3B2v7.51c GML4.3.1 IJHC : 20020527— Prod.Type: com pp.1–24(col.fig.:Fig 2 colour for online) ED:Ravi/Br PAGN: thilakam SCAN: shobha

Int. J. Human-Computer Studies (2002) **56,** 000–000
doi:10.1006/ijhc.2002.0527
Available online at http://www.idealibrary.com.on IDEAL®

Applying models of visual search to menu design†

Baili Liu
Department of Computing and Software Systems, University of Washington Bothell, Box 358534, 18115 Campus Way N.E., Bothell, WA 98011, USA.
email: bliu@bothell.washington.edu

Gregory Francis
Department of Psychological Sciences, Purdue University, USA

Gavriel Salvendy
School of Industrial Engineering, Purdue University, USA

(Received 25 March 2001 and accepted in revised form 23 January 2002)

The Guided Search (GS) model, a quantitative model of visual search, was used to develop menu designs in a four-step process. First, a GS simulation model was defined for a menu search task. Second, model parameters were estimated to provide the best fit between model predictions and experimental data. Third, an optimization algorithm was used to identify the menu design that minimized model predicted search times based on predefined search frequencies of different menu items. Fourth, the design was tested. The results indicate that the GS model has the potential to be part of a system for predicting or automating the design of menus.

KEYWORDS: visual search; human–computer interaction; menu search.

1. Introduction

Menu search tasks are repetitive tasks performed by millions of people every day. Many empirical studies have identified factors such as organization (grouping or ordering), navigation (breadth vs. depth), layout, graphics (using icons), naming and practice (Norman, 1990) that are related to search efficiency in menu search tasks. However, it is often difficult to transfer these empirical findings into design decisions during the creation of a specific menu. Guidelines resulting from these empirical findings sometimes contradict each other. It is not clear how to set priorities among different guidelines because the relative impact of different factors is not generally known and methods to systematically apply the findings are not available.

†This work was carried out at the School of Industrial Engineering, Purdue University.

1071-5819/02/000000 + 00 $35.00/0

Computational models that predict human–computer interaction (HCI) task performance could become, once validated, very useful in evaluating design alternatives or even automating design efforts (e.g. Francis, 2000; Meyer, 2000). Menu search, like other information-processing tasks, includes perceptual, cognitive and motor components. Quantitative models that could predict motor performance based on the law of Fitts (1954) have been used to optimize hierarchical menu layout (e.g. Francis, 2000) to reduce performance time. User models that treated menu search as a lexical-semantic search task were used to investigate display-based competence, highlighting the role of the visual display as an aid to a user's performance (e.g. Howes & Payne, 1990; Payne, Richardson & Howes, 2000). Cognitive models were developed to investigate user strategies in computer menu search (Hornof & Kieras, 1997, 1999).

In our study, we approached menu search as a perceptual search task in order to investigate the effect of the perceptual factors on menu search performance. By holding constant the influence of the cognitive component and varying the perceptual component, we were able to experiment with different design alternatives. The result of our study provides insight in understanding the menu search process and highlights the importance of visually discriminable designs.

The Guided Search (GS) model (Cave & Wolfe, 1990) is a model of visual search from the perceptual literature that quantitatively describes the role of parallel and serial processing in visual search. Visual search is a part of menu search, so variations in visual search times should also affect menu search times. A simulation model that could quantitatively predict visual search time for a menu search task was created with the GS model. A quantitative model that predicts visual search time in a menu search task can be used to identify designs that are optimal with respect to the model factors. The existence of such a model also provides a means to test the relative importance of the model-based perceptual factors on search performance in a menu search task. This was achieved by using the model to generate comparative menu designs through an optimization approach. These comparative designs were then tested in controlled experiments.

This discussion is organized as follows: first, the GS model of visual search in the perceptual literature is reviewed. Then, a method for defining the GS model for a menu search task and estimating the model parameters is described and we show how the model can be used to identify menu designs that optimize expected search time. Finally, two experiments testing the validity of the GS model are presented.

2. The guided search model

In a typical visual search task, a target is searched for among a number of distracters. The total number of displayed items is known as *set size*. On each search trial, a target is presented, then a display is shown. Subjects respond immediately after the target is found. Reaction time (RT) to respond to finding the target is recorded. RT is usually plotted as a function of set size and is used to infer search mechanisms.

A shallow RT vs. set size slope usually implies parallel search and vice versa. One theory that prominently distinguished parallel and serial searches was Feature Integration Theory (Treisman & Galade, 1980), in which feature search (where a target is unique on a single visual feature, for example, the target is red but all the distracters are blue) and conjunction search (the target does not have a unique visual feature thus requiring combined processing of at least two visual features to identify the target from the distracters) were identified. But the strict serial/parallel dichotomy has been questioned. After combining more than 1 million search trials from 2500 visual search experiments, Wolfe (1998) found no evidence of a dichotomous division of RT vs. set size slopes. This led to the proposal of an alternative visual search model: the GS model. The GS model identified parallel and serial stages that are involved in every visual search task. The parallel stage guides the serial stage by choosing the elements to be processed by the subsequent serial stage. The degree of how useful parallel information is in guiding the serial stage determines search efficiency. The GS model identifies efficient vs. less efficient searches rather than serial vs. parallel searches. Simulations of the GS model showed that it reproduced visual search data in a number of different visual search experiments (Wolfe, 1994). Details of the GS model were first given in Cave and Wolfe (1990), with modifications given in Wolfe (1994) for GS2 and in Wolfe and Gancarz (1996) for GS3. The later versions of the GS models take into consideration an eccentricity effect.

The descriptions and formula below show how activation values are calculated for the displayed items. In the parallel stage, separate feature maps that record detection of basic visual features such as color, orientation, and size are created. This process takes place pre-attentively and in parallel. An activation map is then created as a combination of activations from the separate feature maps. A full model that defines the metric of each basic visual feature is needed for calculating the activation values. The metric of each feature map corresponds to the just noticeable difference (JND) value for each category within that feature map. For example, in the orientation feature map, there are four broadly tuned categories: steep (for $-45^\circ < x < 45^\circ$), shallow ($-90^\circ < x < 45^\circ$ and $45^\circ < x < 90^\circ$) right ($0^\circ < x < 90^\circ$), and left ($-90^\circ < x < 0^\circ$). The activation values for each element in each feature map are determined based on the metric. Activation of element i in an individual feature map is defined as:

$$A_i = \exp\left(\frac{p}{n-1} \sum_{\substack{j=1 \\ j \neq i}}^{n} \frac{|f_i - f_j|}{|d_i - d_j|} \right) - q|f_i - t|,$$

where A_i is the activation at element i, p the weight for bottom-up, q the weight for top-down, f_i the activation value of element i, f_j the activation value of element j, t the activation value of target, n the number of items in the visual display, d_i the position value at element i and d_j the position value at element j.

It includes a top-down component and a bottom-up component. The bottom-up component is defined as the exponential of the average absolute difference between activation values for element i and all the other elements in the display. The absolute difference for each pair of elements is weighted by their distances so that the larger the distance between element i and element j, the less influence element j has on the bottom-up activation for element i. The bottom-up component is raised to an exponential, which generally makes it more influential than the top-down component in determining the activation values for the displayed elements. The top-down component is calculated as the absolute difference between the activation values that are expected for the target and element i in that specific feature map. It relates to how different the stimulus elements' features are from the designated target. The weights for top-down and bottom-up components define the relative effectiveness of the two components. Total activation for element i in a specific feature map is the weighted subtraction of top-down from the bottom-up activation because the top-down activation is implemented as inhibitory. Generally, the more distinctive an item is from the others in the display, and the closer it resembles the target, the higher the total activation will be for this item.

