# Attending to and maintaining hierarchical objects in graphics comprehension

Atsushi Shimojima[1], Yasuhiro Katagiri[2], and Megumi Enseki[1]

[1] Doshisha University, 1-3 Tatara-Miyakodani, Kyotanabe, Kyoto, Japan
[2] Future University Hakodate, 116-2 Kamedanakano, Hakodate, Hokkaido, Japan

**Abstract.** Information graphics convey different levels of information, depending on how their elements are grouped into different units of objects. How people set boundaries to graphical objects to be interpreted and how they maintain the object boundaries during the given task are two important problems in understanding the way people comprehend information graphics. Table comprehension process was experimentally investigated in terms of eye gaze control behaviors when people were required to read off information distributed over multiple cells, e.g., a row or a column, of an alphanumerical table. Eye movement asymmetry was observed between vertical and horizontal integration of information. It was also observed that both grid lines orthogonal to the direction of integration and larger inter-cell spacings increase the number of corrective eye movements. We provide an explication of these observations applying recent findings about object-based attention and visual indexing. We emphasize the significance of embodied pictures on human graphics comprehension supported by close interactions between perceptual and cognitive processes.

**Key words:** graphics comprehension, object-based attention, visual index, eye-tracking, embodied cognition

## 1 Introduction

Researchers and designers have noted that information graphics can express "higher-level" information as well as "lower-level" information [2, 3, 36, 10]. The former roughly indicates more abstract information carried by overall patterns formed by multiple graphical elements, while the latter more concrete information carried by individual graphical elements. Designers of information graphics often call their higher-level meanings the "main messages," while Bertin [3] went so far as to state that the communication of higher-level information is the "purpose" of preparing information graphics.

For example, the location of individual dots in Figure 1 indicate the existence of individual data points with specific values. In addition to this lower-level information, the scatter plot expresses higher level-information by "the shape and the density of the cloud" formed by these dots [13]. While the lower-level information is concerned with the values taken by individual samples in the data, the higher-level information is concerned with the overall distribution of the data, such as the strength of correlation between the two variables.

This paper is concerned with the fundamental operations necessary to extract different levels of information from the given graphics. The approach is defined by two leading questions, so we will begin with formulating them in detail (section 2). We will

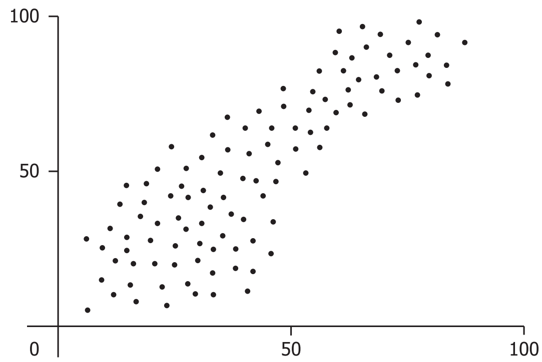



**Fig. 1.** A typical scatter plot

then apply the obtained perspectives to analyze the results from an exploratory study of people engaged in table-reading tasks (section 3). The tasks are designed so that the levels of information to be extracted from the given table may be apparent. Te eye-tracking data, analyzed in the proposed approach, offer a profoundly integrative view of graphics comprehension process, where cognitive process can frequently refer back to perceptual process through active use of saccadic eye-movements (section 4).

## 2 Problem

The distinction of higher-level and lower-level information is quite general over different kinds of information graphics. Just as clouds of dots in a scatter plot can have informative shapes, line segments in a line graph can form an informative slope or curve by connecting individual points on a plane that carry lower-level information. Likewise, when the bars in a bar chart has the shape of "descending staircase" [22], this can mean that the price of a product steadily declined during the period. The information is clearly distinguishable from information carried by the hight of individual bars.

Although less frequently, data maps have been also cited as carriers of multiple-level information. For example, Lowe [18] discussed "secondary structure", where adjacent isobars on a meteorological map together indicate a global trend of the area's barometric situation. Gilhooly et al. [9] found the use of "specialist schemata" in geographers' reading of contour maps, where visual patterns formed by several contour lines indicate some global structures in the area, such as valleys and interlocking spurs.

Node-edge graphs and even tables support derivative meanings. Olivier [20] discussed the case of tree diagrams, where an extended path formed by consecutive edges indicates the presence of a descent or chain in the represented relational structure. In London's tube map, the concentration of edges touching a node indicates the presence of a "hub" station [30]. Many tables are designed to allow the viewer to do "column-wise" or "row-wise" readings, in addition to basic "cell-wise" readings [30].

As the notion of "higher-level information" is applicable to such a wide variety of cases, one may suspect that it might be without content. Shimojima [30] investigated how certain graphics come to carry higher-level information, and identified a general pattern in which additional semantic rules are logically derived from basic semantic conventions in a graphical representation system. Hogher-level and lower-level information are respectively based on the derived semantic rules and the basic semantic rules. Thus, the level difference of information expressed by graphical representations is not just the matter of our subjective judgment, but susceptible to exact semantic characterization.

