Evaluation of Visual, Auditory and Vibro-Tactile Alerts in Supervised Interfaces

Gabrielle Almeida de Souza∗, Laura Amaya Torres∗, Vinicius Stein Dani∗, David Steeven Villa∗, Abel Ticona Larico∗, Anderson Maciel∗ and Luciana Nedel∗
∗ Institute of Informatics (INF)
Federal University of Rio Grande do Sul (UFRGS)
{gasouza, latorres, vsdani, dsvalazar, jatlarico, amaciel, nedel}@inf.ufrgs.br

Abstract—In teleoperation, particularly when controlling systems performing tasks that must be supervised for many hours, human users cannot keep a high level of attention all time. An open issue in the design of such interfaces is to help the user to maintain a situational awareness. In this paper, we compare three types of alert signals aiming to inform users about critical situations that require their full attention. Haptic, visual and auditory signals were assessed as pre-attention mechanisms for tasks in a camera-based supervisory interface scenario. Results show that haptic alerts, for long term supervision, are more effective than auditory and visual ones.

Index Terms—supervisory control systems, situational awareness, multimodal alerts

I. INTRODUCTION

Supervisory control (SVC) systems are used in several different areas of application, such as cockpit interfaces, vigilance systems, process monitoring, and manufacturing operation. These systems allow to monitor and track information from a production process or a physical installation. In the same way, the term “supervisory control” implies that at least one or two persons are continually giving and receiving information from a computer interconnected through effectors and sensors to the controlled process, as defined by Salvendy and Karwowski [30].

Moreover, users’ roles have changed from manually performing activities into the process, to only supervising the whole process or at least a large part of it [25]. It is common for users to get distracted after long periods of usage, due to the information overload continually displayed. It is important that during critical situations – i.e. inadequate application operation – users are alerted in an efficient manner to make them react as soon as possible. As the human attention is the most precious resource in a computational system [8], there are several works facing the problem of obtaining user’s attention by sending them alerts [16], [28]. Additionally, previous reported works compared different alert types such as the work by Hecht and Reiner [10], who studied combinations of auditory, visual and haptic stimuli to analyze sensory dominance phenomenon, concluding the existence of visual dominance when combined with other stimuli, but no dominance when the three stimuli were used together. Ng and Chan [20] investigated the same stimuli in order to obtain information about the human response to these alerts, observing that haptic signals provided the lower response time.

This work aims to study the users’ response time as well, but specifically in supervisory interfaces, evaluating auditory, visual and haptic stimuli. This study helps to determine which of these techniques performs better in terms of response speed to accomplish the objective of capturing the user’s focus of attention in the interface. We also consider the fatigue and the situational awareness of the user when using this type of interface.

Our main contributions with this work are:
• a comparison between three different stimuli
• the development of a haptic device to deliver notifications
• a formal user study with 15 different individuals
• a supervisory interface case study

The remainder of this paper is organized as follows. In the next section, some related works are presented. Then, we introduce the materials and methods used in the conducted user experiments, the achieved results, the conclusions and future work, and finally the acknowledgements.

II. RELATED WORK

There are numerous papers addressing the use of visual, auditory and/or haptic alerts. In 2009, David Hecht and Miriam Reiner [10] studied the sensory dominance phenomenon within combinations of auditory, visual, and haptic stimuli. This study presents a faster response when the signals are used separately because after detecting the presence of the signal(s), participants needed to discriminate these according to their sensory modalities and to choose from different responses the appropriate button or buttons. It is an important result because there were related works showing the opposite, where the combination of different cues improved the reaction time of the user (RT) [5]–[7], [11]. It is often called multisensory enhancement.

In the same work, the authors showed as main conclusions that vision can dominate not only the auditory, but also the haptic sensory modality when combined. However, there is no dominance of the auditory sensory modality over the haptic sensory modality or vice versa. Also, in a tri-sensory combination of audio-visual-haptic signals, participants tend to fail more responding only to two signals (out of three) than they fail to respond to a single stimulus, as the probability of missing two signals is less than the probability of missing just one. In this scenario, there were no bias towards vision
when responding to a single signal, nor when responding to two signals that included the visual element, suggesting that the visual dominance is most likely to be found only in bi-sensory combinations.

On the other hand, Annie W. Y. Ng and Alan H. S. Chan published in 2012 a research related to time response to visual, auditory, and tactile modality stimuli [20]. They concluded that the response time to the tactile stimuli was the shortest, followed by the auditory stimuli and then the visual stimuli.