In the serial stage, the mechanism of operation is that only a limited part of the visual field is examined at one time. The GS model assumes that the order in which elements are processed is determined by their summed activation values. Search starts with the element having the highest activation value and moves to elements having progressively lower activation values, continuing until the target is found. In the case that the element with the highest activation value is the target, this leads to a direct search. Search time is linearly related to the number of steps for locating the target. Assuming the cycle time for processing each element is fixed, the search time could be computed as the cycle time multiplied by the number of steps in the serial stage plus the overhead search time.

3. GS simulation model for a menu search task

The GS model was applied successfully in predicting search time in perceptual search tasks. Visual search is also part of a menu search, although the latter differs considerably from a pure visual search task. We are interested in simulating the GS model on a menu search task. Although the GS model itself may prove to be inadequate in predicting menu search time, it allows us to assess the relative importance of the model-based perceptual factors in determining menu search efficiency.

3.1. CALIBRATION OF A GS SIMULATION MODEL FOR A MENU SEARCH TASK

To accomplish our goal, a simulation model for a menu search task needs to be constructed. This is the first crucial step toward testing the validity of the GS model on menu search. It forms the basis on which the visually discriminable designs are later generated. The process of setting up a GS simulation model for a menu search task included the following steps: (1) selecting model parameters, (2) collecting the experimental data from menu search trials and (3) using the experimental data to quantify free model parameters that need to be estimated.

3.1.1. Model parameters. As described in the previous section, several parameters are needed to define the simulation model. The model parameters depended on the selection of visual features used in the menu search task. In our menu search task, menu options were made of letter strings that varied in length and color and were displayed on a vertical menu. Each string consisted of the letter "A" repeated three (AAA), six (AAAAAA) or nine (AAAAAAAAA) times. The GS model contained three parameters that defined the values given for short, middle and long lengths. Each menu option was one of four colors: red, blue, black or gray. These four colors were chosen because of their relative discriminability. The GS model included four parameters, one for each color.

The GS model also included two parameters that define the top-down weights for length and color information that guide the search process for a known target. Similarly, there were two parameters that weighted the bottom-up information for length and color.

Finally, the GS model included two parameters that scale the output of the model to fit the numerical search times and two other parameters that account for variability in response times. These parameters are overhead time, cycle time, noise mean and noise standard deviation. Overhead time was added to the search time prediction for each search trial. Cycle time was defined as the time for processing each item in the serial stage. Noise mean and noise standard deviation were used to represent the random errors in the processing of parallel information and were added to the total activation for each displayed element. The first 11 parameters are free parameters that need to be estimated by the experimental data. The final four parameters are not directly relevant for the current study because the ultimate goal is to specify the relative differences in search times rather than the absolute differences. For completeness, our procedure did estimate these parameters.

We decided not to include shape characteristics in the simulation model because shape representation is not fully understood. For example, we did not know how to define the bottom-up activation of a word that could be of any combination of different letters. Because of this, same-letter strings (AAA, etc.), instead of words were used as menu options. Using words as menu options would have caused the model to fit additional variability that it could not explain. However, the model could be tested on menu search tasks with words as menu options in later experiments and thus allow us to see how the additional variability in the more realistic menu search tasks would influence the validity of the GS model.

3.1.2. Experimental task. The data used to define the model parameters were gathered by having users search for a target item in a variety of different menu designs. Different menu designs have different combinations of menu option lengths, menu option colors and relative positions of menu options. Menu options were presented vertically on a menu panel that was light gray in color (a typical menu background color). The lengths of menu options were three, six or nine letters (AAA, AAAAAA or AAAAAAAAA). The colors of the menu options were red, blue, black or gray. The number of items on the menu ranged from 3 to 20, which is within the typical range of set sizes for menu search tasks. The target was unique from all distracters in each search trial. Three hundred such menu search trials were randomly generated. A screen snapshot of the menu used in a search trial is presented in Figure 1. Each user went through the 300 search trials in the same order.

On each trial, a user was asked to select a given target from a list of menu options displayed on a single vertical menu. Unlike a routine menu search task, each search trial in this experiment was carried out in the "search and select" paradigm (Armstrong Laboratory, 1994). First the user clicked on a "Target" button to display the target menu option. The user studied the target to memorize it and its visual features. Then, the user clicked on a "Find/Found" button to display the menu. As soon as the target

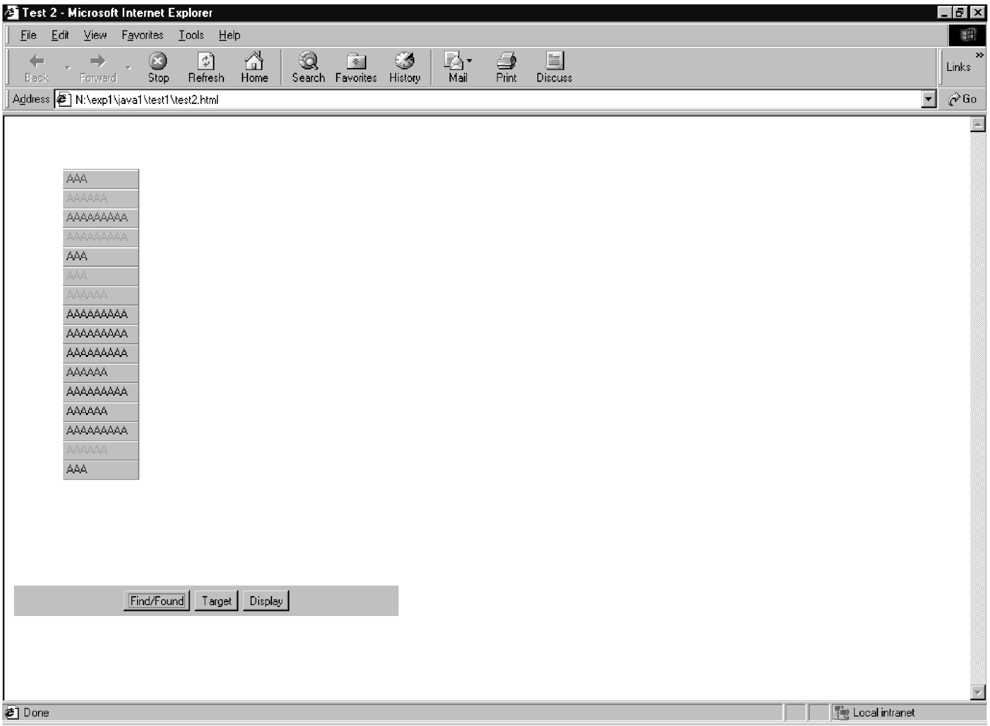

FIGURE 1. Sample screen for a menu search trial in model calibration.

was visually located on the menu, the user clicked on the "Find/Found" button again to indicate that the target was spotted. Finally, the user moved the cursor to the target location in the menu and clicked on it.

The times spent in reading the target, visually searching for the target and moving to the target were recorded separately by the computer. The visual search time was recorded as the time between clicking the "Find/Found" button to show the menu options and clicking the "Find/Found" button to indicate identification of the target. The computer's temporal resolution is accurate to the millisecond.

3.1.3. Participants. Ten participants were recruited for the experiment. All 10 participants were students from engineering departments at Purdue University. Eight were male; two were female. They were natives of different countries. The mean age of the 10 participants was 26.0, and the standard deviation of their ages was 5.09. All participants reported normal color vision.

