Integrality and Separability of Input Devices

ROBERT J. K. JACOB, LINDA E. SIBERT, DANIEL C. MCFARLANE, and
M. PRESTON MULLEN, JR.
Naval Research Laboratory, Washington, D.C.

Current input device taxonomies and other frameworks typically emphasize the mechanical structure of input devices. We suggest that selecting an appropriate input device for an interactive task requires looking beyond the physical structure of devices to the deeper perceptual structure of the task, the device, and the interrelationship between the perceptual structure of the task and the control properties of the device. We affirm that perception is key to understanding performance of multidimensional input devices on multidimensional tasks. We have therefore extended the theory of processing of perceptual structure to graphical interactive tasks and to the control structure of input devices. This allows us to predict task and device combinations that lead to better performance and hypothesize that performance is improved when the perceptual structure of the task matches the control structure of the device. We conducted an experiment in which subjects performed two tasks with different perceptual structures, using two input devices with correspondingly different control structures, a three-dimensional tracker and a mouse. We analyzed both speed and accuracy, as well as the trajectories generated by subjects as they used the unconstrained three-dimensional tracker to perform each task. The results support our hypothesis and confirm the importance of matching the perceptual structure of the task and the control structure of the input device.

INTRODUCTION

A mechanical input device encodes motion into a signal that can be read by the computer. Designing workable input devices requires that the motion clearly convey the intent of the user and complement his or her physical capabilities. The design of current input devices and their interaction techniques have been driven more by what is technologically feasible than from an understanding of human performance. Their success relies, in part, on the well-documented human ability to adapt. To design and select more effective

input devices and interaction techniques, we need to use a deeper under-
standing of task, device, and the interrelationship between task and device
from the perspective of the user.

Such understanding may come from the intuition and judgment of design-
ers and, perhaps, from empirical studies of specific new devices. However,
greater leverage is available by reasoning from a more general predictive
theoretical framework, rather than finding an *ad hoc* answer to each such
question. The present study provides one example of the development and use
of such a theoretical framework to the problem of three-dimensional input
devices (as, for example, Card et al. [1983] have done for two-dimensional
pointing devices).

To do this, we extend the theory of processing of perceptual structure
[Garner 1974; Garner and Felfoldy 1970], first developed with fixed images,
to interactive graphical manipulation tasks. According to this theory, the
attributes of objects in multidimensional spaces can have different dominant
perceptual structures, *integral* or *separable*, as described in more detail
below. The nature of that structure, that is, the way in which the dimensions
of the space combine perceptually, affects how an observer perceives an
object. We posit that this distinction between perceptual structures is a key to
performance of multidimensional input devices on multidimensional tasks.
Hence, two three-dimensional tasks, such as those in our experiment, may
seem equivalent, but if they involve different types of perceptual spaces, they
should be assigned to correspondingly different input devices.

Three-Dimensional Tracker

A three-dimensional tracker, such as the Polhemus 3SPACE or Ascension
Bird, is a three-dimensional absolute-position locator. In contrast, a mouse is
a two-dimensional relative-position locator. The three-dimensional tracker
reports its position in three-space relative to a user-defined origin. (In fact,
the Polhemus and Ascension devices also report their rotational orientation,
but we have focused on position only.) The device allows a user to input three
coordinates or data values simultaneously and to input changes that cut
across all three coordinate axes in a single operation. Such a device is
obviously useful for pointing in three-space, but it is also applicable in any
other situation that involves changing three values simultaneously. A mouse,
in comparison, requires two operations to manipulate three variables. One
commonly used design for mapping three variables (such as x, y, and z for
zooming and panning) onto a mouse allows two of the variables (x and y) to
be input simultaneously in normal operational mode and the third (z) to be
controlled through a mode change button that temporarily turns the mouse
into a one-dimensional slider. A technological view of these two alternatives
suggests that the three-dimensional tracker is logically a superset of the
two-dimensional mouse (provided they are both used as absolute devices, i.e.,
the mouse is not lifted from its pad), since it provides the same two outputs
plus a third. Thus, the three-dimensional tracker should always be used in
place of a mouse since it is always at least as good—and sometimes better.

This also assumes ideal devices (that do not exist) with equal cost and equal accuracy and the absence of other ergonomic differences in such parameters as control-display ratio, nonlinearities, instabilities, size, shape, and weight. Nevertheless, our intuition tells us that this is unlikely to be true—but why? Our goal is to base such judgments on a firmer theoretical foundation.

History of Input Device Frameworks

From the inception of interactive computer graphics, researchers have proposed a number of frameworks that organize knowledge about input devices to simplify the job of selecting an appropriate device for a task. The earliest works focused on equating the physical properties of different input devices to minimize the programming effort in substituting one device for another. Later systems added knowledge about device pragmatics and task requirements to better capture their qualities and use.

An early abstraction is that of the logical device found in device-independent graphics packages developed from standards such as ACM's Core Graphics System [GSPC 1977]. Devices were assigned to logical-device classes based on how they performed fundamental user actions. Foley et al.'s [1984] taxonomy of interaction techniques improved this concept by adding a middle layer that made explicit the fact that more than one interaction technique may perform a given elementary task. However, both approaches considered, for example, selecting with a joystick or trackball equivalent. They hid the crucial pragmatic aspects of haptic input by treating devices that output the same information as equivalent, despite the different subjective qualities they present to the user.

Buxton [1986] recognized the importance of these qualitative differences which he called pragmatic attributes. He developed a taxonomy that organized continuous-input devices by property (position, motion, pressure) and the number of dimensions. A tablet, light pen, and two-dimensional joystick have two-dimensions and sense position, but they differ from a two-dimensional trackball because they sense motion. While this approach can point out that substituting a trackball for a joystick is not equivalent, it does not explain why.

Mackinlay et al. expanded Buxton's taxonomy to include most input devices and complex controls [Card et al. 1991; Mackinlay et al. 1990]. One feature of their approach is an evaluation technique for comparing alternative designs in terms of expressiveness (e.g., is the meaning exactly conveyed) and effectiveness (e.g., is the meaning conveyed with felicity). Their approach both furthers our understanding of the structure of input device space and recognizes that human performance issues are important for understanding how a device actually works in a given situation. Although some relatively straightforward human factors issues are handled formally, such as matching the size of the domain and range of a value, the more subtle pragmatics of input device usage and task characteristics are still handled by a set of specific rules.

Bleser [1991] developed a device taxonomy and input model that explicitly incorporated the physical attributes of input devices, including the notion of

the physical separability of input degrees of freedom, and knowledge about task requirements. The taxonomy and model were used in an interactive design tool to suggest one or more natural interaction techniques based on a description of an interaction task [Bleser and Sibert 1990]. The model introduces a set of heuristic rules and a pattern-matching procedure rather than a more general, theoretical framework, but it highlights the need for such information in the design process.

Therefore, while current frameworks include knowledge about physical structure, pragmatics, and some task requirements, it is not enough to explain why the three-dimensional tracker is not always better. We suggest that what is also needed is to move from an *ad hoc* understanding of input devices and task requirements to the added leverage of incorporating a predictive theoretical framework that allows reasoning about the utility of a device for a particular task. To help put the study of multidimensional input devices on a firmer theoretical footing, we extend the theory of processing of perceptual structure to interactive graphical manipulation [Garner 1974; Garner and Felfody 1970]. Our research hypothesis is that performance improves when the perceptual structure of the task matches the control structure of the device. To test our hypothesis, we conducted an experiment in which subjects performed two tasks that have different perceptual structures, using two input devices with correspondingly different control structures, a three-dimensional tracker and a mouse. The data analysis examined several aspects of performance: task completion time, time to reach an absolute accuracy criterion, and differences among the trajectories generated by subjects using the three-dimensional tracker as they completed each task. The results converge to confirm the utility of matching the perceptual structure of the task and the control structure of the input device.

