Perceiving Window Geometry: An Experimental Study

William Cowan, Sandra Loop

Department of Computer Science
University of Waterloo
Waterloo, Ontario, Canada

Abstract

The importance of incorporating the user's response capabilities into the design of user interfaces is now well established. Many experiments that measure basic user responses have concentrated on motor capabilities, such as Fitts's law [1]. The user's perceptual abilities, as they apply to the organization and interrelating of displayed information, are another important design factor. This paper describes an attempt to measure perceptual capabilities, specifically the ability of users to discriminate borders in window systems. These capabilities are used in parsing visual information that is displayed in multiple windows. Two aspects of this problem are considered, window occlusion, which is important in establishing the depth ordering of windows on the display surface, and window border design, which is important in providing visual separation between windows. The minimum occlusion needed to provide easy depth judgments is measured experimentally. This part of the paper provides an interesting example of closing the loop in interface design. The measured value is related to visual acuity, a very basic aspect of human vision. Additionally, it is used successfully in an experimental interface that is designed to minimize possible inter-window visual conflicts [2]. The experiments on border design are still in progress. However, it is already possible to provide some rules of thumb that are useful in border design.

1. Introduction

The multi-window user interfaces that are common on modern engineering workstations and personal computers present the user with a potentially difficult synchronization problem. Output from a variety of different sources is multiplexed on the display surface, and the user, in understanding it, must segregate the output streams. This visual feat is usually performed almost effortlessly by the mechanisms of selective attention: processing resources are concentrated on the region of the display surface that contains the output stream of current interest, and other areas are ignored. It can fail under two circumstances. First, display organization can be so poor that a user who concentrates all his attention on the problem of discriminating one stream from another cannot do so. Fortunately, this circumstance is unusual: even novice designers who inspect their results see the problem and correct it immediately. Thus, it usually arises not from design flaws but from programming errors, most often a failure to synchronization access to display resources. Second, display organization is good enough that a user who focuses any enough attention can get the correct result consistently, but poor enough that occasional errors occur when attention lapses. Such errors force the user to concentrate on the operation of the window system itself instead of on the tasks to which it provides an interface. Thus, the second type of problem lowers the productivity of users. At the same time it is very difficult to measure and monitor, since experimental subjects concentrate intensively while their behaviour is being monitored.

A problem of just this sort arose during the development of a novel user interface technique [2] in our laboratory. The new technique was designed to provide the user with better discrimination when nearby windows present visually conflicting information. For example, borders of adjacent windows provide visual conflicts if they are inadvertently positioned so as to create unintended patterns. Similarly, the contents of windows provide visual conflicts if visual elements of juxtaposed windows merge into an unintended whole. The spurious patterns produced by these conflicts are a problem because the user provides erroneous inputs based on the unintended patterns, sometimes at considerable cost. The modified window manager detects such patterns and changes the position of windows to remove them. Its implementation requires quantitative limits that describe human perceptual performance. These limits are needed to detect pattern formation and to eliminate it with minimal window movement.Measuring these limits is difficult because they apply to performance when the user is supplying minimal attention to window geometry. Under most experimental conditions users supply full attention to the experimental task, providing performance limits that would apply only if the user abandoned his or her work so as to attend only to the operation of the window system. This paper describes the results of measurements to determine perceptual limits when the user applies minimal effort to the task. Since this is a common requirement of interface performance, the methodology used to isolate the perceptual features of interest and to make measurements in which the subject is allowed to utilize a minimum of processing resources may be of interest in addition to the results.

The first objective is to operationalize the quantities to be measured, thereby creating a quantitative measure that describes performance on the more complicated tasks facing the user. The measure chosen was the perception of window geometry. Most window systems maintain windows internally as a 'two and one-half' dimensional structure [3], with the windows ordered in depth but having no specific depth value. The depth ordering determines occlusion relations when windows overlap. Most window systems are like X [4] in that users are unable to perceive order relations for windows that don't overlap. Thus, they can deduce only a semi-order that is a subset of the total order [5] used by the window system. This semi-order is detected by observation of window overlaps, and users can perceive window overlaps unambiguously only when the visual elements that make up window borders are attributed to the correct window. Thus, the visual conditions under which window geometry can be correctly detected are exactly those under which unambiguous perception of border elements is possible. Since the modified window manager increases either overlap or separation to make unambiguous perception possible the right quantity to measure is the overlap needed to make reliable judgments of window depth orderings.