3.1.4. Model fitting. The average search times across the 10 participants in the 300 random search trials were used to fit the model predictions. Parameter fitting was done with a variation of a hill-climbing method. The GS model was implemented with a computer program written in Matlab. With a given set of model parameters, the properties of the display and the target on a given trial, the model can compute a prediction of the visual search time for locating the target. This prediction was computed for each of the 300 search trials. The predicted search times were then correlated with the experimentally measured search times, and the squared correlation coefficient was computed. A single model parameter was then changed, and the predicted search times for all 300 displays were computed again. These predicted search times were used again to calculate the squared correlation coefficient with the experimental search times. If the parameter change led to an increase in the squared correlation coefficient, the change was kept. If the parameter change led to a decrease in the squared correlation coefficient, the parameter change was undone and the model parameters were returned to the original state. This process iterated until it seemed that no further changes in model parameters led to an increase in the squared correlation coefficient.

The model parameters were first set manually with a similar process described above. However, some judgments were used. For example, it was considered that the difference between red and gray would be bigger than the difference between red and blue. The purpose of the manual adjustment of model parameters was to place the model parameters in a range that seemed to lead to a large squared correlation coefficient. The computational algorithm (hill-climbing) was then used to fine-tune the manually set parameters. This guaranteed that the final set of parameters gave at least a local maximum of the squared correlation coefficient. The program for fitting the parameters ran for about 14 h, which included 100 800 iterations of parameter changes, with about 0.5 s for calculating each squared correlation coefficient for a set of model parameters. Because of the imperfection of the fitting method used, it is not guaranteed that the final set of parameters for the simulation model was the absolute best set of parameters. Instead, we only claim that a working simulation model was constructed to account for a large part of the variation in the users. (Table 1)

3.1.5. Results and discussion. In the manual fitting process, the highest squared correlation coefficient that was obtained by adjusting the model parameters was 0.4858. After running the computer search, model parameters were slightly modified and the squared correlation was raised to 0.4966. The simulation model could account for about 50% of the variation in the user data. This means the correlation between the predicted search times and the experimental search times for the 300 random search trials was above 0.70. It is likely that the GS model is not fitting noise in the data, but rather meaningful patterns as predicted by the model. The resulting set of simulation

TABLE 1
Simulation model parameters for menu search task

Anchor parameter	Color parameter	Length parameter	Weight parameter
overhead = 50	*red* = 10.2	*Long* = 1	*topdown_color* = 3.8
cycle = 25	*blue* = 7	*middle* = 5	*topdown_length* = 2
noisemean = 0.2	*black* = 2	*short* = 4	*bottomup_color* = 5
noisestd = 0.05	*gray* = 2.9		*bottomup_length* = 4

Table 2
Menu options and their search frequencies in experiments 1and 2

Search frequency	10	9	4	3	2
Menu labels	"submarine" "ground" "beauty" "air"	"roadhouse" "map"	"professor" "horseback" "butterfly" "insect"	"heaven" "minute" "log" "toy" "bed" "ear"	"rubber" "circle" "war" "tip"

model parameters is listed in Table 2. The formulas to calculate total activation for each menu option based on these parameters would be

$$Ai(total) = Ai(color) + Ai(length),$$

$$Ai(color) = \exp\left(\frac{p}{n-1}\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{|fi - fj|}{|di - dj|}\right) - q|fi - t|,$$

where $n = 3, \ldots, 20$, $p = 5$ (weight for bottom-up color information in the model), $q = 3.8$ (weight for top-down color information in the model) and f_i, f_j, t correspond to the activation values of element i, j, and the target in the color feature map, respectively.

$$Ai(length) = \exp\left(\frac{p}{n-1}\sum_{\substack{j=1\\ j\neq i}}^{n}\frac{|fi - fj|}{|di - dj|}\right) - q|fi - t|(i \neq j),$$

where $n = 3, \ldots, 20$, $p = 4$ (weight for bottom-up color information in the model), $q = 2$ (weight for top-down color information in the model), f_i, f_j, t correspond to the activation values of element i, j and the target in the length feature map, respectively.

In the calculation of an activation value, the magnitude of a parameter only matters relative to other parameters. For example, the length parameters reveal that the long stimulus feature is notably different from the middle and the short stimulus features. Thus, a long menu option would stand out in a display of otherwise short menu options, and would be only slightly less obvious in a display of middle-length menu options. However, a middle-length menu option would not stand out nearly as well among a set of otherwise short menu options. Likewise, a red or blue menu option will stand out relative to a set of black and/or gray options. However, red and blue menu options are not as easily discriminated, and black and gray menu options are difficult to distinguish. The weight parameters for both the top-down and the bottom-up components indicate that for equal feature differences among the display features, a difference in color is going to be more distinctive than a difference in length. This is especially the case for the bottom-up component because the feature differences in the bottom-up components are raised to their exponentials in the calculations.

Because the fit was based on the squared correlation coefficient, there is no necessary quantitative match between the absolute predicted and experimental search times. The model, in its current form, only predicts the relationships between search times under different conditions. Finally, the model does not predict variability in search times among individual users because we averaged over the users and each user saw a trial only once. If this information is important for a particular application of the model, one would simply modify the procedure we used to ensure that the model parameters fit those details.

3.2. GENERATION OF COMPARATIVE MENU DESIGNS

We are going to test the adequacy of the GS model in predicting menu search time by comparing two distinctive menu designs generated by the model. One design is predicted by the GS model to have fast search times, the other design is predicted by the model to have slow search times. These two designs were both generated through a computerized search among a large number of design alternatives using a simulated annealing algorithm. This section describes the process and the results of generating these two menu designs. These two designs are utilized in the later experiments.

3.2.1. Problem definition. Generation of the model designs was taken as an optimization problem, with the objective function of minimizing or maximizing average search time. The approach is similar in style to Francis (2000), where minimization of predicted movement time was found to improve search through a multifunction display. In the present experiments, *average search time* was defined as the frequency-weighted sum of search times for each option in the menu. The frequency weights were predefined search frequencies for each menu option. The predefined frequencies were set arbitrarily so that some menu options were searched more frequently than other menu options. The labels and the search frequencies associated with each menu option are given in Table 2. The labels were picked from common English words; there should be little difference in familiarity or difficulty associated with these words. The lengths of labels were three, six or nine characters. These lengths correspond to the parameters defined for the length feature in the GS model. Problem definitions for the generating the menu designs generated by the GS model are summarized in Table 3.

3.2.2. Model generated menu designs. A simulated annealing (SA) algorithm was used to generate the model designs. SA is a probabilistic hill-climbing algorithm that is conceptually related to the crystallization of materials when they are first heated to high temperature and then cooled down slowly (Aarts & Korst, 1989). Computer programs in Matlab were written to implement the GS model calculation in the SA algorithm. For any given menu design with color and location of each label specified, the GS model could give a prediction of search time for each menu option when it is the target. A frequency-weighted sum of search times for all menu options can then be calculated. For generating the model predicted fast design, the SA program started with a random

TABLE 3
Definitions of optimization problems for generating menu designs

Menu design	Problem definition	Objective function	Problem size
Model predicted fast design	Assigning 20 labels, defined by $[F_i, L_i]$ ($i = 1, ..20$, F_i is search frequency and L_i is length of the ith label) to 20 menu locations.	Minimize $\Sigma(F_i * T_i)$ $i = 1, ..20$, where F_i is the search frequency of menu i, T_i is the model prediction time for searching menu item i in the specified display	$4^{20} \times 20!$
Model predicted slow design	Each menu location could have one of the 4 colors (red, blue, gray, black) be assigned to it.	Maximize $\Sigma(F_i * T_i)$ $i = 1, ..20$ where $\mathrm{F_i}$ is the search frequency of menu item i, T_i is the model prediction time for searching menu item i in the specified display	$4^{20} \times 20!$

design, calculated its average search time, then picked a menu option, swapped its location with another randomly picked menu option in the design, reassigned their colors and calculated the average search time again. If average search time was decreased, the change in design was kept; otherwise, it was accepted probabilistically, which allows the search to jump out of locally optimal solutions that are not the same as the global optimal solution. The probabilistic acceptance gradually becomes more stringent.