BACKGROUND

This work merges two separate threads of research. One, described above, is the understanding of input devices in human–computer interaction, from the logical-device concept through more recent taxonomies and tools. The other gives insight into task and performance issues; this is the theory of the processing of perceptual structure of a multidimensional space formed by the attributes of an object.

A multidimensional object is characterized by its attributes. A small red circle has size, color, shape, and location. Investigations into spatial structures that describe the perception of such attributes [Attneave 1950; Shepard 1964] led Garner [1974] to observe the attributes of some visual objects combine perceptually to form a unitary whole, while those of other objects remain more distinct and identifiable. Attributes that combine perceptually are said to be *integral*; those that remain distinct are *separable*. For example, value (lightness) and chroma (saturation) of a color are perceived integrally, while size and lightness of an object are perceived separably [Handel and Imai 1972]. The horizontal and vertical positions of a single dot in the middle of an outline square are integral [Garner 1974], but color and shape are separable [Imai and Garner 1968].

Integral and separable define two classes of perceptual structure that mark the endpoints of a continuum rather than forming a sharp dichotomy. Where an object falls can be determined by two operational methods. First, integral and separable objects can be distinguished by direct similarity scaling, a technique that measures the perceived similarity among objects. In this procedure, subjects are asked to compare pairs of objects and derive a value that indicates how alike they are overall. These measures are then compared to the actual Euclidean and city-block distances in the attribute space. Integral objects are those related by the Euclidean metric (distance in Euclidean space); subjects' judgments are based on the overall sameness of the objects. Separable objects are related by the city-block metric which is the sum of the distances measured parallel to the coordinate axes (the distance one would have to travel in a city in which all streets run in an x or y direction). Their distance is the *sum* of distance in each attribute. Second, when asked to form classes of objects that have the same set of attributes (for example, that vary in shape and color), observers will partition an integral set into clusters based on their overall similarity, but a separable set will be partitioned by attribute. Integral objects are said to possess similarity structure, and separable objects dimensional structure. With similarity structure, individual attributes do not dominate; with dimensional structure, the attributes cannot be ignored.

Perceptual Structure Extended to Interaction

The notion of integral and separable can be extended to interactive tasks by observing that manipulating a graphical interaction object is simply the changing of values of its attributes. Since the attributes of an object define a perceptual space, changing these values is the same as moving in real time within the perceptual space of the object. Because integral objects have a similarity structure and follow the Euclidean metric, and separable objects have a dimensional structure and obey the city-block metric, the nature of the interaction path can be predicted: movement in an integral space should be Euclidean, straight-line distance between two points; movement in a separable space should be city-block and run parallel to the axes. In other words, graphical interactive tasks have a perceptual structure that influences how they are completed. Moreover, the type of perceptual structure is the same as the type of perceptual structure of the underlying object.

Control Structure of Input Devices

The control spaces of input devices have similar characteristics. An input device with more than one degree of freedom can be characterized as integral or separable based on whether it is natural (or possible) to move *diagonally* across dimensions. With an integral device, movement is in Euclidean space and cuts across all the dimensions of control. A separable device constrains movement to a stair-step, city-block pattern; movement occurs along one dimension at a time.

METHOD

Our research hypothesis states that performance improves when the structure of the perceptual space of an interaction task mirrors that of the control space of the input device. To test its validity, we examined two interactive tasks, one set within an integral space and one in a separable one, and two devices, one with integral control and one, separable. This yields a two-by-two experiment with four conditions. We predict performance on each task will be superior in the condition where the device matches the task in integrality/ separability. That is, the interaction effect between task and device should far exceed the main effects of task or device alone. We also predict that the perceptual structure of the task determines how the subject manipulates the three-dimensional tracker; the path to target should be more nearly Euclidean with the integral task than the separable one.

We chose two tasks that require adjusting three variables but concentrated on task spaces that do not directly map to three-dimensional physical space and on manipulation rather than pointing tasks, in order to frame a balanced comparison. The integral three-attribute task required changing the x-y location and the size of an object to match a target. The separable task required changing x-y location and the color (lightness or darkness of greyscale). Studies of noninteractive stimuli suggested that position and size are perceived as integral attributes,while position and color are perceived as separable [Garner 1974]. For the integral device, we used the Polhemus tracker which permits input of three integral values. For the separable device, we used a conventional mouse which permits two integral values, to which we added a mode change to enable input of a third, separable, value. Our research hypothesis asserts that the three-degree-of-freedom input device (Polhemus) will be superior to the two-degree-of-freedom plus mode change device (mouse) when the task involves three integral attributes (location/size) rather than two integral plus one separable attribute (location/greyscale). Conversely, the separable mouse plus mode change device will be superior for the separable (location/greyscale) task.

Design

Each subject performed both tasks using both devices, constituting a repeated-measures design. Forty subjects (26 men and 14 women) were randomly assigned to four presentation orders. In each order, one task was completed with both devices before the second task was presented to control for practice and fatigue with a device. Each order had a unique sequencing, and together the four orders formed a latin square design so that each device and task combination appeared only once at each position in the sequence. Subjects were clerical, administrative, and technical personnel from the Information Technology Division of the Naval Research Laboratory who volunteered to participate without compensation. All but two were right-handed. Only two had tried a Polhemus previously. Thirty-seven reported using a mouse daily, two weekly, and one monthly.

Stimulus and Apparatus

On each trial, the subject adjusted a user-controllable object until it matched the location and either size or greyscale level of a target. For the size task, the user-controllable object was a square whose greyscale level was 50% of the total greyscale range and whose location and size could be adjusted (see Figure 1). The target position and size were represented by a black outline square. The range of sizes was 0.7 to 6.2 inches on a side. For the greyscale task, the user-controllable object was a square of size 2.8 inches (the midpoint of the size range) that contained an embedded circle 1.5 inches in diameter (see Figure 2). The greyscale level of the circle was under user control, and the target greyscale level was displayed in the outer region of the user-controllable object to facilitate matching. The target position was a black outline square. In both tasks, the user-controllable object was translucent, so it never obscured the position target; the background was white to maximize the range of remaining available intensities that the human vision system can distinguish; and the monitor gamma correction was set at a level that enhanced the greyscale range of the monitor.

The maximum size of the space of possible three-variable targets is a cube 13.75 inches on a side, the width of the monitor. The restrictions we imposed on size and greyscale cut down this space differently for each task. We required that the entire target square fit within the 13.75-by-11 inch screen display area and that a target not be closer than 5% of the screen height from the edge of the screen to allow for overshooting the target and to eliminate any conflict between the edge of the target and the edge of the screen, and we reserved the greyscale levels near black and white for the target position outline and screen background, cutting the range of allowable intensities by 20%. After taking these restrictions into account, the size target space became a rectangular frustrum with more small than large targets, and the greyscale space, a balanced rectangular prism. We made further adjustments to the two spaces so that they had equal volumes with equal number of targets (giving each target the same probability of being selected) and so that each had a reasonable central value (the greyscale midpoint is the middle of the intensity range; the size midpoint is the centroid of the unbalanced frustrum), stimulus range, and equal range of hand movement.