The second objective is to determine the limit to be measured. As window overlaps increase users' errors decrease until they reach an asymptote that depends on quantities such as display time. The overlap at which the error rate reaches asymptote is a conservative estimate of the level at which performance is no longer limited by simple perceptual factors.

The third objective is to create viewing conditions that examine the user's performance when using limited perceptual processing resources. This choice is difficult because performance can be made arbitrarily bad by degrading either viewing conditions or the perceptual resources the viewer is able to use sufficiently. At some point further degradation is inappropriate, since the experiments then simulate an unusable system. In this set of experiments viewing conditions are chosen by conscious imitation of viewing characteristics that are desirable for a workstation display. For example, the experiments are done in a normal working environment with ordinary ambient lighting and with the background noise of other workers.

The experiments described here have an additional motivation: evaluating the efficacy of different border designs. If an interface uses a window border style that offers the user maximal occlusion discrimination, the amount of rearrangement can be reduced or even eliminated completely. Thus, these experiments also examine the influence of different border styles on the ability of a user to read window geometry. This aspect of the experiments is at present less complete, however, since we lack a suitable categorization of design elements. Speculation on such a categorization and on associated rules of thumb for window borders is included in the discussion.

The next section discusses the experimental design for the whole series of experiments. The following section describes the experiments done to date, with the results compared to measures of visual acuity. The paper then concludes with a discussion of rules of thumb for window design following from the experimental measurements. Using the border results when designing interface components is mostly work for the future. One implementation using the geometry result is already complete, however: the modified window manager discussed above.

2. Methodology

The objective of the experiments described below is to measure how a user's ability to determine the geometry of windows changes as a variety of parameters of the window border change. In particular, the user often must know which window is uppermost on the display, because it is usually the focal window. In a normal window system several cues combine to reinforce one another, including window overlaps, title bar highlighting, insertion point blinking and the user's memory of his or her past actions [6]. The present experiments concentrate on window overlaps because they are an interesting component of other interface issues, as discussed above. Thus, the user (or subject, as he or she is most often called in experimentation of the type described here) is shown a window configuration for a short time and asked which window was uppermost. By acquiring this simple response many times, at only a few seconds per response, the experimenter is able to determine a standard measure of accuracy, proportion correct, which is used as the measure of performance in this series of experiments.

A single trial consists of four components occurring sequentially. First, a fixation point appears in the centre of the display surface. The subject directs his or her attention at this point. Because the fixation point standardizes the direction of gaze from trial the effect of randomly directed attention, which substantially increases the variability of the results, is minimized. The fixation point persists for one second. It then disappears, to be replaced by the stimulus, a configuration of three rectangles similar to Figure 1. The stimulus remains on the display surface a very short time, less than half a second. The shortness of this time has three reasons. 1) The experiment is designed to measure the subject's ability to determine window geometry at a glance. The short exposure makes it impossible to perform a leisurely inspection of the stimulus. 2) Short enough exposures make it impossible for the subject to execute eye movements, so that critical points in the display, the corners where the windows overlap, cannot be foveated. This condition is appropriate because it is important to be aware of window geometry without attending specifically to the window borders. 3) A given window configuration usually persists for tens of seconds or minutes. A fraction of a second is a reasonable proportion of time for the user to attend to window geometry. The stimulus is succeeded by a mask which prevents further processing of the image based on intermediate, or iconic [7], representation of the stimulus in the visual system. The trial ends when the subject indicates the number of the window perceived to be uppermost, and is succeeded immediately by the following trial. The response is made by typing the number of the uppermost window using the numeric keypad of a standard Macintosh keyboard. A variety of details of this procedure merit further explanation.

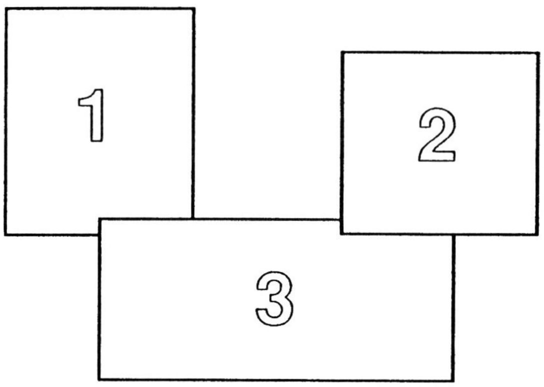

Figure 1. A window configuration similar to the ones used in the experiment. The subject determines the window ordering based on occlusion of the corners. The large black numbers in the diagram are medium grey in the experiment and are used to identify the window for the subject's responses.