Nevertheless, whether a reader can appreciate different levels of information expressed in the given graphical representation is the matter of exact cognitive operations involved in the comprehension process. Specifically, which level of information people extract depends on the following factors:

**(1)** how people set boundaries to graphical objects to be interpreted.
**(2)** (when the reading task is complex enough) how they maintain object boundaries during the task.

For example, in order to extract specific values of individual data points from a scatter plot in Figure 1, one need only interpret the locations of individual dots, whereas one need take the entire cloud of dots as a coherent object in order to evaluate the overall trend of the data. In addition, when the task is to compare the strength of correlation in two data sets (say, the data in the $x$-range 0–50 and those in the x-range 50–100 in Figure 1), one need somehow maintain the boundaries of more than one higher-level graphical objects ("clouds").

The central theme of this paper is an approach focusing on the aspects of graphics comprehension process represented by these two question, namely, the attentional process to different levels of graphical objects and the maintenance process of attended objects.

Specifically, the first question is concerned with the operation of bounded activation, or coloring, in Ullman's sense [35]. Ullman developed a theory of "visual routines," namely, sequences of elemental operations applied to particular locations in the visual scene to analyze visual features and spatial relations holding there. Bounded activation is one of the postulated elemental operations, whose function is to define coherent units of regions in the unarticulated visual scene so that further operations can be applied selectively to the activated regions. Roelfsema and his colleagues made this idea more exact by proposing computational and neurological models of the operation [28, 27], and provided neuro-physiological evidence to its functioning in macaque monkey [29, 15]. "Object-based attention" actively investigated by Duncan [6], Kramer and Jacobson [14] and Driver and Baylis [5] largely overlaps with the operation of bounded activation. Strong empirical evidence for the operation has been accumulated in this tradition too (e.g., O'Craven et al. [19]).

The second question is concerned with the operation of marking in Ullman's sense. According to Ullman, some visual tasks require the application of elemental visual operations to multiple locations of the scene. Some tasks further require one to combine information obtained at different locations. The operation of marking is supposed to meet this demand by maintaining the record of the locations already visited and of summary information associated with these locations. Although it is still an open question how this operation is implemented computationally and neurologically, recent studies of visual indexing operations [23, 24, 1, 33] seem to provide good evidence for the existence of cognitive mechanism with the marking functionality.

Given these supporting theories and findings, the two questions we raised above seem to be susceptible to well-grounded empirical research. They are also essential

| | 2005 | 2006 | 2007 | 2008 | 2009 | 合計 |
|---|---|---|---|---|---|---|
| 商品A | 7,046 | 9,256 | 2,807 | 4,970 | 1,216 | 25,295 |
| 商品B | 4,047 | 3,745 | 2,579 | 7,010 | 4,482 | 21,863 |
| 商品C | 9,934 | 3,009 | 7,057 | 1,318 | 7,727 | 29,045 |
| 商品D | 2,575 | 4,201 | 3,379 | 4,742 | 9,550 | 24,447 |
| 商品E | 9,380 | 7,907 | 3,885 | 8,835 | 9,800 | 39,807 |
| 合計 | 32,982 | 28,118 | 19,707 | 26,875 | 32,775 | 140,457 |

**Fig. 2.** Sales figure table used as a stimulus in the table comprehension task

questions to be investigated in graphics comprehension research, since the levels of information one extracts from a graphical representation heavily depend on the levels of objects one attends to, and many realistic comprehension tasks require one to combine information extracted from different locations at different times.

## 3 An eye-tracking study of table comprehension processes

### 3.1 Method

With (1) and (2) as the leading questions, an eye-tracking study was conducted on a group of participants who were engaged in table comprehension tasks. The task was designed so that participants need to attend to objects in a table at two different hierarchical levels: cells at the small object level, and columns and rows at the large object level. It is expected that attention to different levels of objects should reveal itself as difference in eye movement patterns. Consider a task that requires to read off and integrate information contained in multiple cells in a row (or a column). Those participants who attend to cell level objects would look at each of the cells consecutively with intervening saccades while reading off information, whereas those who attend to row (or column) level objects would fixate their eye gaze on the center of the row (or column) and do not produce much saccadic eye movements.

**Task** A set of hypothetical sales figures were presented in the form of an alphanumerical table shown in Figure 2. Each row corresponds to one of five products A to E, and annual sales figures from 2005 to 2009 were shown in each column. Additional row and column were provided to show the total figures for each year and for each product. Four-digit numbers were used as sales figures for all the cells except for those in the total row and column. Participants were asked to read off sales information from the table. We assigned the following three tasks, which require different kinds of objects to be integrated during the task.

*Row task* Participants were asked to find and count the number of all the products with increasing sales between the year 2005 and 2009. Figures should be scanned horizontally, rows by rows, and the number of rows satisfying the given condition should be counted. Thus, rows of figures must be integrated as objects during this task.

*Column task* Participants were asked to find and count the number of all the years in which products A, B, C, D, and E have increasing sales in this order. This time, figures should be scanned vertically, columns by columns, and the number of columns satisfying the given condition should be counted. Thus, columns of figures must be integrated as objects during this task.

*Figure task* Participants were asked to find and count the number of all the annual sales figures on the order of 9,000. Figures can be scanned either vertically or horizontally, and only the counting of individual figures is necessary. Thus, no objects larger than individual figures have to be integrated during this task.

