



Towards the Right Assistance at the Right Time for Using Complex Interfaces

Blandine Ginon, Simone Stumpf, Stéphanie Jean-Daubias

► To cite this version:

Blandine Ginon, Simone Stumpf, Stéphanie Jean-Daubias. Towards the Right Assistance at the Right Time for Using Complex Interfaces. International working conference on Advanced Visual Interfaces, Jun 2016, Bari, Italy. pp.240 - 243, 10.1145/2909132.2909275 . hal-01402978

HAL Id: hal-01402978

<https://hal.science/hal-01402978>

Submitted on 31 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Towards the Right Assistance at the Right Time for Using Complex Interfaces

Blandine Ginon, Simone Stumpf, Stéphanie Jean-Daubias

► To cite this version:

Blandine Ginon, Simone Stumpf, Stéphanie Jean-Daubias. Towards the Right Assistance at the Right Time for Using Complex Interfaces. International working conference on Advanced Visual Interfaces, Jun 2016, Bari, Italy. pp.240 - 243, 10.1145/2909132.2909275 . hal-01402978

HAL Id: hal-01402978

<https://hal.archives-ouvertes.fr/hal-01402978>

Submitted on 31 Oct 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Towards providing the right assistance at the right time

Blandine Ginon

Simone Stumpf

Stéphanie Jean-Daubias

ABSTRACT

Intelligent assistants which help users in completing their tasks are becoming commonplace yet there are still many challenges to overcome in order to integrate them well into user interfaces. One obstacle that has been recognized is determining the right time to provide the right assistance: too soon might mean the user will be interrupted before intervention is appropriate but too late will leave the user struggling to advance in their task. In order to investigate the timing of assistance, we conducted an empirical study that aimed to uncover what factors influence users' acceptance of proactive assistance. For this purpose, we used an assistance system which monitors a user's tasks and provides proactive assistance when the user deviates from them. Our results describe factors that appear to play a role in personalizing assistance, with a view to developing user models for providing the right assistance at the right time. Our work is a step towards providing effective task-based intelligent assistants.

Authors Keywords

User assistance; predictive model; proactive assistance; activity monitoring; trace-based systems.

ACM Classification Keywords

Human computer interaction (HCI); empirical studies in HCI; user interface management systems.

General Terms

Human Factors; Experimentation; Measurement.

INTRODUCTION

While user interfaces should be as intuitive and easy to learn as possible for a wide audience, there are still circumstances in which users need help and guidance to complete their tasks. Traditional help systems typically are reactive: they require users to recognize that they need help, enter a help mode, and search/browse for the required topic. More recently, research has been directed at providing *proactive assistance* using intelligent user interfaces [42,29,23]. An early example of providing proactive assistance to users in their desktop computer tasks were Office Assistants (Figure 1) [19,18]. Challenges of proactive assistance have been well-documented [38] and some progress has been made in predicting the task that a user is trying to perform [37].

A significant amount of effort has concentrated on designing and implementing assistance in intelligent tutoring systems [1,22] and other task-based systems which

allow guidance to be provided to learners when they are stuck [13,15]. However, identifying the right moment to provide assistance is essential. Too soon and the assistance will interrupt the learner unnecessarily, but helping too late means that the learner is left struggling. Delivering the right assistance is also important: it needs to be relevant to solving the problem and the content needs to be informative. It has been recognized that every learner is different and some progress has been made to personalize assistance by developing user models [6,1].

Our work aims to contribute to our understanding of providing assistance at the right time for each individual user, and to deliver the right help in the right form.

To investigate these issues, we conducted an empirical study using an assistance system that monitored users' interactions with a photo-editing application, and if they deviated from expected tasks, provided assistance for carrying out the next action. We collected interaction and preference data during the study which we analyzed to answer the following research questions:

1. What are the important factors that influence preferences for timing assistance?
2. Can we predict the best time to provide assistance, based on users' characteristics?
3. How should assistance be provided to users?

Our research contributes to a better understanding of the impact of proactive assistance on user satisfaction, and provides first steps toward predicting the right time to provide help to users, leading to better informed system design.

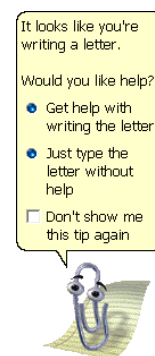


Figure 1. Clippy, an early attempt at providing proactive assistance.

The structure of our paper is as follows: first, we will present related work in providing assistance for completing tasks. We will then describe our study set-up using a prototype system that can deliver task-based assistance and present the results of our study. We will end with a wider discussion of our findings and conclude with a summary of our work.

RELATED WORK

Assistance systems

Assistance facilitates learning and using an application, in a way that is suitable to the user and to the context of use, in order to enable the user to exploit all the possibilities of an application fully [14]. Assistance systems can be classified according to who initiates the assistance [29]: an assistance system is *proactive* if it detects an assistance need and provides assistance to the user; *reactive* if the system relies on the users to ask for help; and an assistance system is *mixed* if the system can either provide assistance at the request of the user or at its own initiative. Because users differ in their needs, some systems are also *configurable* by either users or by the assistance designer, to either deactivate the assistance system or control the level of wanted assistance.

Reactive assistance systems, e.g. help manuals, tool tips, etc, require the user to identify that they have a need for assistance [21,40], and it can take a long time to find the relevant information amongst the available information [24]. Proactive assistance systems, in contrast to autonomous systems, interact with users instead of carrying out actions on their behalf [35,23] but are semi-autonomous in that they embed some intelligence in the user interface. Most research in proactively assisting users has focused on determining the task that users are doing that requires assistance, while there is little research to detect the right time to assist the user.

