Perceptual & Motor Skills: Learning & Memory
2015, 120, 1, 67-83.

SUPERVISED AUTOMATIC INTERPRETATION OF TECHNICAL DOCUMENTS: WHEN INTERRUPTION IS A TIME SAVER[1]

SYLVAIN FLEURY AND ERIC JAMET

CRPCC, University of Rennes 2, France

Summary.—The present study examined human-computer interactions in which the operator has to detect errors made by software designed to automatically recognize technical documents. The goal was to assess the effect of user-initiated interruptions of the recognition process to correct these errors. Participants were asked to check the interpretations, either with or without the possibility of interrupting the process. Results showed that interruptions can improve efficiency by decreasing task duration, especially in the post-recognition verification phase. Interruptions provide an opportunity to correct errors during rather than after the recognition process, which is easier because it requires fewer cognitive resources.

The production of hand-drawn sketches is an important stage in a number of occupations, including architecture (Bilda, Gero, & Purcell, 2006). These sketches then have to be digitized, either manually or automatically. Sometimes, the symbols in paper drawings also have to be recognized and interpreted (*retroconversion*) so that these documents can be converted into forms that can be searched, archived, or modified. Many research projects have been undertaken recently with a view to developing technical document retroconversion software to meet the needs of industry. These technical documents can include not just architectural plans (Ahmed, Liwicki, Weber, & Dengel, 2011) but registry maps (Boatto, Consorti, Del Buono, Di Zenzo, Eramo, Espositio, *et al*., 1992), electrical diagrams (Ouyang & Davis, 2009), and road maps (Chiang & Knoblock, 2013). The present study was conducted under the aegis of the MobiSketch project (ANR 09-CORD-015) to design software capable of recognizing offline digitized plans and rebuilding them in a format compatible with computer-aided design (CAD) software. Although the results of this project will be applicable to all kinds of technical documents, architectural floor plans were chosen.

The processing performed by this type of software inevitably carries a risk of error. Based on a sample of 200 drawings, Lu, Tai, Su, and Cai (2005) estimated that their automatic recognition system had a 90% recognition rate. Assessments by Ahmed, *et al*. (2011) of their own automatic analysis method had a 79% recognition rate (different plans were used in each assessment). Finally, evaluations of a prototype by Ghorbel, Lemai-

[1]Address correspondence to Sylvain Fleury, Research Centre in Psychology, Cognition, and Communication (CRPCC), University of Rennes 2, Place du recteur Henri Le Moal, CS 24 307 Rennes, 35043 France or e-mail (sylvain.fleury@uhb.fr).

DOI 10.2466/22.25.PMS.120v11x9 ISSN 0031-5125

tre, and Anquetil (2012) yielded a 91% recognition rate for simple plans. Results like these mean that an operator must be present and able to correct these errors. Users need to be considered during the design process (Oncina, 2009) to make the product as ergonomic as possible. However, in most current pattern-recognition systems, users are made to wait until the interpretation is complete before correcting the errors (Liu & Boyle, 2009), a design choice that fails to take their characteristics properly into account. Moreover, making corrections during the recognition process might allow subsequent errors to be avoided.

Several experiments have been conducted with a view to facilitating error correction by users. For example, one study found that the task of tracking down errors was completed more quickly when the plan and its software interpretation were superimposed, rather than being displayed side by side with the original (Fleury & Jamet, 2014). In addition, this study showed that making the interpretation appear gradually (i.e., sequentially), rather than all at once, improved error tracking. Superimposition presumably obviates the need for visual searching, thereby making it easier to co-reference the information being compared. As for the positive effect of sequentiality, the authors surmised that it was related to the attentional guidance provided by the gradual appearance of the interpretation. This dictates the order in which users explore that interpretation, leading them to check it more thoroughly. In a study conducted by Fleury, Jamet, Ghorbel, Lemaitre, and Anquetil (2014) and featuring the same type of application, the authors demonstrated that an automated help that flags the risk of errors could facilitate the task. The functioning of the prototype system described in Ghorbel, *et al*. (2012) suggests that the final version of the software will be able to learn in the course of the recognition process, provided that users correct errors *during* rather than *after* it. In the study conducted by Fleury and Jamet (2014), participants could not interrupt the analysis when they detected errors in the sequential display condition. Although they were given the possibility of interrupting the software interpretation to correct errors in a subsequent study (Fleury, *et al*., 2014), the authors did not evaluate the effect of these interruptions.

Making corrections sooner rather than later may improve the software's performance by averting error propagation. There is therefore a technical advantage to allowing users to interrupt the system to correct errors as soon as they appear. In addition to this technical gain, there may be an additional benefit related to the users' cognitive functioning.

Detecting errors (targets), memorizing their location, and possibly interrupting the analysis are three different kinds of activities that constitute three different fields of study. In the present paper, a series of studies of target detection, short-term forgetting, and interruption is described. An experiment was designed to assess the effect of system interruptions by users.

Target Detection

In the study described in the present paper, users had to supervise the automatic recognition of plans. This involved checking the symbol interpretations and detecting any errors. In the field of target detection, only a handful of studies have been conducted with a naturalistic task, generally related to the detection of weapons concealed in luggage, based on X-ray pictures.