The stimuli are displayed on the screen of a Macintosh II as black lines drawn on a white field. The field has a luminance of about 30 candelas per square metre. The subject views the screen from a comfortable viewing distance for workstation use, about 500 mm. To simulate ordinary workstation conditions the experiment is conducted in the Computer Graphics Laboratory under normal lighting and working conditions. Dim ambient illumination illuminates the display surface at about one lux, and there is a low level of background noise from other workers. Reliable stimulus timing is obtained by displaying each part of the trial for a fixed number of frames, drawing it off screen, then transferring it onscreen during vertical refresh.

The subjects are graduate and undergraduate students working in the Computer Graphics Laboratory. All subjects work daily using window systems. Five to ten subjects are used in each experiment, with one exception where extensive data was collected on a single subject, one of the authors (SL).

The relationship of the mask to the stimulus is interesting. The mask consists of randomly positioned horizontal and vertical lines the same contrast and thickness as the window borders. Thus, it effectively masks the overlapping corners that must be processed to determine window geometry. At the same time it interferes minimally with the large grey digits that label the windows. Consequently, the subject can read the appropriate digit after determining the uppermost window even if the stimulus presentation has ended. This arrangement attractively imitates normal window usage, in which the user processes the contents of the focal window after determining which one it is.

Most of the results are presented in terms of pixels. The pixel on the Macintosh screen is about 0.35 mm in diameter, a typical size for a high resolution CRT. At a viewing distance of 500 mm it subtends 2.4 minutes of visual angle. Since both pixel size and viewing distance are standard for workstation usage visual angles vary little if the stimuli are presented on other displays at constant pixel sizes. Since all stimuli are presented in units of pixels and since pixels are easy to incorporate into graphics algorithms, representing results in pixels is simplest and most convenient. When it is necessary to compare the results to data on visual acuity, the visual data is converted to pixels.

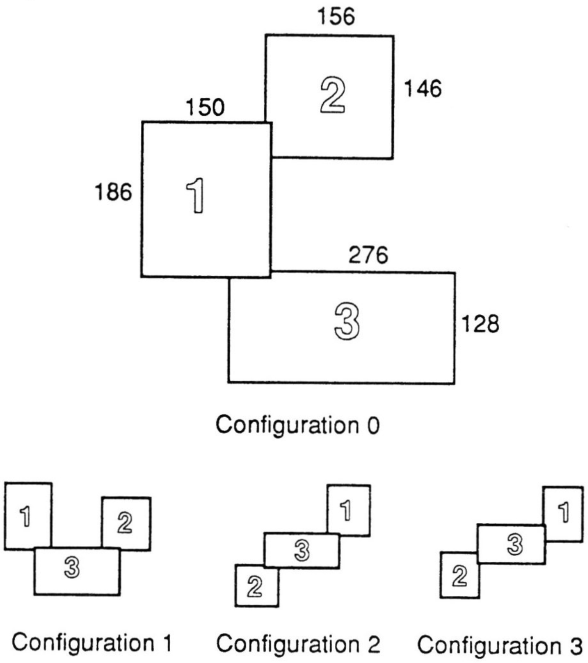

Figure 2. The four window configurations used in the experiments. Configuration 0 shows the height and width of each window, given in pixels.

Trials are grouped into sessions in blocks of several hundred, taking between five and twenty minutes. Each block consists of conditions to be contrasted crossed with all permutations window depth orderings, of window configurations, of which four were used, illustrated in figure 2, and of orientations, horizontal and vertical reflections of each window configuration. The trials were presented in random order, preceded by fifteen practice trials.

Preliminary experiments are used to determine good values for some of the stimulus parameters. When pairs of windows are considered, it is discovered that subject can easily discriminate which window is uppermost under all possible stimulus conditions that are not ambiguous. When three windows are present the discrimination is considerably more difficult, with the difficulty varying as border parameters change. The three windows are arranged as shown in Figure 2. Six window orderings are possible, but two specify the uppermost window ambiguously. Thus they are omitted. Another pair generate identical stimuli since they are ambiguous as to the ordering of the two rearmost windows. Only one of them is used, along with the completely unambiguous pair. Thus, with three windows the number of window orderings is small enough that it can be explored exhaustively. Unfortunately, with more than three windows the number of possible configurations and orderings becomes prohibitively large.