Task assistance

Much research has been carried out to investigate task prediction and task switching (e.g. [37,32,28]) and there are a number of proactive assistance systems that support task or activity management (e.g. [42,17,39,5]). Most of these systems rely on the system being able to monitor and trace the user's activity at a fine-grained level, e.g. [10,12,14,25,43]. Such systems can then provide assistance that takes into account the user's current activity (e.g. [3,31]).

Thus, an assistance system could detect the typical task structure, and proactively provide users with assistance to complete this task and the associated subtasks. However, these intelligent approaches typically require a lot of correct user examples on which to base the task structure, which in a learning context is very difficult to acquire. Hence, most assistance systems in an educational setting have resorted to an *a priori* structuring of the task, by scheduling subtasks that are required to complete the task. This can be done either by the user or by a teacher [11].

Interventions

Some work has been carried out to study the right time for a system to intervene i.e. the best time to interrupt a user to switch tasks (e.g. [20,16]). However, proactive assistance systems do not aim to interrupt the user's task, instead, we are trying to help the user continue with their task.

Other approaches to intervene in the field of Intelligent Tutoring Systems are based on a model of user's affective and motivational aspects (e.g. [8,4,2]). For example, the assistance system can try to detect when a student is disengaging from a pedagogical activity, using pupillary response and other sensor information. As a consequence, they are difficult to use in the everyday life. However, using users' traces, e.g. log data and analysis of low-level actions, could be a fruitful avenue to explore for determining the best time to intervene.

STUDY SETUP

To investigate our research questions, we conducted an empirical study with a prototype assistance system that provided advice on using photo-editing software to complete a task. We manipulated the time interval after assistance was given after the system had determined that the user deviated from the task and needed help. We collected participants' interaction and preference data, together with detailed background information, and analyzed the data quantitatively to investigate our research questions.

Participants and Task

We recruited 144 students and staff from a French university through an email advertisement. No incentives were paid. Each participant was randomly assigned to a group of 12; each group completed the same task but we varied the time that the system used to determine that the user had deviated from their task, i.e. each group experienced different assistance timing. Group A's assistance intervened after 0 seconds i.e. assistance was shown immediately when the user deviated from the task, whereas participants in group K experienced assistance timing of 30 seconds, meaning that they had 30 seconds during which to get back on task. We varied timing between groups in 3-second intervals.



Figure 2. Holiday card to be created as part of the task.

For their task, we asked the participants to create a holiday card (Figure 2) from a photo they were given. To complete this task, they were using PhotoScape (<http://www.photoscape.org/ps/main/index.php>), a freely available photo-editing application.

Task instructions provided to participants asked them to:

- open a given photo in edit mode;
- crop the photo;
- add a speech balloon;
- add a frame;
- save it.

For each of these steps, the participant had to carry out several actions with PhotoScape. For instance, for the step “add a speech balloon”, the participants had to open the “Object” tab, draw a shape, open the shape properties, enter a text, set the font “Verdana” with size 24 points and color blue, set the balloon shape and then save the properties.

The assistance system

The assistance system we used, SEPIA [15,14], can be grafted on to applications without any need to access or modify the application source code. SEPIA can monitor a target application and “trace” all user interactions with this application, e.g. clicking on a button or opening a menu, which can then be leveraged to provide contextualized assistance. User interface enhancements and automated actions can be injected into the application by SEPIA; for example, it is possible to attract the user’s attention to a component by displaying an arrow, or to complete an action on behalf of the user.

In the SEPIA system, the assistance that is provided to end-users is specified *a posteriori* of the target application using an editor that is then executed later on. This approach, often used in the education context [30,11,33,7,29], enables assistance designers or tutors instead of the target application developer to set up an assistance system without programming knowledge. In a first phase, the assistance designer defines a “trace” (or several traces) which comprises a set of low-level events that lead to successfully completing the task. In SEPIA, a trace can be defined by the assistance designer either manually or simply by demonstrating the task in the target application. Second, using this trace, the assistance designer can then associate an assistance action for any low-level events in the trace. In our study, assistance actions were simple help messages explaining to the user what to do next, coupled with a UI enhancement to an object on which the user should act (e.g. an arrow pointing to the button on which to click next). Figure 3 shows an example of such proactive assistance. The assistance designer also determines the maximum

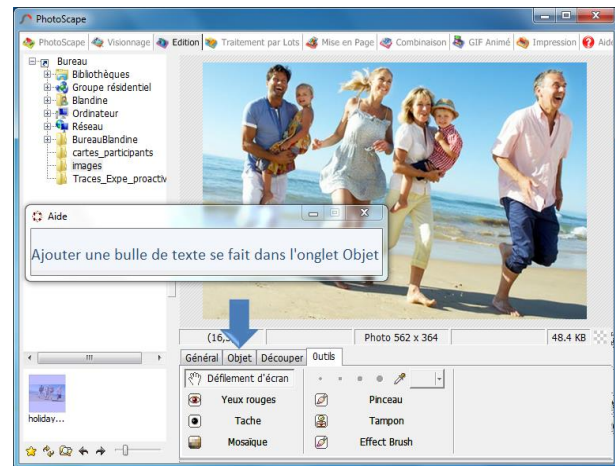


Figure 3. Example assistance actions of either adding a help message and a UI enhancement of an arrow pointing to the next user action required.

amount of time and/or number of actions that the user should spend completing a step before assistance is triggered. In our study, we varied this maximum amount of time for each participant group, meaning that the system delayed triggering the assistance once it had determined deviation from a trace.