In a study conducted by Wiegman, McCarley, Kramer, and Wickens (2006), participants had to look for a knife in each X-ray picture. Only 20% of the pictures actually contained a knife. Some participants had to do this task without any help, but others received an automated aid. The latter consisted of a simulated automatic pattern recognition system capable of identifying a knife in a picture with an accuracy of 0.90 and 0.25 false alarms. This assistance could take the form of either a text signaling the presence or absence of a knife, or a local cue consisting of a circle surrounding the specific area that contained a weapon. This study revealed that local cues were more efficient than general ones. With the same kind of task, Goh, Wiegmann, and Madhavan (2005) showed that target detection is improved by a general indication of whether or not a weapon is present in the image, without specifying its location. However, performances were improved still further when a direct indication was added. The same type of result was obtained by Alberdi, Povyakalo, Strigini, Ayton, and Given-Wilson (2008), in a study of automatic assistance for lesion detection in mammograms. Again, detection performances tended to increase when the assistance took the form of relevant information, but decreased when that information was irrelevant (signaling a location without a lesion). It should be noted that 13.5% of *recognitions* are, in fact, false alarms, and automatic aid has no effect on these false alarms. In another study featuring the same type of material, McCarley, Kramer, Wickens, Vidoni, and Boot (2004) assessed the effect of training on performances on the knife detection task. These authors found that novices were particularly inaccurate. This result was confirmed in a study by Madhavan and Gonzalez (2006), in which training increased the weapon detection rate to 40%. Overall, these studies show that target detection on screen is a difficult task, especially for novices.

Short-term Forgetting

Short-term forgetting has been the subject of extensive research over many years. In a famous paper summarizing experimental results, Miller (1956) claimed that short-term memory can hold around seven chunks. As

Cowan (2000) points out, however, the more recent theories developed by Baddeley (1986) and Richman, Staszewski, and Simon (1995) indicate that short-term memory is limited not in storage capacity but in the amount of time an item can be kept active in memory without rehearsal. In a detailed review of the literature, Cowan (2000) defended the idea that the number of chunks that can be stored in short-term memory falls to four when rehearsal and chunk formation are both impossible.

Elliott and Strawhorn (1976) demonstrated a similarity effect between information stored in memory and a concurrent task. Interference favoring forgetting was greater when the modality of the concurrent task was the same as the modality of the stimulus to be remembered (auditory or visual). Proctor and Fagnani (1978) also showed an effect of similarity on the short-term memorizing of triplets of consonants coupled with a Brown-Peterson distractor task. Results indicated that when the distractor task was in the same modality (auditory or visual) as the main task, memorization was poorer than when the two tasks were in different modalities.

Likewise, performances on an odor recognition task were reduced when participants were exposed to other odors during the period of memory retention (Walk & Johns, 1984), whereas a kinesthetic task did not interfere with a concurrent task of memorizing numbers (Williams, Beaver, Spence, & Rundell, 1969). Interference was also greater when the stimuli were the same kind. For example, word storage interfered more strongly with the processing of other words than with the manipulation of numbers. This was demonstrated by Smyth, Pearson, and Pendleton (1988) in a set of five experiments in which participants were asked to memorize patterns of movements, words, or spatial positions, and perform a motor, spatial, or phonological distractor task. The items had to be recalled in the right order, in a memory span paradigm. Interference was only observed when the main task and distractor task featured the same stimulus type. Other studies have replicated these results with different types of stimuli. For example, Klauer and Stegmaier (1997) showed that a task of sound localization interferes with the storage of spatial information. More recently, Lee, Lee, and Tsai (2007) carried out the same type of experiment with lists of written words. Again, a phonological task performed during the retention period had a negative effect on short-term memory.

In the same kind of experiment, Washburn and Astur (1998) observed both forgetting and interference. This experiment involved a visuospatial matching task. The longer the retention time, the less well the matching task was performed. In other words, the visual patterns were gradually forgotten. In addition, this study showed that a visuospatial distractor (a cursor moving across the screen, generating visual tracking behavior) preventing visual rehearsal during the retention period accentuates the forgetting.

The results of the above studies demonstrating an interference effect solely when the stimuli were the same nevertheless need to be nuanced. Other results suggest that interference also occurs between different stimuli, albeit to a lesser extent. For example, Vergauwe, Barrouillet, and Camos (2009) found interference between processing and storage even when one of the tasks referred to spatial content and the other to visual content. Thus, the processing of visual and spatial stimuli draws on the same pool of cognitive resources. The fact that the memorizing and processing of visuospatial content interfered with each other further suggests the sharing of common resources. This type of result is consistent with the time-based resource-sharing model (Barrouillet, Bernardin, & Camos, 2004). According to this model, storage and processing interfere with each other because they share cognitive resources.

