Testing User Interaction With a Prototype Visualization-Based Information Retrieval System

Sherry Koshman
School of Information Sciences, University of Pittsburgh, 135 North Bellefield, Pittsburgh, PA 15260.
E-mail: skoshman@mail.sis.pitt.edu

The VIBE (Visual Information Browsing Environment) prototype system, which was developed at Molde College in Norway in conjunction with researchers at the University of Pittsburgh, allows users to evaluate documents from a retrieved set that is graphically represented as geometric icons within one screen display. While the formal modeling behind VIBE and other information visualization retrieval systems is well known, user interaction with the system is not. This investigation tested the designer assumption that VIBE is a tool for a smart (expert) user and asked: What are the effects of the different levels of user expertise upon VIBE usability? Three user groups including novices, online searching experts, and VIBE system experts totaling 31 participants were tested over two sessions with VIBE. Participants selected appropriate features to complete tasks, but did not always solve the tasks correctly. Task timings improved over repeated use with VIBE and the nontypical visually oriented tasks were resolved more successfully than others. Statistically significant differences were not found among all parameters examined between novices and online experts. The VIBE system experts provided the predicted baseline for this study and the VIBE designer assumption was shown to be correct. The study's results point toward further exploration of cognitive preattentive processing, which may help to understand better the novice/expert paradigm when testing a visualized interface design for information retrieval.

Introduction

A visual display of queries and/or their resulting data sets offers a novel approach to seeing relationships among retrieved data items for information retrieval (IR). Unlike scientific visualization, which primarily models real-world objects, information retrieval visualization attempts to model intangible concepts usually derived from linguistic information and utilizes an abstract graphical analogue of the document set.

Received February 25, 2004; revised June 21, 2004; accepted June 21, 2004

© 2005 Wiley Periodicals, Inc. • Published online 1 April 2005 in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/asi.20175

Visualization is a means to convey complex data so that users can more easily interpret it, but it imposes its own cognitive load by attaching meaning to symbols, shapes, arrangement, and visual metaphors. The assumption made by designers of visual information retrieval systems is that visual conventions are intuitive to use or easy to interpret by users of the system. Therefore, one of the most pressing questions about visualization-based IR systems is: Can people use them? This research focuses on answering this question for a prototype system named VIBE (Visual Information Browsing Environment). The development of VIBE has progressed under numerous designer assumptions about potential system users. One of the most important assumptions stated by the designers is that VIBE is a tool for a smart user and dependent upon a user's knowledge of the data displayed (Olsen, Korfhage, Sochats, Spring, & Williams, 1991b). A smart user implies an expert VIBE user and it is not known whether or not VIBE is usable by a broader spectrum of users, such as online searching experts and novices.

This assumption inspired the following questions: How well must a user understand the system to interpret the system's display of the data? And how well does the system present itself to facilitate this knowledge? This study is the first formal user evaluation of the PC-based VIBE system and it examined users with different levels of expertise in regard to their interaction with the system. A complementary article presents the results from a comparative study of user performance with VIBE to a traditional text-based system (Koshman, 2004).

Visualization-Based IR System User Studies

User evaluations on visualization systems vary in their formality depending on their overall objective. In most studies the focus is on examining dynamic user interaction with the system that is in a development phase or what may be labeled a prototype system. Some researchers obtained a quick snapshot of user interaction with a new interface (Carey, Kriwaczek, & Ruger, 2000), whereas others designed a more formal experiment (Veerasamy & Belkin, 1996). The TREC (Text REtrieval Conference) Interactive guidelines have been applied to standardize user testing procedures with visualization systems especially in comparison to traditional text-based systems (Osdin, Ounis, & White, 2002; Swan & Allan, 1998).

Studies in visualization system usability generally use small participant groups sometimes representing novice and expert target populations (Sebrechts, Vasilakis, Miller, Cugini, & Laskowski, 1999; Swan & Allan, 1998). Controlled task formulation and limitations on task completion times are characteristic of several studies (Osdin et al., 2002; Sebrechts et al., 1999). A combination of objective and subjective measurements were collected through questionnaires, data logging tools, and verbal protocols as reported by Nowell and colleagues (Nowell, France, Hix, Heath, & Fox, 1996). This study extends the established method of using novice and expert participants to include a VIBE system expert group. Query tasks were supplied for testing, however time limits were not imposed so that the full duration of the task may be measured.

Findings from visualization-based studies are mixed and are largely reported in the context of comparisons with traditional text-based IR systems. Newby's (1993) study showed that subjects rated the visualization system well, but generally preferred the familiar text-based system. Veerasamy and Belkin (1996) found slight trends favoring the visualization tool, but these were not statistically significant. Swan and Allan (1998) showed that the librarians preferred the traditional IR system and the general user group liked the experimental visualization system. Osdin, Ounis, and White's (2002) study showed that the results were positive for the *HuddleSearch* experimental system.

Morse, Lewis, and Olsen (2002) present a basic stepwise (BASSTEP) methodology that divorces the visualization from the system to test the value of the visualization technique itself. This approach tested visualization displays with large participant group sizes ($>$100). The text, iconic, table, graph, and "spring" (VIBE) information retrieval displays were compared for Boolean and vector studies, using two and three terms, performed on paper or online. The study showed that users preferred graphical methods when the task difficulty became greater and that the simplified VIBE display was successful for the most complex task.

One of the primary characteristics of testing users with a visualization-based IR system is that each study is unique because each system's visualization is different. For example, an expanded Venn diagram in *InfoCrystal* provides visualized representations of Boolean query results (Spoerri, 1993). Quite different is *TileBars*, which visually parses full-text documents into document rectangles of varying lengths that correspond to document length. The frequency and distribution of terms within each document rectangle are presented as rows of varying gray-colored squares. The darker the gray square, the greater increase in term frequency (Hearst, 1995, 1999).

