Identifying and Overcoming Attention Limitations in the Detection and Identification of Alarms in Close Temporal Proximity

by

Yuzhi Wan

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Industrial and Operations Engineering)
in the University of Michigan
2019

Doctoral Committee:

Professor Nadine Sarter, Chair
Associate Professor Eytan Adar
Professor Yili Liu
Associate Professor Bernard Martin
Assistant Professor Xi Jessie Yang

"Two things fill the mind with ever new and increasing admiration and awe,
the more often and steadily we reflect upon them:
the starry heavens above me and the moral law within me."

Kant, I. (1788). *The Critique of Practical Reason.*

Yuzhi Wan

wanyz@umich.edu

ORCID iD: 0000-0002-9327-1234

DEDICATION

To mom and dad

ACKNOWLEDGEMENTS

I would like to express my thanks to many people, without whom this dissertation would not have been possible. First and foremost, my deepest gratitude goes to my advisor, Dr. Nadine Sarter, for the constant support, guidance, care, and encouragement over the last five years. She is a researcher who pays attention to details, a supervisor who does not compromise on academic standards, and a colleague who always offers constructive feedback. Thank you for leading me in this journey of exploration, for helping me grow into an independent researcher and a better person, and most importantly, for all the trust and patience you have in me.

I would also like to thank the members of my dissertation committee: Dr. Eytan Adar, Dr. Bernard Martin, Dr. Yili Liu, and Dr. Xi Jessie Yang. I am grateful for the wisdom and knowledge you shared in class, as well as the feedback and support you offered when I needed advice. I have always been inspired by your keenness for knowledge and kindness to students.

I was very fortunate to have been financially supported by the University of Michigan Department of Industrial and Operations Engineering, the Federal Aviation Administration, the Rackham Graduate School, and the Toyota Research Institute. Their support enabled me to conduct my research and share my findings.

A lot of thanks are due to the amazing staff in the Department of Industrial and Operations Engineering. Charles Woolley, Eyvind Claxton, Christopher Konrad, and Olof "Mint" Minto provided me with incredible technical support for my research and for our lab. Robyn Bollman, Rod Capps, Rebekah Smith, Matt Irelan, and Candy Ellis also provided gracious assistance with various things that made my life in IOE easier and more enjoyable. Special thanks goes to Tina Picano Sroka, for always being willing to share her smile, love, and confectionery, to Teresa Maldonado, for patiently helping with so many things that I cannot even begin to count and the great events you organized for the Center for Ergonomics, and to Sheryl Ulin, for your generous help and advice, and the great summer weekend we had with your family.

Thanks to the entire THInC lab family, my Ph.D. life couldn't have happened in a better place. I am so fortunate to have the greatest fellow lab mates – Nadine Moacdieh, Brandon Pitts, Julie Prinet, Yidu Lu, Kevin Lieberman, and Karanvir Panesar – who were not only the best source for support and inspiration, but also lifetime friends with whom I can share happiness, excitement, frustration, and doubt. I can never forget the moments of finding surprise gifts on my desk, going to conferences together, or having lunch together by the pool in summer. I would also like to thank Kejia Xu, Aditya Mannari, Derek Witcpalek, and Grace Miller for their support with simulation development and data analysis. I am grateful that you always made things happen with all the time constraints. Thanks are also due to all the participants who volunteered in my experiments.

I have always appreciated the company of my fellow students in IOE: Rosemarie Figueroa, Vernnaliz Carrasquillo, Yadrianna Acosta-Sojo, Qianyi Fu, Justin Haney, Sol Ie Lim, Ke Liu, and Na Du. Thank you all for being my wonderful peers. I am also thankful to many other friends I met in Michigan, who brought joy and comfort into the long snowy winters here. Special thanks go to Yunxiang Xu, Fengmin Hu, Shuqi Cheng, Changsheng Yao, and Da Li, for sharing your experiences and being supportive all the way.

Many heartfelt thanks to Yu Shi, Chang Gao, Tian Ding, Yeshuyi Qiu, Yuwei Guan, Ziqiang Gao, and Tao Chen – friends who are never far from me regardless of the difference in time and the distance in space.

Last but not least, to my family, thank you for the never-ending love, care, and support. To my maternal grandparents, who were both well-respected teachers, you taught me love and integrity. I hope you can be proud of me as I am of you. To my cousin Yayuan Sun, thank you for your care and support, and especially the time we shared in Ann Arbor. To my parents, thank you for comforting and trusting me through ups and downs. You are my fortress and harbor. All the honors I earned and will earn go to both of you. Xue, I am so lucky to have met you here. I cannot be more grateful for your love and courage to face everything in life with me. My words can barely begin to describe even a fraction of that. I look forward to exploring this world with you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ABSTRACT

In many high-risk domains, such as aviation, driving, and space operations, safety depends greatly on the timely detection and correct identification of alarms. However, due to high levels of system complexity and coupling in those environments, a large number of alarms is sometimes triggered within a short period of time, a so-called alarm flood. Alarm floods pose a challenge for operators at various stages of information processing, including detection, identification and response selection. The detection of alarms during an alarm flood can be difficult due to masking effects. Masking occurs when one signal is obscured by the presence of another simultaneous or asynchronous stimulus. To date, various forms of simultaneous and asynchronous masking, such as change blindness and attentional blink, have been studied almost exclusively for only two stimuli and in single-task conditions. The effects of masking in the presence of a large number of signals and in multi-task environments are not well understood.

Therefore, the goals of this dissertation were to (1) establish the relative contributions of simultaneous and asynchronous masking to missed and misdiagnosed visual and auditory alarms during routine operations and in an alarm flood, (2) identify the stimulus onset asynchronies (SOAs) at which masking effects are observed in demanding multitask environments, (3) investigate the impact of the number and temporal distribution of alarms on alarm detection and identification and (4) develop and test countermeasures to overcome observed performance breakdowns. To this end, a series of four experiments were conducted using a simulation of a drone-based package delivery service.

Both simultaneous and asynchronous masking were observed, intramodally and crossmodally, for visual and auditory alarms. The SOA at which asynchronous masking was observed was longer (around 800ms) than reported in basic studies of the phenomenon. The effects of both simultaneous and asynchronous masking were stronger with an increase in the number of alarms; this increase was more pronounced for simultaneous masking. Simultaneous masking also had a stronger effect on auditory alarms, compared to visual alarms. In clusters of 4 or 6 sequential alarms, the first and last alarms were more likely to be detected and correctly identified. During alarm floods, a speed-accuracy tradeoff was observed, i.e., the response time to alarms was shorter but the identification accuracy was lower, compared to routine operations. Finally, when the criticality of alarms was mapped to preattentive features of the alarm signal (color and pitch), the masking effects were alleviated, especially for clusters of 4 or 6 concurrent visual alarms.

The findings from this dissertation contribute to a better understanding of the effects of two types of masking on the detection and identification of large numbers of critical signals in high workload multi-task environments. They help inform the development and expansion of models and theories of human perception and cognition and also highlight that results from basic research do not necessarily generalize to applied settings. From an applied perspective, the findings will provide guidance for the design and evaluation of alarms and alarm systems. Ultimately, the present research helps prevent catastrophic outcomes due to missed or misinterpreted alarms, and thus improves safety in many real-world environments.

Chapter 1

Introduction

In many complex data-rich domains, such as aviation and process control, system safety depends greatly on the timely detection and correct identification of alarms. However, accomplishing these tasks can be difficult due to the coupling and complexity of systems in these environments which can result in large numbers of alarms within a short period of time – a problem called an alarm flood (Perrow, 2011). Alarm floods have played a role in accidents such as the nuclear meltdown at the Three Mile Island Nuclear Generating Station (March 28, 1979, Dauphin County, Pennsylvania), the explosion and fire at a Texaco Refinery (July 24, 1994, Milford Haven, Wales), and the crash of Air France Flight 447 (June 1, 2009, Atlantic Ocean). An alarm flood is defined as more than 10 alarms in a 10-minute period, but this rate is often exceeded in real-world settings (EEMUA, 1999). For example, in the medical domain, there can be as many as 120 separate alarm devices in one operating room, many of which may sound at the same time (Sykes et al., 2011). In the Texaco Refinery explosion, the two operators had to respond to 275 alarms during the last 10.7 minutes before the explosion, or approximately one alarm every two seconds. As a result, the operators missed critical alarms that could have helped them diagnose and handle the problem (HSE, 1997). In the domain of cyber defense, the situation is even worse with analysts facing thousands of alarms per day (Ben-Asher & Gonzalez, 2015).

Perceptual and attentional limitations contribute to operators missing signals during an alarm flood, including change blindness, attentional blink and, more generally, masking. Masking is the process by which one stimulus is obscured by the presence of another stimulus (Enns, Di Lollo V, & Di Lollo, 2000). Masking can occur when two stimuli are spatially or temporally overlapping or contiguous (Breitmeyer & Öğmen, 2006). One example of temporal masking is change blindness which refers to the failure to detect even salient changes and events when these changes coincide with a transient or brief disruption in visual continuity (Lu, 2014; Rensink, 2002; Simons & Levin, 1997). Another form of temporal masking occurs with attentional blinks which represent an example of asynchronous masking. Attentional blinks involve the failure to detect the second of two target stimuli if both are presented in close temporal proximity (200-600ms apart, according to the basis literature in attention).

The role and relative contributions of these masking effects to problems with alarm detection in real-world environments are not well understood. For one, almost all research on attentional blink and change blindness to date has focused exclusively on the detection and identification of just two stimuli. The vast majority of studies have examined the two phenomena in the visual modality only. And participants in these studies were most often responsible only for the one task of detecting target stimuli. In contrast, alarm floods in real-world domains involve exceedingly large numbers of signals which are presented in both the visual and the auditory channels, and operators in these domains tend to timeshare several tasks. During alarm floods, it is also important to consider the possibility of a serial position effect where the likelihood of detecting an alarm may depend on the nature and timing of surrounding alarms. The proposed research aims to address the above limitations in the literature to gain a better understanding of the nature of, and reasons for breakdowns in alarm detection. It will also explore candidate techniques for overcoming the problem.

One possible way of supporting operators during an alarm flood could be to increase bandwidth through the introduction of multimodal displays, i.e., interfaces that distribute information – in this case, alarms - across multiple sensory channels, including vision, audition and touch. These channels are partially independent of each other, and thus multimodal displays have been used to offload an overburdened modality and were shown to generate faster responses (Lu et al., 2013; Wickens & McCarley, 2007). However, the effectiveness of this approach depends on whether and to what extent the aforementioned masking phenomena affect not only intra-modal, but also cross-modal information processing. Specifically, little is known about the occurrence of attentional blinks and change blindness across non-visual modalities (e.g. Chun & Potter, 2001).

Another way to improve alarm detection is to exploit top-down influences on attention capture and management. Specifically, alarm displays can be designed to support prioritization of alarm signals, so that the operators can choose to allocate more of their limited attentional resources to the detection and identification of the most critical alarms. One way to facilitate this prioritization is to map pre-attentive features of alarms, such as the color of visual alarms and the pitch of auditory alarms, to their urgency or criticality. This approach has been shown to improve the response time to alarms but very little is known about whether this method improves detection performance also (Nikolic, Orr, & Sarter, 2004; Politis, Brewster, & Pollick, 2014; Wickens, McCarley, & Steelman, 2009).

In summary, the overall goal of the reported research is to gain a better understanding of the nature of, reasons for, and possible countermeasures to breakdowns in alarm detection when operators are faced with large numbers of alarms over longer periods of time or during an alarm flood. Specifically, its objectives are:

- To establish the relative contributions of simultaneous and asynchronous masking to missed and misdiagnosed visual and auditory alarms during routine operations and in an alarm flood
- To identify the stimulus onset asynchronies (SOAs) at which masking effects are observed in demanding multitask environments
- To investigate the impact of the number and temporal distribution of alarms on alarm detection and identification
- To develop and test countermeasures to overcome observed performance breakdowns

To this end, a series of four experiments were conducted using an unmanned aerial vehicle (UAV) control simulation. The tasks were designed to simulate those of a commercial drone-based package delivery system. Such a system will likely require operators to serve as supervisory controllers of multiple UAVs, each of which may experience problems and require operator interventions. The findings from the proposed work can be generalized to other similarly data-rich and attention-demanding application domains, such as cyber-defense, air traffic control, and industrial process control.

Alarm displays and alarm floods

In highly automated sociotechnical systems, one of the main responsibilities for operators is to monitor system health. Alarms are a means of assisting operators in this task by alerting them, in a data-driven fashion, to anomalies in the system. Alarm classification systems distinguish between alarms, alerts, and notifications, which differ with respect to the potential results of a failure or delay in responding to the off-nominal event/condition (Sessa, 2013). Alarms tend to be associated with more severe consequences and require an immediate response while alerts do not (e.g., Grosdidier, Connor, Hollifield, & Kulkarni, 2003). Notifications present information that is even less safety-critical than that conveyed by alerts. Since this distinction is not critical for most of the reported research, we will use the word "alarm" to refer to all three types of indications for the remainder of this document, except for the last experiment, where the effect of alarm criticality will be discussed in detail.

Alarm detection and identification pose a challenge to operators because automated systems have become increasingly complex and coupled. As a result, a single anomaly can trigger a series of related alarms in a very short period of time and thus turn into an alarm flood (Perrow, 2011). Industrial standards define an alarm flood as the presentation of more than 10 alarms in a 10-minute period (EEMUA, 1999). Such events have been shown to exceed the processing and response limits of human operators (e.g., Reising, Downs, & Bayn, 2004). Alarm floods were identified as a major contributor to disasters such as the Three Mile Island Nuclear Power Plant accident (1979) and the Texaco refinery explosion (1996). They can be expected to increase in frequency and severity due to the continued growth in digitalization and system complexity (Zwaga & Hoonhout, 1994).

A number of approaches have been explored to reduce the risks associated with alarm floods. Alarm rationalization is a data-mining method that detects statistical similarities among discrete occurrences of alarms, groups correlated alarms based on their degree of similarity and suppresses unnecessary alarms (Noda, Higuchi, Takai, & Nishitani, 2011). Alarm flood pattern recognition is another data-mining technique that tries to obtain common patterns from multiple alarm sequences by finding the optimal alignment between them, using criteria-based search strategies (Vogel-Heuser, Schütz, & Folmer, 2015). Alarm causal analysis uses a combination of alarm log, process data, and connectivity analysis based on transfer entropy (the estimated amount of information transferred from one signal to another) to identify consequence alarms that originate from a shared anomaly and provide a causal inference suggestion (Rodrigo, Chioua, Hagglund, & Hollender, 2016). Alarm summary displays use alternative alarm presentations such as 3D visualization or timeline-based alarm trackers to combine the benefits of list-based displays with time series presentation of alarm information (Laberge, Bullemer, Tolsma, & Reising, 2014). Other approaches to the problem of alarm floods include suppressing nuisance alarms, smart removal of redundant alarms, alarm compression, and alarm screening (Costa, Cachulo, & Cortez, 2009; Kirschen & Wollenberg, 1992; Liu et al., 2003; Mora, Passariello, Carrault, & Le Pichon, 1993; Wollenberg, 1986). The aforementioned alarm configuration practices have been shown to effectively reduce the number of alarms and/or improve detection and response performance but they also share a limitation: by reducing the number of alarms presented to the operator, they risk hiding useful information from the operator if critical messages are misidentified as nuisance or unimportant alarms.

To avoid this problem, and as a possible complement to the above approaches, the present study aims to investigate the alarm flood problem from a different perspective. Its goal is to improve our understanding of human perceptual and attentional phenomena in the face of large numbers of stimuli and use this knowledge to improve alarm presentation and thus support reliable alarm detection and identification.

Alarm detection

Stanton & Stammer (1998) described seven stages involved in processing a visual alarm: detection, discrimination, identification, classification, recognition, scaling and ordering & sequencing (see **Table 1.1**).

Table 1.1. Psychological processes and implications for the design of visual alarms (Stanton & Stammer, 1998)

Psychological process	**Purpose of the process**
Detection	Determining the presence of an alarm
Discrimination	Defining the differences between one alarm and another
Identification	Attributing a name or meaning to an alarm
Classification	Grouping the alarms with a similar purpose or function
Recognition	Knowing what an alarm purports to mean
Scaling	Assigning values to alarms
Ordering & Sequencing	Determining the relative order and priority of alarms

Generally, an alarm system should be designed to support all seven stages. However, in the context of an alarm flood where the sheer number of alarms is likely to exceed the processing limits of the operators, the main concern is preventing failures at the early stages of *detection, discrimination,* and *identification* of critical alarms. In the remainder of this dissertation, two types of performance will be used to describe these early stages of alarm processing. Detection performance is concerned with whether any alarm has been detected and correspond to the first stage, alarm detection. Accuracy is concerned with whether a detected alarm is correctly identified, and thus relies on both the stages of discrimination and identification.

Success at these early stages requires sufficient availability and proper management of attentional resources (i.e., the pool or capacity of mental effort for attention; see Wickens, 2008). The allocation of visual attentional resources has been described by the NT-SEEV model, which was developed to predict the noticeability (N) of, and response time (T) to events (such as alarms), based on visual scan models for supervisory control and visual sampling (Steelman, McCarley, & Wickens, 2011; Wickens, Goh, Helleberg, Horrey, & Talleur, 2003; Wickens et al., 2009; Wickens & McCarley, 2007). The model includes two bottom-up components: the salience (S) of the alarm (e.g., its brightness or hue contrasted against the background), and the effort (E) required to shift attention to the alarm (e.g., the amount of head movement). The other two components of the model are top-down factors: the operator's expectancy (E) of the alarm (for the frequency of information appearing at a certain position) and the assigned value (V) of the alarm. The NT-SEEV model predicts that alarms are more likely to be noticed when 1) they have stronger contrast against the background, 2) they are expected to appear more frequently, 3) they are presented within the field of view and thus require less effort to notice, and 4) they are regarded as more valuable, i.e., important to the task at hand. The computational predictions of the NT-SEEV model were validated with empirical data (eye-tracking and performance data) from pilots in a high fidelity flight simulator (Wickens et al., 2009), showing a high correlation between the empirical results and model predictions. A number of other models have been proposed to describe the processes involved in detection/noticing performance (e.g. Itti, Koch, & Niebur, 1998; Le Meur, Le Callet, Barba, & Thoreau, 2006; Posner, Snyder, & Davidson, 1980; Stanton & Baber, 1995; Stanton, Booth, Stammers, & Stammers, 1992) but the NT-SEEV model is one of the most continually updated models that comprehensively incorporates both top-down and bottom-up factors influencing visual attention.

Masking and alarm detection during an alarm flood

While the NT-SEEV model provides a comprehensive account of the processes involved in the detection of a single event, its application to the alarm flood problem is limited by some factors that it did not cover. Firstly, the NT-SEEV model has focused primarily on the detection of visual events, while a large number of alarms are presented in the auditory channel. Auditory alarms are widely used because they tend to be salient and omnidirectional (i.e., not requiring a specific spatial body/head/eye orientation. More importantly, the use of auditory signals can offload the already-overburdened visual channel by presenting information in a different sensory modality and thus reducing the competition for visual attentional resources (Wickens & McCarley, 2007).

The NT-SEEV modal also does not explain problems experienced during an alarm flood which represents a much more complex multi-signal situation (Steelman et al., 2011). Importantly, during alarm floods, detection performance may be affected by masking among alarm signals (i.e., the process by which one stimulus is obscured by the presence of another stimulus). This was highlighted by a survey of operators from a chemical manufacturer, a confectionery company and a nuclear power plant which identified masking as the most frequently reported reason (nearly 50% of all cases) for missing an alarm (Stanton, 1993). Similarly, a 1995 analysis of the Aviation Safety Reporting System database identified missed alarms as a major contributor to aviation incidents (Bliss, Freeland, & Millard, 1999). Masking can happen when alarm signals appear either simultaneously or in close temporal proximity. These two types of masking effects are called simultaneous masking and asynchronous masking, respectively, and will be further discussed in the following sections. During alarm floods, it is likely that operators experience both types of masking due to simultaneous and closely spaced alarms.

Similar to visual alarms, the detection of auditory alarms can also be influenced by masking effects. The detection of auditory signals has been shown to be highly susceptible to simultaneous masking. For example, Bolton and colleagues have shown that nurses miss auditory alarms even when these alarms comply with industrial standards, a problem that has contributed to medical accidents (Bolton, Edworthy, & Boyd, 2018). Auditory simultaneous masking is especially strong when two or more auditory signals appear in the same spatial location (Carhart, Tillman, & Greetis, 1969; Doll, Hanna, & Russotti, 1992; Fastl & Zwicker, 2006). These masking effects, unlike the other examples of masking as a result of attention or memory bottlenecks, are related to sensory receptors in the ear. These receptors are sensitive to the frequency of vibrations on the tympanic membrane and not capable of differentiating sounds that are highly similar in their base frequency (Bolton et al., 2018).

Depending on task context, masking can happen both at the sensory stage and at later stages of information processing. It is therefore not always possible to identify, with certainty, the underlying mechanisms. However, the behavioral effects of these processes are the same: signals are more likely to be missed when presented in close temporal proximity. The following sections will give some examples of simultaneous and asynchronous masking effects and how they relate to attention limitations.

Simultaneous masking: change blindness and enumeration limits

Change blindness refers to an observer’s inability to notice changes in a visual scene when these changes coincide with a visual transient (Simons & Levin, 1997). This attention deficit is striking due to the fact that even large, salient, and expected changes can go unnoticed. Change blindness has been identified as one reason for operators missing critical events (e.g., Curry, Meyer, & McKinney, 2006; Lee, Lee, & Boyle, 2007; Velichkovsky et al., 2002). Research to date suggests that change blindness is caused by the shortage of both visual short-term memory and attentional resources (Simons & Rensink, 2005). The two most common paradigms used to trigger change blindness are “flicker” and “mudsplash”. In a “flicker” paradigm, the original and the changed scene are separated by a masking stimulus that completely blocks the view, e.g., a blank screen. In a “mudsplash” paradigm (see ***Figure 1.1***), a few small, high contrast shapes are briefly spattered over the view, without covering the change (Lu, 2014; O’Regan, Rensink, & Clark, 1999; Simons & Rensink, 2005).

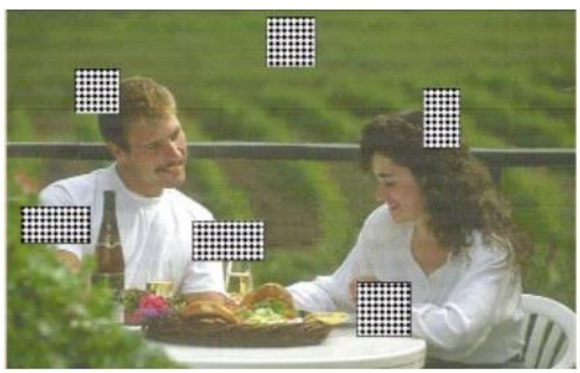

Figure 1.1 The “mudsplash” paradigm (O’Regan et al., 1999)

Change blindness was first and primarily studied in the visual modality but a small number of more recent studies demonstrated the existence of the phenomenon in the auditory and tactile modalities also. For example, change deafness was triggered using auditory masking. When the onset of white noise (comparable to the “mudsplash” in visual stimuli) or silence (comparable to a “flicker” or blank white screen in visual stimuli) coincided with the addition or deletion of an auditory object to/from a complex auditory scene, the participants were prone to miss the change (Pavani & Turatto, 2008). Tactile change blindness can be triggered using a “flicker” paradigm, where a tactile mask is presented between the two tactile signals to be discriminated, or a “mudsplash” paradigm, where a tactile mask is presented concurrently with the change in a continuous tactile signal (Gallace, Tan, & Spence, 2007; Lu, 2014). Finally, crossmodal change blindness has been observed for visual-tactile modality pairs. Signals were presented either in the same modality (visual or tactile) or in different modalities (one visual and the other tactile). The “mudsplash” masks were either tactile or visual. Change blindness was observed for all stimulus-mask combinations (Auvray, Gallace, Tan, & Spence, 2007).

Most empirical work and theory development on change blindness has focused on breakdowns in the detection of one or two targets in various modalities and modality combinations. However, research on enumeration (i.e., reporting the number of targets) shows that participants’ performance drops considerably when more than two targets are present. For example, enumeration studies have shown that, when the number of visual targets exceeds four, performance decreases dramatically (Atkinson, Campbell, & Francis, 1976; Kaufman, Lord, Reese, & Volkmann, 1949; Sathian et al., 1999). The enumeration mechanism was categorized into subitizing, which can enumerate numbers no greater than four by rapidly recognizing learned patterns, and counting, which is a slower but more accurate conscious enumeration process that is dominant when the number exceeded four (Mandler & Shebo, 1982; Revkin, Piazza, Izard, Cohen, & Dehaene, 2008). A similar limitation has been reported in the auditory modality where performance starts to deteriorate rapidly with more than two concurrent auditory stimuli (e.g. Camos & Tillmann, 2008; Kashino & Hirahara, 1996; ten Hoopen & Vos, 1979). These findings are highly relevant for alarm floods because during periods of high alarm frequency, such as alarm floods, the number of alarms that are concurrent or in temporal proximity is very likely to exceed two. Therefore, the ability to detect and identify multiple alarms would be important for system safety.

Asynchronous masking: attentional blink

Masking effects are also observed when two stimuli are asynchronous and non-overlapping. For example, attentional blink (AB) refers to the failure to accurately report the second of two target stimuli (T2) when both target stimuli (T1&T2) are presented in close temporal proximity (Dell'Acqua, Jolicœur, Sessa, & Turatto, 2006; Raymond, Shapiro, & Arnell, 1992; Soto-Faraco et al., 2002). Attentional blink was first studied, and has been examined most thoroughly in the visual modality. In laboratory studies, the phenomenon is often triggered using the paradigm called rapid serial visual presentation (RSVP, see ***Figure 1.2***) where a number of target stimuli (such as letters digits, words, or pictures) are presented sequentially, each for a brief period of time (from 6 to about 20 items per second; typically 10 items per second) (Broadbent & Broadbent, 1987; Raymond et al., 1992). The task for participants in these studies is to identify and report those stimuli (targets T1 and T2) that are different from other distractors (usually the targets are letters while distractors are numbers, or vice versa). In these experiments, an attentional blink occurred with an SOA (stimuli onset asynchrony) of 200-600ms, i.e., when T2 is displayed 200-600ms after T1 (Arnell & Jolicœur, 1999; Brehaut, Enns, & Di Lollo, 1999; Dell'Acqua et al., 2006; Dux & Marois, 2009; Raymond et al., 1992; Ward, Duncan, & Shapiro, 1996). T2 is more likely to be missed if participants are required to report T1, compared to when they can ignore the first target.

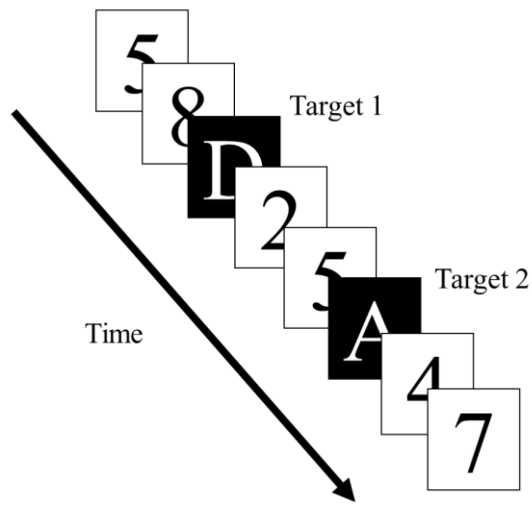

Figure 1.2 Rapid serial visual presentation (RSVP)

In early studies of visual attentional blink, both T1 and T2 were followed by distractors. T1 and T2 masking play different roles in bringing about an attentional blink: T1 masking introduces a delay in processing T2, while T2 masking interrupts the processing of T2 (Brehaut et al., 1999). However, more recent studies have demonstrated that an attentional blink occurs even when just a blank interval separates T1 and T2 (Nieuwenstein, Potter, & Theeuwes, 2009). The detection failure in these studies was explained by rapidly disengaging and re-engaging attention to the target, instead of the disruption being created by distractors.

Recent studies have examined whether the attentional blink is a modality-specific or a supra-modal phenomenon (Arnell & Jolicœur, 1999; Potter, Chun, Banks, & Muckenhoupt, 1998). Experiments on auditory attentional blink have used a method called rapid auditory presentation (RAP) which requires the participant to search for targets in a fast auditory stream of items (Arnell & Jolicœur, 1999). The findings from these studies are not consistent: some researchers succeeded in triggering an auditory attentional blink (e.g. Arnell & Jolicœur, 1999;

Duncan, Martens, & Ward, 1997; Goddard, Isaak, & Slawinski, 1997; Shen & Mondor, 2006), but some did not (e.g. Potter et al., 1998). The reason for these apparently contradictory findings is unclear.

Only two studies have examined tactile attentional blinks (Dell'Acqua et al., 2006; Hillstrom, Shapiro, & Spence, 2002). In both studies, serially presented streams of vibration stimuli were presented to the fingertips, similar to the presentation of visual and auditory targets in the RSVP and RAP paradigms. Both experiments confirmed the existence of tactile attentional blinks by demonstrating a decrease in identification accuracy when the two targets were closer than 360ms in time.

Based on the above findings, it appears that attentional blinks occur intramodally in all three sensory channels – vision, audition, and touch. As with change blindness, one unanswered question is whether attentional blink is observed crossmodally as well. Studies addressing this question have not yielded consistent results. Some studies observed visual-auditory attentional blinks (Arnell, 2006; Arnell & Jolicœur, 1999; Arnell & Larson, 2002; Van der Burg, Brederoo, Nieuwenstein, Theeuwes, & Olivers, 2010; Van Der Burg, Nieuwenstein, Theeuwes, & Olivers, 2013), visual-tactile attentional blinks (Dell'Acqua, Turatto, & Jolicoeur, 2001; Soto-Faraco et al., 2002), as well as auditory-tactile attentional blinks (Dell'Acqua et al., 2001). Other experiments failed to produce the phenomenon. For example, Soto-Faraco and Spence (2002) observed no attentional blinks in their study while Martens, Kandula, & Duncan (2010) confirmed the presence of intramodal visual and auditory attentional blinks, but did not find crossmodal attentional blinks (see also Duncan et al., 1997; Hein, Parr, & Duncan, 2006).

Similar to studies on simultaneous masking, most research on asynchronous masking focused on the interference between only two stimuli. Very little is known about asynchronous masking between multiple stimuli in temporal proximity, the situation encountered during an alarm flood. One study examined how attentional blink contributes to the detection of multiple consecutive targets within close temporal proximity, using a simulated sonar display monitoring task (Boot, Becic, & Kramer, 2007). Participants were required to report the number of new targets appearing on the display. Detection performance decreased with an increase in the number of visual targets, and this performance decrement was even worse than predicted based on the attentional blink literature. The authors also reported that, in contrast to earlier work, the number of targets had a greater effect on detection performance than the SOA between targets.

Research gaps

While the above literature review suggests that difficulties with detecting alarms during an alarm flood result, in part, from masking effects such as change blindness and attentional blink, it is important to highlight aspects of earlier studies that may limit their applicability to complex dada-rich domains. Most change blindness and attentional blink experiments employed a single-task paradigm, involved low levels of workload, and included relatively simple stimuli and displays. Also, most of the studies to date involved search tasks whereas alarm detection is a noticing task. A search task involves a predefined target or some change that is expected and deliberately, in a top-down fashion, sought by the participant. In a noticing task, such as alarm detection during an alarm flood, the target, or at least its exact timing, is not expected or cannot be predicted by the participant, whose attention needs to be drawn largely in a bottom-up fashion.

Another important difference between the experimental setup in the above studies and the alarm floods experienced in real-world settings is the number of stimuli/alarms. An alarm flood is defined as more than 10 alarms per 10 min period, but can often exceed that (EEMUA, 1999). Therefore, it is not uncommon for two or more alarms to occur concurrently or sequentially in very close temporal proximity. In contrast, most studies on change blindness or attentional blink involve two targets only. Based on the studies reviewed earlier, people's ability to report multiple concurrent or temporally close targets decreased with the number of targets (Boot et al., 2007; Burr, Turi, & Anobile, 2010). Therefore, in order to improve alarm detection during periods of high alarm frequency, including alarm floods, the number of alarms in temporal proximity is a critical factor to consider.

As reported, a small number of studies investigated the detection of multiple stimuli in close temporal proximity. However, participants in these studies were required only to report the number of targets, without identifying their nature. In the context of real-world alarms, the nature of an alarm may affect both detection and identification performance, since it relates to the expectancy and assigned value for the alarm. Therefore, in the proposed research, participants will be required to not only detect but also identify the alarms.

Finally, a better understanding is needed of whether and how the serial position of a stimulus affects its likelihood of being detected. For example, previous research has shown that unexpected events (such as an alarm) can lead to surprise and startle, which has been shown to slow down the processing of the unexpected signal and hinder the ability to direct and divide attention properly (Barnett, Wong, Westley, Adderley, & Smith, 2012; Landman, Groen, van Paassen, Bronkhorst, & Mulder, 2017). Thus, an operator may perform worse for the first few alarms in an alarm flood. At the same time, the stress and workload induced by an alarm flood are likely to intensify over time as the operator tries to keep up with the pace of the signals being presented. This can result in an increase in operators' mental arousal. Early on, this may have a positive effect on performance because a medium level of arousal tends to lead to the best performance (Yerkes & Dodson, 1908). As stress and arousal increase further, however, performance may suffer during later stages of the alarm flood. Finally, the processing of the last alarm in a flood may benefit from the fact that there is no interference from subsequent signals. This line of research will gather empirical data on how serial position affects alarm detection and identification during alarm floods which is important because these data may inform the development of countermeasures to missing critical information.

Supporting alarm detection in an alarm flood

Understanding how alarm detection is affected by various attentional limitations, contextual factors and top-down influences is a prerequisite for achieving the final goal of the proposed research: the development of an effective countermeasure to missed or misidentified alarms in the context of periods of high alarm frequency, including alarm floods. As mentioned earlier, suppressing redundant and nuisance alarms can be beneficial, but filtering and eliminating alarms risks hiding potentially valuable information from the operator. Therefore, the proposed research aimed to develop a complementary solution using context-sensitive presentation of alarms to accommodate human information processing abilities and limitations. This solution is based on two design considerations: (1) real-time adjustment of the timing of alarm presentation to avoid temporal masking and (2) pre-attentive criticality mapping to support top-down prioritization of alarms.

The adjustment of alarm timing is based on the findings from the earlier phases in this line of research. The first three studies have identified SOAs that need to be avoided in order to prevent breakdowns in detection performance and how the temporal distribution and the number of alarms interact and affect alarm detection. An intelligent alarm management system would avoid these time intervals by delaying some alarms for a short period of time and thus avoid masking of other alarms. This approach can complement other alarm management techniques because masking can happen even with a relatively small number of alarms.

Instead of suppressing alarms and hiding information from operators, alarm detection may be supported by helping operators classify and prioritize alarms and allocate their limited attentional resources accordingly. To this end, displays can be designed to map some pre-attentive features to properties of the alarms. Preattentive processing is a stage of visual processing where simple features are coded spatially before focused attention is required for later stages of processing (Treisman, 1985). Similar stages of auditory processing also allow people to partition the auditory input to some extent (Alain & Arnott, 2000). The alarm features that can utilize these processing stages and help direct attention to certain alarms are called pre-attentive references. The preattentive signals should be designed to be noticed in parallel with the ongoing task, provide information about the significance and meaning, and allow for evaluation that does not require focused attention (Woods, 1995). Preattentive reference has been proposed as a means to support interruption management and improve situation awareness by presenting the potential interruption and providing partial information of the nature of the interruption that allows for peripheral access (Endsley, 1988; Sarter, 2005; Wickens et al., 2008). In the context of alarm displays, preattentive reference provides support for the identification of alarm signals in a way that can be processed peripherally and in parallel. This type of support is critical for operations during high alarm frequency periods, such as alarm floods, because the operator often cannot afford to process all the alarm signals with focused attention due to the sheer volume of alarms as well as other ongoing tasks. Therefore, facilitating alarm processing at the preattentive stage is important in this context.

