Highlighting items as means of adaptive assistance

Lior Rokach[1], Bracha Shapira[1], Talia Lavie[2], Liat Antwarg[1], Joachim Meyer[2]
,

[1]Department of Information Systems Engineering, Ben-Gurion University of
the Negev and T-Labs @ BGU, Beer Sheva, 84105, Israel
{liatant,bshapira,liorrk}@bgu.ac.il

[2]Department of Industrial Engineering and Management, Ben-Gurion
University of the Negev and T-Labs @ BGU, Beer Sheva, 84105, Israel
{joachim, tlavie}@bgu.ac.il

Abstract

Providing adaptive help during interaction with the system can be used to assist users in accomplishing their tasks. We propose providing guidance by highlighting the steps required for performing a task that the user intends to complete according to the prediction of a system. We present a study aimed at examining whether highlighting intended user steps in menus and toolbars as a means of assisting users in performing tasks is useful in terms of user response and performance. We also examined the effects of different accuracy levels of the relevancy of the provided help and the help format on user response and performance. An experiment was conducted in which 64 participants performed tasks using menus and toolbars of a simulated email application. Participants were offered a highlighted guidance of the required steps in varying levels of accuracy (100%, 80%, 60%, and no guidance). Our results support the benefits of highlighted help both in user performance times and in user satisfaction from receiving such assistance. Users found the assistance necessary and helpful and by the same token not unduly intrusive. Additionally, users felt that such assistance generally helped in reducing performance time on tasks. We did not find a significant difference when users receiving help at 80% accuracy was compared to those receiving help at 100% accuracy; however, such a difference does appear for those receiving 60% accuracy. In such cases we found that the user's satisfaction level, perceived usefulness and trust in the system decreased while their notion of perceived intrusiveness increased. We conclude that assisting users by highlighting the required steps is useful so long as the minimal accuracy level is higher than 60%. Our study has implications on the implementation of highlighting next steps as a means of adaptive help and on integrating probability-based algorithms such as intention prediction to adaptive assistance systems.

Key words

Adaptive assistance, Highlighted help, Accuracy level, Menus and toolbars

1. Introduction

In recent years, applications have continuously increased in functionality and features, thereby complicating the interactions with the applications (Fischer, 2001). Assistance can be provided to help users perform or complete tasks, alert users as to the availability of features, as well as increase their general application knowledge. Effective assistance will allow users to acquire new skill sets and improve application aptitude all while learning to easily operate an unfamiliar device or function.

Assistance can be provided in many forms such as graphical representations, icons and bubbles appearing on the screen, textual tutorials, animated demonstrations and so forth. Assistance can be either requested by the user or offered by the system. Previous research has shown that users do not actively seek help (Novick & Ward, 2006) as they consider it a last resort (Novick et al., 2007). Nevertheless, prior attempts to offer users with help have not proven successful (e.g. Microsoft's assistance, Clippy). Inbar et al., 2009 have examined user responses to four different formats of help offered by the system, varying in the amount of assistance information initially presented to the user. They (Inbar et al., 2009) found that assistance that initially contained more information was perceived to be more intrusive and was less desired by the participants.

Unfortunately, offering users with assistance may be seen as intrusive when it is initiated by the system and not by the user themselves, as assistance might be suggested at inconvenient times (Jameson, 2003). In Microsoft's Office Assistant (Clippy), for example, the decision on when to help was based on a simple rule-based method that did not take into account the user's level of competence and willingness to be interrupted (Jameson, 2003).

Ideally, users would like to be offered assistance according to their specific task (goal). Current technology enables the prediction of user intended tasks by various intention prediction methods and machine learning algorithms (Sun & Giles, 2001). Since prediction of user intention is usually not 100% accurate, in some situations the provided help is not relevant to the user's activity. This situation is similar to other common scenarios where the provided help is not relevant. For example, when a user submits a search query to the help system and the system retrieves a non relevant help item. Shen et al. (2006) claimed that it is essential not to intrude the user with incorrect predictions; it is better to make no predictions if such predictions are inaccurate. Hence, before implementing technologies that might result with an inaccuracy of the help systems (such as intention prediction or help search), it is important to determine the accuracy threshold that differs between intrusive and useful assistance. Finally, due to the probability of receiving inaccurate help, it is important that the help will be provided in a format that will least intrude the user. In this paper we describe experiments conducted to examine the extent of user's tolerance to inaccurate assistance. To minimize intrusiveness we chose highlighting of the next predicted user's step as the investigated adaptive assistance format. We observed user's response to different variations of the format and accuracy levels.

It should be noted that although we have developed our own intention prediction algorithm (Antwarg, 2010), we have not integrated it (or any other intention prediction algorithm) into the experimental system in order to prevent dependence due to performance of specific algorithms. This, in turn, enables us to simulate and manipulate the level of accuracy at which the system provides the assistance. In addition, our results can therefore be generalized to the effect of integrating technologies that are known to be not 100% accurate into adaptive help (i.e., help retrieval) rather than only to intention prediction.

The next section reviews adaptive assistance, followed by a description of adaptive menus and toolbars. We then describe our research aims (section 4) and the experiments performed to achieve them in section 5. Section 6 details the results that are discussed in section 7 and section 8 concludes the paper.

2. Adaptive assistance

Bennet et al. (2001) believe that adaptive systems have the potential to improve system performance and, if designed properly, can help users to complete tasks. Nonetheless, if not properly designed, they may also degrade performance by presenting irrelevant information or by eliminating relevant information. One functionality that can be adapted is guidance (Pierrakos et al., 2003), or in other words, the effort to assist the user in arriving quickly to the information the user is seeking. This adaptation function not only increases the user's loyalty but also alleviates to a great extent the information overload problem that the users may face (Pierrakos et al., 2003).

Research on developing adaptive guidance systems is still facing the problem of recognizing user goals when the user does not perform actions leading towards that goal (Jameson, 2003). In the field of user modeling a number of approaches have been taken for predicting user goals in order to provide relevant guidance: (1) command recommendation systems, such as the system proposed by Davison & Hirsh (1998), that predicts UNIX system commands, (2) intelligent tutoring systems (Encarnacao & Stoev, 1999), and (3) agents similar to the one used in the Lumiere application (Horvitz et al., 1998), which introduced the idea of intelligent assistance to Microsoft users and proposed help topics on the basis of the user's recent actions. Considering that all the above methods are not completely accurate and the provided help is not always relevant, the importance of our results which determine the tolerable level of accuracy has implications on technologies that can be implemented for adaptive help.

According to Fischer (Fischer, 2001), intelligent assistance systems are traditionally divided into two classes: active and passive. In passive assistance, it is the user who initiates the help session by asking for help. In active assistance, the system initiates the help session itself. Our study focuses on the latter and provides users with assistance without them requesting it. This is accomplished by predicting the user’s intentions (goals) and guiding them through the required steps for accomplishing their tasks.

A similar concept was applied by Kunzer et al. (2004) who designed a user support system using action prediction algorithms (APA) which present the user with suggested next actions and hints based on the user's previous behavior. Unlike Kunzer et al.'s (2004) system which presented the suggested actions as a textual list, our system presents the guidance graphically (as will be described in section 5). Additionally, while Kunzer et al. (2004) emphasized the APAs they applied, our study focuses on user attitudes towards the help, independent of the specific algorithm used.

Only few studies have previously examined user attitudes towards adaptive assistance. One such study is a survey conducted in Germany (Rech et al., 2007) that examined attitudes towards intelligent assistance in software engineering activities in German software organizations. They reported that users preferred quickly executable assistance while complex types of assistance are hardly ever used. They also found that the respondents reported that they favored simple and visually perceptible forms of intelligent assistance (tooltips, lists and pictures), while animated and audible forms of assistance are regarded as distracting. Although assistance that is explicitly requested (active assistance) by the user was found to be preferred, one-third of the participants reported that they preferred assistance which is displayed upon the decision of the assistance system (passive assistance). The study also examined the forms of assistance preferred when creating and editing a document. Most respondents showed interest in completion of text, explanation of the currently edited document and highlighting unused parts (Rech et al., 2007).

To conclude, similar to adaptive systems, designing adaptive assistance poses many challenges. Although some previous studies have pointed to various factors affecting the success and failure of different adaptive systems (e.g., Jameson, 2003; Gajos et al., 2006; Findlater & McGrenere, 2008a; Gajos et al., 2008), they mainly addressed the questions of how, when and where to present the help (Rech et al., 2007).