#### Procedure

*Stimuli* A set of six types of stimuli were prepared by changing inter-cell spacings and use of grid lines. Two levels of cell sizes were used to see the influence of inter-cell spacings on table comprehension: dense , 24×72pix and sparse , 37×100pix. Three levels of use of grid lines were used to see the influence of delineating lines: vertical, horizontal, and no grid lines. All alphanumerical tables were constructed with 18pt Tahoma fonts for numerical figures. Three different tables were prepared.

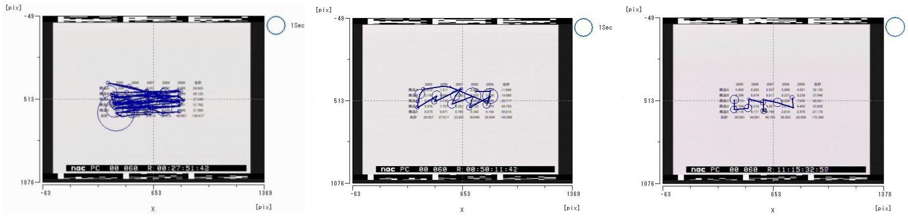
*Presentation steps* After they were informed of the experiment and signed the consent form, participants were instructed to sit in front of a monitor with an eye-tracking equipment. A randomized sequence of stimuli were presented on the display with SuperLab software. 2×3×3 within-subject design was employed for inter-cell spacing, type of grid lines and type of tasks. Task instruction was displayed first, followed by the stimulus table. Participants were asked to press, as quickly and as accurately as possible, a button that corresponds to the answer posed by the task display.

*Measurements* Eye movement patterns of participants while they were reading tables were recorded with an eye tracker NAC EMR-AT VOXER, which has 60Hz temporal precision. Fixation points were extracted from eye movement measurements. Saccades were then identified as connecting two fixations, and they were classified into vertical and horizontal types based on the relative magnitudes of vertical and horizontal displacements. Percentage of correct answers and response times were also recorded.

**Participants** A total of 23 students, 7 males and 16 females, participated in the experiment. Average age of participants were 21.7. Data of participants who did not complete all the experiment steps, and those of participants whose eye movement was not reliably extracted were not included in subsequent analysis.

### 3.2 Results

**Eye movement patterns** Typical eye movement patterns for the three task type (row, column and figure tasks) are shown in Figure 3. A clear difference in eye movement patterns is observed between the row task and the column task. In the row task, eye movement shows a typical reading pattern, which consists of vertical successions of



(a) Row task (b) Column task (c) Figure task

**Fig. 3.** Eye movement patterns in row, column and figure tasks.

horizontal linear movements. The pattern indicates that, to make a judgment on an entire row, the participant scans the row by sequentially fixating on each cell in it for a short time period, and the cumulative row-by-row scans produce the final table comprehension.

In the column task, on the other hand, eye movement consists of a slow horizontal sweep which consists of a sequence of a small number of vertically displaced fixation points. It appears that the participant fixated on the center of each of the columns to make a judgment on the entire column, and produces a column-by-column scan of the entire table during its comprehension. More processing time is required as the column-based scan proceeds to increment and memorize the counting results.

This horizontal-vertical asymmetry between the row and the column tasks indicates that people prefer to attend to smaller, lower level objects, e.g., cells of the table, in the row task, whereas they prefer to attend to larger, higher level objects, e.g., columns of the table, in the column task. It can be argued that, because of the horizontal arrangement of digits in numbers, integrating information across vertically aligned numbers in a column is much easier than integrating information across horizontally ordered numbers in a row. The ease with which numerical information can be integrated in a column encourages people to solve the column task by attending to each column directly. The row task does not afford the same ease of numerical information integration, and people have to resort to cell-based processing of the rows.

The figure task exhibits a similar eye movement pattern as the column task, though the time spent on each column does not vary much across columns. Even though no asymmetry was imposed by the task between vertical and horizontal directions, since the task requires to examine all the cells equally, the pattern exhibited by most of the participants suggests that they employed column task strategy for this task. This was probably made possible because the task of finding figures on the order of 9,000 is relatively easy in a column if we just focus on the initial digits aligned on the vertical line in the column.

The contrast in eye movement patterns observed in the three types of table comprehension tasks clearly indicates that people can and do choose to attend to graphical objects in different levels of hierarchy, depending on the ease of reading off information. As long as information can readily be read off of higher level objects required by the task, those higher level objects are directly attended to. When directly reading information off of task relevant objects is difficult, the information is reconstructed from pieces of information read off of a set of lower level constituent objects.

