_1$ and $C_2$ in Table 3 the number of correct and erroneous sequences from the randomly created sequences and the predefined sequence are combined. According to this contingency table, performance in $C_2$ is significantly lower than in $C_1$ (exact Fisher-test $p < 0.05$).

**Analysis of errors** A look at the errors subjects made in their answers provides 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) [14, 3] 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 such a way that only contexts that contain significant predictive information in a more statistical sense are included into the model [17]. The conventional algorithms do not appropriately consider that sample sizes for different contexts vary in the data. To apply this algorithm




| context → next symbol | interpretation |
|---|---|
| b-6-(7-**9**) → 4,3 | The spatial vicinity, caused subject to mistake 3 for 4 |
| b-(6-**7**) → 9,10 | The spatial vicinity, caused subject to mistake 9 for 10 |
| b-6-(**7-10**) → 3<br>b-6-7-10-(**3**) →10 | **If subject mistakes 10 for 9, he also mistakes 3 for 4.** |
| b-6-7-9-(**4**) →10,11<br>b-6-7-9-(**4-11**) →unif. | Omission of object 10 |

**Fig. 3.** 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.

to the answer sequences the objects in the structure must 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 Fig. 3. In the first column of the table the contexts found by the algorithm are shown in parentheses followed by an arrow and the most probable object occurring next in the answer sequences, if this context is given. E.g. $(7, 10) \rightarrow 3$ means: If subjects had clicked on object 7 followed by object 10, the most probable object they will click next is object 3. Multiple symbols/numbers listed on the right side of an arrow are ordered by their probabilities, with the first in the list being the object that is most probable of being next. On the right of an arrow possible next symbols are listed, as long 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. On the left side of the arrow, also the most probable sequence that lead to the context is given, where the symbol $b$ stands for the beginning of the answer sequence. In structure $C_2$ with the more distant pairs there are more contexts concerned with the confusion of the objects within the pairs of the upper left and lower 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 $(7, 10) \rightarrow 3$. The angle between the line from 7 to 10 and the line from 10 to 3 is similar to the angle between the lines 7 to 9 and 9 to 4. Therefore, this context indicates that subjects used the relative change in the direction of two transitions as a reminder.

### 2.3 Conclusions

The results of these two experiments make the following suggestions with regard to a computational model: first, as Wang et al. [15] suggested, the model should encode spatial object-to-object relations between the previously and currently attended objects as memory chunks. Second, the relation between three objects should also be encoded. As will be discussed in the next section, this can be interpreted since the visual system uses the connection line between two objects as an allocentric reference axis. Third, the results from the comparison of the horizontal and vertical oriented structures in the first experiment suggest that noise in the distance dimensions of spatial memory is distorted towards a higher accuracy in the horizontal direction. Fourth, subjects need not necessarily gaze at objects they are attending in order to assess their locations. During all experiments eye-movement data was collected. Because of a failure of the tracking system the recorded data was very noisy. However, the eye-movement data revealed that in structure $B_2$ subjects tended to fixate on the middle of the screen. Obviously, in $B_2$ it is sufficient to fixate on a location in the middle of the screen to asses most of the spatial object-to-object relations. According to theories of visual attention, moving attention is possible without moving fixation. Therefore, the effort to repeat transitions of the sequences in structure $B_1$ and $B_2$ is similar, yet different in structure $C_2$; here, subjects needed to move fixation to resolve which object within a pair had been highlighted.

## 3 Theory

### 3.1 Rules for the cognitive process

One very popular architecture for cognitive modeling is ACT-R [2]. ACT-R consists of several modules that are controlled by a central production system. These modules are the visual module, memory module, manual module for controlling the hands, and the goal module for keeping track of current goals and intentions. The central production system interacts with these modules by the symbolic content of their specific buffers. The process is described by a set of rules. In each

step of the cognition process one rule that matches the pattern of information in the module buffers is selected. The rule is able to make requests to modules, e.g., to retrieve some information from memory. The retrieved information is loaded into the buffer of the memory and in the next step a new rule applies to the new content of the buffers. The minimal time for one cycle to be completed is 50 ms. The time consumption of a request and the outcome are determined by the design of the module. The working principles of the modules are determined by sub-symbolic mechanisms which may incorporate complex formulas. The most elaborated module of ACT-R is the memory module. The memory is assumed to be a collection of symbolic entities $d^{(i)}$ called chunks. The probability of a successful retrieval and the retrieval time are determined by the activation of the memory chunks, which is calculated by a formula taking into account the time spans in which a chunk has already been retrieved, the strength of its association with the current goal and the similarity of the attributes in a request to the values in the attributes of a memory chunk. This central equation of ACT-R is given by:

