"I WANT TO TREAT THE PATIENT- NOT THE ALARM"
USER IMAGE MISMATCH IN ANESTHESIA ALARM DESIGN

Thesis/Project work submitted in partial fulfillment of the requirements for the MSc in Human Factors and System Safety

Karen Raymer, M.D., FRCP(C)

LUND UNIVERSITY
SWEDEN

Date of submission: 2012-06-04

"I WANT TO TREAT THE PATIENT- NOT THE ALARM"
USER IMAGE MISMATCH IN ANESTHESIA ALARM DESIGN

Karen Raymer, M.D., FRCP(C)

Under supervision of Johan Bergström and James M. Nyce, PhD

ABSTRACT

Although the development of physiologic monitors has improved the safety of patients undergoing anesthesia, a growing body of literature suggests that alarms function sub-optimally in supporting the human operator. A principle of cognitive systems engineering is that there is an image of the user "built in" to the design of any machine used in a joint cognitive (man-machine) system. This study explored the existence of a mismatch between the image of the user built into the anesthesia monitor alarm (machine-embedded image) and the anesthesiologist's own image of himself (user-described image) and whether any such user image mismatch contributed to coordination failure between anesthesiologist and alarm. The Participant Guide for the Carescape B850 Anesthesia monitor was analyzed to define machine-embedded images. Fourteen anesthesiologists were interviewed to explore user-described images and man-machine coordination. User image mismatch was observed in each of three alarm behaviours studied. Furthermore, user image mismatch was found to be related, as symptom of and contributor to the instances of coordination failure described by interviewees.

TABLE OF CONTENTS

LIST OF TABLES AND FIGURES

N.B. The images in Figures 9-12 were produced from a simulated setting.

List of abbreviations

AAMI	Association for the Advancement of Medical Instrumentation
BIS	bispectral index
BP	blood pressure
CSE	cognitive systems engineering
ECG	electrocardiogram
ECRI	Emergency Care Research Institute
$ETCO_2$	end-tidal carbon dioxide
FDA	Food and Drug Administration
FIO_2	fraction of inspired oxygen
GE	General Electric
HR	heart rate
IEC	International Electrotechnical Commission
ICU	Intensive Care Unit
mmHg	millimeters of mercury
OR	operating room
PACU	post-anesthetic care unit
PEEP	positive end-expiratory pressure

1. Introduction

The invasive surgical procedures which are very much a part of modern healthcare could not take place without the benefit of anesthesia. Though preventable mishaps attributable to anesthesia are rare, their importance is magnified by the sheer ubiquity of anesthesia. For example, some 28 million anesthetics are delivered in the United States each year (Stoelting & Miller, 2006). Physiologic monitors are essential for the safe practice of anesthesia, but play a disproportionate role in adverse events (Cooper, Newbower, Long, & McPeek, 2002).

Monitoring of patients under anesthesia involves the interpretation of data related to both the performance of the anesthetic machine and the physiologic condition of the patient. Equipment-related measurements directly capture variables of interest which include flows, circuit pressures, and partial pressures of gases. By contrast, physiologic measurements present unique challenges. Monitored variables[1] are often surrogates for less easily-measured variables that more closely reflect the physiologic process of interest. Furthermore, measurements of those surrogate variables must often be done indirectly and are particularly susceptible to artefactual perturbation (Takla, Petre, & Doyle, 2006). Finally, accepted normal ranges (and therefore, alarm thresholds) are difficult to establish due to both a lack of clinical evidence and the degree of context-sensitivity that applies to the dynamic environment of surgery and anesthesia (Table 1).

Built on such a tenuous foundation, it is not surprising that dynamic fault management presents a major problem in anesthesiology today (Edworthy & Hellier, 2006; Hagenouw, 2007; Schmid et al., 2011; Seagull & Sanderson, 2001). Growing awareness of the multifaceted impact of alarms on patient safety has galvanized the efforts of several large professional organizations, including the Association for the Advancement of Medical

[1] Monitored physiological variables include heart rate (HR) and rhythm, ST segment analysis, blood pressure (BP), oxygen saturation, end-tidal carbon dioxide tension ($ETCO_2$), temperature and bispectral index (BIS).

Instrumentation (AAMI, 2011), and the Emergency Care Research Institute (ECRI), which recently identified alarms as the "number one health technology hazard" for 2012 (ECRI, 2011).

The technical challenges of monitoring are compounded by historical and medico-legal factors. The field of anesthesiology saw its most profound improvement in safety result from the development of two physiological monitors (the oxygen saturation monitor and the end-tidal carbon dioxide monitor) which seems to have fostered an unquestioning belief in the safety benefits of alarms (Imhoff & Kuhls, 2006). Secondly, the use of alarms is mandated by law, and the specifics of design are dictated by the highly detailed, continually evolving and occasionally internally inconsistent standards dictated by the International Electrotechnical Commission (IEC). Thirdly, anesthesia and critical care monitors developed in parallel, lacking site-specific alarm algorithms despite distinct differences in the dynamics and context of each environment. Finally, and perhaps most importantly, manufacturers are motivated to ensure a zero false negative rate, often at the expense of an intolerably high false positive rate (Imhoff, Kuhls, Gather, & Fried, 2009). One result is that despite a net safety effect that may be adverse, (e.g. Bliss & Dunn, 2000; Dixon, Wickens, & McCarley, 2007; Meyer, 2001) risk may be tolerated because liability is displaced from manufacturer to the practitioner when he either disables the alarm audio or becomes immune to the auditory alert (Imhoff &Kuhls, 2006; Weil, 2009; Woods, 1985).

The advances of the past 20 years in microprocessing, artificial intelligence and human factors engineering have been applied to the "alarm problem" in anesthesia. These advances have seen limited penetration into the clinical setting, in large part due to the reluctance of manufacturers to change technologies in the absence of strong evidence supporting improved safety of those newer technologies (Imhoff & Kuhls, 2006; Kiefer & Hoeft, 2010). An opportunity exists to study the alarm problem from the larger perspective of cognitive systems engineering, and in doing so, understand the complexity of this joint cognitive system activity.