**Table 1.** Average number of horizontal and vertical saccades for different table forms.

(a) Column task

| | dense | | sparse | |
|---|---|---|---|---|
| grid line / saccade | horizontal | vertical | horizontal | vertical |
| vertical | 10.3 | 6.5 | 10.9 | 11.2 |
| horizontal | 9.3 | 4.9 | 14.5 | 11.6 |
| no lines | 8.4 | 5.5 | 10.8 | 11.2 |
| (Ave.) | 9.3 | 5.6 | 12.1 | 11.3 |

(b) Row task

| | dense | | sparse | |
|---|---|---|---|---|
| grid line / saccade | horizontal | vertical | horizontal | vertical |
| vertical | 29.9 | 3.2 | 32.7 | 3.9 |
| horizontal | 30.8 | 2.6 | 32.3 | 3.5 |
| no lines | 27.8 | 2.3 | 32.9 | 3.1 |
| (Ave.) | 29.5 | 2.7 | 32.6 | 3.5 |

(c) Figure task

| | dense | | sparse | |
|---|---|---|---|---|
| grid line / saccade | horizontal | vertical | horizontal | vertical |
| vertical | 13.0 | 3.4 | 14.4 | 5.8 |
| horizontal | 13.5 | 3.6 | 12.4 | 5.9 |
| no lines | 14.3 | 3.1 | 14.9 | 6.0 |
| (Ave.) | 13.6 | 3.4 | 13.9 | 5.9 |

**Horizontal and vertical saccades for different table forms** In order to more closely look at the effects of table organization on its comprehension processes, particularly in terms of attention to and maintenance of large graphics objects, average number of horizontal and vertical saccadic eye movements while participants were performing a task was counted. A horizontal saccade is a saccade with a larger displacement in the horizontal direction than in the vertical direction. It is considered to be an eye movement along the rows of the table. A vertical saccade is a saccade with a larger displacement in the vertical direction than in the horizontal direction. It is considered to be an eye movement along the columns of the table. The result is summarized in Table 1.

A four factor analysis of variance was first applied to the entire data set to see the over all trend. Main effects were observed on cell spacings, task types and saccade directions. Sparser cell spacings produced significantly more eye movements than denser cell spacings. Row task required significantly more eye movements than both column and figure tasks. Participants produced significantly more eye movements in horizontal directions than in vertical directions.

To see the complex interactions among factors clearly, 2×3 analysis of variance, with cell spacings and grid line types as independent variables, was conducted independently on each combination of the horizontal/vertical saccade directions and the three task types.

*Across task comparison* It is clear from the tables that the row task induces significantly more horizontal saccadic eye movements when compared with the column and the figure task. This task dependent difference in horizontal saccades underscores the difference we observed in eye movement patterns exhibited in Fig. 3.

**Table 2.** Percentage of correct answers for each of the different table and task conditions.

| | Column task | | Row task | | Figure task | |
|---|---|---|---|---|---|---|
| Grid line / Cell spacing | Dense (%) | Sparse (%) | Dense (%) | Sparse (%) | Dense (%) | Sparse (%) |
| Vertical | 79 | 84 | 87 | 86 | 92 | 97 |
| Horizontal | 90 | 75 | 90 | 95 | 97 | 89 |
| No lines | 73 | 79 | 90 | 84 | 87 | 89 |

*Cell spacing effects* If we compare dense and sparse cell spacings more closely, significant increases in both horizontal and vertical saccades were observed in almost all conditions, except for the horizontal saccades in the figure task. This is expected as the increase in cell spacing results in a larger size table that requires more eye movements in both directions to cover the entire table, except in such cases as when specific visual features can be utilized as landmarks to guide eye movements.

cell spacings make it easier for people to attend to cell spacings make it harder to

*Grid line effects* Examination of saccadic eye movements induced by tables with different patterns of grid lines revealed an interesting regularity. In the column task, use of horizontal grid lines in a table increases the number of horizontal saccades ($F(2, 17) - 3.14, p < .10$), but horizontal grid lines do not produce a corresponding increase in vertical saccades. Similarly, in the row task, use of vertical grid lines in a table increases the number of vertical saccades ($F(2, 17) = 4.6, p < .05$), but vertical grid lines do not produce a corresponding increase in horizontal saccades. In both cases, grid lines run perpendicular to the direction of large graphics objects, whose properties need initially be read off to solve the given task. They run parallel to the direction of the scan, needed in the later stage, along the sequence of large objects to obtain the final answer.

**Percentage of correct answers** The results in terms of percentage of correct responses are shown for each of the conditions in Table 2. 2×3 analysis of variance was conducted independently on each of the three task types. In the column task, no main effect was observed either for cell spacings and for types of grid lines, but there was marginal interaction ($F(2, 19) = 2.64, p < .10$), indicating that higher correct responses were obtained with dense cell spacings when there are horizontal grid lines. Similarly, there were no main effects but the same type of marginal interaction was observed ($F(2, 19) = 3.41, p < .10$) in the figure task. In the row task, marginal effect on the types of grid lines was observed to indicate that horizontal grid lines are of help ($F(2, 19) = 3.04, p < .10$).

**Response time** Response time results are shown for each of the conditions in Table 3. 2×3 analysis of variance was conducted independently on each of the three task types. In the column task, a main effect was observed for cell spacings, indicating that dense spacings produce quicker responses ($F(1, 18) = 4.80, p < .05$). An interaction was also observed to indicate that with horizontal grid lines dense cell spacings give significantly shorter response times ($F(2, 17) = 4.05, p < .05$). No significant differences in response times were observed in the row task and in the figure task.