Our study does not attempt to answer these questions but rather focuses on a number of key factors that are important when implementing adaptive assistance. It should be noted that we relate to the factors that are most relevant to our research purposes, though additional factors not presented here may also be of importance in different contexts.

2.1 Key factors for adaptive assistance

1) *System reliability*. We divide this factor into two dependent components; one component, accuracy, is a system objective feature. The other component, trust, is a user's subjective feature describing their attitude towards the system. The system accuracy affects the user trust as detailed below:

1.1) Accuracy. We use the term accuracy to refer to the percentage of times the system proposes correct and relevant suggestions to the user. Accuracy in adaptive interfaces was found to have a significant impact on user performance (Gajos et al., 2006; Tsandilas & Schraefel, 2005; Findlater & McGrenere, 2008b). Gajos et al. (2008) explored the relative effects of accuracy on the usability of adaptive interfaces using split menus and concluded that improvement in accuracy positively effects performance, utilization and some satisfaction ratings.

Various studies also examined system accuracy in the context of automation. When an automated system is not completely reliable the automation risks being useless, harming performance (Wickens & Dixon, 2007), decreasing user trust in the system and even causing users to ignore suggestions completely or to disable the suggestions offered by the system. Tiernan et al., 2001; Wickens and Dixon 2007 examined the implications of different levels of unreliability or imperfection of diagnostic automation.

The analysis suggests that performance is quite sensitive to the level of imperfection. When at a level of below 70%, unreliable automation is considered worse than no automation at all. In the context of adaptive assistance, which is the focus of the current study, suggesting incorrect guidance to the user, such as guidance to a task other than one user is performing may result in higher levels of intrusiveness and frustration in part of the users.

1.2) *Trust*. Trust in a system is an important factor when considering how to design an interface and a number of taxonomies for trust in human machine interactions have been developed (Barber, 1983; Muir, 1987; Rempel et al., 1985). Trust decreases with declining reliability and evidence suggests that trust declines quite rapidly below a certain level of reliability; the absolute level of this drop-off seems to depend on the system and the context with estimates ranging from 60% to 90% (Lee & See, 2004).

Tsandilas and Schraefel (2005) connected the emotions between frustration and trust. Users felt frustrated when suggestions were incorrect. Authors associate this frustration to a reduction of user's confidence in the system, leading to a reduction in trust (Sears & Shneiderman, 1994; Tiernan et al., 2001). With regard to adaptive assistance, according to Hook (2000), user's trust in the system will decrease drastically following the initial incorrect advice provided by the adaptive system. Consequently, users may abandon the system for long periods of time. On the other hand, Hook also suggests that emotional reactions of these kinds might be a product of culture. Therefore, it may be assumed that users will not use the provided assistance if they lack trust in the system. Moreover, distrust of the help system may also decrease a user's trust in the entire system. Once we get used to having adaptive systems around us, we will also gradually build models of how they work and when they can be trusted (Hook, 2000). These two components of system's reliability (accuracy and trust) also affect each other. Low accuracy levels affect user's trust in the system's suggestions (Tsandilas & Schraefel, 2005).

2) *Intrusiveness*. The adaptive assistance we propose in this study provides user help without user request (i.e., push mode). One of the main concerns with advocating the help function is the burden it places upon the user. The level of intrusiveness depends, of course, on many factors which relate to the format of the assistance, the activity the user is currently involved in, contextual factors, user characteristics and so forth. Much can be learned about intrusiveness from research on notifications. Ho & Intille (2005), for example, summarized 11 factors that impact the perceived burden of an interruption.

Our study examines adaptive assistance provided through menus and toolbars. The motivation behind this decision was that menus are one of the primary controls for issuing commands in graphical user interfaces (Cockburn et al., 2007) and, as of recently, interfaces are also moving towards using more iconic (toolbars) formats (e.g., Ribbon). The study also examines whether there is a difference in providing assistance using these two formats (menus versus toolbars).

3. Adaptive menus and toolbars

While adaptive assistance offers information or advice regarding how to use the application, adaptive menus and toolbars are different ways of helping users to use a system more effectively (Jameson, 2003). The form of adaptation in user interfaces, such as menus and toolbars, are either *spatial* or *graphical* and, in some cases, both (Findlater et al., 2009).

When *Spatial adaptation* is applied (such as split menus), the structure or content of the interface is not constant, resulting in the user frequently needing to adapt to the different layouts (Sears & Shneiderman, 1994). *Graphical adaptation* reduces a visual search time by using graphical techniques such as background colors (Gajos et al., 2006; Tsandilas & schraefel, 2005; Findlater & McGrenere, 2008b) and transparency without changing the structure of the interface (Findlater et al., 2009).

We use graphical adaptations and, in particular, highlight the steps (functions) required for performing tasks as our means of assistance. Highlighting was selected because it seemed to suit the minimal intrusiveness requirement for assistance that is adaptive to user's situation and intentions, and hence should be integrated with probabilistic algorithms that could predict intention or infer the user's situation and cannot be 100% accurate. We postulate that highlighting is a relatively non-intrusive method of assistance because it does not disrupt the user's ongoing task. Moreover, we wanted to maintain the spatial stability of the interface. Spatial stability relates to the frequency with which the adaptive algorithm causes features to change in the interface (e.g., moves or hides a feature) (Findlater & McGrenere, 2008b)). According to Findlater & McGrenere (2008a), a more stable interface should be more predictable to the user because he or she can more easily understand the logic behind the adaptive system. An adaptation strategy that does not alter the familiar parts of the interface, and therefore is more stable, will be at least as good as a non adaptive interface (Gajos et al., 2006). Highlighting the steps required for performing a task maintains this stability. Previous research (Rech et al., 2007) has shown that users respond positively to highlighting as a means of assistance (although Rech et al., 2007, referred to highlighting of unused parts). Finally, highlighting is also known (Biocca et al., 2007) to be a very useful tool for drawing users' attention to the interface, thus it could be effective for providing assistance to users while they are using some interface and are busy trying to complete their task.

In addition, examining probabilistic algorithms that cannot provide 100% accuracy, such as intention prediction algorithms, may be useful for providing assistance of this form. We examined user performances and responses to different levels of accuracy at which the system predicts the task the user wants to perform. Few former studies have examined highlighting techniques and different accuracy levels using menus and toolbars in the context of adaptive menus and toolbars (Findlater et al., 2009; Gajos et al., 2006; Tsandilas & Schraefel, 2005, 2007). A summary of these studies is presented in Table 1.

While the previous studies emphasize advantages and disadvantages of various types of adaptive menus, our study focuses on adaptive assistance using menus and toolbars. This guides the user through the steps he or she is required to perform using the menu or toolbar, rather than just presenting the menu/toolbar items in different forms. Various adaptation techniques can be utilized to enable adaptive assistance (as presented in Table 1). Since we chose to highlight the users predicted next step as the help format, it is most natural to integrate intention prediction algorithms to a system that employs this type of help. On this study, however, we do not limit the findings to any adaptation techniques nor to a specific intention prediction algorithm.

As illustrated above, the particular presentation form we have selected for presenting the guidance is by highlighting the relevant items. This method is similar to the one used by Tsandilas and Schraefel (2005) and Park et al. (2007), however, they have highlighted items in a single menu list to facilitate the selection of items. We, on the other hand, have highlighted the entire process related to the task. Thus, one item in each menu list hierarchy (depth of the menu/toolbar) is highlighted until the user reaches his or her goal. Consequently, the highlights in our study represent a more complex task. Additionally, Tsandilas and Schraefel (2005) have compared highlighting of items in a traditional menu to highlighting in a fisheye menu. Since highlighting was included in both, we cannot infer the value of the highlighting technique. Another difference between prior research and ours concerns the measures examined. Previous research has only examined usability and performance aspects (such as selection time and error rate) and did not examine trust aspects and user perceptions of the help system in terms of how the help assisted them, level of intrusiveness, and so on. Some studies have examined different accuracy levels (Findlater et al., 2009; Gajos et al., 2006; Tsandilas & Schraefel, 2005, 2007), however, their measures also relate mainly to performance and do not explicitly examine how the accuracy level affects user perceptions of the system.