Table 1. Challenges of physiologic monitoring in anesthesia

Surrogate variable that is measured	Variable of true or greater interest	Measurement technique of surrogate variable: Direct or Indirect?	Artefactual influences	Certainty of "Normal Range"
Blood Pressure	Cardiac output (flow)	a) Indirect (blood pressure cuff) b) Direct (arterial cannula placed invasively into a peripheral artery)	Numerous: a) incorrect cuff size b) movement c) incorrect transducer level	Highly uncertain: Dependent on patient and surgical characteristics. Significant lack of guiding evidence
Alveolar carbon dioxide (CO_2) tension	Arterial blood CO_2 levels. To measure CO_2 levels in the blood would require an arterial puncture. Alveolar levels can be measured non-invasively through the expired breath but there is an unpredictable gradient between alveolar and arterial carbon-dioxide levels.	Indirect: unable to sample directly from alveolae; the end of an expiration (end-tidal) best approximates the gas that was in the alveolae.	Numerous: The measured value can be underestimated if true alveolar sample is not achieved, which may occur in rapid and/or shallow breathing pattern. Even if a true alveolar sample is achieved, it may not reflect arterial CO_2 levels in conditions of decreased blood flow (shock, cardiac arrest).	Well-defined , however there is a high degree of context sensitivity depending on both patient (chronic lung disease) and surgical (laparoscopic) factors
Pulse Oximetry	Arterial oxygen saturation; arterial oxygen content	Indirect: compares absorption of two different wavelengths of light across superficial skin capillaries as a reflection of the percentage of oxygenated and de-oxygenated hemoglobin.	Numerous: low temperature, low blood flow, skin pigmentation or nail polish	Well-defined and not context-specific
Bispectral Index (BIS)	Depth of anesthesia; degree of unconsciousness	Indirect: a composite of electro-encephalographic (EEG) signals measured through the scalp: actual formula is proprietary, a "black box" for clinicians	Numerous and poorly-understood	Highly uncertain: very low sensitivity and specificity for predictive value of BIS-derived numeric. No evidence for improved outcome when using BIS monitor to measure depth of anesthesia.

2. Literature Review

2.1 The scope of the problem: Anesthesia alarms in the clinical setting

2.1.1. False alarms, disabled alarms and other challenges to utility

Hagenouw (2007) traces the evolution of anesthesia alarms from the 1960's, when the anesthesiologist[2], finger on the patient's pulse, *was* the monitor. A vast array of devices has been developed since then. However, any gains made by "miniaturization and combination of parameters into one unit [have been] more than offset by the continual introduction of more monitors and devices" (Hagenouw, 2007).

False alarms can be technical, caused by a faulty measurement, or clinical, whereby the measurement is accurate but the alarm thresholds are not appropriate for that particular situation. In the critical care environment, O'Carroll (1986) showed that only 8 of 1455 alarm-soundings represented an actual critical situation. More recently, Imhoff et al. (2009) showed that up to 90% of all alarms in the critical care environment were false positives, while Siebig et al. (2010) indicated that only 15% of alarms were clinically relevant. These results have been replicated in the anesthesia environment. Kestin, Miller, and Lockhard (1988) showed that only 3% of alarms indicated actual patient risk, with 75% of them being patently spurious. Schmid et al. (2011) found that 80% of alarms occurring during anesthesia for cardiac surgery had "no therapeutic consequence". The irony of alarms designed to "never miss" is the paradoxical effect they can have on the user, who may resort to silencing the alarms altogether. Popular media takes note when a patient dies quietly adjacent to a disabled alarm (Kowalcyzk, 2011). Indeed, the Food and Drug Administration (FDA) received 566 reports of deaths related to monitor alarms between 2005 and 2008 in various critical care settings (Weil, 2009). Most incidences were associated with the disabling or silencing of alarm function.

---

[2] Much has changed since the early days of anesthesiology, including the fact that in many jurisdictions, nurse anesthetists monitor the patient, under the supervision of an anesthesiologist who may or may not be present. Indeed, other healthcare providers, including the surgeon, may be in the position of responding to anesthesia monitor alarms. The authors recognize this diversity, but for simplicity use "anesthesiologist" as a collective term.

Though one might wonder why a healthcare provider would take the risk of disabling an alarm, this behaviour has long been recognized. In 1985, the following passage appeared in *Ergonomics*:

> " [*If alarms] become a 'normal occurrence', then the alarm will become part of the routine flow of action, and it will not function to break attention away from other mental and physical activity. In aggregate, trying to make all alarms unavoidable redirectors of attention overwhelms the cognitive processes involved in control of attention and exacerbates the alarm problem. One kind of operational response to this should not really be surprising—practitioners ignore or turn off the alarms." (Woods, 1985)*

Phipps et al. (2008) identified some of the conditions that provoke anesthesiologists to violate the "rules" that would seem to govern their practice. The authors draw from the theory of behavioural economics, where both rule-following and rule-breaking are seen to incur costs as well as reap rewards:

> "*Violations occur when a person's attempts to optimize behavioural resources are not supported by the existing rules and breaching the rules incurs a lower overall cost to these resources. Where the perceived costs of violating-and the perceived benefits of compliance- are more salient, violations are less likely to occur".* (*Phipps et al., 2008)*

Adverse events where the anesthesiologist's actions were negatively influenced by the behaviour of (activated) alarms, either directly or indirectly, are much more elusive than the ones that make the headlines. Hagenouw (2007) outlines the negative effects of false alarms on the anesthesiologist's performance which include distraction, auditory fatigue and increased workload, each of which may impede attention to "real" clinical problems. The sheer ubiquity of alarms limits their utility: Momtahan, Hetu, and Tansley (1993) showed that anesthesiologists were unable to reliably identify the source or acuity of alarms from their daily work environment (cited in Hagenouw, 2007), a result which substantiates Mondor and Finley's (2003) earlier finding that the perceived urgency of auditory alarms did not match their clinical relevance. Healey, Sevdalis, and Vincent (2006) showed that the operating room environment can have many distractions and interruptions, to which false alarms contribute, and through which relevant alarms must be discerned. As disturbing as the FDA's report is, we must resist the superficial solutions. Rather than asking, "How can we design alarms that anesthesiologists *cannot* disable?" we must ask, "How can we design alarms that anesthesiologists will not *want* to disable?"

2.1.2. Human response to false alarms in the experimental setting

Warning alarms have an intuitive appeal, as the human operator is not believed to be well-suited to sustained visual scanning (Baig, Gholamhosseini, Kouzani, & Harrison, 2011). In a simulated cockpit, even an imprecise automated warning system improved pilots' performance compared to those who did not use the system. The automated warnings, though imperfect, seemed to "lead pilots to inspect the raw data more closely" (Xu, Wickens, & Rantanen, 2007). However, this experimental warning system operated with a 17% error rate (misses and false alarms, in equal measure), providing relevance that medical device alarms do not come close to approaching.

The psychology literature has shown us that individuals adjust their responsiveness to alarms sharply according the perceived reliability of the alarms (Bliss, Gilson, & Deaton, 1995; Parasuraman, Hancock, & Olofinboba, 1997). Bliss and his colleagues applied the term "cry wolf effect" to this phenomenon, one that persistently finds its field manifestation categorized as "human error" [3] (e.g. Ferris & Sarter, 2011; Kowalcyzk, 2011; Nyssen & Blavier, 2007; Weil, 2009; Whittingham, 2004).