**Table 3.** Response time for each of the different table and task conditions.

| | Column task | | Row task | | Figure task | |
|---|---|---|---|---|---|---|
| Grid line / Cell spacing | Dense (sec) | Sparse (sec) | Dense (sec) | Sparse (sec) | Dense (sec) | Sparse (sec) |
| Vertical | 13.3 | 12.9 | 15.7 | 15.9 | 8.3 | 7.7 |
| Horizontal | 12.2 | 14.9 | 15.7 | 15.1 | 8.4 | 7.9 |
| No lines | 13.2 | 14.0 | 14.9 | 15.3 | 8.4 | 8.5 |

## 4 Discussions

### 4.1 A picture of graphics comprehension processes

Our table comprehension experiment suggests that, at least, two different types of perceptual/cognitive processes, which we call object integration and object maintenance, are working when we are reading alphanumeric tables.

*Object integration:* Object integration is a process, like Ullman's coloring visual routine, in which certain areas in the visual space are carved up and identified as graphical objects. Properties of the graphical objects are recognized and associated with them. Object integration can take place at different object scales. A 4-digit figure is identified as one graphical object composed of a sequence of four numerals, which has a property of representing a certain number. A column is identified as one graphical object composed of a sequence of 4-digit figures, which has a property of representing an increasing trend. The bigger the graphical objects, the more effort they demand for processing. The integration may take multiple eye fixations with intervening saccadic eye movements.

*Object maintenance:* Once graphical objects are established through object integration, they have to be sustained so that their property associations are maintained to be available in further inferencing. Object maintenance is a process, like Ullman's marking visual routine or Pylyshyn's visual indices, in which graphical objects or certain locations in a graphical space are associated with and used for temporary memory for real or hypothetical properties, which can later be retrieved as needed by visually orienting to the objects/locations. Once a column is marked as representing an increasing trend, this property can then be retrieved by accessing the corresponding index, possibly by gazing at the location. It is not needed to reread and reconstruct the information from scratch again.

We emphasize that both object integration and object maintenance are processes operating on and embedded in the domain of real world graphical objects. They shouldn't be conceived as pure mental operations detached from graphical objects and graphical spaces.

### 4.2 Object integration

Our data show that the frequency of saccades over sequences of figures depended, rather strongly, on the spatial extension of the figure sequences. Figures in our stimulus tables were extended more in the horizontal direction than in the vertical direction, and the number of horizontal saccades (mean 29.5 or 32.6, see Table 1(b)) was much greater in

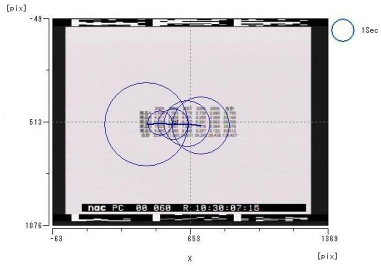


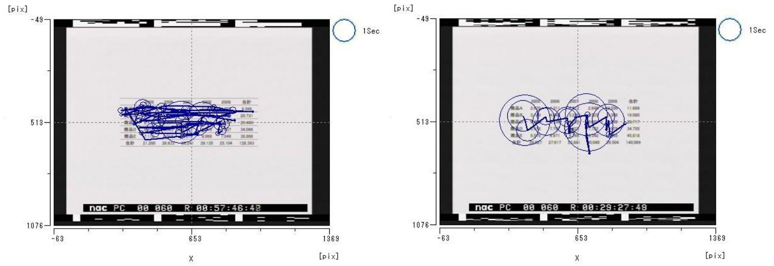

**Fig. 4.** Eye movement pattern for column task on narrow table.

the row task than that of vertical saccades in the column tasks (mean 5.6 or 11.3, see Table 1(a)) and in the figure task (mean 3.4 or 5.9, see Table 1(c)). This indicates that there were a significantly greater number of saccades when scanning horizontal row of figures than when scanning vertical columns of figures, reflecting the difference of spatial extension of rows and columns in our tables. Figures 3(a), 3(b), and 3(c) show typical fixations patterns in the row task with those in the column and the figure task.

The spatial extension of figure sequences also had a significant effect on the frequency of saccades within the same task. For the row task, we found significantly more horizontal saccades on sparse tables (average 32.6) than on dense tables (average 29.5). Significant difference in the frequency of vertical saccades was found between the sparse -table condition and the dense -table condition for the column task (mean 11.3 versus mean 5.6) and the figure task B (mean 5.9 versus mean 3.4).

Under this general tendency, a particularly interesting phenomenon was found in the dense -table, column-task conditions. As shown above, the average number of vertical saccades on dense tables was 5.6 for the column task. Considering each table contains the total of 5 columns of figures to be processed, this means that, on average, less than 1.2 saccades were made per column. Thus, under this condition, eyes move from columns to columns consecutively, often placing only one fixation on each column. Figure 4 shows the typical fixation pattern on narrow tables during the column task. Fairly long fixations were placed on individual columns, but eyes tended to leave a column after no or few vertical movements within the column.