Concerning storage, information in working memory tends to deteriorate gradually. To counter forgetting, the stored information must be reactivated by focusing attention on it (Henson, 1998; Page & Norris, 1998). However, only one item can be rehearsed at a time (Garavan, 1998; Rohrer, Pashler, & Etchegaray, 1998), and during rehearsal attention is focused and unavailable for concurrent processing. Conversely, during processing, attention cannot be used to rehearse information in memory. Therefore, according to Barrouillet, *et al.* (2004), so-called *cognitive resource sharing* actually corresponds to *time sharing*. Participants who have to perform processing and storage concurrently have to divide their attention between the two tasks. They therefore spend some time on the processing task, and allocate the remaining time to the revival of stored information. In this type of double task, attention is switched between the two activities (Barrouillet, *et al.*, 2004). Regardless of the nature or modality of the stimuli used, the period during which attention is focused on the processing activity curtails the amount of time spent reactivating the information to be memorized. Thus, the longer the processes performed during the retention period, the poorer the information storage (Barrouillet, *et al.*, 2004).

Interruption

Since the Zeigarnik effect (Zeigarnik, 1927) was first demonstrated, many studies have focused on the effect of interruption on cognitive task performance. The effects of interruption depend on the type of task and the type and frequency of the interruptions. Although the Zeigarnik effect should mean improved memorization of items when the task is interrupted, the consequences of interruption can sometimes be negative and are task-dependent (Prentice, 1944; Butterfield, 1964). For example, a study by Burmistrov and Leonova (2003) showed that interrupting a text correction task (a phone rings, and the participant has to answer) tends to increase the duration of the correction process if the task is difficult. The same pattern of results was found by Eyrolle and Cellier (2000). These authors found that external interruptions of a database completion task tended to increase task duration, although they did not significantly affect the number of errors. According to Speier, Valacich, and Vessey (1999), the effect of external interruptions depends on the complexity of the main task. These authors showed that interruptions facilitate decision making in a simple task, but hinder it in a complex one. Moreover, increasing the frequency of interruptions reduces decision-making performance in a complex task. In some studies, external interruptions correspond to prospective signals designed to tell the user to stop the main task and perform a secondary one. This kind of interruption was studied by Kvavilashvili, Messer, and Ebdon (2001). In their research, children were asked to name pictures on cards as a main task. They also had to remember to do something (put the card in a box) whenever they saw the target picture. These interruptions had a negative effect on performance of the ongoing task. Dodhia and Dismukes (2009) went further, by considering that interrupted tasks are special cases of prospective memory.

Interruptions during a task can not only reduce performances, but also have a negative effect on the individual's emotional state (Bailey, Konstan, & Carlis, 2000, 2001). Moreover, the effect of interruptions is generally modulated by the user's mental load (Bailey, *et al.*, 2001). Schiffman and Greist-Bousquet (1992) evaluated the effect of interruptions when they were not followed by a resumption of the task. In their study, participants were asked to solve a series of 10 anagrams in two different conditions. In the control condition, participants were simply asked to solve 10 anagrams. In the experimental condition they were asked to solve 20 anagrams, but the task was interrupted after the 10th anagram. Although the actual activity was the same in both conditions, the participants who were interrupted overestimated the task duration compared with those who were not interrupted.

It is important to distinguish external interruptions from self-initiated interruptions. Self-initiated interruptions occur when "the person interrupts herself or himself and breaks from the current task to focus on a different task" (Salvucci & Taatgen, 2010, p. 161). The alert (the reason why the interruption occurs) can be external, even with self-initiated interruptions, but it is the operator who triggers the interruption at a specific point in time.

Self-interruptions resemble external interruptions in terms of the cognitive process they involve. Whenever the operator interrupts him- or herself, the state of the task has to be maintained in memory. This requires mental rehearsal so that the information can be retrieved once the interruption is over (Salvucci & Taatgen, 2010). Self-initiated interruption has been studied less than external interruption, even though these two types of interruption are equally frequent in computer activities (Gonzalez & Mark, 2004). Self-interruption can take many forms. Jin and Dabbish (2009) established a list of self-initiated interruptions by operators working on computers: changing the environment, seeking information about the primary task, breaking to decrease fatigue or frustration, remembering an unrelated task, perceiving a cue to a related task, performing a task out of habit, and filling idle time.

Studies of external interruption often report a negative effect on performances, but results for self-interruption are fewer and less clearcut. To the authors' knowledge, no study has yet explored self-interruption as a strategy in a detection task.

Overview of the Present Study

In the present study, participants had to identify errors made by automatic recognition software in interpreting an architectural plan (i.e., *retroconversion*). This analysis took place sequentially (i.e., symbol interpretations gradually appeared on the screen, one at a time).

Short-term forgetting occurs after just a few seconds (Card, Moran, & Newell, 1986), and is accentuated by an interference task (Washburn & Astur, 1998). When they are unable to interrupt the retroconversion process, users not only have to retain the location of the errors they have identified for several tens of seconds (i.e., until the interpretation is complete), but also have to continue looking for new ones. Searching for new errors can be regarded as both a visuospatial and a phonological (the interpretations are labeled as words) task that competes with the task of maintaining the previously identified errors. As a result, participants locate errors during the recognition process, but are then unable to perform visuospatial rehearsal because of the task of finding new errors.

