User-centred design of an interactive off-line handwritten architectural floor plan recognition

Sylvain Fleury, Achraf Ghorbel, Aurélie Lemaitre, Eric Anquetil, Eric Jamet

User-centred design of an interactive off-line handwritten architectural floor plan recognition

Sylvain Fleury‡, Achraf Ghorbel*, Aurélie Lemaitre†, Eric Anquetil§, Eric Jamet‡

*Université Rennes 1,§INSA de Rennes, †‡Université Rennes 2

*†§UMR IRISA, Campus de Beaulieu, F-35042 Rennes

‡ CRPCC, Université Rennes 2, Place du recteur Henri Le Moal, CS 24 307, 35043 Rennes Cedex

Université Européenne de Bretagne, France

{achraf.ghorbel, aurelie.lemaitre, eric.anquetil}@irisa.fr

{sylvain.fleury, eric.jamet}@uhb.fr

***Abstract*—In this paper, we present the impact of user interaction in the recognition of off-line structured documents. This interaction requires solving two major problems: how interpretation results will be presented to the user, and how the user will interact with analysis process. We propose to study the effects of those two aspects in the context of an interactive method (IMISketch) for off-line handwritten 2D architectural floor plan recognition. The use tests are realized in collaboration with researchers in cognitive psychology (more than 100 persons participated in the tests). The experiments demonstrate that (i) a progressive presentation of the analysis results, (ii) user interventions during it and (iii) the user solicitation by the analysis process are an efficient strategy for interactive recognition of off-line documents.**

***Keywords*—interactive recognition; uses tests; solicitation user; architectural floor plan;**

I. Introduction

In the literature, two types of recognition methods are presented: batch and interactive methods. The *batch* [1] [2] [3] methods are not assisted by the user. Although the batch methods do not need the presence of user during the analysis, these methods lead accumulation errors. The interactive methods avoid the propagation of errors and requires the presence of the user during the analysis phase. Unfortunately, the study of interactivity and computation time gives an additional difficulty for the implementation of such methods.

We have proposed IMISketch method which is an interactive method able to solicit the user if necessary. We previously reported our work on the optimization of calculation time [4] [5]. Now, we want to study the ways of interaction. For that, a series of user tests was conducted with researchers in cognitive psychology and ergonomics in the *Loustic* platform [1] with a user centred development method. This interactivity has been the subject of several studies [6]. Several questions need to be answered. We will give response of two questions: how interpretation results will be presented to the user, and how the user will interact with analysis process. To answer these questions, we will rely on use tests. These tests are done using the *IMISketch* method [7].

The remaining of the paper is organized as follows. In the section II, we introduce the principles of our existing interactive analysis method *IMISketch* for interpreting sketches. The manner to present the interpretation results is described in section III. The evaluation of the interaction between user and analysis process is reported in section IV and V. Finally, section VII concludes the paper.

[1]Loustic is a platform located in Rennes (France) for multidisciplinary research on user-centred design methods

II. Overview of IMISketch

In this section, we summarize the principles of our interactive method and we present the different steps of the existing method IMISketch. IMISketch solicits the user (when necessary) to reduce a verification phase which could be tedious. This method consists of four major blocks: a priori knowledge, primitive extraction, tree construction and decision process shown in Figure 1.