<<Insert Table 1>>

4. Research objectives

1) Examine performance and user response to a form of adaptive assistance that highlights the required steps for performing tasks using menus and toolbars. More specifically, we examine how using highlighted help affects the time to perform a task along with perceptions of the help system in terms of its usefulness, intrusiveness, and how it affects perceived performance and general user satisfaction.

2) Examine performance and user responses to different levels of accuracy of the provided help. More specifically, we examine how receiving help at different levels of accuracy (100%, 80%, and 60% and no help) affects the time to perform a task, user perceptions of the help, and the perceived trust in the help system.

3) Examine the effect of the visual format of the help on performance and satisfaction. We manipulated two help forms, namely, menus and tool bars, and two different sizes of menus (seven and thirteen), assuming that the more items included on a menu, the more difficult the task. Since users have more options to select from, this poses greater uncertainty. We specifically selected seven and thirteen in order to ensure disparity between the conditions. Seven was chosen as number of items on a menu because it is known as the "magical number" (Miller, 1956) (seven + - two) that users easily remember, while thirteen is significantly larger. Consequently, there is a good chance that the users will have difficulties to remember items and would be of greater need for help.

The results of the experiments have implications on the feasibility of utilizing various algorithms to adaptive assistance based on the level of accuracy that they can achieve.

5. Method

We decided to perform a controlled experiment that simulated various levels of accuracy of the help provided to users when they performed tasks using an unfamiliar application. We manipulated the accuracy level as well as the help format. We observed and analyzed users' performance and response to the manipulated conditions. Hence, we were able to draw conclusions about the effect of accuracy and help formats on users' satisfaction and performance. One limitation of our controlled setting is that the performed tasks were dictated to participants who did not need or were not naturally interested in performing. The users therefore had a different motivation to complete the tasks than the natural motivation in a real world situation. While in the controlled settings the users tried to complete the tasks in order to receive the extrinsic motivation that we provided (course credits), in a real world setting, users would try to complete the tasks simply because they needed it. It might well be that this fact had some influence on their attitude towards the help and the system, but we believe that it is not significant, as they had motivation to complete the tasks, and indeed all users completed all tasks.

In addition, since the users were not familiar with the application, we believe that the situation simulates to a meaningful extent the scenario of users performing tasks using unfamiliar applications and need assistance, even if they are not interested in the task. Thus, this setting enables a controlled experiment with measurable results that can point to users' behavior in similar scenarios.

An experiment in a real world setting is not feasible since reliable conclusions can be drawn only when equal conditions apply to all participants. It is, however, not realistic to expect that all participants would perform the same tasks on the same conditions in a reasonable trial period.

5.1 Experimental system

We have developed a PC-based experimental system that simulates an e-mail application in which items were highlighted as a means of assisting users in performing their tasks. Although the system simulates an e-mail application, the functional organization did not parallel any known e-mail application (such as Microsoft's Outlook or Mozilla's Thunderbird) in order to avoid previous familiarity with the application. The main functionalities of the application consist of file operations, edit, view, tools, account and help. The functions appeared in two formats: menu bar and toolbar. Both tasks in the menu bar and toolbar formats included three hierarchical levels where the target item always appeared on the third level. Figures 1(a) and 1(b) present the main screen of the application with the menu bar and toolbar formats respectively. The figure shows the menu and the toolbar option for printing a message. In the menu bar format, the item which needed to be selected at each menu level was highlighted, as presented in Figure 2. In the toolbar format, the relevant item was highlighted in each of the three levels while each level was presented separately (see Figure 3). Therefore, after selecting an item, the next toolbar level appeared and the previous one disappeared. Additionally, for both formats, the second level always included either seven or thirteen items.

<<insert figure 1(a) and figure 1(b)>>

<<Insert Figure 2>>

<<Insert Figure 3>>

The system presented the tasks which participants were requested to perform (such as "Add a new digital signature"). The system guided the participants in performing the tasks by highlighting items they needed to select as their next step (in both menu and toolbar formats) and which corresponded to the steps required for performing the task. The participants proceeded to the next task when they either successfully completed the current task or after a time limit of five minutes.

To simulate an inaccurate prediction of the user's intention, the system either highlighted relevant or irrelevant items according to the desired level of experimented accuracy. The first level was always highlighted accurately in both menu and icon presentation formats. The accuracy level in which the help was offered was controlled by the experimenter and included three levels: 100%, 80% and 60% (e.g., 80% accuracy means that in 80% of the tasks the highlights were correct and relevant to the user's task). The system also included an option to present items without help, thus without highlighting the items.

The experimental system also included two automatic questionnaires that were presented to participants (the questionnaires are attached to the paper as Appendix A and Appendix B):

1) A preliminary questionnaire was presented at the beginning of the experiment which inquired about general demographic information as well as the user's help seeking behavior, i.e. a question about the type of help that they usually use (on screen help, asking other people, search the web or any other), and the frequency of seeking help when using computer applications. In addition we asked the participants about their skill level with computer systems and frequency of using email applications.

2) A questionnaire appeared at the end of the experiment and was comprised of two sections. In the first section, participants rated on a seven point scale their perceptions of the help system by answering five questions inquiring (1) how the help assisted them, (2) whether or not the help was necessary, (3) the intrusiveness of the help, (4) how the help affected their performance, and (5) their general satisfaction with the help. The five questions were presented for each of the presentation formats (menu and toolbar). The second section included (1) questions regarding the help format, (2) questions concerning receiving such help without 100% accuracy *and* trust in the system (the questions are detailed in section 6.2).

The system recorded all user interactions with the system.

5.2 Participants

64 students from the Engineering faculty at Ben-Gurion University of the Negev in Israel participated in the study. Their average age was 26.2. Since the application was not similar to common used email applications, the participants were not familiar with the application. The participants were, however, skillful computer users (with an average rating of six on a seven point computer skill scale) and most of them use email applications frequently throughout the day. Participants received course credit for their participation.

5.3 Procedure

The participants were requested to perform 15 tasks using the system with each of the formats (menu and toolbar), 30 tasks in total, ordered randomly for each participant. Examples for tasks were inserting a digital signature, adding a contact to the address book, going to the deleted items folder and additional tasks usually performed with an e-mail application. The participants were told that they would be asked to use an email application and to perform some dictated task with it. They were told that the email application is similar in functionality to known email applications, but different in terms of interface and organization. They were informed that help would be provided by highlighting the next required step, one that is not necessarily correct, allowing them to choose whether to use it or not. The users that did not receive any help were told the tasks to perform without mentioning any available help. Randomness was applied only to the order of the menu and tool-bar tasks (a participant could receive one menu based task followed by three toolbar based tasks. However, during a task, the format was stable. The tasks were performed only in the menu/toolbar level and the task was completed when the last function in the menu/toolbar was selected. The participants were informed that they can spend up to five minutes on each task, and that the 30 tasks would take them approximately 40 minutes. They were not instructed to perform fast but it took them on average between 11.0-20.9 seconds to complete one task. The participants were divided into eight groups with four accuracy conditions (60%, 80%, 100%, no highlighting). For each accuracy condition participants saw either seven or thirteen items in a list. The accuracy manipulation occurred only on the second and third interface (menu or tool bar). For example, 80% accuracy means that on 24 tasks (out of the 30 tasks each participant performed) the help was correct while 6 tasks were incorrect in the second and third menu or the tool bar on the hierarchy. All highlights on the first interfaces were correct. Randomness occurred between tasks and not during a task, i.e., the interface format was not changed during a task.

The experiment was conducted in a lab with an experimenter present at all times. Following the performance of all tasks the participants were required to complete a questionnaire that interrogated the participants about their satisfaction from the provided help as well as their perception of its effectiveness. The detailed questions and users' responses are presented in the results chapter in section 6.2; subjective questionnaire.

5.4 Experimental design

A 2 * 2 * 4 mixed between-within experimental design was employed. The independent variables included the presentation format (menu, toolbar), number of items in menu/toolbar (seven, thirteen) and accuracy level of help (100%, 80%, 60%, no help). The last two (number of items and accuracy level) were the between-subject variables and the presentation format was the within subject variable (each participant performed tasks using both formats). The dependent variables were (1) objective measures (time to perform tasks and whether the help was used), (2) answers to the questionnaire.