3. Results

The experiments all measure proportion correct as percentage. This measure is corrected for guessing using the equation $p = (3m - 1) / 2$, where p is the proportion correct and m the raw score. Different conditions are compared by analysis of variance using DataDesk, an interactive statistics program for the Macintosh, and the results are shown in tabular form in the sections that follow.

Window occlusion

The main experiment was designed to measure the effect of window overlap on the ability of users to determine the depth ordering of window arrangements. The four window configurations shown in Figure 2 were used, in all depth orderings and orientations with three values of overlap, 3, 6 and 16 pixels, and three values of image duration, 50, 100 and 200 milliseconds. The trials were blocked by duration, with five subjects at 100 and 200 milliseconds and three subjects at 50 milliseconds. There was no main effect of image duration, window arrangement or orientation. There were, however, main effects of overlap and depth ordering, shown in Tables 1 and 2 respectively.

Overlap (pixels)	3	6	16
Per cent correct	48	90	95

Table 1. Percentage of trials on which the uppermost window is identified correctly, corrected for guessing, as the window overlap is varied.

uppermost window	right/top	centre	left/bottom
Per cent correct	84	57	81

Table 2. Percentage of trials on which the uppermost window is identified correctly, corrected for guessing, depending on which window is uppermost.

Note that performance improves as the window overlap increases, and is close to asymptote at 6 pixels. This performance can be understood somewhat better by examining the interaction between image duration and window overlap, shown in table 3. At all durations performance improves when the overlap increases from 6 to 16 pixels but the improvement is small. On this basis it can be concluded that the performance asymptote occurs between the two values. The other interesting feature of this table is the difference at different durations. Although overall performance is the same at each duration the difference between small and large overlaps changes greatly. Presumably subjects did the task with enough attention to get 75-80% correct in a given block, allocating extra resources in the more difficult blocks (50 millisecond durations). The performance difference at 50 milliseconds shows that this condition is very sensitive to changes in the subject's ability to determine window order, so it was used for all other experiments.

	Duration (milliseconds)			
	100	150	250	Average
3	35	49	56	48
Overlap 6	90	94	86	90
(pixels) 16	99	96	92	95
Average	75	79	78	77

Table 3. Percentage of trials on which the uppermost window is identified correctly, corrected for guessing, depending on the duration of the image presentation and the size of overlap.

When interpreting these results on window occlusion it is important to notice that overlap area covaries with the linear measure of overlap used above. A preliminary experiment was run to investigate this relationship, holding a small overlap constant while varying the length of the overlapped region. This experiment showed little variation in performance as the length varied at values greater than 10 pixels. Thus, the above experiments are done with the overlap region having a length of 20 pixels. In balancing different overlaps in situations where trade-offs of a few pixels matter the minimum dimension of the overlap region is of most importance. This rule of thumb breaks down when the overlap region is close to being square with a dimension about ten pixels or smaller. Additional measurements would be needed to parametrize trade-offs for such overlaps, but the case should be uncommon enough and the measurements difficult enough that there was no incentive for doing them as part of this study.

Relationship to visual acuity

One objective of this research is to establish relationships between performance measures such as the one described here and underlying characteristics of human perception. The relationship shown in Table 2 provides the key. Because ambiguous configuration were eliminated configurations with either of two outer windows uppermost (outside configurations) can be discriminated on the basis of a single overlap, but configurations with the centre window uppermost (inside configurations) require correct discrimination of two overlaps. Visual acuity diminishes as the stimulus moves away from the centre of the visual field. Thus, it may be possible to discriminate the outside configurations but not the inside ones because a single overlap region is above the acuity threshold but two separated ones cannot be above simultaneously. This explanation is likely to be true given that the human visual system seems to be capable of establishing depth ordering using all information in the visual field in parallel [3]. A set of measurements in which the size of the centre window varies provides a test of this hypothesis.

Since within subject comparisons are most sensitive for a measurement of this type the experiment is done by having a single subject perform sessions with thirteen different sizes for the centre window. When the centre window is changed in size all windows in the configuration are changed by the same magnification factor. The experiment is done with a three pixel overlap, and only three configurations, since configurations 2 and 3 of Figure 2 are essentially the same for the centre box.