A design approach called urgency mapping has been proposed and empirically validated as promising means of supporting preattentive reference in alarm signals. This approach manipulates the perceived urgency of alarm signals by changing features such as the base frequency, loudness, duration or inter-pulse interval of auditory alarms, the color, wording, or pulse rate of visual alarms, and the pulse rate of tactile alarms (Baldwin & Lewis, 2014; Guillaume, Drake, Rivenez, Pellieux, & Chastres, 2002; Kaufmann, Ohg, Risser, Geven, & Sefelin, 2008; Lewis & Baldwin, 2012; Sanderson, Eunice, Philippe, & Alexandra, 2006). Many of these features, including color, shape, size, location, pitch, and timber, have been shown to be preattentive features (Doll, 1993; Winkler, Sussman, Tervaniemi, Horváth, Ritter, & Näätänen, 2003; Woods, 1995). They are assumed to be processed at the preattentive stage as they have been shown to be processed very quickly (i.e., within 200ms), before visual attention could have been engaged (Healey, Booth, & Enns, 1996). Such manipulation of perceived urgency has been shown to help drivers respond faster to more urgent warnings (Politis et al., 2014), and was also suggested to be beneficial in other domains such as medical alarms (Edworthy & Meredith, 1994). Although urgency mapping has been shown to improve the speed of responses, little is known about whether it can also improve the detection of critical alarms during an alarm flood. However, a benefit in detection performance can be expected based on the contingent orienting hypothesis which states that attention is more likely to be captured by objects that possess properties for which attention is set (Folk, Remington, & Johnston, 1992). Contingent orienting utilizes a combination of the by two types of factors that affect attention allocation: goal-directed (top-down) factors and stimulus-driven (bottom-up) factors (Corbetta & Shulman, 2002). When the operator is aware of certain preattentive features of the alarms and actively look for these features in the searching task, alarms possessing these features are more likely to be noticed and identified correctly. Based on these mechanisms of attention, it was expected that mapping preattentive features to the criticality of alarms would also improve detection and identification performance.

References

Alain, C., & Arnott, S. R. (2000). Selectively attending to auditory objects. *Front. Biosci*, *5*, D202–D212.

Arnell, K. M. (2006). Visual, auditory, and cross-modality dual-task costs: electrophysiological evidence for an amodal bottleneck on working memory consolidation. *Perception & Psychophysics*, *68*(3), 447–457. https://doi.org/10.3758/BF03193689

Arnell, K. M., & Jolicœur, P. (1999). The attentional blink across stimulus modalities:Evidence for central processing limitations. *Journal of Experimental Psychology. Human Perception and Performance*, *25*(3), 630–648.

Arnell, K. M., & Larson, J. M. (2002). Cross-modality attentional blinks without preparatory task-set switching. *Psychonomic Bulletin & Review*, *9*(3), 497–506. https://doi.org/10.3758/BF03196305

Atkinson, J., Campbell, F. W., & Francis, M. R. (1976). The magic number 4±0: A new look at visual numerosity judgements. *Perception*, *5*(3), 327–334.

Auvray, M., Gallace, A., Tan, H. Z., & Spence, C. (2007). Crossmodal change blindness between vision and touch. *Acta Psychologica*, *126*(2), 79–97. https://doi.org/10.1016/j.actpsy.2006.10.005

Baldwin, C. L., & Lewis, B. A. (2014). Perceived urgency mapping across modalities within a driving context. *Applied Ergonomics*, *45*(5), 1270–1277. https://doi.org/10.1016/j.apergo.2013.05.002

Barnett, J., Wong, W., Westley, D., Adderley, R., & Smith, M. (2012). Startle reaction: Capturing experiential cues to provide guidelines towards the design of realistic training scenarios. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 56, pp. 2477–2481). SAGE Publications Sage CA: Los Angeles, CA.

Ben-Asher, N., & Gonzalez, C. (2015). Effects of cyber security knowledge on attack detection. *Computers in Human Behavior*, *48*, 51–61.

Bliss, J. P., Freeland, M. J., & Millard, J. C. (1999). Alarm related incidents in aviation: A survey of the aviation safety reporting system database. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 43, pp. 6–10). SAGE Publications Sage CA: Los Angeles, CA.

Bolton, M. L., Edworthy, J., & Boyd, A. D. (2018). A Formal Analysis of Masking Between Reserved Alarm Sounds of the IEC 60601-1-8 International Medical Alarm Standard. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *62*(1), 523–527. https://doi.org/10.1177/1541931218621119

Boot, W. R., Becic, E., & Kramer, A. F. (2007). Temporal Limitations in Multiple Target Detection in a Dynamic Monitoring Task. *Human Factors*, *49*(5), 897–906. https://doi.org/10.1518/001872007X230244.

Brehaut, J. C., Enns, J. T., & Di Lollo, V. (1999). Visual masking plays two roles in the attentional blink. *Perception & Psychophysics*, *61*(7), 1436–1448. https://doi.org/10.3758/BF03206192

Breitmeyer, B. G., & Öğmen, H. (2006). *Visual masking: Time slices through conscious and unconscious vision*. Oxford University Press.

Broadbent, D. E., & Broadbent, M. H. P. (1987). From detection to identification: Response to multiple targets in rapid serial visual presentation. *Perception & Psychophysics*, *42*(2), 105–113. https://doi.org/10.3758/BF03210498

Burr, D. C., Turi, M., & Anobile, G. (2010). Subitizing but not estimation of numerosity requires attentional resources. *Journal of Vision*, *10*(6), 20.

Camos, V., & Tillmann, B. (2008). Discontinuity in the enumeration of sequentially presented auditory and visual stimuli. *Cognition*, *107*(3), 1135–1143.

Carhart, R., Tillman, T. W., & Greetis, E. S. (1969). Perceptual Masking in Multiple Sound Backgrounds. *The Journal of the Acoustical Society of America*, *45*(3), 694–703. https://doi.org/10.1121/1.1911445

Chun, M. M., & Potter, M. C. (2001). The attentional blink and task switching within and across modalities. In *The Limits of AttentionTemporal Constraints in Human Information Processing* (Vol. 15, pp. 20–35). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198505150.003.0002

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*(3), 201.

Costa, R., Cachulo, N., & Cortez, P. (2009). An intelligent alarm management system for large-scale telecommunication companies. In *Portuguese Conference on Artificial Intelligence* (pp. 386–399). Springer.

Curry, D. G., Meyer, J. E., & McKinney, J. M. (2006). Seeing versus perceiving. *Professional Safety*, *51*(6), 28.

Dell'Acqua, R., Jolicœur, P., Sessa, P., & Turatto, M. (2006). Attentional blink and selection in the tactile domain. *European Journal of Cognitive Psychology*, *18*(4), 537–559. https://doi.org/10.1080/09541440500423186

Dell'Acqua, R., Turatto, M., & Jolicoeur, P. (2001). Cross-modal attentional deficits in processing tactile stimulation. *Perception & Psychophysics*, *63*(5), 777–789. https://doi.org/10.3758/BF03194437

Doll, T. J. (1993). Preattentive processing in visual search. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 37, No. 19, pp. 1291-1294). Sage CA: Los Angeles, CA: SAGE Publications.

Doll, T. J., Hanna, T. E., & Russotti, J. S. (1992). Masking in Three-Dimensional Auditory Displays. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *34*(3), 255–265. https://doi.org/10.1177/001872089203400301

Duncan, J., Martens, S., & Ward, R. (1997). Restricted attentional capacity within but not between sensory modalities. *Nature*, *387*(6635), 808–810. https://doi.org/10.1038/42947

Dux, P. E., & Marois, R. (2009). The attentional blink: a review of data and theory. *Atten Percept Psychophys*, *71*(8), 1683–1700. https://doi.org/10.3758/APP.71.8.1683

Edworthy, J., & Meredith, C. S. (1994). Cognitive psychology and the design of alarm sounds. *Medical Engineering and Physics*. https://doi.org/10.1016/1350-4533(94)90067-1

EEMUA. (1999). *Alarm Systems: A Guide to Design, Management and Procurement*. Engineering Equipment and Materials Users Association London.

Endsley, M. R. (1988). Design and evaluation for situation awareness enhancement. In *Proceedings of the Human Factors Society annual meeting* (Vol. 32, pp. 97–101). SAGE Publications Sage CA: Los Angeles, CA.

Enns, J. T. J., Di Lollo V, & Di Lollo, V. (2000). What's new in visual masking? *Trends in Cognitive Sciences*, *4*(9), 345–352. https://doi.org/10.1016/S1364-6613(00)01520-5

Fastl, H., & Zwicker, E. (2006). *Psychoacoustics: facts and models* (Vol. 22). Springer Science & Business Media.

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary Covert Orenting Is

Contingent on Attentional Control Settings. *Journal of Experimental Psychology*, *18*(4), 1030–1044.
Gallace, A., Tan, H. Z., & Spence, C. (2007). Do “mudsplashes” induce tactile change blindness? *Perception & Psychophysics, 69*(4), 477–486. https://doi.org/10.3758/BF03193905
Goddard, K. M., Isaak, M. I., & Slawinski, E. B. (1997). The auditory attentional blink: Inhibition revisited. *Journal of the Acoustical Society of America*, *101*, 3125.
Grosdidier, P., Connor, P., Hollifield, B., & Kulkarni, S. (2003). A path forward for DCS alarm management. *Hydrocarbon Processing*, *82*(11), 59–68.
Guillaume, A., Drake, C., Rivenez, M., Pellieux, L., & Chastres, V. (2002). Perception of urgency and alarm design. *Proceedings of the 8th International Conference on Auditory Display (ICAD2002)*. Retrieved from Proceedings/2002/GuillaumeDrake2002.pdf
Healey, C. G., Booth, K. S., & Enns, J. T. (1996). High-speed visual estimation using preattentive processing. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 3(2), 107-135.
Hein, G., Parr, A., & Duncan, J. (2006). Within-modality and cross-modality attentional blinks in a simple discrimination task. *Perception & Psychophysics*, *68*(1), 54–61. https://doi.org/10.3758/BF03193655
Hillstrom, A. P., Shapiro, K. L., & Spence, C. (2002). Attentional limitations in processing sequentially presented vibrotactile targets. *Perception & Psychophysics*, *64*(7), 1068–1082. https://doi.org/10.3758/BF03194757
HSE. (1997). *The explosion and fires at the Texaco Refinery, Milford Haven, 24 July 1994. HSE books.*
Itti, L., Koch, C., & Niebur, E. (1998). A model of saliency-based visual attention for rapid scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *20*(11), 1254–1259.
Kashino, M., & Hirahara, T. (1996). One, two, many—Judging the number of concurrent talkers. *The Journal of the Acoustical Society of America*, *99*(4), 2596–2603.
Kaufman, E. L., Lord, M. W., Reese, T. W., & Volkmann, J. (1949). The discrimination of visual number. *The American Journal of Psychology*, *62*(4), 498–525.
Kaufmann, C., Ohg, F., Risser, R., Geven, A., & Sefelin, R. (2008). Effects of simultaneous multi-modal warnings and traffic information on driver behaviour. In *European Conference on Human Centered Design for Intelligent Transport Systems* (pp. 33–42). Retrieved from https://www.researchgate.net/profile/Martin_Baumann2/publication/224998610_The_effect_of_cognitive_tasks_on_predicting_events_in_traffic/links/02e7e52d3d8c521170000000.pdf#page=41
Kirschen, D. S., & Wollenberg, B. F. (1992). Intelligent alarm processing in power systems. *Proceedings of the IEEE*, *80*(5), 663–672.
Laberge, J. C., Bullemer, P., Tolsma, M., & Reising, D. V. C. (2014). Addressing alarm flood situations in the process industries through alarm summary display design and alarm response strategy. *International Journal of Industrial Ergonomics*, *44*(3), 395–406. https://doi.org/10.1016/j.ergon.2013.11.008
Landman, A., Groen, E. L., van Paassen, M. M. (René), Bronkhorst, A. W., & Mulder, M. (2017). Dealing With Unexpected Events on the Flight Deck: A Conceptual Model of Startle and Surprise. *Human Factors*, *59*(8), 1161–1172. https://doi.org/10.1177/0018720817723428

Le Meur, O., Le Callet, P., Barba, D., & Thoreau, D. (2006). A coherent computational Approach to model the bottom-up visual attention. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *28*, 802–817.

Lee, Y.-C., Lee, J. D., & Boyle, L. N. (2007). Visual attention in driving: the effects of cognitive load and visual disruption. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *49*(4), 721–733.

Lewis, B. A., & Baldwin, C. L. (2012). Equating Perceived Urgency Across Auditory, Visual, and Tactile Signals. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *56*(1), 1307–1311. https://doi.org/10.1177/1071181312561379

Liu, J., Lim, K. W., Ho, W. K., Tan, K. C., Srinivasan, R., & Tay, A. (2003). The intelligent alarm management system. *IEEE Software*, *20*(2), 66.

Lu, S. (2014). Tactile and crossmodal change blindness and its implications for display design. University of Michigan.

Lu, S., Wickens, C. D., Prinet, J. C., Hutchins, S. D., Sarter, N., & Sebok, A. (2013). Supporting Interruption Management and Multimodal Interface Design: Three Meta-Analyses of Task Performance as a Function of Interrupting Task Modality. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *55*(4), 697–724. https://doi.org/10.1177/0018720813476298

Mandler, G., & Shebo, B. J. (1982). Subitizing: an analysis of its component processes. *Journal of Experimental Psychology: General*, *111*(1), 1.

Martens, S., Kandula, M., & Duncan, J. (2010). Restricted attentional capacity within but not between sensory modalities: An individual differences approach. *PLoS ONE*, *5*(12), 1–6. https://doi.org/10.1371/journal.pone.0015280

Mora, F. A., Passariello, G., Carrault, G., & Le Pichon, J.-P. (1993). Intelligent patient monitoring and management systems: a review. *IEEE Engineering in Medicine and Biology Magazine*, *12*(4), 23–33.

Nieuwenstein, M. R., Potter, M. C., & Theeuwes, J. (2009). Unmasking the attentional blink. *Journal of Experimental Psychology. Human Perception and Performance*, *35*(1), 159–169. https://doi.org/10.1037/0096-1523.35.1.159

Nikolic, M. I., Orr, J. M., & Sarter, N. B. (2004). Why Pilots Miss the Green Box: How Display Context Undermines Attention Capture. *The International Journal of Aviation Psychology*, *14*(1), 39–52. https://doi.org/10.1207/s15327108ijap1401_3

Noda, M., Higuchi, F., Takai, T., & Nishitani, H. (2011). Event correlation analysis for alarm system rationalization. In *Asia-Pacific Journal of Chemical Engineering* (Vol. 6, pp. 497–502). https://doi.org/10.1002/apj.575

O’Regan, J. K., Rensink, R. a, & Clark, J. J. (1999). Change-blindness as a result of “mudsplashes”. *Nature*, *398*(6722), 34. https://doi.org/10.1038/17953

Pavani, F., & Turatto, M. (2008). Change perception in complex auditory scenes. *Perception & Psychophysics*, *70*(4), 619–629. https://doi.org/10.3758/PP.70.4.619

Perrow, C. (2011). *Normal accidents: Living with high risk technologies*. Princeton University Press.

Politis, I., Brewster, S. A., & Pollick, F. (2014). Evaluating multimodal driver displays under varying situational urgency. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems* (pp. 4067–4076). New York, New York, USA: ACM Press. https://doi.org/10.1145/2556288.2556988

Posner, M. I., Snyder, C. R., & Davidson, B. J. (1980). Attention and the detection of signals.

Journal of Experimental Psychology: General, *109*(2), 160.
Potter, M. C., Chun, M. M., Banks, B. S., & Muckenhoupt, M. (1998). Two attentional deficits in serial target search: The visual attentional blink and an amodal task-switch deficit. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *24*(4), 979–992. https://doi.org/10.1037/0278-7393.24.4.979
Raymond, J. E., Shapiro, K. L., & Arnell, K. M. (1992). Temporary suppression of visual processing in an RSVP task: An attentional blink? *Journal of Experimental Psychology: Human Perception and Performance*, *18*(3), 849–860. https://doi.org/10.1037/0096-1523.18.3.849
Reising, D. V. C., Downs, J. L., & Bayn, D. (2004). Human Performance Models for Response to Alarm Notifications in the Process Industries: An Industrial Case Study. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *48*(10), 1189–1193. https://doi.org/10.1177/154193120404801009
Revkin, S. K., Piazza, M., Izard, V., Cohen, L., & Dehaene, S. (2008). Does subitizing reflect numerical estimation? *Psychological Science*, *19*(6), 607–614.
Rodrigo, V., Chioua, M., Hagglund, T., & Hollender, M. (2016). Causal analysis for alarm flood reduction. *IFAC-PapersOnLine*, *49*(7), 723–728.
Sanderson, P., Eunice, S., Philippe, L., & Alexandra, W. (2006). Auditory alarms, medical standards, and urgency.
Sarter, N. B. (2005). Graded and multimodal interruption cueing in support of preattentive reference and attention management. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 49, pp. 478–481). SAGE Publications Sage CA: Los Angeles, CA.
Sathian, K., Simon, T. J., Peterson, S., Patel, G. A., Hoffman, J. M., & Grafton, S. T. (1999). Neural evidence linking visual object enumeration and attention. *Journal of Cognitive Neuroscience*, *11*(1), 36–51.
Sessa, P. (2013). Guidelines for medical alarm system software design. In *Altran Italia | Technology Review # 08* (pp. 16–27).
Shen, D., & Mondor, T. a. (2006). Effect of distractor sounds on the auditory attentional blink. *Perception & Psychophysics*, *68*(2), 228–243. https://doi.org/10.3758/BF03193672
Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences*, *1*(7), 261–267.
Simons, D. J., & Rensink, R. A. (2005). Change blindness: Past, present, and future. *Trends in Cognitive Sciences*, *9*(1), 16–20. https://doi.org/10.1016/j.tics.2004.11.006
Soto-Faraco, S., & Spence, C. (2002). Modality-specific auditory and visual temporal processing deficits. *The Quarterly Journal of Experimental Psychology. A, Human Experimental Psychology*, *55*(1), 23–40. https://doi.org/10.1080/02724980143000136
Soto-Faraco, S., Spence, C., Fairbank, K., Kingstone, A., Hillstrom, A. P., & Shapiro, K. (2002). A crossmodal attentional blink between vision and touch. *Psychonomic Bulletin & Review*, *9*(4), 731–738. https://doi.org/10.3758/BF03196328
Stanton, N. A. (1993). Operators reactions to alarms: Fundamental similarities and situational differences. *Proceedings of the Conference on Human Factors in Nuclear Safety, Le Meridien Hotel, London*, 84–103. https://doi.org/10.1201/9780203481974-9
Stanton, N. A., & Baber, C. (1995). Alarm-initiated activities: An analysis of alarm handlingby operators using text-based alarm systems in supervisory control systems. *Ergonomics*, *38*(11), 2414–2431. https://doi.org/10.1080/00140139508925276

Stanton, N. A., Booth, R. T., Stammers, R. B., & Stammers, R. B. (1992). Alarms in human supervisory control: A human factors perspective. *International Journal of Computer Integrated Manufacturing*, *5*(2), 81–93. https://doi.org/10.1080/09511929208944518

Stanton, N. A., & Stammer, R. B. (1998). Alarm initiated activities: matching visual formats to alarm handling "tasks." *International Journal of Cognitive Ergonomics*, *2*(4), 331–348.

Steelman, K. S., McCarley, J. S., & Wickens, C. D. (2011). Modeling the Control of Attention in Visual Workspaces. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *53*(2), 142–153. https://doi.org/10.1177/0018720811404026.Copyright

Sykes, D., Barach, P., Basner, H., Belojevic, G., Busch-Vishniac, I., Cavanaugh, W., … Davenny, B. (2011). Clinical alarms and fatalities resulting from "alarm fatigue" in hospitals: perspectives from clinical medicine, acoustical science, signal processing, noise control engineering and human factors. *Paper from the Clinical Alarms*, 4–5.

ten Hoopen, G., & Vos, J. (1979). Effect on numerosity judgment of grouping of tones by auditory channels. *Perception & Psychophysics*, *26*(5), 374–380.

Treisman, A. (1985). Preattentive processing in vision. *Computer Vision, Graphics, and Image Processing*, *31*(2), 156–177.

Van der Burg, E., Brederoo, S. G., Nieuwenstein, M. R., Theeuwes, J., & Olivers, C. N. L. (2010). Audiovisual semantic interference and attention: Evidence from the attentional blink paradigm. *Acta Psychologica*, *134*(2), 198–205. https://doi.org/10.1016/j.actpsy.2010.01.010

Van Der Burg, E., Nieuwenstein, M. R., Theeuwes, J., & Olivers, C. N. L. (2013). Irrelevant Auditory and Visual Events Induce a Visual Attentional Blink. *Experimental Psychology*, *60*(2), 80–89. https://doi.org/10.1027/1618-3169/a000174

Velichkovsky, B. M., Dornhoefer, S. M., Kopf, M., Helmert, J., & Joos, M. (2002). Change detection and occlusion modes in road-traffic scenarios. *Transportation Research Part F: Traffic Psychology and Behaviour*, *5*(2), 99–109.

Vogel-Heuser, B., Schütz, D., & Folmer, J. (2015). Criteria-based alarm flood pattern recognition using historical data from automated production systems (aPS). *Mechatronics*, *31*, 89–100. https://doi.org/10.1016/J.MECHATRONICS.2015.02.004

Ward, R., Duncan, J., & Shapiro, K. (1996). The Slow Time-Course of Visual Attention. *Cognitive Psychology*, *30*(1), 79–109. https://doi.org/10.1006/cogp.1996.0003

Wickens, C. D. (2008). Multiple Resources and Mental Workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 449–455. https://doi.org/10.1518/001872008X288394

Wickens, C. D., Goh, J., Helleberg, J., Horrey, W. J., & Talleur, D. A. (2003). Attentional models of multitask pilot performance using advanced display technology. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *45*(3), 360–380.

Wickens, C. D., & McCarley, J. S. (2007). *Applied attention theory*. CRC press.

Wickens, C. D., McCarley, J., & Steelman, K. S. (2009). NT-SEEV: A model of attention capture and noticing on the flight deck. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 53, pp. 769–773). Sage Publications.

Wickens, C. D., Small, R. L., Andre, T., Bagnall, T., Brenaman, C., Wickens, C. D., … Brenaman, C. (2008). Multisensory Enhancement of Command Displays for Unusual Attitude Recovery. *The International Journal of Aviation Psychology*, *18*(3), 255–267. https://doi.org/10.1080/10508410802073491

Winkler, I., Sussman, E., Tervaniemi, M., Horváth, J., Ritter, W., & Näätänen, R. (2003).

Preattentive auditory context effects. *Cognitive, Affective, & Behavioral Neuroscience, 3*(1), 57-77.
Wollenberg, B. F. (1986). Feasibility study for an energy management system intelligent alarm processor. *IEEE Transactions on Power Systems*, *1*(2), 241–246.
Woods, D. D. (1995). The alarm problem and directed attention in dynamic fault management. *Ergonomics*, *38*(11), 2371–2393.
Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology*, *18*(5), 459–482.
Zwaga, H., & Hoonhout, H. (1994). Supervisory control behavior and the implementation of alarms in process control. In *Human Factors in Alarm Design* (pp. 119–134). Retrieved from https://dl.acm.org/citation.cfm?id=212912

Chapter 2

Establishing the SOA Range for Asynchronous Masking of Visual and Auditory Stimuli in Multitask Settings

As discussed in the Introduction, attentional blink is an example of asynchronous masking that was observed in earlier studies when two stimuli were presented at SOAs ranging from 200 – 600ms. The phenomenon was reported intramodally for both visual and auditory stimuli, as well as crossmodally between the two sensory channels. Attentional blink and other forms of masking may account for observed failures to detect and identify alarms during alarm floods; however, to date, these phenomena have been studied almost exclusively in the context of very simple experimental settings involving only two and mostly visual stimuli. Also, most studies on attentional blink used the RSVP paradigm which differs from real-world alarm signals with respect to the duration and timing of stimuli. In RSVP, each stimulus is presented for around 100ms which is considerably shorter than the duration of most alarms which are typically longer than 500ms (FAA, 2009). With a longer the duration of the stimuli, the effects of masking are more complicated because, on one hand, the stimuli will overlap with a shorter SOA, leading to a bigger risk of simultaneous masking, but on the other hand, a longer stimulus duration increases the chance of it being consolidated into the working memory and thus being detected and identified properly. Finally, most studies on attentional blink do not require participants to perform any concurrent primary task(s) while trying to detect target stimuli in the RSVP sequence. Thus, these studies fail to capture the multitasking attentional demands on operators in real-world workplaces. Based on limited empirical evidence (e.g., Ferris, Penfold, Hameed, & Sarter, 2006), the SOAs at which attentional blink is experienced in more complex and demanding environments appear to be longer. For example, in the Ferris et al. study, the error rate was reported to be higher with an SOA of 1000ms, compared to 500ms. Establishing the SOA range for asynchronous masking of visual and auditory stimuli in these contexts is important for being able to predict and counteract the failure to detect alarms in real-world settings.

To this end, the study reported in this chapter investigates alarm detection in the context of an Unmanned Aerial Vehicle (UAV) simulation. The participants in the experiment played the role of a ground-based operator monitoring multiple autonomous package delivery drones (for an example of such a vehicle, see ***Figure 2.1***). The drones flew predetermined paths to customer residences where they dropped a package on to a delivery pad placed on the ground by the customer. If no pad was detected, the package was not delivered. The drone then continued to the next delivery address or returned to the warehouse to pick up more packages. To be commercially viable, such an autonomous delivery system will require that one operator controls a fairly large number of drones, making it a Single-Operator-Multiple-Agent (SOMA) system. SOMA systems impose significant attentional demands on operators (Prinet, Terhune, & Sarter, 2012), and they involve an increased risk of alarm floods in case multiple vehicles experience problems at the same time. Some of these problems may be interrelated, such as poor visibility in a region which can affect the ability of all vehicles in the area to locate their respective delivery pads. Autonomous package delivery is, therefore, an appropriate application domain for studying the detection and identification of alarms and establishing SOAs for asynchronous masking of visual and auditory signals in a demanding multi-task setting.

Figure 2.1 An autonomous package delivery drone (Amazon.com, Inc.)

Based on a review of findings from previous studies, the following hypotheses were formulated:

- H1: Due to masking, the detection rate and accuracy for the second alarm in an alarm pair will be lower, and the response time would be longer, compared to the first alarm and to alarms that occur by themselves;
- H2: The effective SOA range for attentional blink in a multi-task environment will be longer than the 200 - 600ms range reported in earlier more basic research;
- H3: Masking effects will be observed both intra- and crossmodally for visual and auditory alarms;
- H4: Alarm floods will result in lower detection rates, identification accuracy and longer response times to alarms, compared to routine operations.

Method

Participants

The participants in this study were 26 students from the University of Michigan (16 males and 6 females). The results from the first four subjects were excluded from the final analysis due to changes in the experiment setup. All of the participants were between 20 and 30 years old (mean age = 24.35, SD = 2.50). They all reported normal or corrected-to-normal vision and normal hearing ability. Participants were consented before the experiment and received $25 as compensation for their participation. The three participants who had the best performance received an additional award of $70, $40, and $20, respectively. This study was approved by the University of Michigan Institutional Review Board (UM IRB ID: HUM00125337).

Apparatus and tasks

A simulation of an automated package delivery system was developed in The Human-automation Interaction and Cognition (THInC) Laboratory. The experiment setup consisted of a computer with a keyboard, an optical mouse, a 23-inch LCD monitor, and a pair of stereo speakers. Simulated video feeds from eight drones were presented. In addition, a simple control panel was shown next to each video feed. The central window was used to display visual alarms for the various drones (see ***Figure 2.2***).

Participants were asked to perform two tasks: (1) delivery consent and (2) alarm detection. Delivery consent was required whenever a drone reached one of its destinations. The drone would follow predetermined flight paths to customer residences. Once it reached a delivery address, it hovered in midair. At that point, the brightness of the respective drone window increased to indicate to the participant the need to determine whether the customer had placed a delivery pad (with the letter "H" on it) on the ground. If a delivery pad was present, the participant pressed the top button (a "target" symbol) next to the corresponding window to consent to the package delivery. If no delivery pad was detected, the participant would press the second button from the top (the same 'target' icon but with a line drawn across the icon) to cancel the delivery and make the drone proceed to the next customer. Once the participant had entered their response, the brightness of the window would return to its default lower setting.

Throughout the scenario, alarms were presented whenever there was a temporary link loss between any of the drones and the ground control station. Alarms indicated the number of the affected drone (1-8). They were visual (the number flashed in the central alarm window on the screen) or auditory (synthesized speech announcing the drone number) in nature. Participants were required to detect the alarm and press one of two buttons next to the respective video feed(s). For visual alarms, they would press the third button from the top which showed an eye; the response to auditory alarms was to press the "ear" button at the bottom. The alarms were presented either as single alarms (to determine baseline detection performance) or in alarm pairs (two alarms that were presented in close temporal proximity, at 5 different SOAs).

Figure 2.2 Interface of the UAV control simulation

Cross-modal matching

A technique called cross-modal matching was employed in advance of the experiment. Cross-modal matching refers to "the process of an observer matching the apparent intensities of stimuli across two sensory modalities" (Colman, 2015). Cross-modal matching is important to avoid confounds in studies of multimodal processing, i.e., to ensure that any observed performance differences can be attributed to the modality per se, rather than other stimulus properties (such as salience). In this study, cross-modal matching was performed by presenting each participant with the visual alarm signals at preset brightness levels and asking them to use the keyboard arrow keys to adjust the volume of the corresponding auditory alarm signals to match the salience of both cues (Pitts, Riggs, & Sarter, 2016).

Experiment design and procedure

The experiment employed a 5×4×3 fractional factorial design. The three independent variables, all varied within-subjects, were Stimuli Onset Asynchrony (SOA; the time between the onsets of two alarms in an alarm pair; 200, 600, 800, 1000, or 1200ms), modality pairs (visual-visual, visual-auditory, auditory-visual, and auditory-auditory), and alarm presentation (single alarm or first or second alarm in an alarm pair).

The experiment consisted of two 30-minute scenarios. Each scenario included 27 minutes of routine operations and a 3-minute alarm flood. During the routine period, the delivery consent task was presented once per minute (27 times), either by itself (7 times), followed by a single alarm (10 times) or by an alarm pair (10 times). The alarm flood started either 5 minutes or 25 minutes into the scenario. During the 3-minute alarm flood, 30 single alarms and 30 alarm pairs were presented to the participant. The frequency of the delivery task was the same as during routine operations. The order of alarms was randomized. A total of 30 delivery consent tasks were presented to participants in each scenario.

Upon the participants' arrival at the laboratory, they were given a printed standard IRB consent form that explained the purpose of the study and what participants would experience during the experiment. They were required to sign the form in order to participate in the study. Their chair was then adjusted so that the eyes were about 20 inches from the center of the screen. At that distance, the size of the visual alarms equaled 0.125 rad visual angle. Next, the participants were given instructions on the delivery consent task and the alarm monitoring task. They were told to give equal priority to both tasks. They then received two to three five-minute training sessions to familiarize themselves with the tasks until they reached an overall accuracy of 90% in the two tasks. Following the training sessions, the participant took a short break before starting the two 30-minute experiment sessions. The entire experiment lasted about 100 minutes. Each participant received $25 as compensation upon completion. As an incentive, the three participants with the best combined performance on the two tasks received a bonus of $70, $40, and $20, respectively.

Dependent measures

The three dependent measures in this study were detection rate, accuracy, and response time. Detection rate was defined as the percentage of alarms that the participant responded to, regardless of the correctness of the response. Accuracy was defined as the percentage of correct responses to the alarms, out of all responses. Response time was defined as the time elapsed from the onset of the alarm to the time the participant pressed one of the two buttons next to the video feed.

Results

Detection rate and accuracy for visual and auditory alarms were analyzed separately using mixed model binary logistic regression. The two fixed effect factors were alarm position and SOA and the only random effect factor was subject ID. Sequential Bonferroni corrections were applied for multiple comparisons. A univariate analysis of variance (ANOVA) was conducted for the response time for all alarms including visual and auditory alarms. Alarm position, SOA, and modality were entered as fixed effect factors, but modality was later removed due to a lack of significance. Subject ID was included as a random effect factor, and Bonferroni corrections were again applied for multiple comparisons. For all analyses, only significant results ($p < .05$) are reported here.

Detection rate

Overall, alarm detection rates were significantly lower for visual alarms (92.5%), compared to auditory alarms (99.0%, $X^2\ (\mathrm{df} = 1) = 24.722,\ p < .05$). The average detection rate for a visual alarm was 94.3% when it was presented as a single alarm, and 92.5% and 90.9% when it was the first or second alarm in an alarm pair, respectively. The difference between the detection rates for single visual alarms and second visual alarms in alarm pairs was statistically significant (90.9% versus 94.3%; $F(2, 2482) = 3.097,\ p = .045$). An interaction between SOA and alarm position was observed such that the detection rate for the second alarm was lower when the SOA was 600, 1000 or 1200ms and higher when the SOA was 200ms. For auditory alarms, no significant difference was found between any of the alarm conditions/settings.

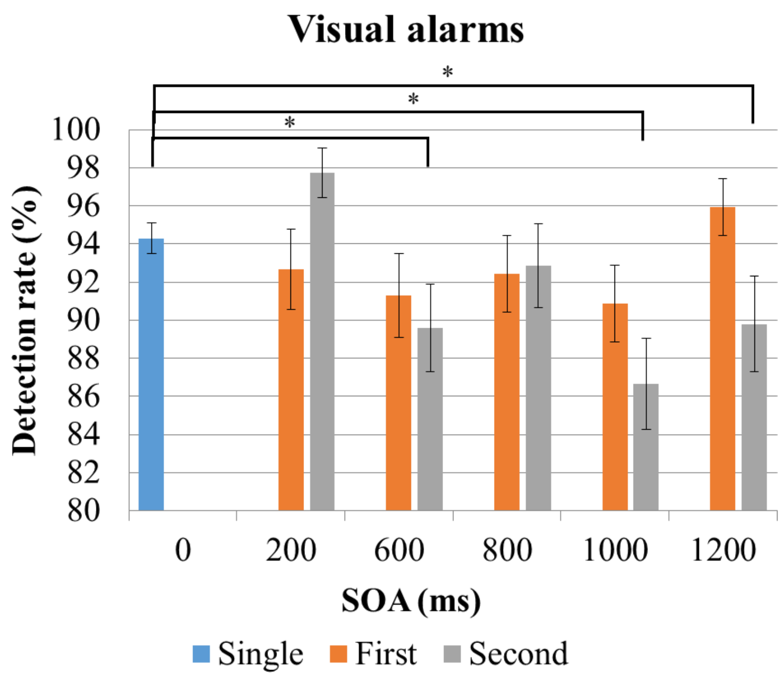

Figure 2.3 Visual alarm detection rates as a function of SOA and alarm position

($^*p < 0.05$)

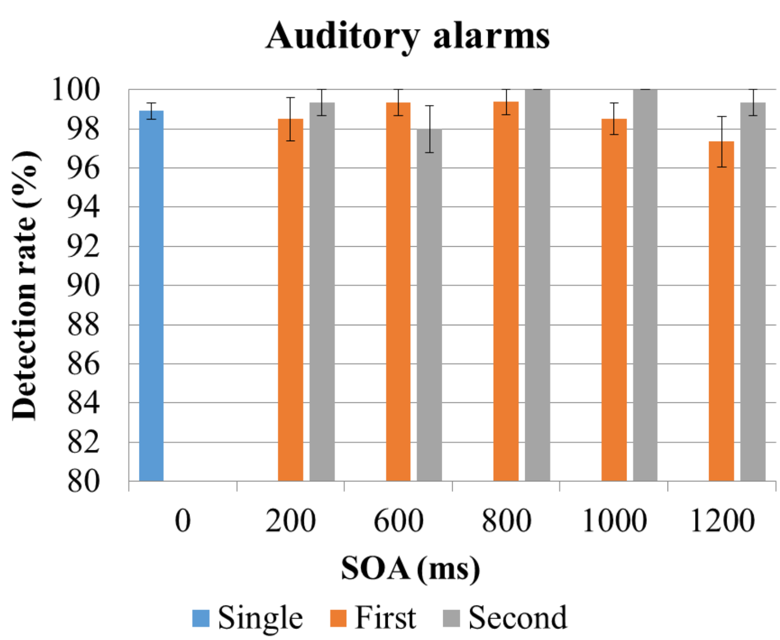

Figure 2.4 Auditory alarm detection rates as a function of SOA and alarm position

($^*p < 0.05$)

Accuracy

Accuracy was lower for visual alarms, compared to auditory alarms (91.0% versus 99.0%, respectively; $X^2(\mathrm{df} = 1) = 77.45,\ p < .001$). The average accuracy for a visual alarm was 93.6% when it was presented as a single alarm, and 94.0% and 86.3% when it was the first or second alarm in an alarm pair, respectively. A main effect of alarm position was observed ($F(2, 2297) = 16.577,\ p < .001$), and post hoc tests revealed a significant difference between the detection rates for single visual alarms and the second visual alarm in alarm pairs (86.3% versus 93.6%, respectively). The effect of alarm position was moderated by the SOA such that detection rates were significantly lower only when the SOA was 600, 1000 or 1200ms. For auditory alarms, no significant difference was found between any conditions.