This suggests an interesting contrast in the way consecutive figures are integrated into an object under different conditions. Under the row task, multiple fixations were placed on individual rows of figures. This indicates that initial attentions were oriented to consecutive sub-regions of the row, each consisting of one or a few figures. Due to the task demand discussed in section 3.1, these initially attended sub-regions were then integrated into a row of five figures, to which a task-relevant judgment was attached and kept attached during the trial. This type of subsequent integration must have taken place in the sparse -table, column-task condition, since multiple fixations were observed on individual columns. In contrast, often under the dense -table, column-task condition, only one fixation was placed on an entire column of five figures. This suggests that a single attention spread over the entire column, integrating five figures into an object already at the time of initial attention. Subsequent integration was not necessary in such a case, and task-relavent judgment could be directly attached to the initially attended object. This could be enormous simplification of the relevant comprehension task, and explains a shorter response time in the dense -table condition during the column task.



(a) With horizontal grid lines (b) Without horizontal grid lines

**Fig. 5.** Effect of horizontal grid line on eye movements.

This account can be applied to explain why shorter response time in the dense -table condition was *not* observed for the row task. As the large numbers of horizontal saccades in horizontal task indicate, figures arranged in a row was not covered by a single attention, and this applies to both dense and sparse tables. Thus, regardless of the table size, the direct tagging of task-relavant judgment to figure sequences could not take place during the row task. Sub-regions of a row had to be subsequently integrated anyway, and the internal operations may not be radically different between the dense-table condition and the sparse -table condition. This explains why response time during horizontal task was not significantly different on sparse and dense tables.

The account is also consistent with the fact that shorter response time in the dense -table condition was not observed for the figure task. As discussed in section 3.1, no objects larger than individual figures have to be integrated during the figure task. Thus, the problem of object integration in the column or row level does not arise in the first place, and the possibility of direct tagging is irrelevant to the task.

### 4.3 Object maintenance

The discussion in the last section was mainly based on the frequency of vertical saccades during the column task and of horizontal saccades during the row task. This section discusses fairly consistent patterns in the frequency of *orthogonal* saccades, namely, vertical saccades during the row task and horizontal saccades during the column task.

One consistent pattern in our data was concerned with the effect of orthogonal grid lines: vertical saccades during the row task were significantly more frequent on tables with vertical grid lines, and horizontal saccades during the column task were significantly more frequent on tables with horizontal grid lines. Figure 5(a) and 5(b) show typical fixations patterns during the column task, on tables with or without horizontal grid lines.

One plausible explanation for this pattern refers to the cost of maintaing large objects previously integrated. On this view, for example, horizontal saccades in the column task increased in the presence of horizontal grid lines because the horizontal grid lines divided vertical sequences of figures rather effectively, making it more difficult to keep them integrated as coherent objects. Thus, extra horizontal saccades were necessary in order to return attention to their locations, reintegrate them if necessary, and check the task-relevant judgment attached to them. Without such additional operations, the attached judgments would have been lost together with the objects to which they had been assigned.

This sort of maintenance processes is a realistic possibility given the visual indexing mechanism investigated by Pylyshyn etc. According to the visual indexing theory, we can assign "indices" to several objects or locations in the visual scene. With these indices, we can quickly return attention to the locations of indexed objects without searching for them. We hypothesize that such an index was attached to a row or column of figures when it was first integrated. The index was then used for the quick return of attention when the maintenance need described above arose. Indeed, the use of eye movements for checking internally attached tags were found also by Shimojima and Katagiri [32, 31], and the present case is another instance of eye movements used in combination with visual indices. The present case is unique in that object groups, rather than individual objects or locations, were indexed and revisited for their tags.

This account also explains the other consistent pattern of orthogonal saccades in our data: vertical saccades in the row task were more frequent on sparse tables than on dense tables, and horizontal saccades in the column task were more frequent on sparse tables than on dense tables. We can explain this effect of table size by noting that, as with grid lines cutting across rows or columns, greater spatial extension of rows or columns could make it more difficult to maintain them as integrated objects. Such rows or columns do not conform to the Gestalt principle of proximity, and would be less appropriate for the coverage by a single object-based attention [6, 14, 5, 28]. Indeed, the data discussed in the last section suggest that it took a greater number of fixations to integrate such a column or row into an object. Thus, it is quite plausible that a row or column with a greater spatial extension is more difficult to maintain as an integrated object as well.

Thus, just as grid lines crossing a row or column necessitated additional returning saccades to the row or column in question, a greater extension of a row or column could necessitate returning attention to the row or column in question. The increased orthogonal saccades reflect this necessity. Visual indices support their orientation as direct indicators to the locations of objects to be maintained.

### 4.4 Related work

The distinction of higher-level and lower-level information is almost a commonplace in graphics comprehension research. In fact, a large portion of the research work in this field is centered on our ability to comprehend different levels of information. Taking bar charts and line graphs as primary examples, Pinker [22] emphasizes the existence of a cognitive process by which graph readers "directly translate a higher-order perceptual pattern...into quantitative trends...it symbolizes" (p. 100). He builds a detailed model of the cognitive mechanism underlying this direct interpretation process. Lohse [16] calls this process "perceptual inference made directly" (p. 382), and adopts it as an important parameter in his computational simulation of the way people comprehend graphical representations. Gattis and Holyoak [8] found that people tend to apply the same direct translation routine even when the representation scheme of the given graph has changed significantly.