Only performance when the centre window was uppermost is considered. The results are shown in Figure 3. Note that performance is essentially perfect when the centre window is small but falls to chance as it increases in size. This performance is consistent with an explanation based on acuity.

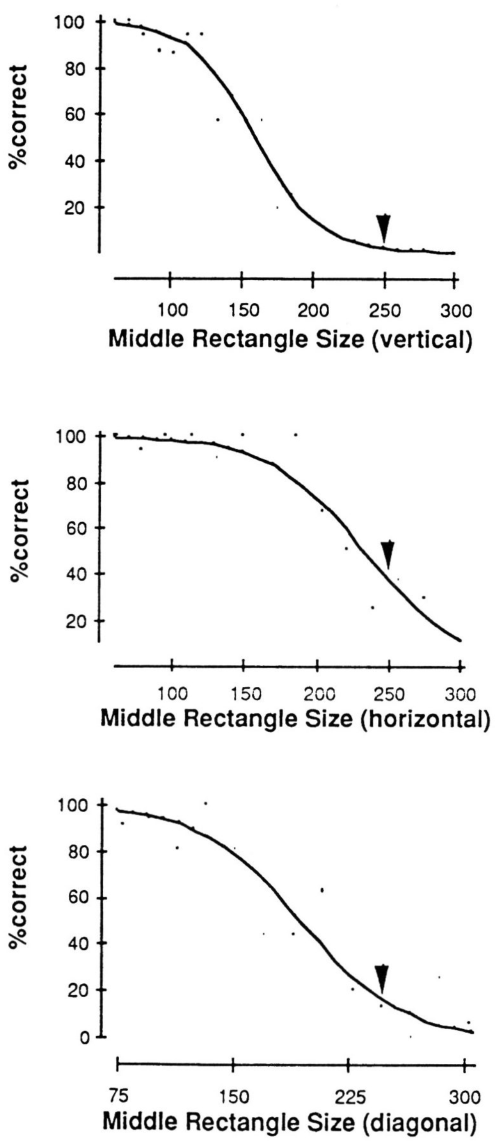

Figure 3. Percentage of responses on which the uppermost window is identified correctly, corrected for guessing. The data is fitted with a hyperbolic tangent to guide the eye. The open arrow indicates the point at which the second corner is far enough into the periphery that the three pixel overlap can no longer be resolved.

A more quantitative test is possible. An overlap of 3 pixels with a 1 pixel border produces a feature 2 pixels, or about 4.8 minutes of visual angle, wide. Under optimal viewing conditions this feature encounters the acuity limit at about 10 degrees from the centre of the visual field [8], which corresponds to 250 pixels. This distance is marked on the curves of Figure 3. It is easily seen to correspond roughly to the limit of the region where performance is at chance. Unfortunately, this test cannot be made much more precise because the stimuli used for determining acuity are so different from the areas of overlap used in this experiment. However, controlled experiments to determine discriminability of the overlap regions could be designed for comparison with the data of Figure 3. Even without such experiments, however, this data points out that holistic depth discrimination must be based on features large enough that they can be perceived regardless of where the user's attention lies on the display surface.

Window borders

The above experiments make use of the simplest possible window border, a one pixel line in a maximally contrasting colour. Real borders are usually more complex. How does this complexity assist or hinder the perception of window geometry? To understand this factor discrimination measurements are made using a variety of different border styles. The eight styles used are shown schematically in Figure 4. All are tested using overlaps of 3 and 7 pixels. The results are shown in Table 4, which shows the very large difference in performance that is produced by small differences in border style.

		Border style							
		1	2	3	4	5	6	7	8
Overlap	3	37	0	0	15	51	45	79	63
(pixels)	7	90	70	45	64	94	85	85	91

Table 4. Percentage of trials on which the uppermost window is identified correctly, corrected for guessing, depending on the border style and the size of overlap.

On the basis of this preliminary data it is already possible to see a few interesting regularities. First, borders perform worst when they create additional edges that interfere with the border edge that delineates the exterior of the window. This interference persists even when the overlap is large, as exemplified by style (3). Second, borders that imitate shadows seem to perform unusually well. Whether this good performance is created by the appearance of adding a feature outside the window, or whether it is the differing visual texture created by using grey for the shadow is currently not known. Measurements to determine optimal parameters for shadow-style borders are continuing in our laboratory.