$$a_i(t) = b_i(t) + \sum_j w_j s_{ji} + \sum_k u_k m_{ki}. \tag{1}$$

The base activation $b_i(t)$ decays logarithmically over time and increases each time the memory chunk is retrieved. The parameters $s_{ji}$ reflect the frequency of how often chunk $d^{(i)}$ has been retrieved if the symbolic value of attribute $\nu_{jg}$ of the goal was identical to the current value. The parameter $m_{ki}$ is the similarity parameter, and we think could best be interpreted as the log-probability $m_{ki} = \ln(P(\nu_{kx} = \nu_{ki}))$ that the value in attribute $\nu_{kx}$ in the request $x$ is identical to the value in the attribute $\nu_{ki}$ of the chunk $d^{(i)}$. The parameters $w_j$ and $u_k$ are weighting factors reflecting the importance of a single attribute. To decide during a simulation which chunk will be retrieved from memory, noise is added to (1), and the random variable

$$A_i = a_i(t) + X + Y \tag{2}$$

is considered. The random variables $X$ and $Y$ are independent normal distributed with a mean of zero and a variance $\sigma_X, \sigma_Y$. The value of the first random variable $X$ is added when the chunk has been created. And the second one $Y$ is added when $a_i(t)$ is reevaluated. The memory chunk with the highest activation will be retrieved. If the activation of no memory chunk exceeds a threshold $\tau_a$ a failure will be retrieved. Because of the decaying of the base activation a memory chunk will be forgotten unless it is not frequently retrieved. The time needed for a successful recall also depends on the activation by the relation $t \propto e^{-A}$. The higher the activation, the faster a memory chunk can be recalled.

The visual model has recently been added to the theory. The current design of the visual module is specialized in reading and finding objects with specific features on the screen. Visual attention is guided by the commands activated by the selected rule. An object location is represented by its coordinates in pixels on the computer screen. In the following, we will present our approach to adding the concept of noisy spatial relations to the visual module.

### 3.2 Locations and reference systems

As mentioned above, the visual module of ACT-R encodes object-locations in the reference-system of the screen, which is equivalent to creating all spatial object relations to one edge of the screen. The recently proposed extension called ACT-R/S [6] focuses on an egocentric frame of reference. However, according to Mou & McNamara [10] humans also use reference systems concerning the intrinsic axis of the object configuration. E.g., two salient objects create an axis that is used to specify the location of other objects. The most natural way to integrate this into the concept of attention of the visual module is to consider the last two attended objects as an axis of reference. This is an extension to the proposal of Johnson et al. [8, 15] considering only the previously attended object in creating object-to-object relations. This means that the distance is represented in a pure environmental reference system, and the direction in respect to the egocentric perceived up-vector, which is defined by the orientation of the retina in space. In this sense, we call this reference system "semi-allocentric" because the encoded information is not independent from the position and orientation of the observer. However, creating object-location memory chunks in this "semi-allocentric" reference system is less effort for the visual system because it only needs to keep track of two objects, whereas in the case of the pure allocentric reference system, three objects are needed. This point will be discussed in more detail below in the context of visual indices. We considered all three different reference systems that are summarized in Fig. 4. The introduction of object-relations based on three objects is important for three reasons. First, it fits well with the concept of intrinsic axes in the object configuration as reported by Mou & McNamara. Second, the concept of angles is essential to most cognitive operations in geometrical tasks. Third, it is the simplest percept for spatial memory chunks that allows reconstructing object-locations, even if the whole configuration is rotated.

### 3.3 Noise

The variances in pointing errors of recalled object-locations require the dimensions stored in the memory chunks to be noisy. A clear definition of the reference systems is needed to integrate noise into the stored dimensions. Huttenlocher et al. [7] showed that the distribution of pointing errors supports the assumption that subjects imagine object-locations on a plane relative to a center in polar coordinates. We generalized this to use spherical coordinates in respect to an extension of the visual module in three dimensions. This also has some interesting implications on the representation of locations on a screen though, as will be discussed in the following. Spherical coordinates are a system of curvilinear coordinates that are natural for describing positions on a sphere or spheroid. Generally, $\theta$ is defined to be the azimuthal angle in the $xy$-plane from the $x$-axis, $\phi$ to be the polar angle from the $z$-axis and $r$ to be distance (radius) from a point to the origin. In the case of the allocentric reference system this means that if the three points $v_{-2}$, $v_{-1}$, $v_0$ were attended and $v_0$ has to be represented

