Incorporating Perceptual Task Effort into the Recognition of Intention in Information Graphics

Stephanie Elzer[1], Nancy Green[2], Sandra Carberry[3], and James Hoffman[4]

[1] Dept of Computer Science, Millersville Univ., Millersville, PA 17551 USA
elzer@cs.millersville.edu
[2] Dept of Computer Science, Univ. of Delaware, Newark, DE 19716 USA
[3] Dept of Mathematical Sciences, Univ. of North Carolina at Greensboro, Greensboro, NC 27402, USA
[4] Dept of Psychology, University of Delaware, Newark, DE 19716 USA

Abstract. The rapidly increasing availability of electronic publications containing information graphics poses some interesting challenges in terms of information access. For example, visually impaired individuals should ideally be provided with access to the knowledge that would be gleaned from viewing the information graphic. Similarly, digital libraries must take into account the content of information graphics when constructing indices. This paper outlines our approach to recognizing the intended message of an information graphic, focusing on the concept of perceptual task effort, its role in the inference process, our rules for estimating effort, and the results of an eye tracking experiment conducted in order to evaluate and modify those rules.

1 Introduction

Information graphics (line graphs, bar charts, etc.) are pervasive in popular media such as newspaper and magazine articles. The rapidly increasing availability of electronic publications poses some interesting challenges in terms of information access. For example, individuals with impaired eyesight have limited access to graphical displays, thus preventing them from fully utilizing available information resources. Information graphics also provide a challenge when attempting to search the content of mixed-media publications within digital libraries.

Our research involves recognizing the graphic designer's communicative intention for a particular information graphic. Our analysis of a corpus of information graphics from popular media sources indicates that information graphics generally have a communicative goal and that this intended message is often not conveyed by accompanying text. Thus recognizing the intended message of an information graphic is crucial for full comprehension of a mixed-media resource. Our project's overall goal is two-fold: 1) to provide alternative access to information graphics for visually impaired users and 2) to provide access to publications in digital libraries via the content of information graphics. For visually impaired users, we are designing an interactive natural language system that provides an initial summary that includes the information graphic's intended message along with notable features of the graphic, and then responds to follow-up questions from the user [1]. For digital libraries, the initial summary of the graphic will be used in conjunction with the document text to provide a more complete representation of the content of the document to be used for searching and indexing.

Although some projects have attempted to make images accessible to visually impaired viewers by reproducing the image in an alternative medium, such as soundscapes [14], these approaches are ineffective with complex information graphics; moreover, they require the user to develop a "mental map" of the information graphic, which puts congenitally blind users at a disadvantage since they do not have the personal knowledge to assist them in the interpretation of the image [8]. The underlying hypothesis of our work is that alternative access to what the graphic looks like is not enough — the user should be provided with the message and knowledge that one would gain from viewing the graphic in order to enable effective and efficient use of this information resource.

This paper first outlines our overall approach to inferring the communicative message of an information graphic as well as various types of evidence (caption, highlighting, and a user model) that can aid the inference process. It then focuses on one specific type of evidence, perceptual task effort, discusses its role in recognizing the graphic's intended message, describes our rules for estimating perceptual task effort, and presents the results of an eye tracking experiment conducted in order to evaluate and revise our effort estimates.

2 Recognizing the Graphic Designer's Intended Message

As Clark [3] noted, language is more than just words. It is any "signal" (or lack of signal when one is expected), where a signal is a deliberate action that is intended to convey a message. Language research has posited that a speaker or writer executes a speech act whose intended meaning he expects the listener to be able to deduce, and that the listener identifies the intended meaning by reasoning about the observed signals and the mutual beliefs of author and interpreter [6, 3]. Applying Clark's view of language to information graphics, it is reasonable to presume that the author of an information graphic similarly expects the viewer to deduce from the graphic the message that he intended to convey by reasoning about the graphic itself, the salience of entities in the graphic, and mutual beliefs.

Beginning with the seminal work of Allen [16] who developed a system for deducing the intended meaning of an indirect speech act, researchers have applied plan inference techniques to a variety of problems associated with understanding utterances, particularly utterances that are part of a dialogue. Given domain knowledge in the form of operators that decompose goals into a sequence of subgoals, along with evidence in the form of an observed action (such as an utterance), a plan inference system chains backwards on the plan operators to deduce one or more high-level goals that might have led the agent to perform the observed action as part of an overall plan for achieving his goal(s). The high-level communicative goals in the plan capture the utterance's intended meaning.