A different stimulus set was randomly generated for each of the four conditions but constrained by a script so that the total three-dimensional distance between one stimulus and the next was the same across conditions. Corresponding trials thus had equal distances but different absolute locations. We also required each stimulus to differ from the preceding one by at least 0.5 inches in each dimension to avoid degenerate trials that could be completed without exercising all three dimensions of motion. Each stimulus trial was represented in the computer program as a three-dimensional point (x, y, z), and the software driving the experiment was identical except for how the z dimension was displayed on the monitor. The x and y dimensions of both tasks were mapped as position, but the z was mapped as either size or greyscale level.
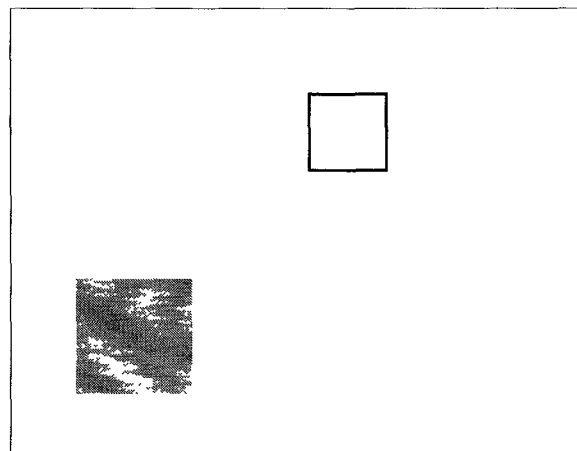
Fig. 1. Stimulus for the integral (size) task. The outline square is the target. The user adjusts the location and size of the solid grey square to match the target.
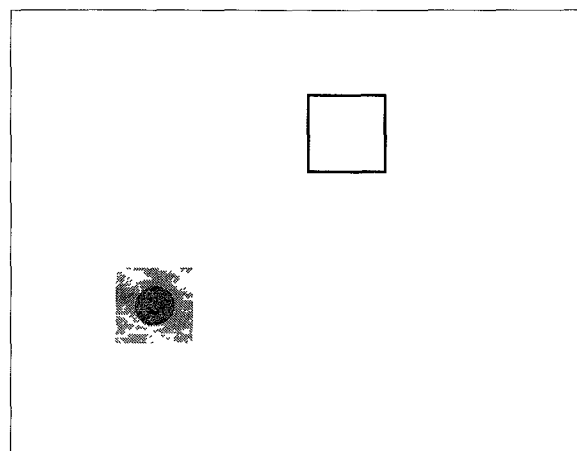

Fig. 2. Stimulus for the separable (greyscale) task. The outline square gives the target location, and the outer area of the solid square gives the target color. The user adjusts the location of the solid grey square to match the target outline and the color of the inner circle on the grey square to match that of the outer area.

The three-dimensional tracker used in the experiment was the Polhemus 3SPACE magnetic tracker. It consists of a transmitter that generates electro-magnetic fields and a wand housing three orthogonal coils that sense the fields. The position and orientation of the wand are transmitted to the host computer. In this experiment, only the three position values were used. The Polhemus was configured as an absolute device with the origin at what would be the forward end of the arm of the chair. The source was permanently fixed

under the subject's chair. The control–display ratio was one inch of device movement to one inch of screen movement. To improve the performance of the Polhemus, the raw position data was smoothed with a small moving average filter, and we eliminated as much metal as possible from the surrounding area to reduce interference with the electromagnetic fields. The experiment was located away from electrical wiring and metal poles; the furniture was wooden; and the metal mouse pad was removed when not in use. The Polhemus was initialized from one of two precalibrated files (one for left-handed use, one for right) so that (1) the operating area was consistent across subjects and (2) its axes were orthogonal. Movement of the wand in a plane parallel to the screen moved the user-controllable object in x and y. Moving the wand toward the screen made the object either bigger or darker; away made it smaller or lighter.

The mouse was a standard relative-position optical mouse with its control–display ratio set to minimize the need for stroking (picking up and repositioning the mouse while moving the cursor over long distances) in order to make its behavior close to that of an absolute-position device like the Polhemus. Two of the three input values needed were the standard x and y mouse coordinates. The third was a mode change into a one-dimensional slider, in which the mouse was moved along the vertical axis of the pad while any of the three mouse-buttons was depressed. We chose this strategy because it is similar to those currently used in applications that require mouse input of more than two variables. Movement over the optical pad corresponded to moving the user-controllable object in x and y. Movement while holding the mouse button corresponded to changing the size or lightness of the user-controllable object; movement toward the top of the optical pad (toward the display screen) made the object either bigger or darker; toward the bottom made it smaller or lighter.

The computer was a two-processor (16 MHz) Silicon Graphics Iris Workstation, model 4D/120G. The program was divided into two processes, which ran concurrently on the two processors. Most other system processes and all network daemons were eliminated. One of our two processes continuously monitored the Polhemus over a serial port and fed data into an event queue in shared memory, while the other drew the images and supervised the experiment. This architecture was effective in greatly reducing the often-observed lag in response to movements of the Polhemus because handling of the heavy serial port data transfer from the Polhemus was performed by a separate processor, communicating via shared memory. Position data from the mouse or Polhemus were recorded continuously throughout the experiment, approximately every 20 milliseconds. The monitor was a 19-inch Hitachi, and the gamma correction was set at 2.7.

Procedure

The arrangement of the apparatus is shown to scale in Figure 3. The subject was seated on a straight-backed chair in front of a monitor that was placed on a standard office desk. The subject was approximately 56 inches from the
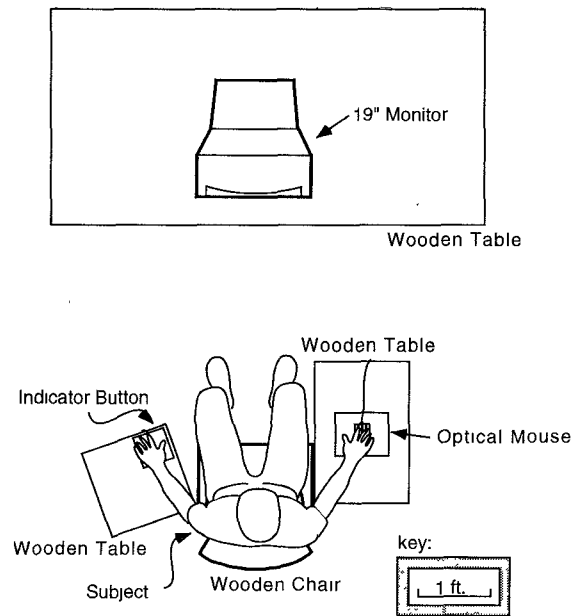
Fig. 3. Room layout, precisely to scale; note how the subject sits 56 inches from the monitor. This illustration shows the mouse condition. For the Polhemus condition, the wooden table and mouse were removed, and the subjects held the Polhemus wand in their right hand. The indicator button and input device were reversed for left-handed subjects.

monitor, a distance that gave the subject ample room to manipulate the Polhemus and reduced the interference between the electromagnetic fields of the monitor and the Polhemus. The mouse pad was placed on a 27.5-inch high table located under the subject's preferred hand. The indicator button was located on another 27.75-inch high table under the subject's other hand. The tables and chair were located in the same position for each subject. The subject, however, was allowed to adjust the position of the mouse pad and the button upon the table tops for comfort. The experimenter sat at a terminal to the left rear of the subject. The experiment was conducted in a special-purpose laboratory designed for such work [Achille 1990].