in a local allocentric reference system, the point $v_{-1}$ defines the origin, the polar axis is given by ($v_{-1}$,$v_{-2}$), and the local spherical y-axis points orthogonal into the screen. For the semi-allocentric reference system, $v_{-1}$ is again the origin, but the polar axis is parallel to the vertical axis of the screen and the $x$-axis is parallel to its horizontal axis. The viewpoint of the subject is the origin in the case of the egocentric reference system. In the typical scenario of a user interacting with symbols on the screen the differences in the angles and distances between symbols represented in the egocentric system are very small compared to the differences if represented in an allocentric or semi-allocentric reference system. Therefore, if the same magnitude of noise is assumed in all reference systems, memory chunks represented in the egocentric reference system would be far more inaccurate compared to object-locations represented in the other two reference systems, and as a result, can nearly be neglected. The next question is, if $\theta$, $\phi$, and $r$ should be considered as single, independent memory chunks. Because it is impossible to imagine a distance without a direction and an angle without corresponding lines, it is reasonable to combine distance and angle as one percept in one memory chunk. This does not mean that the dimensions cannot be separated later. E.g. , it should be possible to extract the $r$-dimension as a distance and apply it to a different direction, as originally perceived. This kind of transformation corresponds to a mental rotation, but these are probably post processing activities of the cognitive system.

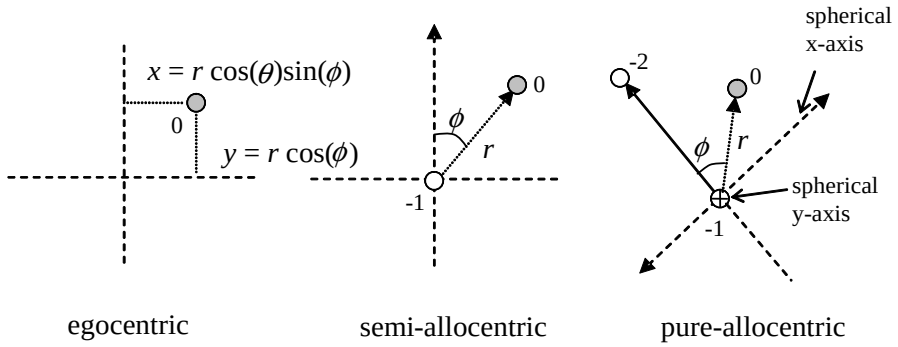



**Fig. 4.** Three different frames of reference, that can be defined according to how many attended object locations are considered. The objects are attended in the order ($v_{-2}$,$v_{-1}$,$v_0$).

For modeling the noise it is assumed that the dimensions of a location are buffered in the neural system and the representation is noisy. A spatial location is represented by $d(r, \theta, \phi, \theta', \phi')$ , where $r, \theta, \phi$ are the spherical coordinates of the pure allocentric reference system and $\theta', \phi'$ additionally hold the polar and azimuth angles in the semi-allocentric reference system. These values are summarized into one symbol representing this relation as one percept. The angles

of the semi-allocentric reference system are integrated into the symbol according to the assumption that we are not able to imagine any angle between two directions without the two directions themselves. In this sense, the symbol of a location encoded into a general allocentric reference system contains both the semi-allocentric and the pure-allocentric reference systems. However, we do distinguish a symbol that encodes only the dimension in the semi-allocentric reference system to reflect that the visual system may only focus on a single direction. Finally, the values of the dimensions are interpreted as the mean values of a noisy neural representation.

$$P(r_x, \theta_x, \phi_x, \theta'_x, \phi'_x | D = d(r, \theta, \phi, \theta', \phi')) = \\ f(r, r_x) f(\theta, \theta_x) f(\phi, \phi_x) f(\theta', \theta'_x) f(\phi', \phi'_x) \tag{3}$$

Where $f(a, a_0)$ is a logistic distribution with mean $a_0$ and variance $\sigma_a$, the noise in the $r$-dimension is biased. We used the logistic distribution for computational reasons. The bias depends on the vertical or horizontal orientation of distance to be estimated. The noise $\sigma_r$ is relative to $r$, which means that it is scale invariant. As the final noise in the $r$-dimension we use:

$$\sigma_r(\phi', r) = (f_{\sigma_r} + (1 - f_{\sigma_r}) \cos^2(\phi')) \sigma_r r \tag{4}$$

If during a simulation the cognitive system requests an object-location based on a symbolic entity $d(r, \theta, \phi, \theta', \phi')$, the values of the dimensions are set to random values $\tilde{d}$ according to the noise given by (3). After the noise has been added to the location request it is decided if the values are latched on possible features in the display. Therefore, the object-locations $x^{(i)}(r, \theta, \phi, \theta', \phi')$ of all features in question are calculated in the current local reference system corresponding to the reference system in the request. The probability $P_{x^{(i)}}$, that visual attention is caught by feature $x^{(i)}$, and the probability $P_0$ that it is not, are given by:

$$P_{x^{(i)}} = \frac{P(\tilde{d}^{(x)} | x^{(i)})}{V^{-1} + \sum_i P(\tilde{d}^{(x)} | x^{(i)})} \qquad P_0 = \frac{V^{-1}}{V^{-1} + \sum_i P(\tilde{d}^{(x)} | x^{(i)})} \tag{5}$$