In their work on intelligent multimedia generation, the AutoBrief group proposed that speech act theory can be extended to the generation of graphical presentations[9]. When designing an information graphic, the designer has one or more high-level communicative goals. Consequently, he constructs an information graphic that he believes will enable the viewer to perform certain perceptual and cognitive tasks which, along with other knowledge, will enable the viewer to recognize the intended message of the graphic [9]. By *perceptual tasks* we mean tasks that can be performed by simply viewing the graphic, such as finding the top of a bar in a bar chart; by *cognitive tasks* we mean tasks that are done via mental computations, such as computing the difference between two numbers.

In our research, we extend plan inference techniques (that have been used successfully on natural language discourse) to inferring intention from information graphics. Our plan operators capture knowledge about how the graphic designer's goal of conveying a message can be achieved via the viewer performing certain perceptual and cognitive tasks, as well as knowledge about how perceptual and cognitive tasks decompose into sets of simpler tasks. Using these plan operators, we can chain from evidence provided by the information graphic to eventually reach a high-level goal that captures the underlying message of the graphic in the same way that plan inference systems chain from a speech act to the probable goals of an utterance. Input to our plan recognition system consists of an XML representation of the graphic as provided by a vision module [1].

In extending plan inference techniques to the recognition of intentions from information graphics, we need to identify the types of evidence that will be used in the plan inference process. In plan recognition systems involving dialogue, the evidence is naturally centered around the utterances, and the inference process proceeds incrementally as the dialogue unfolds, using evidence such as the surface form of the utterance, the focus of attention in the dialogue, etc. When dealing with information graphics, the viewer is presented with the entire information graphic, and a decision needs to be made as to which aspects of the graphic should be used as evidence of the graphic designer's intentions. Following AutoBrief [9], we contend that when constructing the graphic, the designer made certain design decisions in order to make "important" tasks (the ones that the viewer is intended to perform in getting the graphic's message) as *easy* or as *salient* as possible. By reasoning about these design decisions, we can glean information about the graphic designer's intended message for the graphic. The graphic designer can make a task *easy* for the viewer to perform by the choice of graphic type (for example, bar chart versus line graph[20]) and the organization and presentation of data. This observation has led us to include perceptual task effort as one of the sources of evidence in our plan inference process; this particular type of evidence is the focus of this paper, and is discussed further beginning in Section 3. The graphic designer might also intend a task to be particularly *salient* to the viewer. We have identified three sources of evidence which allow us to reason about the tasks that the graphic designer intended to be salient for the viewer: captions, highlighted entities in the information graphic, and a model of mutual beliefs about entities of interest to members of the viewing audience.

Well-chosen captions can be useful indicators of the intended message of an information graphic. Consider, for example, the graphic on the left in Figure 4. If this graphic had the caption "Penny Pinching in 2000," this would indicate that the bar representing 2000 is a particularly salient item in the graphic, whereas if the caption read "Capital Expense Peaks in 1999" this would indicate the salience of the bar representing 1999 and the task of finding the maximum in the graph. Therefore, we use noun phrases in captions as an indication of the salience of particular items in the graphic, and verb phrases to indicate the salience of particular tasks. One might wonder why we do not deal almost exclusively with captions to infer the intentions of the information graphic. Corio[5] performed a large corpus study of information graphics and noted that captions often do not give any indication of what the information graphic conveys. Our examination of a collection of graphics supports his findings. Thus we must be able to infer the message underlying a graphic when captions are missing or of little use.

Graphic designers also use techniques to highlight particular aspects of the graphic, thus making them more salient to the viewer. Such techniques include the use of color or shading for elements of a graphic, annotations such as an asterisk, an arrow pointing to a particular location, or a pie chart with a single piece "exploded." Our working hypothesis is that if the graphic designer goes to the effort of employing such attention-getting devices, then the highlighted items are almost certainly part of the intended message. Thus we treat the highlighted entities as suggesting instantiations of primitive perceptual tasks that produce particularly salient tasks. Suppose for example that there was no caption on the information graphic shown on the left in Figure 4, but that the bar for 2000 was highlighted by shading it darker than the other bars. This suggests that this bar is particularly relevant to the intended message of the graphic. Consequently, we use the attributes of the bar (such as its label) to instantiate primitive perceptual tasks and produce tasks that are hypothesized to be salient.

A model of the intended recipient of the information graphic also plays a role in the plan recognition process. In designing the information graphic, the graphic designer takes into account mutual beliefs about entities that will be particularly salient to his audience. For example, if an information graphic appears in a document targeted at residents of Cambridge, then both the designer and the viewer will mutually believe that entities such as Cambridge, its sports teams, etc. will be particularly salient to the viewer. Our viewer model captures these beliefs, and our approach is to treat them in a manner similar to the way in which we handle noun phrases in captions.