The second phase involves the application's users during assistance execution. During the execution, a user's interaction with the application is monitored by SEPIA; the assistance system then analyses these interaction “traces” to provide contextualized assistance based on the triggers specified in the assistance specification. SEPIA looks through the low-level actions performed by the user and checks whether they conform to the trace specified in the assistance editor. After each correct action i , the assistance system is waiting for action $i+1$; if the user does not perform action $i+1$ in the time defined by the assistance designer, or if he does more “wrong” clicks than the number defined by the assistance designer for step $i+1$, then the assistance system will intervene and provide the assistance associated with step $i+1$.

Procedure and Data Collection

Each session included 12 participants at a time. Upon arrival, they were briefed and signed consent forms. They then filled in a background questionnaire, including details about their demographics, personality and help preferences. They then completed the task using PhotoScape; no tutorial was given how to use this application to succeed. Finally, we administered an exit questionnaire capturing the participants' feedback regarding the assistance provided.

Background Questionnaire

The background questionnaire captured participants' characteristics that might be useful in determining their propensity for wanting to be assisted and that could be used as variables in predicting timing of advice. Based on previous research into help-seeking and intelligent tutoring systems [1,27,26], we developed a set of questions that asked for participants' gender, age and previous experience in photo-editing. We captured participants' self-efficacy in completing a computer photo-editing task, based on [9], and their self-esteem [34]. We also developed a set of questions that probed their help-seeking behavior, based on factors identified by [1,27,26], such as locus of control, need for achievement, authoritarianism, mastery and patience. We also asked participants to rate their perseverance when faced with a difficulty in the use of software and wished assistance frequency.

Interaction Logging

During the use of PhotoScape, all the participants' actions were traced, as well as the assistance actions. Thus, we were able to determine what the participants did in PhotoScape, when and how often the assistance system provided help and for which subtask(s). We also captured how long they took and how many low-level actions participants carried out before following the advice.

Exit Questionnaire

The aim of this questionnaire was to measure the participant's satisfaction with the assistance provided by the system. We measured perceived frequency (1 – not frequent enough at all, 5 – far too frequent) and timeliness (1 – far too slow, 5 – far too quick). We also captured participants' assessment of the relevance of the assistance, their effectiveness and efficiency on 5-point Likert scales. Finally, we asked participants to give us feedback about the way SEPIA provided assistance in terms of the messages' clarity, content and layout, again on 5-point Likert scale.

RESULTS

To investigate how to provide the right advice at the right time, we analyzed participants' traces and questionnaire data. We excluded 3 participants from our analysis because they never received any assistance. We first investigated when assistance was followed using the logged traces, then we analyzed participants' subjective feedback about appropriate timing, and conclude with an analysis of the role of the right assistance in timing advice.

Following Assistance Provided

In order to investigate the right time to assist, we first wondered how often participants followed the assistance given. Understanding the time point at which users start to ignore assistance could give us a heuristic for adjusting the timing of advice. Thus, we analyzed participants' traces to investigate what participants did after receiving each piece of advice.

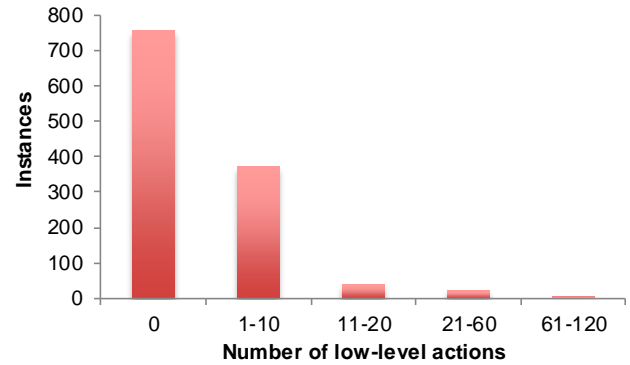


Figure 4. Number of low-level actions before advice is followed.

Each intervention of the assistance system suggested a low-level action to participants that directed them back to the task path, e.g. a click on a button, selecting a checkbox, etc. Thus, using the traces, it is possible to see if the suggestion was acted on. On average, all participants followed the assistance provided in less than 2 low-level actions and within 20 seconds. Overall, 64% of all instances when assistance was provided were followed *immediately*, i.e. 0 low-level actions before carrying out the suggested action (Figure 4). We investigated whether there were any differences in following the assistance based on the timing of the advice experienced. We found that participants in group A with an assistance timing of 0 seconds tended to follow the advice more quickly (mean=1.09 low-level actions), whereas the number of low-level actions participants carried out increased as assistance was delayed (max mean=3.59 in 24 seconds timing interval). However, there was no difference in the number of low-level actions before advice was followed whatever the timing interval that participants experienced ($F=1.145$, $p=0.335$).

We found that participants varied in how many low-level actions they carried out before following the advice, with 18 participants on average carrying out more than 3 low-level actions before following advice. We noticed that sometimes participants had a very large number of low-level actions in a particular step in the task; in one extreme case, a participant carried out 62 actions before following the advice. For participants that had very large number of low-level actions before following the assistance, we noted from their traces that it was because they deviated from the task instructions we had given them, for example, a participant added a black and white effect to the photo. In these circumstances, participants ignored the assistance we provided because they did not require it as they were doing a different task. Hence, instances with a high number of low-level actions indicate that our assistance was maybe unnecessary.