These equations express the posterior probability $P(x^{(i)} | d^{(x)})$ that if a noisy location $d^{(x)}$ from the neural representation in memory is given the location results from the feature $x^{(i)}$. The parameter $V^{-1}$ describes the weight of a noisy background. The likelihood functions $P(d^{(x)} | x^{(i)})$ are the truncated logistic distribution as if the feature $x^{(i)}$ would have been the stimulus and are similar to (3). The process of the projection of a noisy neural location representation from memory onto a new percept is illustrated in Fig. 5. The most rational choice would be the feature with the maximum probability according to (5). However, we assume this decision to be noisy as well, so the visual system maps the request on the features with a probability given by (5). A similar mapping scheme can be applied to map a perception to already existing memory chunks. Whenever a location has been encoded into the symbolic entity $d$ it will be stored in memory and the retrieval is determined by the memory module and equation (1). This noise model has two interesting properties. First, because the truncated

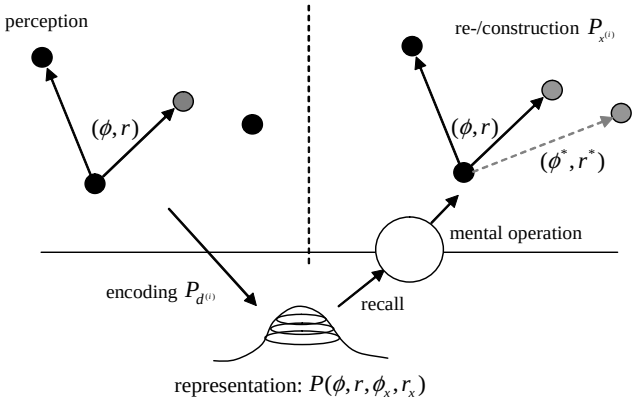



**Fig. 5.** Perception, representation and re-/construction of a location.

logistic distribution is asymmetric, the expected report of an object-location is biased away from the reference axis. This is the same effect as reported at categorical boundaries by Huttenlocher et al. [7]. Second, for object-locations on a flat screen the values of $\theta$ are discrete $\theta = \{\pi/2, 0, -\pi/2\}$ and encode whether the object-location in question is on the left side or the right side, or aligned, when facing into the direction of the reference axis. This is consistent with the assumption of interpreting the reference axis as a categorical boundary, where $\theta$ encodes the category. Thus, the categorical boundaries of Huttenlocher et al. are simply projected egocentric reference axes.

### 3.4 Visual indices

It is evident that humans browsing a graphical structure encode environmental characteristics of object-locations, say the spatial relations to objects nearby. The crucial point in encoding such environmental features is that after some objects in the environment have been attended, attention needs to return to a specific location previously attended. If this return depends on such noisy operations as those so far described, the cognitive system will hardly return to a specific reference point. At this point the concept of visual indexing, or FINST - FINger INSTantiation [11], is needed. According to this theory, the cognitive system has *"access to places in the visual field at which some visual features are located, without assuming an explicit encoding of the location within some coordinate system, nor an encoding of the feature type"*. Experiments show that the number of FINSTs in the visual system is limited to four or five. To implement environmental scan patterns, FINSTs need to enable the visual system to access previously attended locations without or at least with minimal noise. In the visual module of ACT-R the concept of FINSTs is currently only used to determine if a location has already been attended, but it gives no direct access

to such an indexed location. In our simulations we gave the cognitive system direct access to an indexed location by determining the visual index through the sequential position in the chain of attended locations. This index can be used to return (or avoid returning) attention to a particular location in the chain of attended locations.