Subjects were required to designate a preferred hand and use it to manipulate both devices. Information about handedness and choice of preferred hand was recorded, as was gender of the subject. The information was used to balance the assignment of subjects to experimental conditions to control for variation caused by gender differences and handedness.

The experiment was divided into four presentation orders, one for each device and task combination. Within each order, the subject was first given 33 practice trials to stabilize performance before beginning 88 data trials. The data trials were subdivided into sets of 11, with the first in each set not scored because it measured the time to home the user-controllable object from an unknown starting position. The home position was located in the middle of

the screen with the object being either midsized or midcolored. Each trial required that the subject change the location and either size or greyscale color of the user-controllable object using one of the two devices until it matched the position and size or color of the target. The subject indicated that a match was complete by pushing the indicator button. It would have been easy to have the computer terminate each trial automatically, as soon as a match of sufficient accuracy was achieved. However, as discussed below, we wanted to collect data that would allow us to investigate a wide range of different accuracy criteria through a retrospective analysis procedure, and thus we did not want the experimental procedure to constrain the choice of the criterion in advance.

Within a set of trials, as soon as a subject pushed the indicator button, that trial was ended and another target presented for the next trial. An instruction screen separated the sets of trials, and the subject determined when to start a new set. Subjects were instructed to rest as long as they needed between sets of trials. Subjects were encouraged to ask questions during practice but not during the data trials. They were instructed that accuracy and speed were of equal importance. Subjects were asked to complete a short questionnaire at the conclusion of the experiment to learn their opinions about the devices and tasks and about their prior experience with selected input devices. Each subject took approximately 1.5 hours to finish the experiment.

RESULTS AND DISCUSSION

The data analysis is in three parts. The first looks at the time and accuracy, when the subject ended the trial by pressing the indicator button, that is, when the subject thought he or she was satisfied with the match. The second examines time to reach a fixed absolute criterion (before the very end of the trial); this combines time and accuracy into a single measure. The last analyzes the difference between the trajectories that subjects followed with the Polhemus as they completed each task. To facilitate the latter two analyses, we recorded the position of the mouse or Polhemus approximately every 20 ms. during each trial. All processing and aggregation was conducted after the data collection part of the experiment was completed. The analysis of questionnaire data can be found in Jacob and Sibert [1992].

Time to End of Trial

The first analysis looked at time and accuracy when a subject ended a trial. Time was measured from the appearance of the target until the subject pressed the indicator button. Accuracy was total Euclidean distance across all three dimensions of the stimulus space (x, y, and size or lightness) from the center of the user-controllable object to the center of the target. The research hypothesis predicts a strong interaction effect between task and device; the two task and device combinations in which both are integral or separable should give superior performance.

Average (mean) time to button press, shown in Figure 4 and graphed in Figure 5, suggests that, as predicted, neither task nor device alone produced as large an effect as the interaction of the two. These observations were evaluated with a repeated-measures analysis of variance. Order of presentation was not significant ($F(3, 39) = 1.09$, $p > 0.30$), allowing aggregation of the four orders. There were significant effects for both task ($F(1, 39) = 7.40$, $p < 0.01$) and device ($F(1, 39) = 7.80$, $p < 0.01$) and a highly significant effect for interaction between task and device ($F(1, 39) = 69.34$, $p < 0.0001$). An omega-squared analysis indicated that 20% of the total variance was accounted for by this interaction; this is high given the large variation from individual differences.

The accuracy results, shown in Figure 6, exhibit a strong task difference, with subjects less accurate on the greyscale task. (The small device effect is confounded by a task and device interaction and so was not considered.) These observations were also evaluated with a repeated-measures analysis of variance. Order of presentation was again not significant ($F(3, 36) = 2.29$, $p > 0.09$), allowing aggregation. The task difference was highly significant ($F(1, 39) = 117.11$, $p < 0.0001$). Both device ($F(1, 39) = 11.32$, $p < 0.01$) and a task and device interaction ($F(1, 39) = 6.52$, $p < 0.05$) were present. Although the mean scores differed by task, the variances of their accuracies are approximately the same, indicating a stability in a subject's underlying judgment process.

Since the button press indicates when a subject thinks the match is good enough, performance time is linked to the subject's own maximum accuracy criterion. The data showed that although the subjects selected a wide range of criteria, they used a consistent stopping criterion across tasks, as indicated by the standard deviations of the accuracy results. The strong interaction effect in performance times and the high percentage of the variance accounted for by the interaction strongly support the conclusion that matching integrality or separability of the device to that of the task provides an advantage. A common denominator to improving performance is matching the structure of the device and task.

Time to Absolute Criterion

The second analysis measured the time required by subjects to reach a fixed accuracy criterion on each trial. This can be viewed as simulating an experiment in which the subject was required to reach a certain accuracy, at which point the trial was terminated automatically. This approach combines speed and accuracy into a single measure and removes the effect of an individual subject's *personal* accuracy criterion for terminating trials. It also allows removing the effect of a speed–accuracy tradeoff, since performance to a fixed accuracy criterion is required. Accuracy was, again, the overall Euclidean distance to the target in the three-dimensional space of stimuli.

The raw position data, recorded approximately every 20 ms. during each trial, were transformed using linear interpolation into a time series with a 10-ms. period. A retroactive data analysis algorithm then calculated the *last*

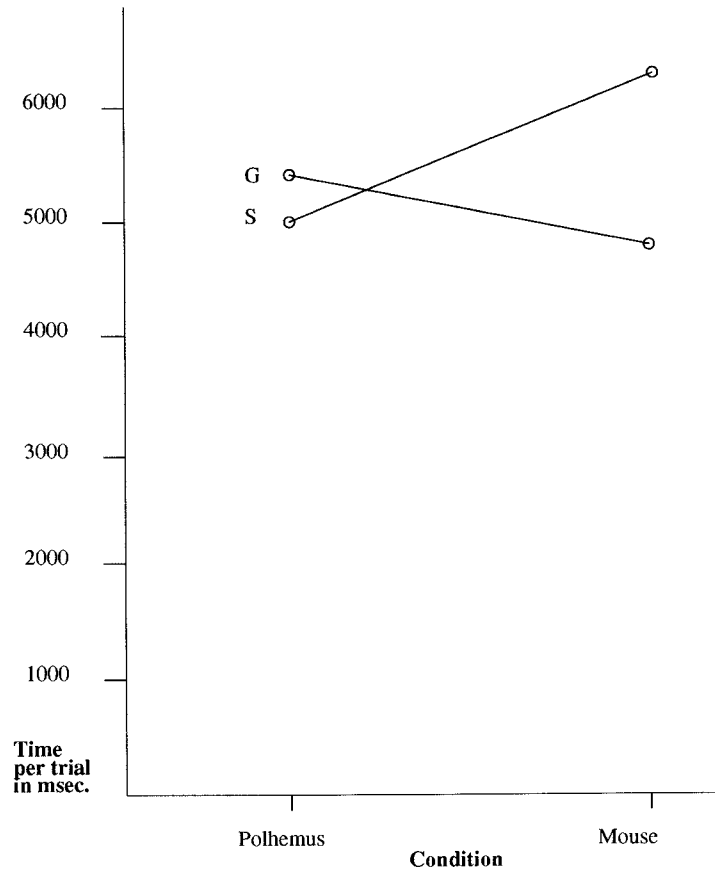| Task	Device			
	Integral (Polhemus)		Separable (Mouse)	
	Mean (ms.)	Std. dev.	Mean (ms.)	Std. dev.
Integral (Size)	4981	2065	6274	2518
Separable (Grey)	5357	1613	4838	1269

Fig. 4. Time per trial in msec.