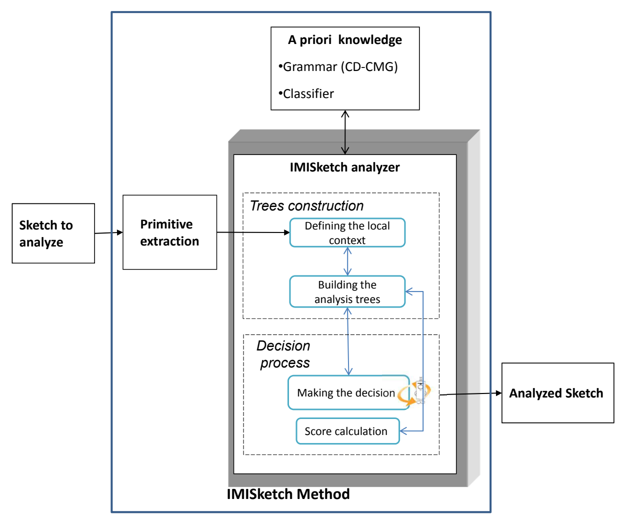


Fig. 1. IMISketch Processing

A. IMISketch characteristics

IMISketch method is based on a two-dimensional visual language through production rules to express a priori structural knowledge of the document. In addition, *IMISketch* adopts a hybrid exploration unlike classical syntactic analysis methods based on hybrid exploration. Each branch of the tree is a possible hypothesis. The uncertainty is formalized by the attribution of scores to each hypothesis. If the ambiguities can not be resolved in an automatic manner, the user will be solicited by the analyser to resolve the ambiguity. Two types of ambiguities can be managed: structural ambiguity and form ambiguity. Structural ambiguity is an ambiguity between two symbols having a different structure. In the case of architectural plans, we can find a structural ambiguity between a wall and an opening (door, window,...). Ambiguity of form is an ambiguity between two symbols having the same structural context : for example furniture.

B. Analysis progress

The first step is the extraction of basic primitives required for analysis. The aim of this phase is to extract all the basic primitives that will be used to analyse the document. These primitives are line-segments and polygons. These primitives will be interpreted based on CD-CMG grammars [8].

The primitive interpretation depends on its neighbourhood in structured documents: the structured document analysis requires a two-dimensional context. The analyser begins by defining a spatial contextual focus that aims to limit the combinatory exploration due to the hybrid exploration of the analysis tree.

This two-dimensional local context is defined for an analysis tree as the maximum distance between the elements of the root and the elements of any leaves. The choice of the size of the local context depends on the application domain. For example, to interpret an architectural plan, we suggest a local context with a size corresponding to the maximum size of entity in the document.

Once the local context is defined, the method of analysis builds analysis trees. Indeed, the analyser explores all the possible hypotheses of interpretation in the spatial context using a set of two-dimensional rules that describe the structure of the document. Each primitive can be interpreted in several ways which led to a construction of an analysis tree. In the building of the analysis tree, the analyser explores all the possible hypotheses of interpretation using hybrid exploration in the spatial context with the algorithm described in [4]. Each root is the production rule that would consume this primitive. Each node or leaf is the application of a production rule deduced from the previous node. The number of analysis trees corresponds to the number of possible interpretations for the current primitive.

Each leaf or node of the tree has a score calculated from both its local score and the score obtained from the preceding nodes. Each score determines the adequacy degree to validate a production. It is calculated from each rule. The production score can also be deduced from a classifier. Each branch (hypothesis) is characterized by a score.

Once the tree is well constructed, we start the decision phase. The role of the decision process is to validate the right hypothesis among a set of competing hypotheses generated with a descending hybrid analysis. It is a structural decision. The decision process also validates the recognition of symbol shapes. Sometimes the decision process is not sure to make the right decision by validating the best hypothesis (because it has a too low score or it goes into confusion with the other hypotheses). In this case, an interaction between analysis process and user is required. In the remainder of this paper, we will focus on the properest way to interact with the user.

III. Presentation of interpretation on the screen

A. Experimental Design

A first test has been conducted to provide an answer to the following problem: how to present information to users in order to make it most likely to identify all errors? This test has been passed by 54 volunteers (19 men and 35 women) from 18 to 31 years. The IMISketch method, was used to interpret three plans, like the one of Figure 2. It recognized walls, doors and windows. We synthetically introduced many mistakes in the recognition process in order to study how the participants would identify them. On doors and windows, only 50% of the symbols were correctly interpreted.

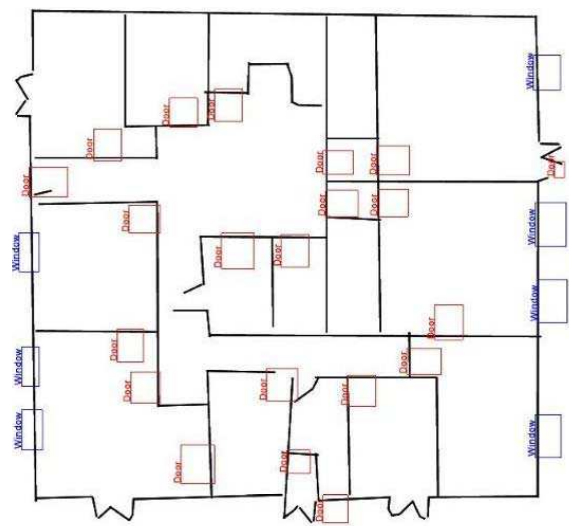


Fig. 2. Example of an interpreted plan

Three experimental conditions were compared. In the *separate condition*, the manuscript plan appeared on the left side of the screen, and the interpretation appeared on the right side of screen. In the *integrated condition*, the manuscript plan appeared on the middle of the screen. Then, the interpretation plan appeared superimposed to it.