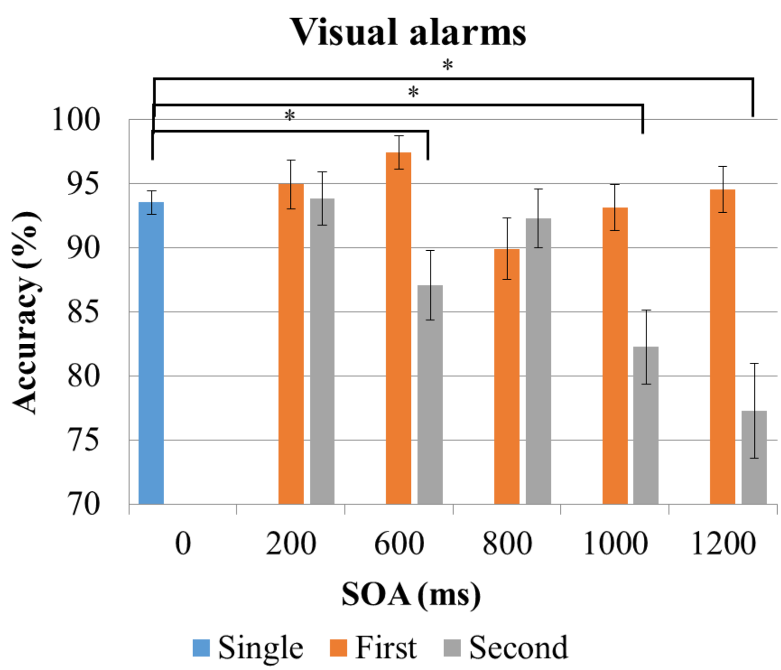

Figure 2.5 Visual alarm accuracy as a function of SOA and alarm position

(*p < 0.05)

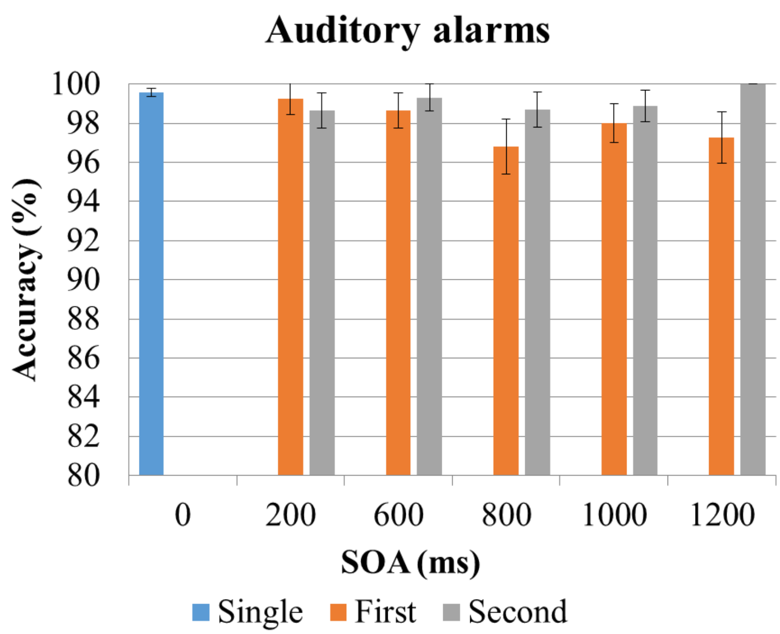

Figure 2.6 Auditory alarm accuracy as a function of SOA and alarm position

(*p < 0.05)

Response time

There was no significant difference between the response times to visual alarms and auditory alarms. However, a main effect of alarm position was observed ($F(1, 4356) = 94.781$, $p < .001$; see ***Figure 2.7***). Response times to both alarms in an alarm pair were longer than those to single alarms, except at SOAs of 1200ms. Post-hoc tests (with Bonferroni correction) revealed that this effect was more pronounced for the second alarm than the first alarm in an alarm pair ($p < .001$). A main effect of SOA was also observed, where the response time to all alarms was longer with a shorter SOA ($F(4, 4356) = 40.987,\ p < .001$). Finally, there was an interaction effect between SOA and alarm position, such that longer response times at short SOAs were observed more often for the second alarm ($F(4, 4356) = 14.788,\ p < .001$).

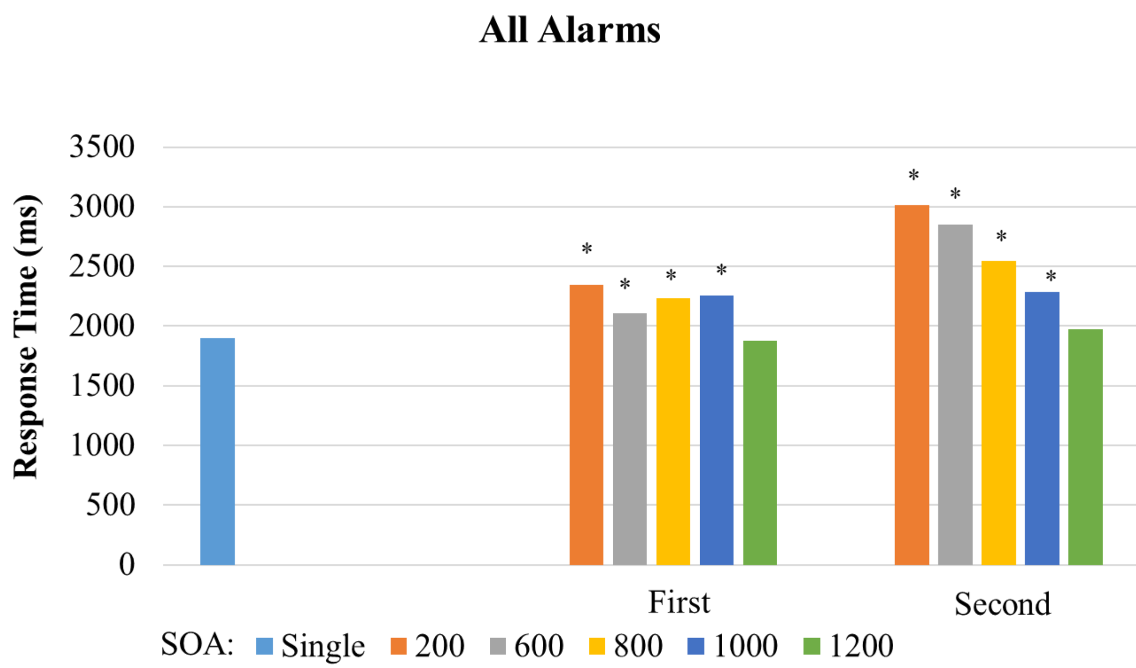

Figure 2.7 Response time as a function of SOA and alarm position

(*p<0.05, asterisk alone indicates significant difference from single alarms)

Alarm flood analysis

During alarm floods, detection rates for any of the alarms did not differ from those during routine phases of the scenario (see ***Figure 2.8***). Accuracy was slightly lower during alarm floods (94.7% versus 96.7%), but this difference was only marginally significant ($F(1, 4588) = 3.320$, $p = .068$). The average response time decreased by more than 200ms during the alarm floods, compared to routine periods (2160ms versus 2318ms; $F(1, 4588) = 6.284,\ p = .001$). Finally, there was no difference between trials where the alarm flood occurred after 5 minutes and those where the alarm flood started after 25 minutes.

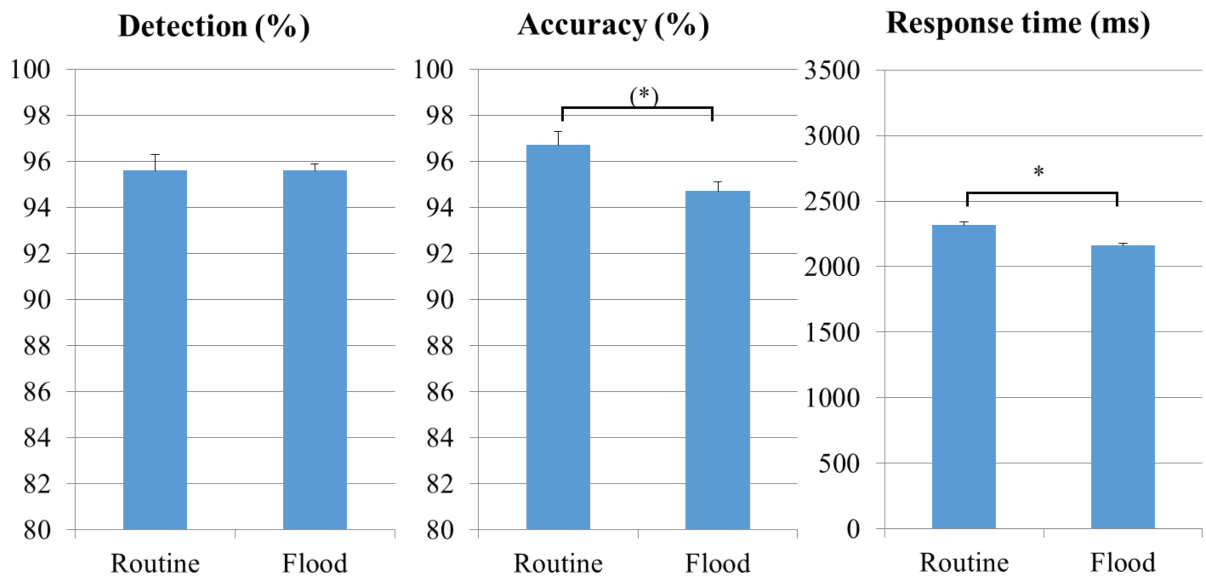

Figure 2.8 Performance metrics during routine periods vs. alarm floods

(*p < .05)

To identify possible performance changes over the course of the alarm floods, these were divided into an early phase (the first 25 alarms), a middle phase (the 30 alarms in the middle of an alarm flood), and a late phase (the last 25 alarms). The detection rates, accuracy and response times were compared across the three phases (see ***Figure 2.9***). There was a significant main effect of phase on detection rate ($F(2, 3809) = 7.136$, $p = .001$) and response time ($F(2, 3451) = 13.067$, $p < .001$) and a trend towards lower accuracy across the three phases ($F(2, 3643) = 2.096$, $p = .123$). Detection rates were higher during the last two phases but did not differ between phase 2 and 3. Response times were significantly lower in the last phase, compared to the first two phases which did not differ.

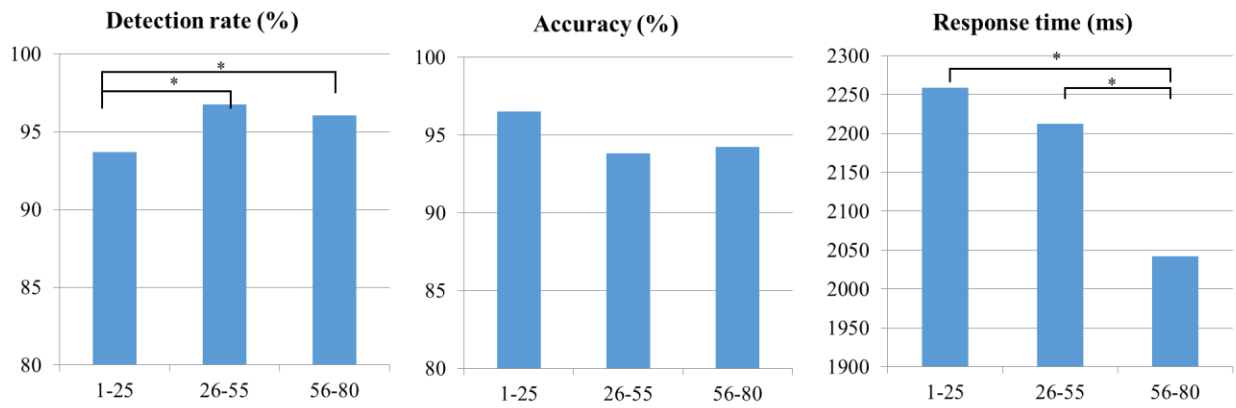

Figure 2.9 Performance as a function of phases during the alarm flood

Discussion

Masking effects, such as attentional blink, have been shown to affect the detection of visual and auditory stimuli at SOAs of 200 – 600ms when using rapid serial visual or auditory presentation of short simple stimuli in single-task settings. Very few studies have examined, however, whether and at what SOAs these effects may be observed also with longer visual and auditory alarms in a more complex multi-task environment. The present study sought to fill this gap.

Asynchronous masking was expected to result in decreased detection rates for the second alarm in alarm pairs. This hypothesis was partially confirmed by the results of the present study. Lower detection rates and accuracy were observed for second alarms but only when these were visual alarms, regardless of the modality of the preceding alarm. This finding is in agreement with results from an earlier empirical study that showed attentional blink affecting the detection of pairs of visual targets in a sonar display monitoring task (Boot, Becic, & Kramer, 2007). In the present study, the absence of masking effects for auditory alarms may be explained by the fact that there was no competition for auditory attentional resources. Both the detection of the two types of visual targets (the delivery pads and the visual alarms) and the manual response required visual attention. As a result, a ceiling effect was observed for auditory alarms, with detection rates and accuracy well above 95% in all the cases. Visual alarms may have been missed more easily or misidentified also because participants started moving their eyes away from the center of the screen to locate the response buttons as soon as they detected a visual or auditory alarm. Therefore, if the second alarm in an alarm pair was a visual alarm, it appeared in peripheral vision. According to the NT-SEEV model, eccentricity requires increased effort to notice a signal and thus reduces its noticeability (Wickens, McCarley, & Steelman, 2009). Finally, a phenomenon called inattentional blindness may have contributed to the failure to detect the second, visual alarm. Inattentional blindness refers to the failure to notice an unexpected signal or event due to the engagement of attention on other aspects of the visual scene (in this case, the response button) (Mack & Rock, 1999).

The fact that both visual and auditory alarms showed masking effects on the following visual alarm suggested potential crossmodal links in the processing of multimodal alarms. Similar crossmodal links have been observed in spatial attention. For example, a salient stimulus in one modality can orient attention in other modalities to the same location (Driver & Spence, 1998; Spence & Driver, 2004). When a tactile cueing signal and a visual or auditory target appear in the same or similar location, participants' response time to the target is significantly shorter than when they appear in opposite directions (Ho, Tan, & Spence, 2005). These results suggested a crossmodal link in orienting spatial attention. Similarly, results from the current study suggested that the attention resource required for the detection and identification of visual and auditory alarms are not entirely independent. This result is consistent with findings from previous studies that have shown that asynchronous masking occurs crossmodally, especially when the second target is visual (Arnell & Larson, 2002). It suggests that the underlying mechanism is likely associated with some resource or process that is not totally modality specific. Some researchers explain attentional blink as a bottleneck of consolidating targets into working memory. This bottleneck has been suggested to be amodal (Arnell, 2006; Vogel, Woodman, & Luck, 2006).

Response time was expected to be longer for the second alarm in an alarm pair, based on the results from previous attentional blink studies. According to these studies, a delay in the consolidation of the second target is responsible for the longer response time (e.g. Arnell, 2006). The current results show that the increase in response time to the second alarm was indeed largely proportional to the decrease in SOA. For example, the average response time was 2285ms when the SOA was 1000ms, and it was 3016ms when the SOA was 200ms. The difference in response times closely matched the difference in SOA. This supports the claim of a consolidation bottleneck which suggests that the consolidation of the second alarm has to wait until that of the first alarm has finished (Vogel & Luck, 2002). Therefore, no matter when the second alarm was presented, it had to wait for the consolidation of the first to end. The response time to the second alarm therefore increased with a shorter SOA, while the sum of the response time and the SOA remained about the same.

The expectation that the SOA at which asynchronous masking occurs would be longer in this study than the 200 – 600ms suggested by earlier more basic research on attentional blink was confirmed. Both the detection and identification of visual alarms were affected not only at an SOA of 600ms but also at SOAs as long as 1000 – 1200ms. This finding could be attributed to the longer time required to process the alarm signals due to increased task demands in this experiment. As mentioned above, it has been proposed that attentional blink occurs when a second target is presented while the first target is still being consolidated from its perceptual representation into working memory. This consolidation process happens sequentially for each target. Thus, if the perceptual representation of the second target cannot outlast this consolidation bottleneck, it will not enter into working memory (Arnell, 2006; Vogel & Luck, 2002). It has been proposed that the speed of consolidating an item into working memory varies as a function of task complexity, from 50ms in a single-task paradigm to 500ms in a dual-task setting (Chun & Potter, 1995; Vogel et al., 2006). The consolidation bottleneck in the present study may have been even longer than 500ms due to the demanding nature of the tasks. This longer bottleneck may have prevented the detection of a second alarm signal even with SOAs above 600ms (Arnell, 2006; Nieuwenstein, Potter, & Theeuwes, 2009). It should be noted that the masking effect was not as strong at an SOA of 800ms. The reason for this gap in the effective SOA range warrants further investigation.

Although SOA did not affect the detection rate and accuracy for auditory alarms, there was a main effect of SOA on response times to both visual and auditory alarms. Participants were slower to respond to alarm pairs than to single alarms when the SOA was shorter than 1200ms. This effect was more pronounced the shorter the SOA, especially for the second alarm in an alarm pair. These results are consistent with previous studies on so-called double stimulation tasks, where the participants were presented with one or two groups of lines and required to judge the length of the lines were. Results from these studies indicated that a shorter inter-stimuli interval (ISI, another metric used to quantify the time between two stimuli; ISI is essentially SOA minus the duration of the first stimuli) led to longer response time (for ISIs between 33 – 1000ms), but had no effect on accuracy (Aykin & Aykin, 1987; Aykin, Czaja, & Drury, 1986). In addition, in this experiment, response time did not differ significantly between visual and auditory alarms, indicating that the speed of processing visual and auditory alarms, once detected, is similar.

It was expected that detection and identification performance would decrease, and response times would increase during an alarm flood due to interference between a large number of alarms in a short period of time. However, the observed performance changes were very different. During alarm floods, the detection rate did not differ from that during routine periods, accuracy decreased slightly, and, surprisingly, the response time to alarms was significantly shorter. Thus, we did not see a net performance decrease, but rather a speed-accuracy tradeoff. A speed-accuracy tradeoff, i.e., an inverse relation between speed and accuracy, is observed when responses to a task are required to be both rapid and accurate, as in the present study. Speed-accuracy tradeoffs have been observed with a wide range of tasks, including psychomotor tasks and visual search tasks (MacKenzie & Isokoski, 2008; Schlegel & Storm, 1983; Wickens, Hooey, Gore, Sebok, & Koenicke, 2009). The observation of this tradeoff has been explained as a result of limited mental resources and the fact that these tasks can only be processed serially (Drury, 1994), which also applies to the tasks used in the current study.

A further breakdown of the alarm flood into three phases revealed performance changes over time. The detection rate increased, and response times became faster. These changes could be the result of participants investing more effort in the alarm detection task once they realized that they entered into an alarm flood. The performance improvements could also be explained by increased arousal due to the high pace of events during the alarm flood. According to previous studies on arousal, a certain level of mental workload is required to maintain attention and interest in order to achieve optimal arousal and performance (Matthews & Davies, 1998; Molloy & Parasuraman, 1996; Warm, Parasuraman, & Matthews, 2008; Wiener, Curry, & Faustina, 1984; Yerkes & Dodson, 1908). This finding is important as it suggests that operators need support especially during the early stage of an alarm flood to help them cope with possible startle effects and the transition from slow to fast operations.

References

Arnell, K. M. (2006). Visual, auditory, and cross-modality dual-task costs: electrophysiological evidence for an amodal bottleneck on working memory consolidation. *Perception & Psychophysics*, *68*(3), 447–457. https://doi.org/10.3758/BF03193689

Arnell, K. M., & Larson, J. M. (2002). Cross-modality attentional blinks without preparatory task-set switching. *Psychonomic Bulletin & Review*, *9*(3), 497–506. https://doi.org/10.3758/BF03196305

Aykin, N., & Aykin, T. (1987). Complex Task Performance under Speed-Accuracy Tradeoff: Single Task versus Dual Task. In *Proceedings of the Human Factors Society Annual Meeting* (Vol. 31, pp. 161–165). Los Angeles, CA: SAGE Publications.

Aykin, N., Czaja, S. J., & Drury, C. G. (1986). A simultaneous regression model for double stimulation tasks. *Human Factors*, *28*(6), 633–643.

Boot, W. R., Becic, E., & Kramer, A. F. (2007). Temporal Limitations in Multiple Target Detection in a Dynamic Monitoring Task. *Human Factors*, *49*(5), 897–906. https://doi.org/10.1518/001872007X230244.

Chun, M. M., & Potter, M. C. (1995). A two-stage model for multiple target detection in rapid serial visual presentation. *Journal of Experimental Psychology: Human Perception and Performance*, *21*(1), 109–127. https://doi.org/10.1037/0096-1523.21.1.109

Colman, A. M. (2015). *A dictionary of psychology*. Oxford University Press, USA.

Driver, J., & Spence, C. (1998). Attention and the crossmodal construction of space. *Trends in Cognitive Sciences*, *2*(7), 254–262. https://doi.org/10.1016/S1364-6613(98)01188-7

Drury, C. G. (1994). The speed—accuracy trade-off in industry. *Ergonomics*, *37*(4), 747–763.

FAA. (2009). *Practice Alarms and Alerts in the Technical Operations*.

Ferris, T. K., Penfold, R., Hameed, S., & Sarter, N. (2006). The Implications of Crossmodal Links in Attention for the Design of Multimodal Interfaces: A Driving Simulation Study. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(3), 406–409. https://doi.org/10.1177/154193120605000341

Ho, C., Tan, H. Z., & Spence, C. (2005). Using spatial vibrotactile cues to direct visual attention in driving scenes. *Transportation Research Part F: Traffic Psychology and Behaviour*, *8*(6), 397–412. https://doi.org/10.1016/j.trf.2005.05.002

Mack, A., & Rock, I. (1999). Inattentional Blindness: An Overview By Arien Mack & Irvin Rock.

MacKenzie, I. S., & Isokoski, P. (2008). Fitts' throughput and the speed-accuracy tradeoff. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1633–1636). ACM.

Matthews, G., & Davies, D. R. (1998). Arousal and vigilance: Still vital at fifty. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 42, pp. 772–776). SAGE Publications Sage CA: Los Angeles, CA.

Molloy, R., & Parasuraman, R. (1996). Monitoring an automated system for a single failure: Vigilance and task complexity effects. *Human Factors*, *38*(2), 311–322.

Nieuwenstein, M. R., Potter, M. C., & Theeuwes, J. (2009). Unmasking the attentional blink. *Journal of Experimental Psychology. Human Perception and Performance*, *35*(1), 159–169. https://doi.org/10.1037/0096-1523.35.1.159

Pitts, B. J., Riggs, S. L., & Sarter, N. (2016). Crossmodal Matching: A Critical but Neglected Step in Multimodal Research. *IEEE Transactions on Human-Machine Systems*, *46*(3), 445–

450.
Prinet, J. C., Terhune, A., & Sarter, N. B. (2012). Supporting Dynamic Re-Planning In Multiple Uav Control: A Comparison of 3 Levels of Automation. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 56, pp. 423–427). SAGE Publications.
Schlegel, R. E., & Storm, W. F. (1983). Speed-Accuracy Tradeoffs in Spatial Orientation Information Processing. *Proceedings of the Human Factors Society Annual Meeting*, *27*(5), 359–363. https://doi.org/10.1177/154193128302700505
Spence, C., & Driver, J. (2004). *Crossmodal space and crossmodal attention*. Oxford University Press.
Vogel, E. K., & Luck, S. J. (2002). Delayed working memory consolidation during the attentional blink. *Psychonomic Bulletin & Review*, *9*(4), 739–743.
Vogel, E. K., Woodman, G. F., & Luck, S. J. (2006). The time course of consolidation in visual working memory. *Journal of Experimental Psychology: Human Perception and Performance*, *32*(6), 1436.
Warm, J. S., Parasuraman, R., & Matthews, G. (2008). Vigilance requires hard mental work and is stressful. *Human Factors*, *50*(3), 433–441.
Wickens, C. D., Hooey, B. L., Gore, B. F., Sebok, A., & Koenicke, C. S. (2009). Identifying Black Swans in NextGen: Predicting Human Performance in Off-Nominal Conditions. *Human Factors*, *51*(5), 638–651. https://doi.org/10.1177/0018720809349709
Wickens, C. D., McCarley, J., & Steelman, K. S. (2009). NT-SEEV: A model of attention capture and noticing on the flight deck. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 53, pp. 769–773). Sage Publications.
Wiener, E. L., Curry, R. E., & Faustina, M. Lou. (1984). Vigilance and task load: In search of the inverted U. *Human Factors*, *26*(2), 215–222.
Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology*, *18*(5), 459–482.

Chapter 3

Establishing the SOA Range for Asynchronous Masking of Visual and Auditory Stimuli in a High Workload Multitask Setting

The experiment described in Chapter 2 demonstrated that asynchronous masking is a likely contributor to alarms being missed in demanding multi-task settings and during alarm floods. The detection rate and identification accuracy were lower for visual alarms that were presented following another visual or auditory alarm in close temporal proximity. The SOA range at which asynchronous masking was experienced was longer than the 200 – 600ms suggested by the basic literature on attentional blink, a specific case of asynchronous masking. The study involved limitations, however, which were associated with the design of the participants' task. Specifically, participants were required to respond to an alarm by clicking on a button next to the window of the affected drone. This required a re-orientation of their gaze away from the center of the screen where visual alarms were presented. This may have interfered with the detection of visual, but not auditory alarms. Another related concern is that the alarm detection task, and the overall task set, were not sufficiently difficult. In particular, the detection rate and accuracy for auditory alarms were above 95% in all conditions. To address these limitations, and establish the role of masking in a context that is more representative of real-world operations, the study reported in this chapter differed from the previous experiment in terms of:

1. Increased task demands to test the effects of asynchronous masking under high workload;
2. Elimination of the need for forced re-orientation of gaze so that the detection of visual alarms would not be at a disadvantage

Specifically, we changed the response mechanism to, and required classification of alarms. Participants were required to verbally report alarms, instead of clicking on a button. By using verbal responses, they were able to maintain their gaze and monitor the center of the screen for other visual alarms. The difficulty of the tasks was increased by introducing different types of alarms that were color-coded or differed with respect to the gender of the voice used to present the alarms. Participants had to verbally announce both the color/gender and the number of the affected drone. We also added an air traffic control monitoring task to introduce competing attentional demands in the auditory channel.

Method

Participants

The participants in this study were 15 students recruited from the College of Engineering at the University of Michigan. The number of participants was determined by performing a statistical power analysis based on the results of the previous experiment. All participants were between 20 and 35 years old (mean = 24.2 years, SD = 3.4 years; 8 males and 7 females) and had not participated in the previous study reported in Chapter 2. They all reported normal or corrected-to-normal vision, normal color vision, and normal hearing ability. This study was approved by the University of Michigan Institutional Review Board (UM IRB: HUM00144319).

Apparatus and tasks

The simulation used in this study was a modified version of the ground-based drone control system described in Chapter 2. It consisted of a computer with a keyboard, an optical mouse, a 23-inch LCD monitor, and a pair of stereo speakers. An audio recorder was added in this study to record the participants' verbal responses to the visual and auditory alarms.

During the experiment, participants were required to perform three tasks: delivery consent, alarm detection and identification, and air traffic control monitoring. Delivery consent was required whenever a drone reached a customer residence. The participant needed to search for the delivery pad with the letter "H" on it. Upon detection of the pad, s/he had to press the top button (with a "target" symbol) next to the corresponding drone window to give consent to the delivery of the package. If the pad was not present, the participant needed to click on the second button from the top (showing the same "target" symbol but with a line drawn across it) to cancel the delivery. The drone would then proceed to the next customer without dropping the package. In this study, the brightness of the drone window remained unchanged when the UAV reached a residence in order to increase the overall task difficulty.

Throughout the experiment, visual and auditory alarms were presented whenever there was a temporary link loss between any of the drones and the ground control station. Visual alarms were presented in the central window of the screen, which was divided into a 3x3 grid (see ***Figure 3.1***). The outer rectangles of the central alarm grid mapped spatially onto the eight drone windows. A visual alarm was displayed as a small red or blue box in the rectangle adjacent to the affected drone (see ***Figure 3.1*** for an example of an alarm related to drone #8; the color coding is explained below). Auditory alarms were presented using a synthesized male or female voice.

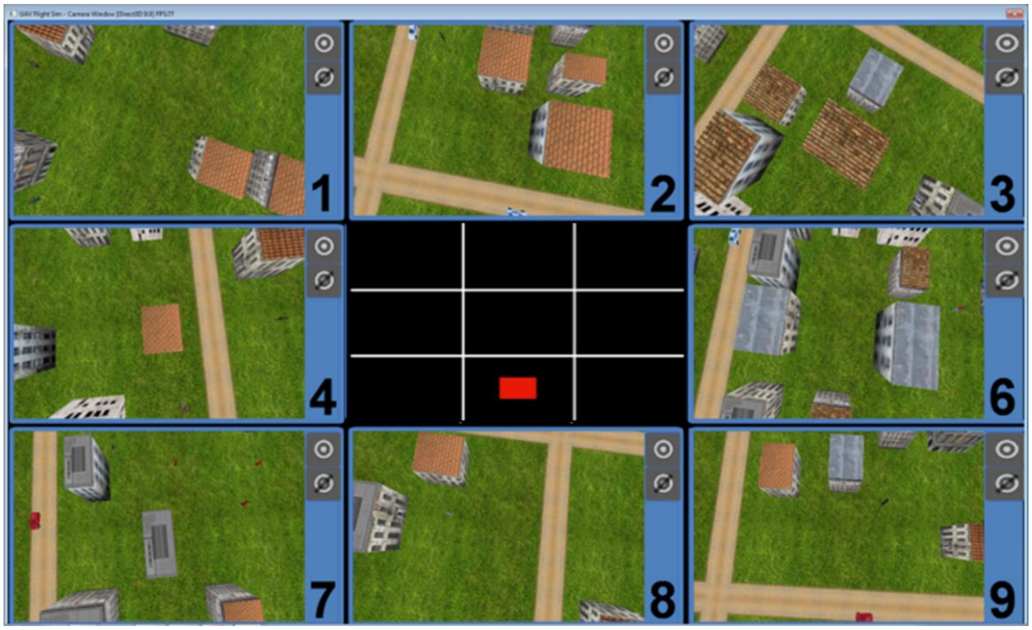

Figure 3.1 Drone simulation interface

(displaying a red visual alarm for drone 8; requiring the response “R8”)

There were two categories of alarms which were defined by color (visual alarms) or the gender of the voice (auditory alarms). Type A alarms included red visual alarms and male voice auditory alarms for drones 1, 3, 6, 8, as well as blue visual alarms and female voice auditory alarms for drones 2, 4, 7, 9. Type B alarms included the remaining alarms (see **Table 3.1**). The two alarm categories required different responses. For type A alarms, the participants were required to report the first letter of the alarm color or voice (“R” for red visual alarms, “B” for blue visual alarms, “M” for male voice auditory alarms, and “F” for female voice auditory alarms), followed by the number of the affected drone. For type B alarms, the required response was the reverse; participants had to report the number of the affected drone first, followed by the first letter of the color or voice of the alarm. This manipulation served to increase the difficulty of the alarm identification and response tasks.

Table 3.1 Alarm type as a function of drone number, modality, and color/voice

	Drone number							
Alarm type	1	2	3	4	6	7	8	9
Visual Red	A	B	A	B	A	B	A	B
Visual Blue	B	A	B	A	B	A	B	A
Auditory Male	A	B	A	B	A	B	A	B
Auditory Female	B	A	B	A	B	A	B	A

The third task required participants to monitor a continuous audio recording of air traffic control at the Detroit airport. They were monitoring for three pre-recorded messages containing their call sign “DOI51”. These messages were randomly distributed throughout the Detroit tower recording. To increase the difficulty of the task, another three messages were added that contained similar information but did not include the call sign “DOI51”. Whenever participants heard their call sign, they were required to press the space bar on the keyboard as quickly as possible. This task was introduced to create resource competition in the auditory channel, similar to the competing attentional demands of the delivery consent and the visual alarm monitoring tasks.

Experiment design and procedure

The experiment employed a 5×4×3 fractional factorial design. The three independent variables, all varied within subjects, were Stimuli Onset Asynchrony (SOA; the time between the onsets of two alarms in an alarm pair; 200, 600, 800, 1000, or 1200ms), modality pairs (visual-visual, visual-auditory, auditory-visual, or auditory-auditory) and alarm presentation (single alarm or first or second alarm in an alarm pair).

The experiment consisted of three 20-minute scenarios, named early (alarm) flood, late (alarm) flood, and (alarm) flood only, respectively (see ***Figure 3.2***). Each scenario included 17 minutes of routine operations and a 3-minute alarm flood. The only difference between the early and late flood scenarios and the flood only scenario was that, in the latter, alarms were presented only during the 3-minute alarm flood. In each scenario, the delivery consent task was presented 30 times (about once every 40 seconds). During routine operations in the early and late flood scenarios, 20 of the delivery consent tasks were followed by a single alarm (10 times) or an alarm pair (10 times). In the flood only scenario, no alarms were presented during routine operations. The alarm flood started 5, 10, or 15 minutes into the early flood, flood only, and late flood scenario, respectively. During the 3-minute alarm flood, 30 single alarms and 30 alarm pairs were presented to the participant. The order of alarms was randomized. Air traffic control messages were presented about once per minute throughout the scenario.