3 Estimating Perceptual Task Effort

Given a set of data, the graphic designer has many alternative ways of designing a graphic. As Larkin and Simon note, information graphics that are *informationally* equivalent (all of the information in one graphic can also be inferred from the other) are not necessarily *computationally* equivalent (enabling the same inferences to be drawn quickly and easily) [12]. Peebles and Cheng further observe that even in graphics that are informationally equivalent, seemingly small changes in the design of the graphic can affect viewers' performance of graph reading tasks[15]. Much of this can be attributed to the fact that design choices made while constructing an information graphic will facilitate some perceptual tasks more than others. Following the AutoBrief work on generating graphics to achieve communicative goals, we hypothesize that the designer chooses a design that best facilitates the tasks that are most important to conveying his intended message, subject to the constraints imposed by competing tasks [9].

In order to identify the perceptual tasks that the graphic designer has best enabled in the graphic, our methodology is to apply the results of research from cognitive psychology to construct rules that estimate the effort required for different perceptual tasks within a given information graphic. Our working hypothesis is that the *easiest* tasks are good candidates for tasks that the viewer was intended to perform, since the designer went to the effort of making them easy to accomplish. We can then use this set of the easiest perceptual tasks along with any unusually salient tasks as a starting point for our inference process. By reasoning about the more complex tasks in which these perceptual tasks play a role, we can hypothesize the message that the graphic designer intended the viewer to extract from the graphic. The component of our system that is responsible for estimating effort is called APTE (Analysis of Perceptual Task Effort).

3.1 Analysis of Perceptual Task Effort

The goal of APTE is to determine whether a task is easy or hard with respect to other perceptual tasks that could be performed on an information graphic. In order to estimate the relative effort involved in performing a task, we adopt a GOMS-like approach [2], decomposing each task into a set of component tasks. Following other cognitive psychology research, we take the principal measure of the effort involved in performing a task to be the amount of time that it takes to perform the task, and our effort estimates are based on time estimates for the component tasks. In this sense, our work follows that of Lohse [13] in his UCIE system, a cognitive model of information graphic perception intended to simulate and predict human performance on graphic comprehension tasks. However, we are not attempting to develop a predictive model of our own – our aim is to identify the tasks that the designer would expect to have best facilitated by his design choices in order to utilize that information in the plan inference process.

Structure of Rules APTE contains a set of rules that estimate how well a task is enabled in an information graphic. Each rule captures a perceptual task that can be performed on a particular type of information graphic (line graph, bar chart, etc.), along with the conditions (design choices) that affect the difficulty of performing that task. The conditions for the tasks are ordered so that the conditions producing the lowest estimates of effort appear first. Often several conditions within a single rule will be satisfied – this might occur, for example, in the rule shown in Figure 2 which estimates the effort of determining the exact

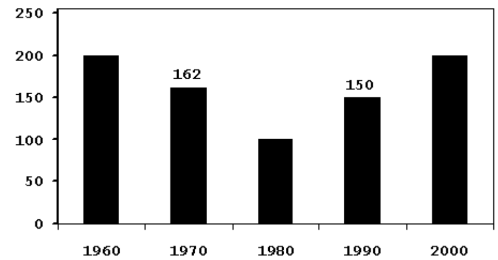

Fig. 1. Information Graphic Example

value represented by the top of a bar in a bar chart. Condition-computation pair B1-1 estimates the effort involved when the bar is annotated with the value; this condition is illustrated by the second and fourth bars in Figure 1. The second condition-computation pair, B1-2, is applicable when the top of the bar aligns with a labelled tick mark on the dependent axis; this condition is illustrated by all bars except the second bar in Figure 1. If the top of a bar both falls on a tick mark and has its value annotated at the top of the bar (as in the fourth bar in Figure 1), the easiest way to get the value represented by the top of the bar would be to read the annotated value, although it could also be obtained by scanning across to the tick mark. When multiple conditions are applicable, the first condition that is satisfied will be applied to calculate the effort estimate, thereby estimating the least expected effort required to perform the task.

Rule-1:Estimate effort for task
Perceive-dependent-value(<viewer>, <g>, <att>, <e>, <v>)
Graphic-type: bar-chart
Gloss: Compute effort for finding the exact value <v> for attribute <att>
represented by top <e> of a bar <b> in graph <g>
B1-1: IF the top <e> of bar <b> is annotated with a value,
THEN effort=150 + 300
B1-2: IF the top <e> of bar <b> aligns with a labelled tick mark on
the dependent axis, THEN effort=scan + 150 + 300