Finally, in the *sequential format*, the interpretation plan was built gradually over the manuscript plan. As in the integrated condition, the interpretation was superimposed to the manuscript plan. The recognition process was perceived by participants. For each experimental condition, participants compared 3 pairs of different plans.

B. Results

Task duration was subjected to an ANOVA[2]. We compute the time token by each participant to localize the mistakes in each condition. The results showed that the response to integrated condition was shorter (170.65 seconds) than task duration measured in the separate condition (228.12 seconds, $F(1, 32) = 13.586$, $MSE = 2066.426$, $p = .001$ (Table I)). That represents a decreased of 20%.

[2]ANOVA is a statistical method assessing the significance of a difference between averages

	Task duration	
	M	SD
Separated	228.12	48.182
Integrated	170.65	42.560
Sequential	103,72	39,716

TABLE I. TASK DURATION (MEANS AND STANDARD DEVIATIONS)

Error detections require to mentally link each symbol with it interpretation. In the *separate condition*, participants encode a visuo-spatial information from the manuscript plan to match it with the interpretation plan. Results suggest that the superposition of plans removes these steps of visual search. This interpretation is coherent with previous studies on parallel processing of visual information in the field of multimedia learning [9] [10]. In the *integrate condition*, when the participant looks at a particular point in the interpreted plan, he has sufficient information to identify an error if there is one.
Comparison of performance of integrate with sequential format is not made for the duration of the task. In the sequential condition, although participants did not have the opportunity to surround errors during the interpretation, they can still mentally begin the task. In the integrate condition and separate one, the participants cannot begin to search for errors before the end of the retroconversion. Therefore, duration differences probably could be explained by the simple fact that participants from sequential group began the task before the two others groups did. Concerning errors detection accuracy (Figure 4), likelihood ratio showed no significant difference between separate and integrate format (LR(1, N = 34) = .485, p =.485). However, participants in the sequential group were significantly more likely to have identify all errors (88% of them) than those from integrate group (39%, LR(1, N = 35) = 6.660, p = .01).

The explanation for this result implies to understand that the sudden appearance of an element on the screen (here, the interpretation of a given symbol) triggers an attentional capture [11] [12] [13] [14] [15]. Therefore, each item appearance should generate a visual saccade on it and a verification of the interpretation relevance. The gain obtained in the sequential condition could be due to an attentional shift toward the different part of the sequential display. Following this order ensures participants to monitor the full plan. In the integrate condition, participants looked freely increasing the probability to make some errors. The sequential format seems to be the most advantageous because it allows on integration of both plans and a gradual appearance of interpretation which improves error detections. However, participants are led to identify errors in a given period and correct them later. They must memorize the errors and are likely to forget some. Some interruptions with corrections during the task could improve the performance.

IV. INTERRUPTION OF INTERPRETATION BY THE USER

A. Experimental Design

Retroconversion software automates a task that previously would have been made manually by human who are "out-of-the-loop" [16]. Bringing-back users "into-the-loop" could improve performance by monitoring the recognition. Then, the software would learn depending of corrections done by users. A second test was conducted to assess the impact of interruptions of interpretation by users on the interaction. In this test, 36 volunteers (10 males and 26 females) aged 18 to 33 years were asked to surround the misinterpretations of three successive plans (Figure 5) that where synthetically introduced in the interpretation of IMISketch method. Half of participants performed this task in the *sequential format* equivalent to that of the previous experiment. The second half had the possibility to *interrupt the system* during the interpretation to surround errors directly. To do that, they only needed to press a pause button on the screen. Interpretations of the three used plans each contained six errors.