The time that the optimization program ran was determined by the setting of the SA parameters. These parameters determined how fast the temperature would cool down, that is, how soon the optimization process would converge to a local optimal solution and how wide the range of the random search process would be. The approximate time for generating each menu design by the SA program was about 7 h, which included about 252 000 iterations with 0.1 s for calculating the average search time for each random design. Ideally, if the temperature is decreased logarithmically and the SA program is set to run for a long enough time, it will be able to find the globally optimal solution. However, because of the limits of time and CPU speed, it is not guaranteed that the SA program would find the global optimal solution based on the optimization problem defined. Despite these indeterminacies, the model predicts substantial differences between the comparative designs generated by the SA program.

3.2.3. Characteristics of model generated menu designs. We used the above approach to generate two types of menu designs that the model predicts should have very different search times. Design layouts for model predicted fast and model predicted slow menu designs are shown in Table 4. To understand the model predicted differences in the designs, consider the menu label "professor". The predefined search frequency of this label is 4. In the predicted fast design, when searching for the target "professor", which is located on top of the menu and with a red color, its bottom-up color activation should be calculated based on the formulas given in Section 3.2. That is, take the absolute difference in color activation between menu option "professor" and each of the other menu options divided by the corresponding distances in locations between "professor" and each of the other menu options and add them up. The activation value for each color is specified in the model parameters, so the result would be $0/1 + 3.2/2 + 3.2/3 + 3.2/4 + 3.2/5 + 7.3/6 + 7.3/7 + \cdots = 25.524$. Multiply this number by the bottom-up weight for color ($p = 5$), divide it by the number of menu options in the design minus 1 ($n - 1 = 19$) and raise it to its exponential (exp (5*25.524/19) = 826.204). This is the bottom-up color activation for menu option "professor".

The top-down color activation for menu option "professor" is calculated by multiplying the top-down weight for color and the absolute difference in color activations between menu option "professor" and the target, which is zero because the target is menu option "professor".

Bottom-up length activation for menu option "professor" would be calculated in the same way except that the differences in length are used in the calculation. The result is $\exp\ (4 * (0/1 + 0/2 + 0/3 + 3/4 + 6/5 + 6/6 + 6/7 + 3/8 + \cdots)/19) = 5.122$. Top-down length activation for "professor" is zero because "professor" is the target.

Adding the activations from all bottom-up components and subtracting all the top-down components results in the total activation value for menu option "professor"

Table 4
Screen layout of model menu designs

Model predicted fast				Model predicted slow				Designer predicted fast			
Pred	Freq	Layout	Color	Pred	Freq	Layout	Color	Pred	Freq	Layout	Color
175	4	professor	Red	500	10	air	Black	225	4	professor	Blue
100	9	roadhouse	Red	150	2	circle	Red	250	9	map	Black
200	4	butterfly	Blue	175	3	log	Black	200	10	beauty	Red
125	10	submarine	Blue	200	3	toy	Red	250	10	air	Red
225	3	minute	Blue	125	2	rubber	Red	150	10	submarine	Red
100	9	map	Blue	350	4	insect	Gray	100	10	ground	Red
100	10	air	Gray	400	9	map	Gray	100	9	roadhouse	Blue
375	3	log	Gray	75	2	war	Red	175	4	insect	Black
225	4	insect	Black	325	3	ear	Black	200	4	butterfly	Black
425	2	war	Gray	225	3	minute	Blue	75	4	horseback	Blue
175	3	toy	Black	350	4	professor	Black	325	3	bed	Black
125	3	bed	Gray	500	10	beauty	Black	350	3	toy	Black
75	10	ground	Red	475	10	submarine	Gray	375	3	ear	Black
125	3	heaven	Gray	300	4	butterfly	Blue	400	3	log	Black
100	10	beauty	Black	450	9	roadhouse	Black	400	3	heaven	Black
400	2	rubber	Gray	400	4	horseback	Black	425	3	minute	Black
350	2	circle	Gray	100	2	tip	Red	400	2	tip	Gray
200	3	ear	Blue	250	3	heaven	Black	425	2	war	Gray
425	2	tip	Blue	250	3	bed	Red	450	2	circle	Gray
150	4	horseback	Blue	550	10	ground	Blue	500	2	rubber	Gray

$(826.204 - 0 + 5.122 - 0 = 831.326)$. This is the total activation value of menu option "professor" in the model predicted fast design when the target is "professor."

The pop-out levels of other menu options when searching target "professor" could be calculated in similar way, but they would have non-zero subtractive top-down signals for the feature that does not match the designated target. Among the pop-out levels (measured by total activation) of all the menu options, "professor" ranks fifth. Thus, the predicted search time for searching "professor" in the model predicted fast design is $175(50 + 25 * 5)$ units time.

In the model predicted slow design, menu option "professor" is placed at the 11th menu location, with black color. When searching for "professor," the bottom-up activation for color feature is $\exp\ (5 * (0/10 + 8.2/9 + 8.2/7 + 8.2/6 + 0.9/5 + 0.9/4 + 8.2/3 + 0/2 + 5/1 + 0/1 + 0.9/2 + 5/3 + 0/4 \cdots)/19) = 79.964$. The bottom-up activation for length feature is $\exp\ (4 * (6/10 + 3/9 + 6/8 + 6/7 + 3/6 + 3/5 \cdots)/19) = 62.631$. The top-down activations in color and length for "professor" are zero because the target is "professor". The total activation for "professor" in the model predicted slow design when the target is "professor" is $62.631 + 79.964 = 142.596$. When this activation value is compared with the total activations of other menu options in the design, it ranks 12th. So the predicted search time for searching "professor" in model predicted slow design is 350 $(50 + 12 * 25)$ units time.

When these calculations are done for every menu option and the resulting predicted search time is weighted by the specific frequency of searching for that menu option, we can produce an average search time. The predicted average search time by the GS model is 148 units time in the predicted fast design and 393 units time in the predicted slow design. Figure 3 plots the predicted search time with respect to different frequency groups. In the model predicted fast design, predicted search time decreases with increased search frequency. This reflects the task assigned to the optimization. The control of search time is done by the assignment of color and location to the menu options. To minimize average search time, the menu items searched for most frequently should have the shortest search times. Depending on the frequency weights, it may be worthwhile to increase the search time for less frequently searched for menu options if such a change allows for a large enough decrease in search times for more frequently searched for menu options.

These effects are also evident in the model predicted slow design. Here, the effect of frequency is the opposite of that in the model predicted fast design: menu options are given stimulus features so that higher frequency items have longer search times. Note that the optimization task is fairly constrained. To get the longest predicted average search time, the design set menu colors and locations so that high-frequency targets take a long time to find. To achieve this, the system made low-frequency targets very distinctive (e.g. elements in frequency group 2 have very short search times). This arrangement makes the low-frequency menu items effective distracters when searching for high-frequency menu items. In the model, the "observer" will always be distracted by the low-frequency items and must disengage from them before finding the target. Thus, the model predicts that the slow design is not necessarily slow for all menu options, but that resources are allocated differently compared with the fast design.

Indeed, the model predicts that menu options in frequency group 2 should be found faster in the slow design than in the fast design.

When looking at the model predicted fast and model predicted slow menu designs, it is difficult to pinpoint how the layout characteristics accelerate or hinder the overall search process because the predicted search times are determined quantitatively by the GS model. In general, in the GS model, items that have distinctive visual features (in terms of both color and length in the search tasks) from nearby items should have higher activations. Table 4 shows that some of the frequently searched items are placed beside items that are visually distinctive from them in the model predicted fast design. In the model predicted slow design, less frequently searched items are placed beside items that are visually distinctive from them, which makes them effective distracters when searching for high-frequency items.