Fig. 2. A rule for estimating effort for the perceptual task *Perceive-value*

Developing Effort Estimates Researchers have examined many different perceptual tasks, although often studying individual perceptual tasks in isolation. As mentioned earlier, we have followed Lohse's approach [13] in breaking down our tasks into component tasks. We then utilize existing time estimates (primarily those applied in Lohse's UCIE system) for the component tasks wherever possible. For some perceptual tasks, this has been a sufficient foundation for our rules. For example, we developed effort estimates for the rule shown in Figure 2 in this manner. In the case of condition-computation pair B1-1 (finding the exact value for a bar where the bar is annotated with the value), the effort is estimated as 150 units for discriminating the label (based on work by Lohse [13]) and 300 units for recognizing a 6-letter word [7]. In the case of B1-2 (finding the exact value for a bar where the top of the bar is aligned with a tick mark on the axis), the effort estimate includes scanning over to the dependent axis (measured in terms of distance in order to estimate the degrees of visual arc scanned [11]) in addition to the effort of discriminating and recognizing the label.

For more complex tasks that have not been explicitly studied by cognitive psychologists, we have applied existing principles and laws in the development of our rules for estimating perceptual effort. An example of this is the class of comparison tasks (for example, comparing the tops of two bars to determine the relative difference in value), where the proximity compatibility principle defined by Wickens and Carswell [19] plays a major role. This principle is based on two types of proximity; *perceptual proximity* refers to how perceptually similar two elements of a display are (in terms of spatial closeness, similar annotations, color, shape, etc.) while *processing proximity* refers to how closely linked the two elements are in terms of completing a particular task. If the elements must be used together (integrated) in order to complete a task, they have close processing proximity. The proximity compatibility principle states that if there is close processing proximity between two elements, then close perceptual proximity is advised. If two elements are intended to be processed independently, then distant perceptual proximity is advised. Violating the principle will increase the effort required for a viewer to process the information contained in the display.

Rule-3: Estimate effort for task
    Perceive-relative-diff(<viewer>, <g>, <e1>, <e2>, <b1>, <b2>, <r>, <d>)
Graphic-type: bar-chart
Gloss: Compute effort for finding the relative difference <r> in value (greater than /less than/equal to) and degree <d> of difference (high/low/none) represented by the tops <e1> and <e2> of two bars <b1> and <b2> in graph <g>
  B3-1: IF bar <b1> and bar <b2> are adjacent and the height difference is >10% THEN effort=92 + 230 + 150
  B3-2: IF bar <b1> and bar <b2> are not adjacent and the height difference is >10% THEN effort=92 + 460 + 150
  B3-3: IF bar <b1> and bar <b2> have height difference >5% THEN effort=92 + 920 + 150

Fig. 3. A rule for estimating effort for perceptual task *Perceive-relative-diff*

We assume that the graphic designer attempted to follow the proximity compatibility principle in designing the information graphic so as to facilitate intended tasks and make them easier to perform than if the principle were violated. This assumption is reflected in the rule in Figure 3, where the effort required to perform the integrated task of determining the relative difference between two bars is different based on the bars' spatial proximity. For adjacent bars, the effort required will generally be lower than if the bars were not adjacent.

Weber's Law [4] has also played a critical role in our rules. Many of the tasks for which we have had to develop effort estimates involve discriminating between two or more graphical elements; these tasks require the viewer to make comparative judgments of length, area, and angle. In order to define the conditions affecting the complexity of these judgments, we have applied Weber's Law

[4]. One of the implications of Weber's Law is that a fixed percentage increase in line length or area is required to enable discrimination between two entities (and the probability of discrimination is affected not by object size, but by the percentage increase). Weber's Law has influenced the thresholds used in rules for estimating effort such as *Rule-3* in Figure 3 where thresholds in the percentage difference in the height of the bars influence the effort required to perceptually discriminate the relative difference between the values represented by the bars.

In some cases, the optimal combination of component tasks does not take into account the escalating complexity represented by the conditions of the rule. For example, our eye-tracking experiments showed that viewers performed an average of four saccades if the bars to be compared differ in height by 5% to 10% and an average of two saccades if the non-adjacent bars' height difference was greater than 10%. In both cases, one saccade (from the top of the lowest bar to the top of the highest bar) would be optimal in the sense of providing the necessary information. Our rules capture the expected number of saccades required by the average viewer in order to perform the necessary perceptual judgment. The effort estimates in Figure 3 show the estimate of 92 units to perform a perceptual judgment [18] along with a multiple of 230 units where 230 represents the estimate for a saccade [17] and the effort of discriminating the top of the higher bar (150 units based on [13]). The conditions and estimates in *Rule-3* (Figure 3) reflect the results of our experiment (described in Section 4). The eye-tracking data guided the development of the thresholds and showed that when the height difference was small (between 5% and 10%), the bars' adjacency did not have a discernable effect on the number of required saccades.