6. Results

We first present the objective measures examining the time it took to perform each task and whether the help was used, followed by the subjective questionnaires. It should be noted that the participants who received no help were not asked to answer the following questions and their data were mainly used as a baseline comparison for the effect of the help in terms of performance times.

6.1 Objective measures

A Multivariate Analysis of Variance (MANOVA) was used to examine the effect of the independent variables on the two objective criterion variables: times and clicks with the percent of help accuracy (100%, 80% and 60% and no help) and number of items in a menu (7, 13) as the between-subject variables and the format (menu, toolbar) as the with-in subject variable.

Results of the MANOVA indicated significant differences between the various accuracy levels with Pillai‘s Trace = .473, F(6,114) =5.8, $p < 0.05$, Partial Eta Squared=0.237 and Observed Power=0.997. Moreover the format was also found to be significant with Pillai's Trace=.885, F(2,56)=214.854, $p<0.05$, Partial Eta Squared=0.885 and Observed Power=1.000. The following subsections present the univariate ANOVA results.

6.1.1 Performance times

We found significant differences among the different levels of help accuracy; F(3,56) =11.6, $p<0.01$. Performance times were longer when the percentage of help accuracy decreased and were shortest for the 60% (mean performance times 100% - of 11.9 and standard deviation of 2.3, 80% - 15.2 and standard deviation of 3.1, 60% - 21.9 seconds and standard deviation of 3.5).

In addition, performance time was significantly longer for the condition in which no help was presented at all (mean performance time of 30 seconds). Tukey post hoc analyses revealed significant differences between the 60% and 100% ($p<0.01$), the 60% and no help condition ($p<0.05$), and the 80% and no help condition ($p<0.01$). No significant difference was found between the 60% and 80% and between the 100% and 80% conditions. Hence, in terms of performance times, it did not matter whether the system presented help at 100% accuracy or at 80% accuracy. Conversely, presenting help at 60% accuracy increased performance times significantly compared to 100%, but not compared to 80%. Presenting help at all accuracy levels was still better when compared to not presenting any help.

The format variable (Menu, Toolbar) was found to be significant with $F(1,56)=374.8$ and $p<0.05$ although the performance times of the toolbar format was only slightly longer than the menu format (19.3 and standard deviation of 1.9 seconds vs. 19 seconds and standard deviation of 1.5).Performance time also increased with the number of items in the menu/toolbar $F(1,56) = 2.943$, $p=0.09$. It was higher when 13 items were presented in the menu/toolbar, compared to 7. For 13 items in the list the mean performance times was 21.7 seconds and standard deviation of 3.8, per task for the four groups (i.e., 100%, 80%, 60%, accuracy, no help), as opposed to 17.8 seconds average performance time and standard deviation of 3.1 for 7 items on the list.

6.1.2 Using the help

We also wanted to examine whether or not the accuracy level in which the help was presented affected the users' decision to use the help. Since we did not explicitly ask participants if they used the relevant help following each task, we inferred it from the number of clicks performed for each task. We assumed that participants who used the relevant help clicked on no more than three items (according to the highlighted guidance), whereas for those who did not use the help, more clicks were performed. In particular, we found significant differences in the number of clicks among the different levels of help accuracy; $F(3,56) = 12.06$, $p<0.01$.

We counted the number of clicks per user per task when using the system version in which no help was provided. In 66% of the tasks these users produced more than 3 clicks, compared to the users that were provided with 100% accurate help, who produced more than 3 clicks only on 5.8% of the tasks. This result indicates that this type of help is indeed helpful and that users tend to use it. It also indicates that the tasks were not trivial to users (without help). We explain that by the fact that the structure of the menus and toolbars of our email application are not similar to other email applications and participants were not familiar with it.

During the sessions with inaccurate help, we found that for the 80% accuracy level, users clicked only three clicks during 74% of the events (given that only 80% of the help instances were accurate means that on most of the correct events users used the help, but we cannot estimate whether they used help during the inaccurate help suggestions). For the 60% accuracy level, the users clicked only 3 clicks during 60% of the events. Since we could not record if the participants actually used the help, this is just an estimate assuming that once they used the correct help it lead them to exactly 3 clicks.

6.2 Subjective questionnaires

6.2.1 Questions asked for each of the presentation formats

A MANOVA was performed with the percent of help accuracy (100%, 80% and 60%) and number of items in a menu (7, 13) as the between-subject variables and the format (menu, toolbar) as the within-subject variable. The results of the MANOVA indicates significant differences between the various percent of help accuracy with Pillai‘s Trace = .51, F(10,80)=2.183, $p < 0.05$. The effect size is estimated using the "Partial Eta Squared". The Eta Squared of "percent of help accuracy" is 0.214. The value of Eta Squared for the intercept is 0.977, which means that there is a lot of variance in subjects' responds. In addition, no significant differences were found between the presentation formats and between the number of items in the menu/toolbar for any of the questions. The following results relate to the level of help accuracy. Table 2 summarizes the mean ratings and standard deviations for all the help accuracy levels for all the questions. The table also lists the results that were found to be significant in the post-hoc analysis including the squared Eta and Observed power (for alpha=5%).

<<Insert Table 2>>

Q1 -The first question asked participants if the help assisted them (one presented not at all, and seven presented very much).

Overall participants felt the help assisted them with mean ratings ranging from 4.9 to 5.8 (on a 7-point scale) and standard deviation of 0.3 .

Tukey post hoc tests showed no significant difference using all help accuracy levels.

Q2 -The second question inquired whether the help was necessary or not (one represented not at all and seven very much).

Overall participants found the help necessary using all help accuracy levels (mean ratings ranging between 4.8 and 5.5). No significant differences were found between the help accuracy percentages or between any of the other variables examined.

Q3 - The third question inquired whether the help was found to be intrusive (one represented not at all and seven very much).

The intrusiveness level varied according to the help accuracy, F(1,42)=3.1, p=0.05, decreased with the percent of help accuracy, and was highest for the 60% accuracy level (see Table 2). Nevertheless, the Tukey post hoc tests revealed a significant difference only between the 100% accuracy (mean ratings of 2.5) and 60% accuracy (mean ratings of 3.9), p=0.04. Thus, the help was perceived to be more intrusive only when presented with 60% accuracy. The intrusiveness ratings were fairly low for the 100% and 80% accuracy (mean ratings of 2.5 and 3.2 respectively).

Q4 -The fourth question asked how participants perceived the effect of help on their performance times (one presented shortened and seven lengthened).

A significant main effect was found for the help accuracy, F(1,42)=3.1, p<0.01 (see Table 2). Participants perceived help with 60% accuracy as significantly prolonging their performance time compared to the 80% accuracy, (p<0.01).

Tukey Post-hoc tests did not show any significant differences between the 80% and 100% accuracies.

This finding is not completely consistent with user's perception of the effect of accuracy on performance times. Both the objective and subjective time measures indicate that at the level of 60% accuracy, help was time consuming. However, for the 80% accuracy of help, users did not feel the time increase in their performance, even though their performance was decreased. This result is, however, consistent with the results of (Q3) that interrogated the participants about the intrusiveness of the help; at an accuracy rate of 60%, users felt that the help was intrusive. At an accuracy rate of 80%, users were likely to notice and enjoy the benefits of said help over feelings of intrusion or performance disturbance; hence this fact may not be reflected in the objective measure. The effect of difference between user perception and user performance when experience is positive is noted (Nemeth et al., 2004).

Q5 - Participants were asked to rate their general satisfaction from the help (one represents low and seven high).

No significant effects were found for this question and overall satisfaction from the help was relatively high, ranging from 4.3 to 6.

5.2.2. General questions concerning the help

Participants were asked a number of general questions at the end of the experiment regardless of the format used for the help.

In the first two questions, we performed chi-square analysis. For the other questions, we performed 2-way ANOVAs with the percent of help accuracy and number of items in a menu as the independent variables.

Q1 - Would you prefer to receive this type of help for each task you perform? (1-Yes; 2-only for difficult tasks; 3- only for tasks I don't know how to perform; 4- No, in any case)

A significant difference was found in the participant's preferences, $\chi^2(3, 48) = 39.5$, $p<0.01$. Most of the participants reported that they were interested in such help only for tasks they do not know how to perform (30 participants), eleven participants reported they want this help only for difficult tasks, five were interested in it always, and only two said they would never desire help assistance (see Figure 4). No significant differences were found in their preferences between the different accuracy levels.