Fig. 5. Graph of mean time per trial in msec., illustrating interaction effect. Line marked S shows performance on the integral (size) task, G, the separable (grey) task.

time a subject reached a given criterion on each trial. Simply setting an accuracy cutoff would underestimate trial completion time by terminating a trial when a subject just nicks a target distance unintentionally while over-shooting the desired accuracy. With our retroactive analysis, however, time is measured until the subject reached our criterion for the last time during a trial, a better approximation of performance to that distance from the target.

Task	Device			
	Integral (Polhemus)		Separable (Mouse)	
	Mean (ms.)	Std. dev.	Mean (ms.)	Std. dev.
Integral (Size)	0.0805	0.0308	0.0630	0.0334
Separable (Grey)	0.1088	0.0303	0.1034	0.0422

Fig. 6.   Accuracy in inches.

Retroactive analysis also allowed us to produce a series of simulated experiments, each having a different accuracy criterion. Each such simulated experiment is a snapshot of performance to a selected distance from the target. The results from each experiment are the average performance times for the four conditions. Two hundred evenly spaced distances from the target were chosen in order to investigate how performance changed as the distance to the target decreased (i.e., as accuracy increased). These 200 experiments cover the full range of behavior, from distances that were too far from the target to be considered even a crude match, to distances that required the match to be very good. As noted, we could have run any *one* of these experiments by terminating trials automatically at a fixed accuracy criterion. The retrospective analysis, however, allows us to simulate many different experiments from the same data. (The simulation is not entirely precise, since subjects' behavior would have been somewhat different under different stopping criteria.)

The analysis involves two stages. First, the average performance of all subjects was computed for each task and device combination at each criterion distance from target. A plot of these for criteria near the target is shown in Figure 7.

Next, the four averages were ranked according to completion time. The bar drawn in Figure 7 highlights the region in which the rankings follow our predictions, that is, for the size task, where Polhemus always beats mouse, and for the greyscale task, where mouse always beats Polhemus. This region contains all the accuracies that one would consider to be successful completions of the task, from very inaccurate task performance at the distance 0.24 inches (Figure 8 illustrates a 0.24-inch match) down to an almost perfect match. The results from several criteria were analyzed further to evaluate the significance of these findings, and the results were all similar. We present results for 0.099 inches, an accuracy reached on 70% of all trials. These results are shown in Figure 9 and graphed in Figure 10. These observations were evaluated with a repeated-measures analysis of variance. Neither task $(F(1, 39) = 2.48, p > 0.12)$ nor device $(F(1, 39) = 3.18, p > 0.08)$ was significant, but the interaction of task and device $(F(1, 39) = 52.46, p < 0.0001)$ was highly significant as predicted. Omega squared indicated that 22% of the variance was accounted for by this interaction. In contrast, task (1%) and device (2%) accounted for very little. The highly significant interaction supported by the high percentage of variance accounted for by the interaction
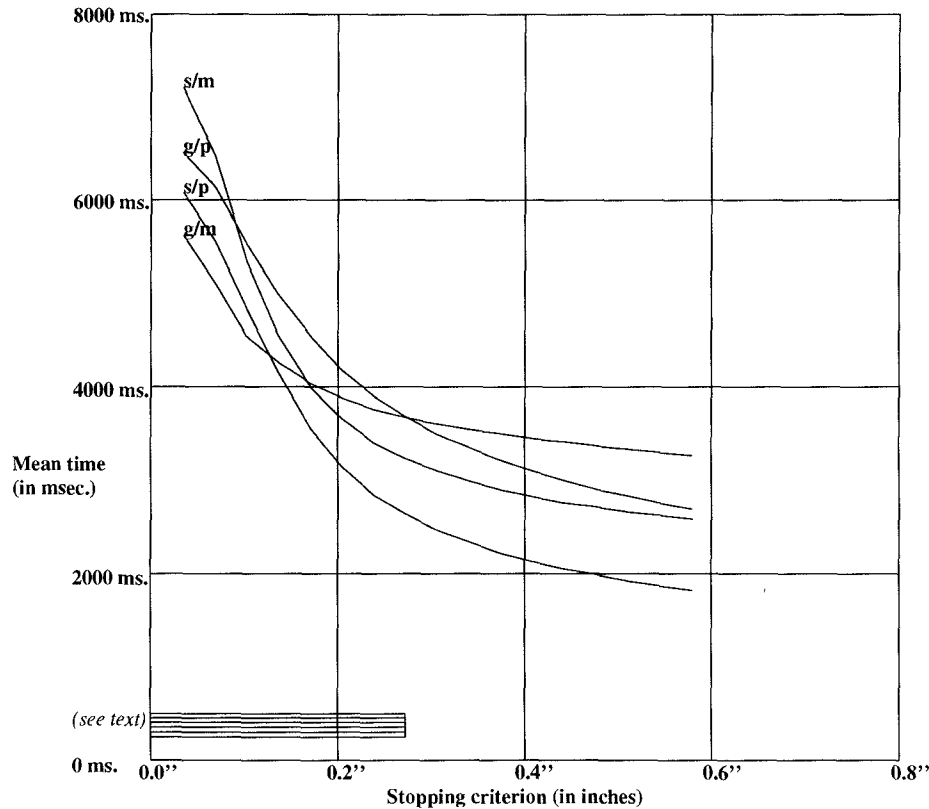
Fig. 7. Average performance by condition over a range of stopping criteria covering 0.579 to 0.034 inches (Euclidean distance to target). The line marked **s / p** gives the data for the size task using the Polhemus; **s / m** is size task using mouse; **g / p** is the greyscale task using Polhemus; and **g / m** is greyscale task using mouse. The striped bar at the bottom indicates the criteria where performance time rankings followed our predictions. The value 0.24 inches at the right end of the bar corresponds to the criterion at which **g / m** and **g / p** cross.

again supports the hypothesis that matching task and device in integrality or separability leads to better performance.