Hypothesis 1. The interruption group would detect more errors than the no-interruption group.

Hypothesis 2. The no-interruption group would try to compensate by spending more time checking the interpretation at the end than those participants in the interruption group who actually interrupted the process. Participants detecting errors during the recognition process probably would do it faster than those detecting them after it, owing to the attentional guidance afforded by the sequential display of the interpretation.

Hypothesis 3. The total interaction time in the interruption group (calculated by summing the interruption and verification times) would be shorter than the verification time in the no-interruption group. Since the duration of the recognition process was always the same, participants who could not interrupt the process would spend more time overall on the task than those who could interrupt it.

Method

Participants

The participants in this study were 36 volunteers (10 men, 26 women) who were all novices in handling architectural plans. The youngest was 18 years old and the oldest 33.3 years old ($M = 24.4$ yr., $SD = 2.8$). All the participants had normal or corrected-to-normal vision and French was their native language.

Material and Procedure

Every task in this experiment was performed on an ASUS Eee Slate 12.1 tablet. Each participant had to look for and circle errors of interpretation made by the software in three successive plans (Fig. 1). Eighteen participants performed the experiment with the interruption version of the software, and 18 with the no-interruption version. The two groups were balanced in terms of sex and age. Software version (interruption vs no-interruption) was the between-participants factor. The source plan was immediately displayed on the screen and the interpretation was gradually superimposed on it. With the no-interruption version, participants had to wait until the recognition process was complete before they could start circling the errors. The interruption version of the software was exactly the same, except that participants could interrupt the process by clicking a pause button and circle the errors before the interpretation was complete.

There were three floor plans, containing 33, 45, and 44 symbols, respectively (doors, sliding windows, casement windows, furniture). Without any interruption, the recognition process lasted 109, 96, and 70 sec., respectively. These plans were interpreted by the prototype, and the participants had to detect all the errors of interpretation (Fig. 2). Each participant checked the interpretations of the three different plans in the same experimental condition. Using three plans, each with six errors, reduced the chances of participants finding either all the errors or none of them. The order of appearance of the three plans was counterbalanced within each experimental condition, to avoid an effect of plan. Performance on the error detection task was measured as an efficacy score (number of errors detected) and an efficiency score (task duration). It should be noted that, for reasons of experimental control, interruptions did not enhance the recognition algorithm of the prototype software during the analysis. Also, for reasons of experimental control, participants were not allowed to

FIG. 1. The three interpreted plans that participants had to check

use external aids such as notepads during the task. This type of software was designed to be used in a mobile context (i.e., with a tablet and no other equipment).

For the no-interruption group, the first step was the interpretation (or recognition) phase, in which the symbol interpretation software analyzed the plan. Recognition time for a given plan was controlled and held constant for all participants. When this phase was complete, the participants embarked on the verification phase, in which they had to identify and circle the recognition errors. When they deemed that all the errors had been corrected, they had to end the task by clicking the Finish button. Verification time corresponded to the length (in seconds) of this error correction phase. Consequently, the total duration of the task for this group was equal to the duration of the interpretation and verification phases.

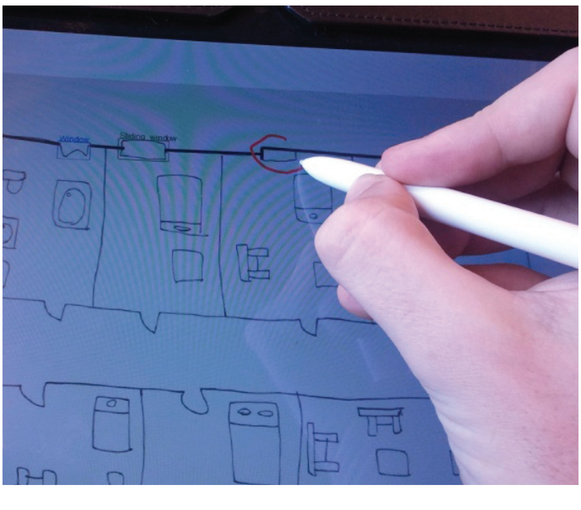

FIG. 2. Participant circling an error after interrupting the retroconversion process

In the interruption group, the only difference was that participants were able to interrupt the analysis (i.e., the interpretation phase) to circle errors. Interruption time was calculated by summing the duration of all the interruptions. When the plan had been fully interpreted by the software, participants proceeded to the verification phase, which lasted until the end of the task. A total detection duration was calculated by summing the interruption and verification times. Consequently, the total duration of the task for this group was equal to the duration of the interpretation and detection phases.

Results

Interruption Behavior

A nonparametric Mann-Whitney *U* test revealed that the task duration for the four participants in the interruption group who chose not to interrupt the process was significantly different from that of participants who did interrupt it ($U=7.00$, $p=.03$), but not from that of participants who could not do so ($U=27.00$, $p=.44$). This result tends to support the idea that users who do not make interruptions are more comparable to those who *cannot* make interruptions than to those who *do* make them. Thus, in subsequent statistical analyses, these four participants who did not interrupt were added to the no-interruption group.