<<Insert Figure 4>>

Q2 –In which presentation format was the help more useful? (Menu display, toolbar display; equally useful/ not useful).

The chi-square analysis did not reveal any preferences. We found that 16 participants preferred the highlighted help for the menu format, 18 for the toolbar format and 14 did not mind. We also performed a crosstabs analysis to examine if there were differences between the conditions of the accuracy levels however no significant difference was found.

Q3 - Would you like to receive such help in the future? (definitely Yes – 1, absolutely No – 7).

No significant differences were found for this question. Overall, participants were interested in receiving such help in the future with average ratings ranging from 4 to 5.8 (mean 4.7 and standard deviation of 0.6)

5.2.3 Receiving irrelevant help

The following questions inquired about the effects of receiving inaccurate help, namely, help for the 80% and 60% accuracy levels. We therefore only examined the responses for 80% accuracy and 60% accuracy (since 100% accuracy means that the users received only relevant help).

Q1 - Receiving help that was not relevant to my task was annoying (one represented not at all and seven very much)

Although the annoyance ratings were higher for the 60% accuracy (a mean ratings of 5.5 and standard deviation of 0.8 compared to 4.7 and standard deviation of 0.7), the difference was not significant.

Q2 - Receiving help that was not relevant to my task decreased my trust in the system (one represents not at all and seven very much)

A significant difference was found for the help accuracy, F(1,28)=7.9, p<0.05, and the level of trust in the system when receiving help with 60% accuracy decreased significantly more (mean ratings of 5.7 and standard deviation of 0.7) when compared to receiving help with 80% accuracy (mean ratings of 4.8 and standard deviation of 0.8).

A summary of the most important results from this study is presented in table 3.

<<Insert Table 3>>

7. Discussion

Presenting users with help may assist them in performing their tasks. In some situations, on-line adaptive help is not accurate or relevant to the user's tasks, such as when the user presses a hot-key for help, or any other forms where the system needs to infer the user's exact situation or intention in order to provide help. In this paper we propose providing help by highlighting the required steps for performing such tasks. The underlying assumption is that it is possible to predict a user's task by using intention prediction algorithms. One objective of the study was to examine the effectiveness of this specific help format. In addition, this is one example of a situation that might result in inaccurate help, since intention prediction usually involves probabilistic inference and it is not realistic to predict a user's intention with 100% accuracy. Accordingly, on this study we tried to examine the level of tolerance to errors that users allow when provided with inaccurate help. For this purpose we measured task performance times using highlighted adaptive assistance with various accuracy levels and asked users how they perceived the systems in terms of general satisfaction, necessity, usefulness, intrusiveness and trust. We manipulated the level of accuracy to examine the minimal level of accuracy for user's acceptance of such a system. This research is a first step towards understanding how users perceive such help.

Overall, when the help appeared at 100% accuracy, participants demonstrated high levels of satisfaction. Participants found that the help was necessary, assisted them, was not intrusive and aided in reducing the time to perform the tasks. Participant's actual performance times strengthen the participant's last postulation; their performance times were indeed significantly shorter compared to their performance without help (i.e. using a regular non adaptive menu). That said, when questioned regarding the circumstances in which they would like to receive such help, only a few participants reported they were interested in receiving this help at all times, while most were interested in it mainly for tasks they did not know how to perform themselves. Actually, the circumstances in our experiment fit the definition of tasks the users did not know and for which they would be interested in receiving help. This explains their satisfaction from the help. We believe an adaptive highlighted help system may also be of value for users experienced with the system. In this case, the help system can shorten the user's interaction by assisting them in selecting the items quicker instead of scanning long lists. We leave this examination as future work.

As opposed to Findlater et al., (2009), who found adaptive highlighted menus to be preferred on static menus, but not advantageous in terms of performances, our results found improved performance times using highlighted help. The difference in performances may be attributed to the menu structure related to the task. While we highlighted items in all three levels of the menus, Findlater et al. (2009) only highlighted items in the main menu (one level), resulting in a more simple menu structure and task for the latter. Perhaps highlighting is more advantageous for more complex tasks and menu structures. Our research did not question participants regarding their preferred menu type (highlighted or standard) since each participant used a different type of menu. Additionally, while Findlater et al., (2009) only asked participants about their preferences, our study examined how participants respond to adaptive highlighted help in menus in terms of its usefulness, necessity, intrusiveness, and trustworthiness.

Since the adaptive highlighted help is based on an intention prediction that cannot predict user goals with 100% accuracy, we wanted to examine whether the benefits diminish in lower accuracy levels. We found that the 80% accuracy did not significantly reduce the satisfaction from the help, did not make it more intrusive and participants did not feel it increased the time to perform the task when compared to the 100%. It did, however, decrease participants trust level in the system. Thus, users are able to ignore inaccurate help so as not to interrupt their activities while still trusting the system less.

On the other hand, the impact of receiving help with 60% accuracy was much more noticeable. Here, participants felt the help was not very useful, was more intrusive and did not shorten performance times as it did when using 100% accurate help. In addition, participant's trust in the system decreased more significantly in the 60% and, although not significant, they perceived the help to be more annoying.

These results correspond to Lee and See (2004) who found that trust declines with decreasing reliability, especially at around 60% to 90% (depending on the system goals and context). This result is also in line with Wickens and Dixon (2007) who found a 70% accuracy to be the threshold for system reliability. It seems that similar levels of reliability that are acceptable when using automated systems, as demonstrated by Lee and See (2004) and Wickens and Dixon (2007), also apply to adaptive systems.

We also found that receiving irrelevant help did not affect the user’s decision on whether or not to use the suggested help. Thus, the decrease in the participant’s trust in the system did not affect their actual actions and in most cases they still decided to use the help. Our intuitive explanation for this disparity would suggest that users would not ask for such help once they lose trust in the system, but as the help is provided anyway (items are highlighted without the user's request), the users are more likely to use it. Another factor we looked at was the format of presentation. Many systems and applications are moving from the conventional menu format to a more iconic/toolbar presentation (for example Microsoft's Ribbon, Apple's iPhone). One of our aim was to examine if presenting highlighted help in different formats affects performance times and subjective evaluations. We did not find any differences in presenting the functions in a menu format or toolbar. We conclude that this type of help may be applicable for both formats, at least for new users in the system.

Finally, it may be assumed that highlighted help will be more beneficial for longer menus in which it is more difficult to find the relevant function. In terms of performance times, we indeed found that it took participants significantly more time to perform the tasks when more items were included in the menus for all help groups (60%, 80%, 100%, no help).

This, however, did not affect their perceptions of the help in all accuracy levels and the help was equally beneficial for both shorter and longer menus. This may result from the fact that each participant in our study experienced a different system condition. If each participant would have experienced help in both short and long menus, they could more easily have compared between the two.

8. Summary, conclusions and future work

Our study focused on assisting users by guiding them through the steps required to perform their tasks by using highlights of items as a form of assistance. We introduced two key factors for implementing successful adaptive assistance: One is system reliability and the other is minimal intrusiveness. Reliability stems from two dependent components, namely, maintaining reasonable accuracy levels in which the system provides the help, and as a result of maintaining user trust in the system. We believe we have addressed all these issues in our study. In terms of the accuracy level of the system assistance, the accuracy rate does not necessarily need to be very high. It appears that users are willing to accept suggestions that are not relevant to their task to some degree. Our results show that guiding users by highlighting the relevant steps is useful as long as the assistance appears at accuracy levels of higher than 60%. Users respond to the assistance positively and demonstrate interest in receiving such help in the future. With regard to maintaining trust in the system, as expected, receiving less accurate assistance lowers user's trust in the system, even for the 80% accuracy condition. Surprisingly, users are relatively tolerant to inaccuracy and are willing to accept irrelevant assistance even though it reduces their trust in the system. In the future, more accuracy levels between 60% and 80% should be examined in order to achieve a more specific threshold in general, or for a specific algorithm. This will provide a stronger basis for the decision of whether or not to use a specific intention prediction algorithm in an adaptive system.

Participants did not consider highlighted guidance to be intrusive, even when the assistance was not 100% accurate. It was interesting to observe the interaction between the factors. A lower level of accuracy, however, did negatively affect the perceived intrusiveness.