Vibration, visual, and auditory stimuli were also studied by Adrian Ramos and Domenico Prattichizzo in a surgical robotic context in 2013 [24]. The user, in this case, was a doctor operating and the alerts pretended to make them avoid certain areas that should not be touched. Like the Annie W. Y. Ng and Alan H. S. Chan work, they conclude that vibration is the most effective type of alert, causing the user to retract faster. Also, strong auditory alerts are good but they are unpleasant to the user.

Other works demonstrate an advantage of using haptic feedback [31] [2]. The first study shows that reaction time is shorter with tactile warnings than with visual warnings, suggesting that tactile stimulus produces faster driver responses than visual ones. The second study gives users the capability to perform 1D and 2D task without any visual cues.

As far as the reaction time from different stimuli is concerned, the literature is mature but does not consider the specific case this paper is dealing with. In the supervised interfaces, the user has tasks such as high-level interface perception, task planning and action, and overall system supervision. In this situation, we have the union of the best of the two extremes, the autonomy of the tool and the power of decision of the user throughout various situations [13]. In part of the supervised interfaces, some operations can last a long time, which leaves the operator tired and susceptible to errors. Thus, the interaction in these interfaces should promote flexibility and a more comfortable use of the system [33]. In some cases [9], [35], the user must continuously monitor a large amount of numbers and indicators, and through this information take quick decisions, making it indispensable to be aware of the situation.

Considering all this, an alert system in a supervised interface needs to call the attention of the users to solve the operating problems that are requested, guaranteeing their situational awareness and addressing the issue of user fatigue when using this type of interface.

III. MATERIALS AND METHODS

In our experiment, users have the simple task of monitoring a supervisory interface that displays surveillance videos. There, a set of specific events previously defined occur at different moments (see Figure 1 middle). When an event occurs, the user receives an alert to help them to know that something is happening in the interface. Besides supervising, the user needs to report, using a second interface (see Figure 1-right), the moment when they perceive the received alert and in which camera they found the event. After reporting the event, the user continues waiting for another event to occur.

The purpose of this type of monitoring task was to simulate the workload users can experience during long duration supervision tasks, where their attention can be easily lost due to the monotony of the task itself.

The system registers the experiment’s start time and four performance measures for the user:

- Time to identify the alert: time when the user pressed the button to indicate they perceived the alert
- Time to identify the event: time when the user pressed the button to indicate the identified event
- Number of errors: number of mistakes the user made while identifying the camera number

A. Interfaces

In order to implement the testbed for our experiment, we designed two interfaces. The first one shows 12 surveillance videos (see Figures 1 (middle) and 2) created specifically for the tests and depicting events with an uniform duration of 5 seconds. These videos were created using the free tool provided by Steam – The Source Filmmaker – that allows the creation of movies within the game source engine including some downloaded virtual scenarios and characters. From there, we selected some scenarios and characters to compose the
12 videos of the interface. Regarding the chosen scenarios, everyone maintains a sense of urban environment that users are probably already familiar with.

Fig. 2. Main surveillance interface showing 12 synthetic videos imitating surveillance cameras.

The events that occur in each of the screens are the passage of one or more human characters in the scene (Fig. 3). To provide specific characteristics to the events that the users have to pay attention, we divided them into two types: events with characters of red clothes (Fig. 3 top) and events with characters of blue clothes (Fig. 3 bottom). Thus, all events including characters in red clothes are the ones that should be notified. The lighting conditions of the videos were also different, in such a way that the events are heterogeneous. This interface was displayed in an LG FLATRON W2363D 23” monitor.

Fig. 3. Events’ Characters: character in the scene from camera 1 (left); character appearance (right). Top row shows a red character and bottom row depicts a blue one.

The second interface is a touch pad used for reporting the events. It consisted of a button to indicate the perceived alert, and another 12 buttons representing each camera for the user to indicate where the perceived event occurred. It was displayed in a 9.7” Apple iPad tablet (see Figure 1 right).

The different modalities we use to alert the user that an event is happening are visual, auditory, and haptic (more specifically vibrotactile), all of them with a single intensity.

Visual alerts should display a brief important message to attract the user’s attention without interrupting the task [15]. Taking this into account, the visual alert is a notification delivered directly into the interface window to the top right as in Warnock et al. work [37]. Since it is important to define the visual alert color to represent an emotion to be transmitted to users [21], the chosen color was red to convey the message of needing the user’s immediate attention, as defined in the Windows Dev Center design guidelines [17]. Figure 4 shows the visual alert used in this evaluation.

Fig. 4. Visual alert that appear at the upper right corner of the main interface screen whenever and event occurs in any of the 12 windows.