4. Conclusions

This paper reports work in progress aimed at putting the design of window borders on an experimental foundation. At this point in the project it is already possible to draw several conclusions and to see some future directions where interesting questions can be answered. These experiments examine a user's ability to tell 'at a glance' how windows are arranged in depth.

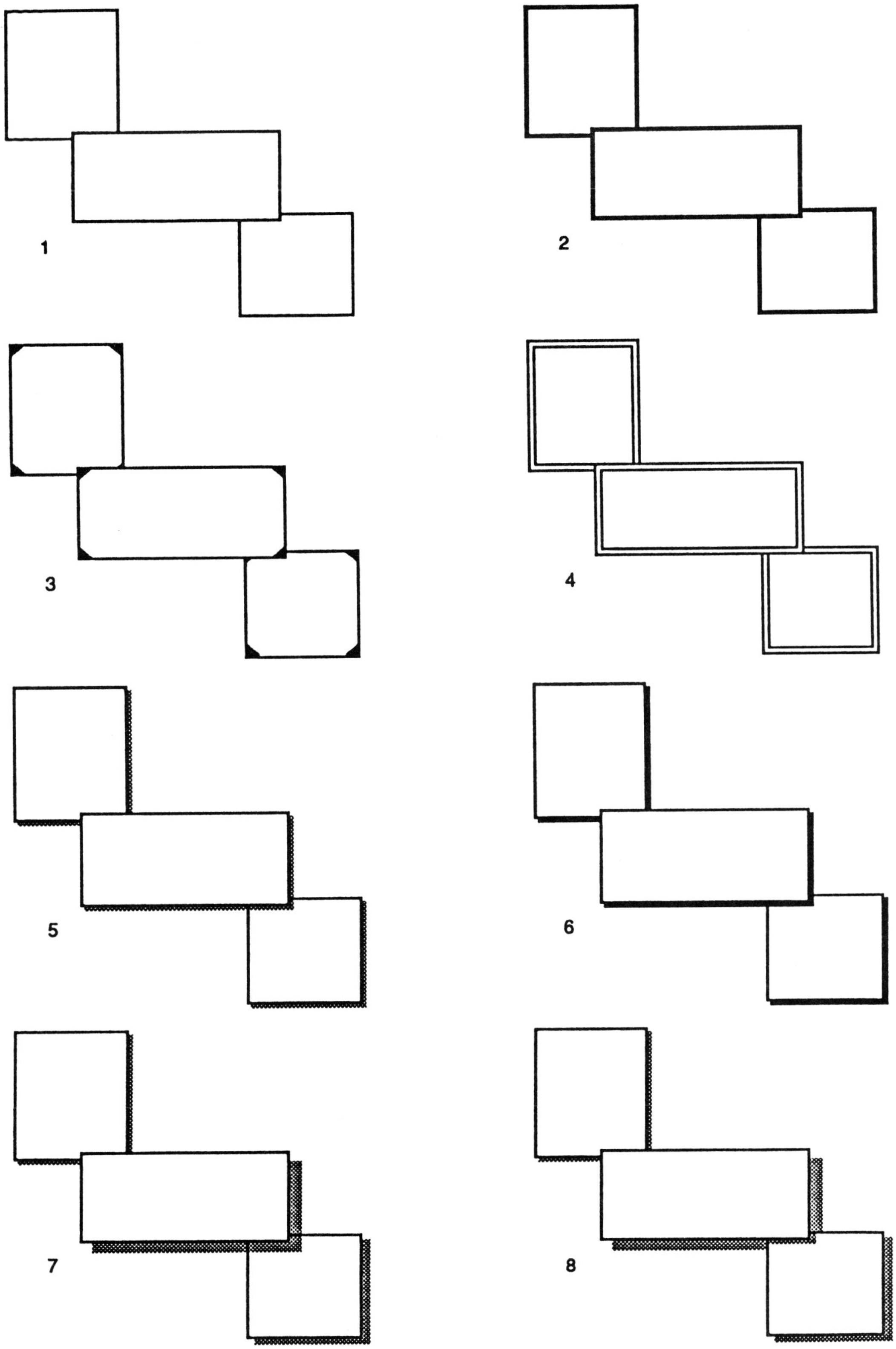

Figure 4. Schematic illustrations of the different border styles used in the final experiment. They include simplified examples of many border styles that are common in commercial user interfaces.