The visualization interface is distinct in its formal modeling, choice of icons, and display methods. The iconic representation may be arbitrary to new users and needs to be learned, but the visualization display itself is rooted in human information processing. Ware (1999) outlines gestalt laws for pattern perception that users automatically employ to see patterns in a data visualization display. The gestalt principles provide a useful overview for examining key design precepts in visualization-based IR system interfaces from the user's perspective. The laws discussed here include proximity, closure, and continuity.

Proximity or how people perceive the grouping and spacing of elements is one of the most useful principles for interface design (Ware, 1999). It is relatively straightforward for users to distinguish groups or clusters of items in an interface display. Many IR-based visualization system displays exhibit proximity when spatial relationships are established among the document sets and queries.

The earlier *Bead* interface displays patterns of articles in space based on particle force and motion theory (Chalmers & Chitson, 1992). *Galaxies* offers a night sky view that uses document and cluster similarity to define the spatial proximity of clusters and documents as stars in two-dimensional space (Wise et al., 1999). *Lyberworld's* relevance sphere capitalizes on visualizing the "attraction" of documents to key terms and spatially clusters the geometric document icons on this basis (Hemmje, Kunkel, & Willett, 1994). From the user's vantage point, proximity processing is used to view VIBE's rectangular document icons' positions relative to the query's key words. Similar to Lyberworld, VIBE's document icons are positioned closer to the keywords that influence those documents. VIBE's geometric document icon display indicates term frequency for each document and is represented as groups of varying sized rectangles that aid visual discrimination among document icons.

The *Radial* visualization introduced by Carey, Kriwaczek, and Ruger (2000) relies on not only the user's proximity processing to view the document set, but also the Gestalt Law of Closure when the brain perceives boundaries to enclose an object or an interface object has contours that separate space. We use closure properties to process a typical pie chart. The Radial display features a circle that has keywords plotted along its contour. The inside of the circle contains document icons represented as small x's, which are situated according to the keyword's strength in the documents.

Carey et al.'s work (2000) also exemplifies closure in the Tree Map visualization that uses black lines to separate "super clusters" and white lines to distinguish "sub clusters" depicted as varying sized rectangles. Each cluster contains documents and is labeled with the most frequently occurring keyword. Similarly, the early self-organizing semantic map makes manifest the closure law to delineate term frequencies and associations. The rectangular map contains fixed-spaced nodes, some of which are numbered to show the number of documents on that node. The map is separated by lines into asymmetrical "word areas" and the size of the area corresponds to term frequency—the larger the area, the higher the occurrence and the position of the area associates it with neighboring word areas (Lin, Soergal, & Marchionini, 1991).

The Gestalt Law of Continuity refers to our ability to organize visual entities on the basis of continuous attributes, such as arched lines (Card, Mackinley, & Shneiderman, 1999). Ware points out that researchers have expanded this concept to include connectedness, where one object may be perceived as connected to another by means of a line to group them. Sebrechts et al.'s (1999) two-dimesional (2-D) *Nirve* interface provides connectedness by means of colored straight lines drawn between document cluster boxes to depict conceptual similarities on a flattened 2-D sphere. Similarly, the Hyperbolic Tree Browser connects concept boxes with slightly curved lines to visually aid user navigation in a large information tree structure (Pirolli, Card, & Van Der Wege, 2001).

Although the Gestalt laws for visual processing are well suited for understanding better the user's overall perceptual pattern building when using VIBE and other visualization-based IR interface displays, these principles do not help with the deciphering and decoding of the interface's icons or symbols. Users must focus their attention on the display's glyphs to process the information display and learn their meaning. Ware (1999) defines a glyph as a graphical object that symbolizes a data entity with multiple dimensions. Examples include cluster boxes or VIBE's polygonal icons. How effectively novice and expert users interpret and interact with VIBE's specific visualization constructs for information retrieval tasks are explored in this investigation.

Study Design

The research question is: What are the effects of the different levels of the user expertise on VIBE usability? The level of expertise is the independent variable and the operationalized components of usability are the dependent variables in this study. Level of expertise is defined in accordance with two aspects of Nielsen's (1993) model of computer user experience, which includes experience with the system and experience with computers in general. A third facet for level of expertise is experience with information retrieval techniques and online retrieval systems such as Dialog. Novice computer users, online searching experts, and VIBE system experts were used for this study.

Usability testing refers to the process of ascertaining that users can "find and work with the functions that meet their needs" (Dumas & Redish, 1993). In this study usability is operationalized by five measures outlined by Shneiderman (1998) as quantifiable human factor goals. They include system familiarity time, task performance speed, errors in task assignments, system feature retention, and subjective satisfaction. The first four variables were measured through user data logs, task assignment responses, and timing data. The remaining variable, subjective satisfaction, was examined through the participants' answers to post-search questionnaires and will be examined in a subsequent article.

It is hypothesized that users representing the three different levels of expertise (VIBE experts, online searchers, and novices) will exhibit significant performance differences in completing VIBE tasks. It is predicted that participants with expertise in using VIBE will serve as a baseline and perform faster in task completion, have low task error rates, and exhibit strong retention of system features. Online experts are predicted to exhibit better performance on all measures than the novice group. It is assumed that the online searching expert group will transfer their retrieval knowledge when using VIBE to facilitate faster performance times.

This investigation uses a quasi-experimental design that seeks to control as many factors as possible, but allows the entire system to be tested and evaluated by users in as close to a real operating mode as possible. The study is not a strict experimental design because subjects were not randomly selected for the testing although they were representative of a larger population who may eventually use this software. A repeated measures structure is used since the same participants are measured in each of the conditions (Kirakowski & Corbett, 1990).