The rankings also contain a second region of interest, further from the target. Although the distance to target in this region is too great to be considered a match, the results shed light on the process of completing a match. This region begins at approximately 3.65 inches from target and ends at the 0.24-inch mark (the distance at the right end of the bar in Figure 8). In this early behavior, for each task, the Polhemus always performed better than the mouse (although still better on the integral size task than the separable greyscale one). That is, for very crude, *quick-and-dirty* matches, the Polhemus was always faster. Again, several points in the range were examined, and the data for 0.987 inches are presented here. The results are shown in Figure 11 and graphed in Figure 12. A repeated-measures analysis of variance was again conducted. Both task $(F(1, 39) = 143.23, \ p < 0.0001)$
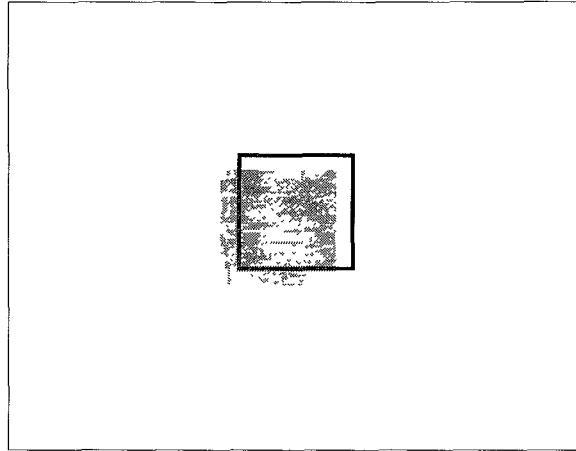
Fig. 8. Illustration of the accuracy implied by the 0.24-inch criterion referenced in Figure 7, shown to scale for a 1.2-inch-wide target and user-controllable object. Shaded square is 0.24 inches in Euclidean distance from the outline square in x and y only, in the stimulus space. This criterion distance is too large to be considered a match for most purposes.

	Device			
	Integral (Polhemus)		Separable (Mouse)	
Task	Mean (ms.)	Std. dev.	Mean (ms.)	Std. dev.
Integral (Size)	4062	1523	4501	1424
Separable (Grey)	4902	1427	4035	938

Fig. 9. Time in msec. to reach 0.099-inch criterion accuracy.

and device $(F(1, 39) = 97.60, \ p < 0.0001)$ were highly significant, and there was no interaction $(F(1, 39) = 0.83, \ p > 0.37)$. An omega-squared analysis showed that 35% of the variance was accounted for by task and 32% by device in this region.

Trajectory Analysis

Our research hypothesis implies that the nature of the task influences how it is perceived which, in turn, alters how the task is performed. To examine this notion, we investigated whether a plot of the trajectory taken by the subject performing a match reflects the perceptual structure of the task. The theory of perceptual structure suggests the shape of the trajectory for each task type; that is, integral stimuli exhibit an Euclidean pattern (subjects' movement cuts diagonally across the dimensions) and separable stimuli, a city-block pattern (movement occurs along one dimension at a time). Specifically, in our tasks, if z (size or greyscale) is traversed along with x and/or y (position), then the trajectory follows a Euclidean pattern. If z is manipu-
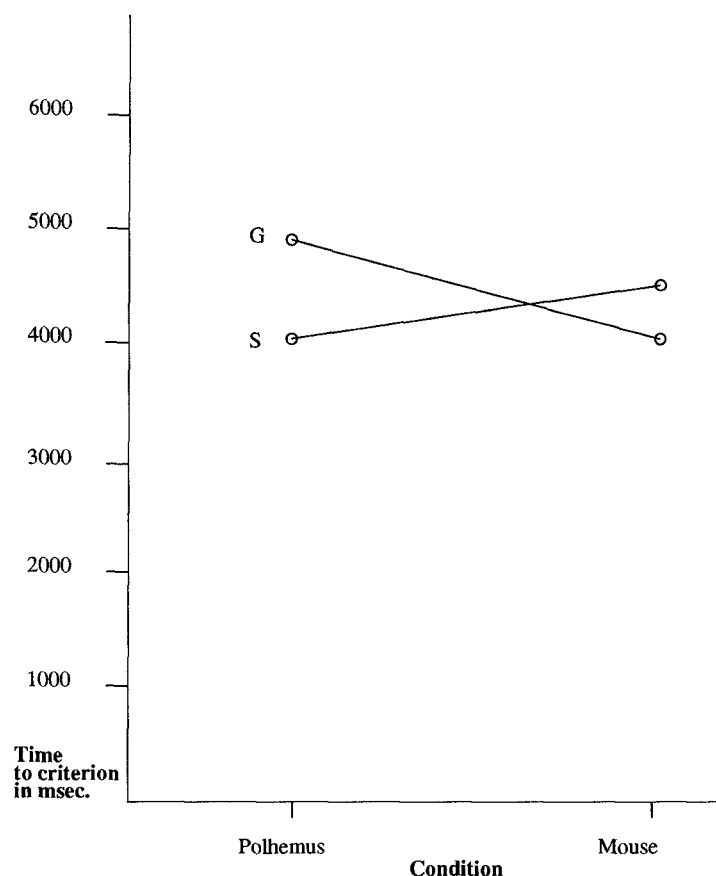
Fig. 10.   Graph of mean time to 0.099-inch criterion in msec., illustrating interaction effect. Line marked **S** shows performance on the integral (size) task, **G**, the separable (grey) task.

	Device			
	Integral (Polhemus)		Separable (Mouse)	
Task	Mean (ms.)	Std. dev.	Mean (ms.)	Std. dev.
Integral (Size)	1442	438	2113	558
Separable (Grey)	2147	416	2929	688

Fig. 11.   Time in msec. to reach 0.987-inch criterion accuracy.

lated separately from x and y in a stair-step manner, then it follows a city-block pattern.

To demonstrate this requires two conditions, both of which were present in the two Polhemus conditions of this experiment. First, the subject's range of motion must not be constrained; both tasks must be performed with a device
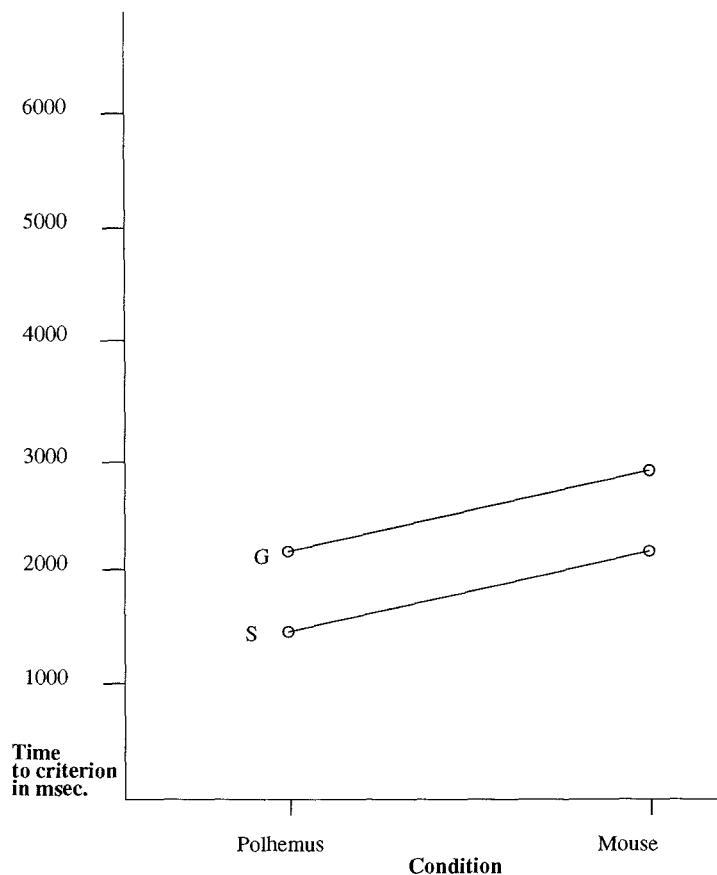
Fig. 12. Graph of mean time to 0.987-inch criterion in msec., illustrating lack of interaction effect. Line marked **S** shows performance on the integral (size) task, **G**, the separable (grey) task.

that allows freedom to move in any direction. The Polhemus meets this requirement. In contrast, the combination of two-dimensional mouse movements and one-dimensional slider does not because it constrains the subject to city-block movement by not allowing all dimensions to be manipulated concurrently. Second, the tasks should differ only minimally to isolate key differences. In this experiment, the software driving the two tasks was identical except for how one parameter was visually presented. In both tasks, the x and y dimensions were displayed as position coordinates. The difference lay solely in how the z dimension was displayed, as either size or greyscale.