Taken together, the 18 participants in the interruption group made an average of 4.17 interruptions ($SD=2.83$) during the task (see Fig. 3 for the distribution of participants according to the number of interruptions they made). As indicated above, four of them chose not to use the interruption functionality. When only those participants who made at least one interruption were taken into account, the mean number of interruptions rose to 5.36 ($SD=2.79$). These participants identified a total of 15.64 ($SD=1.98$) errors (i.e., 86.7% of all errors). This represents an average of one interruption for every 2.92 errors.

Effect of Interruptions

The error detection performances of participants who made interruptions were compared with those who did not. A Levene's test used to analyze homogeneity of variance did not reveal any significant difference between variances on the scores for the detected errors ($F_{1,34}=2.25$, $p=.14$) between the interruption group ($M=15.64$, $SD=1.99$) and the no-interruption group ($M=15.27$, $SD=4.23$). An analysis of variance (ANOVA) conducted on the participants' error detection performances also failed to reveal any significant difference ($F_{1,34}=0.93$, $MSE=1.17$, $p=.76$). The first hypothesis of improved error detection in the interruption condition was therefore not supported.

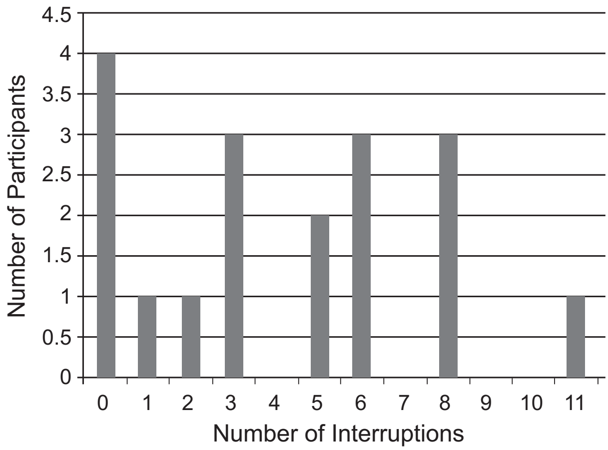

FIG. 3. Distribution of participants according to the numbers of interruptions they made

Figure 4 shows the mean duration of each of the three phases of the task: recognition/interpretation by the software (the duration of this process was the same for all participants), interruptions (only for participants who could use this functionality), and verification (error detection once the interpretation was complete). A correlation analysis revealed significant negative correlations between the duration of the verification phase and both the duration (r_{14}=–.48, p=.05, 95%CI=–.77, –.01) and number (r_{14}=–.72, p<.001, 95%CI=–.89, –.39) of interruptions. The more interruptions participants made, the less time they spent looking for errors at the end. Moreover, the more time they spent on interruptions, the less time they spent on verification. This result was consistent with the second hypothesis, that interruption would reduce the duration of the verification phase.

The third hypothesis was that interruptions would reduce the total detection (interruption + verification) duration for the interruption group compared with the verification duration for the no-interruption group. Levene's test assessed homogeneity of variance for these two durations. This did not reveal any significant difference ($F_{1,34}$=2.80, p=.10) between variances for the interruption group (M=175.26, SD=63.71) and the no-interruption group (M=242.99, SD=105.82). An ANOVA conducted on the two durations revealed a significant difference ($F_{1,34}$=4.63, MSE=39176.49, p=.04, η^2=0.12). The mean detection duration for participants who interrupted the interpretation process was significantly lower than the verification duration for participants who did not. This result was consistent with the third hypothesis.

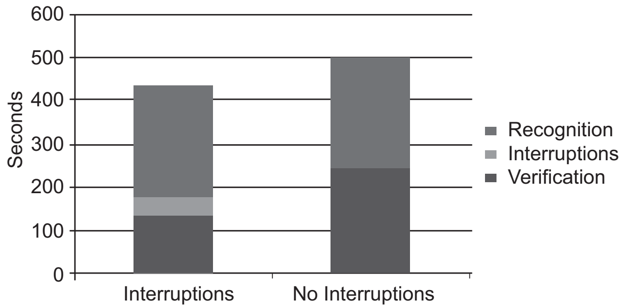

FIG. 4. Mean duration of the three phases of the task: recognition, interruptions, and verification

DISCUSSION

The first hypothesis of this study was that interruptions improve efficacy (i.e., error detection rates). This hypothesis was not supported. The difference in the number of errors detected by participants in the interruption vs no-interruption groups was very small and statistically nonsignificant. There are two possible explanations for this. The first is that participants who could not (or did not) interrupt the analysis may have been aware that they risked forgetting some of the errors. Consequently, they undertook a second exploration of the interpretation at the end. This explanation is supported by the finding that participants who did not interrupt the process spent more time on verification than those who did. The second possible explanation is that there was a ceiling effect in the error detection score: the average number of errors detected was 15.64 with interruptions and 15.27 without interruptions, out of a total of 18 errors.