Two pilot studies examined novice user interaction with VIBE prior to the final study. Fifty-three library and information science students participated in the pilot studies and the findings were used to refine the methodology, procedures, and data collection instruments for this investigation. Pilot studies ensured the study's reliability in tracking errors inherent in the design and instruments.

System

VIBE originated with Kai Olsen at Molde College in Norway and was developed in conjunction with Robert Korfhage and other researchers in the Department of Information Sciences at the University of Pittsburgh (Olsen, Williams, Sochats, & Hirtle, 1991a; Olsen et al., 1991b; Olsen, Korfhage, Sochats, Spring, & Williams, 1993). When a user evaluates search output with VIBE, the system presents the entire retrieved set of items (all those with one or more terms in common with the query). Information is visualized by the retrieved set of items and the query's keywords, which are graphically depicted on the screen as geometric icons. A VIBE user may examine how the documents are influenced by the query's keywords. The influence is visualized spatially by the proximity of the document icons to the user-defined points of interest (POIs), which represent keywords used in the query.

Figure 1 shows a screen display from PC-based VIBE, version 3.5. The query's keywords or POIs are round icons beside the terms "gorbachev," "soviet," "economy," "leadership," and "power." The different-sized rectangles represent documents that are plotted in the 2-D space. The size of the document icon is determined by term frequency calculations for each selected keyword (POI) in the document.

Document icons are positioned on the screen to show their relationship based on the distance and relative position to one or more POIs. If the document is influenced by only one POI, then the document icon will be situated on that POI. If the document is influenced by two or more POIs, then a positioning algorithm begins by calculating the document

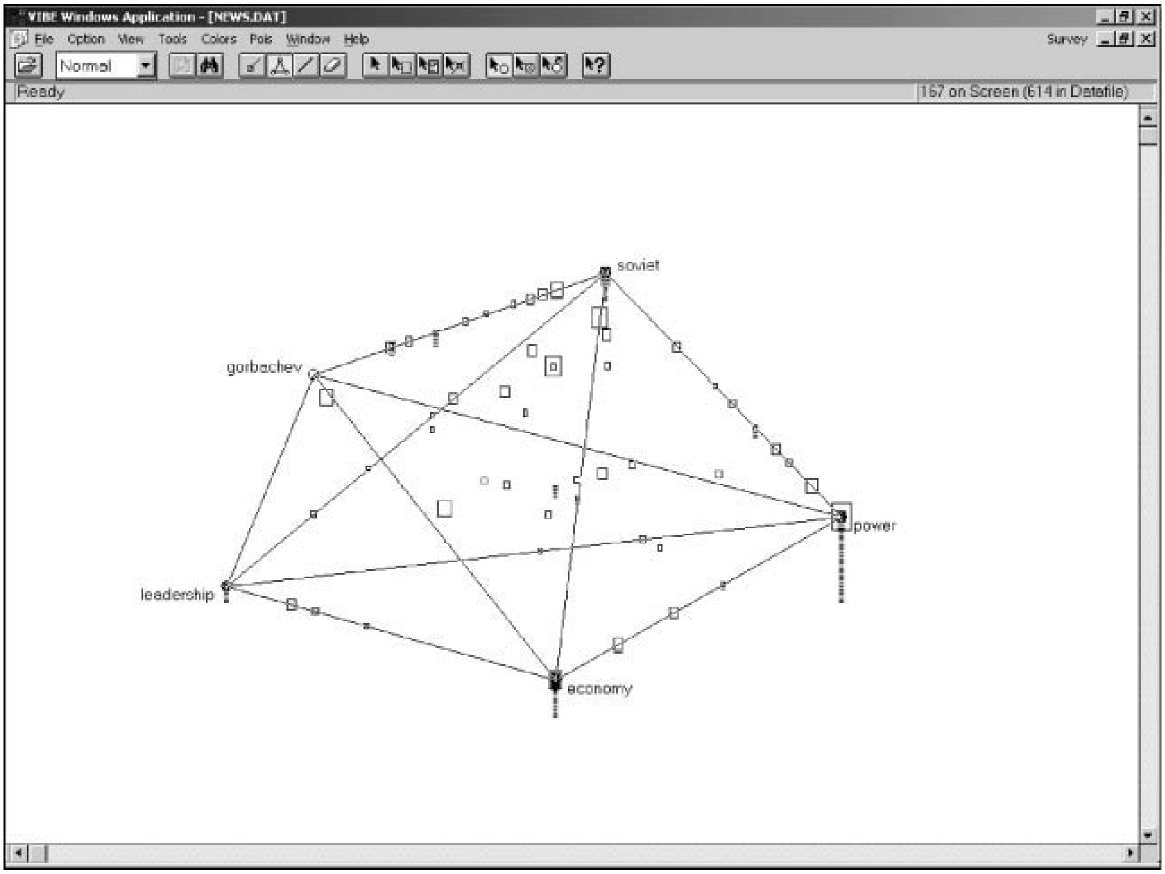

FIG. 1. VIBE screen display using the net feature.

icon position between the first two POIs. This is done according to the term frequency score of those POIs and the POIs' positioning data (*x*–*y* coordinates). The term frequency scores are added and the combined score, along with the new position information, is then used to form another combined score with the third POI and positioned according to the third POI's coordinates. This process is iterated until the document icon is placed in position to all the POIs on the screen that influence it (i.e., those that are contained in the document). Multiple documents containing only one POI are represented by stacked dashes underneath the POI icon.