The pictures in Figures 13 and 14 show the average performance of all subjects on one selected trial out of the 80 total scored trials. Like-numbered trials in each condition had the same distance and so could be compared. The trial presented is one of the longer ones (5.885 inches out of the range 2.946 inches to 6.875 inches) and one of the last completed in a session (number 74
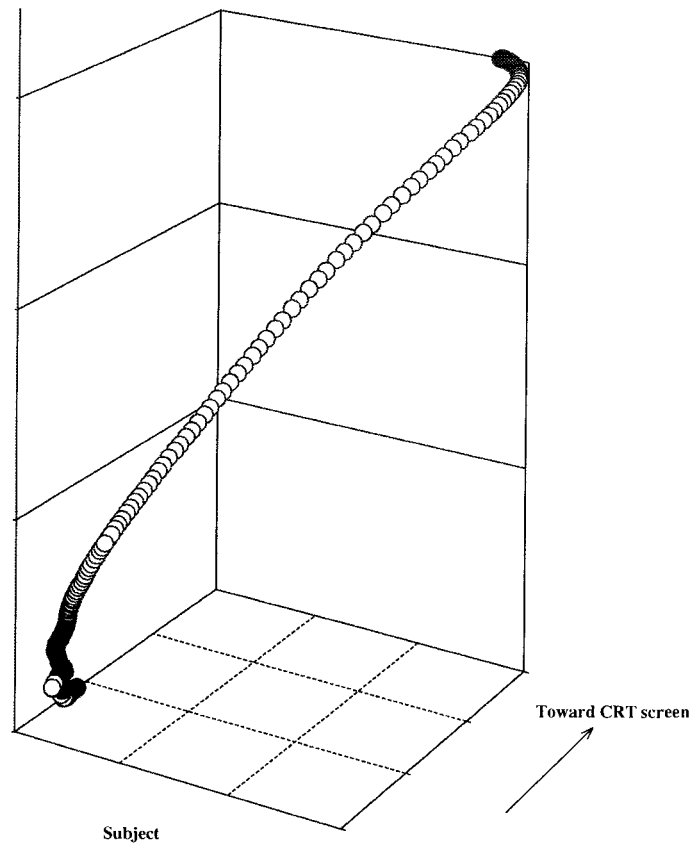
Fig. 13. Picture of average Polhemus trajectory for all subjects on one trial in the integral (size) condition. Trajectory shows that subjects cut across all three axes in an integral fashion. The trial begins at the upper right corner. Fine tuning is the small *knot* seen at the end of the trial (to the left).

out of 80). This trial was selected because it clearly illustrates the observed difference in subject performance between tasks. The Polhemus trajectory for the size task on this trial (Figure 13) suggests that the subject was cutting across all three axes. In contrast, the greyscale task (Figure 14) was completed first by resolving the discrepancy in x and y followed by the discrepancy in z. In other words, we do not see simultaneous motion in the x-y plane and along the z axis. These tasks differed only in how z was displayed. The perceptual structures of the tasks were thus strong enough to alter how subjects performed them physically. Other trials that were sampled exhibited a similar but less pronounced effect, that is, smooth, free-form, integral characteristics with the size task and stair-step, separable characteristics with the greyscale task.

While Figures 13 and 14 visually suggest the effect of perceptual space on trajectory for a selected trial, we wish to measure it in a more objective way,
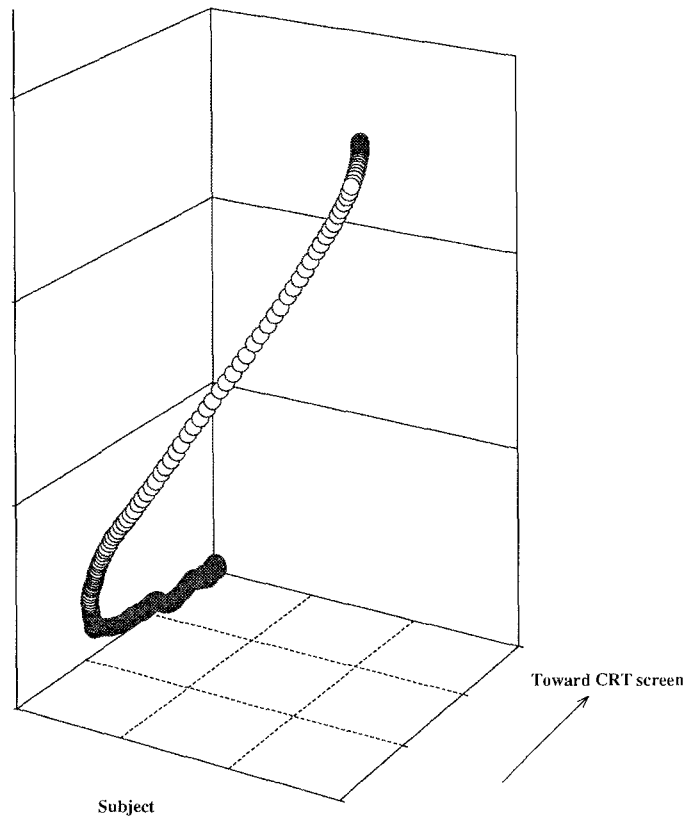
Fig. 14. Picture of average trajectory for all subjects on one trial in the separable (greyscale) condition. Trajectory shows that subjects did not move simultaneously in the x-y plane and along the z axes, consistent with separable performance. The trial begins from the top right and moves down to the left.

over all trials. We therefore developed a technique to quantify how a subject manipulated the Polhemus during a trial. The idea is to divide a trajectory into small segments, ask if the movement in each segment is Euclidean or city-block, and then tally the proportion of Euclidean-versus-city-block segments. The procedure involves four steps:

—First, the raw event data series was transformed into a time series with a 10-ms. period. It was then smoothed using a 15-point low-pass filter to remove high-frequency equipment noise and hand tremor.

—Second, the data were truncated to isolate the area of interest. A subject's behavior on a trial typically divides into two parts: strong, quick movements toward the target followed by back-and-forth fine-tuning behavior. The latter can be seen in the small *knot* at the left in Figure 13; the view

angle in Figure 14 obscures this behavior. We truncated each trial at 0.3 inches from the target to remove this highly variable end part. (As discussed below, the sensitivity of our results to this specific choice of cutoff value was very small; other choices yielded similar results.)

—Third, the data for a trajectory were segmented, and each trajectory was passed through a classification algorithm that labeled the dimensions as having changed location (movement) or not (no movement). A position change of more than 0.0008 inches in one 10-ms. time step was considered movement. (Again, as discussed below, the sensitivity of the results to the choice of this threshold value was small.)

—Fourth, a recognition algorithm computed the amount of Euclidean and city-block movement used to complete that trial. A segment was classified as Euclidean if its trajectory showed movement in more than one dimension (except for movement in the x-y plane, since the two tasks are both integral in this respect); it was classified as city-block if it showed movement in any single axis or in the x-y plane. The ratio of Euclidean to city-block segments was then calculated for each of the experimental conditions.