Hence, we started to look deeper into when assistance was followed immediately. Figure 5 shows that as the timing interval increased there was a decrease in the number of

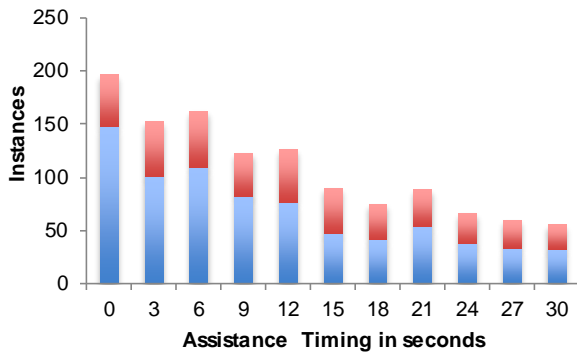


Figure 5. Number of instances followed immediately (blue) or after a carrying out at least one low-level action (red), per timing interval experienced.

times advice was given; on average, participants were assisted 14 times if they were in the group that experienced assistance timing of 0 second whereas participants who had a timing of 30 seconds only saw assistance 4.58 times. We found that there is a difference in the number of interventions ($F=36.326$, $p<0.0001$), indicating that as the timing increased, advice was less often given. This could have only occurred if they had already figured out how to complete the steps in this task, otherwise the assistance would have been triggered. We found that 75% of suggestions were followed immediately by participants who experienced a timing delay of 0 (i.e. they were shown advice immediately when they deviated from the path), whereas this dropped to 58% when advice was shown after 30 seconds (Figure 5). This indicates that more of the advice was ignored the longer it was delayed.

So where is that "sweet spot" at which assistance does not come too late or too early? We assumed that following advice immediately is an indication that the timing was right. We trained a decision tree (J48, an implementation of ID3 in weka) using the traces and participants' background variables that we captured in the pre-task questionnaire. We achieved accuracy of 62.92%, with 95.7% of immediately followed advice correctly predicted. The decision tree showed that participants' background variables did not seem

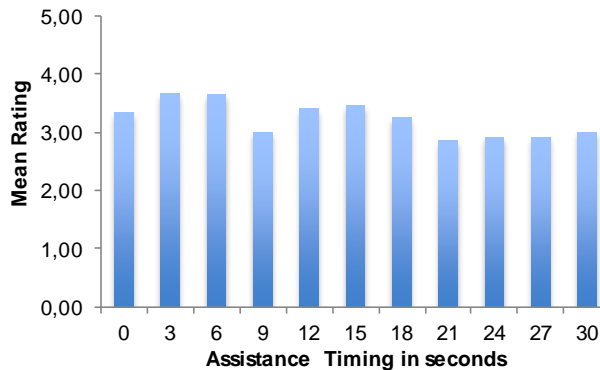


Figure 6. Perceived timeliness ratings for each group, per timing interval experienced.

to matter in most instances: 619 out of 1176 (52.64%) were correctly classified simply by keeping the timing below 9 seconds. Once timing reached above 9 seconds, then the participants only followed the advice immediately if they rarely edited photos, with 401 out of 1176 correctly classified (34%).

Taken together, this means that an overall rule of thumb could be to provide assistance within 9 seconds of determining that users require help, to ensure that the assistance provided is useful at the time it is provided. However, this does not take into account whether participants felt that the assistance was provided at the right time, which we turn to next.

Perceived Right Time of Assistance

Perceived Timeliness

We asked participants how they felt about the timing of the advice through the perceived timeliness ratings in the exit questionnaire, based on a 5-point Likert scale (recall that 1 means "far too slow", and 5 means "far too quick"). Figure 6 shows the distribution of perceived timeliness ratings for each group, indicating that on average participants, whatever the timing of assistance they experienced, rated the timeliness as very close to 3, i.e. as "at the right time". Using an ANOVA, we found that there was no significant difference between ratings based on the timing of when interventions were made ($F=1.106$, $p=0.363$). However, timeliness decreased as timing increased ($r=-0.205$, $p=0.015$). This correlation is only weak, indicating that there also appeared to be other factors that played a role in participants' perceived timeliness.

We wondered whether the timeliness ratings might be related to the frequency of interventions. First, recall that as timing increased the number of interventions dropped. We found that there is no significant correlation between the timeliness ratings and the number of interventions that participants experienced ($r=0.069$, $p=0.422$). We also investigated *perceived* frequency of advice given, based on the feedback of the participants. (Recall that we measured perceived frequency on a 5-point Likert scale, with 1

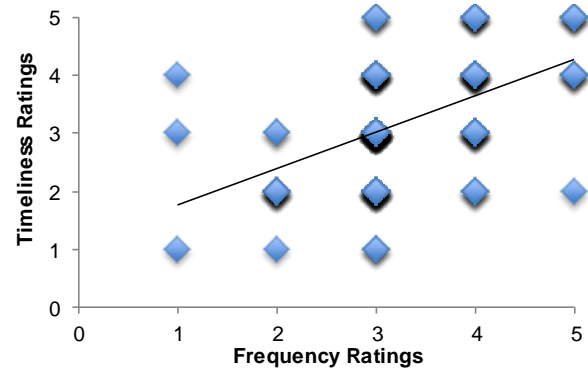


Figure 7. Participants' ratings on perceived frequency against perceived timeliness.

| | Coefficients | p |
|-----------------------------|--------------|-------|
| Constant | 4.524 | 0.000 |
| Assistance timing (sec) | -0.027 | 0.003 |
| Age | -0.002 | 0.800 |
| Gender | 0.078 | 0.688 |
| Expertise in photo editing | -0.164 | 0.045 |
| Self efficacy | 0.008 | 0.199 |
| Self esteem | 0.003 | 0.672 |
| Help-seeking | -0.012 | 0.187 |
| Perseverance | 0.142 | 0.106 |
| Wished assistance frequency | -0.386 | 0.001 |

Table 1. Factors in timeliness regression model (shaded shows significant)

meaning “not frequent enough at all” and 5 meaning “far too frequent”.) We found that participants' timeliness and frequency ratings are indeed significantly correlated ($r=0.494$, $p<0.0001$). This means as participants perceived advice to be given too frequently, they also perceived it as too quickly presented (Figure 7). However, perceived frequency and actual number of interventions are not correlated ($r=-0.132$, $p=0.123$), showing that what participants judged as being “too frequent” is very subjective.