However, we caution the reader against trying to take an intuitive approach for understanding the basis of the designs. The description given is not always accurate. It is fairly easy to find individual counterexamples in the predicted fast and slow designs. This does not invalidate the optimization or the model, but instead indicates that the menu search time measured by the model is a global property of the entire design. It is likely that, when a high-frequency item is not notably distinctive from nearby items, moving it to an alternative position or changing its color would lead to a worse overall search time. Perhaps such a change would improve search for that item but would increase search times for many other items. The overall search time cannot be understood by looking at only the features of individual menu items or only the details of search for a single item. Instead, the design must be based on the interacting effects of the search process and the frequencies of searching for different items. It seems nearly impossible for a designer to simultaneously consider all the possibilities. This is our motivation for using the GS model and our optimization approach to handle these interactions in a quantitative manner.

4. Experiment 1

In Experiment 1, model predicted fast and model predicted slow menu designs were compared. The design layout in Table 4 illustrates the two menu designs used in the experiment. In the actual experiment, menu options were presented on a vertical menu with light gray background color, as shown in the screen snapshot of the model predicted fast design in Figure 2. The model predicted fast design should minimize expected visual search time, while the model predicted slow design should maximize expected visual search time. If the GS model is valid in modeling important perceptual aspects of a menu search task, then the model predicted fast design will help users locate targets easily, whereas the model predicted slow design will hinder users from locating targets easily. Failure to find such a difference will indicate one of two conclusions: either the specific GS model we have defined is not capturing the important perceptual aspects that determine visual search time in this display, or the perceptual aspects of the display are not important in determining visual search time for this task. The latter is possible because it could be that cognitive factors such as memory are more important

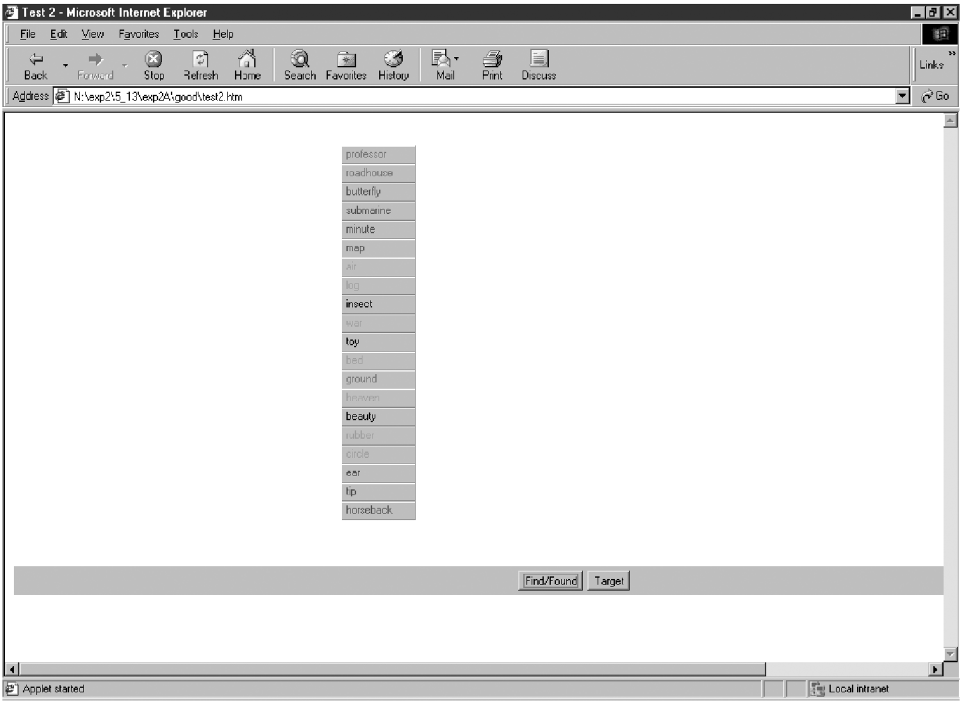

FIGURE 2. Sample screen for model predicted fast design.

than perceptual aspects of the display. Finding the predicted difference will indicate that the model has captured some important issues in menu design.

4.1. METHOD

4.1.1. Participants. Twenty participants were recruited from students at Purdue University. The mean age of the participants was 24.9, and the standard deviation of age was 4.45. Seven of the participants were male, 13 were female. There were seven undergraduate students, seven master students, and six doctoral students. They were from 11 different countries and 11 different departments. All participants reported having normal color vision.

4.1.2. Experimental task. Each participant went through 100 search trials in both menu designs (model predicted fast and model predicted slow). The order of performing the search trials in the two menu designs was randomized for the participants: 10 performed the search trials in the model predicted fast design first and the other 10 performed the search trials in the model predicted slow design first. In the 100 search trials performed on each menu design, each of the 20 menu options was searched as a target by the predefined frequency used to generate the model designs. The sequence of target appearance was randomized in the two designs but was the same for all the participants.

4.1.3. Dependent variables. Search time, error and satisfaction were measured as dependent variables. Search time variables were measured in three ways: mean search time for the 100 search trials in each design group, frequency-weighted search time of the first trial for each menu option and frequency-weighted search time of the second trial for each menu option. The first variable is useful for comparing the designs in a straight-forward manner. However, search times for high-frequency items are likely to be different from search times for low-frequency items both because of the design and because of practice effects. Thus, the overall mean search time does not allow for a comparison of frequency effects that are related only to differences in menu designs. The latter two terms allow us to compare search times for labels with different search frequencies without interference from practice effects. Because menu designs were generated with the objective to minimize or maximize average search time, frequency-weighted search time variables correspond to model predictions of average search time.

Satisfaction scores were also collected for each menu design by a satisfaction questionnaire. They were calculated as the average of seven questions that were asked about the arrangement of colors and locations of menu options in reducing average search time, and participant's general satisfaction toward task performance.

Error rates were counted as the number of errors made in the 100 search trials in each design group.

4.1.4. Independent variable. The independent variable was type of menu design. Two groups of menu design were compared: model predicted fast vs model predicted slow. They were described in Section 3.2.

4.1.5. Experimental design. A within-subject experimental design was used. Each participant used both menu designs. Half of the participants used the model predicted fast design first and then the model predicted slow design, the other half of the participants used the model predicted slow design first and then the model predicted fast design. The experiment design was a three-factor crossover design. The three factors were design, participant and order of using menu design. The corresponding statistical model could be written as:

$$Y_{ij} = D_i + S_j + O_k (i = 1, 2, j = 1, 2 \ldots 20, k = 1, 2),$$

Where Y_{ij} is the performance measurements, D_i the menu design effect, S_j the subject effect and O_k the order effect.

4.1.6. Procedure. An experimental session lasted about 50 min. Each participant went through a training session with 50 search trials first. The menu layouts in the search trials were randomly generated with another set of words, but with similar physical presentation as the experimental ones, that is, a vertical menu with the same number of menu options, the same number of color options for each menu options and the same menu background color. The purpose of the training was to let the participants become familiar with the procedure of performing the search task, especially the search and select paradigm. After the training, each participant performed two sets of search trials in the designated order using the two designs. After performing 100 search trials in each menu design, the participant filled out a satisfaction questionnaire.

4.2. RESULTS

The predicted search times and the average search times of the first two trials, broken down by design type and search frequency, are shown in Figure 3.