3.2 Applying Effort Estimates

After applying the APTE rules to identify the set of the easiest perceptual tasks for a given information graphic, and then reasoning about the more complex tasks in which the these perceptual tasks play a role, we can hypothesize the message that the graphic designer intended the viewer to extract from the graphic. Consider, for example, the graphic shown in Figure 1. Because the graphic designer has chosen to annotate the two bars representing 1970 and 1990 with their exact values, the task of perceiving the exact value for these two bars will appear in the set of the easiest perceptual tasks for this graphic. We then infer that higher-level (more complex) tasks that include these tasks as subgoals are good candidates to represent the possible communicative intention of the graphic designer. A task that would be considered a good candidate in this example is comparing the values represented by the two bars, since this task not only includes two of the easiest perceptual tasks, but the instantiation of the parameters in this task are appropriate according to the proximity compatibility principle since the two bars with close processing proximity (the two bars being compared) have close perceptual proximity (are both annotated with their values). Other lesser candidates would include finding the relative difference in capital expense in 1970 versus 1980, 1980 versus 1990, and 1990 versus 2000. These candidates would be supported by the fact that the tasks of finding the relative difference between these pairs of bars are in the set of easiest perceptual tasks for the graphic, since the pairs of bars are adjacent and have large percentage differences in height. If other evidence, such as a helpful caption, highlighting techniques or a relevant user model are available, this evidence will also be taken into account in the inference process.

4 Evaluating and Modifying APTE

This section describes an eye tracking experiment that was conducted to evaluate the APTE rules for bar charts and to suggest revisions to these rules. A set of rules describing tasks in which we were interested was developed based on the cognitive principles described in Section 3.1. Information graphics were then designed to test the various conditions of the tasks. The results from the experiment were used both to verify that the cognitive principles that guided the development of the rule set were appropriately applied and to suggest modifications to individual conditions of the rules within the rule set.

4.1 Method

Eleven participants[5] were asked to perform various tasks using vertical (column) bar charts shown to them on a computer monitor while their eye fixations were recorded. Each task was completed by seven of the participants. Examples of the tasks include finding the bar representing the maximum value in the bar chart and finding the exact value represented by the top of a particular bar in a bar chart. For each task, participants were shown a screen with the instructions for the task displayed. The instructions for each task included some specific action to be taken by the participants to indicate the results of the task. These actions fell into two categories; in the first category, the result of the task was indicated by the participants clicking on an element of the information graphic, while results of tasks in the second category were indicated by the participants clicking on one of three buttons shown below the information graphic. Both categories of tasks included a mouse movement and a mouse click, and the time of the start of the mouse movement and the time and location of the mouse click were recorded as part of the data collected during the experiment. When the participants had read and felt that they understood the instructions, they clicked the mouse. The next screen that the participants were shown contained only a fixation point. After clicking on the fixation point, the participants were shown the bar chart on which they were to perform the prescribed task. The participants moved to the next task by clicking on a "Done" button shown at the bottom right corner of the screen.

4.2 Design

The experiment was designed to obtain the average time required to complete a given task across participants. Six bar charts were constructed that displayed a

[5] A twelfth participant could not be calibrated on the eye tracker.

variety of different characteristics (increasing versus decreasing trends, varying numbers of bars, sorted versus unsorted labels, bars sorted by height or unsorted, etc.). We call these six bar charts the “base” bar charts, since the actual bar charts used in the experiment were variants of these. The APTE rule set for bar charts currently contains ten rules describing various perceptual tasks that can be performed using a bar chart. For a given base bar chart, only a subset of the rules and their conditions could be analyzed. For example, there are three rules describing trends – one for increasing trends, one for decreasing trends and one for stable trends. In this experiment, each base bar chart contained only one trend, so at least two of the rules would not apply to a given base bar chart. However, other rules might have multiple conditions that could all apply to a given base bar chart (for example, applying the rule shown in Figure 3 to different pairs of bars). The set of tasks being evaluated varied between base bar charts, but always included at least one condition of each applicable rule.

In order to prevent participants from becoming familiar with the six base bar charts being analyzed, the actual test graphics were variants of the base bar charts. In designing the test graphics, characteristics of the base bar chart that were extraneous to the task being evaluated were altered. For example, if the task was to locate and read the label of a given bar in the base bar chart, the attribute name displayed on the y-axis and the heights of the bars not involved in the task would be altered in the test graphic (see Figure 4). The order in which the participants completed the tasks was also varied so as to avoid effects of familiarity with the content of the specific information graphics and expertise obtained through practice in performing the requested tasks.