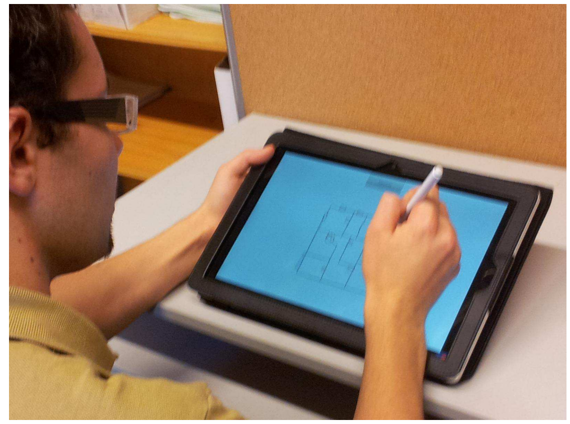


Fig. 3. A participant who performs the task

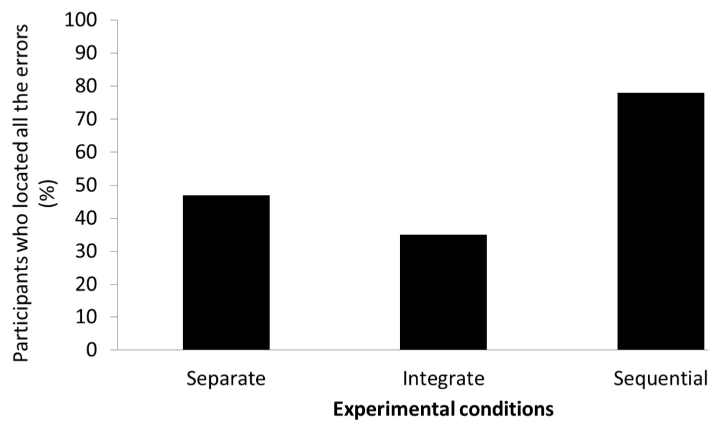
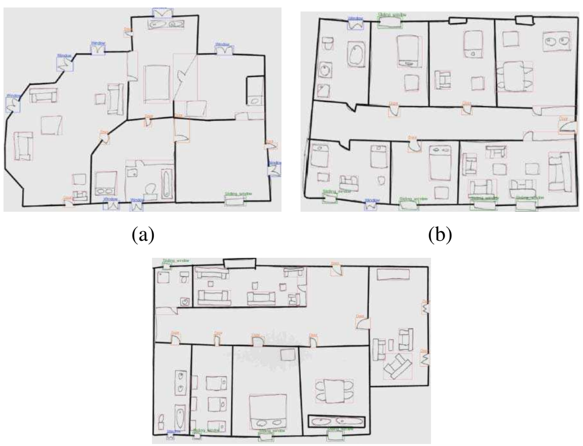


Fig. 4. Percentage of participants who detected all the errors

B. Results

Likelihood ratio for identifying every errors reveals no significant difference between interruptions by users group and without interruption group (LR(1, N = 36) = .874, p = .350). In contrast, an ANOVA revealed that participants who performed system interruptions during the task finished faster (448.71 seconds on average) than those who did not (515.61 seconds, F(1, 35) = 4.745, MSE = 41017.161, p = .036).

The average time saving was amount to 13% of total time (Table II). Indeed, using the pause button saves time. This result is confirmed by the negative correlations of the duration of posterior correction with interruptions duration and with interruptions amount. Indeed, as shown by [17], memorization of visuospatial material is subject to a progressive forgetting that can be amplified by a visual distractor preventing rehearsal. Credible interpretation of our results could be that participants who do not use the pause button forget previously identified errors because of the interfering task that is the search of other errors. They must recrawl the entire plan after the end of the retroconversion and so tend to verify errors twice. Participants who interrupted the system did not need to perform a final recrawl. From an applied point of view, it seems to be preferable to allow users to intervene on live when they identify an error. The interruption of the system by users increases the interactions efficiency. One may wonder whether an interruption that would be performed by the system itself could also generate a performance improvement.



(a) (b) (c)

Fig. 5. The three plans controlled by participants

	Task duration	
	M	SD
With interruptions	448,71	67,138
Without interruptions	517.95	105,837

TABLE II. TASK DURATION (MEANS AND STANDARD DEVIATIONS)

V. INTERRUPTION OF INTERPRETATION BY THE SYSTEM

A. Experimental Design

A third user test study was realized with 18 volunteer students (12 females and 6 males) aged 18 to 25 years. They were all novices when it came to processing architectural plans. In this test, participants supervised three plans interpretations. They were asked to surround the errors each time they detected one. Participants were able to surround the errors after the end of the interpretations or in blocking it by clicking on the pause button, visible on the screen as in the previous experiment. The prototypal software was equipped with an additional functionality of interrupting itself for soliciting users. When confronted with a symbol which interpretation was characterized by a low likelihood score, the prototypal software blocked interpretation. Then, it highlighted the targeted symbol and the corresponding uncertain interpretation and opened a window containing the written question ”Is it a mistake?”. Participants had to click ”Yes” to automatically surround the error and restart interpretations or ”No” to resume interpretations without surrounding the symbol. Each plan's interpretation was synthetically interrupted six times by the system. Three interruptions concerned real errors and the three others were relevant interpretations. Thus, there are six misinterpretations located on each plan. Three of them were identified by the system interruptions and the three others were not.