---

[3] Whittingham (2004, pp.171-175) describes a fatal rail crash where the conductor repeatedly "acknowledged" an alarm that was indicating the increasing proximity of a train ahead, without taking action the corrective action of slowing the train. The authors concluded that his action of depressing the acknowledge button had become a "conditioned response", a phenomenon that was not uncommon amongst conductors. They also found that the alarm system was "inadequate for its intended purpose." The conductor survived, but served 18 months in jail for manslaughter.

The global performance impact of alarms has been studied in a laboratory where volunteers were asked to perform a task (tracking a ball on a screen) while simultaneously monitoring a gauge denoting the status of a parallel system. The gauge had an alarm, of variable reliability, to indicate when the system was failing. The authors found that *overall system performance* was more negatively impacted by a false-alarm prone gauge than a miss-prone gauge (Dixon et al., 2007).

Bliss and Dunn (2000) examined the effect of "increasing primary task and alarm workload on alarm mistrust as reflected by alarm and primary task performance". In addition to supporting earlier work showing that individuals "probability-match their response rates to alarm system reliability", they found that "increasing primary task and alarm task workload degraded alarm response performance".

The parallels to the healthcare setting in many of the experimental studies are limited by both the setting (non-naturalistic) and the design, which has participants performing tasks that are unrelated to the alarm task. In an experimental simulation study inspired by the observed responses of health-care professionals to the (mostly false) alarms in a medical intensive care unit (ICU), Bitan and Meyer (2007) underscored the complex role that warning signals can play in guiding the operator's actions. Operators' responses (compliance, reliance and overall frequency of interventions) depended on both the characteristics of the warning system as well as how frequently the system required intervention. Even in an artificial environment, they were able to show that a warning signal was integrated with other data from the external environment and the operators' own internal resources.

2.1.3. Anesthesiologist and monitor alarms: An "uneasy relationship"

Though the operating room makes for a disobliging laboratory, several observational trials have added to our understanding of the "uneasy relationship" (Hagenouw, 2007) that anesthesiologists have with alarms. Seagull and Sanderson (2001) found that the response to alarms varied significantly depending on the type of surgery and phase of anesthesia, though the most common response across all phases was "ignoring". Alarms were most frequently ignored during anesthetic induction and emergence, the most dynamic and treacherous phases of care, but also the phases that engage the anesthesiologist most directly with the patient. The authors identified "the inability to filter out irrelevant alarm information in a useful way" as a major annoyance for practitioners, as silencing measures had to be non-selectively applied to all alarms. Although their data argued for improving context-sensitivity of alarms, the authors cautioned that "there will always be exceptions to a pre-programmed plan or pattern-recognition algorithm because of patient condition or because of conditions that emerge during a surgical procedure that would make a device's understanding of context incorrect.....making such systems unsafe" (Seagull & Sanderson, 2001).

Smith, Goodwin, Mort, and Pope (2003) used a variant of ethnography to explore how anesthesiologists enact sense-making in response to alarms. They found that anesthesiologists seamlessly integrate information from monitors and the physical examination of the patient, and that "which type is favoured at a particular moment ...depend[s] on context, circumstances and the [anesthesiologist's] prior expectations". The attitude towards alarms highlighted the novice-expert divide, with novices responding reflexively to outlying parameters while experts used the electronic data along with their own internal resources to form an image of the patient's current condition and where that condition might be headed. In experts, the authors observed scepticism toward out-of-range monitor data. Even when not patently false, alarms were often redundant, the anesthesiologist having already detected the changing clinical state. The authors concluded that while a monitor may meet its manufacturer's specifications, it requires significant input from the practitioner to enable it to "work" within the human-machine system. Corroborating findings were published by Wright, Dawodu, Segall , Taekman, and Mark (2011) who examined the perceived value of alarms in alerting the anaesthesiologist (or nurse-anesthetist) to the onset of a crisis. Among nine critical incidences discussed, none of the six anesthesia providers was able to recall being alerted to the problem by an alarm. The importance of waveform interpretation and the use of multiple data points in recognizing the onset of a crisis were recurring themes among the interviewees. Importantly, the confusing nature of the many false alarms that occurred, even in a real crisis, was another recurrent theme. Nyssen and Blavier (2007) examined the nature of error-detection in the anesthesia setting and found that routine monitoring of the environment was the most frequent method that resulted in detection of errors. Different detection methods were used for different types of errors, with alarms being associated most frequently with the detection of technical and procedural errors. Echoing the work of Smith et al. (2003), this study showed that experts had a richer repertoire of error detection methods compared with novices.

Clearly, though anesthesiologists are mandated by law to use alarms, they cannot be mandated to find them useful. The technological advances of the previous two decades have yielded scant onsite progress in the ability of alarms to support the joint human-machine cognitive system. What stands in the way of change are not technological issues. Rather it is the risk-reduction imperative of equipment manufacturers results in a commitment to both the status quo and a zero- false-negative rate. Accordingly, anesthesiologists today are using systems that would be easily recognized by the pioneers of the specialty who practiced when those physiological monitors were first developed and put to use (Kiefer & Hoeft, 2010; Solet & Barach, 2012).

2.1.4. Current strategies for improving alarm performance

Imhoff and Kuhls (2006) reduced the problem to two issues: "The correct identification of a situation that needs to be alarmed...and the consistent and unambiguous annunciation of this alarm." Existing strategies aimed at improving alarm performance remain largely within that framework and can be categorized as follows:

- *Detection and decision problem:* Efforts to reduce technical and clinical false positives are outlined in detail (Imhoff & Kuhls, 2006). They include technical measures (improving device hardware), simple statistical methods (linear filtering or weighted averaging of data) and sophisticated approaches that utilize artificial intelligence and fuzzy logic. Integration of device data can reduce false positives. For example, an electrocardiogram (ECG) asystole alarm could be overridden if the oximeter waveform showed pulsatility. Other approaches have attempted to imitate the flexibility of the human thought processes, but with a greater response speed. Baig et al. (2011) developed an elegant fuzzy logic approach to the early detection of hypovolemia in surgical patients. Other simpler approaches have been shown to reduce false positive alarms. Watson and Sanderson (1998) modelled that eliminating low-priority alarms during induction and emergence would reduce the alarm rate by 70 and 90 percent respectively (cited in Seagull & Sanderson, 2001). This approach is constrained by the degree to which pre-programmed alarm priorities could successfully reflect the moment-to-moment importance of those alarms through a range of clinical situations. Graham and Cvach (2010) described a quality improvement project undertaken in an ICU setting in order to address perceived alarm desensitization on their unit. The intervention, whereby nurses were trained to individualize patients' alarm parameter limits, resulted in a 43% reduction in critical alarm incidence. Gorges, Markewitz, and Westenskow (2009) showed that delaying alarms by fourteen seconds reduced the incidence of critical alarms by 50%.
- *User interface and human factors problem:* The IEC has attempted to increase the salience of alarms through standardization of the "melody" attached to each individual alarm. Unfortunately, very low recognition rates have been reported, even after training (Sanderson, Wee, & Lacherez , 2006; Wee & Sanderson, 2008). Speech alarms are highly salient but limited by their potential to cause anxiety in conscious patients (Edworthy & Hellier, 2006). Regarding interface, efforts have centred on attempts to reduce demands

on auditory perception and visual scanning while taking into account the requirements for multi-tasking in the anesthesiologists' workflow. For example, head-up multisensory displays have been modelled after those used by fighter-pilots (Kiefer & Hoeft, 2010). Other efforts have focussed on hybridization of monitors and alarms, where physiological data is continuously broadcast, through sound (sonification) or vibrotactile display, allowing the clinician to decide the boundaries of safety in any given situation (Ferris & Sarter, 2011).