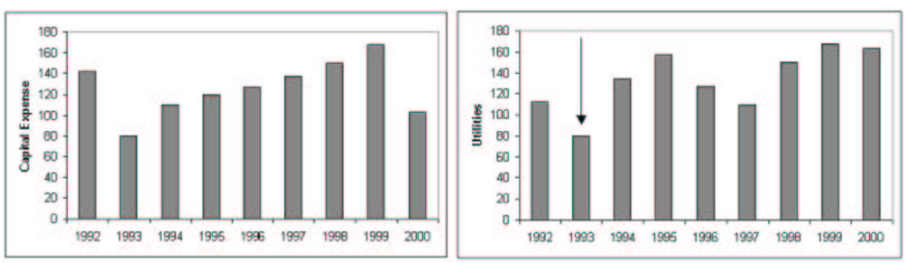

Fig. 4. Base bar chart example (left) and test graphic for get-label (right)

4.3 Procedure

Each trial began with a series of five practice tasks. Participants were informed that the first five tasks were for practice, and were allowed to ask any questions about the format of the experiment during the warmup period. Participants were given several points of instruction. For tasks that required participants to choose an answer shown on one of three response buttons, participants were instructed to look only at the graphic in completing the task, and to determine the answer to the task before reading the labels on the buttons. For all tasks, participants were asked to only move the mouse when they were ready to make a response. After the fifth practice graphic, the participants were presented with the series of tasks comprising the experiment. The participants' eye fixations were measured using an Iscan Model RK-716PCI eye tracking processor operating at 60Hz.

4.4 Data Analysis

The aim of this experiment was to obtain the average completion time of all participants for a given task, and to compare the rank order of those average completion times to the rank order of estimates produced by the APTE rules for those same tasks. The completion time for each set of data was determined based on a combination of the time of the initial mouse movement and the pattern of the participant's eye fixations. In order to obtain the best possible measure of the completion time of the task, it was determined that a combination of the initial mouse movement and the pattern of the participants' eye fixations would be used. For tasks that required the user to click on a button, task completion time was recorded as the beginning of the mouse movement if that movement was just prior to the participant moving his gaze to the region of the screen where the buttons were located. If the mouse movement did not coincide with the shift in gaze away from the graphic and towards the buttons, the task completion time was recorded as the end time of the final fixation within the information graphic. For tasks that required the participant to click on an element of the information graphic, task completion time was recorded as the beginning of the mouse movement if that movement took place during the same fixation in which the participant "clicked" on the appropriate element. If the mouse movement did not coincide with this eye fixation, the completion time of the task was recorded as the beginning of the fixation during which the participant selected the appropriate graphical element.

Recall (Section 4.2) that variants of each base bar chart were constructed to avoid participants becoming familiar with particulars of the bar chart. The base bar charts were divided into two sets. For each set, the participants evaluated all of the tasks related to the base bar chart using the same variants. For each of the six base bar charts, the list of average task completion times[6] (over all participants) was sorted. These sorted task lists were compared to the sorted lists of the effort estimates produced by the APTE rules. The emphasis in this analysis was on the relative rank of each task within the sorted lists, rather than on the actual values. Any discrepancies in the task ordering for an individual graph were noted. Of particular interest were any perceptual tasks where the rankings did not correlate across different base bar charts. The eye fixations for these tasks were then analyzed in closer detail in order to detect patterns in the fixations that would support a change in the APTE rule for the task. Several

[6] Data was excluded from the results if the participant's gaze left the information graphic to view the labels on the buttons before completing the task, if the participant responded incorrectly to the task, or if it was clear that the participant was performing processing not required by the task.

APTE rules were modified based on the data gathered in the experiment, and these changes, along with an overall analysis of the data are discussed in the next section.

4.5 Results and Discussion

The comparison of the rankings of average completion times and the corresponding effort estimates for each of the six base bar charts showed strong support for the cognitive principles on which the APTE rules are based. For example, the hypothesis that it would require less effort to compare bars that are adjacent (based on the proximity compatibility principle [19]) was upheld, as was the application of Weber's Law [4] in developing rule conditions based on thresholds in the percentage of height difference of the bars. However, as intended, the results of the experiment also provided evidence of ways in which the APTE rules could be modified in order to improve the quality of the effort estimates produced.

One area in which we applied the results of the experiment was in the effort estimates for perceiving trends in bar charts. Our initial APTE rules for trends represented the perception of a trend in terms of simple scans of the bar chart, while the eye fixation data showed a less smooth, slower processing of the data. By altering our APTE trend rules to represent the pairwise perceptual judgments supported by the eye fixation data, we were able to more accurately assess the effort required for trend recognition.