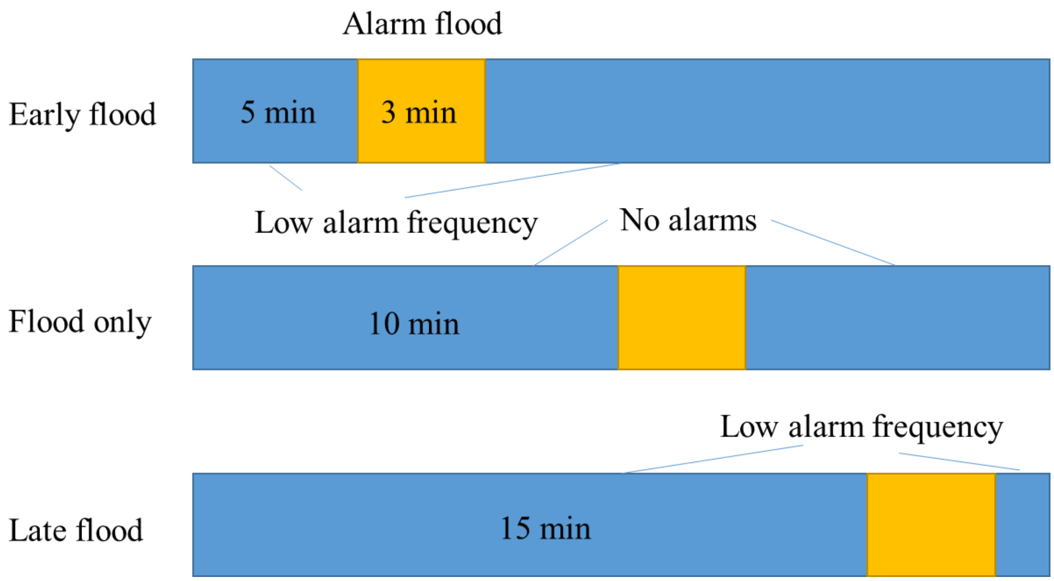

Figure 3.2 Three scenarios and the distribution of alarms

Upon arrival at the laboratory, participants were asked to read and sign a printed consent form. Their chair was then adjusted so that their eyes were about 20 inches from the center of the screen. At this distance, the size of the visual alarms equaled 0.125 rad visual angle. Next, participants were given instructions for their three tasks, and they were told to give equal priority to all tasks. They then received four to five 5-minute training sessions to familiarize themselves with the tasks until they reached a combined accuracy of 90% or higher across all three tasks. Following the training sessions, the participant took a short break before starting the three 20-minute experiment sessions. The order of the three sessions was counterbalanced between participants. Before each session, cross-modal matching was performed as in the previous experiment, by adjusting the volume of the auditory alarms to match the intensity of the visual alarms (Colman, 2015; Pitts, Riggs, & Sarter, 2016). The entire experiment lasted about 100 minutes. Upon completion, each participant received $25 as compensation. As an incentive, the three participants with the best combined performance on the three tasks received a bonus of $50, $30, and $20, respectively.

Dependent measures

The dependent measures in this study were detection rate and accuracy for the alarm monitoring task, accuracy and response time for the delivery consent task, and accuracy and response time for the air traffic control monitoring task. For the visual and auditory alarms, detection rate was defined as the percentage of alarms that the participant responded to, regardless of the accuracy of the response. Accuracy was defined as the percentage of correct responses to the alarms, out of all responses. For the delivery consent task, accuracy was defined as the percentage of correct responses (giving consent only when a delivery pad was present and rejecting when the delivery pad was absent), divided by all responses. Response time was defined as the time from when the drone reached the residency and stopped to when the participant pressed one of the two buttons next to the drone window. For the air traffic control monitoring task, accuracy was defined as the percentage of correct responses (only responding when the call sign “DOI51” was included in the message). Response time was defined as the time from the start of the call sign to the time when the space key was pressed.

Results

Mixed model binary logistic regression was used to analyze the detection rate and accuracy separately for visual and auditory alarms, with alarm position and SOA as fixed effect factors. The same model was also used to analyze the accuracy for the ATC monitoring task, with scenario and alarm frequency as fixed effect factors. In all the models mentioned above, subject ID was included as a random effect factor and Sequential Bonferroni corrections were applied for multiple comparisons.

Response time to the delivery consent task was analyzed using a univariate analysis of variance (ANOVA), with scenario and alarm frequency as fixed effect factors and subject ID as a random effect factor. Again, Bonferroni corrections were applied for multiple comparisons. For all analyses, the significance level was set at $p < .05$. Only significant results are reported here.

Detection rate

Alarm number (single alarms vs. alarm pairs) had a significant main effect on the detection of visual alarms ($F(1, 2233) = 42.254,\ p < .001$). Both the first and second visual alarms in an alarm pair were less likely to be detected by participants, compared to single alarms (59.5% and 63.8% vs. 74.7%, respectively). There was also a significant main effect of SOA on the detection of visual alarms ($F(4, 1475) = 4.863,\ p = .001$). The detection rate was lower when the SOA was 600 or 800ms (56.6% and 54.5%, respectively), compared to the other three SOAs (see ***Figure 3.3***) which did not differ significantly from each other. There was no significant interaction between the two factors.

Auditory alarms were also less likely to be detected when they were presented in alarm pairs than when they were presented as single alarms (see ***Figure 3.4***; 63.9% vs. 82.3%; $F(1, 2263) = 83.287,\ p < .001$). In contrast to visual alarms, the position of an auditory alarm in an alarm pair had a significant effect on its likelihood of detection. Specifically, the second alarm in an alarm pair was more likely to be missed than the first alarm (59.2% vs. 68.8%; $F(1, 1505) = 17.628,\ p < .001$). SOA also had a significant effect on the detection rate for auditory alarms. When the SOA was 200, 800, or 1000ms, the detection rate was lower (57.1%, 56.6%, and 63.6%, respectively) than for the other two SOAs ($F(4, 1505) = 7.287,\ p < .001$). No significant interactions between the various factors were observed.

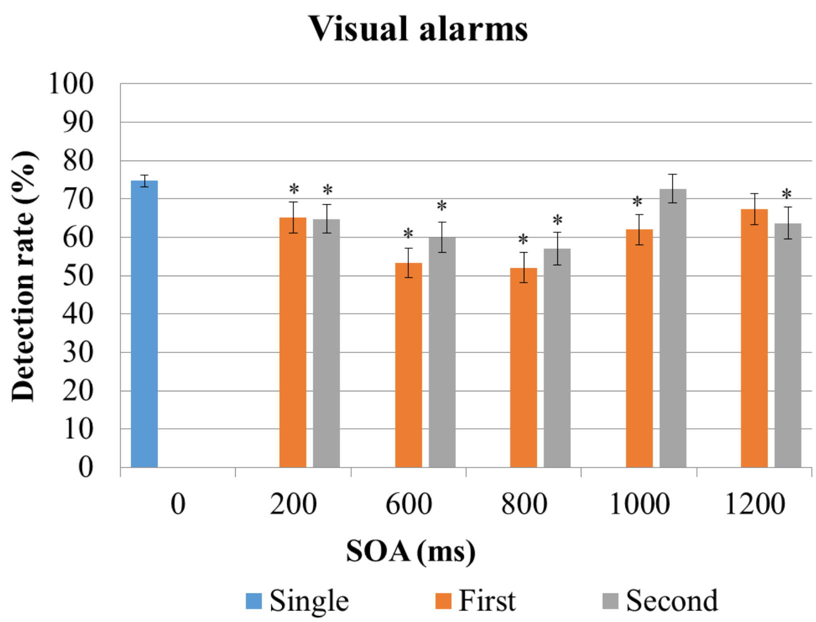

Figure 3.3 Detection rates of visual alarms as a function of position and SOA

(**p* < .05 when compared with single alarms)

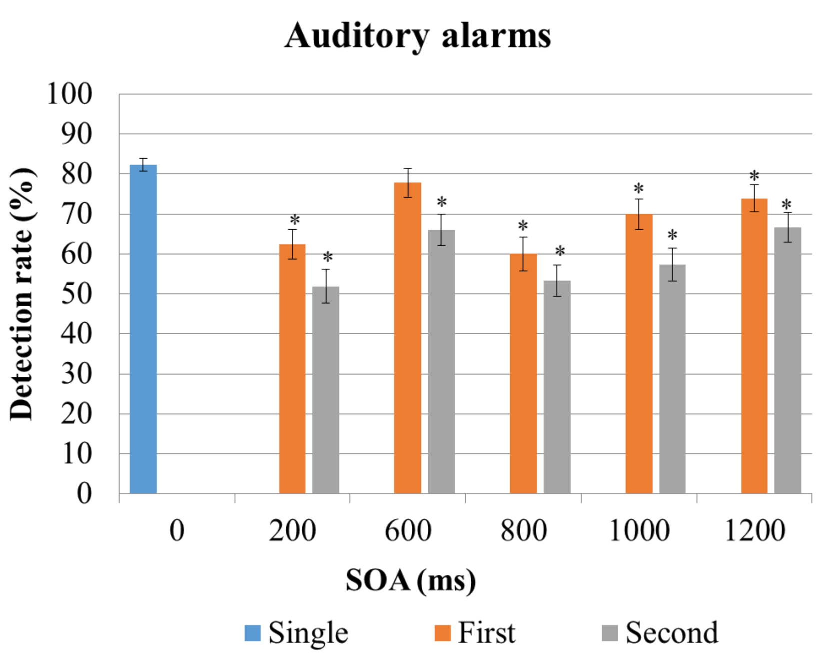

Figure 3.4 Detection rates of auditory alarms as a function of position and SOA

(**p* < .05 when compared with single alarms)

Accuracy

Overall, the accuracy was higher for auditory alarms than for visual alarms (93.3% vs. 77.6%; $F(1, 3128) = 85.346,\ p < .001$). For visual alarms, accuracy did not differ significantly between singles and alarm pairs. Within alarm pairs, accuracy was affected by both SOA ($F(4, 926) = 4.731,\ p = .001$) and alarm presentation ($F(1, 926) = 13.609,\ p < .001$). Accuracy was higher at an SOA of 1200ms (86.7%), compared to all other SOAs (see ***Figure 3.5***). It was also higher for the second, as compared to the first alarm in an alarm pair. A significant interaction effect was observed between SOA and position ($F(4, 926) = 3.803,\ p = .004$), such that accuracy was lower for the first visual alarms in an alarm pair when the SOA was 200, 600, or 1000ms, but higher for the second visual alarm at an SOA of 600 or 1200ms (see ***Figure 3.5***), compared to single alarms. No significant effects of SOA or position were observed for auditory alarms (see **Figure *3.6***).

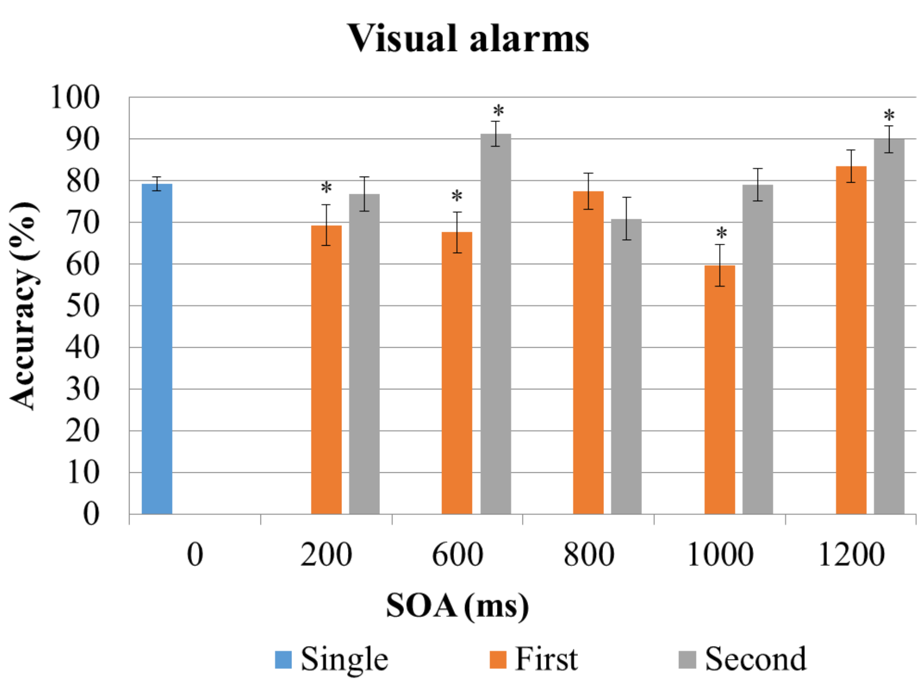

Figure 3.5 Accuracy of visual alarms as a function of position and SOA

(**p* < .05 when compared with single alarms)

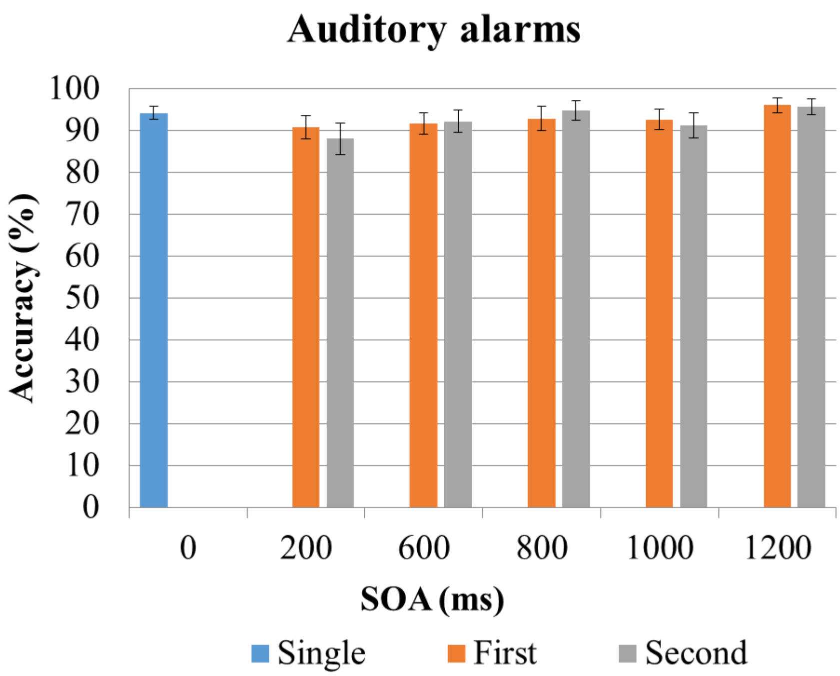

Figure 3.6 Accuracy of auditory alarms as a function of position and SOA

(**p* < .05 when compared with single alarms)

Crossmodal effects

Figure 3.7 and ***Figure 3.8*** show a comparison of the detection rate and accuracy for first and second alarms, respectively, as a function of the modality of the preceding or the following alarm in an alarm pair. A significant difference between intra- and crossmodal pairs was observed only when the first alarm was a visual alarm. The detection rate and accuracy for these visual alarms were both lower when they were followed by an auditory alarm, compared to another visual alarm ($F(1,748) = 3.998,\ p = .046$ and $F(1,456) = 4.192,\ p = .041$, respectively).

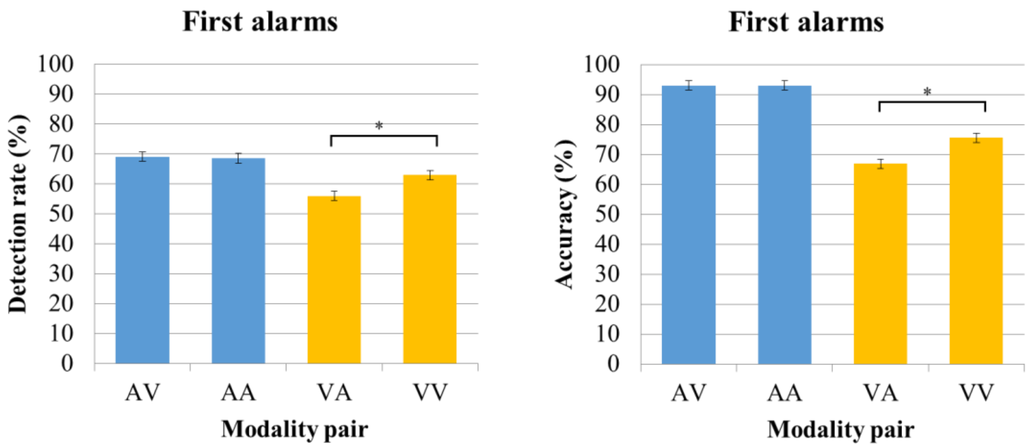

Figure 3.7 Detection rate and accuracy of first alarms as a function of the modality of the following alarm

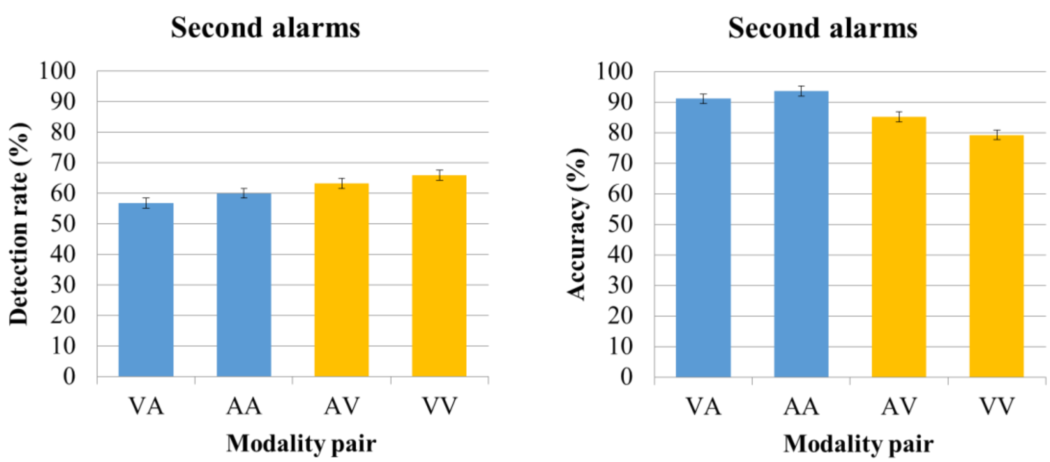

Figure 3.8 Detection rate and accuracy of second alarms as a function of the modality of the preceding alarm

Alarm flood analysis

The overall detection rate for all alarms was significantly lower during alarm floods, compared to routine periods (64.6% vs. 98.7%; $F(1, 4498) = 86.352,\ p < .001$). Accuracy was also significantly lower during alarm floods (85.0% vs. 90.1%; $F(1, 3128) = 9.066,\ p = .003$; see ***Figure 3.9***).

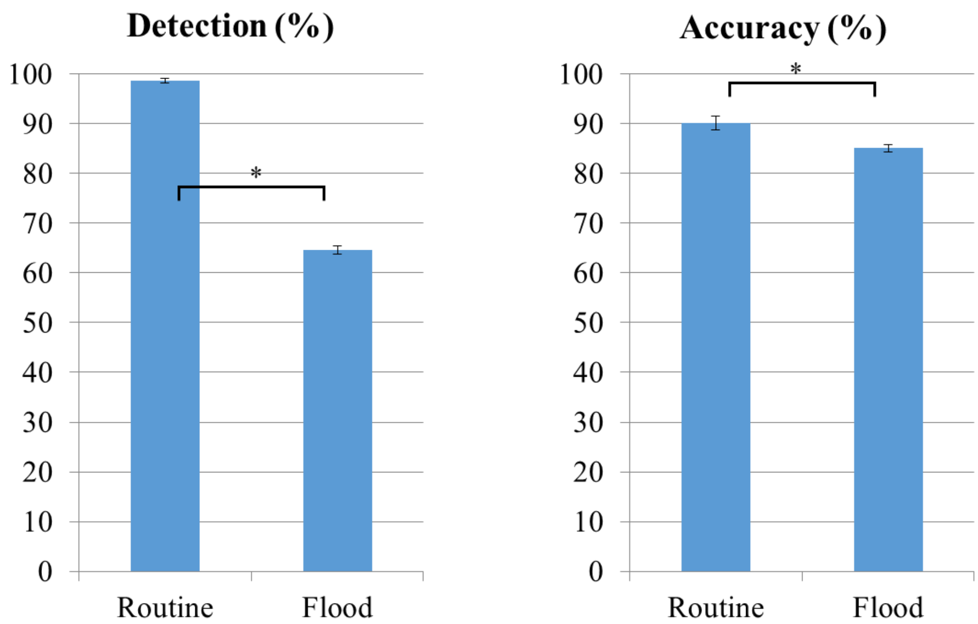

Figure 3.9 Detection rate and accuracy during alarm floods versus routine periods

(*p < .05)

Figure 3.10 and ***Figure 3.11*** show the detection rate and accuracy for each of the 30 single alarms and 30 alarm pairs as a function of their serial position in the alarm flood (for all three scenarios). Detection rate and accuracy were normalized by dividing the performance for each alarm or alarm pair by the average performance for all alarms/alarm pairs of the same type (e.g., the detection rate for a single visual alarm was divided by the average detection rate for all single visual alarms). The first and last alarms or alarm pairs during a flood were more likely to be detected than the rest of the alarms (a paired t-test between these two positions and other alarms yielded $t(4048) = 5.551$, $p < .001$). The first alarm or alarm pair in this study were either a single auditory or visual alarm, or an auditory alarm pair; the last one was either a single auditory alarm or a visual alarm pair. Accuracy was not affected by the serial position of the alarm.

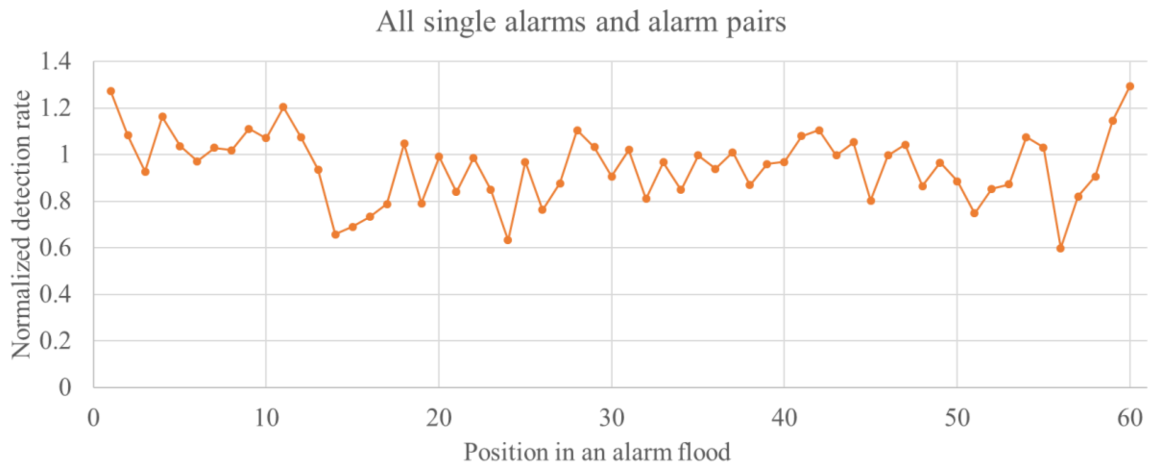

Figure 3.10 Normalized detection rate of single alarms and alarm pairs as a function of serial position in an alarm flood

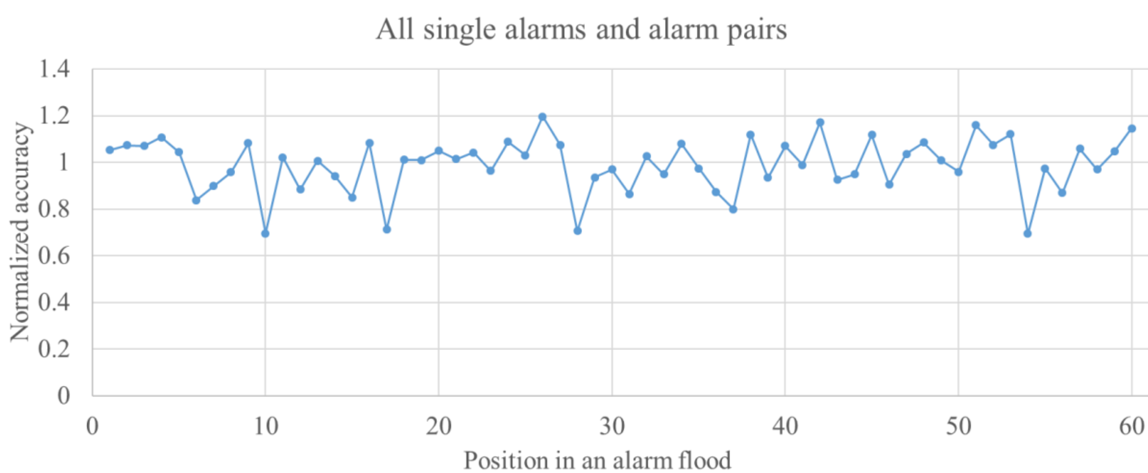

Figure 3.11 Normalized accuracy of single alarms and alarm pairs as a function of serial position in an alarm flood

Delivery consent task

Across all scenarios, 98.7% responses to the delivery consent task were accurate. However, the response time to the delivery consent task varied as a function of workload (routine vs. flood) and flood scenario (early flood, late flood, flood only). Response times were longer during alarm floods ($F(1, 1433) = 102.316,\ p < .001$), compared to routine operations. There was also a main effect of flood scenario ($F(2, 1433) = 4.949,\ p = .007$) and an interaction effect between workload and scenario ($F(2, 1433) = 7.342,\ p = .001$), such that response times were shorter during the alarm flood in late flood scenarios, compared to early flood and flood only scenarios (see ***Figure 3.12***).

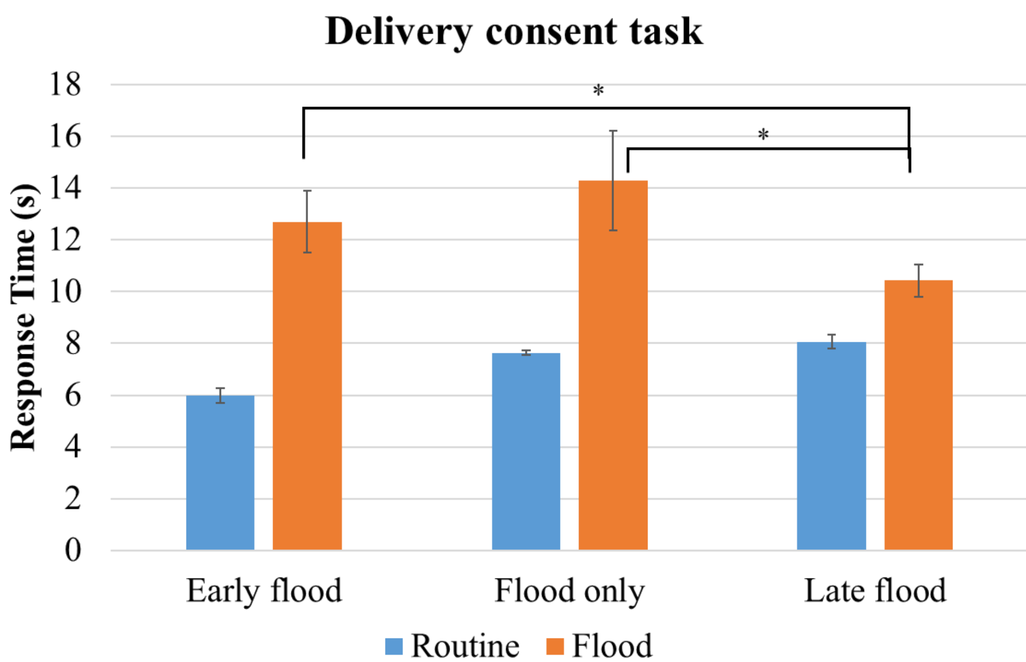

Figure 3.12 Response time to delivery consent as a function of workload and scenario (*p < .05, main effects are not shown in the figure)

Air traffic control monitoring task

The detection rate for the air traffic control monitoring task was significantly lower during the alarm floods, compared to the routine periods (14.8% vs. 91.3%, respectively; $X^2(1) = 522.209,\ p < .001$, see ***Figure 3.13***). Since most of the target messages were missed during the floods, there was no sufficient data for an analysis of response time.

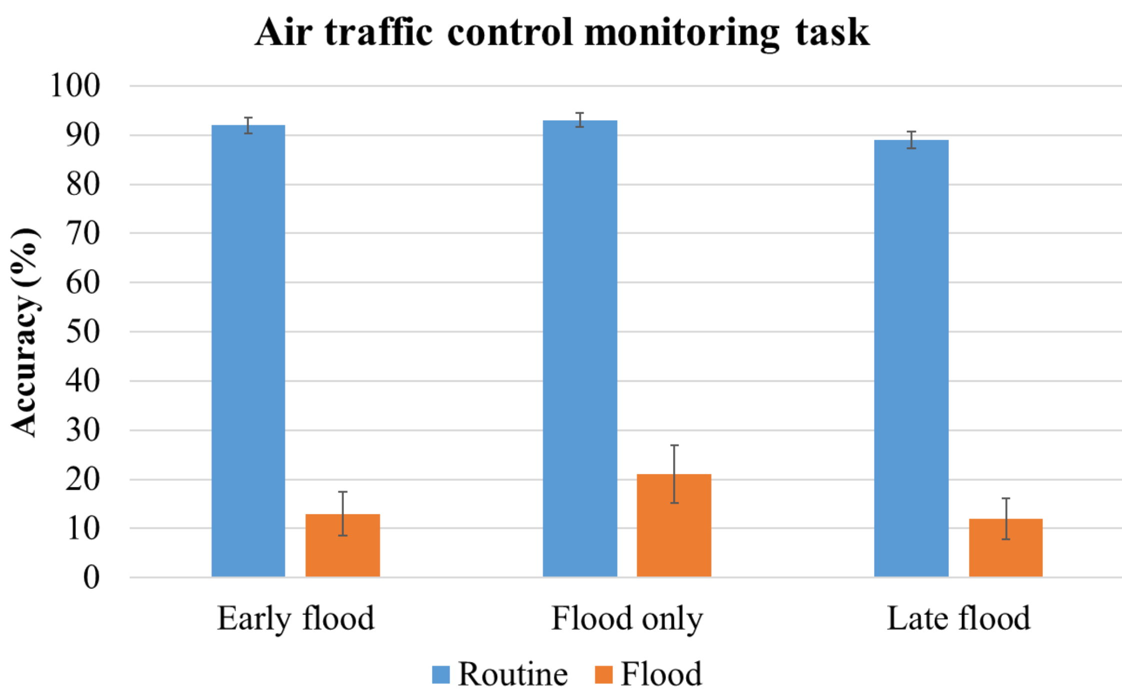

Figure 3.13 Accuracy of the air traffic control monitoring task as a function of workload and scenario

Discussion

The study reported in Chapter 2 demonstrated asynchronous masking effects for visual, but not for auditory alarms. This finding may have been the result of the particular implementation of the two types of alarms and the task set in the experiment. The current study therefore increased task difficulty overall and introduced an air traffic control monitoring task to create competition for auditory attentional resources and thus make the difficulty of alarm detection more comparable for the two modalities.

As a result of these changes, the detection of both visual and auditory alarms was now affected by asynchronous masking. Unlike in the previous experiment, where only the second (visual) alarm in an alarm pair was affected by temporal masking, detection performance now suffered for both the first and second visual and auditory alarms in an alarm pair. In other words, both forward masking (the masking stimulus precedes the target stimulus) and backward masking (the masking stimulus follows the target stimulus) were observed. Previous studies have shown that both forms of masking interrupt the processing of visual stimuli at very short SOAs, ranging from 0 – 200ms (Bachmann & Francis, 2013; Breitmeyer & Ogmen, 2000; Eriksen, 1966; Ogmen, Breitmeyer, & Melvin, 2003). However, at SOAs shorter than 150ms, backward masking was shown to be more detrimental to noticing a visual stimulus than forward masking (Bachmann & Francis, 2013). In contrast, in the present study, the effects of forward and backward masking were nearly equally strong for visual alarms. This observation may be explained by earlier findings showing that the effect of backward masking decreases faster than forward masking as the SOA increases (Schiller, 1966). Thus, it can be expected that forward and backward masking may have comparable effects when the SOA is as long as it was in the current study. For auditory stimuli, it has been reported that forward masking has a stronger effect than backward masking (Elliott, 1962; Wilson & Carhart, 1971). This was confirmed in the current study.

The effect of serial position in an alarm pair on the identification of visual alarms was also different from what was observed in the previous experiment. In that study, accuracy was lower for the second visual alarm in an alarm pair whereas results from the current study show the opposite effect, lower accuracy for the first visual alarm. This difference may be attributed to the different responses required in the two studies. In the first study, after noticing an alarm, participants moved their gaze away from the center of the screen to the response buttons next to the respective drone window; as a result, the second alarm was more likely to appear in their peripheral vision. The detection of the alarm was not necessarily affected as peripheral vision is highly tuned to noticing onsets and luminance changes; however, reading the actual number of the affected drone was not possible in peripheral vision. In the current study, participants did not need to look at a specific location in order to make a response; they could continue to monitor the central window where the second alarm would appear. Thus, performance for the second alarm benefited once the participant had noticed the appearance of the first alarm and their visual attention was allocated to the central window. However, the first alarm may have suffered as a result of participants' attention being focused on other tasks/windows at the time. In this study, the identification of an alarm required color perception which is rather poor in peripheral vision as it relies on regions of the retina where fewer cones (capable of color discrimination) are located (Wickens, Gordon, & Liu, 1998). Auditory alarms are omnidirectional; i.e., they do not require a particular eye or head orientation; as a result, position effects were not observed for these alarms.

The comparison of intra- and cross-modal alarm pairs revealed a stronger backward masking effect for auditory alarms following a visual alarm, compared to intramodal visual-visual masking. This effect was not observed with other modality pairs. It may be explained by the fact that auditory signals have an advantage in terms of storage in short-term memory. A large body of literature suggests that, in an immediate recall task, performance is consistently better with auditory signals, compared to visual ones. This difference is greatest for the most recent item (e.g. Conrad & Hull, 1968). This modality effect has been attributed to a special type of short term memory called the pre-categorical acoustic storage (PAS) (Crowder & Morton, 1969). PAS stores information for the most recent item before recognition; its visual counterpart decays too fast to be utilized in a recall task (Craik, 1969). There is also evidence that crossmodal interference exists between PAS and other types of visual sensory memory. In a visual-auditory alarm pair, it can be expected that the advantage of the second, auditory alarm at the pre-recognition stage may interfere with the storage of visual information and therefore reduce the detection rate and accuracy of the preceding visual alarm. This effect was observed only for V-A pairs but not A-V pairs because PAS has a very limited capacity and may store only the most recent piece of information (Frankish, 1985).

The alarm floods in this study greatly undermined participants' ability to perform their three tasks. The detection rate for alarms plummeted from 98.7% to 64.6%. Response times to delivery consent requests almost doubled (see ***Figure 3.12***), and nearly all air traffic control messages with the call sign "DOI51" were missed (see ***Figure 3.13***). These performance decrements indicate a strong competition for attentional resources. For the delivery consent task, this competition resulted in increased response times because the drone waited until the participant responded. On the other hand, the (auditory) air traffic control messages were transient and had to be noticed immediately. These messages were more likely to be missed than auditory alarms because they were embedded in a stream of very similar (in terms of acoustic characteristics such as pitch and timbre) ATC messages and thus less salient than the auditory alarms (Huang & Elhilali, 2017; Shinn-Cunningham, 2008).

A breakdown of the detection rate for alarms and alarm pairs as a function of their position in an alarm flood (see ***Figure 3.10*** and ***Figure 3.11***) showed that the first and last alarms/alarm pairs were more likely to be noticed, while the detection rate for the remaining alarms fluctuated throughout the flood, with no clear upward or downward trend. To explain this finding, it is helpful to distinguish between the detection and the reporting of a signal. The first alarm/alarm pair probably benefited from the fact that it was a salient signal against the 'quiet' routine phase of the scenario and thus captured attention reliably. It did not interfere with, or was masked by other alarm signals. Thus, it was indeed more likely to be detected. The last alarm/alarm pair in a flood may have shown better performance not because it was more likely to be detected than alarms in the middle of the flood but because its reporting was not interfered with by other, subsequent alarms. Although response times were not measured in the current study, results from the experiment reported in Chapter 2 indicate that the average response time to an alarm was around 2.5 seconds. Thus, there could be a potential overlap between the response to one alarm and the presentation of the following alarm.

One explanation for the poor and decreasing performance throughout the flood may be high levels of stress which were building up as more and more alarms were presented. Stress is known to be associated with a reduced capacity of working memory and a reduced ability to sustain attention – two cognitive processes required for signal detection and reporting (Hancock, 1989; Matthews, 1996; Matthews & Campbell, 1999; Stokes & Raby, 1989).

References

Bachmann, T., & Francis, G. (2013). *Visual masking: Studying perception, attention, and consciousness*. Academic Press. https://doi.org/10.1016/B978-0-12-800250-6.00001-7

Breitmeyer, B. G., & Ogmen, H. (2000). Recent models and findings in visual backward masking : A comparison , review , and update. *Perception & Psychophysics*, *62*(8), 1572–1595.

Colman, A. M. (2015). *A dictionary of psychology*. Oxford University Press, USA.