4.2.1. Difference of designs. If the GS model captures factors that matter to performance in a menu search task, then search performance should be better in the model predicted fast design than in the model predicted slow design. The overall mean search time ($F_{1,18} = 3.04$, $p < 0.1$) was suggestively shorter in model predicted fast design (1.440 s) than in model predicted slow design (1.541 s). Frequency-weighted search time variables for both the first ($F_{1,18} = 5.18$, $p < 0.05$) and the second ($F_{1,18} = 9.27$, $P < 0.01$) trials were significantly shorter in model predicted fast design (1.829 s for the first trial, 1.505 s for the second trial) than in model predicted slow design (2.107 s for the first trial, 1.718 s for the second trial). Thus, the primary experimental finding validates the model's prediction that there is a difference in search time between the designs.

No significant differences were present in satisfaction scores and errors between the two designs, a finding that was replicated throughout our studies. This may reflect the inadequacy of the questionnaire or may indicate that participants were unaware of performance differences across the designs. Generally, the number of errors was small because the search task is relatively simple, thus, less prone to errors.

The primary finding is that the model captures important components of the search process, and that our optimization algorithm could utilize the model to build designs of differing quality.

4.2.2. Frequency effects. The GS model also predicts differences in search times for menu items of different search frequencies. There are two types of frequency effects: within design and between design. The effect of frequency within each design on each of the three search time variables (mean search time of overall trials, first trial, second trial) generally matches the predicted pattern. A frequency effect is significant in the model predicted fast design for all three search time variables ($F_{4,75} = 12.15$,

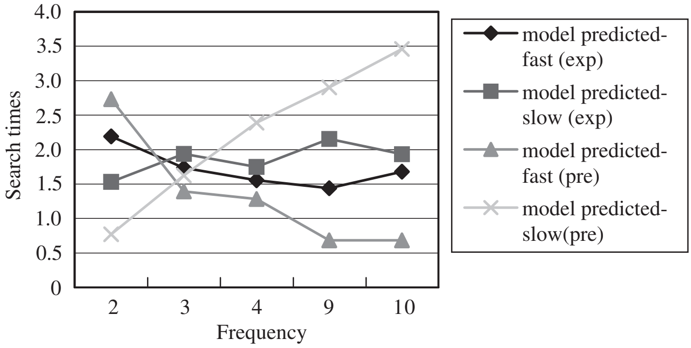

FIGURE 3. Plot of search times by frequency in Experiment 1. Experimental search times are based on the average of the first two trials and are measured in seconds. Predicted search times are measured in time units and are scaled so that the mean experimental search time was equivalent to mean predicted search time.

$p < 0.01$; $F_{4,75} = 5.53$, $p < 0.01$; $F_{4,75} = 4.57$, $p < 0.01$). In the model predicted slow design, there was also a significant frequency effect in the mean search time of overall trials and first trial ($F_{4,75} = 2.62$, $p < 0.05$; $F_{4,75} = 3.88$, $p < 0.01$). Furthermore, Duncan range tests showed that in the model predicted fast design, more frequently searched items (frequency group 9) have shorter search times than less frequently searched items (frequency group 2). In the model predicted slow design, less frequently searched items (frequency group 2) have shorter search times than more frequently searched items (frequency group 10). This pattern indicates that search times were related to the predefined search frequency. Because the results hold before practice effects could develop, the differences must be related to the menu design, as predicted by the model. As predicted, the contrast of search time difference among different frequency groups in the model predicted slow design is in the opposite direction from the contrast in the model predicted fast design. The results were consistent with the GS model predictions on the two designs.

The GS model predicts that there should be differences in search times for the extreme search frequency groups. Significant design effects on all three search time variables were found in frequency groups 2 and 9. In frequency group 9, mean search time of overall trials ($F_{1,18} = 20.5$, $p < 0.01$), the first trial ($F_{1,18} = 22.88$, $p < 0.01$), and the second trial ($F_{1,18} = 5.36$, $p < = 0.05$) were significantly longer in the model predicted slow design than in the model predicted fast design. This finding reflects the general effect of the designs, whereby searches are faster in the predicted fast design than in the predicted slow design. However, the GS model also predicts that search times for menu items in frequency group 2 should be faster in the predicted slow design than in the predicted fast design. The data support this prediction. In frequency group 2, the mean search times of overall trial ($F_{1,18} = 24.68$, $p < 0.01$), first trial ($F_{1,18} = 44.45$, $p < 0.01$), and second trial ($F_{1,18} = 11.03$, $p < 0.01$) were significantly shorter in the model predicted slow design than in the model predicted fast design.

4.2.3. Discussion. All of these results indicate that the GS model is capturing some important characteristics for the first few trials that a user searches a menu. The experimental data validate the model predictions that the model predicted fast design should lead to faster searches than the model predicted slow design and that the relative ordering of search efficiency within a design are related to search frequency. Moreover, comparing search times across designs for the lowest frequency menu items validates the model's prediction that the predicted slow design is slow because observers must disengage from a potent distracter that is itself rarely the target.

5. Experiment 2

Experiment 2 serves three functions. First, it considers whether the model is considering factors that human designers also consider when they create menu designs. We had a human designer create a menu design using the same stimuli and situations as the model. We then used the GS model to interpret the design created by the designer. If the design is interpreted in the GS model as having systematic trends related to the menu search task, this can be taken as evidence that the designer and the GS model are, in part, optimizing similar display characteristics.

Second, the experiment considers the relative worth of the model's optimized design by directly comparing it with the design created by the human designer. This is an important comparison because although Experiment 1 demonstrates differences in model predicted fast and model predicted slow designs, it does not indicate whether either of the designs is good in an absolute sense. To make such a comparison, a model-based design must be compared with a standard design of high quality. The human designer created such a standard design.

Third, if the design created by the human designer is superior to the model design, we can look for differences between the designs. This will help identify factors that are important in menu search but are not accounted for by the GS model. This is important because it will help drive the development of better models and will identify situations where the current approach is likely to be beneficial. Toward these ends, a human designer was told to minimize the average search time based on the predefined search frequencies of menu items. The menu items and associated search frequencies were the same as in Experiment 1.

We had no *a priori* expectations as to whether the design created by the model or the human designer would be superior. The model has the advantage of being able to quantitatively consider a large number of possible designs. Given the combinations of the design task and the complicated interactions of menu search, it seemed plausible that the human designer would only consider a small subset of possible designs. From a computational perspective, the model seemed to have an advantage. On the other hand, the model is limited by its inability to exploit cognitive aspects of search strategy. A human designer could possibly consider both the perceptual and cognitive factors in a menu design. It seems plausible that, if the cognitive factors are more important than the perceptual factors in the menu search task, the design created by the human designer would be better. It is also possible that the design created by the human designer would incorporate perceptual factors that the GS model does not consider. These factors would seem to give the designer an advantage. Thus, taking all things into consideration, it was unclear whether the model or the human designer would produce a better design. The advantage to one or the other surely depends on the details of the menu search task and the complexity of the menu.

5.1. METHOD

5.1.1. Participants. For Experiment 2, 20 different participants were recruited. The mean age of the 10 participants in Experiment 2 was 26.8, with a standard deviation of 8.28. All participants reported normal color vision.

5.1.2. Experimental task. Each participant went through 100 search trials in both menu designs (model predicted fast and designer predicted fast). The order of performing the search trials in the two menu designs was randomized for the participants: 10 performed the search trials in the model predicted fast design first and the other 10 performed the search trials in the designer predicted fast design first. In the 100 search trials performed on each menu design, each of the 20 menu options was searched as a target by the predefined frequency used to generate the model design. The sequence of target appearance was randomized in the two designs but was the same for all the participants.

5.1.3. Dependent variables. The dependent variables were search time, error and satisfaction. The same search time variables were measured in the same way as in Experiment 1: mean search time for the 100 search trials in each design group, frequency-weighted search time of the first trial for each menu option, and frequency-weighted search time of the second trial for each menu option. Satisfaction scores were calculated from the satisfaction questionnaire for each design. Error rates were counted as the number of errors made in the 100 search trials for each design group.