Predicting Perceived Timeliness

We have already shown that timeliness was correlated with the assistance timing but we wondered what other factors might matter in perceived timeliness. It has been surmised that background factors play an important role in whether help is sought [1] and we therefore investigated the impact of participants' background factors on perceived timeliness.

Using a multiple regression, we found that the regression model significantly predicted perceived timeliness ($r=0.454$, $r^2=0.206$, $p=0.000298$) and that there were three important factors that matter in the prediction (Table 1): the timing interval the participant experienced, the expertise in photo editing the participants had, and the amount of assistance the participants wanted to receive. We discuss these important factors in detail now.

First, in our model, assistance timing had a negative impact on the perceived timeliness rating ($B=-0.027$) i.e. the slower advice was given after the participant deviated from the task, the slower they also perceived it. However, the coefficient shows the contribution of timing is quite low.

Second, their previous experience with carrying out the task mattered and had a negative impact on perceived timeliness ratings ($B=-0.164$). Hence, the higher their self-assessed expertise rating, the lower the perceived timeliness. This implies that the more they knew about photo-editing

previously, the slower they perceived the assistance to be given.

Third, the amount of help they wanted appeared to matter, again in a negative relationship ($B=-0.386$): the higher their rating on required assistance the lower the rating on perceived timeliness. This means that the more they wanted help, the slower they perceived help to arrive.

Last, it should be noted that there is a very important “anchor” from which participants seemed to judge the timeliness of advice. This baseline is represented by the constant ($B=4.524$), sitting very close to extreme end of the 5-point Likert scale, meaning that participants started out as perceiving advice given as “too quick” in most cases, and then decreased their ratings based on other factors. Indeed, 53 out of 141 (38%) participants rated the timing of the advice as “too quick”, whereas 54 rated it as “right” (38%), and only 34 (24%) as “too slow”.

These results indicate that while the timing of the advice and some personal background characteristics seem to matter to some degree in how users perceive the timeliness of advice. However, assistance was frequently judged as being given too quickly, whatever the background of the user.

Predicting the “Right Time”

We trained a decision tree (J48, an implementation of ID3 in weka) to predict the right timeliness, i.e. perceived timeliness of 3, from participants' background variables that we captured in the pre-task questionnaire and the assistance timing they experienced. Using 10-fold cross-validation, the resulting decision tree correctly classified 57.97% instances. Overall, it was easier to predict the wrong time to present the advice, than the right time: 78.8% of all wrong timeliness ratings were predicted corrected whereas the decision tree was correct for only 24.5% of all right timeliness ratings.

The decision tree is rather complex. We concentrate on the four subtrees that together accounted for 26 of 54 correct classifications for perceived timeliness:

Perseverance ≤ 4 & Gender = female & Timing ≤ 12 seconds & Self-esteem ≤ 70 and wished assistance frequency ≤ 3

(8 correctly classified instances)

Perseverance ≤ 4 & Gender = male & Age ≤ 41 & Expertise in Photoediting ≤ 2 & Self-efficacy ≤ 76 & Help-seeking > 64 & Frequency of photoediting > 1

(8 correctly classified instances)

Perseverance ≤ 4 & Gender = male & Age > 41

(5 correctly classified instances)

Perseverance <=4) & Gender = male & Age <=41 & Expertise in Photoediting > 2 & Perseverance >1 & Frequency of photoediting > 2 & Help-seeking > 48 & Timing >21

(5 correctly classified instances)

Note that in two of these, assistance timing does not matter at all, i.e. those participants were happy with the advice whenever it was given.

It should be noted that this is a very crude way of predicting the right time. First, timeliness is an aggregated measure for judging the timing that does not take into account individual instances of assistance that the participant experienced. For example, participant A1 saw 14 interventions but possibly some of the advice given was well-timed whereas others might have been offered too quickly or not quickly enough. Second, our training data is not ideally balanced; only 38% of participants rated the timing of the advice "right", whereas the majority thought the advice did not arrive at the right time. Last, we are also only using two classes – either right or wrong – instead of more subtle distinctions in timing the advice.

Note that when we add the variables captured in the final questionnaire to the decision tree (e.g. relevance, content and layout of advice, etc.), overall accuracy improves to 67.4%, with 64.2% of right timings correctly classified. However, these are variables that we would not usually be able to determine before users have experienced the assistance system, limiting their usefulness in practice.

A common approach in user modeling is to "stereotype" users and determine the behavior of the systems by how well a new user fits this stereotype. One way this could be done is by dividing data of users into clusters. We decided to use the three factor described as important in predicting the perceived timeliness in a cluster analysis.

We produced five clusters over the data set containing 141 participants, giving us reasonably distributed and separated data (see Table 2). We can identify three different approaches to timing advice: increase the time after which advice is given, speed up giving advice, or instances when the timing was about right.

| Attribute | Cluster# | | | | |
|-----------------------------|----------|---------|--------|---------|---------|
| | 0 | 1 | 2 | 3 | 4 |
| N | 15 | 38 | 31 | 23 | 34 |
| Assistance timing (sec) | 9.6 | 25.1842 | 3.0968 | 11.8696 | 18.0882 |
| Expertise in photo editing | 3.4 | 2.2368 | 1.5161 | 1.913 | 1.2941 |
| Wished assistance frequency | 2.0667 | 3.0526 | 3.2258 | 2.6522 | 2.5 |
| Timeliness | 4 | 2.4211 | 3.2581 | 2.4348 | 4.2647 |

Table 2. Cluster centroid information

Cluster 0 (N=15) and cluster 4 (N=34) both contained participants who rated the advice as being given "too quickly" (4 or 4.2647, respectively). Hence it might be an idea to increase the timing for users similar to these participants. In our study, cluster 0 contained participants who had a timing interval of just over 9 seconds, with medium experience of photo editing and not a great wish to be assisted. In cluster 4, even though they seemed to be beginners at photo editing and already had a timing interval of about 18 seconds, they wanted less help.