2.2. Theoretical foundations

2.2.1. Single-stage signal detection theory

The ideal anesthesia monitor alarm would detect all abnormal events (100% sensitivity) without ever sounding an inappropriate alert (100% specificity). Unfortunately, that "all reasoning and decision-making takes place in the presence of some uncertainty" (Heeger, 1997) is a truth that can be applied to machines as well as to the humans who design them. During the design of any alarm system, choices are made that result in a trade-off between sensitivity and specificity. Signal detection theory helps us to understand the factors that both inform those choices and determine their impact on performance.

Most physiologic processes can be representing as occurring within a normal range. Alarms are triggered when the reading lies outside that preset range. Each machine response can be categorized in one of four possible ways (Figure 1).

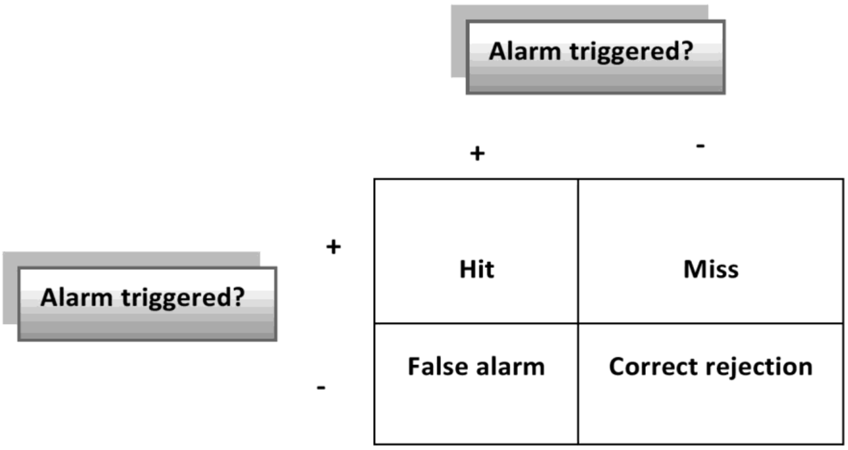


Figure 1. Categories of alarm performance

High and low alarm limits present separate signal detection challenges, creating additional complexity when compared to an alarm, such as a smoke alarm, that alerts abnormality in a single direction. Heart rate (HR) is of the most fundamental variables monitored under anesthesia. The principles outlined in Heeger's (1997) review of signal detection theory can be exemplified using a heart rate monitor alarm.

Even when the heart rate is normal, an outlying measurement may occur to trigger an alarm. This would be classified as *noise*. Figure 2 shows hypothetical probability of occurrence curves for a machine that is monitoring HR in a group of normal patients and a group of tachycardic[4] patients. If the "noise" and "signal plus noise" distributions shared no common territory then the monitor would reliably discriminate, providing only "hits" and "correct rejections".

[4] A tachycardic patient has an abnormally fast heart rate.

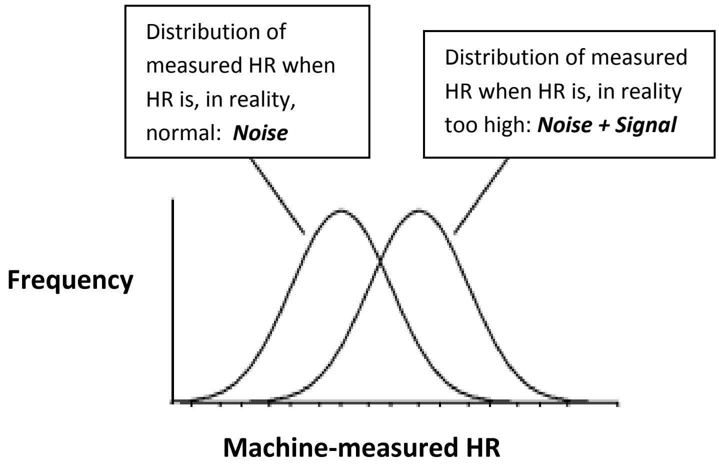


Figure 2. Probability of occurrence curves for normal and tachycardic patients. Reproduced from Heeger (2007) with permission.

The discriminability index, (d'=separation/spread) represents how effectively the monitor can distinguish signal from noise. Improving the strength of the information will increase the separation between curves[5]. Reducing noise will narrow the spread of each curve. Either manoeuvre will serve to reduce the overlap between the two curves, increasing d'. It is such an overlap that forces compromises to be made, compromises that are defined by the "criterion".

---

[5] It is not easy to make improvements in d', which is often fixed by both hardware limitations and our own limited understanding of the ways in which a true event differs from normal with respect to the patterns it will present to the monitor. This is an area in which more research clearly needs to be done if clinically appropriate alarms are to be developed.

The criterion is the strength of response required to trigger an alarm. The alarm designer sets the "criterion" according to the perceived importance of avoiding a miss[6]. A very low criterion will ensure that few or no abnormalities are missed, but will also result in many false positives. Figure 3 shows us the impact of shifting criteria, while illustrating the reality that as long as there is overlap between curves, no criterion can produce "zero misses and zero false alarms".