Some events were preceded by an auditory stimulus that lasted 500ms. The sound were a 1000 Hz beep tone. These characteristics were taken from the US Department of Defense standard, which defines a range of Hz and duration in ms for sound alerts that should be used in military systems. This standard was created, among other reasons, to promote the performance of military systems and equipment operators [4].

B. Vibrotactile wristband

While video monitor and PC speaker are standard peripherals for visual and auditory feedback, a non-conventional device should be designed to provide haptic stimuli.

Several works have studied the vibrotactile stimuli response through the body. Forearm feasibility was compared against torso by Oakley et al. [22]. Besides, an older publication shows a comparison between more than 30 regions [38]. Showing that the hand is the most vibration sensitive region of the body, followed by the sole of the foot, Cholewiak and Collins published a study of sensitivity in the arm [3]. They show that the elbow and wrist are the most sensitive places in the arm.

This hypothesis is supported by further works [14], [36]. Thus, we selected the wrist to design and implement our prototype. Frequency is also an important variable. The Pacinian corpuscles are the mechanoreceptors responsible for perceiving the frequency stimuli. They are embedded in the whole body and they rapidly adapt to intensity changes [27]. Moreover, they are sensitive in the range of 10 to 500 Hz. In the hand, the major sensitivity is reached when the frequency is greater than 100Hz [19] and the highest perception is achieved in 250 Hz [1], [26].

Other authors studied about stimuli duration. Several works recommend stimuli not superior to two seconds [18], [32], [34]). A vibration of maximum 200 ms is recommended by Kaaresoja and Linjama for short vibration stimulus in pocket placed devices [12]. The users in the experiments mentioned
that a stimulus longer than 200 ms is “strong” or “too strong”. The values of 200 ms and 600 ms are recommended by Saket et al. [29] for short and long vibration, respectively.

We based our prototype on the approach presented by Oliveira et al. [23] to convey navigational information using vibrotactile actuators. We placed eight tactile actuators (ROB-08449) equally distributed through a velcro strap. This allows us to adapt our wristband to the wrist of each user. The operating voltage of actuators is within 2.5V and 3.8V and their amplitude vibration is 0.8G. The diameter of each actuator is 3.4mm. We used an Arduino Uno to control the tactors by means of individual drivers. Each driver works as an interface between the digital outputs of the Arduino board and the actuator, letting us reduce the size of the setup and improving the control over the array of vibrators (see Figures 1 left and 5).

![Fig. 5. Experimental setup: haptic wristband (left); driver box and haptic wristband (right)](image)

C. Experimental Protocol

Every experiment started with a tutorial to explain to the users the kind of events they could find in the surveillance videos and which were the targeted ones (events where characters in red clothing appear). In this phase, the users could also make themselves familiar with the different types of alerts.

After that, the test starts presenting a total of 16 events distributed along 32 minutes. The type of events’ alert distribution corresponded to four auditory, four visual, four vibrotactile, and another four events without any alert. Alerts were given in a random order that was predefined to be the same for every user. Subjects were allowed to use their cellphone or any other type of distracting device during the test.

Fifteen persons took part in the experiment (three female), most of them with advanced informatics knowledge. Half of the users had minor vision problems. The average age of the participants was 25 years old.

At the end of the experiment, users were asked to fill a questionnaire to express their qualitative evaluation about the experience and the alerts they received.

IV. RESULTS AND DISCUSSION

A. Quantitative analysis

In this section, we present an analysis of the data collected to evaluate our hypotheses.

**H1: The user is going to need the alerts to accomplish the supervision task without missing an event.**

<table>
<thead>
<tr>
<th>Identified</th>
<th>without alert</th>
<th>with alert</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>49</td>
<td>81.7%</td>
<td>160</td>
<td>209</td>
</tr>
<tr>
<td>Missed</td>
<td>11</td>
<td>20</td>
<td>31</td>
</tr>
</tbody>
</table>

The users’ responses were evaluated in two situations:

- **Event without alert**: events sent without any stimulus to see the reaction time of users identifying the event without the alerts help.
- **Event with alert**: events accompanied by any type of alert (visual, auditory, haptic) to help the user to identify the event.

To validate H1 we compared the two situations and obtained the results shown in Table I, in which 81.7% and 88.9% are the percentage of identified events without and with alert respectively. We see that with the alerts, the user performance is improved by 7.2 points. However, our hypothesis cannot be proved because the users are still missing events even with the help of alerts.