Points of interest are always selected by the user to construct a query. Points of interest may be placed on the screen directly by the user, or by the system in a default circular arrangement on the screen. The VIBE user is allowed to manipulate the display by repositioning the POIs, removing POIs, or ignoring POIs. PC VIBE contains a well-developed set of interface options to explore the document space and a list of the most significant software features is provided in Table 1.

PC VIBE v3.5 was installed on the local area network of the School of Information Sciences (SIS) Teaching Laboratory. A subset of 614 Associated Press (AP) news articles derived from the TREC collection dated December 28–31, 1989 made up the database. The articles and 28 POIs (key terms) were loaded into VIBE.

Participants

Thirty-one participants representing three levels of expertise were tested with VIBE. Fifteen novice users possessing general computer experience were drawn from students in the Application of Microcomputers course offered by the Department of Library Science at the University of Pittsburgh. Novice users have minimal experience in online searching techniques and no prior experience with VIBE.

A group of 12 online searching experts was recruited from the pool of librarians working for the University of Pittsburgh Library System, Carnegie Mellon University, and special libraries in the Pittsburgh area. Online searchers were recruited by letters and posters that asked for professional librarians who had at least 3 years of online searching experience and extensive experience with computers. Online searcher participants had not used VIBE prior to this study.

The four VIBE experts were drawn from the Department of Information Science thus constituting the entire VIBE expert population from the University of Pittsburgh with the exception of the programmer and faculty designer. The experts had worked with VIBE for more than one year and were familiar with VIBE's internal structure. Additional facets of VIBE expertise may have been derived from attending VIBE demos and/or using similar systems. VIBE experts also had general experience with computers and

TABLE 1. VIBE display manipulation features.

VIBE feature	Description
Color	Used to color a POI; the document icons that contain that POI are automatically changed to the same color. If two or more different colors are used to color the POIs, then the result is an intersection set color of red which shows the document icons that contain two or more of the key terms (POIs).
Displacement	Shows the influence of one POI on a set of documents. The user manipulates the pointer to move the POI position and lines emanate from the document icons to show their previous positions and length of movement. Longer lines show that the POI had a greater influence on that particular document.
Lasso	Draws a box around a group of document icons so that a cluster of related documents may be viewed.
Lens	Displays the document identifier (e.g., title) in a pop-up window that lists all the document identifiers or headlines when the document icon is selected. May be used to view full-text documents.
Lines	Visualizes the relationships between all the POIs and document icons on the screen. The strength of the relationships varies from weak (infrequent term occurrences) which are represented by dashed lines to strong (frequent term occurrences) which are represented by solid bold lines.
Net	Permits the user to display axes between the POIs and is useful for determining the influence of two particular POIs on a document icon.
Star	Allows the user to click on any document icon to discover which POIs influence the document. This relationship is visualized by lines emanating from the document. Each line represents the POI it is pointing toward. The length of the line depicts the POI strength on the document.

experience with information retrieval. It was expected that VIBE experts would provide a standard performance level for the tasks.

Materials

A user profile questionnaire, answer sheets, task assignment forms, diskettes for logging data, and post-search questionnaires were used in this study. The user profile questionnaire contained 35 demographic factors to gather background information on the participants. Fixed response questions regarding work experience, educational background, online resource searching, frequency, length, and type of computer usage provided background data.

Task Assignments

Participants were given four sets of seven VIBE tasks: two sets in session 1 and two sets in session 2, resulting in 28 individual tasks. They logged their system familiarity time and their assigned tasks with VIBE using the "Survey" menu, which was written into the VIBE interface specifically for this study. The data logging (survey) tool was tested successfully during the second pilot study and it captures the VIBE-specific features accessed and used by participants.

The tasks were designed to test the system's ability to make its features accessible and understandable to the user. The user's responsibility was to select the appropriate features to complete the task. The tasks identified the POIs to be selected by the participants and were designed to test their interpretation of VIBE's document display. Two of the seven tasks were constructed to specifically test the user's visual discrimination of individual documents presented by VIBE.

Dumas and Redish's (1993) task construction model was applied to define the task scenario, the task objective, task set-up, and task description for each task. They suggest that two or three scenarios be used to measure similar tasks to control for user errors. The final study contained different scenarios (different AP news story topics) for the four sets of seven different tasks. An example of this model applied to VIBE is shown in Table 2.

Participants were tested in two different sessions that were held approximately a week apart in a group setting in the Teaching Laboratory at the School of Information Sciences (SIS) at the University of Pittsburgh. Each group—the VIBE experts, online searching experts and novices—was tested separately. A half-hour system description and demonstration of VIBE features was given to the novice and online searcher groups. The VIBE experts did not require training. This investigation controlled the environment variable as much as possible by using training scripts for each session.

The pilot studies revealed interesting dynamics during group testing such as the participants viewing others' screen displays and conferring during task assignments. The final investigation successfully controlled for these effects by randomly varying the task assignment order among participants.

TABLE 2. Task construction model example. The scenario for this investigation builds on selecting a popular Associated Press story, providing a brief synopsis of the story and asking participants to find related news articles. Participants were asked to conduct this IR task to find more information on the subject of the news story.

Scenario: Please find more news articles related to the news article summary provided to you. This scenario applies to all tasks, but the subject matter changes for each task group. Participants are provided with the keywords to use in the query.

Task 1

Task objective: Can users find document items that relate to all keywords in their query?

Test set-up: How many newspaper articles contain all the keywords or POIs identified in the query?

Task description: Finding the star feature, invoking the star feature, counting the lines emanating from the document icon, examining the status bar for the headline.

Task 2

Task objective: Can users execute a Boolean intersection and find the number of documents items related to two or more aspects of the query?

Test set-up: How many newspaper articles relate to the keywords 'communist' and 'reform'?

Task description: Selecting 1 color option for 1 POI and another color option for another POI, counting the document icons influenced by the system collision color—red.