### 3.5 Competitive chunking

As a human subject learns object-locations in a graphical structure he/she becomes familiar with the structure after some time. This means that he/she recognizes environmental features faster and is able to link environmental features more efficiently to object-locations. This implicit learning is similar to the effect that subjects are able to learn sequences of letters more efficiently if the sequences contain well-known words or syllables. Servan-Schreiber & Anderson [12] discussed this effect in the context of a symbolic theory as a competitive chunking (CC) mechanism. According to the theory of CC, a memory chunk for the sequence can be compressed by replacing elements of the sequence by references to subchunks having a high activation, and therefore can be retrieved quickly and reliably from memory. Memory chunks need to be declared in ACT-R in advance. Subsequently, the mechanism of emerging subchunks is not part of ACT-R. To investigate such a mechanism in the context of object-location memory we extended the formulae 1 for the activation of chunks in memory by a term $a_i^{CC}$, calculating correlations across attributes of chunks in memory, which results in virtual sub-chunks supporting their container chunks. We derived this term heuristically and it is given by:

$$a_i^{CC} = a_i + c_{CC} \sum_{m=1}^{n_s} \sum_{n=1}^{n_s} \sum_{k=1}^{N} I_{mni} K_{mnik} \ln(1 + e^{c_d b_k}) \tag{6}$$

The index $k$ runs through the chunks of the same kind; the index $m$ and $n$ through the slots of the chunk type. The parameter $K_{mnik}$ compares the similarity of the slot values and is given by:

$$K_{mnik} = \begin{cases} P(\nu_{mk} = \nu_{mi}) P(\nu_{nk} = \nu_{ni}) & \text{, if } P(\nu_{mk} = \nu_{mi}) P(\nu_{nk} = \nu_{ni}) > c_\tau \\ 0 & \text{, otherwise} \end{cases} \tag{7}$$

Hence, $K_{mnik}$ is the probability that both values are equal. The parameters $P(\nu_{ab} = \nu_{cd}) = p_{abcd}$ express the probability that the values $\nu_{ab}$ and $\nu_{cd}$ both result from the same source. For scalar values they can be calculated by a Bayesian approach. To limit the contributions, $K_{mnik}$ is cut by a threshold $c_\tau$. Thus, roughly speaking, the sum of the $K_{mnik}$ over the slot pairs is a measure of how many equal slot values chunks $d^{(i)}$ and $d^{(k)}$ share. If only $K_{mnik}$ is used as a factor for the CC, slots also contribute, whose values are equal over all chunks. This means that they do not carry any information. Therefore, we introduced the factor $I_{mni}$ that estimates how much normalized information the knowledge

of the value $V_m = \nu_{mi}$ in attribute $m$ of memory chunk $d^{(i)}$ contains about the values $V_n$ in attribute $n$ of the other chunks.

$$I_{mni} = 1 - \frac{H(V_n|V_m = \nu_{mi})}{H(V_n)} \tag{8}$$

If $\nu_{mi}$ contains no information about $V_n$, $I_{mni}$ is zero. If $V_m$ is fully determined by the knowledge of $\nu_{mi}$ then $I_{mni}$ is 1. If the slots only contained clearly distinguishable symbolic values, the entropies in (8) could be calculated by their frequencies. In the case of spatial memory chunks though, the similarities have to be taken into account. With the abbreviation $P(\nu_{ab} = \nu_{cd}) = p_{abcd}$ the entropies can be estimated by

$$H(V_n) = -\frac{1}{N}\sum_{k=1}^{N} \ln \frac{\sum_{k'}^{N} p_{nknk'}}{N} \tag{9}$$

and

$$H(V_n|V_m = \nu_{mi}) = -\frac{1}{\sum_{k'}^{N} p_{mimk'}} \sum_{k=1}^{N} p_{mimk} \ln \frac{\sum_{k'}^{N} p_{nknk'} p_{mimk'}}{\sum_{k'}^{N} p_{mimk'}} \tag{10}$$

In the limit $P(\nu_{nk} = \nu_{nk\prime}) = \delta(\nu_{nk}, \nu_{nk\prime})$ of clearly distinguishable slot values the equations (9) and (10) are identical to a formula estimating the probabilities of the entropies for the information by the frequencies of the slot values. Furthermore, the contribution of each chunk is weighted by a factor according to its basis activation $b_k$ with a lower bound to zero and approximating $b_k$ for large activations. Due to the additional term (6) in the activation equation, virtual subchunks emerge through the clustering of attribute values, which support their container chunks (Fig. 6).

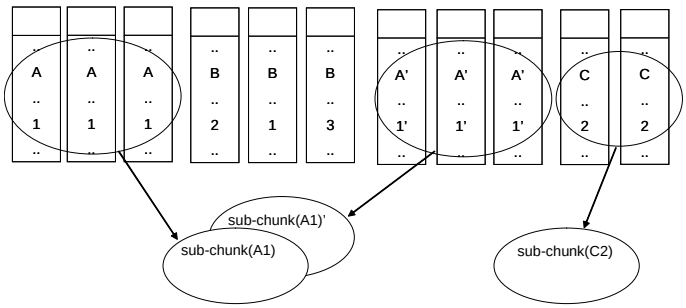



**Fig. 6.** Virtual sub-chunks from correlations in attributes across parent chunks. The rectangles symbolize the parent chunks with their attributes.