Conrad, R., & Hull, A. J. (1968). Input modality and the serial position curve in short-term memory. *Psychonomic Science*, *10*(4), 135–136.

Craik, F. I. M. (1969). Modality effects in short-term storage. *Journal of Verbal Learning and Verbal Behavior*, *8*(5), 658–664.

Crowder, R. G., & Morton, J. (1969). Precategorical acoustic storage (PAS). *Perception & Psychophysics*, *5*(6), 365–373.

Elliott, L. L. (1962). Backward and forward masking of probe tones of different frequencies. *The Journal of the Acoustical Society of America*, *34*(8), 1116–1117.

Eriksen, C. W. (1966). Temporal luminance summation effects In backward and forward masking. *Perception & Psychophysics*, *1*(2), 87–92.

Frankish, C. (1985). Modality-specific grouping effects in short-term memory. *Journal of Memory and Language*, *24*(2), 200–209.

Hancock, P. A. (1989). A Dynamic Model of Stress and Sustained Attention. *HUMAN FACTORS*, *31*(5), 519–537. Retrieved from http://journals.sagepub.com/doi/pdf/10.1177/001872088903100503

Huang, N., & Elhilali, M. (2017). Auditory salience using natural soundscapes. *The Journal of the Acoustical Society of America*, *141*(3), 2163–2176.

Matthews, G. (1996). Individual Differences in Driver Stress and Performance. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *40*(12), 579–583. https://doi.org/10.1177/154193129604001205

Matthews, G., & Campbell, S. E. (1999). Individual Differences in Stress Response and Working Memory. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *43*(11), 634–638. https://doi.org/10.1177/154193129904301103

Ogmen, H., Breitmeyer, B. G., & Melvin, R. (2003). The what and where in visual masking. *Vision Research*, *43*(12), 1337–1350. https://doi.org/10.1016/S0042-6989(03)00138-X

Pitts, B. J., Riggs, S. L., & Sarter, N. (2016). Crossmodal Matching: A Critical but Neglected Step in Multimodal Research. *IEEE Transactions on Human-Machine Systems*, *46*(3), 445–450.

Schiller, P. H. (1966). Forward and backward masking as a function of relative overlap and intensity of test and masking stimuli. *Perception & Psychophysics*, *1*(3), 161–164.

Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences*, *12*(5), 182–186. https://doi.org/10.1016/j.tics.2008.02.003

Stokes, A. F., & Raby, M. (1989). Stress and Cognitive Performance in Trainee Pilots. *Proceedings of the Human Factors Society Annual Meeting*, *33*(14), 883–887.

Wickens, C. D., Gordon, S. E., & Liu, Y. (1998). An introduction to human factors engineering.

Wilson, R. H., & Carhart, R. (1971). Forward and backward masking: interactions and additivity. *The Journal of the Acoustical Society of America*, *49*(4B), 1254–1263

Chapter 4

Simultaneous and Asynchronous Masking between Multiple Visual and Auditory Stimuli

The experiments reported in Chapters 2 and 3 established the effects of asynchronous masking on the detection of single visual and auditory alarms and alarm pairs in multitasking conditions. Detection rate was shown to be lower for pairs of alarms presented in temporal proximity, compared to single alarms. Both forward and backward masking were observed but only in high workload conditions. These masking effects occurred at SOA's that were significantly longer (up to 1200ms) than the 200 – 600ms range suggested by basic research on masking effects, such as attentional blink. Based on the findings from the last two experiments, a fixed SOA of 800ms was chosen for the next two studies because this SOA resulted in the lowest detection rates for both visual and auditory alarms.

Given that one of the main goals of this line of research is to examine and support the detection and identification of alarms in an alarm flood, which is defined as more than 10 alarms in a 10-minute period, the experiment reported in this chapter was designed to study masking effects when more than two alarms are presented. To date, very few studies have examined this question. Boot, Becic, & Kramer (2007) studied the detection of up to 4 visual targets in a radar monitoring task, with SOAs ranging from 0 to 300ms. They reported main effects of both SOA and the number of targets in temporal proximity. Detection performance was worse with more targets and at shorter SOAs. An interaction effect between SOA and the number of targets was also reported, such that the effect of SOA on target detection became more pronounced with an increase in the number of targets. Note that, in that study, participants were asked to report the total number of targets but they did not have to identify or discriminate between targets.

The present study went beyond earlier research in that it examined the effect of masking on both the detection and identification of multiple alarms. Specifically, it investigated the role of both asynchronous (which was the focus of the two previous experiments) and simultaneous masking (which becomes more likely with the large numbers of alarms in an alarm flood) in the presence of up to 6 alarms and during an alarm flood. Simultaneous masking is experienced when two stimuli overlap temporally and thus reduce each other's detectability. An example of simultaneous masking is change blindness where observers fail to notice even large changes to objects or scenes when these changes coincide with a brief visual disruption (Simons & Levin, 1997). This phenomenon has been observed intramodally for both visual and auditory stimuli, as well as crossmodally between simultaneous signals in the two sensory channels (Auvray, Gallace, Tan, & Spence, 2007; Gallace, Tan, & Spence, 2006). In fact, auditory stimuli are even more susceptible to simultaneous masking – referred to as change deafness – than visual signals because auditory masking happens at both the attention level and the acoustic level (Carhart, Tillman, & Greetis, 1969; Doll, Hanna, & Russotti, 1992; Fastl & Zwicker, 2006). During an alarm flood, multiple alarm signals that appear simultaneously can be viewed as "transients" masking each other. For example, when two visual alarms appear at the same time, one alarm can be the "transient" and the other can be the "change" to be noticed or vice versa. In this way, simultaneous presentation of alarms increases the risk of missing alarms even when they are expected.

In addition to findings from change blindness studies, research on enumeration (i.e., reporting the number of targets presented) also suggests that people's ability to detect and report multiple concurrent targets is limited (Atkinson, Campbell, & Francis, 1976). Specifically, studies have shown that, when the number of simultaneous targets exceeds four, it becomes much more difficult for the observer to report the correct number of targets. The ability to rapidly recognize (and report the number of) four or fewer targets is called subitizing, while reporting larger numbers of targets relies on counting (slower) or estimation (less accurate) (Revkin, Piazza, Izard, Cohen, & Dehaene, 2008). When the targets are presented rapidly, the observer is forced to estimate the number and, as a result, their report accuracy will decrease. The ability to enumerate is highly related to the detection and reporting of multiple concurrent alarms.

The goals of the current study are:

1. To compare the detection and identification of multiple visual and auditory alarms that are presented concurrently or in close temporal proximity;
2. To investigate the effect of the number of alarms on their detection and identification;
3. For alarms that are presented in close temporal proximity (but not simultaneously), to establish the effect of their serial position on their detection and identification.

The following hypotheses were generated based on findings from previous studies:

1. Alarm detection and identification will suffer more with concurrent alarms than with sequential alarms. This effect is expected to be even more pronounced for auditory alarms.
2. Alarm detection and identification will deteriorate as the number of alarms increases. This effect will be stronger for concurrent alarms.

3. The first and last alarm in an alarm flood is more likely to be detected and correctly identified than other alarms in the same sequence.

Method

Participants

The participants in this study were 15 students recruited from the College of Engineering at the University of Michigan. The required number of participants was determined by performing a statistical power analysis based on the results of the previous experiment. The participants were between 20 to 35 years old (mean age = 23.7 years, SD = 3.1 years; 11 males and 4 females) and had self-reported normal or corrected-to-normal vision. They were also required to have self-reported normal hearing ability and color vision and could not have participated in the experiments reported in Chapter 2 and 3. This study was approved by the University of Michigan Institutional Review Board (UM IRB: HUM00144319).

Apparatus and tasks

The apparatus used in the current study was similar to that used in Chapter 3, which consisted of a computer with a 23-inch LCD monitor, a pair of stereo speakers, a mouse, a keyboard, and an audio recorder. A set of driving pedals (part of the Logitech® MOMO® Racing suite) was added to the setup.

Participants were required to perform three tasks in the context of a package delivery scenario: delivery consent, alarm monitoring, and air traffic control monitoring. The delivery consent task was the same as in Chapter 3. The delivery drones hovered in mid-air whenever they reached a customer residence. Participants were required to search the corresponding video feed for a delivery pad with the letter "H" on it. If the pad was present, participants had to push the confirm button at the top (with a target symbol on it; see ***Figure 4.1***) to give consent to the delivery; else, they were instructed to push the cancel button below (with a target symbol and a line across it).

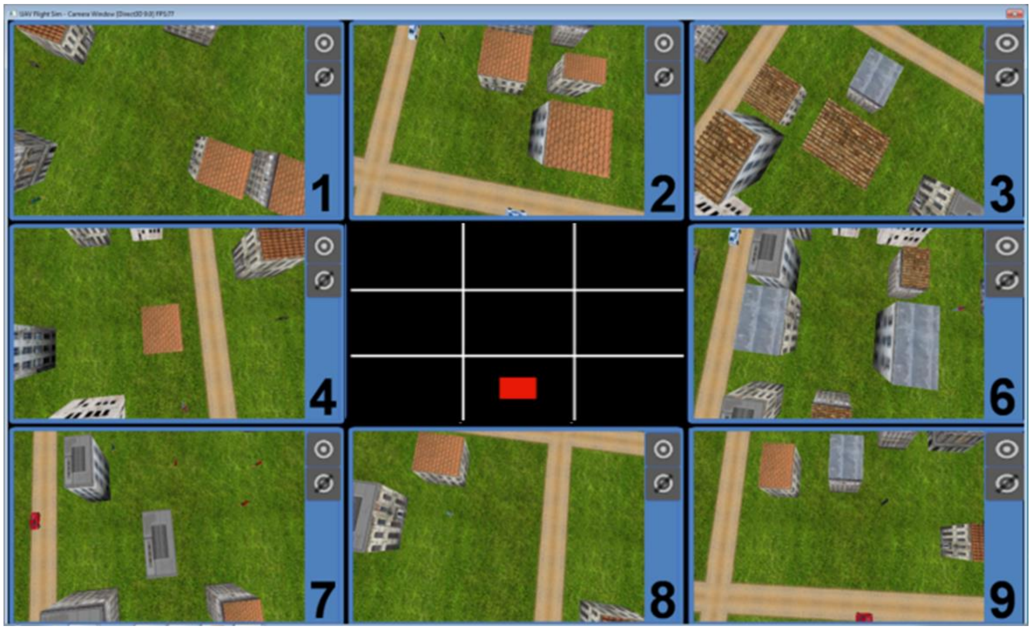

Figure 4.1 Simulation interface

(displaying a red visual alarm for drone 8; required response: R8)

The alarm monitoring task was different from the one described in Chapter 3 in order to include concurrent alarms. Visual alarms were presented in the center of the screen as red (R223, G25, B7; HSL: 5, 94%, 45%), green (R75, G223, B7; HSL: 101, 94%, 45%), or blue (R7, G136, B223; HSL: 204, 94%, 45%; same saturation and brightness) squares in the small grid cell closest to the affected drone. Auditory alarms consisted of a male, female, or children's voice stating the number of the affected drone. These colors and voices were chosen to ensure that they were equivalent in salience but also easily differentiable from each other. Upon detection of a visual or auditory alarms, the participants were required to press the space key on the keyboard as soon as possible (to record response time) and then verbally report the type of alarm (R for red, B for blue, G for green; M for male, F for female, and K for kid), followed by the number of the drone. The alarms were presented either as single alarms or as groups of multiple alarms (alarm clusters) that appeared either concurrently or closely spaced in time. **Table 4.1** shows all combinations of alarms in this experiment. Note that, for auditory alarms, the number of concurrent alarms was limited to two because extensive pilot testing with various types of auditory signals showed that it was impossible for participants to detect and differentiate more than two auditory alarms.

Table 4.1 Types of alarm clusters

Number of alarms	**Visual**		**Auditory**		**Cross-modal**	
	Concurrent	Sequential	Concurrent	Sequential	Concurrent	Sequential
2	Included	Included	Included	Included	1V+1A	1V1A or 1A1V
4	Included	Included		Included		
6	Included	Included		Included		

The air traffic control (ATC) monitoring task was slightly modified also. The ATC recording and the call sign "DOI51" were the same as in Chapter 3; however, the required response was different. Instead of pressing the space key, participants were required to press the foot pedal as soon as possible when they noticed the call sign. This change was made to avoid confusion between the responses to alarms and call signs, while still allowing us to record the response time to both events without requiring participants to look away from the screen.

Experiment design and procedure

The experiment employed a 4×2×2 fractional factorial design. The three factors were cluster size (single, 2, 4, and 6), cluster type (concurrent and sequential), and modality (visual and auditory).

Similar to Chapter 3, the current experiment included three 20-minute scenarios, which were the early flood, late flood, and flood only scenarios. The early flood and late flood scenarios consisted of 17 minutes of low alarm frequency periods and a 3-minute alarm flood (high alarm frequency). The alarm flood scenario consisted of 17 minutes without any alarms and the 3-minute alarm flood. A total of 15 alarm clusters or single alarms (including one of each possible alarm cluster listed in **Table 4.1**, one single visual alarm, and one single auditory alarm) were presented during the low alarm frequency periods of the early and late flood scenarios. During each alarm flood, a total of 30 alarm clusters or single alarms (two of each possible item) were presented. The delivery consent task was presented about once every 40 seconds in each scenario, for a total of 30 times. The air traffic control target message "DOI51" was presented about once every minute throughout each scenario, totaling 20 times.

Upon arrival in the laboratory, participants were consented and received instructions on all three tasks. After reading the instructions, they received four or five 5-minute training sessions to familiarize themselves with the tasks. They needed to achieve a combined accuracy of 90% in the three tasks in order to continue their participation. After the training, the participant completed three 20-minute scenarios, with optional 5-minute breaks between two scenarios. Cross-modal matching was performed before each session using the same method as described in Chapter 3. At the end of the experiment, participants completed a debriefing questionnaire asking them to estimate their own task performance, rate task difficulties, and provide any other feedback they wanted to share. Each participant received $25 compensation, and the three participants with the best combined performance in the three tasks received an additional $50, $30, and $20, respectively. The entire experiment lasted about 100 minutes.

Dependent measures

The dependent measures for the delivery consent task were accuracy and response time. Accuracy was defined as the percentage of correct responses out of all responses. Response time was the time from when the delivery address was reached to when the participant pushed the 'target' or 'no target' button. For the alarm monitoring tasks, detection rate, accuracy, and response time were measured. Detection rate was measured at the cluster level, defined as the number of detected alarms, divided by all alarms in the cluster. Response time was also measured at the cluster level, as the time from the onset of the alarm cluster/single alarm to the time when the participant first pressed the space key. Accuracy was defined at the alarm level, as the percentage of correctly reported alarms out of all alarms. The dependent measures for the air traffic control monitoring task were detection rate and response time. Detection rate was defined as the percentage of detected call signs, out of all call signs presented. Response time was defined as the time from the presentation of the call sign to the time when the foot pedal was pushed.

Results

The analyses of detection rate and accuracy for visual and auditory alarms were conducted separately using mixed model binary logistic regression. Univariate analysis of variance (ANOVA) was used to analyze the response time for visual and auditory alarms. In all the analyses, cluster type and cluster size were included as fixed effect factors and subject ID was included as a random effect factor; multiple comparisons were corrected with sequential Bonferroni procedures; the significance level was set at $p < .05$ and only significant results are reported here.

Detection rate

As shown in ***Figure 4.2***, the detection rate for visual alarms in an alarm cluster decreased significantly as the number of alarms in the cluster increased ($F(3, 833) = 13.888, p < .001$). There was no significant difference in detection performance between concurrent and sequential clusters, but an interaction between cluster type and cluster size was observed. The detection rate dropped significantly more when the size of the cluster reached 6 concurrent, as compared to 6 sequential alarms ($F(2, 833) = 3.312, p = .037$). The detection rate for auditory alarms (see ***Figure 4.3***) also decreased with two concurrent auditory alarms ($F(1, 595) = 16.530, p < .001$) and as the number of sequential alarms in an alarm cluster increased ($F(3, 595) = 9.813, p < .001$), compared to single auditory alarms.

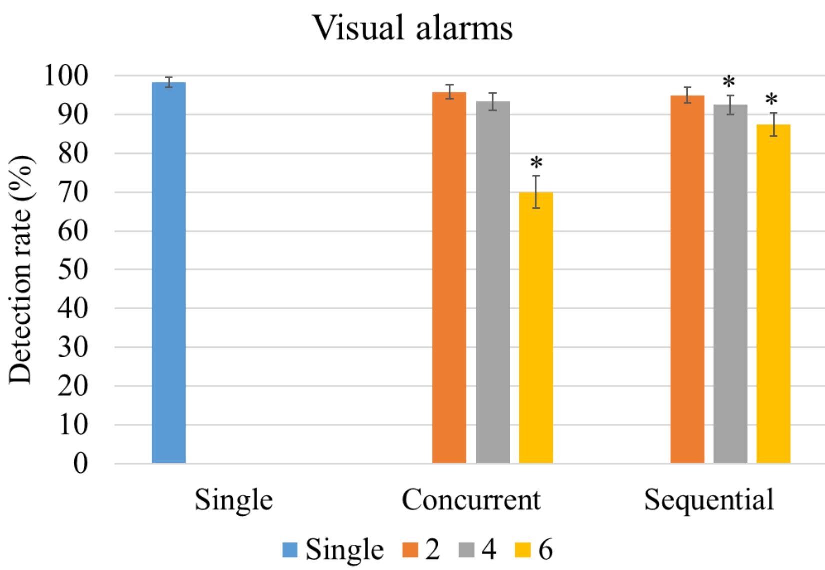

Figure 4.2 Detection rates of visual alarms as a function of cluster type and size

(*p < .05 compared to single alarms)

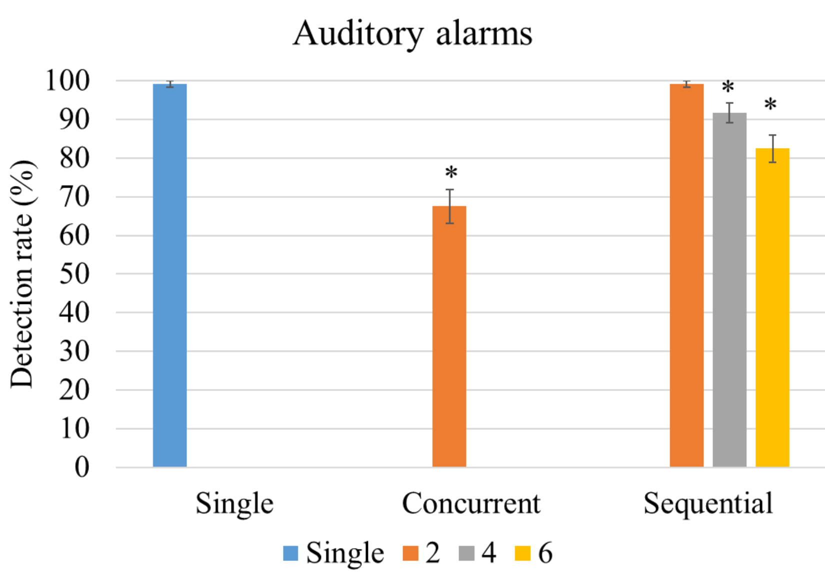

Figure 4.3 Detection rates of auditory alarms as a function of cluster type and size

(*p < .05 compared to single alarms)

Accuracy

As shown in ***Figure 4.4***, the accuracy for visual alarms decreased as a function of cluster size. This effect was more pronounced for concurrent alarms, compared to sequentially presented alarm signals. A further breakdown of accuracy for individual alarms in sequential visual alarm clusters shows that accuracy was significantly affected by cluster size ($F(2, 678) = 3.562, p = .029$; see ***Figure 4.5***), such that accuracy decreased as the number of alarms in a cluster increased. In clusters that included 6 sequential visual alarms, accuracy was higher for the first and last alarms in a cluster, compared to all other positions ($t(718) = 3.422$, $p < .001$).

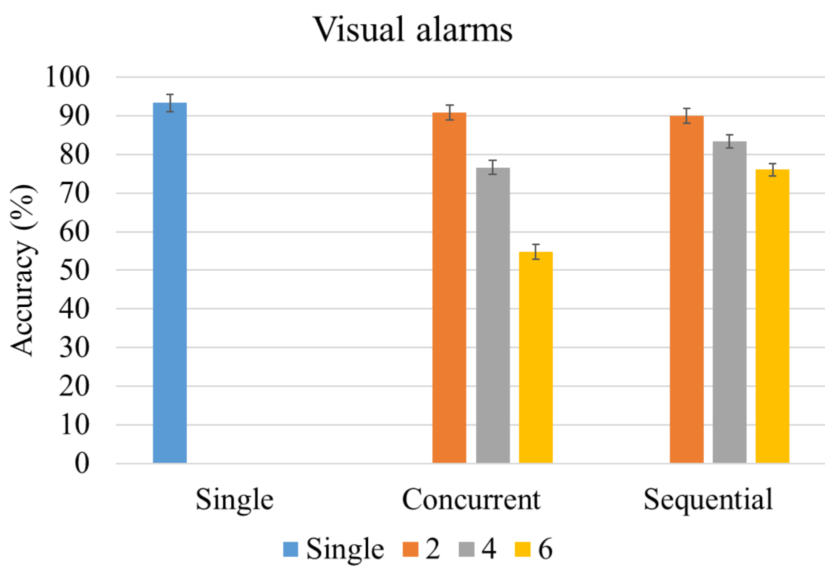

Figure 4.4 Accuracy of visual alarms as a function of cluster type and size

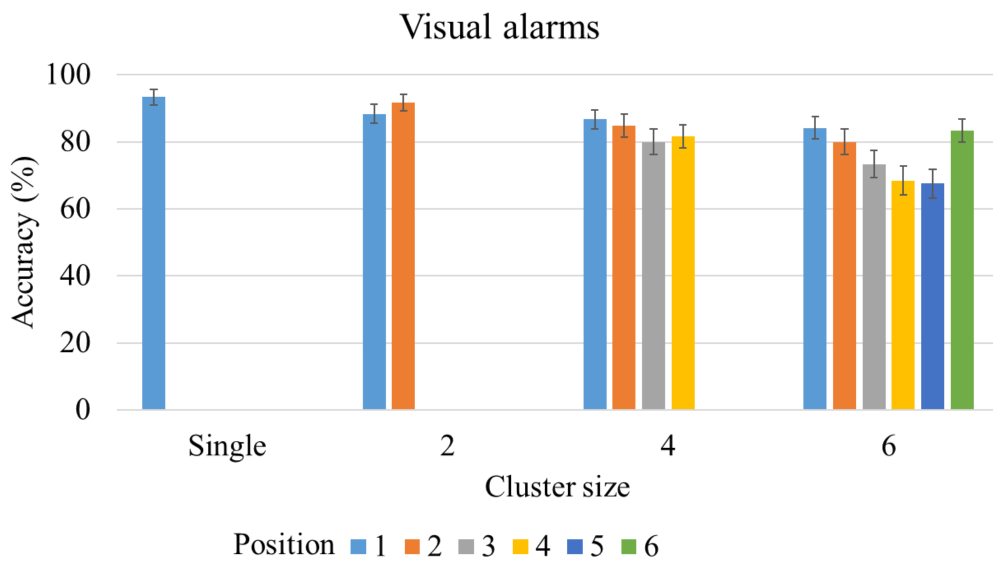

Figure 4.5 Accuracy of visual alarms as a function of cluster size and alarm position

For auditory alarms, accuracy also decreased with an increase in cluster size (see ***Figure 4.6*** and ***Figure 4.7****)*. The lowest accuracy was observed for two concurrent auditory alarms. A further breakdown of individual auditory alarms in sequential clusters revealed a significant main effect of cluster size ($F(2, 678) = 10.736, p < .001$), such that accuracy was lower for clusters that had more alarms. In alarm clusters with 4 or 6 alarms, the first and last alarm signals showed the highest accuracy ($t(1198) = 6.502$, $p < .001$).

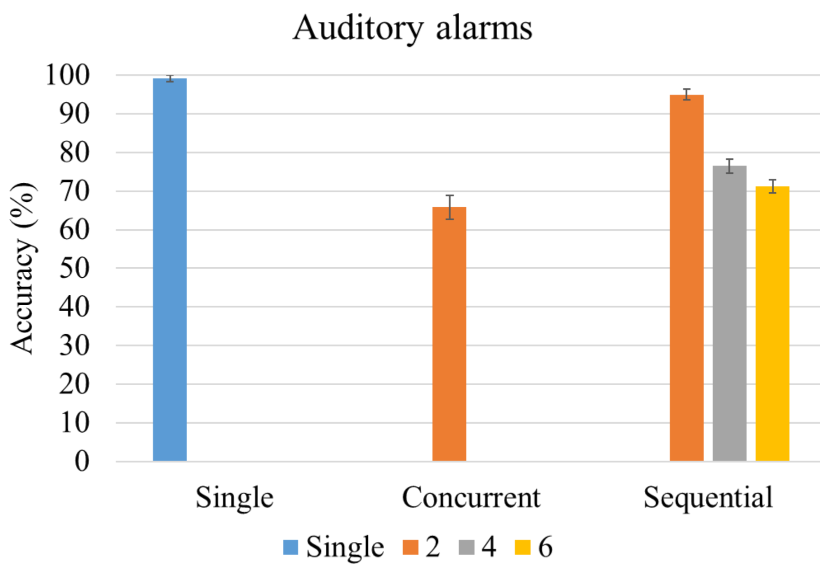

Figure 4.6 Accuracy of auditory alarms as a function of cluster type and size

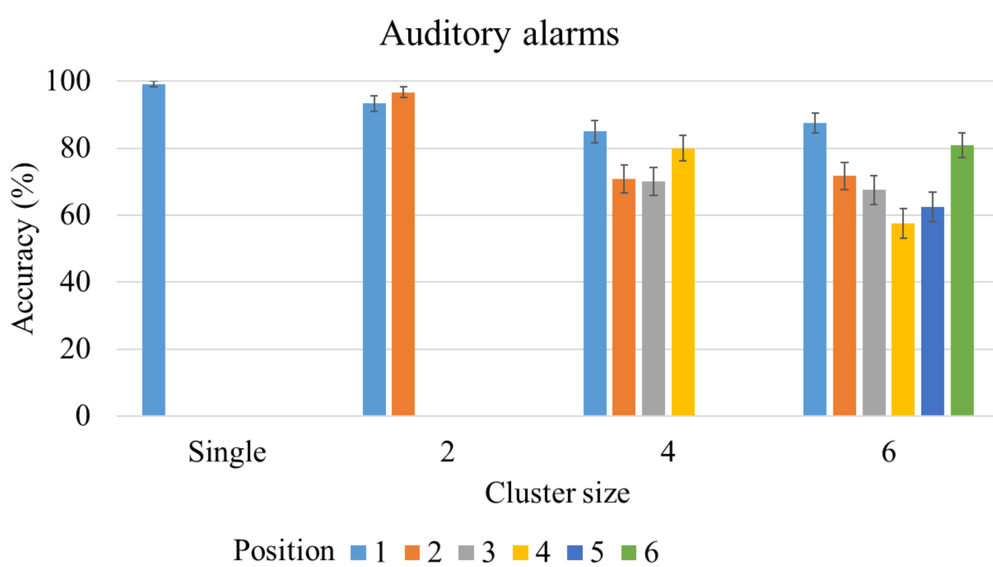

Figure 4.7 Accuracy of auditory alarms as a function of cluster size and alarm position

Response time

The response time to visual alarms in a cluster increased with cluster size (see ***Figure 4.8***; $F(3,677) = 6.139, p = .001$). This effect was more pronounced for concurrent alarm clusters than for sequential ones ($F(2,677) = 4.768, p = .017$). The response time to auditory alarms in clusters was affected only by the type of cluster. Specifically, the response time to concurrent alarm clusters was significantly longer than to sequential ones overall ($F(1,677) = 5.261, p = .038$). Pairwise comparisons between each group of sequential alarm clusters showed that the response time to the 6-alarm clusters was significantly longer than to the 2- and 4-alarm clusters.

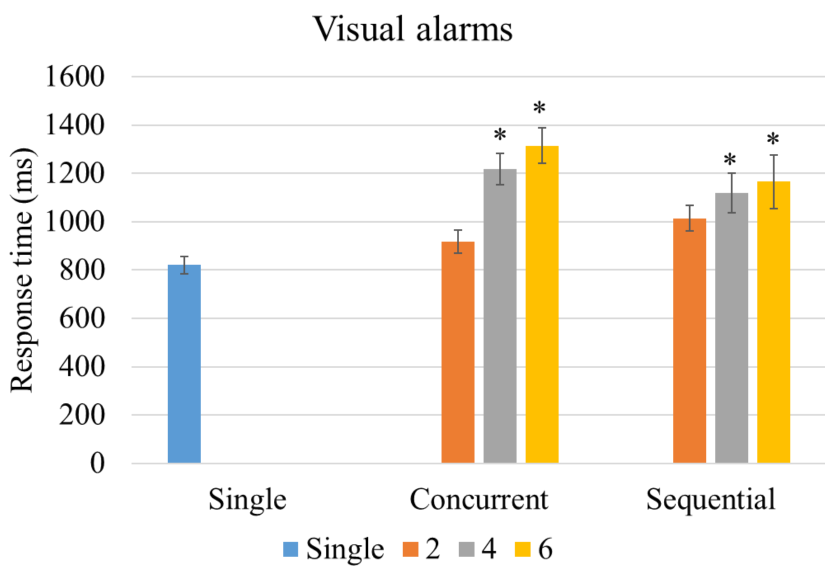

Figure 4.8 Response time to visual alarms as a function of cluster type and size

(*p < .05, compared to single alarms)

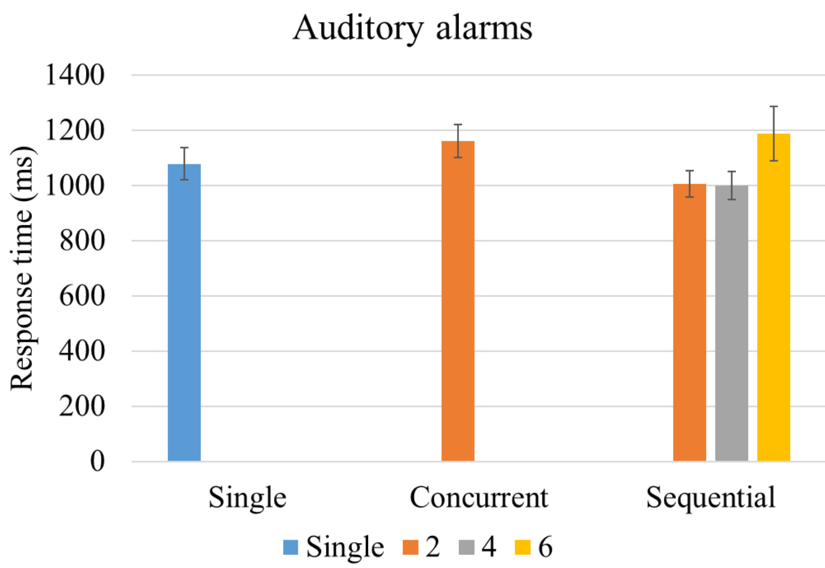

Figure 4.9 Response time to auditory alarms as a function of cluster type and size

Subjective ratings

Participants were asked to estimate the percentage of alarms they had detected and reported correctly, detected but reported incorrectly, and missed. Their estimates are compared to their actual performance in ***Figure 4.10*** which shows that they underestimated both their detection rate and accuracy.

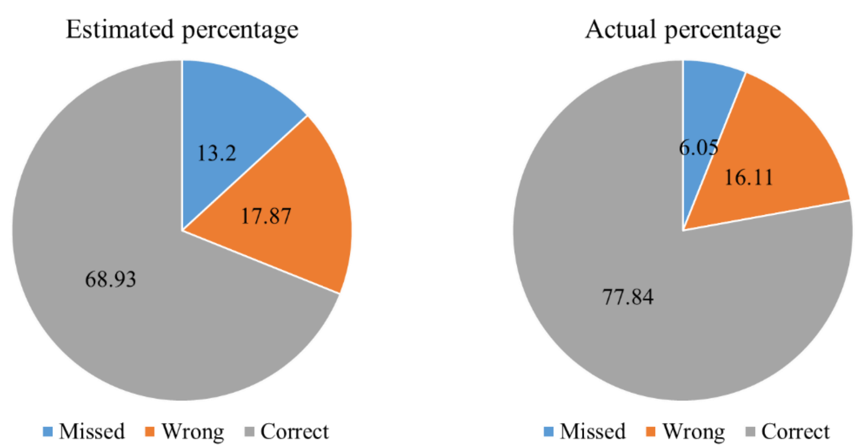

Figure 4.10 Estimated and actual performance of the alarm monitoring task

Participants also rated the difficulty of detecting different types of alarm clusters. Concurrent visual alarms were rated as being equally or slightly more difficult to detect than sequential visual alarms. The difficulty rating of sequential auditory alarm cluster increased the most rapidly in the three types of alarm clusters. However, overall speaking, the ratings did not differ significantly between the three types. The size of the cluster seemed to play a dominant role. Concurrent auditory pairs were rated as a lot more difficult than other 2-alarm clusters, but not as difficult as the 5- and 6-alarm clusters.

Crossmodal alarms

There were three different types of crossmodal alarm clusters, all of which consisted of two alarms: concurrent visual-auditory, sequential visual-auditory, or sequential auditory-visual. Neither detection rate, accuracy nor response time differed significantly for the three types of alarm clusters. The detection rates for these three types of alarm clusters were 95.8%, 95.0%, and 94.2%, respectively. Accuracy was 91.7%, 91.3%, and 90.0%, respectively. And the response times to theses alarm clusters were 1037.3ms, 944.9ms, and 1008.9ms, respectively. However, participants' difficulty ratings were much higher for crossmodal (4.00, 3.13, and 3.47 respectively), compared to intramodal pairs (all of which were between 1 and 2, except for concurrent auditory pairs which were rated 4.80).

Alarm flood analysis

A comparison of three performance metrics between the low alarm frequency periods and the alarm floods is shown in ***Figure 4.11***. The detection rate was slightly lower during alarm floods, but the difference was not significant ($F(1, 1798) = 1.519, p = .218$). However, accuracy was significantly lower during the alarm floods ($F(1, 5518) = 23.947, p < .001$), and response time was significantly shorter ($F(1, 1798) = 13.710, p < .001$). A detailed analysis of each alarm cluster was conducted to reveal the change of performance during the alarm flood, as shown in ***Figure 4.12***. The performance metric used was a normalized final report rate. The final report rate was calculated as the percentage of alarms that were both detected and correctly identified. This metric was then averaged by each type of alarm cluster, and this average performance was used to normalize the performance of each alarm cluster in the alarm flood. The best performance was observed with the last alarm cluster and the worst performance was observed with clusters 7 and 8.