Preliminary experiments showed this task to be easy under any conditions when only two windows are to be ordered. But a larger number of windows necessitates integration of information from different parts of the display surface. The experiments discussed above provide significant evidence that the limiting factor in perceiving occlusion relations is the visibility of the overlaps that provide information about them. This conclusion is derived from extensive examination of perception of only three windows. Experiments using more complicated window configurations should now be done to confirm the validity of this hypothesis for more realistic displays. The existence of the well-defined hypothesis we have deduced from three window experiments will make it possible to examine a small number of well-chosen cases where more windows are present. They will suffice to confirm or contradict the present model.

Knowing the factors used to organize multi-window displays becomes increasingly important as display size increases. Current 21-inch displays provide a viewing angle of more than 45 degrees. The model favoured by the data in this paper suggests that a display that is easy for a user to understand 'at a glance' requires features like window borders, overlaps, etc. to be big enough to be visible that far into the periphery of the user's visual system. This requirement necessitates feature dimensions upwards of thirty pixels. Since features of these dimensions often produce significant waste of screen real estate attributes that can be used to provide an organizational capability that operates in parallel, such as colour [9], are likely to be increasingly important to help the user organize the large amount of disparate information that can be presented on a modern multi-window system.

It also suggests that designs for visual elements that are based on a quantitative understanding of visual effectiveness will become an increasingly important aspect of user interface design. In this respect the preliminary experiments on window borders indicate an important future direction for research into the user interface. Window borders play a more complicated role than window overlaps, since they are the basic element that allows the user to parse the display into areas that contain related information. Thus, it will be important to determine how border designs help users to perform a variety of different tasks: separating, identifying and so on. Their multi-potential role means that any specific border design is likely to represent a trade-off among several different roles. Making such a trade-off in a quantitative way, or having it made by an algorithm, requires a significant data base of experimental information. Thus, the experiments reported here represent the tip of an iceberg.

Of course, the ultimate test of the perceptual limits measured in these experiments is the successful operation of user interfaces based on them. One user interface, the modified window manager discussed above utilized these results, found 25 pixels to be an amount large enough to provide unambiguous information about window arrangement. The usability tests reported in [2] seem to show that this value is large enough, and that smaller values of 5 and 10 pixels, which the present experiments show to be at best marginal on the 21 inch screen used in the tests, are insufficient for easy use of the system.

5. Acknowledgments

This research was partially supported by grants from the Natural Sciences and Engineering Research Council. In addition, financial support and motivational discussions were provided by the Human Interface Group at Apple Computer.

We would like to thank Kevin Schlueter and Pierre Jolicoeur for many helpful discussions during the course of these experiments.

Macintosh is a trademark of Apple Computer. DataDesk is a trademark of Odesta Corporation. NeWS is a trademark of Sun Microsystems.

6. References

[1] S.Card, W. English and B. Burr, 'Evaluation of mouse, rate-controlled isometric joystick, step keys and text keys for text selection on a CRT', *Ergonomics*, 21, 127–148, 1978.

[2] K. G. Schlueter, *Perceptual Synchronization in Window Systems*, M. Math. Thesis, Department of Computer Science, University of Waterloo, 1990. K. G. Schlueter and Wm Cowan, 'Synchronizing the User: A Problem in the User Interface of Windowing Systems', manuscript in preparation.

[3] D. Marr, *Vision*, Freeman: San Francisco, 1982.

[4] O. Jones, *Introduction to the X Window System*, Prentice-Hall: New York, 1989.

[5] Most text-books on the foundations of mathematics contain useful descriptions of semi-orders and total orders. A good example is P. R. Halmos, *Naive Set Theory*, Springer-Verlag: New York, 1976.

[6] W. Cowan and M. Wein, 'History and state in user interfaces', *Interact'90*, Cambridge, 1990; published in *Human-Computer Interaction—INTERACT '90*, edited by D. Diaper et al., pp. 555–560, 1990.

[7] B. G. Breitmeyer, *Visual Masking: an Integrative Approach*, Clarendon: Oxford, 1984.

[8] L. L. Sloan, 'The photopic acuity-luminance function with special reference to parafoveal vision', *Vision Research*, 8, 901–911, 1968.

[9] W. Cowan and C. Ware, *Colour Perception Tutorial Notes*, SIGGRAPH '85, San Francisco, 1985.