Data Analysis—Demographics

The user profile questionnaire gathered information on participants to validate their placement in the three groups (novices, online experts, and VIBE experts). The gender distribution for the entire group was 25.8% male and 74.2% female. Online searching experts represented an older group of participants in comparison to the novices and VIBE experts. Overall, 22.6% of participants were in the 20–30 age groups, 35.5% in the 31–40 age groups, 25.8% in the 41–50 age group, 9.7% in the 51+ age group.

All novice participants were enrolled in the Application of Microcomputers course and were pursuing a Master in Library Science degree. A large percentage of the online searching group (67.7%) had completed a Masters degree in Library Science, one participant had a PhD in Library Science (LS), and three were working toward a PhD in Library Science. All VIBE experts were pursuing doctorates in Information Science (IS).

The distribution of participants who took computer and statistics courses did not reveal any striking patterns. The only meaningful finding is that almost half (46.7%) of novices reported "never" using online services. In comparison, 83.3% of the online searchers used online services "frequently" and none of them reported "never" using online services. Overall, the online searching group demonstrated the highest percentage of frequent usage for online resources when compared to the novices. The online searchers and VIBE experts also reported longer and more frequent use of computers than the novices. These data confirmed the participant groupings.

Feature Log Analysis

Data log files were analyzed to discover which VIBE features the participants used during familiarity time and task completion. The log tool recorded 27 VIBE interface options that were activated by subjects. Recorded actions are specific to VIBE and the tool does not log any options running under the standard Microsoft Windows menus. There were a few limitations inherent in the data logging tool for VIBE features. Findings regarding the help option were not totally accurate since the use of help could not be consistently recorded for all participants.

Log files were analyzed by means of a Unix shell script and features used during the VIBE familiarity time were examined. The VIBE experts' behavior was predictable in that most of them used all of VIBE's features. Varying percentages of novices tried all the features identified in the log list, and the online expert group tried all but the adjust feature. Both novices and online experts selected many features to familiarize themselves with VIBE.

A higher percentage of online searchers than novices gained familiarity with overall system control features such as removing POIs and removing attributes, as opposed to the more specific features that the novices used for system exploration. An interesting finding is that the online experts demonstrated a consistent pattern for selecting and deselecting the same feature, whereas a large percentage of novices selected a feature, but did not follow through to deselect the feature to see how it could be invoked and revoked, providing an overall sense of system control.

Four features used most frequently were: the icon hit, which allows users to click on a document icon (15.5%); the star feature, that allows the user to see which POIs influence the document (13.5%); the lens feature, which accesses the full text of the document (10.2%); and the color option used to visualize an intersecting set of documents (8.5%).

Task Completion Times

Data were collected to measure system familiarity and task performance times. Mean times and their standard error (SEM) are presented graphically. Novice and online searcher group timings were analyzed for significant differences using a two-way, nonparametric ANOVA. VIBE experts were not included in the statistical analysis since they provide a standard for the other two groups.

No significant differences were found between novices and online searchers in their familiarity times with the systems. Both the novices (13:48 minutes:seconds [mm:ss]) and online searchers (10:54 mm:ss) approximated the mean familiarity time shown by the VIBE system experts (11:46 mm:ss).

Task times using VIBE showed improvement between sessions 1 and 2. Significant differences were found between the first session (VIBE1) and second sessions (VIBE2) for four of the seven (*a*–*g*) individual task timings. Performance times between tasks for the first and second trials favored the second trial with the VIBE interface for task *a* (1.79 vs. 1.21, mean rank of VIBE1 and VIBE2 respectively, $p < 0.01$), task *b* (1.77 vs. 1.23, $p < 0.05$), task *c* (1.79 vs. 1.21, $p < 0.05$), and task *g* (1.87 vs. 1.13, $p < 0.01$).

Overall, the novices and online experts did not exhibit statistically significant differences in their performance times across tasks for VIBE. Significant differences were found between the novices and online searchers for task *b* ($p < 0.05$) and task *d* ($p < 0.01$) in their second trial with VIBE2 (Figure 2).

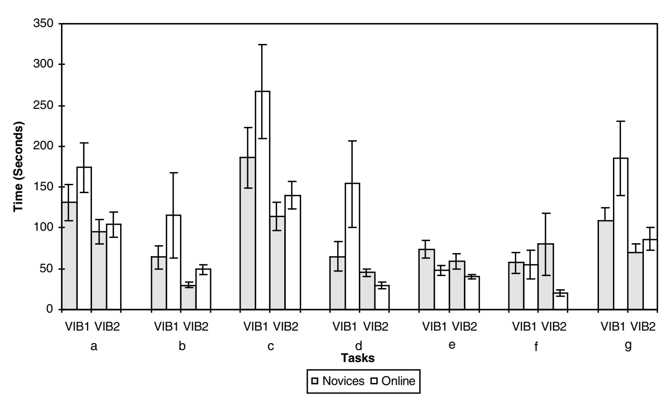

FIG. 2. VIBE sessions 1 and 2 mean times ± SEM.

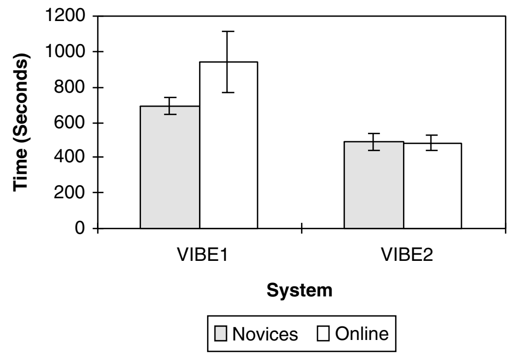

FIG. 3. Total task mean times ± SEM for VIBE sessions.