A related hypothesis is that the proficiency or expertise of reading graphics consists in the ability to appreciate the informational relation underlying this direct translation routine. For example, Lowe [17, 18] compared the performance of meteorologists and non-meteorologists in the reproduction task of meteorological maps, and found that the maps reproduced by meteorologists more accurately preserve those combinatorial features of the original maps carrying higher-level information. In the same line of interest, Kinnear and Wood [12] and Gilhooly et al. [9] compared the performance of experienced readers of topographic maps and less experienced readers, and Halpern and Bower [11] compared the performance of musicians and non-musicians in the recall task on musical scores.

Note that these lines of research are built on the view that the existence of direct translation routines profoundly affects the relative efficiency with which people comprehend information graphics. However, in order for the direct translation process to occur, a particular visual feature to interpret must be isolated, and this presupposes that the boundary of an object having this feature must be set. It is this very early phase of graphics comprehension that our research focuses on. Our data suggest that the performance in this phase can make a significant difference to reaction time, depending on whether a single attention is sufficient to bound an object with the task-relevant size. We also find an indication that the object-boundary is something to be maintained, sometimes calling for returns of overt attention via visual indexing mechanism. This implies that graphics comprehension process is not a one-directional process, where the perceptual process of obtaining a visual feature and the cognitive process of interpreting it are sequentially ordered. The obtained visual feature, and hence its interpretation, can be at risk of collapsing when the boundary of the relevant object is hard to maintain. In such a case, graphics comprehension is a tight coupling of perceptual process and cognitive process, with an active use of eye movements for the maintenance of object boundaries.

In this regard, we suggest an integrative model of graphics comprehension slightly extending the one proposed by Carpenter and Shah [4]. Carpenter and Shah found that visual chunks (such as function lines in a line graph) are interpreted and integrated into knowledge in an incremental way, so that after one visual chunk receives the cycle of interpretation and integration process, the cycle restarts on the next visual chunk. Our data suggests, in addition, that the phase of setting object boundary can be *iterative*, possibly returning to the same object for maintenance.

Freedman and Shah [7] further proposes to incorporate the effect of prior knowledge of the represented domain into the integrative model. Trickett and Trafton [34] emphasize the importance of mental spatial transformation processes in complex reading tasks on complex information graphics. It is very plausible that prior knowledge and on-going mental transformation influence the way a graph reader sets and maintains the boundary on existing graphical elements. And more directly, it is known that the purposes of reading (the differences of reading task) influence the levels of information to be extracted from the given graphics [10, 25]. This indicates that reading purpose has a direct influence on the object boundary. The exploratory study reported in this paper only considered the influence of perceptual features of the given graphics (sparcity/density of figures and the direction of grid lines in tables), and the influence of higher processes on the setting and maintenance of object boundaries have to be left for future work.

## 5 Conclusions

We argued that in order to fully understand how people extract higher-level and lower-level information, graphics comprehension research should explicitly consider the issues of how they set the boundaries of graphical objects to be interpreted, and how the object boundaries are maintained during the task. Our exploratory study with table-reading tasks illustrates the utility of this approach to graphics comprehension processes, especially when it is equipped with the supporting theories of object-based attention and visual indexing.

Overall, the study suggests a fundamentally integrative view, where the cognitive process not only works on visual features obtained in the perceptual process, but in the course of retaining the interpretative result, sometimes refers back to the basic perceptual process of object bounding. This referring-back appears to be mediated by eye-movements back to the actual location of the object in question. In this regard, graphics comprehension seems to be a more embodied process, relying on the persistence of the given graphics as parts of the external world [21, 26].

## References


1. D. H. Ballard, M. M. Hayhoe, P. K. Pook, and R. P. N. Rao. Deictic codes for the embodiment of cognition. *Behavioral and Brain Sciences*, 20(4):723–767, 2001.
2. J. Bertin. *Semiology of Graphics: Diagrams, Networks, Maps*. The University of Wisconsin Press, 1973.
3. J. Bertin. *Graphics and Graphic Information*. Walter de Gruyter, 1981. Originally published in France in 1977.
4. P. Carpenter and P. Shah. A model of the perceptual and conceptual processes in graph comprehension. *Journal of Experimental Psychology: Applied*, 4(2):75–100, 1998.
5. J. Driver and G. C. Baylis. Attention and visual object segmentation. In J. Driver, G. C. Baylis, and R. Parasuraman, editors, *The attentive brain*, pages 299–325. The MIT Press, 1998.
6. J. Duncan. Selective attention and the organization of visual information. *Journal of Experimental Psychology: General*, 113(4):501–517, 1984.
7. E. G. Freedman and P. Shah. Toward a model of knowledge-based graph comprehension. In M. Hegarty, B. Meyer, and N. H. Narayanan, editors, *Diagrammatic Representation and Inference, 2nd International Conference, Diagrams 2002*, LNAI 2317, pages 18–30. Springer-Verlag, 2002.
8. M. Gattis and K. J. Holyoak. Mapping conceptual to spatial relations in visual reasoning. *Journal of Experimental Psychology*, 22(1):231–239, 1996.
9. K. J. Gilhooly, M. Wood, P. R. Kinnear, and C. Green. Skill in map reading and memory for maps. *Quarterly Journal of Experimental Psychology*, 40A:87–107, 1988.
10. J. T. Guthrie, S. Weber, and N. Kimmerly. Searching documents: Cognitive processes and deficits in understanding graphs, tables, and illustrations. *Contemporary Educational Psychology*, 18:186–221, 1993.
11. A. R. Halpern and G. H. Bower. Musical expertise and melodic structure in memory for musical notation. *American Journal of Psychology*, 95(1):31–50, 1982.
12. P. R. Kinnear and M. Wood. Memory for topographic contour maps. *British Journal of Psychology*, 78:395–402, 1987.
13. S. M. Kosslyn. *Elements of Graph Design*. W. H. Freeman and Company, 1994.
14. A. F. Kramer and A. Jacobson. Perceptual organization and focused attention: the role of objects and proximity in visual processing. *Perception & Psychophysics*, 50(3):267–284, 1991.
15. V. A. F. Lamme, V. Rodriguez-Rodriguez, and H. Spekreojse. Separate processing dynamics for texture elements, bundaries and surfaces in primary visual cortex of the macaque monkey. *Cerebral Cortex*, 9:406–413, 1999.
16. G. L. Lohse. A cognitive model for understanding graphical perception. *Human-Computer Interaction*, 8:353–388, 1993.