We can also identify two clusters, cluster 1 (N=38) and cluster 3 (N=23), where assistance timing was too slow and therefore we could decrease the assistance timing. Cluster 1 contained participants who requested a moderate amount of assistance, were not greatly experienced with photo editing and had a long timing interval of more than 25 seconds. In contrast, cluster 3's participants were equally not that experienced with photo editing but wanted less assistance and had a shorter timing interval of advice of about 11 seconds.

When did we actually get it right? Cluster 2 (N=31) seems to contain most of the participants who judged the advice as coming at the right time. These individuals were not very experienced in photo editing but wanted more help. In this case they experienced assistance timing of about 3 seconds.

To summarize, it was difficult to determine the right time to intervene based on perceived timeliness. We found that the accuracy of our predictive model was low; possibly because few participants thought assistance arrived at the right time and their feedback was not fine-grained enough. Using a clustering approach might yield better results but further work is needed to evaluate this approach in practice.

Perceived right assistance

It seems that timing advice is very difficult and benefits of assistance at any time might outweigh the cost of an interruption. Timing the advice right appears to be only part of the solution, and participants also expected the right advice. Indeed, some participants commented that this was the most important aspect in receiving advice e.g. "a useful assistance is nearly always welcome" (participant D10).

Most participants found the assistance appropriate to what they wanted to achieve (Figure 8): 77% of them found it relevant or very relevant and 81% stated that it helped them to achieve the task more quickly. Slightly fewer participants considered the advice useful: 65% of the participants said that the advice was effective. We found that there is no significant difference in perceived relevance between any of the groups (Figure 9), adding evidence to advice being relevant to users at any time it is provided.

A part of good advice is its clarity, content and layout (Figure 10); our way of providing assistance proved to be very popular with participants. We found that 82% of participants in our study found the assistance clear or very clear. Our advice was displayed using simple statements and 64% appreciated how we communicated the advice,

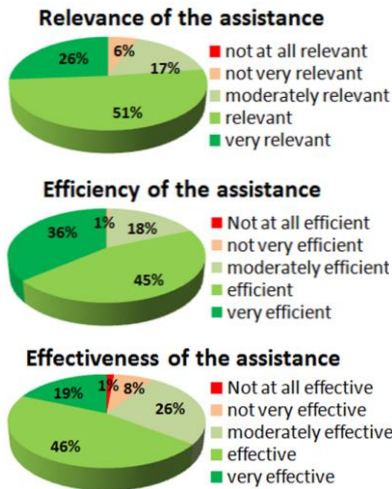


Figure 8. Evaluation of the relevance, efficiency and effectiveness of the assistance by the participants.

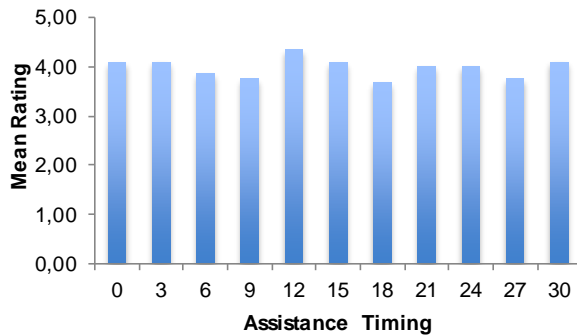


Figure 9. Mean perceived relevance for each group, per timing interval experienced.

while 87% appreciated the way we enhanced the UI with the advice.

DISCUSSION

We have collected interaction and preference data through a study which offered assistance to participants during a photo-editing task. Even though our study provides interesting results toward timing assistance right, it also has several limitations. First, the task we asked participants to carry out, although engaging and easily understandable, was very structured and constrained to a narrow subset of functionality within PhotoScape. Thus, assistance was triggered whenever participants deviated from this narrow task, which might have inflated the number of interventions that would be necessary in real-life use. Second, we did not account for task difficulty in our study design and some of the steps in our task instruction were very complex to do. It is possible that participants welcomed the given assistance at any time only for these steps, even though they might not have been timed well. Third, this also points to weakness in our preference data collection. We only captured an overall rough measure of timeliness of assistance from participants instead of feedback about each intervention. This means

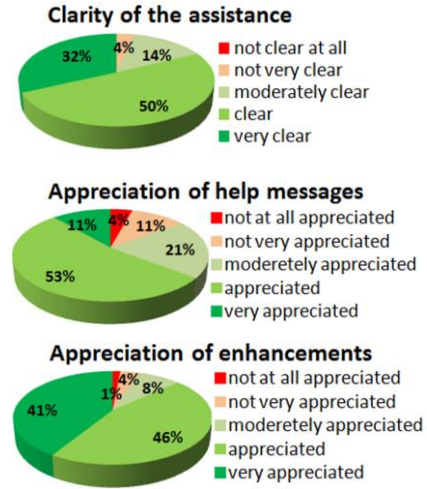


Figure 10. Evaluation of the clarity of the assistance and of help messages and enhancements by the participants.

that there is a substantial amount of "noise" in our data which makes prediction and modeling difficult. Last, we varied the assistance timing in 3-second intervals, introducing gaps in the evaluation of timing preference. Ideally, finer-grained intervals or choosing timings randomly for each participant possibly would have allowed us to build a better model.