The second hypothesis was that interruptions reduce the duration of the post-recognition verification phase. Results supported this hypothesis, as a negative correlation between the interruption and verification durations was observed. Participants who spent the most time detecting errors during the recognition process were those who spent the least time doing so after it. Moreover, participants who made interruptions and those who did not detected an equivalent number of errors. These results can therefore be interpreted as a simple displacement of the error detection activity. Participants who detected errors during the recognition process had fewer errors to detect during the verification phase. This final phase therefore took less time.

The third hypothesis was that interruptions improve efficiency. This hypothesis was confirmed by the results, as participants spent 27.87% less time on error detection when they made interruptions. The amount of time spent on interruptions was lower than the amount of time saved in the verification phase. The fact that this difference concerned detection duration, but not the number of errors detected, suggests that participants who did not make any interruptions were aware of the risk of forgetting some of the errors. They therefore compensated for this by taking more time to make checks afterwards. This interpretation is consistent with the results of Washburn and Astur (1998), who highlighted an interference effect between a visuospatial distractor and the storage of visual configurations. If some participants forgot the location of errors because of the interference, they had to check the interpretation a second time at the end. This second exploration during the verification phase was done without the guidance of the sequential display (Fleury & Jamet, 2014), and therefore took longer than the error search carried out during the recognition process. This result is also coherent with the time-based resource-sharing model (Barrouillet, *et al.*, 2004). During the retroconversion, participants had to process the information to identify errors. While they were doing that, they could not reactivate the errors they had previously stored in memory. They therefore forgot them and had to re-explore the document afterwards. Each time they interrupted the system, participants obtained time in which they did not have to process new information and could devote themselves to detecting the errors.

These results may seem surprising, in view of the literature on task interruption. In general, interruptions tend to have a negative effect on the main task (Bailey, *et al.*, 2001). In particular, they tend to increase task duration (Burmistrov & Leonova, 2003). In contrast, in this study interruptions reduced the overall duration of the task. This incongruence could be explained by the nature of the task. The interruptions that occurred during this experiment fall into the category of "Perceiving a cue to a related task" in Jin and Dabbish's (2009) classification. Interruptions of this type usually waste time as they relate to a second task, i.e., not really linked to the first one, and therefore require participants to spend time remembering or rediscovering the status of the first task. In the current study, interruptions were used to perform a secondary task that was a part of the main task. Interruptions therefore saved time, because the subtask of surrounding the errors was easier to perform during the recognition phase (with interruptions) rather than all at the end. By interrupting the analysis and surrounding errors, participants could lessen their memory load because they no longer had to memorize the locations of the errors.

Analyses of the results were performed by grouping the four participants from the interruption group who did not make any interruptions with the 18 participants who were not given the opportunity to do so. Comparisons of their performances confirmed that these four participants were closer to those who could not make interruptions than to those who could and did. However, the experiment did not allow the authors to explain why these four particular participants chose not to use the interruption functionality. It is possible that they overestimated their ability to memorize the errors and thus thought that they did not need to use the interruption functionality. The second main limitation of this study was that the sample was made up of novices, whereas in a real-life situation users would probably be experts. This undoubtedly had an effect on results.

To conclude, the interruption functionality allows users to detect errors during the recognition process. While most pattern recognition systems allow users to make corrections afterward, it seems preferable to allow them to perform these corrections during the processing. First, this may prevent the system from making subsequent errors. Moreover, it improves efficiency in the case of a sequential display, as users can benefit from the attentional guidance it affords (Fleury & Jamet, 2014). From an application point of view, giving users the ability to interrupt the analysis has many advantages. First, the attentional guidance saves them time. In addition, it can also improve system performances, for some types of software are capable of learning from their mistakes, so when a user corrects an error the system does not repeat that error again. In a fully automated system, where the software does not allow the user to interrupt the recognition process, the human is entirely out of the control loop. Out-of-the-loop performance can have several negative consequences on the functioning of human-machine systems (Kaber & Endsley, 1997), including reduced vigilance, overconfidence in the machine, and decreased situation awareness. By enabling users to intervene, the system moves to the supervised control level on the scale of automation (Kaber & Endsley, 1997). This adjustment in the level of automation partially reintroduces the user into the loop and can thus improve vigilance and situation awareness (Endsley & Kaber, 1999). Generally speaking, the level of automation should be tailored to the task. Automatic processes that risk making errors require the intervention of operators to correct them. Rather than correcting these errors at the end, it may be better to reduce the level of automation by enabling users to intervene during the processing. Efficiency is thus improved and the user is partially reintroduced into the control loop. However, because working memory obviously plays a role in identifying and maintaining errors, it would be interesting to make direct measurements of it. This would allow determination of how working memory can predict the use of the interruption capability and the efficacy and efficiency of error detection. This could be assessed in future research.

REFERENCES

Ahmed, S., Liwicki, M., Weber, M., & Dengel, A. (2011) Improved automatic analysis of architectural floor plans. *Proceedings of the 2011 International Conference on Document Analysis and Recognition*, Beijing, China, September 18-21. Pp. 864-869.