The experiment also yielded some unexpected insights into the way in which participants process information graphics. For example, our initial APTE rule for finding the exact value represented by the top of a bar when the top of the bar is aligned with a tick mark on the axis (Figure 2) was expected to require far less effort to complete than the task of perceiving the data required to interpolate the value represented by the top of the bar when the top of the bar does not align with a tick mark. However, the patterns of eye fixations of the participants showed that when the top of the bar is aligned with a tick mark, participants frequently repeat the task (presumably to ensure accuracy). This resulted in a change to the rule shown in Figure 2 so that the effort estimate for B1-2 is now equal to 230 + (scan + 150 + 300) x 2. Similarly, in graphics where the labels along the primary key axis are unsorted, our initial APTE rule for finding the top of a bar given the bar's label described the viewer as performing a left-to-right scan and sampling of the unsorted labels. In analyzing the results of the experiment, we found that viewers were far more likely to begin in the middle of the axis and search to the left or right, then saccade to the other half of the axis if they are unsuccessful in their initial search. Without being able to analyze the patterns of participants' eye fixations, we would have been unable to capture this search process and the resultant effort estimate.

The somewhat surprising results outlined above give rise to some interesting questions about what should actually be captured by the APTE rules. As described previously, our use of the APTE rules and the resulting ranking of effort estimates reflects our hypothesis that since the graphic designer has many alternative ways of designing a graphic, the designer chooses a design that best facilitates the tasks that are most important to conveying his intended message, subject to the constraints imposed by competing tasks [9]. Underlying this hypothesis is the assumption that the graphic designer is competent, and that a competent graphic designer has a fairly accurate model of what is required to perceptually facilitate the tasks. We have based this assumption on the wealth of resources describing ways in which graphic designers can and should facilitate tasks for their viewers ([4] and [10], for example) and the observation that many of the techniques described in these resources correspond to the cognitive principles upon which we based our APTE rules. However, it is unclear that graphic designers would have an accurate model of some of the less expected results, such as the similarity in effort between reading a value from a tick mark and gathering the information necessary to interpolate the value. It seems reasonable that a graphic designer wishing to facilitate the task of determining the exact value would align the top of the bar with a tick mark rather than forcing the viewer to interpolate the value.[7] This problem of what to represent in the APTE rules accentuates the distinction between the actual effort which viewers expend in performing particular tasks versus the difficulty the viewer might reasonably be expected to have in performing the task. Viewers tend to repeat the process of locating, discriminating and reading the value on a tick mark, but the expected difficulty of that task would not include this repetition. Another example of this is the task of finding a particular label amongst the unsorted labels along the primary key axis of the bar chart. A graphic designer might reasonably place a bar first on the axis, expecting that this will reduce the difficulty of locating the bar (expecting the viewer to use a left-to-right search). In practice, it seems that viewers begin in the middle of the axis, so the task might be better facilitated by placing the bar in the middle of the bar chart.

However, in order to base the APTE rules on the expected difficulty of the tasks rather than the actual effort required to complete them, we would need to have evidence of what is contained in the stereotypical graphic designer's model of expected difficulty. Up to this point, we have been using the resources and guidelines published for graphic designers along with the assumption of competence of the graphic designer in order to draw reasonable conclusions about the knowledge of the stereotypical graphic designer. Unfortunately, the resources that we have examined do not provide guidelines for the tasks in question. Lacking evidence to support this intuitive distinction between expected difficulty and the actual effort expended in performing tasks, we have based our APTE rules on the solid evidence regarding the effort expended by participants performing these tasks that we have collected during this experiment. An interesting future area of research would be to investigate the difference between expected difficulty and actual effort and its influence on the choices made by graphic designers.

[7] Of course, following that same argument, it would seem even more reasonable that the graphic designer wishing to facilitate the finding of the exact value represented by the top of a bar would annotate the top of the bar with the value, and we did find that task to require substantially less effort than the other tasks being described.

Statistical Analysis of Correlation Having modified the APTE rules to better reflect the patterns of eye fixations demonstrated by viewers, we produced a new set of effort estimates based on the modified rules. We then performed a statistical analysis to test the correlation between the average completion times and the effort estimates produced by our rules. We performed two types of correlation tests. The Spearman Rank-Order Correlation (rho), used to determine whether two sets of rank-ordered data are related, is an especially appropriate choice for analyzing this data since we are primarily interested in ranking the effort estimates generated by APTE. However, since we use the gaps in the effort estimates generated by APTE in order to identify the set of easiest perceptual tasks, we also did a Pearson Product-Moment Correlation (r) on the actual values. Figure 5 shows the results of both correlations between the average completion times and our effort estimates for each of the six base bar charts (values approaching 1 show a strong correlation). The p-values are one-tailed and were calculated by t-approximation. The results of the Spearman Rank-Order Correlation show a very high and significant correlation between the ranking of task effort provided by the APTE rules and the actual completion times. Interestingly, the Pearson correlation also shows a very strong correlation between the average completion times from the experiment and our APTE effort estimates. We intend to run additional experiments to further validate the modified rules with new data.