5.1.4. Independent variable. The independent variable was menu design. The human designer was given the same design task as was given to the model for producing a fast design. The human designer was a top Master's student from the industrial design department at Purdue University. He was not aware of the principles of the GS model and was excited to challenge the model design when given the task. The design layout for the human designer predicted fast design is presented in Table 4. The actual menu being searched in the experiment was displayed as a vertical menu (background color light gray) with the color and location layout as shown in Table 4. The comparative designs in Experiment 2 were the model predicted fast design and the human designer predicted fast design. The model predicted fast design was the same as the one used in Experiment 1.

5.1.5. Experimental design. The experiment design for Experiment 2 was the same as that for Experiment 1, that is, a three-factor within-subject design. The three factors were design, participant and order of using the two designs.

5.1.6. Procedure. An experimental session lasted about 50 min. Each participant went through a training session with 50 search trials, and then performed the two sets of search trials in the designated order, with 100 trials performed in the model predicted fast design and 100 trials performed in the human designer predicted fast design. There was a brief break after performing each test set, during which participants filled out the satisfaction questionnaire.

5.2. RESULTS

5.2.1. Difference of designs. Our analysis starts with a model-based analysis of the two designs. Figure 4 plots the predicted search times and experimental search times for each frequency group for the two designs. The model predicts that for both designs, those items with higher search frequency will have shorter search times. This is not surprising for the model-based design because this pattern is entirely consistent with the optimization that produced the design. However, there was no *a priori* reason that the design created by the human designer would show the same tendency. If the human designer were using design principles dramatically different from the model, then the model predictions would likely show no systematic effect of search frequency. That the

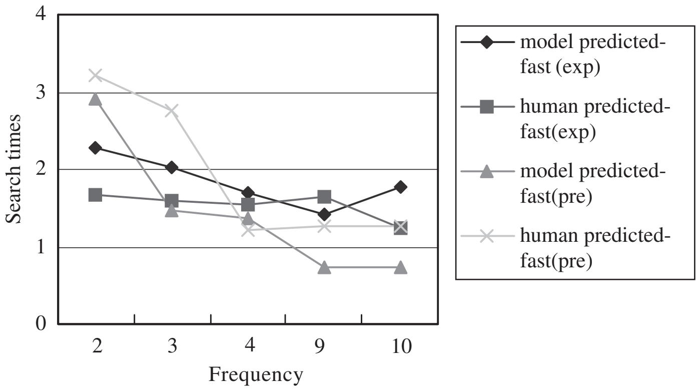

FIGURE 4. Plot of search times by frequency in Experiment 2. Experimental search times are based on the average of the first two trials and are measured in seconds. Predicted search times are measured in time units and are scaled so that the mean experimental search time was equivalent to mean predicted search time.

two designs have a similar pattern of model predicted search times suggests that the human designer is, in part, assigning colors and locations to menu items in a manner that taps into the same properties that the model was using. This correspondence between the model design and the design created by the human designer suggests that the model is attacking the design problem properly.

However, a comparison of the designs and their predicted search times also identifies differences. The average frequency-weighted search time predicted by the GS model for the model predicted fast design is 148 units time. The average frequency-weighted search time predicted by the GS model for the designer predicted fast design is 247 units time (they were rescaled in Figure 4). Thus, if only the model-based factors are important, the model predicted fast design should be better than the human designer predicted fast design.

By looking at the design created by the human designer, we can identify factors that seem to promote certain search strategies. For example, more frequently searched items were put in the upper part of the menu, and they were given a red or blue color. Thus, the observer could adopt a strategy of generally focusing on the top of the menu or on the red and blue colors and thereby reduce the effective number of distracters. Because of the menu design, this would generally reduce search time (although it would increase it for low-frequency targets). Strangely though, the design is not entirely consistent with regard to this strategy (e.g. menu items "professor" and "horseback"), so the human designer may not have consciously planned for this particular search strategy. Alternatively, the human designer may have implemented this strategy, but then revised it when it resulted in a conflict with perceptual factors. Regardless of the human designer's reasoning, the final design seems to include a possible cognitive component. Whether the inclusion of the cognitive component is used by observers and whether its benefits offset what the model predicts would be a poorer use of perceptual components is an empirical issue that is addressed next.

Contrary to the model's predictions, the design created by the human designer produced slightly faster search times than the model predicted fast design. The mean search time for all trials was 1.559 s for the model design and 1.272 s for the design created by the human designer ($F_{1,18} = 4.42$, $p < 0.1$). The frequency-weighted search time for the first trial ($F_{1,18} = 4.13$, $p < 0.1$) was suggestively shorter in the design created by the human designer (1.515 s) than in the model design (1.959 s). There was no statistical difference in search times for the frequency-weighted search times in the second trial between the design created by the human designer (1.326 s) and the model design (1.611 s). There were also no significant differences in satisfaction scores and error rates between the two designs.

Thus, there is evidence that the design created by the human designer is superior to the model design, but the difference is not exceptionally large. We tentatively interpret these results as indicating that the model produces menu designs that compare favorably with the design created by the human designer.

5.2.2. Frequency effects. Unlike Experiment 1 in which the analysis of frequency effects included both within-design and between-design comparisons, here we only consider within-design frequency effects. This is because the model predicted search times in Figure 4 do not suggest any particular between-design differences that are likely to be statistically significant. For the model predicted fast design, the effect of frequency is generally the same as for Experiment 1. The frequency effect on search time was significant in the overall trials ($F_{4,35} = 4.29$, $p < 0.01$) and the second trial ($F_{4,35} = 2.60$, $p < 0.05$) in the model predicted fast design. The frequency effect on search time was also significant in the overall trials ($F_{4,35} = 3.61$, $p < 0.05$) and the second trial ($F_{4,35} = 2.49$, $p < 0.1$) in the designer predicted fast design. Duncan multiple range tests on the frequency effect suggest that the pattern of differences in search time among frequency groups were similar in the model predicted fast design and the designer predicted fast design. Both designs have shorter search times for items that have higher search frequency (frequency 9, 10) than items of lower search frequency (frequency 2, 3).

5.2.3. Discussion. The results of Experiment 2 suggest several important findings. First, the common pattern of predicted search times with variation in search frequency, as shown in Figure 4, suggests that the GS model and the human designer are utilizing some common methods to manipulate search times. Second, the model design is good, but not quite as good as the design created by the human designer. This indicates that the GS model with the estimated parameters includes factors that are important for menu use, but does not include some factors that also significantly contribute to menu use. These additional factors could be perceptual factors that the model does not include or they could be cognitive factors that are outside the domain of the GS model. We now consider what the details of some of these factors could be. This analysis is necessarily speculative, but we feel it is worthwhile to hypothesize if only it suggests variations of our approach to menu design.

Menu search tasks are different from traditional visual search tasks in that the items remain fixed across trials. In a traditional visual search task, the distracters change on every trial and the target remains unchanged. The target and distracters also change location across trials. As a result, an observer cannot rely on memory to guide the search process. In a menu design, the target varies across trials, but the distracters remain fixed. All menu items remain in a fixed location across trials. Thus, an observer can rely on memory to guide the search process. Thus, an observer may see the target and not remember its exact location but remember that it was near the bottom of the menu. Such knowledge could simplify the search process by effectively removing the number of distracters, if the observer could ignore the top half of the menu. A designer could take advantage of such memory-related strategies by designing the menu to promote memory of item locations. This could include alphabetical arrangement of menu items (which seems not to have been used in the current situation) or the assignment of perceptual features (e.g. the design created by the human designer seems to use color to group frequently used items). We cannot tell if the design created by the human designer promotes memory formation because we have no direct experimental measure of memory for menu item locations.