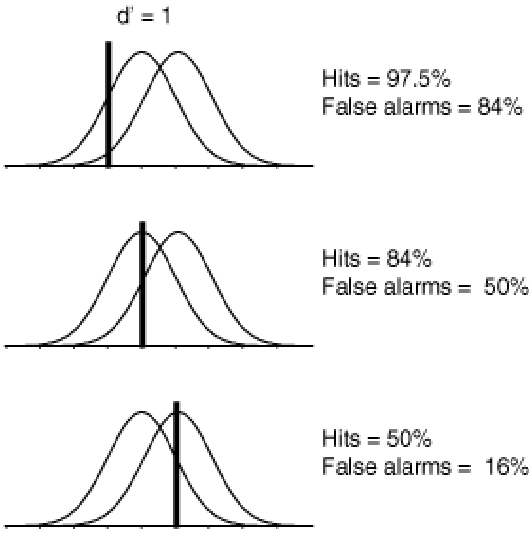


Figure 3: Effect of shifting the criterion. Reproduced from Heeger (2007) with permission

2.2.2. Signal detection in a machine-alerted human monitoring system

With devices like anesthesia monitor alarms, the machine is just one part of a human machine-alerted monitoring system. Sorkin and Woods (1985) expand signal detection theory to model a two-staged monitoring system in an effort to shed light on "how human or machine subsystem characteristics affect the performance of the overall alerted-monitor system." They developed a series of modified receiver operator curves to define the properties of a two-staged system and characterize theoretical interactions between the automated and human monitors

[6] The criterion determination is influenced by the a-priori probability of a signal and the risk/benefit ratio of each of the decision outcomes elaborated in Figure 1.

within that system. By necessity, assumptions were made. Each of the following findings should be viewed as reflecting a specific set of assumptions:

(1) Assuming equivalent sub-system discriminability (d'), overall system performance is 1.414d'. This 41% advantage over the performance of a single-staged system falls short of the sum of the performance of the two parts.[7]

(2) If the human operator increased his criterion in response to a high false positive rate (low criterion) in the machine, the dual system would cease to be effective except within a very limited range of low output rates from the machine.

(3) When the operator is overloaded (by frequent alerts in the face of other concurrent roles), he might use strategies to task-shed. He might attend only a subset of machine responses. Alternatively, he might observe all alerts, but with reduced detectability (d').

The authors postulate that in the first and third situations, the human's performance (criterion choice and discriminability) are *dependent on* the machine criterion. In this setting, the system would become constrained, optimally performing within very limited boundaries.

This highly theoretical research leads to a most practical conclusion, namely: "Optimal criterion placement for one stage of a two-stage detection system may not be identical to optimal placement in the single-stage case" (Sorkin & Woods, 1985). In a system where a machine is alerting a human supervisor, the machine's criterion should be set to optimize overall system performance. The work of Sorkin and Woods (1985) was referenced in a later study by Bustamante, Bliss, and Anderson (2007), who used signal detection theory to examine the effects of varied alarm thresholds and workloads on humans performing a complex task. As expected, their results confirmed that humans respond more quickly to alarms when using a system with the lowest rate of false alarms. Unexpectedly, and contradicting the earlier work of Bliss and

---

[7] Pollacks and Madans (1964) examined the performance of a two-stage detection system, finding that optimal system performance occurred when each stage had equal discriminability and equal criteria. Interestingly, they found that system performance was 1.2d', just slightly less than the theorized maximal performance (cited in Sorkin & Woods, 1985).

Dunn (2000), they found that the benefit on overall performance was lost, especially under high workload conditions, where the operators tended to miss conditions that were not alarmed. The authors postulated that when the system was more reliable (low false alarm rate), "operators over-relied on it to warn them about imminent danger but disregarded the fact that the system would likely miss more potential problems" (Bustamante et al., 2007). It is important to note that the system with the highest rate of false alarms in this study had a 65% false alarm rate, still much lower than those reported in health care. We don't know what impact a 90% false alarm system would have had on performance in the high workload condition in this experimental setting, or how these factors would manifest their impact in a naturalistic setting.

The two-staged signal-detection model, with machine and human operating in series, must be applied cautiously to the anesthesia environment. This is because an anesthesiologist often acts in *parallel* to the machine, monitoring the machine's responses as well as the noisy raw input data. Furthermore, the anesthesiologist's interpretation of that mutually-available data is modified by his own biases, past experiences, and any number of other human factors. Finally, rarely does an anesthesiologist view a parameter as "normal" or "abnormal", but more often as a shade of grey in between. Masalonis and Parasuraman (2003) addressed this latter issue as applied to air traffic control in their application of fuzzy signal detection theory techniques. They found that by assigning each event a fuzzy membership (between 0 and 1) they were able to reduce the computed false alarm rate in both human and machine systems.

The discussion of two-staged signal detection hints at the complexity of interplay between alarm, operator, task and workload. To better understand how the machine might influence and improve the performance of the human operator, we now turn to the cognitive systems literature.

2.2.3. Cognitive systems engineering

Early machines extended man's *physical* capabilities. The impact of the physical relationship of man and machine came to be addressed through the field of ergonomics. With the advent of microprocessors, machines also became an extension of man's *mental* function. The architecture of the interaction between man and machine was such that man no longer controlled the machine, as with a machine that manufactured widgets; instead, man "now had to control a process-or to monitor a self-controlling process" (Hollnagel & Woods, 1983).

With the introduction of cognitive systems engineering (CSE), Hollnagel and Woods (1983) addressed the cognitive impact of the machine on the man-machine system. Together, the human and machine comprise a joint cognitive system[8]: It receives data from the environment then makes a plan to achieve a goal. Its arrived-upon plan is both data-driven and concept-driven, the latter based on its own knowledge of the world apart from the specific input data at any given moment. Hollnagel and Woods (1983) outline principles of joint cognitive system design that remain relevant today:

- Man is a naturalistic (not "rational") decision maker.
- Human fault-finding performance varies with the degree to which the operator is engaged within the system control-loop. Whether system- engagement of the operator improves or degrades performance depends on workload.
- The total performance of the joint cognitive system cannot be explained by the performance of the individual sub-systems; accordingly, tasks should be assigned (to man or machine) with the function of the whole system in mind.
- Human "errors" are usually a symptom of a flawed system. Merely automating what humans are "not good at" can merely shift or expose other system vulnerabilities.

Concept-driven behaviour hinges on one's internal representation of the environment. That the human has a "model" of the system in which he is operating (that may be less than complete or accurate) seems intuitive. What this model is and the role it plays in human-machine

[8] The machine itself is a cognitive system, as is the operator. Working together, they are a joint cognitive system.

interaction has been the issue on which CSE research has traditionally focused. Hollnagel and Woods (1983) have extended (and reversed) this logic and posited that the machine itself has its own "image" of the operator built into the very design of the machine. Unfortunately, "the system's image of the user...is virtually never explicitly designed to enhance the joint function of man and machine", and thus creates "mismatches between man and machine" (Hollnagel & Woods, 1983).

The authors stress that machine-designers must first understand the nature of human cognitive functioning in order to design machines that can be compatible with and support the cognitive functioning of the human, and ultimately, that of the joint cognitive system. Without that key step, the human will be left to adapt to the machine, with predictable results. Applying these concepts to the field of anesthesia monitors, Kiefer and Hoeft (2010) state that anesthesia-related adverse events are "not random or fateful events, but can be viewed as a systematic result of man-machine interaction. Thus, devices should map to the human thought process, not vice versa…"

2.2.4. Data overload

Nearly thirty years after the work of Hollnagel and Woods, designing machines that adapt to the cognitive functioning of machine remains an elusive goal. One example of machine-human mismatch is data overload.