Partly because of our limited data, predicting the right time to intervene proved to be very difficult. Also, it appears that the timing of assistance is very complex: we already indicated its relationship to perceived and wished frequency, expertise in the relevant task, and also other factors that have been shown important in help-seeking, such as age, gender and self-esteem. Further work to develop more accurate models for predicting the right time, which would involve a larger sample of users, finer-grained data and task difficulty [20], is warranted. However, we have shown that users' background does influence perceived timeliness and these factors should be included in any model to choose assistance timing.

Our findings also have implications for the design of assistance systems. First, proactive assistance removes control from the user in order to automate some actions. However, offering *mixed assistance*, in which on-demand assistance is provided in addition to proactive assistance, might improve the effectiveness of the assistance from the user's point of view. Second, identification of the user's task is currently quite basic, based on a deviation from an expected task path which has to be demonstrated by the assistance developer. However, if the current task along with an expected sequence could be predicted from users' actions, assistance could be made more accurate. Of course, gathering enough examples for task prediction might be challenging in this context but possibly previous work in detecting frequent procedures could be useful in these circumstances [36]. Last, we have shown that the style of

the assistance we provided was appreciated. This means that short, directive contextual help is a viable design option for these kinds of systems.

CONCLUSION

We conducted an empirical study to investigate how best to time assistance to users, capturing feedback by participants through logged interactions and their subjective ratings. Even though assistance timing proved very difficult to predict, our results showed that:

- When assistance was promptly given, advice was followed almost immediately. However, participants overwhelmingly rated assistance as too frequently and too quickly given.
- If assistance timing was longer than 9 seconds, following advice tailed off. We found that most appropriate results were clustered around 3 seconds for participants who wanted and needed more help with the task.
- The right time to intervene is very difficult to predict; if advice is felt to be relevant then it is welcome at any time. However, some aspects of users' background and their attitude to help-seeking do seem to matter in timing assistance right.

Our work has shown some early indications how to adjust assistance timing based on a user's preferences, however, further work is needed to fully understand the right time to provide the right assistance.

REFERENCES

1. Aleven, V., Stahl, E., Schworm, S., Fischer, F. and Wallace, R. (2003). Help Seeking and Help Design in Interactive Learning Environments. *Review of Educational Research*, 73(3), 277-320.
2. Baker, R. Corbett, A. and Koedinger, K. (2004). Detecting Student Misuse of Intelligent Tutoring Systems. *Proceedings of Seventh Int'l Conf. Intelligent Tutoring Systems*, 531-540.
3. Bardram, J., Bunde-Pedersen, J. and Soegaard, M. (2006). Support for activity-based computing in a personal computing operating system. *Proceedings of the SIGCHI conference on Human Factors in computing systems*, ACM, 211-220.
4. Beck, J. (2005). Engagement Tracing: Using Response Times to Model Student Disengagement. *Artificial Intelligence in Education: Supporting Learning through Intelligent and Socially Informed Technology*, Looi C. et al., eds., IOS Press, 88-95.
5. Brusilovsky, P. and Cooper, D. W. (2002). Domain, task, and user models for an adaptive hypermedia performance support system. *Proceedings of the 7th international conference on Intelligent user interfaces*, ACM, 23-30.
6. Bull, S. and Kay, J. (2010). Open Learner Models. *Advances in Intelligent Tutoring Systems*, R. Nkambou, J. Bourdeau, R. Mizoguchi (Eds.), Springer Berlin Heidelberg, 301-322.
7. Carlier, F. and Renault, V. (2010). Educational webportals augmented by mobile devices with iFrimousse architecture. *ICALT*, Sousse, Tunisia.
8. Cocea, M. and Weibelzahl, S. (2011). Disengagement detection in online learning: validation studies and perspectives. *Learning Technologies*, IEEE Transactions on, 4(2), 114-124.
9. Compeau, D. R. and Higgins, C. A. (1995). Computer self-efficacy: Development of a measure and initial test. *MIS Quarterly*, 189-211.
10. Dragunov, A. N., Dietterich, T. G., Johnsrude, K., McLaughlin, M., Li, L. and Herlocker, J. L. (2005, January). TaskTracer: a desktop environment to support multi-tasking knowledge workers. *Proceedings of the 10th international conference on Intelligent user interfaces*, ACM, 75-82.
11. Dufresne, A. and Paquette, G. (2000). ExploraGraph: A Flexible and Adaptive Interface to Support Distance Learning. *Ed-Media*, Victoria, Canada, 304-309.
12. Fenstermacher K. D. and Ginsburg, M. (2002). A Lightweight Framework for Cross-Application User Monitoring. *Computer* 35, 3, 51-59.
13. Ginon, B. Thai, L. V., Jean-Daubias, S., Lefevre, M. and Champin, P.-A. (2014). Adding epiphytic assistance systems in learning applications using the SEPIA system. *Ec-Tel*, Graz, Austria.
14. Ginon, B., Champin, P. A. and Jean-Daubias, S. (2013). An approach for collecting fine-grained use traces in any application without modifying it. *Workshop EXPORT of ICCBR*, New-York, USA.
15. Ginon, B., Jean-Daubias, S., Lefevre, M. and Champin, P. A. (2014). Adding epiphytic assistance systems in learning applications using the SEPIA system. *Ec-Tel*, Austria, Graz, 138-151.
16. Grandhi, S. and Jones, Q. (2010). Technology-mediated interruption management. *International Journal of Human-Computer Studies*, 68(5), 288-306.
17. Haigh, K. Z., Kiff, L. M. and Ho, G. (2006). The independent lifestyle assistant: Lessons learned. *Assistive Technology*, 18(1), 87-106.
18. Hedberg, S. R. (1998). Executive Insight: Is AI Going Mainstream at Last? A Look Inside Microsoft Research. *IEEE Intelligent Systems*, 13(2), 21-15.
19. Horvitz, E., Breese, J., Heckerman, D., Hovel, D. and Rommelse, K. (1998). The Lumiere project: Bayesian user modeling for inferring the goals and needs of software users. *Proceedings of the Fourteenth conference*

on *Uncertainty in artificial intelligence*, Morgan Kaufmann Publishers Inc, 256-265.