Alberdi, E., Povyakalo, A. A., Strigini, L., Ayton, P., & Given-Wilson, R. (2008) CAD in mammography: lesion-level versus case-level analysis of the effects of prompts on human decisions. *International Journal of Computer Assisted Radiology and Surgery*, 3(1-2), 115-122.

Baddeley, A. D. (1986) *Working memory*. Oxford, UK: Oxford Univer. Press.

Bailey, B. P., Konstan, J. A., & Carlis, J. V. (2000) Measuring the effects of interruptions on task performance in the user interface. *2000 IEEE Internal Conference on Systems, Man, and Cybernetics*. Vol. 2. Piscataway, NJ: Institute of Electrical and Electronics Engineers. Pp. 757-762.

Bailey, B. P., Konstan, J. A., & Carlis, J. V. (2001) The effects of interruptions on task performance, annoyance, and anxiety in the user interface. In M. Hirose (Ed.), *Human-computer interaction—INTERACT 2001 conference proceedings.* Amsterdam, The Netherlands: IOS Press. Pp. 593-601.

Barrouillet, P., Bernardin, S., & Camos, V. (2004) Time constraints and resource sharing in adults' working memory spans. *Journal of Experimental Psychology: General*, 133(1), 83-100.

Bilda, Z., Gero, J. S., & Purcell, T. (2006) To sketch or not to sketch? That is the question. *Design Studies*, 27(5), 587-613.

Boatto, V., Consorti, V., Del Buono, M., Di Zenzo, S., Eramo, V., Espositio, A., Melcarne, F., Meucci, M., Morelli, A., Mosciatti, M., Scarci, S., & Tucci, M. (1992) An interpretation system for land register maps. *IEEE Computer Magazine*, 25(7), 25-33.

Burmistrov, I., & Leonova, A. (2003) Do interrupted users work faster or slower? The microanalysis of computerized text editing task. In J. Jacko & C. Stephanidis (Eds.), *Human-computer interaction: theory and practice (Part 1)—proceedings of HCI International* (2003). Vol. 1. Mahwah, NJ: Erlbaum. Pp. 621-625.

Butterfield, E. C. (1964) The interruptions of tasks: methodological, factual, and theoretical issues. *Psychological Bulletin*, 62(5), 309-322.

Card, S. K., Moran, T. P., & Newell, A. (1986) The model human processor: an engineering model of human performance. In K. Boff (Ed.), *Handbook of perception and human performance*. New York: Wiley. Pp. 1-35.

Chiang, Y-Y., & Knoblock, C. A. (2013) A general approach for extracting road vector data from raster maps. *International Journal on Document Analysis and Recognition*, 16(1), 55-81.

Cowan, N. (2000) The magical number 4 in short-term memory: a reconsideration of mental storage capacity. *Behavioral and Brain Sciences*, 24, 87-185.

Dodhia, R. M., & Dismukes, R. K. (2009) Interruptions create prospective memory tasks. *Applied Cognitive Psychology*, 23, 73-89.

Elliott, L. A., & Strawhorn, R. J. (1976) Interference in short-term memory for vocalization: aural versus visual modality differences. *Journal of Experimental Psychology: Human Learning and Memory*, 2(6), 705-711.

Endsley, M. R., & Kaber, D. B. (1999) Level of automation effects on performance, situation awareness and workload in a dynamic control task. *Ergonomics*, 42(3), 462-492.

Eyrolle, H., & Cellier, J-M. (2000) The effects of interruptions in work activity: field and laboratory results. *Applied Ergonomics*, 31(5), 537-543.

Fleury, S., & Jamet, E. (2014) Facilitating the comparison of multiple visual information on screen: the example of electronic architectural plan correction. *Applied Ergonomics*, 45(3), 601-607.

FLEURY, S., JAMET, E., GHORBEL, A., LEMAITRE, A., & ANQUETIL, E. (2014) Application of the resources model to the supervision of an automated process. *Human-Computer Interaction*, 30(2), 103-121.

GARAVAN, H. (1998) Serial attention within working memory. *Memory & Cognition*, 26(2), 263-276.

GHORBEL, A., LEMAITRE, A., & ANQUETIL, E. (2012) Competitive hybrid exploration for off-line sketches structure recognition. *2012 International Conference on Frontiers of Handwriting Recognition*, Bari, Italy, September 18-20. Pp. 571-576.

GOH, J., WIEGMANN, D. A., & MADHAVAN, P. (2005) Effects of automation failure in a luggage screening task: a comparison between direct and indirect cueing. *Proceedings of the Human Factors and Ergonomics Society 49th Annual Meeting.* Santa Monica, CA: Human Factors and Ergonomics Society. Pp. 492-496.

GONZALEZ, V. M., & MARK, G. (2004) Constant, constant, multi-tasking craziness: managing multiple working spheres. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York: ACM Press. Pp. 113-120.

HENSON, R. N. (1998) Short-term memory for serial order: the start-end model. *Cognitive Psychology*, 36(2), 73-137.

JIN, J., & DABBISH, L. A. (2009) Self-interruption on the computer: a typology of discretionary task interleaving. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. New York: ACM Press. Pp. 1799-1808.