While today we can collect ever increasing volumes of data, technology has been less effective in helping the human unearth meaning from the mound of data to find the item that might advance or hinder the operator's goal(s). Despite considerable effort, further technological advances seem to have only deepened this "data-availability paradox" (Woods, Patterson, & Roth, 2002). In the operating room environment, "the continuous development of additional, better and more complex sensors and a tremendous increase in the number of displayed parameters, curves and alarms…has led to a widening gap between the load of information and the quality of its delivery" (Kiefer & Hoeft, 2010). The characterization of data-overload as a problem of "finding the significance of data when it is not known a priori what data from a large data field will be informative" (Woods et al., 2002) has particular relevance to this discussion as monitor alarms themselves can be seen as a "data-overload" condition, where the alarms themselves are the data items, and the operator has to extract meaning from noise.

One approach to the data-overload problem has been to harness the power of computers, which, unlike humans, "have the capability to monitor large volumes of diverse data rapidly" (Baig et al., 2011). Though appealing, this approach can overlook the true challenge of data overload which is context sensitivity, the fact that data finds its meaning in its relationship to other data, to goals and to expectations. Rather than seeing humans as a weak link, Woods et al. (2002) assert that people are "highly skilled at directing attention to aspects of the perceptual field that are of high potential relevance", even when what is relevant reflects changing contexts and goals. Woods et al. (2002) examine how people can focus on what is important to better understand how machines get in the way of those natural abilities. To this end, the authors highlight three fundamental skills unique to human cognition:

- Perceptual organization (lumping the data into meaningful groups)
- Attentional control (knowing where to focus; keeping focus simultaneously granular and global)
- Anomaly-based processing (noting departures from typicality as well as relative changes or trends)

Woods et al. (2002) conclude that machines must be designed to support (not replace) man's innate context sensitivity and only then will the problem of data overload recede. They outline necessary design constraints that emphasize the organization and conceptualization (but not the minimization) of data. The authors maintain that humans alone are able to achieve context-sensitivity. They identify attempts to offload context sensitivity to machines as misguided, yielding brittle and limited results. They categorize the methods of building "context sensitivity" into machines as follows:

- Reducing the amount of data available to the operator
- Static prioritization: deciding in advance which classes of data will take priority
- Trying to build intelligence into the machine so that it can indicate only what is important
- Syntactic approaches, where keywords within the content trigger alarms priority

Yet is an alarm not already, by definition, a machine-generated attempt at context sensitivity, the machine trying to tell the operator what is interesting? It is not surprising then, that improved context-sensitivity is a frequently-applied approach to the alarm problem in anesthesia. In order to better understand how to manage the problems associated with data overload in the joint cognitive activities of anesthetic care we can turn to the literature on the coordination among participants in joint cognitive activity.

2.2.5. Coordination in joint cognitive activity

Coordination has been defined as "the intent to work together to align goals and to invest effort to sustain common interests" (Klein, Feltovich, Bradshaw, & Woods, 2005). Participants may have to relax or even sacrifice "some shorter-term local goals in order to permit more global and long-term goals to be addressed". The authors define three requirements for effective coordination in joint activity:

- *Interpredictability*: One must be able to predict the actions of the other in any given situation. Interpredictability is the result of parties having a "shared script", from which they "form expectations about how and when others will behave". Interpredictability is also contingent on each party being able to view the situation from the other party's perspective.

- *Common ground*: The parties must negotiate a shared set of knowledge, beliefs and assumptions which "support(s) the interdependent actions in...joint activity". Key facets that require solid common ground include "the roles...of each participant; the skills and competencies of each participant; the goals of the participants; and the stance of each participant (e.g. his or her perception of time pressure, level of fatigue and competing priorities)".
- *Directability*: Directability refers to the ability of one actor within the joint cognitive system to influence the actions of another actor due to shifting goals and conditions. A simple, fixed response to a given stimulus creates an inflexible system. Directability is one way of allowing moderation of the programmed response when the setting dictates.

During joint activity, communication between participants can occur on two levels: There is the "task-work" that relates directly to the joint activity and the "teamwork" which serves to keep the choreography of the activity on track. Coordination is particularly important as the activity transitions through phases. Participants must have ways of signaling to each other and judging when the other participant is at an "interruptible" phase of his own tasks. The toll exacted, in time and effort, by these efforts to maintain choreography has been termed "coordination cost".

The authors identify the loss of common ground as the most frequent cause of joint activity breakdown. Some of the most salient causes of loss of common ground are:

- Discrepant access to data
- Lack of clarity regarding the joint goal
- Lack of awareness of differing stances between workers (workload, competing priorities)
- Confusion over who knows what

Klein et al. (2005) highlight six unique challenges[9] that must be overcome to allow successful joint cognitive system activity between man and machine. For example, "by making machine agents more adaptable (to context), we also make them less predictable". Another challenge lies in the area of attention management: How can the machine communicate status changes to its human partner while allowing humans to work without unnecessary interruption?

The analytic focus on coordination shifts the object of analysis from humans and technology as arbitrary self-contained entities of the joint operating system. Instead, humans as well as technological artifacts are all actors in the joint cognitive activity in which they engage.

[9] The six challenges outlined by Klein et al. (2005) for joint man-machine activity are:

(1) Achievement of basic compact
(2) Mutual predictability
(3) Goal negotiation
(4) Phase coordination
(5) Attention management
(6) Control of coordination costs

3. Thesis Question

To date, methods to address the alarm problem in anesthesia have been centred within the existing framework of alarm operation: Improving physiologic data detection, fine-tuning decision-making, finessing context-sensitivity and streamlining user interface. Though critical to the development of alarms that could respond more accurately and less intrusively, these granular level efforts are predicated on a linear interaction between man and machine (Figure 4).