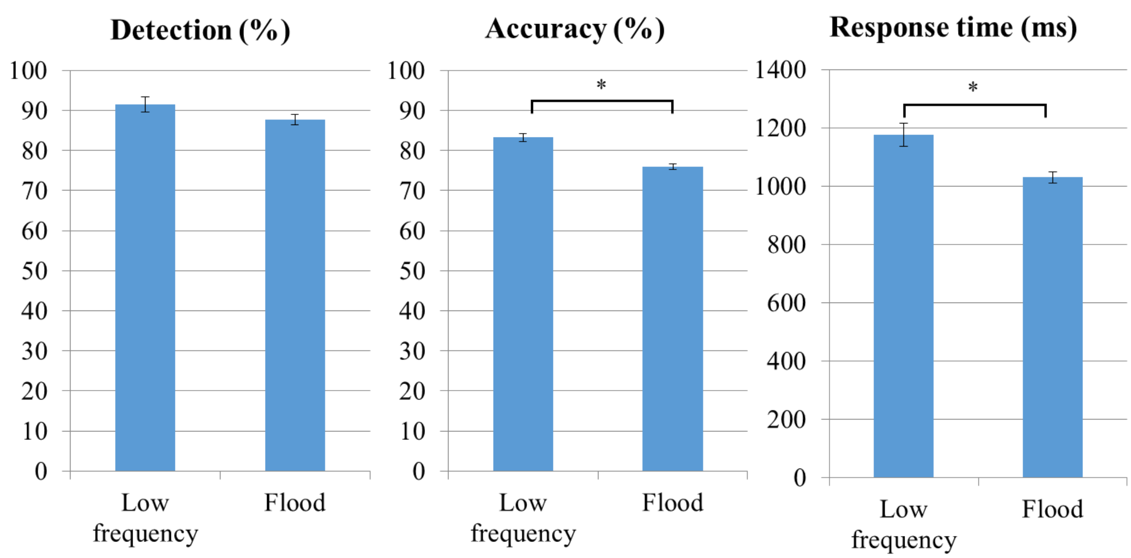

Figure 4.11 Performance comparison between alarm floods and low alarm frequency periods

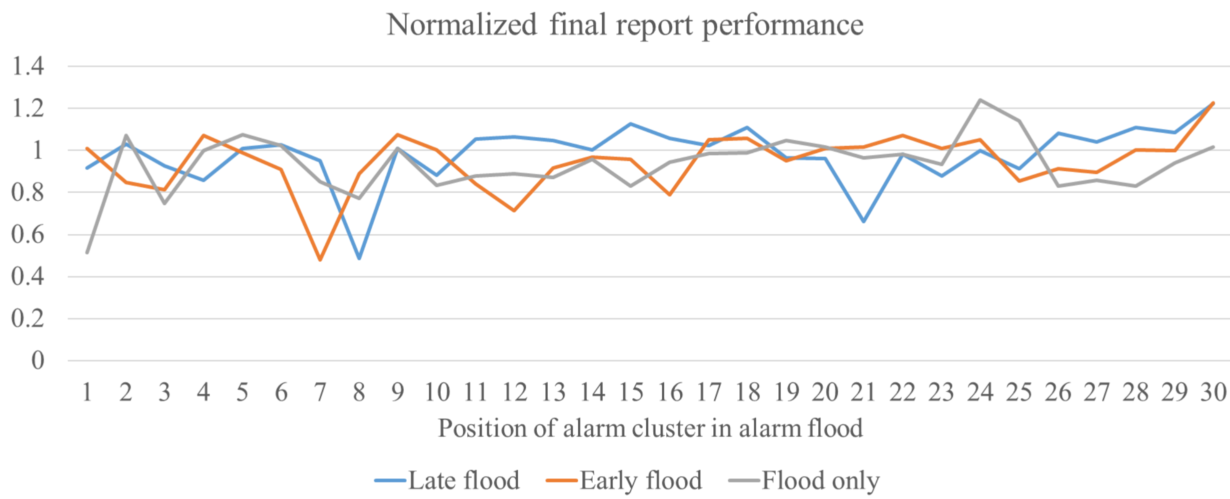

Figure 4.12 Normalized alarm report performance as a function of cluster position in alarm flood

Discussion

The current study was based on and expanded on the research reported in Chapters 2 and 3 which focused on asynchronous masking effects on the detection of single alarms and alarm pairs. In this experiment, concurrent alarms were introduced to study the effect of simultaneous masking, and the number of alarms in temporal proximity was increased from two to six. These changes were made to investigate performance in a context more representative of an alarm flood.

The results from this study indicate that the detection rate and accuracy for visual alarms decreased as the number of alarms in a cluster increased. As expected, this performance decrement was more pronounced with concurrent alarms than with sequential ones. This finding is in agreement with the few studies that, to date, have compared synchronous and asynchronous masking. For example, Beanland & Pammer (2012) investigated attentional blink (asynchronous) and inattentional blindness (the failure to notice unexpected visual targets when attention is engaged with other targets; synchronous) in two experiments. Although the tasks in these studies were not entirely equivalent in nature, the authors reported detection rates of 26% for the inattentional blindness task and 62% for the attentional blink task. This suggests a much stronger performance effect of simultaneous masking. A more equivalent comparison was made by Boot et al. (2007) who examined various numbers of targets and SOA's in the context of a target detection task using a simulated radar monitoring interface. They reported lower detection rates with simultaneous targets than with sequential targets, especially when the number of targets exceeded three. Detection rates in their study were comparable for pairs of targets with SOA's ranging from 0 to 300ms. Similarly, in the present study, participants' performance deteriorated rapidly with an increase in cluster size for concurrent visual alarm clusters; however, their performance was very similar for concurrent and sequential alarm clusters when the cluster size was small (2 alarms). One possible explanation for these findings is that two visual alarms are processed as one visual stimulus, thus avoiding or reducing competition for attentional resources (Akyürek & Hommel, 2005; Shapiro, Schmitz, Martens, Hommel, & Schnitzler, 2006). For 4-alarm clusters, the detection rate remained similar, but accuracy was lower for concurrent alarms. This result is consistent with findings from studies on enumeration of targets. As mentioned earlier, these studies have shown that it is possible for people to report the number of targets very quickly when that number is no greater than four (Mandler & Shebo, 1982; Revkin et al., 2008). This ability, termed subitizing as opposed to counting, has been likened to the recognition of figural patterns that contain small numbers of targets (Von Glasersfeld, 1982). Since subitizing is based on processing targets as one figural pattern, it does not support the identification and differentiation of multiple alarm signals. As a result, the detection rate for 4-alarm concurrent visual alarm clusters was comparable to that for sequential alarm clusters because the participants could quickly subitize the number of alarms, but the accuracy for 4-alarm concurrent visual alarm clusters was lower than for sequential alarm clusters because subitizing does not support identification.

One difference between the visual and auditory modalities that was observed in this study was that both detection rate and accuracy were higher for auditory alarms in case of 2-alarm clusters but decreased more rapidly with an increase in cluster size for sequential auditory alarm clusters, compared to visual ones. There are two possible explanations for this modality difference. It could result from the smaller capacity of the auditory working memory, compared to visual working memory (Saults & Cowan, 2007), and/or it could be attributed to the increased interference between the auditory presentation of the stimulus and the required verbal response. Generally, there is a tendency toward better performance when stimulus and response are compatible, i.e., share the same processing code (Wickens, Vidulich, & Sandry-Garza, 1984). However, this benefit may disappear due to the temporal overlap between stimulus and response when the number of alarms increases.

The auditory concurrent alarm clusters led to the worst performance – an average detection rate of only 67.5% – amongst all cluster types, even though these clusters consisted of only two alarms. As mentioned before, auditory alarms are more susceptible to simultaneous masking because their detection suffers not only due to attentional limitations, but also because of acoustic interference between different sounds, at the sensory stage. Unlike visual perception which can attend to spatially independent objects at the same time, auditory perception relies on receptors that are highly sensitive to differences in frequency (Bolton, Edworthy, & Boyd, 2018; Greenwood, 1971). The alarm signals used in the current study were close in their base frequency and identical in their duration; as a result, the simultaneous masking effect was very strong. Similar problems could arise in many application domains (such as medicine) where auditory displays or alarms are widely used but not necessarily well coordinated as they are associated with systems that are developed by different companies (Momtahan, Hetu, & Tansley, 1993).

Interestingly, despite their poor detection performance, participants did not assign a very high difficulty rating to the detection of two concurrent auditory alarms. In fact, it was rated as being easier than the detection of 6-alarm sequential auditory clusters (4.80 vs. 7.07) even though performance contradicted these perceptions (67.5% vs. 82.5%). This suggests that participants' ratings were based on the perceived effort required to perform the task, rather than its actual difficulty. Since participants did not receive feedback on their performance throughout the experiment, they were not aware of the fact that they had actually missed many of the concurrent auditory alarm pairs.

Participants overestimated the difficulty of detecting alarms in crossmodal alarm pairs. Performance was actually comparable to intramodal visual pairs but participants were less comfortable with crossmodal pairs, especially concurrent ones. This result seems to contradict MRT which predicts that processing information in different sensory modalities reduces interference and (perceived) task difficulty and workload. For other alarm clusters (concurrent visual, sequential visual, and sequential auditory), participants' difficulty ratings were largely proportional to the size of the cluster and did not differ much even though performance did. Taken together, these findings suggest that subjective ratings should not be used as a basis for alarm design because system safety should always be the priority of alarm system design. Some other studies have similarly reported dissociation between subjective preference and performance due to various reasons, including familiarity, excitement, comfort, and effort (Andre & Wickens, 1995).

For both visual and auditory alarms, the response time was longer with larger alarm clusters and also increased more rapidly with concurrent alarm clusters. Because response time was measured at the cluster level, the response time to a cluster is essentially the response time to the first alarm in the cluster. The processing of the following or other concurrent alarm signals delayed the response. Note that this effect was different from the delay effects observed in Chapter 2, where the manual response required the participant to look at the button to be pushed. In the current study, the response was made by pressing the space key, and participants were instructed to rest their left hand on the space key at all times. Participants could respond to the first alarm while processing subsequent alarms because it did not require a re-orientation of their gaze. Therefore, the delay observed in this study is likely the result of cognitive interference, rather than a motoric or oculomotor constraint. This delay was not reported in previous studies as they did not ask participants to respond as quickly as possible (Beanland & Pammer, 2012; Boot et al., 2007). However, this delay is worrisome when viewed not in absolute terms (e.g., 821ms for single visual alarms versus 1314ms for six concurrent visual alarms) but as a 60% increase in response time. In real-world environments where operators multitask more complex tasks, absolute response times tend to be longer, and a 60% increase can result in operationally relevant delays.

For clusters consisting of 4 or 6 sequential alarms, the first and last alarms were reported more accurately than other alarms in the same cluster. This was true for both visual and auditory alarms. These results are consistent with the pattern observed in free recall tasks, where subjects are presented with a series of items and required to recall the items, regardless of the order of presentation. Items closer to the beginning or the end of the sequence are more likely to be recalled. These effects were termed the primacy effect and the recency effect, respectively (Murdock Jr, 1962). The primacy effect was explained as a result of covert rehearsal of the items in the sequence. The closer an item is to the beginning of the sequence, the more it gets rehearsed and possibly enters long-term memory. Therefore, the first items in a sequence are more likely to be recalled (Fischler, Rundus, & Atkinson, 1970; Laming, 2010). The recency effect was attributed to the fact that these items had just entered into short-term memory and were thus more accessible (Tzeng, 1973). There were two ways in which the participants reported a cluster of multiple alarms: report immediately after each alarm or wait until the end to avoid reporting during the cluster. In the first way, the report of one alarm overlapped and interfered with the presentation of the following alarms, thus giving the early alarms an advantage because they were less affected by such interference. In the second way, the early alarms had a similar advantage as those in the free recall studies because they were rehearsed more. The advantage of the last alarm was also comparable to that in the free recall studies because it was the last item to enter the short-term memory and therefore easier to retrieve. Therefore, in larger clusters, attention is not the only resource that limited report accuracy. The capacity of short-term memory could also play an important role in this process.

Finally, the comparison between alarm floods and routine operations yielded results that were mostly consistent with those from Chapters 2 and 3. During the alarm floods, the detection rate was slightly, but not significantly lower; accuracy was lower, and response times were shorter. These results confirm that participants change their response strategy when exposed to very large numbers of alarms in close temporal proximity, leading to a speed-accuracy tradeoff. To keep up with the pace of the alarm flood, they sped up their responses, but this was achieved at the cost of accuracy. This strategy could lead to potentially catastrophic misdiagnoses of critical alarms during an alarm flood in a real-world setting.

The performance changes during the alarm floods differed from those seen in the study reported in Chapter 3. The advantage of the last alarm/alarm cluster was still observed, but the first alarm cluster did not show better detection performance. This difference is possibly due to the increase in the number of alarms in each cluster (specifically, in all three scenarios, the second alarm cluster was a cluster of 6 alarms) and, as a result, the rapidly increased interference between alarms. Therefore, the first alarms did not benefit from a lower level of stress and interference between alarm clusters because the difficulty of the task was different. On the contrary, the performance for the first alarm in the flood only scenario was much lower compared to the other alarms. Participants might be surprised by the onset of the alarm flood in the flood only scenario because there were no other alarms in the entire scenario. Previous research has also shown that unexpected events can lead to surprised or even startled reactions, which can slow down the processing of unexpected signals and even hinder the ability to direct attention properly (Barnett, Wong, Westley, Adderley, & Smith, 2012; Landman, Groen, van Paassen, Bronkhorst, & Mulder, 2017). This was not observed in the previous study because the first alarm to appear in the alarm flood only scenario was a single visual alarm in the previous study, while a pair of concurrent auditory alarm was the first to appear in the current study. The difficulty of detecting concurrent auditory alarms have been shown to be much higher than a single alarm. When the participant was least prepared, the steep increase of difficulty resulted in strong surprising or startling effects rather than performance benefits.

It should be noticed that in all three scenarios, as well as in the previous study, performance was the lowest for alarms or alarm clusters that appeared about 40 – 50s into the alarm floods. In the current experiment, these were the 7^{th} – 8^{th} alarm clusters out of all 30 alarms or alarm clusters; in the previous experiment, they were the 14^{th} – 16^{th} alarms or alarm pairs. Initially, participants probably tried to cope with the demands associated with the alarm flood. However, this was difficult because of the unexpected sudden onset of the alarm flood which meant that they were not mentally ready. Their stress level continued to increase as they realized that they failed to keep up with the pace of the alarms. At about 40 to 50 seconds into the alarm flood, participants' arousal level had likely increased and they may also have invested more effort into the task at this point, leading to better performance. Similar to what has been discussed in Chapter 2, an appropriate level of arousal (neither too high or too low)is required to achieve optimal performance (Matthews & Davies, 1998; Molloy & Parasuraman, 1996; Warm, Parasuraman, & Matthews, 2008; Wiener, Curry, & Faustina, 1984; Yerkes & Dodson, 1908).

The findings from the studies reported in Chapters 2-4 highlight the importance of avoiding certain temporal patterns in alarm presentation. An SOA of 800ms was shown to be particularly detrimental to detection performance. The simultaneous presentation of six visual alarms also resulted in an increased risk of missed alarms. And concurrent auditory alarms were very difficult to detect and identify. Intelligent alarm systems should be designed with these challenges in mind. Such a human-centered design approach can complement previous efforts which have focused on adjusting the threshold of alarms (the value at which alarms are triggered) based on the context of system operations (Pollard, 2010; Schmid et al., 2017; Welch, 2011). These systems are "adaptive" to the state of the system but not to the cognitive limitations of the human operator.

References

Akyürek, E. G., & Hommel, B. (2005). Target integration and the attentional blink. *Acta Psychologica*, *119*(3), 305–314. https://doi.org/10.1016/j.actpsy.2005.02.006

Andre, A. D., & Wickens, C. D. (1995). When Users Want What's not Best for Them. *Ergonomics in Design*, *3*(4), 10–14. https://doi.org/10.1177/106480469500300403

Atkinson, J., Campbell, F. W., & Francis, M. R. (1976). The magic number 4±0: A new look at visual numerosity judgements. *Perception*, *5*(3), 327–334.

Auvray, M., Gallace, A., Tan, H. Z., & Spence, C. (2007). Crossmodal change blindness between vision and touch. *Acta Psychologica*, *126*(2), 79–97. https://doi.org/10.1016/j.actpsy.2006.10.005

Barnett, J., Wong, W., Westley, D., Adderley, R., & Smith, M. (2012). Startle reaction: Capturing experiential cues to provide guidelines towards the design of realistic training scenarios. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 56, pp. 2477–2481). SAGE Publications Sage CA: Los Angeles, CA.

Beanland, V., & Pammer, K. (2012). Minds on the blink: The relationship between inattentional blindness and attentional blink. *Attention, Perception, & Psychophysics*, *74*(2), 322–330.

Bolton, M. L., Edworthy, J., & Boyd, A. D. (2018). A Formal Analysis of Masking Between Reserved Alarm Sounds of the IEC 60601-1-8 International Medical Alarm Standard. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *62*(1), 523–527. https://doi.org/10.1177/1541931218621119

Boot, W. R., Becic, E., & Kramer, A. F. (2007). Temporal Limitations in Multiple Target Detection in a Dynamic Monitoring Task. *Human Factors*, *49*(5), 897–906. https://doi.org/10.1518/001872007X230244.

Carhart, R., Tillman, T. W., & Greetis, E. S. (1969). Perceptual Masking in Multiple Sound Backgrounds. *The Journal of the Acoustical Society of America*, *45*(3), 694–703. https://doi.org/10.1121/1.1911445

Doll, T. J., Hanna, T. E., & Russotti, J. S. (1992). Masking in Three-Dimensional Auditory Displays. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *34*(3), 255–265. https://doi.org/10.1177/001872089203400301

Fastl, H., & Zwicker, E. (2006). *Psychoacoustics: facts and models* (Vol. 22). Springer Science & Business Media.

Fischler, I., Rundus, D., & Atkinson, R. C. (1970). Effects of overt rehearsal procedures on free recall. *Psychonomic Science*, *19*(4), 249–250.

Gallace, A., Tan, H. Z., & Spence, C. (2006). Numerosity judgments for tactile stimuli distributed over the body surface. *Perception*, *35*(2), 247–266.

Greenwood, D. D. (1971). Aural combination tones and auditory masking. *The Journal of the Acoustical Society of America*, *50*(2B), 502–543.

Laming, D. (2010). Serial position curves in free recall. *Psychological Review*, *117*(1), 93.

Landman, A., Groen, E. L., van Paassen, M. M. (René), Bronkhorst, A. W., & Mulder, M. (2017). Dealing With Unexpected Events on the Flight Deck: A Conceptual Model of Startle and Surprise. *Human Factors*, *59*(8), 1161–1172. https://doi.org/10.1177/0018720817723428

Mandler, G., & Shebo, B. J. (1982). Subitizing: an analysis of its component processes. *Journal of Experimental Psychology: General*, *111*(1), 1.

Matthews, G., & Davies, D. R. (1998). Arousal and vigilance: Still vital at fifty. In *Proceedings*

of the Human Factors and Ergonomics Society Annual Meeting (Vol. 42, pp. 772–776). SAGE Publications Sage CA: Los Angeles, CA.
Molloy, R., & Parasuraman, R. (1996). Monitoring an automated system for a single failure: Vigilance and task complexity effects. *Human Factors*, *38*(2), 311–322.
Momtahan, K., Hetu, R., & Tansley, B. (1993). Audibility and identification of auditory alarms in the operating room and intensive care unit. *Ergonomics*, *36*(10), 1159–1176.
Murdock Jr, B. B. (1962). The serial position effect of free recall. *Journal of Experimental Psychology*, *64*(5), 482.
Revkin, S. K., Piazza, M., Izard, V., Cohen, L., & Dehaene, S. (2008). Does subitizing reflect numerical estimation? *Psychological Science*, *19*(6), 607–614.
Saults, J. S., & Cowan, N. (2007). A central capacity limit to the simultaneous storage of visual and auditory arrays in working memory. *Journal of Experimental Psychology: General*, *136*(4), 663.
Shapiro, K., Schmitz, F., Martens, S., Hommel, B., & Schnitzler, A. (2006). Resource sharing in the attentional blink. *Neuroreport*, *17*(2), 163–166. https://doi.org/10.1097/01.wnr.0000195670.37892.1a
Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences*, *1*(7), 261–267.
Tzeng, O. J. L. (1973). Positive recency effect in a delayed free recall. *Journal of Verbal Learning and Verbal Behavior*, *12*(4), 436–439.
Von Glasersfeld, E. (1982). Subitizing: The role of figural patterns in the development of numerical concepts. *Archives de Psychologie*.
Warm, J. S., Parasuraman, R., & Matthews, G. (2008). Vigilance requires hard mental work and is stressful. *Human Factors*, *50*(3), 433–441.
Wickens, C. D., Vidulich, M., & Sandry-Garza, D. (1984). Principles of SCR compatibility with spatial and verbal tasks: The role of display-control location and voice-interactive display-control interfacing. *Human Factors*, *26*(5), 533–543.
Wiener, E. L., Curry, R. E., & Faustina, M. Lou. (1984). Vigilance and task load: In search of the inverted U. *Human Factors*, *26*(2), 215–222.
Yerkes, R. M., & Dodson, J. D. (1908). The relation of strength of stimulus to rapidity of habit-formation. *Journal of Comparative Neurology*, *18*(5), 459–482.

Chapter 5

The Effects of Masking on the Detection and Identification of Alarms: Summary of Findings from Experiments 1 – 3

Chapters 2 to 4 reported three studies that aimed to establish the effects of various factors – SOA, the number of alarms, and the serial position of an alarm – on alarm detection and identification during periods of low and high alarm frequency (routine operations vs. alarm floods). Specifically, the three experiments were carried out with the following goals in mind: the first two studies examined whether asynchronous masking is experienced with single and pairs of visual and auditory alarms presented in a multi-task setting. They also established the SOAs at which such masking is experienced. The third study investigated both simultaneous and asynchronous masking of visual and auditory alarm signals as a function of the number and serial position of alarms in clusters of up to 6 sequential alarms.

Experiments 1 and 2: Asynchronous masking of single or pairs of alarms in multi-task settings

The main findings of experiment 1 were:

- Asynchronous masking was observed for visual but not auditory alarms
- Only forward masking was observed

- The effective SOAs included 600, 1000, and 1200ms, with 1000ms leading to the worst detection and identification performance

Main findings from the second study which included a secondary auditory task and a different response mechanism that did not require reorientation of gaze were:

- Asynchronous masking was now observed for both visual and auditory alarms
- SOAs of 600 and 800ms led to the worst detection performance for visual alarms
- The worst performance for auditory alarms was observed with SOAs of 200, 800, and 1000ms

In the following studies, only one SOA – 800ms – was used to induce performance breakdowns because this SOA led to the worst detection performance for both visual and auditory alarms under high workload.

Experiment 3: Simultaneous versus asynchronous masking and serial position of alarms

This study included alarm clusters that consisted of up to 6 alarms.

- For both concurrent and sequential visual alarm clusters, masking was stronger as the number of alarms increased. This effect was most pronounced with 6 concurrent visual alarms
- The same effects were observed for auditory alarm clusters; however, note that only single and pairs of concurrent auditory alarms were employed
- Response times to visual alarms were longer with an increase in the number of alarms. This effect was again stronger with concurrent, compared to sequential alarms
- The response time to auditory alarms was not affected

- For clusters consisting of 4 or 6 sequential alarms, identification accuracy was the highest for the first and last alarms in the sequence. This effect was observed for both visual and auditory alarms

Performance during alarm floods

In all three studies, alarm signals were presented in either low alarm frequency periods (routine operations) or as part of an alarm flood. Each alarm flood lasted three minutes and included 90 alarms (30 single alarms and 30 alarm pairs) in the first two studies and 92 alarms in the third study (including single alarms and clusters of 2 – 6 alarms).

- Overall, when compared to the low alarm frequency periods, alarm floods resulted in lower or comparable detection rates, lower accuracy, and shorter response times
- Performance of concurrent tasks suffered during alarm floods: response time to the delivery consent task nearly doubled, and over 90% of the messages with the call sign “DOI51” were missed
- In all three studies, the worst performance was observed with alarms that were presented about 40 – 50 seconds into the alarm flood. Towards the end of the alarm floods, a consistent improvement in alarm detection and identification was observed.

Chapter 6

Improving Alarm Detection and Identification through Preattentive Criticality Mapping

The effects of simultaneous and asynchronous masking on alarm detection and identification have been described and discussed in Chapters 2 – 4 and summarized in Chapter 5. The three experiments reported in those chapters showed that detection and identification are greatly affected by the temporal distribution of alarm signals. Alarms that are closely spaced in time are less likely to be detected and correctly identified due to asynchronous masking. Alarms presented concurrently suffer from simultaneous masking. Both masking effects become stronger with an increase in the number of alarms in temporal proximity. These effects were observed for both visual and auditory alarms. These results informed the design of alarm displays by suggesting avoiding certain temporal distributions of alarm signals. For example, an adaptive alarm display may briefly delay an alarm or change the modality in which it is presented to support operators in noticing and identifying signals when faced with a high frequency of alarms, such as during an alarm flood. When multiple alarms are to be presented in temporal proximity, the order of presentation could also play a role in the probability of detection and identification. Results from previous chapters have shown, for example, that between two alarms, auditory forward masking is likely stronger than backward masking. When more than two alarms are presented in a sequence, the first and last alarms are more likely to be detected and correctly identified. These effects should be considered carefully when the proposed adaptive alarm display is to delay the presentation of some alarms. For example, if two auditory alarms were to be presented at the same time, the less critical one should be delayed so that the more critical one is more likely to be detected. Such an adaptive design has been indirectly demonstrated as an effective countermeasure to the masking problem by the performance differences observed with different temporal distributions.

The next and final study in this line of research aims to improve alarm detection and identification in a different approach. Two main approaches have been proposed to resolve the conflict between the high demand of alarm floods and the limited perceptual and cognitive abilities of the operators: (1) reduce the number of alarm signals or (2) increase the bandwidth of alarm processing.

Many studies on alarm management have adopted the former approach and attempted to reduce the number of alarms by eliminating false alarms (alarms that are triggered when there is no actual underlying problem or emergency) and/or nuisance alarms (alarms that are redundant or have little operational value). For example, Oberli and colleagues (1999) developed a fuzzy logic-based expert system (one that is able to work with incomplete or noisy information) that combined several patient vital signs to reduce the number of alarms in operating rooms. Traditional alarm systems generate alarms whenever any one of the vital signs exceeds a critical threshold, whereas the proposed expert system integrated data from several sensors and only generated an alarm when a meaningful combination of signs suggested an actual common problem. The judgment was made based on historical data and medical knowledge. The expert system was able to reduce the percentage of false alarms from 75% to 1%. Another approach to reducing the number of nuisance alarms is to adjust the alarm limit depending on the context of the alarms. For example, customizing alarm settings to each patient can significantly reduce the number of nuisance alarms in a medical context (Graham & Cvach, 2010). Zhu and colleagues (2014) proposed automated alarm limit adjustments based on Bayesian estimation from the previous states of the system in order to reduce alarm floods in the context of chemical processes. A number of methods have also been developed to detect and suppress nuisance or chattering alarms (repeated and rapid transitions between alarms and normal states), including data mining methods to identify the correlation between alarm-related events and statistical methods to reconfigure alarm settings dynamically when system state changes (Burnell & Dicken, 1997; Noda, Higuchi, Takai, & Nishitani, 2011; Srinivasan, Liu, Lim, Tan, & Ho, 2004; Wang & Chen, 2013).

A different approach to support alarm detection and identification is to increase the bandwidth of information processing. This can be accomplished, for example, through the use of multimodal alarm displays, i.e., displays that present alarms via several different sensory channels. The development of multimodal displays is motivated by Multiple Resource Theory (MRT). MRT posits that partially independent pools of attentional resources are associated with three dimensions of information processing: processing stage (perception/cognition/responding), modality (visual/auditory/tactile), and processing code (spatial/verbal). Tasks that draw from different pools of resources (e.g., tasks that involve the processing of visual versus auditory information) can be better performed in parallel as resource competition and interference are minimized or eliminated (Wickens, 2008). Other benefits of distributing information across different modalities are reduced workload and increased bandwidth of information processing (e.g., Ferris, Penfold, Hameed, & Sarter, 2006; Giang et al., 2010; Nikolic, Sklar, & Sarter, 1998).

Information processing can be facilitated also by supporting preattentive reference, i.e., the processing of (alarm) signals that occurs before attentional selection (Treisman, 1985; Woods, 1995). For example, ambient displays are a type of preattentive reference. They present subtle changes in features of a continuous display, such as its brightness or pitch (Streitz et al., 2003; Wisneski et al., 1998). These changes can be processed without requiring focused attention, "in the background of awareness". Ambient displays have been used, for example, to support the monitoring of patient vital signs by changing the appearance of tools and decorative items in nurses' working environment (Tentori, Segura, & Favela, 2009). Ambient displays, however, are designed to present changes in continuous variables, rather than for presenting discrete alarm signals.

A more appropriate technique for supporting preattentive processing in the context of alarms is urgency mapping. Studies to date have shown that the perceived urgency of alarm signals can be manipulated by changing features such as the color, wording, or pulse rate of visual alarms; the base frequency, loudness, duration or pulse rate of auditory alarms; and the pulse rate of tactile alarms (Baldwin & Lewis, 2014; Guillaume, Drake, Rivenez, Pellieux, & Chastres, 2002; Kaufmann, Ohg, Risser, Geven, & Sefelin, 2008; Lewis & Baldwin, 2012; Sanderson, Eunice, Philippe, & Alexandra, 2006). Mapping the perceived urgency of an alarm to the actual urgency of the underlying event helps operators attend selectively and respond faster to more critical alarms (Chancey, Brill, Sitz, Schmuntzsch, & Bliss, 2014; Edworthy & Meredith, 1994; Politis, Brewster, & Pollick, 2014; Suied, Susini, & McAdams, 2008).

While there is an abundance of evidence that proper mapping of alarm urgency improves response time, it is not clear whether such mapping also improves the detection of alarms. To date, most studies on the benefits of urgency mapping measured response time only because participants in these experiments were not required to multitask or cope with large numbers of alarms; alarms were therefore not likely to be missed (e.g., Guillaume et al., 2002; Haas & Edworthy, 1996; Mondor & Finley, 2003; Suied et al., 2008). Limited evidence from a study requiring participants to perform a cognitively demanding primary task shows that high urgency alarms (associated with red signals) were more likely to be reported than low urgency alarms (color-coded in green) (Bliss, Gilson, & Deaton, 1995). In this study, responses to high urgency alarms were also more accurate than to low urgency alarms. These findings may be explained by the interaction of two different factors directing attention: goal-directed (top-down) factors and stimulus-driven (bottom-up) factors (Corbetta & Shulman, 2002; Oliva, Torralba, Castelhano, & Henderson, 2003). The contingent orienting hypothesis (Folk, Remington, & Johnston, 1992) postulates that stimulus-driven attention capture is contingent on attentional control settings (goal-directed attention control). In other words, a signal (such as an alarm) is more likely to involuntarily capture attention if its features match those the observer is looking for. In the case of preattentive urgency mapping of alarms, the operator is tuned to certain features (e.g., they are aware that red alarms are more important; a top-down influence), and the alarms are designed to possess those features (a bottom-up influence). Preattentive mapping not only helps increase the bandwidth of information processing, but also facilitates prioritization of the alarm signals based on their importance. The prioritization of alarm signals is important because it is likely that operators will not be able to respond to all alarms during an alarm flood and therefore have to make decisions about which alarms they should attend to.

Method

Participants

The participants in this study were 15 students recruited from the College of Engineering at the University of Michigan. The number of participants in this study was determined based on a statistical power analysis using the results from the previous experiment. All participants were between 20 and 35 years old (mean = 22.47 years, SD = 1.96 years; 7 males and 8 females) and had not participated in the previous studies reported in Chapters 2, 3, and 4. They all reported normal or corrected-to-normal vision, normal color vision, and normal hearing ability. This study was approved by the University of Michigan Institutional Review Board (exempt and not regulated; UM IRB: HUM00156091).

Apparatus and tasks

The simulation used in this study was the same ground-based drone control system described in Chapter 4. It consisted of a computer with a keyboard, an optical mouse, a 23-inch LCD monitor, a pair of stereo speakers, a set of driving pedals, and an audio recorder. During the experiment, participants were required to perform the same three tasks as in the previous studies: (1) delivery consent, (2) alarm detection and identification, and (3) air traffic control monitoring.

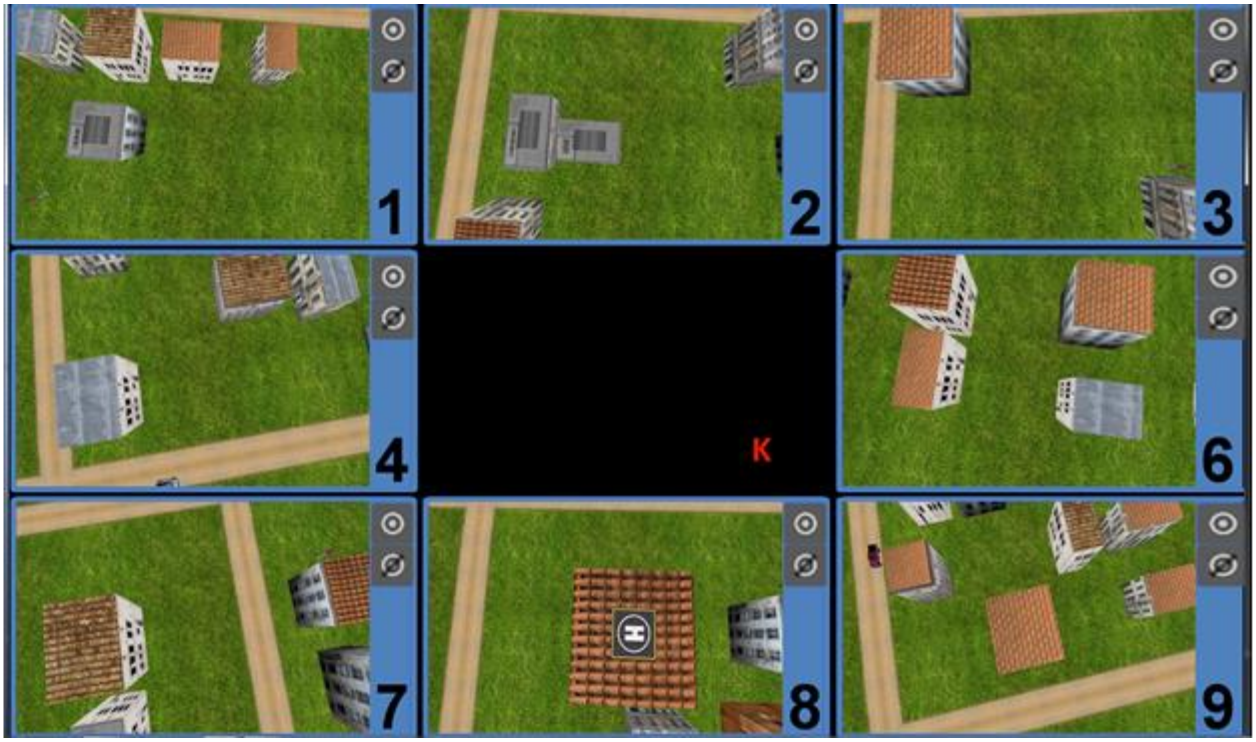

Figure 6.1 Drone simulation interface

(displaying a high criticality visual alarm for drone 9, motor K; requiring the response “9K”)

Participants were informed that each drone had six motors denoted with different letters (B, J, K for visual alarms; X, Y, or Z for auditory alarms) and that visual or auditory alarms would be presented when one or more of the motors was not functioning correctly. The alarms were presented in one of three ways: control, scored, or mapped. Participants were randomly assigned to one of three groups, and each group of participants was exposed to only one of the three methods of alarm presentation. Note that, in contrast to much of the earlier literature in this area, a distinction is made in this study between criticality and urgency where criticality – the focus in this experiment – is defined as the severity of the outcome should an alarm be missed whereas urgency is defined as the time available to respond to the alarm.

In the control group, visual alarms were presented in the central window of the screen as a grey letter referring to one of the drone's motors (B, J, or K; see ***Figure 6.1*** for an example of a visual alarm for the motor K in drone 9). Auditory alarms were presented using a synthesized female voice that read the number of the drone, followed by the letter of the motor (X, Y, or Z). The voice was accompanied by a 440Hz pure tone (middle C note). No criticality information was provided. In the scored group, the alarm presentations were the same as in the control group, but participants were told in advance of the experiment that the motors performed different functions and were therefore not equally important for the safety of flight. **Table 6.1** shows the criticality levels associated with the different motors. Participants were informed that the criticality of an alarm determined the associated performance score and thus how much incentive they would receive at the end of the experiment (for detailed scores, see **Table 6.1**). Notice that each motor corresponded to exactly one level of criticality, so the letter alone was sufficient to determine the criticality of an alarm. The letters were chosen based on two criteria: all the letters were visually and acoustically distinct from each other, and criticality increased with alphabetic order. In the mapped group, the alarms were not only assigned different criticalities, but also displayed in different colors (note that all visual alarms were equal in brightness and saturation and differed only with respect to hue) or accompanied by different tones (auditory alarms; see **Table 6.1**). The three auditory frequencies were selected to ensure that the differences between the low, medium, and high pitch were perceived to be equal (Houtsma, 1995). Participants in all three groups were required to press the space key on the keyboard as soon as they detected an alarm, and then verbally report the number of the affected drone first, followed by the letter of the affected drone. In the scored and mapped groups, participants were reminded and encouraged to prioritize the more critical alarms.