# 4 Simulation

## 4.1 The model

The model we developed is based on the theory sketched in the previous section. It is similar to a conventional model of list memory [1] and it describes the encoding and retrieval stage of the memorizing task. The previously highlighted objects up to the current location are mentally rehearsed during the encoding of the sequentially presented object-locations. This rehearsal serves to boost the activation of the corresponding memory chunks so they can be recalled reliably later on. Different to common models of list memory, the model encodes environmental-features of the object-locations (say, spatial relations to objects in the vicinity) during the rehearsal or checks if one of the objects retrieved to the environmental feature matches the environment of the currently attended object. If the environment does not match, the reference system is restored through the visual indices, and a new guess is made excluding the denied object-location. The environmental features are encoded in competing chunks with a symbolic tag to the corresponding object-location and spatial relations to objects in its neighborhood. To check an environmental feature is time consuming because it must be retrieved from memory. Therefore, the cognitive system needs to find a tradeoff between the loss of time in the rehearsal and the reliability of the location of a rehearsed object. In ACT-R it is possible to let different rules compete. Which rule will be selected is determined stochastically proportional to parameters reflecting the probability that the selection of each rule has induced a feeling of success in the past. These parameters can be learned during the simulation. In our simulation one rule for skipping and one rule for actually performing the validation of a location by an environmental feature compete. The answer stage is equal to the rehearsal stage, except that environmental features are not encoded anymore and are validated for every location because time pressure is no longer present. Overall, the model contains 142 rules. This unexpected high number of rules results from the time pressure set on the task. At any possible stage the model needs to check if a new object is highlighted, which leads to a lot of exceptions. The ACT-R parameters for retrieval of memory chunks were set to the defaults reported in literature [2]. The variance $\sigma_{(\theta,\phi)}$ of the noise for the angular dimension was set to 0.06 radians and to 0.08 for the $r$-dimension. The skewing factor $f_r$ in (4) of the noise in the $r$-dimension was chosen to be 0.8. The parameter for the background noise was set to $V = 2.e3$. The same parameters and rules were used for all simulations and graphical structures.

## 4.2 Validation of the model

We simulated the experiments with 30 subjects ten times and compared the mean values to the data. The model satisfyingly reproduces the overall performance (Fig. 7) of the subjects ($R^2 = 0.83$), though there is one disturbing discrepancy: the model exhibits a different performance for structure $A_2$ and $A_3$, yet a statistical test reports no difference for the data. The structures $A_2$ and

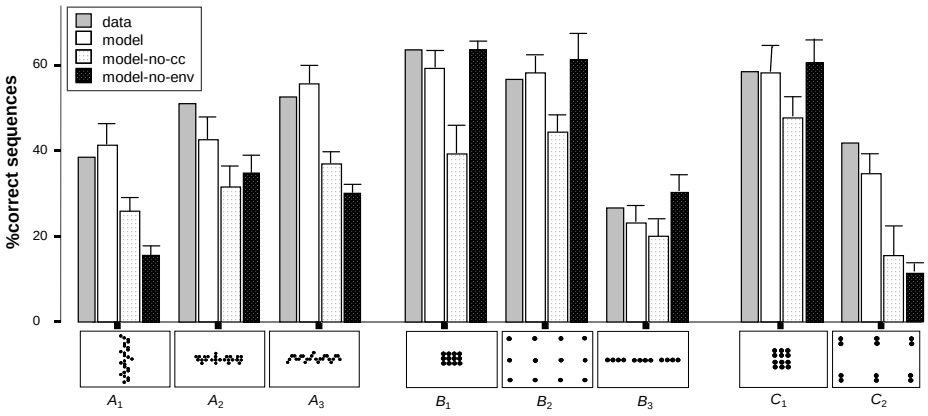



**Fig. 7.** Performance of different models compared to the data (model: model with encoding of environmental features and competitive chunking [CC] mechanism enabled, model-no-cc: model with encoding of environmental features, but without the CC. model-no-env: model without encoding environmental features).