This study integrated objective and subjective measures for increasing knowledge about how users perceive help by highlighting items in systems. This study, however, is not without some limitations. First, all the participants in our study were skillful computer users. It would be interesting to examine the results with non-skilled computer users that might take more time to learn new applications and perhaps are in need of more guidance. Secondly, we only looked at users inexperienced with the system. It may be beneficial to also examine how such help can benefit users experienced with the system. Obviously it may benefit them differently.

Another limitation concerns the application we used for this experiment which was artificially built and contained tasks in a fixed length of three actions. Although this type of application does not exactly simulate real world applications, we believe it does simulate a scenario of users unfamiliar with an application that requires to perform tasks they are not familiar with. As indicated in our results, users use help only for tasks that they are not familiar with. Hence, our experiment simulated a situation where users tend to use help, therefore we believe that the findings are sufficient for deducing the results we obtained. It is important to mention again there are limitations stemming from the lab-based controlled nature of our experiments that might have affected users' behavior with the application, mainly due to the difference in motivation to complete the tasks. While in our setting the motivation is extrinsic, i.e., course credits, in a real world setting, users are motivated intrinsically by their own needs. Our setting can be justified by its feasibility to control the factors to be examined and to enable the drawing of results based on sufficient instances (from the statistical point of view), as opposed to real world setting. Thus, it seems more reasonable to apply a controlled setting here and to be careful with generalizing the results due to a possible effect of the type of motivation on the user's behavior. Now that we gained some insights about the factors for implementing an adaptive help system, a future real world pilot is planned. In such a real world setting we will also integrate our intention prediction algorithm rather than the simulated inaccuracy. Additionally, we used a specific form of assistance (i.e., guided highlights). Intention prediction algorithms may be implemented in other forms of assistance as well (as done by Kunzer et al. (2004), for example). The strengths of intention prediction methods should therefore be examined using additional assistance formats.

To conclude, the results of this study have implications on the implementation of algorithms that are not 100% accurate to adaptive assistance systems, especially intention prediction algorithms. The results demonstrate that even though the provided help suggestions are not made with 100% accuracy, such implementations can assist users in performing their tasks.

9. References

Antwarg, L. 2010. Character driven Hidden Markov Model trees for Intention prediction and adaptive assistance. Msc. Dissertation Ben-Gurion University, Israel

Barber, B. 1983. The logic and limits of trust. Rutgers University Press: New Brunwick, New Jersey.

Bennet, K.B., Cress, J.D., Hettinger, L.J. and Stautberg, D. 2001, A theoretical analysis and preliminary investigation of dynamically adaptive interfaces. The International Journal of Aviation Psychology, II(2), 169-195,

Biocca, F., Owen, C., Tang, A., and Bohil, C. 2007, Attention issues in spatial information systems: Directing mobile users' visual attention using augmented reality, Journal of Management Information Systems 23(4), pp. 163-184.

Cockburn, A., Gutwin,C., and Greenberg, S. 2007. A predictive model of menu performance. In Proc. CHI’07, 627–636.

Davison, B.D. and Hirsh, H. 1998. Predicting Sequences of User Actions. In notes of the AAAI/ICLM workshop on predicting the future: AI Approaches to time-series analysis.

Encarnacao, L.M. and Stoev, S.L. 1999. An application-independent intelligent user support system exploiting action-sequence based user modeling. Proceedings of 7th International Conference on User Modeling. Wien, New York: Springer Verlag.

Findlater, L. and McGrenere, J. 2008a. Comprehensive User Evaluation of Adaptive Graphical User Interfaces. CHI 2008, April 5 – April 10, 2008, Florence, Italy. Workshops and Courses: Usable Artificial Intelligence.

Findlater, L. and McGrenere, J. 2008b. Impact of screen size on performance, awareness, and user satisfaction with adaptive graphical user interfaces. Proc. CHI '08, (2008).

Findlater, L., K. Moffatt, J. McGrenere and Dawson, J. 2009. 'Ephemeral adaptation: the use of gradual onset to improve menu selection performance. In Proceedings of the 27th international Conference on Human Factors in Computing Systems (Boston, MA, USA, April 04 - 09, 2009). CHI '09. ACM, New York, NY, 1655-1664.

Fischer, G. 2001. User modeling in Human-computer interaction. User Modeling and User Adapted interaction, 11(1-2), 65-86

Gajos, K. Z., M. Czerwinski, D.S. Tan and Weld, D.S. 2006. Exploring the design space for adaptive graphical user interfaces. In Proceedings of the Working Conference on Advanced Visual interfaces (Venezia, Italy, May 23 - 26, 2006). AVI '06. ACM, New York, NY, 201-208.

Gajos, K.Z., K. Everitt, D.S. Tan, M. Czerwinski and Weld, D.S. 2008. Predictability and accuracy in adaptive user interfaces. Proc. CHI 2008.

Ho, J. and Intille, S.S. 2005. Using context-aware computing to reduce the perceived burden of interruptions from mobile devices. Human Factors in Computing Systems: Proceedings of CHI'05, New York: ACM Press, 909-918.

Hook, K. 2000. Steps to take before intelligent user interfaces become real. Interacting With Computers, 12(4), 409–426.

Horvitz, E., J. Breese, D. Heckerman, D. Hovel, and Rommelse, K 1998. The Lumière project: Bayesian user modeling for inferring the goals and needs of software users. Proceedings of Fourteenth Conference on Uncertainty in Artificial Intelligence, Madison, WI, pp. 256-265

Inbar,O Lavie, T., Meyer, J. 2009. Acceptable intrusiveness of online help in mobile devices. In Proceedings of the 11th International Conference on Human-Computer Interaction with Mobile Devices and Services (MobileHCI '09). ACM, New York, NY, USA, Article 26, 4 pages.

Jameson, A. 2003. Adaptive Interfaces and Agents, in Jacko, J.A., Sears, A. (Eds.), Human-Computer Interface Handbook. Mahwah, NJ, Erlbaum, pp. 305-330.

Kunzer, Aa., F. Ohmann and Luczak, H. 2004. Anticipated user modeling based on action prediction algorithms. Working with computer systems. In H.M. Khalid, M.G., Halender, A.W. Yeo (Eds.). Kuala Lumpur: Damai Sciences.

Lee, J., and. See, K. 2004. Trust in automation: designing for appropriate reliance. Human Factors, 46, 50-80.

Miller G.A. 1956. The Magical Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information. Psychological Review, 63(2), 81-97.

Muir, B.M. 1987. Trust between humans and machines, and the design of decision aids. International Journal of Man-Machine studies, 27(5-6), 426-432.

Nemeth Y., Shapira B., and Maimon M. 2004. Evaluation of the real and perceived value of automatic and interactive query expansion, Proceedings of Sheffield SIGIR 2004. The Twenty-Seventh Annual International ACM SIGIR 2004. 526 - 527

Novick D.G, Ward, K. 2006. Why Don’t People Read the Manual? Proceedings of the 24th annual ACM international conference on Design of communication, 11 – 18

Novick, D., E. Elizadle and Bean, N. 2007. Toward a More Accurate View of When and How People Seek Help with Computer Applications. SIGDOC '07: Proceedings of the 25th annual ACM international conference on Design of communication, 95 - 102

Park, J., Han, S. H., Park, Y. S., and Cho, Y. 2007. Adaptable versus adaptive menus on the desktop: Performance and user satisfaction. International Journal of Industrial Ergonomics, 37(8), 675-684.

Pierrakos, D., G. Paliouras, C. Papatheodorou and Spyropoulos,C. 2003. Web usage mining as a tool for personalization: A survey. User Modeling and User-Adapted Interaction, 13(4), pp. 311-372(62)

Rech, J., E. Ras and Decker, B 2007. Intelligent assistance in German software development: A survey. IEEE Software, 24 (4), pp.72-79.[4] Fischer, G. (2001). User modeling in Human-Computer Interaction. User Modeling and User-Adapted Interaction, 11 (1-2), 65-86.

Rempel, J.K., J.G. Holmes and Zanna, V. 1985. Trust in close relationships. Journal of personality and social psychology, 49, 95-112.

Sears, A., and Shneiderman, B. 1994. Split menus: effectively using selection frequency to organize menus. ACM Transactions on Computer–Human Interaction 1 (1), 27–51.