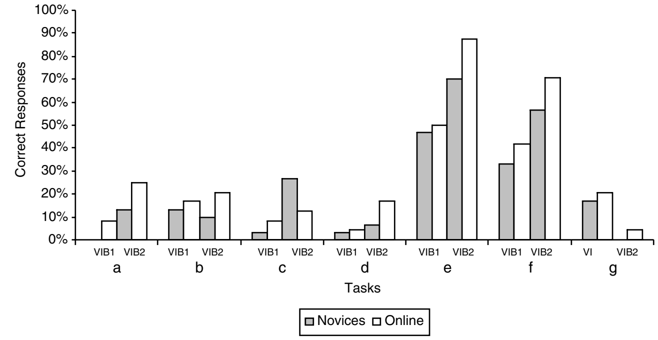

FIG. 4. Percentage of correct responses between sessions 1 and 2.

Total mean times ± SEM for the complete VIBE task sets over both sessions favor the second session (1.95 vs. 1.05 mean rank for VIBE1 and VIBE2, respectively [$p < 0.0001$]). The online searchers who were initially slower showed greater improvement as indicated by the significant interaction ($p < 0.05$) found between sessions (Figure 3).

The VIBE system experts provided a solid baseline for timing measurements in that they performed their tasks more quickly than the two other participant groups with the exception of the last task in the set where their mean time exceeded the online expert and novice groups, perhaps due to an inherent limitation in that task design for the expert group.

Task Error Rates

In session one, the novices did not solve many tasks correctly for VIBE task sets 1 and 2. None of the novices correctly solved the first task that required a document icon count. The percentage of incorrect responses was high for most of task set 1 except for the visually oriented tasks (*e* and *f*). No significant differences were found in the task error rates between VIBE task sets in the first session. For task sets in the second session with VIBE, a high percentage of novices experienced task errors on all tasks except for the visually oriented tasks once again. There were no significant differences between the task sets executed in the second session.

Online experts presented a similar pattern to novices for task error results. The percentage of online participants who gave incorrect answers was high for most of task set 1, except for the visual tasks. A slightly different pattern emerged during the online experts' task responses with VIBE in task set 2, but this result was not statistically significant. The task solution pattern for online experts in the VIBE portion of session two was similar to session one except that a larger percentage of online searching participants solved the visual tasks (*e* and *f*) more correctly than the other tasks in the set (Figure 4).

The first visual task (*e*) used icon size as a visual cue to select the correct answer. The other task (*f*) relied upon the participants' ability to access VIBE's full-text data through the lens feature and find a specific author's name. VIBE experts solved more tasks correctly in session two than in session one. In the second set of VIBE tasks, all tasks except *d* and *g* were solved correctly by 75% or more of the VIBE experts.

The most important finding is that participants solved the visually oriented tasks with VIBE better than the other tasks in the set. These tasks relied on visual cues by asking participants to find the largest document icon on the screen and to find the author of the article represented by that icon.

Discussion and Conclusion

The VIBE experts solved the tasks more accurately than novice and online expert participants and provided baseline performance measures that were predicted in this study. In general, the level of expertise did not constitute a notable impact on conducting typical IR tasks using VIBE, but had interesting implications for visual discrimination IR tasks. Statistical tests on task errors showed that overall there were no significant differences between sets of tasks presented for VIBE in session one between the online searchers or novices. In the second VIBE session, the magnitude of correct responses for some of the tasks increased between task sets in the same session. This is interesting since repeated use of VIBE points toward an increased level of task solution.

Total VIBE task set times showed substantial improvement between participants' first and second sessions with VIBE. This difference was statistically significant and there was a trend toward a time decrease in most of the individual tasks performed from session one to session two and in the total task performance times. The participants' more time efficient performance with VIBE in the second session indicated some system feature retention which points toward VIBE being a learnable system. This factor was also supported by the improved task error rate during the second session.

The feature log analysis provided valuable insight into the selection of interface options used by participants to familiarize themselves with the system and to resolve the tasks assignments. This tool objectively recorded user choices while they interacted with the system. In some instances, participants selected the appropriate VIBE features to conduct the task, but were unable to complete the task successfully. This finding presented an interesting situation.

The selection of interface features may be analogous to the selection of power tools for a household task. Imagine multicolored leaves strewn across a lawn in the autumn. To collect the leaves, a person might choose a leaf blower over a rake, although they may not completely understand its mechanical operations, nor picture exactly what its final result on the lawn may look like. Does the machine cluster, scatter, or pile leaves? It is possible that the participants understood which VIBE interface option was appropriate for the task, but experienced difficulty in understanding the feature's application or in determining if they had the correct results after using that feature.

Users employ predictive cognitive modeling when interacting with a system and these predictions are more accurate if they are based upon previous experience or if the system is more frequently used. For example, many end users can probably describe what a typical Google Web search will display. New users of VIBE do not have a basis to form a predictive model either from a real-world analogy or frequent visualization system use so their interpretations of the interface projections are less realistic. The overall results of this investigation indicate that although the display takes advantage of basic perceptual organization, VIBE's visual pattern of glyphs may require more strenuous cognitive processing for novices than for the system experts. VIBE's geometric glyphs encode the query terms' influence on a resulting document, the intensity of that influence reflected in the size of the icon and the relevance of each document in the set to each other.

Ware (1999) discusses the merits of glyphs in terms of integral and separable display dimensions that are used for categorizing tasks. An integral dimension is one where two or more facets are perceived in a holistic manner. A separable dimension requires the user to make a separate decision about each element of the graphical facet. Figure 5 borrows from Ware's example and applies it to the VIBE polygonal three document set display.