17. R. K. Lowe. Speech strategies and inference in the exploration of scientific diagrams. *Educational Psychology*, 9(1):27–44, 1989.
18. R. K. Lowe. Diagram prediction and higher order structures in mental representation. *Research in Science Education*, 24:208–216, 1994.
19. K. M. O'Craven, P. E. Downing, and N. K. Kanwisher. fMRI evidence for objects as the units of attentional selection. *Nature*, 401(584–587), 1999.
20. P. Olivier. Diagrammatic reasoning: an artificial intelligence perspective. In A. F. Blackwell, editor, *Thinking with Diagrams*, pages 63–78. Kluwer Academic Publishers, 2001.
21. K. J. O'Regan. Solving the "real" mysteries of visual perception: The world as an outside memory. *Canadian Journal of Psychology*, 46(3):461–488, 1992.
22. S. Pinker. A theory of graph comprehension. In *Aritificial Intelligence and the Future of Testing*, pages 73–126. L. Erlbaum Associates, 1990.
23. Z. Pylyshyn. The role of location indexes in spatial perception: A sketch of the FINST spatial-index model. *Cognition*, 32:65–97, 1989.
24. Z. Pylyshyn. *Seeing and Visualizing: It's Not What You Think*. The MIT Press, 2003.
25. R. M. Ratwani, J. G. Trafton, and D. A. Boehm-Davis. Thinking graphically: Extracting local and global information. In *Proceedings of the Twenty-fifth Annual Conference of the Cognitive Science Society*, pages 958–963, 2003.
26. D. C. Richardson and M. J. Spivey. Representation, space, and hollywood squares: Looking at things that aren't there anymore. *Cognition*, 76:269–295, 2000.
27. P. R. Roelfsema. Cortical algorithms for perceptual grouping. *Annual Review of Neuroscience*, 29:203–227, 2006.
28. P. R. Roelfsema, V. A. F. Lamme, and H. Spekreijse. The implementation of visual routines. *Vision Research*, 40:1385–1411, 2000.
29. P. R. Roelfsema, V. A. F. Lamme, and H. Spekreojse. Object-based attention in the primary visual cortex of the macaque monkey. *Nature*, 395:376–381, 1998.
30. A. Shimojima. Derivative meaning in graphical representations. In *Proceedings of the 1999 IEEE Symposium on Visual Languages*, pages 212–219. IEEE Computer Society, 1999.
31. A. Shimojima and Y. Katagiri. An eye-tracking study of exploitations of spatial constraints in diagrammatic reasoning. In G. Stapleton, J. Howse, and J. Lee, editors, *Diagrammatic Representation and Inference, 5th International Conference, Diagrams 2008*, Lecture Notes in Artificial Intelligence 5223, pages 74–88. Springer-Verlag, 2008.
32. A. Shimojima and Y. Katagiri. Hypothetical drawing in embodied spatial reasoning. In *Proceedings of the 30th Annual Meeting of the Cognitive Science Society*, pages 2247–2252, 2008.
33. M. J. Spivey, D. C. Richardson, and S. A. Fitneva. Thinking outside the brain: Spatial indices to visual and linguistic information. In J. M. Henderson and F. Ferreira, editors, *The Interface of Language, Vision, and Action: Eye Movements and the Visual World*, pages 161–189. Psychology Press, 2004.
34. S. B. Tricket and J. G. Trafton. Toward a comprehensive model of graph comprehension: Making the case for spatial cognition. In D. Barker-Plummer, R. Cox, and N. Swoboda, editors, *Diagrammatic Representation and Inference, 4th International Conference, Diagrams 2006*, LNAI 4045, pages 286–300. Springer-Verlag, 2006.
35. S. Ullman. Visual routines. *Cognition*, 18:97–159, 1984.
36. H. Wainer. Understanding graphs and tables. *Educaional Researcher*, 21:14–23, 1992.