Graph	Tasks	Pearson	p-value	Spearman	p-value
1	12	0.92	$<$0.0001	0.89	$<$0.0001
2	11	0.85	0.0005	0.94	$<$0.0001
3	10	0.89	0.0003	0.99	$<$0.0001
4	9	0.97	$<$0.0001	1	$<$0.0001
5	9	0.93	0.0002	0.95	$<$0.0001
6	11	0.96	$<$0.0001	0.98	$<$0.0001

Fig. 5. Correlations between average completion time and APTE effort estimates

5 Conclusion

In this paper, we have outlined our approach to a novel application of plan recognition: recognizing the intended message of information graphics. The ability to infer the intended message of a graphic plays a vital role in 1) providing alternative access to information graphics for visually impaired viewers, and 2) providing access to publications in digital libraries via the content of information graphics. Although we utilize multiple types of evidence (caption, highlighting, user model) in the intention recognition process, this paper focused on one specific type of evidence, perceptual task effort. We discussed the role of perceptual task effort in recognizing the designer's intended message, described our rules for estimating effort, and presented the results of an experiment which provides strong evidence of the correlation between the effort estimates generated by the modified APTE rules and the relative difficulty that humans have in performing the corresponding perceptual tasks. In future work, we will expand our APTE rules to encompass other graph types, such as line graphs and pie charts.

References

1. S. Carberry, S. Elzer, N. Green, K. McCoy, and D. Chester. Understanding information graphics: A discourse-level problem. In *Proc. of SigDial 2003*. Assoc. for Computational Linguistics, pp. 1–12, 2003.
2. S. K. Card, T. P. Moran, and A. Newell. *The Psychology of Human-Computer Interaction*. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, 1983.
3. H. Clark. *Using Language*. Cambridge University Press, 1996.
4. W. S. Cleveland. *The Elements of Graphing Data*. Chapman and Hall, New York, 1985.
5. M. Corio and G. Lapalme. Generation of texts for information graphics. In *Proc. of the 7th European Workshop on Natural Language Generation*, pp. 49–58, 1999.
6. H. P. Grice. Utterer's meaning and intentions. *Philosophical Review*, pp. 147–177, 1969.
7. B. E. John and A. Newell. Toward an engineering model of stimulus response compatibility. In R. W. Gilmore and T. G. Reeve, editors, *Stimulus-response compatibility: An integrated approach*, pp. 107–115. North-Holland, New York, 1990.
8. A. R. Kennel. Audiograf: A diagram-reader for the blind. In *Second Annual ACM Conference on Assistive Technologies*, pp. 51–56, 1996.
9. S. Kerpedjiev and S. F. Roth. Mapping communicative goals into conceptual tasks to generate graphics in discourse. In *Intelligent User Interfaces*, pp. 157–164, 2000.
10. S. M. Kosslyn. *Elements of Graph Design*. W. H. Freeman and Company, 1994.
11. S. M. Kosslyn. Understanding charts and graphs. *Applied Cognitive Psychology*, 3:185–226, 1989.
12. J. H. Larkin and H. A. Simon. Why a diagram is (sometimes) worth a thousand words. *Cognitive Science*, 11:65–99, 1987.
13. G. L. Lohse. A cognitive model for understanding graphical perception. *Human-Computer Interaction*, 8:353–388, 1993.
14. P. B.L. Meijer. An experimental system for auditory image representations. *IEEE Transactions on Biomedical Engineering*, 39(2):112–121, 1992.
15. D. Peebles and P. C-H. Cheng. Modeling the effect of task and graphical representation on response latency in a graph reading task. *Human Factors*, in press.
16. C. R. Perrault and J. F. Allen. A plan-based analysis of indirect speech acts. *American Journal of Computational Linguistics*, 6(3-4):167–182, 1980.
17. J. E. Russo. Adaptation of cognitive processes to eye movement systems. In J. W. Senders, D. F. Fisher, and R. A. Monty, editors, *Eye movements and higher psychological functions*. Lawrence Erlbaum Associates, Inc., Hillsdale, NJ, 1978.
18. A. T. Welford. Attention, strategy, and reaction time. In S. Kornblum, editor, *Attention and Performance IV*, pp. 37–54. Academic, New York, 1973.
19. C. D. Wickens and C. M. Carswell. The proximity compatibility principle: Its psychological foundation and relevance to display design. *Human Factors*, 37(3):473–494, 1995.
20. J. Zacks and B. Tversky. Bars and lines: A study of graphic communication. *Memory & Cognition*, 27(6):1073–1079, 1999.