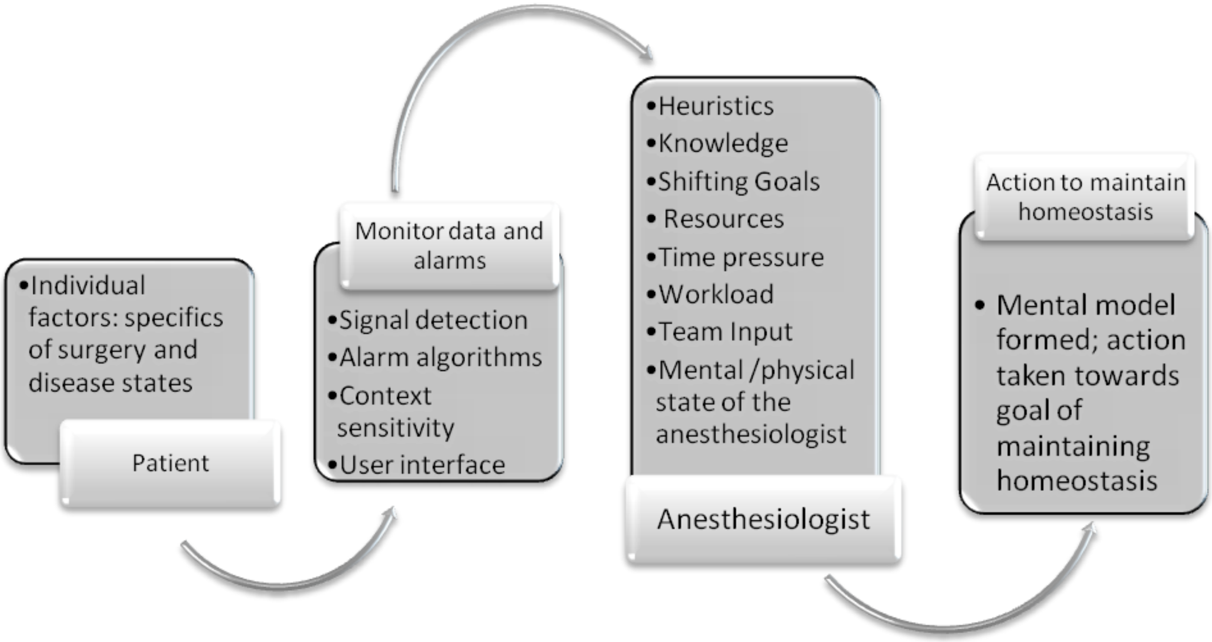


Figure 4. Linear modeling of the interaction between the patient, machine and anesthesiologist

The role of the alarm in a dual monitoring system calls for a more nuanced approach to design and development than the simple alerting of the operator to out-of-range variables. Understanding how cognitive systems engineering applies to anesthesia alarms could provide a theoretical base from which to build alarms that function as "team players" within a machine-supported human monitoring system.

Anesthesia alarms can be viewed as having an image of the user's characteristics that is built-in, or embedded within the alarm (Hollnagel & Woods, 1983). This "machine-embedded image" is the intentional or unintentional result of choices made during the design of the alarm and is expressed through the function of the alarm. The user himself has a corresponding image, how he defines himself with respect to those same characteristics. This is the "user-described image". Encompassing the previously described engineering concepts of user image and coordination, this project will study the following question:

> ***Is there a mismatch between the machine-embedded image and the user-described image of the anesthesiologist and if so, is the impact of that mismatch observable within the framework of coordination failure?***

4. Methods

4.1. Study Overview and Hypothesis

Anesthesia alarms remain an unresolved clinical problem with which the anesthesiologist grapples on a daily basis. In this study, "the alarm problem" was examined from the cognitive systems perspective, based on the hypothesis that the machine-embedded image is imperfectly aligned with the anesthesiologist's user-described image, creating a user image mismatch (Figure 5). It was further hypothesized that user image mismatch plays a role in coordination failures between anesthesiologist and anesthesia machine.

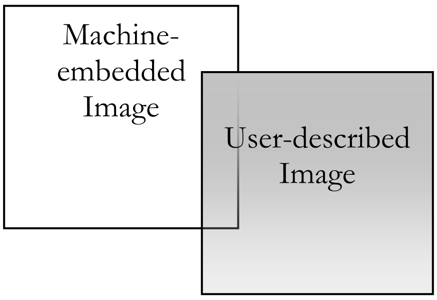


Figure 5. "User image mismatch" in anesthesia alarms

The exploration of user image mismatch occurred from two perspectives. Firstly, the investigator determined whether user image mismatch could be identified between a commonly-used anesthesia monitor, the General Electric (GE) Carescape B850, and a group of anesthesiologists, using the following methods:

- Document analysis of the participant guide for the Carescape B850 anesthesia monitor was used to identify the machine-embedded image. Confirmation of the machine-embedded image was made through investigator observation of alarm performance in the clinical setting.

- Insight into the specific design choices that resulted in the relevant machine-embedded image features was sought through focused interview(s) of the principal alarm engineer from GE.
- Corresponding user-described image were determined through focused interviews of fourteen anesthesiologists.
- Comparison of the machine-embedded image with the user-described image was used to confirm or refute the existence of user image mismatch between the Carescape B850 and a group of anesthesiologists who use that same system.

Secondly, the impact of user image mismatch was explored through its role in coordination failure between anesthesiologists and anesthesia alarms.

- (The same) fourteen anesthesiologists participated in in-depth interviews where their interactions with anesthesia alarms were explored.
- Those interactions were interpreted within the theoretical framework of "coordination" provided by Klein et al. (2004).
- The investigator identified whether user image mismatch played a role in the experiences of coordination failure (Figure 6).

It is important to understand the defined "images" (both machine-embedded and user-described) are hypothetical. Of first order generation, this concept has not been previously defined in the literature. The machine-embedded images were drawn from a merging of the investigator's experience with the above-identified resources and were identified prior to conducting the interviews.