**H2: The longer a user supervises the interface, the more they are going to need the alerts to keep track of what is happening.**

We analyzed the alerts both in time response and accuracy through time to look for a correlation between them. First, we performed a Pearson product-moment correlation obtaining a correlation coefficient of -0.07410773 (Figure 6). The value is very near zero, suggesting that there is not a direct correlation in the data. Next, we plotted a graphic showing the data’s behavior confirming there is no relation between duration time and time response. Then, people do not benefit more from alerts as time goes on. Although we would have to account for the learning curve here, and the test took only 32 minutes, the raw data collected suggest that the need for alerts decreases with time.

![Fig. 6. Correlation Analysis](image)
Then, we also analyzed the behavior by alert type (Figure 7), showing a similar path. As there is not a relevant correlation, we could not prove H2.

**H3: The use of vibrotactile (haptic) alerts will lead to a smaller reaction time compared with the other types of stimuli.**

To test this hypothesis, performed a post-hoc analysis, separately confronting: (1) vibrotactile vs auditory alert (Figure 8), and (2) vibrotactile vs visual alert (Figure 9). In the first case, we can observe a non-normal distribution through the Shapiro-Wilk normality test, whose resulting p-value is \( p = 2.826e-07 \) (i.e., \( p \)-value \( < 0.05 \)). Similarly, in the second case the resulting p-value \( p = 2.212e-06 \). Considering non-normal samples, we used the non-parametric Mann-Whitney U test to identify if there were statistically significant differences between the two groups. As a result, in the first case the vibrotactile alert reaction time was significantly smaller that the auditory one (\( p = 0.005164 \)). For the second case, significantly smaller reaction times were also found for the vibrotactile stimuli in comparison with the visual alert (\( p = 0.03013 \)). Such results allow us to prove H3 and conclude that haptic alerts lead to faster reactions.

**B. Qualitative analysis**

In the post-test questionnaire, users were asked which type of alert they preferred and which alert they disliked, as well as which alert they found most effective and which one they found less effective to draw their attention to the system.

The visual alert was the least preferred by the participants. In some comments made by users in the questionnaire, they mentioned that the visual alert disturbed the visualization of the event in camera 4, which may be one reason for its low qualification. In addition, the visual alert obtained lower classification regarding its effectiveness in drawing the attention of the user. This may be due to the loss of focus on the screen. In some comments, it was also mentioned that, in the real world, where the focus is probably more easily lost on supervisory systems, the visual alert may not be as effective as the others. In contrast, the vibratory alert was the best rated by users in terms of both preference and effectiveness. The auditory alert obtained a medium score.

When asked if participants, at some point in the experiment, lost focus on the system, 73% said yes and 27% said no. In one comment, one participant emphasized that, because it was a short experiment, he kept an eye on the interface all the time, but that in the real world, in a long time using this type of system, the alerts would have more effect.

Regarding the other comments, only one participant mentioned that the auditory alert was a very acute and irritating sound. About the vibratory alert, one participant said that it would be better to use a stronger vibration to draw more attention. Finally, some participants said that, even if the alert caught their attention, they still did not know in which camera the event was happening.
V. CONCLUSIONS AND FUTURE WORK

In this work, a visual, auditory, and vibratory alert evaluation was performed within the framework of supervisory interfaces.

In general, the haptic alert was pointing qualitatively as the most preferred and most effective in the subject’s opinion. Quantitatively, the haptic alert has also shown a significantly lower reaction time. The other alerts did not have significant results regarding their effectiveness in reaction time. However, the visual alert was the least preferred, and was evaluated as the least effective by the subjects.

A limitation of our experiment, however, was the time that the subjects were submitted to the test. While the target applications require the user attention for many uninterrupted hours, our experiment took no more than 30 minutes. We tried to make the test not too long and uncomfortable for the volunteers, but still being boring and tiresome, inducing in some way the user to distract himself. Results pointed out future improvements that should be made during the next studies with a longer duration.

For future work, we suggest the evaluation should be redone adding natural distractions to the subjects (multitasking), which will increase the loss of focus in the monitoring system without having to increase the experiment time. We also noticed that some videos were too dark and thus very difficult for the users to distinguish colors. This eventually led to not all events being detected. In future tests, the system videos must be adjusted on brightness and contrast to avoid this effect. Other types of remote, uncontrolled tests, with duration of at least 8 hours, will be considered to measure the loss of focus and fatigue levels of the subjects.

ACKNOWLEDGMENT

This study was financed by Petrobras (Project Annelida) and partially by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001. We also acknowledge FAPERGS (project 17/2551-0001192-9) and CNPq-Brazil (project 311353/2017-7).

REFERENCES