Table 6.1 Alarm criticality in the scored and mapped scenarios

	Visual		Auditory	
Criticality	**Drone**	**Color (if mapped)**	**Drone**	**Pitch (if mapped)**
Low (1 point)	B	Green	X	Low (220 Hz)
Medium (3 points)	J	Orange	Y	Med (440 Hz)
High (5 points)	K	Red	Z	High (880 Hz)

The air traffic control monitoring task was the same as in Chapter 4. Participants listened to the air traffic control recording and monitor for messages containing their call sign "DOI51". Participants were required to press the driving pedal as quickly as possible whenever they heard their call sign.

Experiment design and procedure

The experiment employed a 3×3 fractional factorial design. The two independent variables were alarm presentation and alarm criticality. Alarm presentation (control, scored, and mapped) was varied between subjects to avoid confusion. Alarm criticality (low, medium, high) was varied within subjects. All alarms in the control group were treated as medium criticality for the purpose of the analysis.

The experiment consisted of two 20-minute scenarios, named 'mixed alarm frequency' (Mixed) and 'alarm flood only' (Flood), respectively. The Mixed scenario included two low alarm frequency periods (minutes 0 – 14 and minutes 18 – 20) and one high alarm frequency period (minutes 15 – 17). Each type of alarm cluster shown in **Table 6.2** (in addition to single visual and auditory alarms) was presented once during the low frequency periods and twice during the high alarm frequency period. In the Flood scenario, alarms were presented only during the alarm flood (minutes 10 – 12), when each type of cluster and single alarm were presented twice. The order of the alarms and alarm clusters was randomized. A total of 92 alarms were presented in each alarm flood, and 46 alarms were presented in the low alarm frequency period of each mixed alarm frequency trial.

Table 6.2 Alarm cluster compositions

Cluster size	**Visual**		**Auditory**		**Cross-modal**	
	Concurrent	Sequential	Concurrent	Sequential	Concurrent	Sequential
2	Included	Included	Included	Included	1V+1A	1V1A or 1A1V
4	Included	Included		Included		
6	Included	Included		Included		

In both scenarios, the delivery consent task was presented 30 times (about once every 40 seconds). The air traffic control task was presented about once per minute throughout the scenario (a total of 20 times), including the messages that included their call sign and distractor messages.

Upon arrival at the laboratory, participants were asked to read and sign a printed consent form. Their chair was then adjusted so that their eyes were about 20 inches from the center of the screen. At this distance, the size of the visual alarms equaled 0.125 rad visual angle. Next, participants were given instructions for their three tasks, and they were told to give equal priority to all tasks. They then received four to five 5-minute training sessions to familiarize themselves with the tasks until they achieved a combined accuracy of 90% or higher in the three tasks. Following the training sessions, the participant took a short break before starting the three 20-minute experiment sessions. Before each session, cross-modal matching was performed by adjusting the volume of the auditory alarms to match the intensity of the visual alarms (Colman, 2015; Pitts, Riggs, & Sarter, 2016). The entire experiment lasted about 100 minutes. Upon completion, each participant received $25 as compensation. As an incentive, the three participants with the best combined performance on the three tasks received a bonus of $50, $30, and $20, respectively.

Dependent measures

The dependent measures in this study were detection rate, accuracy, and monitoring score for the alarm monitoring task, accuracy and response time for the delivery consent task, and accuracy and response time for the air traffic control monitoring task.

For the visual and auditory alarms, detection rate was defined as the percentage of alarms that the participant responded to, regardless of the accuracy of the response. Accuracy was defined as the percentage of correct responses to the alarms. Monitoring score was defined as the assigned score of the correctly reported alarms, aggregated over the entire scenario.

For the delivery consent task, accuracy was defined as the percentage of correct responses (giving consent only when a delivery pad was present and rejecting when the delivery pad was absent), divided by all responses. Response time was defined as the time from when the drone reached the residency and stopped to when the participant pressed one of the two buttons next to the drone window.

For the air traffic control monitoring task, accuracy was defined as the percentage of correct responses (only responding when the call sign “DOI51” was included in the message). Response time was defined as the time from the start of the call sign to the time when the space key was pressed.

Results

For the alarm monitoring task, detection rate and accuracy for alarms were analyzed with mixed model binary logistic regression, and response time was analyzed using univariate analyses of variance (ANOVA). These analyses were conducted separately for visual and auditory alarms. In all these analyses, the fixed effect factors were alarm presentation and alarm criticality, and the random effect factor was subject ID. Sequential Bonferroni corrections were applied for multiple comparisons. Alarm performance scores were also analyzed with univariate analyses of variance (ANOVA) but, in this case, with alarm frequency and alarm presentation as the two fixed effect factors.

For the delivery consent and air traffic control monitoring tasks, response times were analyzed with univariate analyses of variance (ANOVA). The accuracy for the ATC monitoring task was analyzed with mixed model binary logistic regression. In these analyses, the fixed effect factors were alarm frequency and alarm presentation, and the random effect factor was subject ID. The accuracy for delivery consent tasks was not analyzed because no mistakes were made. For all three tasks, only significant results ($p < .05$) are reported here.

Detection rate

Alarm criticality had a significant effect on the detection rate for visual alarms ($F(2, 2091) = 3.280$; $p = .038$; see ***Figure 6.2***). Post-hoc tests (with Bonferroni correction) revealed a significant difference between low and medium as well as low and high criticality alarms. No significant interaction was found. The detection rate for auditory alarms was not significantly affected by alarm criticality or alarm presentation (see ***Figure 6.3***).

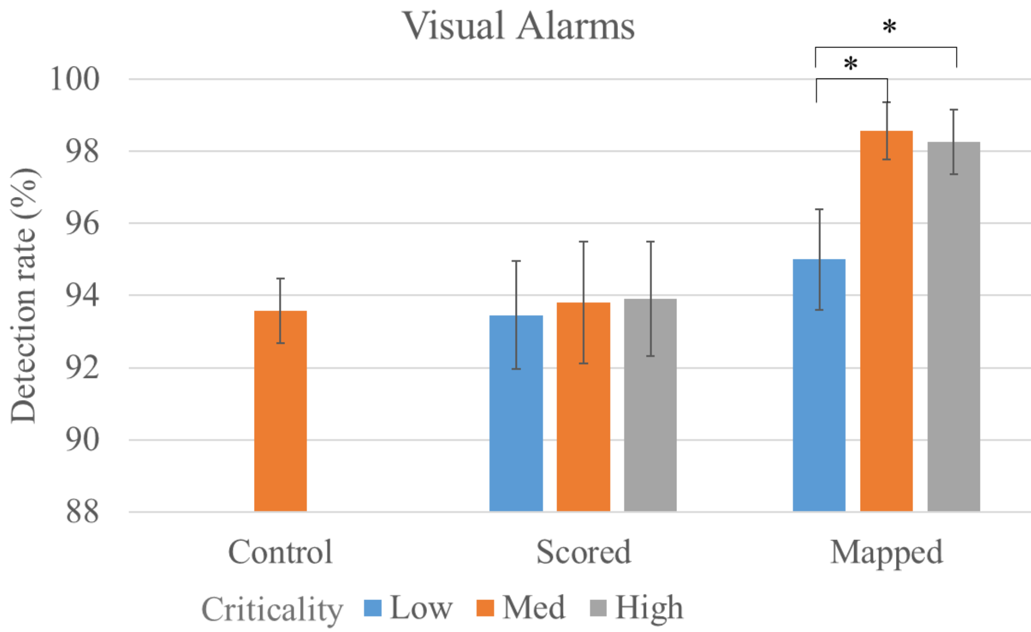

Figure 6.2 Detection rates of visual alarms as a function of alarm presentation and criticality (*p < .05)

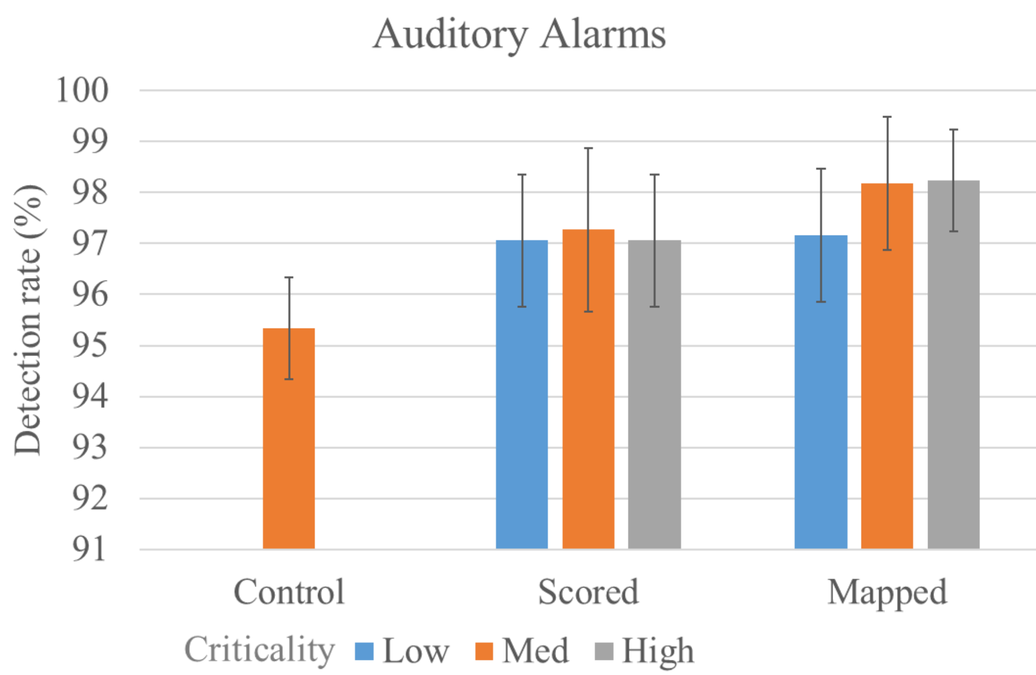

Figure 6.3 Detection rates of auditory alarms as a function of alarm presentation and criticality (*p < .05)

Accuracy

The accuracy for visual alarms was significantly affected by both alarm presentation ($F(2, 2091) = 24.182$; $p < .001$) and alarm criticality ($F(2, 2091) = 4.782$; $p < .001$; see ***Figure 6.4***). There was also a marginally significant interaction between these two factors ($F(4, 2091) = 2.004$, $p = .091$), such that accuracy for visual alarms differed significantly only between low- and medium-criticality alarms and low- and high-criticality alarms (but not between medium and high criticality) in the mapped group. The accuracy for auditory alarms was affected only by alarm presentation ($F(2, 1341) = 6.598$; $p = .001$). Accuracy was higher for the scored and mapped groups, compared to the control group (see ***Figure 6.5***).

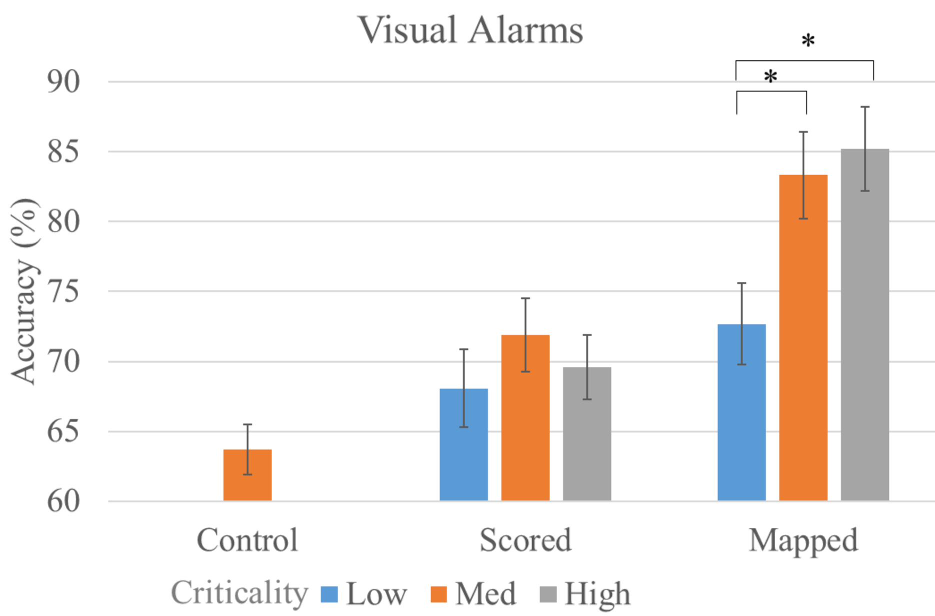

Figure 6.4 Accuracy of visual alarms as a function of alarm presentation and criticality

(*p < .05)

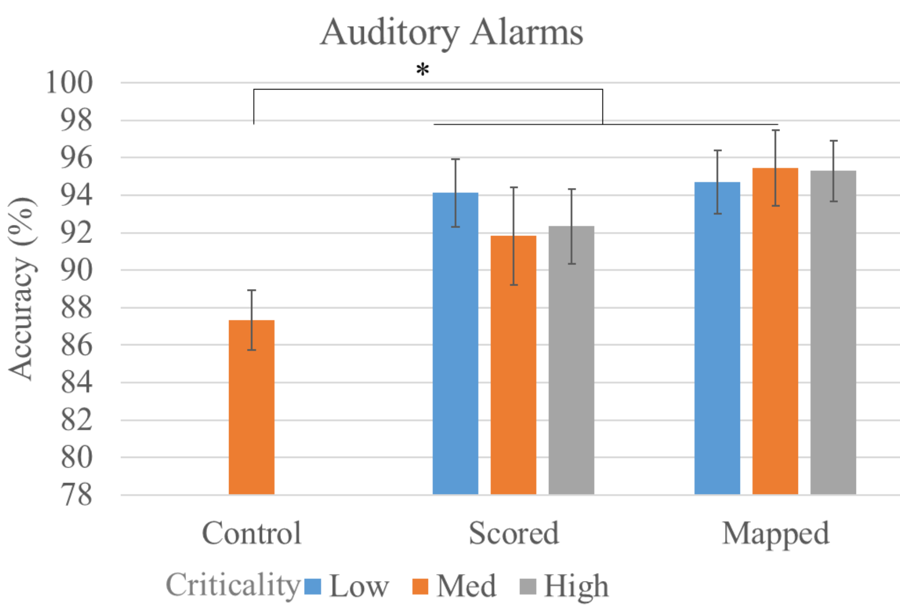

Figure 6.5 Accuracy of auditory alarms as a function of alarm presentation and criticality

(*p < .05)

Response time

The response times to visual and auditory alarms were not affected by alarm presentation nor by alarm criticality (see ***Figure 6.6*** and ***Figure 6.7***, respectively).

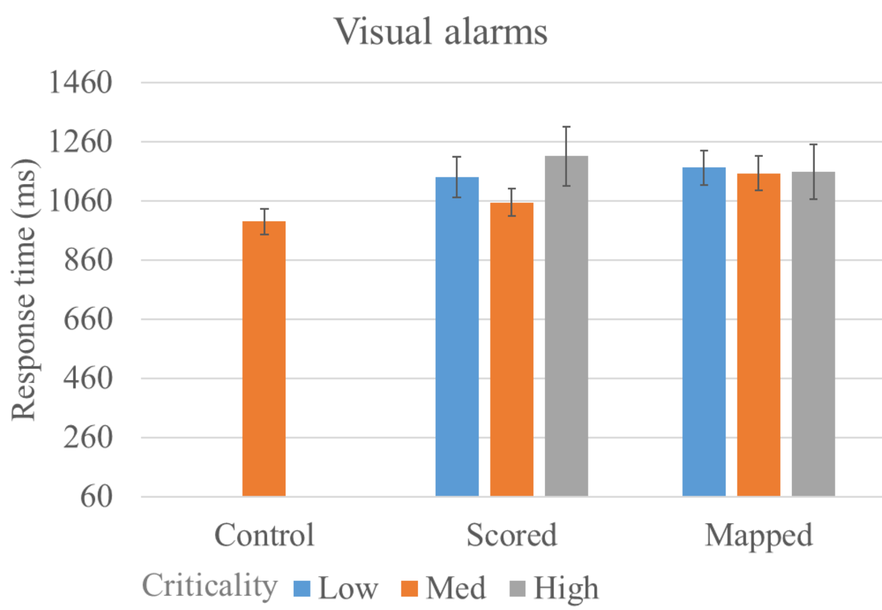

Figure 6.6 Response time to visual alarms as a function of alarm presentation and criticality

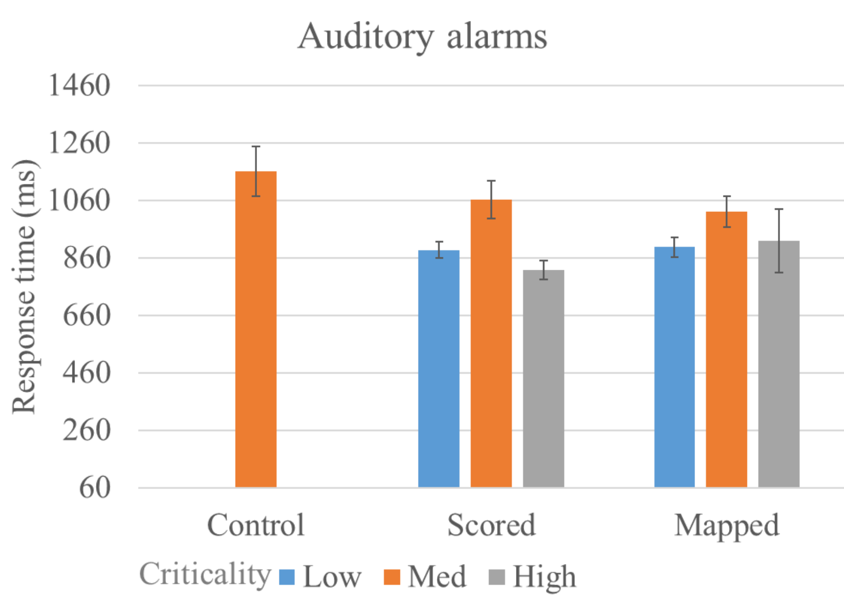

Figure 6.7 Response time to auditory alarms as a function of alarm presentation and criticality

Alarm Monitoring Score

The average scores on the alarm monitoring task were higher when the alarm frequency was low ($F(1, 2760) = 6.731,\ p = .013$; see ***Figure 6.8***). Alarm presentation also had a significant effect on the average score ($F(2, 2760) = 7.259,\ p = .002$) which was higher for the mapped group, compared to both the scored and the control groups.

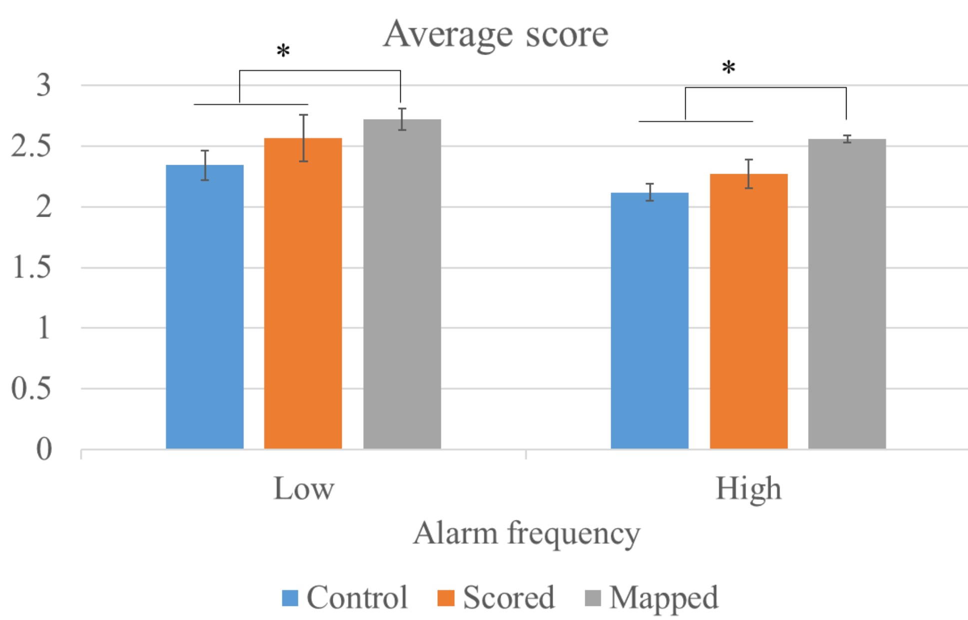

Figure 6.8 Average alarm score as a function of alarm frequency and alarm presentation

Figure 6.9 and ***Figure 6.10*** show the average monitoring score for each alarm cluster or single alarm as a function of alarm presentation for the mixed and the flood scenario, respectively. The scores were normalized by dividing each score by the average score of the same type of clusters or single alarms (e.g., the score for a particular cluster consisting of two concurrent visual alarms was divided by the average score for all concurrent two visual alarm clusters). This normalization served to determine the relative difference between scores regardless of the difficulty of each type of cluster. Alarm clusters for which the normalized scores differed by more than 0.5 between the three types of alarm presentation are listed in **Table 6.3**.

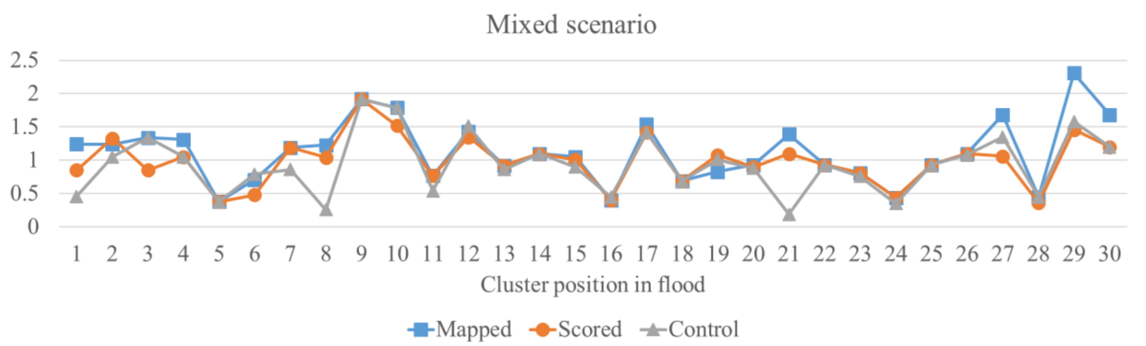

Figure 6.9 Normalized average score as a function of cluster position in high alarm density periods (Mixed)

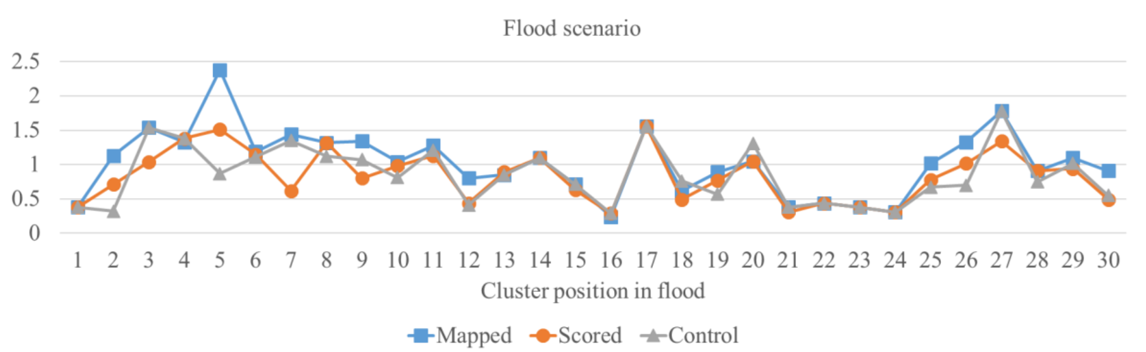

Figure 6.10 Normalized average score as a function of cluster position (Flood)

Table 6.3 Alarm clusters most affected by alarm presentation

Scenario	Cluster #	Modality	Cluster type	Cluster size	Difference
Mixed	1	Visual	Concurrent	4	.78
	8	Visual	Concurrent	6	.97
	21	Visual	Concurrent	6	1.20
	27	Auditory	Concurrent	2	.62

	29	Visual	Concurrent	4	.85
Flood	2	Visual	Concurrent	6	.81
	5	Visual	Concurrent	2	1.51
	7	Auditory	Concurrent	2	.73
	26	Visual	Concurrent	4	.63

It is interesting to note that the above alarm clusters are all concurrent clusters and that the visual alarm clusters consist of more than two signals (except for cluster 5 in the Flood scenario). For all clusters, the score was highest with the mapped alarm presentation. ***Figure 6.11*** shows a comparison of the normalized average scores for the alarm clusters that were most affected, as a function of alarm presentation. A one-way ANOVA revealed significant main effects of alarm presentation ($F(2,720) = 10.494,\ p < .001$) and cluster type ($F(2,720) = 12.839,\ p = .019$).

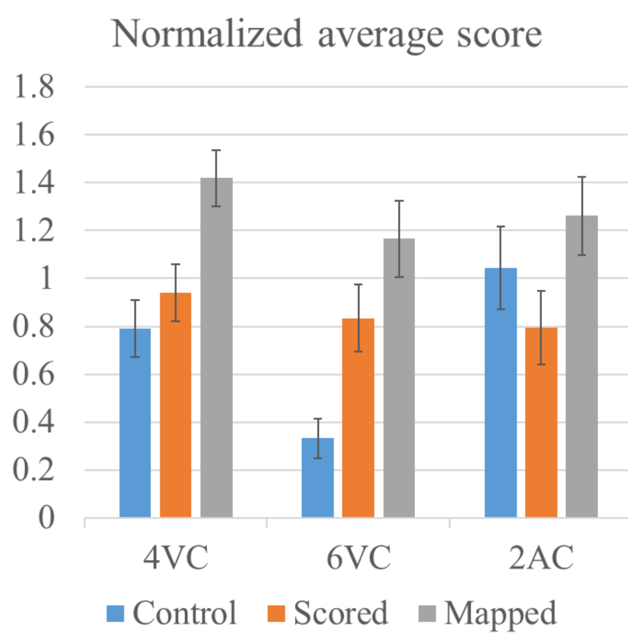

Figure 6.11 Normalized average score of selected clusters

(4VC-4 visual concurrent, 6VC-6 visual concurrent, 2AC- 2 auditory concurrent)

Delivery consent task

Overall, the response times to the delivery consent task were longer during the high alarm frequency periods (2125ms vs. 1514ms; ($F(1, 1182) = 35.039,\ p < .001$)). An interaction between alarm presentation and frequency was observed, such that the response time was shorter in the mapped group compared to the other two groups ($F(2, 1182) = 3.054,\ p = .048$).

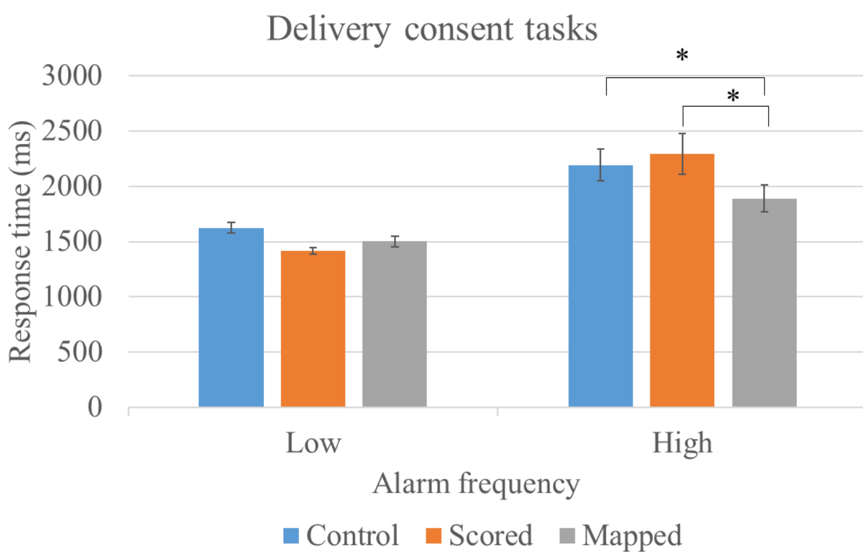

Figure 6.12 Response time to delivery consent as a function of alarm presentation and frequency

Air traffic control monitoring task

The response time for the air traffic control monitoring task was not affected by either factor. Accuracy was significantly lower during the high alarm frequency periods, compared to the low alarm frequency periods (70.3% vs. 90.2%; $F(1\ , 1194) = 47.003,\ p < .001$). No interaction was observed.

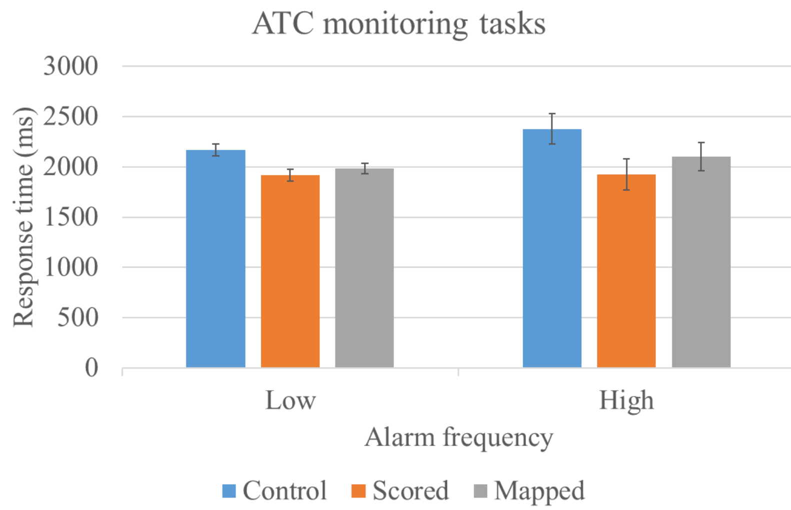

Figure 6.13 Response time for the ATC monitoring task as a function of alarm presentation and frequency

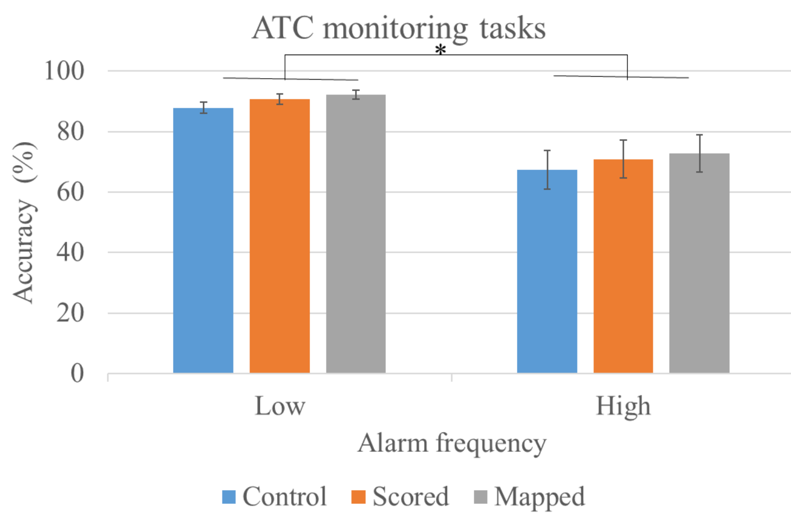

Figure 6.14 Accuracy for the ATC monitoring as a function of alarm presentation and frequency

Subjective rating

Participants were asked to rate, on a 1 – 10 scale, how much each design helped them determine the importance of an alarm. For the visual alarms, the color mapping (mean = 7.40, SD = 1.96) was rated as much more important than the letters (mean = 4.17, SD = 1.57). For the auditory alarms, the pitch mapping (mean = 6.20, SD = 2.56) was also rated as more important than the letters (mean = 5.33, SD = 2.69). The difference between the two visual features was greater than that between two auditory features. Participants were also asked to rate the impact of each feature on their final performance. The rating was again higher for the color mapping (mean = 7.40, SD = 1.96) than for the visual letters (mean = 6.00, SD = 1.10). However, the pitch mapping (mean = 5.80, SD = 2.48) was rated lower than the auditory letters (mean = 6.60, SD = 1.74).

Discussion

The goal of this study was to facilitate the detection and identification of alarm signals using criticality mapping. Two preattentive features, color and pitch, were implemented for visual and auditory alarms, respectively. It was expected that both detection and identification performance would improve, especially for alarms of high criticality. To ensure this effect was due to the criticality mapping of the alarms, another group (the scored group) was informed about alarm criticality per se in advance of the experiment but this alarm property was not mapped to any preattentive features of the alarm signal itself.

The performance for visual alarms was largely consistent with expectations. In the mapped group, both detection rate and accuracy for medium and high criticality alarms were greatly improved, compared to the control group. Participants were clearly able to prioritize the more critical alarms. This performance gain for medium and high criticality alarms was not achieved at a cost to low criticality alarms. In fact, the detection rate for low criticality alarms remained the same, and accuracy even improved with preattentive mapping. The observed improvements in accuracy may be partially attributed to the use of redundant color coding in the mapped group. In the control and scored groups, alarms could be identified only by the letter of the motor while, in the mapped group, color also identified the alarm, regardless of its criticality. It has been suggested that color is the best feature for encoding nominal (but not quantitative) information, compared to shape and size, and that color encoded information supports rapid processing (Christ, 1975; Nowell, Schulman, & Hix, 2002). This redundant encoding of information is useful only after an alarm has been detected. Therefore, the low criticality alarms benefitted from the color mapping in terms of accuracy but not detection rate.

Another piece of evidence that the observed benefits were attributable to preattentive mapping, rather than alarm criticality per se, is that no significant performance benefits were observed for the scored group. Compared to the control group, the detection rate for the scored group was the same and accuracy was only slightly improved. These results indicate that knowledge of criticality per se did not help participants prioritize highly critical alarms even though they were instructed to do so. The need to respond to alarms as quickly as possible may have led to this outcome. In the debriefing questionnaire, some participants mentioned that they “did not pay much attention to this (criticality)” or “paid more attention to the fact that an alarm had been detected rather than the importance of that alarm”. In the scored group, the participants had to consciously determine the criticality of an alarm, which was very difficult under time pressure. It required activation of a learned rule (the relation between letters and criticality) which is much more effortful than relying on the intuitive mapping between color and criticality (intuitive in the sense of following convention).

In the mapped condition, performance of the delivery consent and air traffic monitoring tasks did not suffer, compared to the control group. In fact, response times to the delivery consent task were faster in the mapped group, compared to both the control and the scored groups. This also suggests that the mapped alarm features were processed at a preattentive level, thus freeing up attentional resources and supporting multitasking (see Woods, 1995).

The results for using criticality mapping of auditory alarms were different. The detection rate of auditory alarms was not affected by alarm presentation or criticality. Accuracy was, in general, slightly higher in both the scored and mapped groups, compared to the control group, but there was no difference between alarms of different criticality. Also, in the subjective ratings, the use of pitch mapping was rated as less useful than the color mapping. These results suggest that pitch may not be a good candidate for preattentive criticality mapping when it requires absolute judgments of pitch levels. The absolute judgment of pitch is known to require training and experience, and it also depends on innate abilities (Wickens, Hollands, Banbury, & Parasuraman, 2015). Another factor that may have contributed to difficulties with making pitch judgments is that, in this study, the pure-pitch tone and the voice were two distinct auditory objects, whereas the color coding was a feature of the letter itself and thus part of the same object. This explanation is suggested by the fact that different features of the same object have been shown to be easier to process in parallel in both the visual and auditory modalities (Shinn-Cunningham, 2008; Treisman, Kahneman, & Burkell, 1983). Further investigation is needed to determine whether other auditory features, such as tempo or gender, could yield better results for preattentive criticality mapping. It should be noticed that two other acoustic features – the tempo and pitch of the voice – has been tested and ruled out in the piloting phase of this experiment. Tempo is perceived to be closely associated with urgency rather than criticality. It also affects the discriminability of the signal and makes the faster (more critical) alarms less audible. Pitch of the voice is more difficult to manipulate because a major change in the pitch of the voice changes the perceived identity of the voice and elicits different cultural and emotional reactions from different people.