In example 1, the two dimensions defining the glyph are height and width, and in this instance B and C are perceived as more alike. In example 2, the dimensions are height and shading. A and B, which are shaded gray, are perceived to be more alike than C, which is white. Ware points out that integral dimensions are processed more efficiently than the separable dimensions for holistic recognition tasks. Separable dimensions require users to process separately the size and the color of the glyphs and are more useful for analytical tasks.

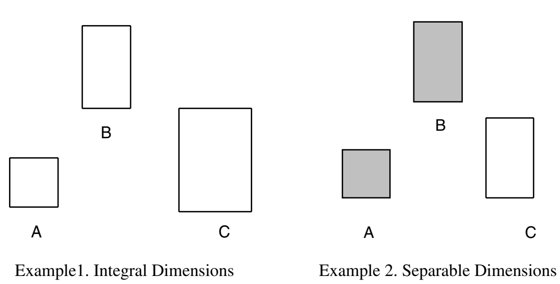

FIG. 5. Integral and separable dimensions using rectangles.

In VIBE, the document set glyphs inherently exhibit example 1's integral dimensions in the default display. In contrast to the glyphs themselves, the document icons' *x*, *y* positioning relative to the keywords' circular icons and the document icons' positioning data relative to other document icons renders the display dimensions as highly separable and thereby requires more of the user's attention to process each dimension.

Novices' inability to interpret VIBE's document display effectively may be a result of the complexity introduced when processing the separable dimensions of the glyphs in the document display. To solve tasks successfully, the user must visually categorize the document icons according to size and document position relative to the query terms and the other documents in the set. In comparison, the VIBE system experts have successfully learned one or more of these glyph dimensions. This gain toward more automatic dimensional encoding could result in the expert's faster display interpretation and task resolution.

This notion conforms to Anderson's (1990) research into the stages of skill acquisition that transform into expert performance. The autonomous stage is reached when the skill becomes more rapid and automatic. In the VIBE scenario, the expert skill may be the ability to efficiently process the dimensional encoding. Further research is warranted to test this theory and to learn which of these dimensions was primarily used or in what order the dimensions were parsed by the experts to solve the tasks. Also, more research is required into the user's cognitive investment when processing multivariate data display dimensions for more than three features in the VIBE glyphs.

From another perspective, participants who are new to VIBE solved specific visual tasks more accurately than broader (categorizing) tasks. The categorizing tasks that required the user to provide a document set metric are more cognitively challenging than the visual perception tasks. To understand how many documents contained terms *x*, *y*, and z, the user needs to formulate a mental model of the positioning algorithm of the document set. The thought process needs to revolve around the query terms' influence on the document icons. If only two terms influence the document, then it is positioned between them, a line can be inserted in the display and the documents can be counted. If there are three terms, then the document icons are shifted accordingly to represent the document set, and so forth. In addition, the user must manipulate VIBE features to identify the number of document icons influenced by the query terms specified.

This is a much more complex task than visually identifying the largest object on the computer screen. For example, preschool children are easily able to identify the largest item in a set of objects and this object recognition is a fast perceptual decision referred to as preattentive processing in which the human perceives a visual object before it is given focused attention. One of the most compelling ideas to

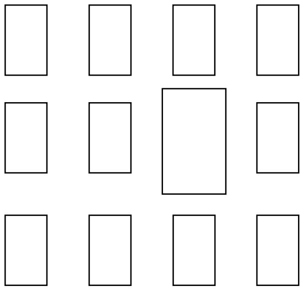

FIG. 6. Pop-out display using rectangles.

support preattentive theory in the design of information visualizations system for IR is a "pop-out." Pop-out refers to an object that can capture the user's attention against a background of distracting objects. Figure 6 shows an example of a pop-out display using a size feature for rectangles that resemble the document icons used in VIBE.

Pop-outs may be accomplished in an interface display by using other attributes such as color, orientation, shape, number, enclosure, motion, and more to exemplify the target item (Ware, 1999). Pirolli, Card, and Van Der Wage (2001) conducted eye tracking studies using the Hyperbolic Tree Browser and take into account the display's density of background distractors to enable visual target identification.

VIBE often offers users a high-density document display as a result of the query posed to the system. Its document icon size feature is a useful pop-out visual discriminating tool for users as shown in this investigation. Also, users were able to extrapolate their general knowledge of computer systems by using the mouse to click on the largest document icon to identify a document attribute (in this case, the document author). The combination of the pop-out visual aid with a familiar procedure to execute an option appears to be a powerful VIBE feature. Further testing of pop-out features against backgrounds of varying density for visual IR task resolution presents an interesting avenue for research.

While the overall task activities are realistic for information retrieval systems, there are problems in testing environments due to the lack of realistic query construction (Ledwith, 1992). Search questions developed for this study are realistic in terms of their overall IR activities, but the questions themselves are not derived from real users and this is standard practice among researchers in this field. A visualization information retrieval system's usability must not only be tested for basic retrieval tasks, but also must take into account human factor design which adds complexity to the testing process. This study lays the foundation for understanding better the role of task formulation in the context of studying user interaction with visualization systems for IR.

In conclusion, this investigation showed that the designer assumption that VIBE is a tool for smart or expert users is correct. Effective use of VIBE depends upon the expert's knowledge of the data and their system familiarity. This does not diminish the importance of applying visualization tools, such as VIBE, to general information retrieval since some of its unique features enabled efficient use of the system among novice users. However, the cognitive processing and interface design efforts needed to develop expertise with a visualization-based IR system, such as VIBE, require further consideration.

Acknowledgments

Comments and feedback on this paper from Edie Rasmussen are gratefully acknowledged. This investigation was a result of doctoral research conducted at the School of Information Sciences, University of Pittsburgh. I am grateful for the guidance and vision of my committee members Edie Rasmussen, Michael Lewis, and the late Robert Korfhage.