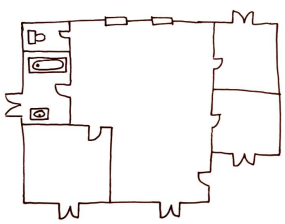
B. Results

After the end of the task, the experimenter asked participants if the interruptions by the system had rather been a help or a hindrance to them and why. To this question, ten participants responded that it helped them, seven responded that it had disturb them and one said neither one nor the other. Those that answered that interruptions by the system disturbed them explained that the difficulties arose from some graphic characteristics of the interface. For example, some suggested that the location of the window appearing to ask them to confirm the error was irrelevant. Other said that the highlight of the targeted area was too dark or the incoherent apparition order of interpretations. Finally, some mentioned the rate of relevant interruptions (i.e. the insufficient number of times an interruption corresponds to an error and errors that do not trigger interruption). The principle of interruption by the system was never mentioned but peripheral aspects always were. It is thus reasonable to conclude that this functionality was thought well of by users, but it requires some improvements. For example, it is likely that the impact of this solicitation on interaction is strongly related to the number of requests on a real error, and also to the number on errors not specified by this device. In our experiment, even though the system is not very accurate for these solicitations (it detects 3 errors of 6 and 3 times solicits users for symbols without error), it is rather considered as an aid by users. An ongoing study seeks to assess the effects of an interruption made by the system on task performance, its duration and its error detection accuracy.

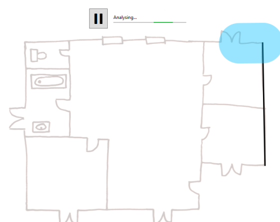
VI. EXPERIMENTAL RESULTS AND DISCUSSION

We used a user centred design method to design the introduction of interaction in IMISketch method. Ergonomics recommendations from user tests have been applied to the software. Indeed, the first test which is describe in section III showed that the superimposition of the manuscript plan and it interpretation allows the users to save time. Second, this test revealed that the gradual appearance improves the accuracy of errors tracking by the users. As shown on Figures 6(a) and 6(d), the current interface displays the interpretation superimposed on the manuscript plan (sequential, integrated). In addition, the interpretation appears gradually on the screen.

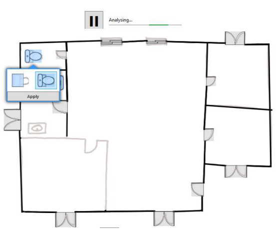
The interface contains a pause button (Figures 6(b) and 6(c)) as recommended by the test presented in the section IV. Thus, the users save time because they can correct the errors during the interpretation process. The users solicitation has also been implemented. This functionality was considered as assistance by the users (see the third test in the section V). The window displaying the possible corrections appears directly next to the targeted area as suggested by the tests participants. Finally, since the test participants talked about the unpredictable order of recognition, the area processed by the current system is highlighted in blue during the recognition (Figure 6(b)). This focus allows the users to know which symbol will be the next to be interpreted. Some tests participants said they had trouble in watching the gradual appearance of the interpretation. Thanks to this focus, the users are guided.



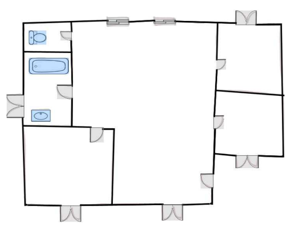
(a) Original document



(b) System focus by blue highlight



(c) Upgraded software's interface



(d) An interpreted document

Fig. 6. Analysis process screenshots

The previous experimentations [4] have shown that the presence of the user during the analysis improves the recognition rate. Indeed, the experiments on 24 architectural handwritten plans (Figure 5) shows that the structural recognition rates increase from 91% without user solicitations to 96% with user solicitations. We can notice that the best compromise (recognition rates / user solicitations) is obtained with 12 user solicitations per image: it means that 4% of primitive interpretations require the user solicitation. 49% of user solicitations are useful to take the right decision which is not the best hypothesis proposed by the analyser.