KABER, D. B., & ENDSLEY, M. R. (1997) Out-of-the-loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety. *Process Safety Progress*, 16(3), 126-131.

KLAUER, K. C., & STEGMAIER, R. (1997) Interference in immediate spatial memory: shifts of spatial attention or central-executive involvement. *The Quarterly Journal of Experimental Psychology*, 50A, 79-99.

KVAVILASHVILI, L., MESSER, D. J., & EBDON, P. (2001) Prospective memory in children: the effects of age and task interruption. *Developmental Psychology*, 37(3), 418-430.

LEE, Y. S., LEE, H. M., & TSAI, S. H. (2007) Effects of post-cue interval on intentional forgetting. *British Journal of Psychology*, 98, 257-272.

LIU, S., & BOYLE, I. M. (2009) Engineering design: perspectives, challenges, and recent advances. *Journal of Engineering Design*, 20(1), 7-19.

LU, T., TAI, C-L., SU, F., & CAI, S. (2005) A new recognition model for electronic architectural drawings. *Computer-Aided Design*, 37, 1053-1069.

MADHAVAN, P., & GONZALEZ, C. (2006) Effects of sensitivity, criterion shifts, and subjective confidence on the development of automaticity in airline luggage screening. *Proceedings of the Human Factors and Ergonomics Society 50th Annual Meeting*. Santa Monica, CA: Human Factors and Ergonomics Society. Pp. 334-338.

MCCARLEY, J. S., KRAMER, A. F., WICKENS, C. D., VIDONI, E. D., & BOOT, W. R. (2004) Visual skills in airport-security screening. *Psychological Science*, 15(5), 302-306.

MILLER, G. A. (1956) The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological Review*, 101(2), 343-352.

ONCINA, J. (2009) Optimum algorithm to minimize human interactions in sequential Computer Assisted Pattern Recognition. *Pattern Recognition Letters*, 30(5), 558-563.

OUYANG, T. Y., & DAVIS, R. (2009) A visual approach to sketched symbol recognition. In H. Kitano (Ed.), *Proceedings of the 21st International Joint Conference on Artificial Intelligence*. San Francisco, CA: Morgan Kaufmann Publishers. Pp. 1463-1468.

PAGE, M. P., & NORRIS, D. (1998) The primacy model: a new model of immediate serial recall. *Psychological Review*, 105(4), 761-781.

PRENTICE, W. C. (1944) The interruption of tasks. *Psychological Review*, 51(6), 329-340.

PROCTOR, R. W., & FAGNANI, C. A. (1978) Effects of distractor-stimulus modality in the Brown-Peterson distractor task. *Journal of Experimental Psychology: Human Learning and Memory*, 4(6), 676-684.

RICHMAN, H. B., STASZEWSKI, J. J., & SIMON, H. A. (1995) Simulation of expert memory using EPAM IV. *Psychological Review*, 102(2), 305-330.

ROHRER, D., PASHLER, H., & ETCHEGARAY, J. (1998) When two memories can and cannot be retrieved concurrently. *Memory & Cognition*, 26(4), 731-739.

SALVUCCI, D. D., & TAATGEN, N. A. (2010) *The multitasking mind*. New York: Oxford Univer. Press.

SCHIFFMAN, N., & GREIST-BOUSQUET, S. (1992) The effect of task interruption and closure on perceived duration. *Bulletin of the Psychonomic Society*, 30(1), 9-11.

SMYTH, M. M., PEARSON, A., & PENDLETON, L. R. (1988) Movement and working memory: patterns and positions in space. *The Quarterly Journal of Experimental Psychology*, 40A(3), 497-514.

SPEIER, C., VALACICH, J. S., & VESSEY, I. (1999) The influence of task interruption on individual decision making: an information overload perspective. *Decision Sciences*, 30(2), 337-360.

VERGAUWE, E., BARROUILLET, P., & CAMOS, V. (2009) Visual and spatial working memory are not that dissociated after all: a time-based resource-sharing account. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 35(4), 1012-1028.

WALK, H. A., & JOHNS, E. E. (1984) Interference and facilitation in short-term memory for odors. *Perception & Psychophysics*, 36(6), 508-514.

WASHBURN, D. A., & ASTUR, R. S. (1998) Nonverbal working memory of humans and monkeys: rehearsal in the sketchpad? *Memory & Cognition*, 26(2), 277-286.

WIEGMANN, D., MCCARLEY, J. S., KRAMER, A. F., & WICKENS, C. D. (2006) Age and automation interact to influence performance of a simulated luggage screening task. *Aviation, Space, and Environmental Medicine*, 77(8), 825-831.

WILLIAMS, H. L., BEAVER, W. S., SPENCE, M. T., & RUNDELL, O. H. (1969) Digital and kinesthetic memory with interpolated information processing. *Journal of Experimental Psychology*, 80(3), 530-536.

ZEIGARNIK, B. (1927) Über das Behalten von erledigten und underledigten Handlungen [About the keep of completed and pending actions]. *Psychologische Forschung* [*Psychological Research*], 9, 1-85.

Accepted December 10, 2014.