$A_3$ are both horizontally oriented, but $A_2$ is more regular than $A_3$ and contains some symmetrical features. Therefore, we expected that in $A_2$ subjects become familiar with the structure earlier by the CC mechanism. On the other hand, the environmental features are more distinct in the less symmetric structures and therefore provide more information for the validation of an object in the sequence. As reported in Sect. 2.1, the analysis of the learning curves of the subjects demonstrated these effects as a tendency, but could not be significantly demonstrated. In the overall performance these two effects should have been balanced to result in the observed error rate. Obviously, this is not the case. We primarily integrated the encoding and validation of environmental characteristics into the model in order to investigate the effect of becoming familiar with the structure. The model, however, did not show these effects either (see Fig. 8a). In the end it seems that the observed data could have been explained better without considering the validation of objects by their environmental characteristics. To examine the contribution of the encoding of environmental characteristics and CC we made simulation runs without CC and without the encoding of environmental features. The results are displayed in Fig. 7 and show that the encoding of environmental features does not have the same effect for all structures. Interestingly, the performance of the model not encoding environmental features is similar to the complete model for the structures $B_1$, $B_2$, and $C_1$. Performance is even lower for these structures if environmental features are encoded without CC. However, for the structures $A_1$, $A_3$, and $C_2$ the model seems to significantly benefit from the encoding of environmental features. This effect can be explained by the observation that during the encoding of environmental features the model loses a lot of time, especially if activation of memory chunks for environmental features are low. In this case the sequences cannot be rehearsed so

often to boost activation. If spatial relations of preceding objects in the sequence are as distinct as in $B_1$, $B_2$, and $C_1$, the advantage of encoding environmental features clearly does not compensate the disadvantage of time loss. Only if CC is assumed is the activation of environmental features in memory high enough to make the encoding of environmental features paying. For structures $A_1$, $A_3$, and $C_2$ the cognitive model benefits from encoding environmental features, even if no CC supports their activation. For the unsymmetrical structures $A_1$ and $A_3$ the environmental features are more distinct and therefore more effective for the validation of object-locations. Altogether, the attempt to fit the model without encoding of environmental features to the data required lowering the variances in the angular dimensions to 0.01 radians, which is only approximately 15% of the variance assumed for the model with encoding of environmental features, and therefore much lower than reported in literature [7]. Accordingly, the consideration of strategies to validate object-locations are mandatory. Unfortunately, the model encoding environmental features became very complex, and the new integrated mechanism, such as the CC, should be further evaluated. Regardless, the complete model reproduces some other effects that may result from encoding and validating environmental features. This is displayed in Fig. 8b. Here the number

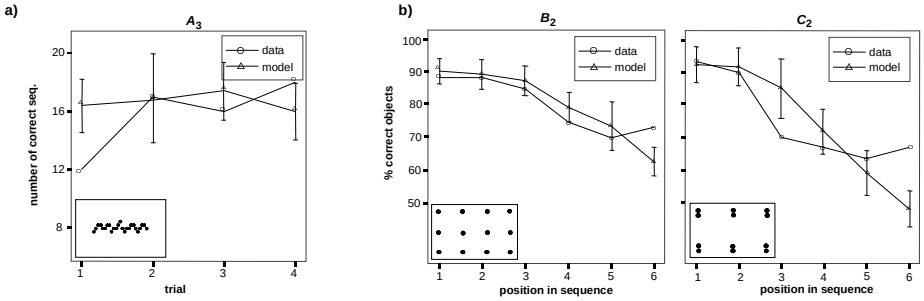



**Fig. 8.** Performance of subjects and model. Left: Overall performance. Right: Dependency of the performance from the position in the sequence.

of correct repeated objects is plotted over the position of the object in the sequence for two structures. Both curves show a plateau at the beginning up to the third object. In the model this plateau results from the environmental features. Because one object of the sequence is mainly learned by the spatial relation to its predecessor, each correct reconstructed object of the sequence strongly depends on a successful reconstruction of its predecessor. As a result, the cognitive system focuses on the first objects in the rehearsal to boost the activation of the corresponding memory chunks for a reliable recall. An analysis of the model shows that the stronger decline after the second and third object in the curve of the model results from an unreliable recall of environmental features rather

than a failed recall of the primary spatial relation. The curves of the subjects show an additional plateau at the end of the sequences. The error rate at the end of the sequence is lower than for the whole sequence. This means that some objects of the sequence have been reconstructed based only on the environmental characteristics here. This hypothesis is supported by the observation that this plateau is more distinct for structure $C_2$. The model only validates an object by the environmental features and repeats applying the spatial relation as long as the validation fails. It contains no rules for a strategy for skipping one object of the sequence and trying to identify the next object only by its environmental features. Therefore, the model only reproduces the plateau at the beginning. Nevertheless, when gathering the measuring points of all curves for all structures together the model achieves a correlation to the data with $R^2 = 0.92$. Hence, the performance of the model is generally satisfying. Additionally, the parameters for the variances in the location representation are in the magnitude reported by Huttenlocher et al. [7]. Huttenlocher et al. reported $\sigma_\phi = 0.17$ (the model: 0.06) and $\sigma_r = 0.025$ (the model: 0.08). The higher variance in the angular dimension reported by Huttenlocher et al. may be explained by a missing reference point in their experimental design. Subjects had to infer the reference point as the center of a circle which increased uncertainty. Subjects might encode the distance as a distance to opposite reference points on the border of the circle. This might decrease the overall measured variance in the distance of the answers. In the model, the parameter for the variance refers to one single representation. If strategies of subjects are assumed using multiple representation, the overall variance in the answers is decreased as well.