References

Anderson, J.R. (1990). Cognitive psychology and its implications. New York: W.H. Freeman.

Card, S.K., Mackinley, J.D., & Shneiderman, B. (Eds.). (1999). Readings in information visualization: Using vision to think. San Francisco: Morgan Kaufmann Publishers.

Carey, M., Kriwaczek, F., & Ruger, S. (2000). A visualization interface for document searching and browsing. Retrieved October 20, 2003, from citeseer.nj.nec.com/382997.html

Chalmers, M., & Chitson, P. (1992). Bead: Explorations in visualization. In Proceedings of the Fifteenth Annual International ACM/SIGIR Conference on Research and Development in Information Retrieval (pp. 330–337). New York: ACM Press.

Dumas, J.S., & Redish, J.C. (1993). A practical guide to usability testing. Norwood, NJ: Ablex.

Hearst, M. (1995). TileBars: Visualization of term distribution information in full text information access. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 59–66), New York: ACM/Addison-Wesley.

Hearst, M. (1999). User interfaces and visualization. In R. Baeza-Yates & B. Ribeiro-Neto (Eds.), Modern information retrieval (pp. 257–323). Harlow, England: Pearson Addison-Wesley Longman.

Hemmje, M., Kunkel, C., & Willett, A. (1994). Lyberworld—A visualization user interface supporting fulltext retrieval. In Proceedings of the Seventeenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 249–259). Berlin/Heidelberg/New York: Springer-Verlag.

Kirakowski, J., & Corbett, M. (1990). Effective methodology for the study of HCI. Amsterdam/New York: North-Holland Elsevier.

Koshman, S. (1997). User testing of a prototype visualization-based information retrieval system. Unpublished doctoral dissertation, University of Pittsburgh, Pittsburgh, PA.

Koshman, S. (2004). Comparing usability between a visualization and text-based system for information retrieval. Journal of Documentation, 60(5), 265–280.

Ledwith, R. (1992). On the difficulties of applying the results of information retrieval research to aid in the searching of large scientific databases. Information Processing & Management, 24(3), 249–255.

Lin, X., Soergal, D., & Marchionini, G. (1991). A self-organizing semantic map for information retrieval. In Proceedings of the Fourteenth Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 262—269). New York: ACM Press.

Morse, E., Lewis, M., & Olsen, K. (2002). Testing visual information retrieval methodologies case study: Comparative analysis of textual, icon, graphical and "spring" displays. Journal of the American Society for Information Science and Technology, 53(1), 28–40.

Newby, G. (1993). Towards navigation for information retrieval. Unpublished doctoral dissertation, Syracuse University, Syracuse, New York.

Nielsen, J. (1993). Usability engineering. San Diego/Boca Raton: Academic Press.

Nowell, L.T., France, R.K., Hix, D., Heath, L.S., & Fox, E.A. (1996). Visualizing search results: Some alternatives to query-document similarity. In Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 67–75). New York: ACM Press.

Olsen, K.A., Williams, J.G., Sochats, K.M., & Hirtle, S.C. (1991a). Ideation through visualization: The VIBE System. Pittsburgh: School of Library and Information Sciences, University of Pittsburgh.

Olsen, K.A., Korfhage, R.R., Sochats, K.M., Spring, M.B., & Williams, J.G. (1991b). Visualization of a document collection: The VIBE System. Pittsburgh: School of Library and Information Sciences, University of Pittsburgh.

Olsen, K.A., Korfhage, R.R., Sochats, K.M., Spring, M.B., & Williams, J.G. (1993). Visualization of a document collection: The VIBE System. Information Processing & Management, 29(1), 69–81.

Osdin, R., Ounis, I., & White, R. (2002). Using hierarchical clustering and summarization approaches for Web retrieval: Glasgow at the TREC 2002 Interactive Track. Retrieved October 20, 2003, from citeseer.nj.nec.com/580060.html

Pirolli, P., Card, S.K., & Van Der Wege, M. (2001). Visual information foraging in a focus + context visualization. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 506–513). New York: ACM Press.

Sebrechts, M.M., Vasilakis, J., Miller, M.S., Cugini, J.V., & Laskowski, S. (1999). Visualization of search results: A comparative evaluation of text, 2D, and 3D interfaces. In Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 3–10). New York: ACM Press.

Shneiderman, B. (1998). Designing the user interface: Strategies for effective human-computer interaction. Reading, MA: Addison-Wesley Longman.

Spoerri, A. (1993). InfoCrystal: A visual tool for information retrieval & management. In B. Bhargava, T. Finin, & Y. Yesha (Eds.), Proceedings of the Second International Conference on Information and Knowledge Management (pp. 11–20). New York: ACM Press.

Swan, R.C., & Allan, J. (1998). Aspect windows, 3-D visualizations, and indirect comparisons of information retrieval systems. In Proceedings of SIGIR '98, Conference on Research and Development in Information Retrieval (pp. 171–181), New York: ACM Press.

Veerasamy, A., & Belkin, N.J. (1996). Evaluation of a tool for visualizaton of information retrieval results. In Proceedings of the 19th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 85–92). New York: ACM Press.

Ware, C. (1999). Information visualization: Perception for design. San Francisco: Morgan-Kaufmann.

Wise, J.A., Thomas, J.J., Pennock, K., Lantrip, D., Pottier, M., Schur, A., et al. (1999). Visualizing the non-visual: Spatial analysis and interaction with information from text documents. In S.K. Card, J.D. Mackinlay, & B. Shneiderman (Eds.), Readings in information visualization: Using vision to think (pp. 442–450). San Francisco: Morgan Kaufmann Publishers.