VII. Conclusion

In this paper, we have presented several use tests for an interactive off-line handwritten architectural floor plan recognition. These tests were applied to centred design of IMISketech method. The aim of these tests is to determine the best way to present the results to the user interpretation and the best manner to interact with the user. The experiments, made on more than 100 persons, have shown that display the interpretation result of documents in a progressive manner is most appreciated by the participants. In addition, other experimental recommend interacting with the user by soliciting him if needed.

Currently, IMISketch method only allows the solicitation of users. Future work will focus on extending the IMISketch method. Future work will focus on extending the method IMISketch by adding the ability of the user to pause the analyser to correct errors not detected by the analyser.

Acknowledgment

The authors would like to thank all the people who took part in the experiments. This work partially benefits from the financial support of the ANR Project Mobisketch (http://mobisketch.irisa.fr).

References

[1] K. Chan and D. Yeung, "An efficient syntactic approach to structural analysis of on-line handwritten mathematical expressions," *Pattern Recognition*, vol. 33, no. 3, pp. 375–384, 2000.

[2] J. Fitzgerald, F. Geiselbrechtinger, and T. Kechadi, "Mathpad: A fuzzy logic-based recognition system for handwritten mathematics," in *ICDAR 2007*, vol. 2, 2007, pp. 694 –698.

[3] S. Mao, A. Rosenfeld, and T. Kanungo, "Document structure analysis algorithms: a literature survey," in *Proc. SPIE Electronic Imaging*, vol. 5010, 2003, pp. 197–207.

[4] A. Ghorbel, A. Lemaitre, and E. Anquetil, "Competitive hybrid exploration for off-line sketches structure recognition," in *Proceedings of International Conference on Frontiers of Handwriting Recognition (ICFHR'2012)*, 2012.

[5] A. Ghorbel, E. Anquetil, and A. Lemaitre, "Optimization analysis based on a breadth-first exploration for a structural approach of sketches interpretation," in *10th IAPR International Workshop on Document Analysis Systems (DAS'12)*, 2012, pp. 240–244.

[6] S. Brennan and E. Hulteen, "Interaction and feedback in a spoken language system: A theoretical framework," *Knowledge-Based Systems*, vol. 8, no. 2, pp. 143–151, 1995.

[7] A. Ghorbel, S. Macé, A. Lemaitre, and E. Anquetil, "Interactive competitive breadth-first exploration for sketch interpretation," in *International Conference on Document Analysis and Recognition (ICDAR'2011)*, 2011, pp. 1195–1199.

[8] S. Macé and E. Anquetil, "Eager interpretation of on-line hand-drawn structured documents: The dali methodology," *Pattern Recognition*, vol. 42, no. 12, pp. 3202–3214, 2009.

[9] J. Holsanova, N. Holmberg, and K. Holmqvist, "Reading information graphics: The role of spatial contiguity and dual attentional guidance," *Applied Cognitive Psychology*, vol. 23, no. 9, pp. 1215–1226, 2009.

[10] S. Erhel and E. Jamet, "How can positive effects of pop-up windows on multimedia learning be explained?" *Journal of Educational Multimedia and Hypermedia*, vol. 20, no. 2, pp. 135–156, 2011.

[11] R. A. Abrams and S. E. Christ, "Motion onset captures attention," *Psychological Science*, vol. 14, no. 5, pp. 427–432, 2003.

[12] S. D. Craig, B. Gholson, and D. M. Driscoll, "Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features and redundancy." *Journal of Educational Psychology*, vol. 94, no. 2, p. 428, 2002.

[13] R. Godijn and J. Theeuwes, "Oculomotor capture and inhibition of return: Evidence for an oculomotor suppression account of ior," *Psychological research*, vol. 66, no. 4, pp. 234–246, 2002.

[14] A. P. Hillstrom and Y.-C. Chai, "Factors that guide or disrupt attentive visual processing," *Computers in human behavior*, vol. 22, no. 4, pp. 648–656, 2006.

[15] C. J. Ludwig, A. Ranson, and I. D. Gilchrist, "Oculomotor capture by transient events: A comparison of abrupt onsets, offsets, motion, and flicker," *Journal of Vision*, vol. 8, no. 14, 2008.

[16] D. B. Kaber and M. R. Endsley, "Out-of-the-loop performance problems and the use of intermediate levels of automation for improved control system functioning and safety," *Process Safety Progress*, vol. 16, no. 3, pp. 126–131, 1997.

[17] D. A. Washburn and R. S. Astur, "Nonverbal working memory of humans and monkeys: Rehearsal in the sketchpad?" *Memory & cognition*, vol. 26, no. 2, pp. 277–286, 1998.