## 5 Conclusions

This paper described extensions to the visual-module of the ACT-R theory which enable the development of very detailed models for the visual working memory. The concepts were derived from well-known effects in experimental psychology. Overall, the modeling gave us a deeper insight into the mechanisms and bottlenecks of encoding object-locations. One challenge in modeling the memorizing task was the limited number of FINSTs. The number of FINSTs limits the complexity of environmental features that can be encoded. This is interesting with respect to visual working memory in three dimensions. Encoding of an object-location in a real allocentric local reference system in three dimensions needs at least three object-locations to define a reference plane. This reduces the number of free FINSTs in an encoding task. This might explain why spatial reasoning in three dimensions is more difficult for most people than spatial reasoning in two dimensions. In future work we will extend the concepts described in this paper to three dimensions. The future work will elaborate the chosen Bayesian approach and will try to develop an integrated model of human cognition when interacting with graph-based structures. Furthermore, we are aiming at additional validation studies based on very simple geometrical tasks without time pressure.

## References


1. Anderson, J. R., Bothell, D., Lebiere, C., Matessa, M.: An integrated theory of list memory. Journal of Memory and Language 38 (1998) 341-380
2. Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., Qin, Y.: An integrated theory of the mind. Psychological Review 111 (4) (2004) 1036-1060
3. Bühlmann, P., Wyner, A. J.: Varaible Length Markov Chains. Annals of Statistics 27 (1) (1999) 480-513
4. Cockburn, A.: Evaluating spatial memory in two and three dimensions. International Journal of Human-Computer Studies 61 (3) (2004) 359-373
5. Ehret, B. D.: Learning where to look: Location learning in graphical user interfaces. CHI Letters, 4(1) (2002) 211-218
6. Harrison, A. M., Shunn, C. D.: ACT-R/S: Look Ma, No "cognitive-map"!, In International Conference on Cognitive Modeling, (2003) 129-134
7. Huttenlocher, J., Hedges, L. V., Duncan, S.: Categories and Particulars: Prototype Effects in Estimating Spatial Location. Psychological Review Vol. 98 No. 3 (1991) 352-376
8. Johnson, T. R., Wang, H., Zhang, J., Wang, Y.: A Model of Spatio-Temporal Coding of Memory for Multidimensional Stimuli. In: Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society Mahweh, NJ: Lawrence Erlbaum Associates (2002) 506-511
9. McNamara, T. P.: How are the locations of objects in the environment represented in memory? In: Freksa C, Brauer W, Habel C and Wender K (Eds.) Spatial cognition III: Routes and navigation, human memory and learning, spatial representation and spatial reasoning. Springer Berlin (2003) 174-191
10. Mou, W., McNamara, T. P. Intrinsic frames of reference in spatial memory. Journal of Experimental Psychology: Learning, Memory, and Cognition 28 (2002) 162-170
11. Pylyshyn, Z. W.: The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. Cognition 32 (1989) 65-97
12. Servan-Schreiber, E., Anderson, J. R.: Chunking as a mechanism of implicit learning. Journal of Experimental Psychology: Learning, Memory, and Cognition 16 (1990) 592-608
13. Tavanti, M., Lind, M.: 2D vs. 3D, Implications on Spatial Memory. In: Proceedings of the IEEE Symposium on Information Visualization (2001) 139-145
14. Ron, D., Singer, Y., Tishby, N.: The Power of Amnesia: Learning Probabilistic Automata with Variable Length. Machine Learning 25 2/3 (1996) 117-149
15. Wang, H., Johnson, T. R., Zhang, J., Wang, Y.: A study of object-location memory. In: Gray W. and Schunn C. (Eds.). In: Proceedings of the Twenty-Fourth Annual Conference of the Cognitive Science Society. Lawrence Erlbaum Associates Mahweh NJ (2002) 920-925
16. Winkelholz, C., Schlick, C., Motz, F.: 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 Montreal Canada (2003)
17. Winkelholz, C., Schlick, C.: Statistical Variable Length Markov Chains for the Parameterization of Stocastic User-Models from Sparse Data. In: Proceedings of IEEE International Conference on Systems, Man and Cybernetics (2004)
18. Winkelholz, C., Schlick, C., Brütting, M.: The effect of structure on object-location memory. In: Proceedings of the Twenty-Sixth Annual Conference of the Cognitive Science Society Mahweh, NJ: Lawrence Erlbaum Associates (2004) 1458-1463