Most previous studies on alarm urgency mapping observed faster response times to more urgent alarms (e.g., Baldwin & Lewis, 2014; Haas & Edworthy, 1996; Suied et al., 2008). In contrast, in this study, response times to both visual and auditory alarms were not affected by alarm presentation or alarm criticality. This discrepancy may be explained by two differences between the current study and earlier work: the mapped property and the required tasks. Previous studies mapped preattentive features to the urgency of the alarms, whereas the current study mapped those features to the criticality of the alarms. Recall that, in most studies, the term "urgency" refers to both the severity of the outcome if an alarm is not properly handled and also to the need to respond to the alarm as quickly as possible. In this study, in contrast, participants in the scored and mapped groups were told that different motors (indicated by different alarms) were of different importance for the flight task, but the time available to respond to these alarms was the same. Therefore, participants were not motivated to respond faster to more critical alarms. The second difference between the current experiment and previous studies is that the current study required multitasking which resulted in very high workload and strong competition for mental resources, while previous studies usually employed a single task only. All three groups of participants in the present experiment were motivated to minimize their response time to alarms simply because of the need to perform the other two tasks. The criticality of the alarms provided very little additional incentive.

Overall, average alarm monitoring scores were higher for the mapped group, compared to the scored group and the control group, during both low and high alarm frequency periods. This finding confirms that preattentive mapping improves alarm detection and identification under high and low workload conditions. The clusters that benefited the most from preattentive mapping consisted of 4 or 6 concurrent alarms. These are the same types of alarm clusters that showed the worst performance in the experiment described in Chapter 4. In combination, these findings suggest that preattentive reference is more beneficial when parallel processing is needed and the competition for short term memory and attentional resources is strong. They are also consistent with previous studies that found preattentive processing to be useful for rapid visual target detection and numerical estimation (e.g., Healey, Booth, & Enns, 1996). These studies showed that, when participants were presented with a large set of items and required to rapidly estimate the percentage of items with a certain feature, they were able to do so within as little as 105ms, which is much faster than the engagement of visual attention (suggested to be no less than 200ms).

This experiment developed and tested preattentive criticality mapping as a countermeasure to observed breakdowns in alarm detection and identification due to simultaneous and sequential masking. The color of visual alarms and the pitch of auditory alarms, respectively, were mapped to the criticality of alarms. This mapping improved detection and identification performance for visual alarms of medium- and high-criticality, without affecting performance for low-criticality alarms. These benefits were most pronounced with clusters of 4 or 6 concurrent visual alarms, which had been shown to suffer the most from masking effects. The observed benefits of preattentive mapping suggest that the identification of alarm signals can precede and affect their detection, a reversal of the first two stages in the seven-stage model of alarm processing proposed by Stanton & Stammer (1998) and introduced in Chapter 1. However, the effect of pitch mapping for auditory alarms was much less pronounced compared to the color mapping for visual alarms. To sum up, the proposed design serves as a useful countermeasure to reduce the risk of missed and misdiagnosed alarms due to masking, but a more effective mapping feature for auditory alarms remains to be explored.

References

Baldwin, C. L., & Lewis, B. A. (2014). Perceived urgency mapping across modalities within a driving context. *Applied Ergonomics*, *45*(5), 1270–1277. https://doi.org/10.1016/j.apergo.2013.05.002

Bliss, J. P., Gilson, R. D., & Deaton, J. E. (1995). Human probability matching behaviour in response to alarms of varying reliability. *Ergonomics*, *38*(11), 2300–2312.

Burnell, E., & Dicken, C. R. (1997). Handling of repeating alarms. In *Stemming the Alarm Flood (Digest No: 1997/136), IEE Colloquium on* (pp. 11–12). IET.

Chancey, E. T., Brill, J. C., Sitz, A., Schmuntzsch, U., & Bliss, J. P. (2014). Vibrotactile Stimuli Parameters on Detection Reaction Times. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*(1), 1701–1705. https://doi.org/10.1177/1541931214581355

Christ, R. E. (1975). Review and Analysis of Color Coding Research for Visual Displays. *Human Factors*, *17*(6), 542–570. https://doi.org/10.1177/001872087501700602

Colman, A. M. (2015). *A dictionary of psychology*. Oxford University Press, USA.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*(3), 201.

Edworthy, J., & Meredith, C. S. (1994). Cognitive psychology and the design of alarm sounds. *Medical Engineering and Physics*. https://doi.org/10.1016/1350-4533(94)90067-1

Ferris, T. K., Penfold, R., Hameed, S., & Sarter, N. (2006). The Implications of Crossmodal Links in Attention for the Design of Multimodal Interfaces: A Driving Simulation Study. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(3), 406–409. https://doi.org/10.1177/154193120605000341

Folk, C. L., Remington, R. W., & Johnston, J. C. (1992). Involuntary Covert Orenting Is Contingent on Attentional Control Settings. *Journal of Experimental Psychology*, *18*(4), 1030–1044.

Giang, W., Santhakumaran, S., Masnavi, E., Glussich, D., Kline, J., Chui, F., … Zelek, J. (2010). *Multimodal interfaces: Literature review of ecological interface design, multimodal perception and attention, and intelligent adaptive multimodal interfaces*. DTIC Document.

Graham, K. C., & Cvach, M. (2010). Monitor alarm fatigue: standardizing use of physiological monitors and decreasing nuisance alarms. *American Journal of Critical Care*, *19*(1), 28–34.

Guillaume, A., Drake, C., Rivenez, M., Pellieux, L., & Chastres, V. (2002). Perception of urgency and alarm design. *Proceedings of the 8th International Conference on Auditory Display (ICAD2002)*. Retrieved from Proceedings/2002/GuillaumeDrake2002.pdf

Haas, E. C., & Edworthy, J. (1996). Designing urgency into auditory warnings using pitch, speed and loudness. *Computing & Control Engineering Journal*, *7*(4), 193–198.

Healey, C. G., Booth, K. S., & Enns, J. T. (1996). High-speed visual estimation using preattentive processing. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *3*(2), 107–135.

Houtsma, A. J. (1995). Pitch perception. Hearing, 6, 262.

Kaufmann, C., Ohg, F., Risser, R., Geven, A., & Sefelin, R. (2008). Effects of simultaneous multi-modal warnings and traffic information on driver behaviour. In *European Conference on Human Centered Design for Intelligent Transport Systems* (pp. 33–42). Retrieved from https://www.researchgate.net/profile/Martin_Baumann2/publication/224998610_The_effect_of_cognitive_tasks_on_predicting_events_in_traffic/links/02e7e52d3d8c521170000000.pd

f#page=41
Lewis, B. A., & Baldwin, C. L. (2012). Equating Perceived Urgency Across Auditory, Visual, and Tactile Signals. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *56*(1), 1307–1311. https://doi.org/10.1177/1071181312561379
Mondor, T. a, & Finley, G. A. (2003). The perceived urgency of auditory warning alarms used in the hospital operating room is inappropriate. *Canadian Journal of Anaesthesia = Journal Canadien d'anesthesie*, *50*(3), 221–228. https://doi.org/10.1097/01.sa.0000108474.38017.92
Nikolic, M. I., Sklar, A. E., & Sarter, N. B. (1998). Multisensory feedback in support of pilot-automation coordination: the case of uncommanded mode transitions. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 42, pp. 239–243). SAGE Publications.
Noda, M., Higuchi, F., Takai, T., & Nishitani, H. (2011). Event correlation analysis for alarm system rationalization. In *Asia-Pacific Journal of Chemical Engineering* (Vol. 6, pp. 497–502). https://doi.org/10.1002/apj.575
Nowell, L., Schulman, R., & Hix, D. (2002). Graphical encoding for information visualization: an empirical study. In *IEEE Symposium on Information Visualization, 2002. INFOVIS 2002.* (pp. 43–50). IEEE.
Oberli, C., Urzua, J., Saez, C., Guarini, M., Cipriano, A., Garayar, B., … Irarrazaval, M. (1999). An expert system for monitor alarm integration. *Journal of Clinical Monitoring and Computing*, *15*(1), 29–35.
Oliva, A., Torralba, A., Castelhano, M. S., & Henderson, J. M. (2003). Top-down control of visual attention in object detection. In *Proceedings 2003 International Conference on Image Processing (Cat. No. 03CH37429)* (Vol. 1, p. I-253). IEEE.
Pitts, B. J., Riggs, S. L., & Sarter, N. (2016). Crossmodal Matching: A Critical but Neglected Step in Multimodal Research. *IEEE Transactions on Human-Machine Systems*, *46*(3), 445–450.
Politis, I., Brewster, S. A., & Pollick, F. (2014). Evaluating multimodal driver displays under varying situational urgency. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems* (pp. 4067–4076). New York, New York, USA: ACM Press. https://doi.org/10.1145/2556288.2556988
Sanderson, P., Eunice, S., Philippe, L., & Alexandra, W. (2006). Auditory alarms, medical standards, and urgency.
Shinn-Cunningham, B. G. (2008). Object-based auditory and visual attention. *Trends in Cognitive Sciences*, *12*(5), 182–186. https://doi.org/10.1016/j.tics.2008.02.003
Srinivasan, R., Liu, J., Lim, K. W., Tan, K. C., & Ho, W. K. (2004). Intelligent alarm management in a petroleum refinery. *Hydrocarbon Processing*, *83*(11), 47–54.
Stanton, N. A., & Stammer, R. B. (1998). Alarm initiated activities: matching visual formats to alarm handling "tasks." *International Journal of Cognitive Ergonomics*, *2*(4), 331–348.Wickens, C. D. (2008). Multiple Resources and Mental Workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 449–455. https://doi.org/10.1518/001872008X288394
Streitz, N., Prante, T., Röcker, C., Van Alphen, D., Magerkurth, C., Stenzel, R., & Plewe, D. A. (2003). Ambient displays and mobile devices for the creation of social architectural spaces. In *Public and Situated Displays* (pp. 387–409). Springer.
Suied, C., Susini, P., & McAdams, S. (2008). Evaluating warning sound urgency with reaction

times. *Journal of Experimental Psychology. Applied*, *14*(3), 201–212. https://doi.org/10.1037/1076-898X.14.3.201

Tentori, M., Segura, D., & Favela, J. (2009). Monitoring hospital patients using ambient displays. In *Mobile Health Solutions for Biomedical Applications* (pp. 143–158). IGI Global.

Treisman, A. (1985). Preattentive processing in vision. *Computer Vision, Graphics, and Image Processing*, *31*(2), 156–177.

Treisman, A., Kahneman, D., & Burkell, J. (1983). Perceptual objects and the cost of filtering. *Perception & Psychophysics*, *33*(6), 527–532.

Wang, J., & Chen, T. (2013). An online method for detection and reduction of chattering alarms due to oscillation. *Computers & Chemical Engineering*, *54*, 140–150.

Wickens, C. D. (2008). Multiple Resources and Mental Workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 449–455. https://doi.org/10.1518/001872008X288394

Wickens, C. D., Hollands, J. G., Banbury, S., & Parasuraman, R. (2015). *Engineering psychology & human performance*. Psychology Press.

Wisneski, C., Ishii, H., Dahley, A., Gorbet, M., Brave, S., Ullmer, B., & Yarin, P. (1998, February). Ambient displays: Turning architectural space into an interface between people and digital information. In *International Workshop on Cooperative Buildings* (pp. 22-32). Springer, Berlin, Heidelberg.

Woods, D. D. (1995). The alarm problem and directed attention in dynamic fault management. *Ergonomics*, *38*(11), 2371–2393.

Zhu, J., Shu, Y., Zhao, J., & Yang, F. (2014). A dynamic alarm management strategy for chemical process transitions. *Journal of Loss Prevention in the Process Industries*, *30*, 207–218.

Chapter 7

General Discussion and Conclusions

The safe operation of highly automated sociotechnical systems depends greatly on operators' ability to detect, identify, and respond to alarms in a timely and accurate fashion. One major challenge for alarm management is an alarm flood which happens when a large number of alarms are presented in a short period of time and exceed the operators' perceptual and attentional capabilities. An alarm flood is defined by industrial standards as more than 10 alarms in a 10-minute period (EEMUA, 1999), but this limit is often exceeded, as illustrated by accidents such as the Three Mile Island nuclear disaster (Rogovin, 1979) and the Texaco Refinery explosion (HSE, 1997). In these accidents, the failure to detect and identify critical alarms and subsequently diagnose the problem has been identified as a major contributor. While research abounds on methods to reduce the number of unnecessary and unwarranted alarms, such as false alarms and nuisance alarms (e.g. Burnell & Dicken, 1997; Edworthy, 2013; Laberge, Bullemer, Tolsma, & Reising, 2014), less is known about the role of perceptual and attentional limitations in the detection and identification of valid alarms.

One such limitation is masking. Masking refers to the situation where one stimulus is obscured by the presence of another stimulus when they are simultaneous or sequential in close temporal proximity (Breitmeyer & Öğmen, 2006; Enns, Di Lollo V, & Di Lollo, 2000). Previous research has shown that both types of masking can lead to missing even expected signals or events. For example, change blindness is an example of simultaneous masking where observers fail to notice even large changes to objects or scenes when these changes coincide with a brief visual disruption (Simons & Levin, 1997). Attentional blink is an example of asynchronous masking which is experienced when two signals are presented in very close temporal proximity and observers miss the second signal (Shapiro, Raymond, & Arnell, 1997). Most studies on masking were conducted in a single-task setting (the single task being the detection of the target stimuli) and examined the interaction between only two signals. In contrast, operators in real-world environments are usually required to multitask and have to cope with larger numbers of alarm signals which creates competition for mental resources. As a result, it is not clear that the findings from this basic research can inform the design of safe and effective alarm systems. For example, limited empirical evidence suggests that, in a more complex environment, the SOA at which asynchronous masking is experienced may be longer than the 200 – 600ms window suggested by basic research on attentional blink (Boot, Becic, & Kramer, 2007; Ferris, Penfold, Hameed, & Sarter, 2006). The first phase of this line of research therefore studied the detection of single and multiple visual and auditory alarms in a demanding multitask setting and, based on the results from the early experiments, examined the effectiveness of preattentive criticality mapping to better support operators in alarm detection and identification. Specifically, the goals of this dissertation were to:

1. Establish the SOA range at which asynchronous masking is experienced in a multitasking, high workload environment;
2. Compare the effects of simultaneous and asynchronous masking on alarm detection and identification;
3. Investigate how alarm detection and identification performance changes as a function of the number of alarms and the serial position of each alarm;

4. Develop and test a countermeasure to masking and the resulting failure to detect and identify alarms.

In order to achieve these goals, a series of 4 experiments was conducted in the context of a simulated automated package delivery system. Chapters 2 and 3 describe two studies on the effect of asynchronous masking on alarm detection and identification in low and high workload settings. In both studies, asynchronous masking was observed with SOAs of 800 – 1000ms, which was longer than the 200 – 600ms range suggested by basic attentional blink literature (Shapiro et al., 1997). This was consistent with previous studies that suggested "task and domain complexity can alter the occurrence and magnitude" of such effects (Ferris et al., 2006). In low workload conditions, asynchronous masking was observed only with visual alarms and forward masking, regardless of the modality of the preceding alarm. When workload increased, both forward and backward masking were observed, with both visual and auditory alarms.

Chapter 4 reports a study that addresses goals 2 and 3 of this line of research. Here, not only asynchronous but also simultaneous alarm presentation was included. The latter becomes more likely, the larger the number of (possible) alarms in a system. The number of alarms in a concurrent alarm cluster, and the number of alarms in temporal proximity (in a sequential alarm cluster) was increased to include 2, 4, and 6 alarms. For both visual and auditory alarms, detection and identification performance decreased as the number of alarms in a cluster increased. This effect was stronger with concurrent alarms than sequential alarms, for both visual and auditory alarm clusters. These findings were consistent with other studies that reported a similar interaction effect between the timing and number of targets to be detected (Boot et al., 2007). For concurrent visual alarms, performance deteriorated significantly when the number of alarms reached 6; in contrast, for auditory alarms, performance breakdowns were observed with only 2 concurrent alarms. In clusters of 4 or 6 sequential visual or auditory alarms, accuracy was higher for the first and last alarms, compared to alarms presented in the middle. This serial position effect may be explained as a result of the interference between the response to one alarm and the presentation of a following alarm. The first alarms had an advantage because their presentation was affected less by the response to the preceding alarms. The performance benefit for the last alarm was a result of the fact that it is the last to enter short-term memory and not interfered by any following alarms, making it easier to retrieve (Laming, 2010).

The studies described in Chapters 2 to 4 included low alarm frequency periods (routine operation) and high alarm frequency periods (alarm floods). A speed-accuracy tradeoff was observed during alarm floods. Also, during alarm floods, they were able to maintain similar detection performance (compared to that in low alarm frequency periods) under low-to-moderate workload (Chapter 2 and 4) but missed more alarms when workload was high (as in Chapter 3). A more detailed analysis of performance as a function of serial position in an alarm flood revealed that performance deteriorated approximately 40 – 50s into the flood. At that point, participants seemed to suffer the most from the surprising effect from the onset of the alarm floods and the resulting increased interference and stress. Beyond that point, they were able to adjust to the pace of events and may have invested more effort in the task which was reflected by a slight improvement in performance.

The last experiment addressed the fourth goal of this dissertation – the development and testing of a countermeasure to observed breakdowns in alarm detection and identification. Specifically, the criticality of alarms was mapped to the color or pitch of the signal to allow faster, preattentive processing of the alarm signals. Detection and identification performance improved for medium- and high-criticality alarms, without negatively affecting the high performance for low-criticality alarms. While the present study does not prove that these benefits were the result of preattentive processing, color and pitch were employed for indicating criticality based on earlier studies suggesting that these two features are processed before engagement of selective attention (Treisman, 1985; Woods, 1995). Also, the fact that performance of the delivery consent and air traffic monitoring tasks did not suffer, and even slightly improved for one task in the mapped condition provides indirect evidence that the criticality information was processed preattentively. Compared to other studies on the mapping of alarm urgency, where the response time to alarms was usually the concern (e.g. Baldwin & Lewis, 2014; Edworthy, 1994), the current study did not yield a benefit in response time because the preattentive features were mapped to criticality but not urgency. This result highlighted how similar display designs could yield different performance benefits when the context is different. Such a mapping technique was shown to be most effective with 4 or 6 concurrent visual alarms, which suffer the most from masking effects. Therefore, the proposed design serves as a useful countermeasure to reduce the risk of missed and misdiagnosed alarms due to masking.

Intellectual merits and broader impacts

The findings from this dissertation contribute to a better understanding of the effects of different types of masking on the detection and identification of critical signals in high workload multi-task environments. They highlight that results from basic research on phenomena like attentional blink and change blindness do not necessarily generalize to applied settings but need to be re-examined instead, taking into consideration differences in tasks and signals. The results from this line of research also inform the development and expansion of models and theories such as MRT (Wickens, 2008), NT-SEEV (Wickens, McCarley, & Steelman, 2009), crossmodal links in attention (Driver & Spence, 1998), and top-down vs. bottom-up factors directing attention (Corbetta & Shulman, 2002; Oliva, Torralba, Castelhano, & Henderson, 2003). For example, results from this dissertation confirmed that some factors already used in the NT-SEEV model, such as visual eccentricity (related to the "effort" component of NT-SEEV) and criticality (related to the "value" component of NT-SEEV), affect alarm detection and identification also in multi-task and multi-alarm settings. At the same time, findings from this line of research suggest additional factors to be included in the model to account for noticing in the context of multiple signals, such as the number of alarms and their temporal distribution. While this dissertation focused primarily on the detection and identification of alarms, most of its findings related to masking generalize to the detection of other types of signals, such as targets in radar surveillance or cybersecurity. Finally, a countermeasure to breakdowns in alarm detection and identification based on preattentive reference and contingent orienting was also developed, tested, and proven to be effective.

From an applied perspective, the findings will provide guidance for the design and evaluation of alarms and alarm systems in data-rich high-risk domains, such as aerospace, cyber defense, medicine, and process control. Even though there is an abundance of alarm design guidelines and standards, most of these guidelines focus on ensuring the noticeability and usefulness of individual alarms by prescribing their desirable salience and informativeness (Russ et al., 2014; Wogalter, Conzola, & Smith-Jackson, 2002). Little guidance is available on how to create effective alarm systems that consider the interaction between multiple alarms, especially during an alarm flood. As complex systems become more digitized (e.g. power plants), automated (e.g. manufacturing), and autonomous (e.g. aviation or driving), the number of alarms and the scale and complexity of alarm systems are inevitably increasing. There is an increased risk of alarm floods which tax human perceptual and attentional processes and abilities. The present research helps prevent that missed or misinterpreted alarms contribute to catastrophic outcomes, and thus improves safety in many real-world environments.

Future work

As with all research, this dissertation answers a number of important questions but also raises new ones that suggest directions for future work. First, only the early stages of alarm handling – detection and identification – were examined in this line of research. Later stages, such as information integration, diagnosis and response selection, are equally critical. The cognitive processes involved in these tasks need to be studied in more detail and supported through technologies that support rule and knowledge activation as well as decision making.

As mentioned earlier, adaptive alarm displays could be developed that adjust the temporal distribution of alarms to avoid masking. Another interesting approach could be to develop an adaptive system based on information about the operator. For example, eye-tracking and other physiological measures have been widely used to detect changes in the mental state of a person, such as fatigue, stress, or attention narrowing (Horng, Chen, Chang, & Fan, 2004; Prinet, 2016; Stankovic, Aitken, & Clark, 2014). These changes are likely to affect alarm detection and interpretation. Dynamic changes in signal salience and means of focusing or broadening attention may need to be developed. Finally, the use of non-traditional channels for alarm presentation, such as tactile signals, should be explored. These signals, in combination with visual and auditory alarms, could be used to create an alarm system that consists of well-coordinated alarms, alerts and notifications, each with its appropriate representation.

Finally, as alarm management systems become smarter and more autonomous, it is important to consider the system and the human operator as a collaborative team. When an alarm flood occurs, the system should support the interpretation and diagnosis of the problem not only by means of providing recommendations and aids but also offering explanations and justifications for the recommendations provided. These explanations can increase the transparency of the system, enable the user to decide whether to trust the system's suggestions, and improve the acceptance of such systems and, ultimately, system performance and safety (Mercado et al., 2016; Sinha & Swearingen, 2002).

References

Baldwin, C. L., & Lewis, B. A. (2014). Perceived urgency mapping across modalities within a driving context. *Applied Ergonomics*, *45*(5), 1270–1277. https://doi.org/10.1016/j.apergo.2013.05.002

Boot, W. R., Becic, E., & Kramer, A. F. (2007). Temporal Limitations in Multiple Target Detection in a Dynamic Monitoring Task. *Human Factors*, *49*(5), 897–906. https://doi.org/10.1518/001872007X230244.

Breitmeyer, B. G., & Öğmen, H. (2006). *Visual masking: Time slices through conscious and unconscious vision*. Oxford University Press.

Burnell, E., & Dicken, C. R. (1997). Handling of repeating alarms. In *IEE Colloquium on Stemming the Alarm Flood (Digest No: 1997/136)* (p. 12/1-12/4). https://doi.org/10.1049/ic:19970751

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature Reviews Neuroscience*, *3*(3), 201.

Driver, J., & Spence, C. (1998). Attention and the crossmodal construction of space. *Trends in Cognitive Sciences*, *2*(7), 254–262. https://doi.org/10.1016/S1364-6613(98)01188-7

Edworthy, J. (1994). Improving auditory warning design: Relationship between warning sound parameters and perceived urgency. *Human Factors*, *33*(2), 205–231.

Edworthy, J. (2013). Medical audible alarms: a review. *Journal of the American Medical Informatics Association*, *20*(3), 584–589. https://doi.org/10.1136/amiajnl-2012-001061

EEMUA. (1999). *Alarm Systems: A Guide to Design, Management and Procurement*. Engineering Equipment and Materials Users Association London.

Enns, J. T. J., Di Lollo V, & Di Lollo, V. (2000). What's new in visual masking? *Trends in Cognitive Sciences*, *4*(9), 345–352. https://doi.org/10.1016/S1364-6613(00)01520-5

Ferris, T. K., Penfold, R., Hameed, S., & Sarter, N. (2006). The Implications of Crossmodal Links in Attention for the Design of Multimodal Interfaces: A Driving Simulation Study. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *50*(3), 406–409. https://doi.org/10.1177/154193120605000341

Horng, W.-B., Chen, C.-Y., Chang, Y., & Fan, C.-H. (2004). Driver fatigue detection based on eye tracking and dynamic template matching. In *IEEE International Conference on Networking, Sensing and Control, 2004* (Vol. 1, pp. 7–12). IEEE.

HSE. (1997). *The explosion and fires at the Texaco Refinery, Milford Haven, 24 July 1994*. *HSE books*.

Laberge, J. C., Bullemer, P., Tolsma, M., & Reising, D. V. C. (2014). Addressing alarm flood situations in the process industries through alarm summary display design and alarm response strategy. *International Journal of Industrial Ergonomics*, *44*(3), 395–406. https://doi.org/10.1016/j.ergon.2013.11.008

Laming, D. (2010). Serial position curves in free recall. *Psychological Review*, *117*(1), 93.

Mercado, J. E., Rupp, M. A., Chen, J. Y. C., Barnes, M. J., Barber, D., & Procci, K. (2016). Intelligent agent transparency in human–agent teaming for Multi-UxV management. *Human Factors*, *58*(3), 401–415.

Oliva, A., Torralba, A., Castelhano, M. S., & Henderson, J. M. (2003). Top-down control of visual attention in object detection. In *Proceedings 2003 International Conference on Image Processing (Cat. No. 03CH37429)* (Vol. 1, p. I-253). IEEE.

Prinet, J. (2016). Attentional Narrowing: Triggering, Detecting and Overcoming a Threat to

Safety.
Rogovin, M. (1979). Three mile Island. *Technology and Society*, *7*(28), 9–11. https://doi.org/10.1109/TS.1979.6500414
Russ, A. L., Zillich, A. J., Melton, B. L., Russell, S. A., Chen, S., Spina, J. R., … McManus, M. S. (2014). Applying human factors principles to alert design increases efficiency and reduces prescribing errors in a scenario-based simulation. *Journal of the American Medical Informatics Association*, 21(e2), e287–e296.
Shapiro, K. L., Raymond, J. E., & Arnell, K. M. (1997). The attentional blink. *Trends in Cognitive Sciences*, *1*(8), 291–296. https://doi.org/10.1016/S1364-6613(97)01094-2
Simons, D. J., & Levin, D. T. (1997). Change blindness. *Trends in Cognitive Sciences*, *1*(7), 261–267.
Sinha, R., & Swearingen, K. (2002). The role of transparency in recommender systems. In *CHI'02 extended abstracts on Human factors in computing systems* (pp. 830–831). ACM.
Stankovic, A., Aitken, M. R. F., & Clark, L. (2014). An Eye-tracking Study of Information Sampling and Decision-making Under Stress: Implications for Alarms in Aviation Emergencies. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *58*(1), 125–129. https://doi.org/10.1177/1541931214581027
Stanton, N. A., & Stammer, R. B. (1998). Alarm initiated activities: matching visual formats to alarm handling "tasks." *International Journal of Cognitive Ergonomics*, *2*(4), 331–348.Wickens, C. D. (2008). Multiple Resources and Mental Workload. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, *50*(3), 449–455. https://doi.org/10.1518/001872008X288394
Treisman, A. (1985). Preattentive processing in vision. *Computer vision, graphics, and image processing*, 31(2), 156-177.
Wickens, C. D., McCarley, J., & Steelman, K. S. (2009). NT-SEEV: A model of attention capture and noticing on the flight deck. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* (Vol. 53, pp. 769–773). Sage Publications.
Wogalter, M. S., Conzola, V. C., & Smith-Jackson, T. L. (2002). Research-based guidelines for warning design and evaluation. *Applied Ergonomics*, 33(3), 219–230.

APPENDICES

APPENDIX A Debriefing Form for Experiments 1 and 2

The detection of visual and auditory alarms in close temporal proximity - Debriefing Questionnaire

Your responses will be accessible only to the researcher listed on the consent form.

Part 1: Participant background

1. **Age**

   ____

2. **Gender**
   Mark only one oval.
   - ( ) Male
   - ( ) Female
   - ( ) Prefer not to say
   - ( ) Other: ______________________

Part 2: Experiment feedback

3. **Did you find it more difficult to detect two alarms that appeared close together in time, compared to alarms that were separated further?**
   Mark only one oval.
   - ( ) Yes
   - ( ) No
   - ( ) Maybe

4. **Please explain.**

   ____
   ____
   ____
   ____
   ____

5. **Which alarm pair was the most difficult to detect?**
Mark only one oval.

- Visual followed by visual
- Visual followed by auditory
- Auditory followed by auditory
- Auditory followed by visual
- Not sure

6. **Please explain.**

7. **How difficult was it to perceive and report alarms in the order they appeared?**
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Extremely easy											Extremely difficult

8. **Please explain.**

9. **Were you surprised by the onset of the alarm flood?**
Mark only one oval.

- Yes
- No

10. **Did you find it more difficult to detect alarms during the alarm flood?**
Mark only one oval.

- Yes
- No
- Maybe

11. **Please explain.**

____
____
____
____
____

12. **Was it more difficult to detect alarms early or late during the alarm flood?**
Mark only one oval.

- ⬭ Earlier
- ⬭ Later
- ⬭ No difference
- ⬭ Not sure

13. **Please explain.**

____
____
____
____
____

14. **Did your ability to detect alarms change over time throughout the experiment?**
Mark only one oval.

- ⬭ Yes
- ⬭ No
- ⬭ Maybe

15. **Please explain.**

____
____
____
____
____

16. **Please rate your level of fatigue before the alarm flood.**
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Not at all	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Extreme fatigue

17. **Please rate your level of fatigue during the alarm flood.**
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Not at all	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Extreme fatigue

18. **Please rate your level of fatigue after the alarm flood.**
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Not at all	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Extreme fatigue

19. **Please rate the difficulty of the judgment task (locating/identifying the delivery pad).**
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Extremely easy	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Extremely difficult

20. **Please rate the difficulty of the monitoring task over all (detecting the visual and auditory alarms).**
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Extremely easy	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Extremely difficult

21. **Please explain what made either task challenging.**

Other questions

22. **Please feel free to add any other comments about the simulation and the experiment in general.**

____
____
____
____
____

23. **Are you willing to receive recruitment emails for our future experiments (including this series of experiments in drone monitoring and other experiments in driving, industrial control, etc.)**

Mark only one oval.

- ⬭ All experiments
- ⬭ Only this series
- ⬭ No   *Stop filling out this form.*

Contact information

24. **Please enter your email address to receive recruitment emails.**

____

Powered by
Google Forms

APPENDIX B Debriefing Form for Experiment 3

Temporal limitations in the detection of concurrent and sequential alarms - Debriefing Questionnaire

Your responses will be accessible only to the researcher listed on the consent form.

Part 1: Participant background

1. **Age**

   ____

2. **Gender**

   Mark only one oval.

   - ◯ Male
   - ◯ Female
   - ◯ Prefer not to say
   - ◯ Other: ______________________

Part 2: Experiment feedback

Please rate the difficulty of the following alarm combinations.

3. **Two concurrent visual alarms**

   Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Extremely easy	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Extremely difficult

4. **Three concurrent visual alarms**

   Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Extremely easy	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Extremely difficult

5. **Four concurrent visual alarms**
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Extremely easy	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Extremely difficult

6. **Five concurrent visual alarms**
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Extremely easy	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Extremely difficult

7. **Six concurrent visual alarms**
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Extremely easy	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Extremely difficult

8. **Two sequential visual alarms**
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Extremely easy	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Extremely difficult

9. **Three sequential visual alarms**
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Extremely easy	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Extremely difficult

10. **Four sequential visual alarms**
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Extremely easy	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	⬭	Extremely difficult

11. **Five sequential visual alarms**

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Extremely easy											Extremely difficult

12. **Six sequential visual alarms**

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Extremely easy											Extremely difficult

13. **Two concurrent auditory alarms**

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Extremely easy											Extremely difficult

14. **Three concurrent auditory alarms**

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Extremely easy											Extremely difficult

15. **Two sequential auditory alarms**

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Extremely easy											Extremely difficult

16. **Three sequential auditory alarms**

Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Extremely easy											Extremely difficult

17. **Were you surprised by the onset of the alarm flood?**
Mark only one oval.

- Yes
- No

18. **Did you find it more difficult to detect alarms during the alarm flood?**
Mark only one oval.

- Yes
- No
- Maybe

19. **Please explain.**

20. **Was it more difficult to detect alarms early or late during the alarm flood?**
Mark only one oval.

- Earlier
- Later
- No difference
- Not sure

21. **Please explain.**

22. **Did your ability to detect alarms change over time throughout the experiment?**
Mark only one oval.

- Yes
- No
- Maybe

23. **Please explain.**

24. **Please rate your level of fatigue before the alarm flood.**
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Not at all	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Extreme fatigue

25. **Please rate your level of fatigue during the alarm flood.**
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Not at all	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Extreme fatigue

26. **Please rate your level of fatigue after the alarm flood.**
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Not at all	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Extreme fatigue

27. **Please rate the difficulty of the judgment task (locating/identifying the delivery pad).**
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Extremely easy	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Extremely difficult

28. **Please rate the difficulty of the communication task (responding to the call sign).**
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Extremely easy	◯	◯	◯	◯	◯	◯	◯	◯	◯	◯	Extremely difficult

29. **Please rate the difficulty of the monitoring task over all (detecting the visual and auditory alarms).**
Mark only one oval.

	1	2	3	4	5	6	7	8	9	10	
Extremely easy											Extremely difficult

30. **Please explain what made these tasks challenging.**

Other questions

31. **Please feel free to add any other comments about the simulation and the experiment in general.**

32. **Are you willing to receive recruitment emails for our future experiments (including this series of experiments in drone monitoring and other experiments in driving, industrial control, etc.)**
Mark only one oval.

- All experiments
- Only this series
- No *Stop filling out this form.*

Contact information

Please leave your email address in the excel sheet provided to be contacted for future studies.

APPENDIX C Debriefing Form for Experiment 4

Default Question Block

Your responses will be accessible only to the researcher listed on the consent form.

Participant background

Age

Gender

- Male
- Female
- Other

Block 1

Overall assessment

Please rate the difficulty of the following tasks. (0 - extremely easy; 10 - extremely difficult)

	0	1	2	3	4	5	6	7	8	9	10
The judgment task (locating/identifying the delivery pad)											
The communication task (responding to the call sign DOI51)											
The monitoring task (detecting the visual and auditory alarms)											
Please rate the difficulty of the monitoring task with varying criticality (detecting the visual alarms withc varying colors and auditory alarms with varying pitches)											

Please explain what made these tasks challenging.

Please estimate the percentage of alarms (without varying criticality) that you:

0	10	20	30	40	50	60	70	80	90	100

	0	10	20	30	40	50	60	70	80	90	100
missed											
detected but reported incorrectly											
detected and reported correctly											
Total:											

Please estimate the percentage of alarms that you:

	0	10	20	30	40	50	60	70	80	90	100
missed											
detected but reported incorrectly											
detected and reported correctly											
Total:											

Block 2

Alarm criticality

Do you think the color or pitch of the alarms improved your performance? Please explain.

Can you describe any strategy you used in the tasks?

Block 3

Alarm flood

Did you find it more difficult to detect alarms during the alarm flood?

- Yes
- Maybe
- No
- It depends

Please explain.

Was it more difficult to detect alarms early or late during the alarm flood?

- Earlier
- Later
- No difference
- Not sure

Please explain.

What is most difficult about the alarm flood?

Block 4

Other questions

Please feel free to add any other comments about the simulation and the experiment in general.

Are you willing to receive recruitment emails for our future experiments (including this series of experiments in drone monitoring and other experiments in driving, industrial control, etc.)

- ○ All experiments
- ○ Only this series

○ No

Block 5

Please leave your email address in the excel sheet provided to be contacted for future studies.

Powered by Qualtrics