Shen, J., L. Li, T.G. Dietterich and Herlocker, V. 2006. A hybrid learning system for recognizing user tasks from desktop activities and email messages. In Proceedings of the 11th international Conference on intelligent User interfaces (Sydney, Australia, January 29 - February 01, 2006). IUI '06. ACM, New York, NY, 86-92.

Sun, R. and Giles, V. 2001. Sequence learning: From recognition and prediction to sequential decision making. IEEE Intelligent Systems, 16(4), 67–70.

Tiernan, S.L., E. Cutrell, M. Czerwinski and Hoffman, H. 2001. Effective Notification Systems Depend on User Trust. INTERACT'02. Tokyo. pp. 684-685.

Tsandilas, T. and Schraefel, M.C. 2005. An empirical assessment of adaptation techniques. In CHI '05 Extended Abstracts. ACM Press. 2009-2012.

Tsandilas, T. and Schraefel, M.C. 2007. Bubbling Menus: A Selective Mechanism for Accessing Hierarchical Drop-Down Menus. ACM CHI'07, p. 1195-1204.

Wickens, C. D., and Dixon, S.R. 2007. Is there a magic number 7 (to the minus 1)? The benefits of imperfect diagnostic automation: A synthesis of the literature'. Theoretical Issues in Ergonomics Science, 8(3), 201–212.

Appendix A – preliminary questionnaire

Please state your skill level with computer systems:

Very low 1 2 3 4 5 6 7 Very high

At what frequency do you use email applications

1-a couple of times a day

2-once a day

3-two-three times a week

4-once a week

5-less than once a week

How do you usually seek help when you encounter a problem in a computer application

1-on screen help from the application

2-asking other people

3-search the web

4-other

How often do you tend to request help when using a computer application?

Never 1 2 3 4 5 6 7 Very often

Appendix B – Post experiment questionnaire

You have been presented with tasks and help in two different presentation formats: a menu presentation and an icon presentation. Please answer the following questions for each type of display and the general questions that refer to both types of displays.

Menu presentation:

Did the help assist you?

Not at all 1 2 3 4 5 6 7 very much

Was the help necessary?

Not at all 1 2 3 4 5 6 7 very much

Was the help intrusive (disturbing)?

Not at all 1 2 3 4 5 6 7 very much

How do you think using the help affected your performance times?

Shortened 1 2 3 4 5 6 7 lengthened

Rate your general satisfaction from the help:

Low 1 2 3 4 5 6 7 High

Icon presentation:

Did the help assist you?

Not at all 1 2 3 4 5 6 7 very much

Was the help necessary?

Not at all 1 2 3 4 5 6 7 very much

Was the help intrusive (disturbing)?

Not at all 1 2 3 4 5 6 7 very much

How do you think using the help affected your performance times?

Shortened 1 2 3 4 5 6 7 lengthened

Rate your general satisfaction from the help:

Low 1 2 3 4 5 6 7 High

General questions

Would you like to receive such help in the future?

Yes 1 2 3 4 5 6 7 No

Would you prefer to receive this type of help for each task you perform?

1-Yes 2-only for difficult tasks 3- only for tasks I don't know how to perform 4- No, in no case

Receiving help that was not relevant to my task was annoying

Not at all 1 2 3 4 5 6 7 very much

Receiving help that was not relevant to my task decreased my trust in the system

Not at all 1 2 3 4 5 6 7 very much

To what presentation format was the help more useful?

Menu display Icon display equally useful (not useful)

Figure captions

Figure 1a- menu-base interface

Figure 1b - toolbar based interface

Figure 2 - Illustration of guidance for performing a task using the Menu bar. The items needed to be selected are highlighted at each menu level

Figure 3 – Illustrations of guidance for performing a task using the toolbar. Each toolbar level is presented separately with the required item highlighted. Pressing the item leads to the next level. The task is completed at level 3

Figure 4 – Distributions of participants' preferences for the help context

Table 1

Summary of experiments on graphical adaptive interfaces

Experimented interfaces	*Variables*	*Significant results*	*Source*
Different types of menus: 1) Normal highlight - changes background color of items 2) Shrink highlight- highlight items and shrinks the font size of non-suggested items	1)Highlighted Normal vs. highlighted Shrink menus 2) Accuracy levels - 60%, 80%, 100%. 3) Number of items	- longer selection time in lower accuracy levels - Shrink was slower than normal in selecting items not highlighted - Normal highlight technique generated fewer errors when accuracy was imperfect	[33]
Different types of toolbars: 1) Static without adaptation 2) Split - extra toolbar with copied important functions 3) Moving split -moves promoted functionality to main toolbar 4) Highlight - colors the background of items	1) Interface types 2) Different tasks 3) Adaptation model: recently used or frequency 4) Accuracy levels - 30%, 70%	- satisfaction data showed significant preferences for the split interface condition over no adaptation and highlights - In split and moving interfaces task completion time was shorter with the 70% accuracy than with 30%.	[11]
Different types of menus: 1) Static without adaptation 2) Highlight - colors the background of items 3) Ephemeral (momentary) - provide initial adaptive support, which then gradually fades away	1) Interface types 2) Accuracy levels - 50% and 79% (when comparing between ephemeral and highlighted only 79% accuracy was used).	- Ephemeral adaptive menus were faster than static menus in high accuracy levels, and were not significantly slower in lower - Ephemeral adaptive menus were faster than adaptive highlighting - Highlight was preferred to static	[9]
Different types of menus: 1) Static without adaptation 2) Bubbling - combines the bubble cursor with directional mouse gesture techniques to facilitate the access to certain items	1) Interface types 2) Accuracy levels - 28.6%, 85.7%	Bubbling menus improved mean selection time by 20% when accuracy was high and reduced mean performance by 14% when accuracy was low	[34]
Different types of menus: 1) Static without adaptation 2) Adaptable - items can be moved with drag-and-drop 3) Adaptive split - top section shows 3 most frequently selected items and the bottom section contains the others. 4) Highlight - colors the background of items	1) Interface types 2) Menu label Categories	- Adaptable menu was faster than static and adaptive highlight - Adaptive split was slower than the others. - The adaptable and adaptive highlight menus were more recognizable than the others. - The adaptable menu was perceived more efficient than the adaptive highlight or traditional. - The static was found to be the least efficient and preferred.	[25]

Table 2

Means and standard deviations of ratings in all help accuracy levels for all tasks. The squared Eta and Observed power (for alpha=5%) were calculated regarding the "percent of help accuracy" variable.

Question	100%		80%		60%		Significance for variable "percent of help accuracy"	Squared Eta	Observed Power
	Mean	SD	Mean	SD	Mean	SD			
Q1- help assisted?	5.6	0.3	5.8	0.3	4.9	0.3	-	0.102	0.466
Q2 – help necessary?	4.8	0.3	5.5	0.3	5.0	0.3	-	0.042	0.203
Q3 – help intrusive?	2.5	0.3	3.2	0.3	3.9	0.3	100-60	0.126	0.567
Q4 – perceived time	2.6	0.4	2.5	0.4	4.4	0.4	80-60, 100-60	0.209	0.837
Q5 - satisfaction	5.4	0.3	5.2	0.3	4.5	0.3	-	0.056	0.282