20. Iqbal, S. T. and Horvitz, E. (2010). Notifications and awareness: a field study of alert usage and preferences. *Proceedings of the 2010 ACM conference on Computer supported cooperative work*, 27-30.
21. Kantner L. and Rusinsky L. (1998). Designing a WinHelp project for quick conversion to lowest-common-denominator HTML-based help: a case study. *Conference SIGDOC*, Quebec, Canada, 214-218.
22. Koedinger, K. R. and Alevan, V. (2007). Exploring the Assistance Dilemma in Experiments with Cognitive Tutors. *Educational Psychology Review*, 19(3), 239-264.
23. Lieberman H. (1997). Autonomous interface agents. *ACM SIGCHI Conference on Human factors in computing systems*, New York, USA, 67-74.
24. Matthews M., Pharr W., Biswas G. and Neelakandan H. (2000). USCSH: An Active Intelligent Assistance System. *Artificial Intelligence Review*, Kluwer Academic Publishers, 14, 121-141.
25. Mille, A., Champin, P. A., Cordier, A., Georgeon, O. and Lefevre, M. (2013). Trace-based reasoning-modeling interaction traces for reasoning on experiences. *The 26th International FLAIRS Conference*, 1-15.
26. Nadler, A. (1997). Personality and Help Seeking. *Sourcebook of Social Support and Personality*, G. R. Pierce, B. Lakey, I. G. Sarason and B. R. Sarason (Eds.), Springer US, 379-407.
27. Nadler, A. and Jeffrey, D. (1986). The Role of Threat to Self-Esteem and Perceived Control in Recipient Reaction to Help: Theory Development and Empirical Validation. *Advances in Experimental Social Psychology*, L. Berkowitz (Ed.), Academic Press, 19, 81-122.
28. Oliver, N. Smith, G. Thakkar, C. and Surendran, A.C.. (2006). SWISH: semantic analysis of window titles and switching history. *Proceedings of the 11th international conference on Intelligent user interfaces*. ACM, New York, NY, USA, 194-201.
29. Paquette, G., Pachet, F., Giroux, S. and Girard, J. (1996). EpiTalk, a generic tool for the development of advisor systems. *IJAIED*, 349-370.
30. Paquette, G., Rosca, I., Mihaila, S. and Masmoudi, A. (2007). TELOS: A Service-Oriented Framework to Support Learning and Knowledge Management. *E-Learning Networked Environments and Architectures*, Springer London, 79-109.
31. Ramachandran, A. and Young, R. M. (2005). Providing intelligent help across applications in dynamic user and environment contexts. *Proceedings of the 10th international conference on Intelligent user interfaces*, ACM, 269-271.
32. Rattenbury, T. and Canny, J. (2007). CAAD: an automatic task support system. *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM, 687-696.
33. Richard, B. and Tchounikine, P. (2004). Enhancing the adaptivity of an existing Website with an epiphyte recommender system. *New review of hypermedia and multimedia*, 31-52.
34. Rosenberg, M. (1965). Society and the adolescent self-image. Princeton, NJ: Princeton
35. Selker T. (1994). COACH: a teaching agent that learns. *Communication of the ACM*, 37 (7), 92-99.
36. Shen, J., Fitzhenry, E., Dietterich, T. (2009). Discovering Frequent Work Procedures From Resource Connections. *Proceedings of the International Conference on Intelligent User Interfaces*.
37. Shen, J., Irvine, J., Bao, X., Goodman, M., Kolibaba, S., Tran, A. and Dietterich, T. G. (2009). Detecting and correcting user activity switches: algorithms and interfaces. *Proceedings of the 14th international conference on Intelligent user interfaces*, Sanibel Island, USA, ACM, 117-126.
38. Swartz L. (2003). Why People Hate the Paperclip: Labels, Appearance, Behavior, and Social Responses to User Interface Agents, *Master's report*, Stanford University, Stanford, USA.
39. Tambe, M. (2008). Electric elves: What went wrong and why. *AI magazine*, 29(2), 23.
40. Vivacqua, A. and Lieberman, H. (2000). Agents to assist in finding help. *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, ACM, 65-72.
41. Willis, M. (2006). Building effective help systems: modelling human help seeking behaviour. *Proceedings of the 18th Australia conference on Computer-Human Interaction: Design, Activities, Artefacts and Environments*, ACM, 433-436.
42. Yorke-Smith, N., Saadati, S., Myers, K. L. and Morley, D. N. (2009). Like an intuitive and courteous butler: a proactive personal agent for task management. *Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems*, International Foundation for Autonomous Agents and Multiagent Systems, 1, 337-344.
43. Zarka, R., Champin, P-A., Cordier, A., Egyed-Zsigmond, E., Lamontagne, L. and Mille., A. (2012). TStore: A Web-Based System for Managing, Transforming and Reusing Traces. *ICCBR 2012 TRUE and Story Cases Workshop*, Luc Lamontagne, Juan A. Recio-Garc ed. Lyon, France, 173-182.