A designer might also take advantage of stereotyped behaviors among observers. For example, people may follow a certain scanning order such as top to bottom. The GS model does not consider such a tendency, but the human designer seemed to have taken advantage of this tendency by placing the most frequently searched for items in positions that are likely to be at the beginning of the scan. Using perceptual or cognitive factors, a designer could also suggest a particular scanning pattern.

We did ask the human designer about the strategies he was using in creating the predicted fast design. According to the designer, he basically used color grouping and frequency grouping. That is, the most frequently searched items were put in the upper part of the menu and were assigned the most salient color (red), less frequently searched items were placed in the middle of the menu with color black and the least frequently search items were located in the lower part of the menu with color gray. Color blue was used to highlight menu options with the longest words. These menu options were placed at the border of a different frequency group to create a pop-out effect. We are not clear as to why the grouping principles were not consistently followed. But the human designer did believe this would create a fast design that would beat the model's design. However, it would be interesting to contrast both of these designs with designs with consistent assignment methods (ordering or perceptual grouping).

6. Conclusions and discussions

6.1. FINDINGS

We have taken a model of visual search from the perceptual psychology literature and applied it to menu design. To do this we defined a GS model, gathered data to estimate parameters of the model, used the model and an optimization routine to build menu designs, and tested the menu designs. In Experiment 1, we experimentally measured search times for menu designs that the model predicts should be extremes (fast and slow) for a given set of menu items and search frequencies. The pattern of experimental results matched the predicted pattern. In Experiment 2, we compared the model predicted fast menu design to a corresponding menu design generated by a human designer. The design created by the human designer was slightly better than the model-based design. An analysis of the design created by the human designer suggests that it uses some of the same principles as the model-based design, but also includes additional design principles.

This study is an explorative attempt to improve the graphic design of menu display. It tested the adequacy of the GS model in a related HCI area. In conclusion, the GS model was able to highlight the importance of visually discriminable designs. In practice, it suggested that to create effective menu displays, we should not violate perceptual rules but we should also not restrict ourselves to visual perception.

6.2. LIMITATIONS AND RECOMMENDATIONS

We realized that there are some inherent limitations of applying the GS model to menu design. First, there is still a distance before the GS model can actually be used as a tool for the designer to evaluate design alternatives or find optimal designs with respect to visual perception. One of the reasons is that our model deals only with a single menu at a time, more specifically, arrangement of colors and placement of menu options within a vertical menu. Second, it does not address memory or search strategies. And third, the requirement that users know the exact visual appearance of a target before the search does not sound realistic either.

We also identify some ways that could possibly improve the performance of the current model. As discussed in Section 3.1, shape characteristics were not included in the current model. However, for top-down activation, a similarity score of two words might be calculated by comparing letters at corresponding locations from a confusion matrix (Van Der Heijen, Malhas & Van Den Roovaaart, 1984). The performance of the model might be improved with some of the shape characteristics included. The current model did not include certain scanning strategies, such as starting at the top and moving down, because all users may not consistently follow it. Research shows people use both random and systematic strategies in menu search (Hornof & Kieras, 1997). However, certain scanning strategies could be incorporated into an elaboration of the GS model. For example, if there is a general scanning order to menu search, then an elaborated GS model could include location as a stimulus feature and assign parameters to different locations. Such a model would need to be able to predict when the pop-out effects from differences in features would overcome the default scanning strategies. Although there are details to sort out, the creation of such a model seems possible. On the other hand, the ability to predict memory effects seems more daunting. Current models of memory are able to account for general statistical properties of memory but are unable to predict recall or recognition in a particular situation. Such prediction would be necessary for the model to be applied to menu design.

Because the GS model only deals with visual search, it is natural to think of applying it in more visually based interaction, such as icon search. This involves building the model with additional variables such as shape, size or orientation. This could allow the GS model to be applied to search of icons or hardware panels. However, this is still different from the approach of modeling the propositional representation of a target in icon search (May, Tweedie & Barnard, 1993) because there is no assumption about the user's semantic knowledge of the target.

The basic approach we have used here can also become a tool to investigate the importance of different variables. For example, we could use the GS model we defined here to assign colors to a menu design that has fixed the location of the menu items. We could then study the benefit of color vs. black and white menu designs. By systematically varying what is optimized and what is constrained in the design process, it may be possible to identify the relative importance of different stimulus characteristics.

We thank Haolong Ma for programming the experimental software.

References

Aarts, E. & Korst, J. (1989). *Simulated Annealing and Boltzmann Machines*. Chichester: John Wiley & Sons.

Armstrong Laboratory Human Engineering Division. (1994). *User's Guide: Computer Aided Systems Human Engineering Performance Visualization System (CASHEPVS)*, Version 1.0 for Apple Macintosh. Wright-Patterson AFB.

Cave, K. R. & Wolfe, J. M. (1990). Modeling the role of parallel processing in visual search. *Cognitive Psychology*, **22,** 225–271.

Fitts, P. M. (1954). The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology*, **47,** 381–391.

Francis, G. (2000). Design multifunction displays: an optimization approach. *International Journal of Cognitive Ergonomics*, **4,** 107–125.

Hornof, A. J. & Kieras, D. E. (1997). Cognitive modeling reveals menu search is both random and systematic. *Proceedings of ACM CHI 97: Conference on Human Factors in Computing Systems*, pp.107–114. New York: Association for Computing Machinery.

Hornof, A. J. & Kieras, D. E. (1999). Cognitive modeling demonstrates how people use anticipated location knowledge of menu options. *Proceedings* of *ACM CHI 99: Conference on Human Factors in Computing Systems.* New York: Association for Computing Machinery.

Howes, A. & Payne, S. J. (1990). Display-based competence — towards user models for menu-driven interfaces. *International Journal of Man Machine Studies*, **33,** 637–655.

May, J., Tweedie, L., & Barnard, P. J. (1993). Modeling user performance in visually based interactions. In J. L. Alty, D. Diaper & S. Guest, Eds. *People and Computers VIII*, pp. 95–110. Cambridge: Cambridge University Press.

Meyer, J. (2000). Performance with tables and graphs: effects of training and a visual search model. *Ergonomics*, **43,** 1840–1865.

Norman, K. L. (1990). *The Psychology of Menu Selection*. Norwood, NJ: Ablex Publishing.

Payne, S. J., Richardson, J. & Howes, A. (2000). Strategic use of familiarity in display-based problem solving. *Journal of Experimental Psychology: Learning*, **26,** 1685–1701.

Treisman, A.M. & Gelade, G. (1980). A feature-integration theory of attention. *Cognitive Psychology*, **12,** 97–136.

Van der Heijden, Malhas & Van den Roovart (1984). An empirical interletter confusion matrix for continous-line capitals. *Perception and Psychophysics*, **35,** 8548.

Wolfe, J. M. (1998). What can 1 million trials tell us about visual search? *Psychological Science*, **9,** 33–39.

Wolfe, J. M. (1994). Guided search 2: a revised model of visual search. *Psychonomic Bulletin and Review*, **1,** 202–238.

Wolfe, J.M. & Gancarz, G. (1996). Guided search 3.0: a model of visual search catches up with Jay Enoch 40 years later. In V. Lakshminarayanan, Ed. *Basic and Clinical Application of Vision Science*, pp. 189–192. Dordrecht, Netherlands: Kluwer Academic.

(Paper accepted for Publication by Associate Editor, Dr Michael Harrison

AUTHOR QUERY FORM

HARCOURT PUBLISHERS

JOURNAL TITLE: IJHC DATE: 18/2/2002
ARTICLE NO. : 20020527

Queries and / or remarks

Manuscript Page/line	Details required	Author's response
	Please check the citation of Table 1 in the text, It is ok?	