Table 3

Summary of results

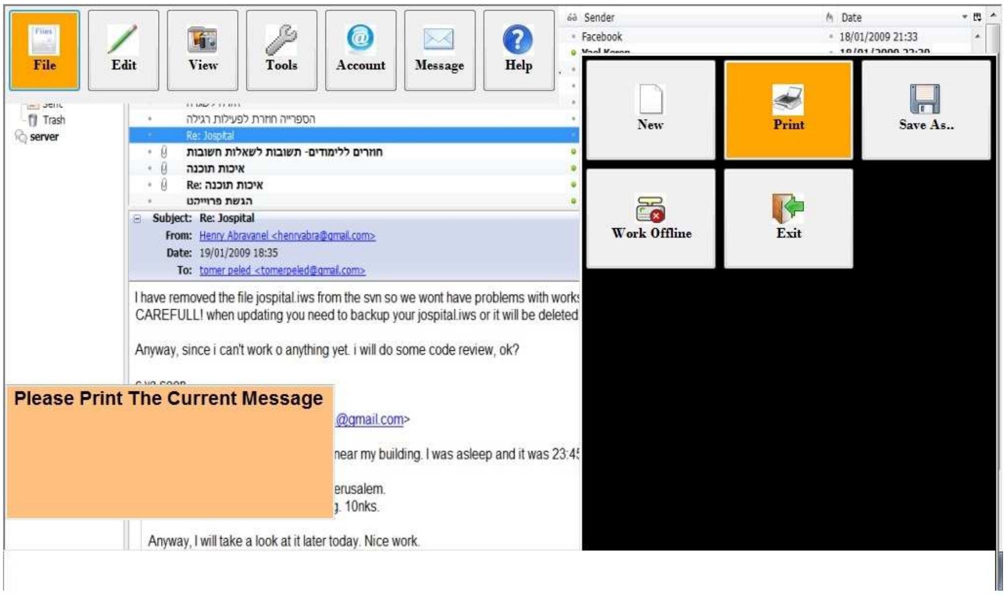
Variable	***General results***	
Formats (menu/toolbar)	No significant differences were found between the presentation formats in both objective and subjective measures	
Decision whether to use help	Receiving the help in 60% did not significantly affect the decision whether to use the help, compared to the 80% accuracy	
Interest in such help in the future	Most participants were interested in receiving such help only for tasks they do not know how to perform followed by receiving this help only for difficult tasks. Only few were interested in it always and even fewer were never interested in it	
Variable	***General results***	***Differences in accuracy levels (post hoc)***
Performance times	-Were significantly longer for the condition in which no help was presented compared to all other conditions -were higher with 13 items in a menu compared to 7 items	- were significantly longer when the percentage of help accuracy decreased and were slowest for the 60%. -Presenting help at 60% increased performance times significantly compared to the 100%, but not compared to the 80%
Degree of assistance	Participants felt the help assisted them in all help accuracy levels. However, accuracy of help affects the degree of assistance perceived. Low accuracy is perceived as less assisting.	A significant difference was found only between the 80% and the 60%
Necessity of help	Participants found the help necessary using all help accuracy levels	No significant differences
Intrusiveness of help	The intrusiveness level decreased with the percent of help accuracy and was highest for the 60% accuracy level.	A significant difference was found only between the 100% accuracy and 60% accuracy
General satisfaction	The overall satisfaction from the help was relatively high with no significant differences between the conditions	No significant differences
*Annoyance level		No significant differences
*Perceived affects on performance times		Participants perceived the help with 60% accuracy as significantly prolonging their performance time compared to the 80% accuracy
*Trust level in the system		Decreased when receiving help with 60% accuracy significantly more compared to when receiving help with 80% accuracy

* We examined only the differences between the 80% accuracy and 60% accuracy

Deleted: Accuracy of help affects

Deleted: Participants felt the help assisted them in all help accuracy levels

The main screen of the application with the menu bar
79x53mm (72 x 72 DPI)



Facebook 18/01/2009 21:33

הספרייה חוזרת לפעילות רגילה

Re: Jospital

חוזרים ללימודים- תשובות לשאלות חשובות

איכות תוכנה

Re: איכות תוכנה

הגשת פרוייקט

Subject: Re: Jospital
From: Henry Abravanel <henryabra@gmail.com>
Date: 19/01/2009 18:35
To: tomer peled <tomerpeled@gmail.com>

I have removed the file jospital.iws from the svn so we wont have problems with works
CAREFULL! when updating you need to backup your jospital.iws or it will be deleted

Anyway, since i can't work o anything yet. i will do some code review, ok?

Please Print The Current Message

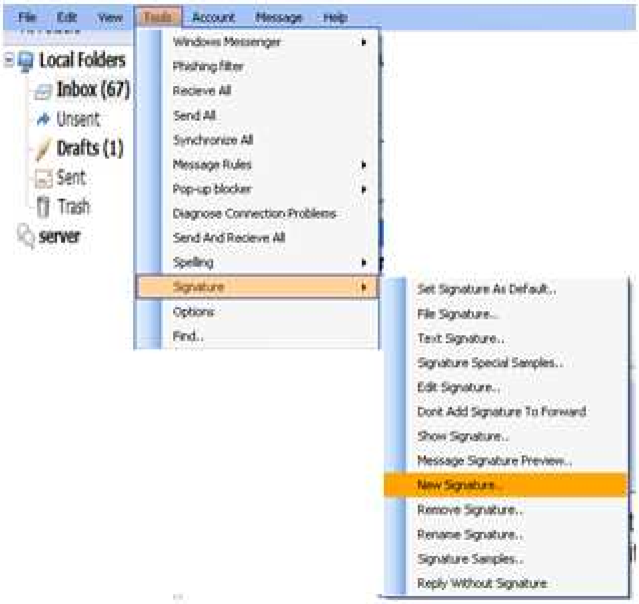
@gmail.com>

near my building. I was asleep and it was 23:4

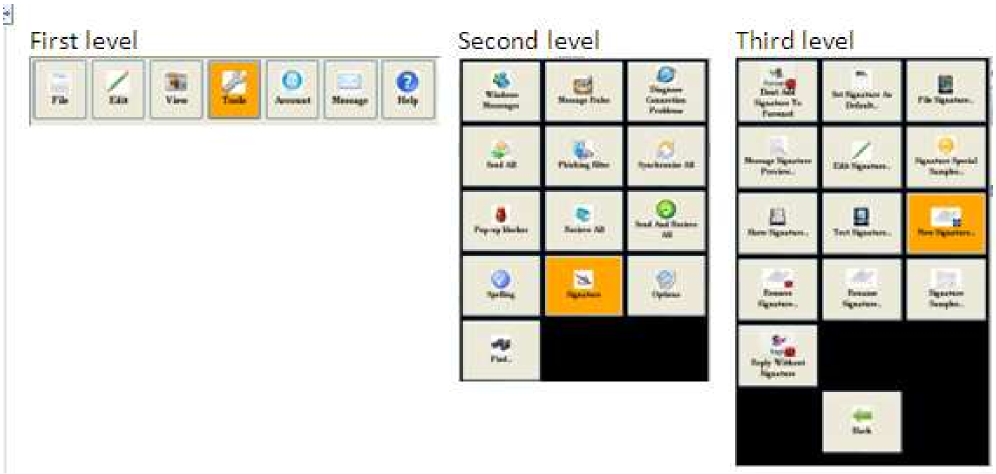
erusalem.
. 10nks.

Anyway, I will take a look at it later today. Nice work.

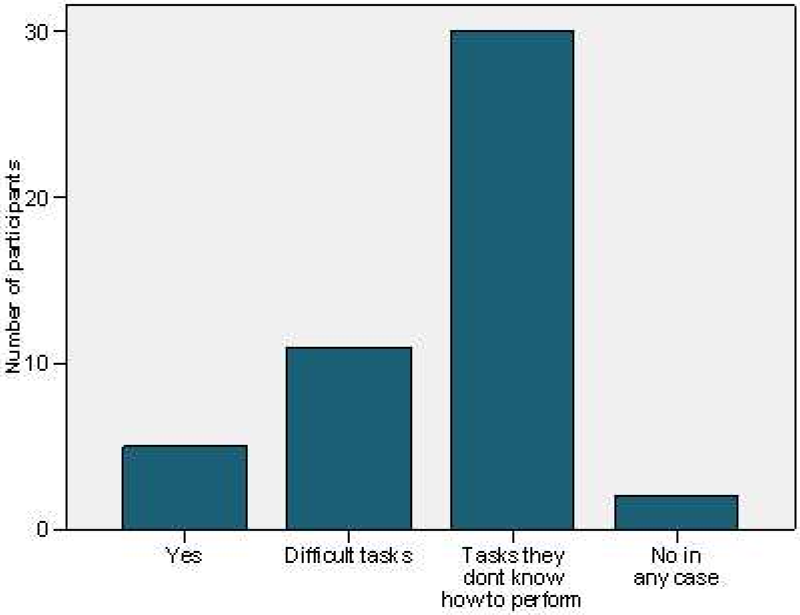
The main screen of the application with the toolbar
270x183mm (96 x 96 DPI)



Highlighting Illustration
89x81mm (96 x 96 DPI)



Highlighting illustration for toolbar format
151x71mm (96 x 96 DPI)



Response Distribution
115x85mm (96 x 96 DPI)