For the selected criterion of 0.3 inches and threshold of 0.008 inches, the average ratio of Euclidean to city-block behavior was 1.408 for the size task versus 1.234 for the greyscale task. A one-tailed pair comparison t-test showed a highly significant difference ($t(39) = 4.297$, $p < 0.0001$), supporting our hypothesis that the size task is completed in a more integral manner than the greyscale task. We performed a sensitivity analysis to give confidence in our choice of the two parameters, criterion from the truncation step and threshold from classification by computing a t-value for a range of criterion and threshold pairs. A series of one-tailed paired comparison t-tests of these pairs was significant throughout the range 0.217 to 0.38 (criteria) and 0.003 to 0.015 (threshold).

The results confirm that the way in which a user physically manipulates the Polhemus to complete a task depends on the perceptual structure of the task. The integral task induces more motion that cuts across all three dimensions. The separable task alternates more between changes to location and changes to greyscale.

APPLICATION

An example application of this work would be in designing controls for zooming and panning of a geographic display. Zooming and panning, taken together, involve three degrees of freedom. The most common design uses a mouse or trackball for two-dimensional panning and a separate control for zooming. We claim that a user typically does not really think of zooming or panning operations separably, but thinks rather of integral operations like "focus in on *that* area over there." The space is thus Euclidean, like that of the size task in the experiment, and, therefore, making the user do the two

separately violates perceptual compatibility. It would be more natural to permit a user to make a gesture that performs the overall operation he or she had in mind, using an integral three-dimensional input device. The user moves the puck around in a volume directly in front of the display screen. Moving it in the x or y direction parallel to the display surface causes panning; moving it perpendicular to the display (directly toward or away from it) causes zooming. The user typically moves the puck in all three dimensions simultaneously, resulting in some combination of zooming and panning and directly reaches the view of interest. We have demonstrated a mockup of this application.

CONCLUSIONS

Our research hypothesis is that performance improves when the structure of the perceptual space of a graphical interaction task mirrors that of the control space of the input device. We have presented an experiment to test this hypothesis and three different analyses in support. The first looked at time and accuracy at button press. Button press captures a subject's behavior when he or she is satisfied with the match. These outcome measures describe terminal actions and give insight into performance at a subject's maximum accuracy criterion. The results support our hypothesis that completing the integral size task with the integral Polhemus and the separable greyscale task with the separable mouse plus mode change lead to faster performance. The second approach removes the variation caused by individual differences in accuracy as well as speed–accuracy tradeoff and allows us to simulate a range of different experiments retroactively. We studied performance of a series of fixed criteria that combine time and accuracy into one measure. This approach allowed us to investigate performance at a number of criteria to determine the stability of performance over time. We found that within the range of plausible matches, completion times are faster when the structure of the device and task match rather than for one device or one task uniformly. The third analysis examined the trajectory of the Polhemus on both tasks. The research hypothesis suggests that if the device used by the subject is not restrictive, the path taken will be influenced by the structure of the perceptual space of the task. Our analysis confirms that the perceptual structure of the task determines how an input device is used.

The overall conclusion confirms our prediction that choosing an input device for a task requires looking at the deeper perceptual structure of the task, the device, and the interrelationship between task and device. We have shown that perceptual structure determines how a user approaches an interactive manipulation task. Tasks in which the perceptual structure is integral operate by similarity and follow the Euclidean metric. The attributes of an integral object do not dominate and are viewed as a unitary whole. They are manipulated as a unit. Tasks set in a separable space operate by dimensional structure and obey the city-block metric. The attributes of a separable object cannot be ignored and are manipulated along each attribute

in turn. If the input device supports the type of motion required by the task, then the task can be performed in an efficient manner. If the device limits necessary motion or does not restrict motion appropriately, efficiency can decrease. The interplay between task and device was more important in determining performance than either task or device alone.

Current input device taxonomies and other frameworks developed to aid device selection have typically started from the point of view of input device structure. They recognize the need for what Buxton calls pragmatics, Mackinlay, Card, and Robertson term expressiveness and effectiveness ratings, and Bleser includes in her input model, but they relegate these crucial pragmatic aspects of haptic input to rules based primarily on ad hoc testing or expert judgment. We suggest incorporating a predictive theoretical framework to allow formal reasoning about selecting an input device. We offer the approach of extending a perceptual theory and provide as an example our extension of the processing of perceptual structure to interactive stimuli.

ACKNOWLEDGMENTS

REFERENCES

ACHILLE, L. B.   1990.   Considerations in the design and development of a human computer interaction laboratory. NRL Rep. 9279, Naval Research Laboratory, Washington, D.C.

ATTNEAVE, F.   1950.   Dimensions of similarity. *Am. J. Psychol. 63*, 516–556.

BLESER, T. W.   1991.   An input device model of interactive systems design. Doctoral Dissertation, The George Washington Univ., Washington, D.C.

BLESER, T. W. AND SIBERT, J. L.   1990.   Toto: A tool for selecting interaction techniques. In *Proceedings of the ACM UIST '90 Symposium on User Interface Software and Technology*. ACM Press, New York, 135–142.

BUXTON, W.   1986.   There's more to interaction than meets the eye: Some issues in manual input. In *User Centered System Design: New Perspectives on Human-Computer Interaction*. Lawrence Erlbaum, Hillsdale, N.J., 319–337.

CARD, S. K., MACKINLAY, J. D., AND ROBERTSON, G. G.   1991.   A morphological analysis of the design space of input devices. *ACM Trans. Inf. Sys. 9*, 2, 99–122.

CARD, S. K., MORAN, T. P., AND NEWELL, A.   1983.   *The Psychology of Human-Computer Interaction*. Lawrence Erlbaum, Hillsdale, N.J.

FOLEY, J. D., WALLACE, V. L. AND CHAN, P.   1984.   The human factors of computer graphics interaction techniques. *IEEE Comput. Graph. Appl. 4*, 11, 13–48.

GARNER, W. R.   1974.   *The Processing of Information and Structure*. Lawrence Erlbaum, Potomac, Md.

GARNER, W. R. AND FELFOLDY, G. L.   1970.   Integrality of stimulus dimensions in various types of information processing. *Cog. Psychol. 1*, 225–241.

GSPC.   1977.   Status report of the Graphics Standards Planning Committee. *Comput. Graph. 11*.

HANDEL, S. AND IMAI, S.   1972.   The free classification of analyzable and unanalyzable stimuli. *Percep. Psychophy. 12*, 108–116.

IMAI, S. AND GARNER, W. R.   1968.   Structure in perceptual classification. *Psychonom. Monograph Suppl. 2*, 9, Whole No. 25.

JACOB, R. J. K. AND SIBERT, L. E.   1992.   The perceptual structure of multidimensional input device selection. In *Proceedings of the ACM CHI '92 Human Factors in Computing Systems Conference*. ACM Press, New York, 211–218.

MACKINLAY, J. D., CARD, S. K., AND ROBERTSON, G. G.   1990.   A semantic analysis of the design space of input devices. *Hum. Comput. Interact. 5*, 145–190.

SHEPARD, R. N.   1964.   Attention and the metric structure of the stimulus space. *J. Math. Psychol.* 1, 54–87.