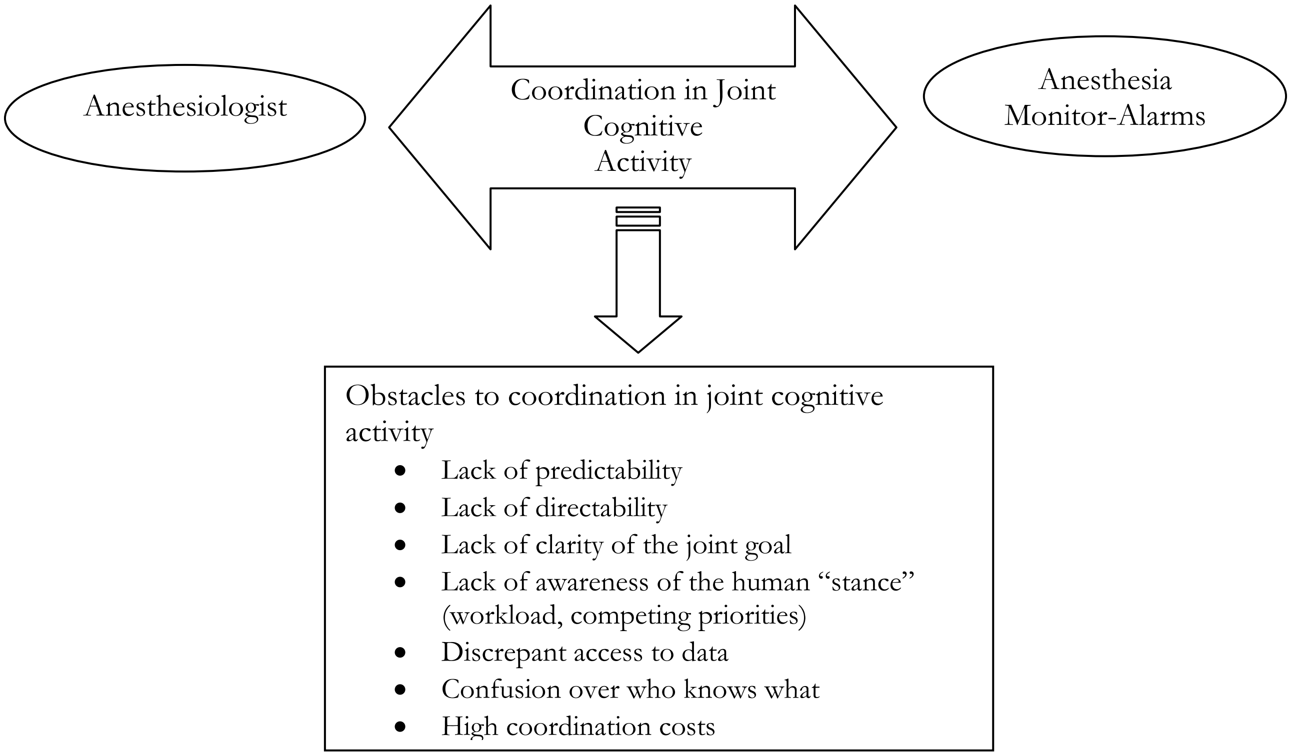


Figure 6. Obstacles to coordination in joint cognitive activity as applied to anesthesia alarms (Klein et al., 2004)

4.2. Sample

- Fourteen volunteer informant/interviewees from Hamilton Health Sciences, a large (1500 bed) university-affiliated teaching hospital in Canada. All participants were in active practice and regularly used the Carescape B850 monitor which had been installed in their hospital in August, 2010
- The investigator-observer, an anesthesiologist familiar with the above equipment
- The principal alarm engineer from GE who worked on the Carescape B850 monitor
- The "Participant Guide" for the Carescape B850 monitor

4.3. Materials

Four distinct instruments were used in the collection of data:

a) *Document analysis*

b) *Observation*

c) *Focused interviews*

d) *In-depth (ethnographically-inspired) interviews*

4.4. Procedures

The study design was a single case study. The unit of analysis was the "user image", from the perspective of both machine and user. The user image was analyzed for:

- mismatch (between man and machine)
- the role of mismatch in coordination failure of the joint cognitive system

The study was conducted as follows:

i. Research Ethics Board approval at the Hamilton Health Sciences was sought and received.

ii. Through analysis of the "Participant Guide" for the Carescape B850 Monitor (specifically Chapter 5), three unique machine-embedded images were constructed.

iii. The investigator confirmed that the behaviour of the alarms in the clinical setting matched the description in the manual and that the three identified machine-embedded images were associated with commonly-occurring clinical situations.

iv. The investigator pursued interviews with the GE alarm engineer regarding the rationale for the design choices that resulted in the machine-embedded images.

v. A focused interview tool was developed to elicit the user's own self-image(s). An in-depth interview tool was developed to explore the nature of coordination failure between anesthesiologist and anesthesia alarms in the clinical setting and the possible role of user image mismatch in coordination failure (Appendix A).

vi. Fourteen interview participants were identified after a "request for participation" letter was sent to anesthesiologists (Appendix B). Interviewees were given a consent form to read and sign (Appendix C).

vii. Interview participants completed both focused and in-depth interviews, conducted by the investigator. Interviews were conducted over a ten week period. In-depth (open-ended) interview questions were administered first, in order that the interviewees were not biased by the content of the focused questions. Interviews were recorded and later transcribed as permitted by the interviewee. In all cases, a "table shell" was used for the purpose of written data collection during the interviews.

viii. Data was analyzed. Machine-embedded and user-described images were compared to identify user image mismatch. Coordination failures identified in the in-depth interviews were coded within the framework provided by Klein et al. (2004). The investigator identified instances where user image mismatches were observed to play a role in those coordination failures.

4.5. Generalizability, Bias, Validity and Reliability

4.5.1. Generalizability

Though there will be no statistical generalizability resulting from this project, the investigator aims to achieve analytic generalizability (Yin, 2009) through the underpinning of results to a theoretical framework described above. The interviews involved subjects from a single, large tertiary care department of Anesthesiology being questioned regarding their experience with one specific type of Anesthesia machine, a design which might call into question the generalizability of results to other anesthesiology communities. This concern can be addressed on two levels: First, the similarity of anesthesia training and practice across North America as well as the dominant market share of the anesthesia equipment in question are both factors which support the wider relevance of these results. Second, even if we consider the study community as an isolated, unique culture, the demonstration of the existence and relevance of user-image mismatch (at all) would be a significant finding. Attempts to identify similar patterns in other anesthesia settings, or indeed, in other socio-technical environments in general could then be made.

4.5.2. Bias

No attempt was made to minimize selection bias of volunteer-interviewees because the aim of the study was not to determine the incidence of a user image mismatch but rather to understand the nature and impact of a mismatch when it does occur.

As an anesthesiologist studying a group of anesthesiologists, the investigator was an “insider” researcher. The benefit was that she had insights into technical and practical aspects that no lay researcher would be able to attain. She also had pre-existing experiences, opinions and biases that affected:

- The initial choice of topic and thesis question
- The pre-existing expectation of what the answer to the question would be
- The targeted machine-embedded images
- The questions generated for the interview tool, the way the interview questions were delivered and the nature of the follow-up questions
- The interpretation of answers, analysis of data and conclusions drawn

It is impossible for any researcher to be entirely objective, and, as stated by Blaxter, Hughes, and Tight (2009), “the play of emotions between researcher, researched, and research is often something to be welcomed and celebrated. Yet there is a need to be aware of your influence on your research and to be as open as you can in recording and recognizing these effects” (p. 83). The impact of bias was minimized by using pre-existing research to provide analytic distance. In addition, the careful use of “Level 1”[10] questioning avoided the interviewee

[10] Level 1 questions are posed to the interviewee. Level 2 questions are the questions that the case study itself seeks to answer. The answers to the Level 2 questions must be arrived at indirectly through the use of a set of distinct

being lead towards a perceived desired answer (Yin, 2009). Also, the careful ordering of in-depth then focused interviews avoided "priming" the interviewees to pinpoint user-image mismatches within their stories of coordination failures. Finally, the researcher made a conscious effort to be open to the instances where the a-priori hypotheses did not hold true, such as when:

- The machine-embedded and user-described images were well-aligned
- User image mismatch existed, but did not appear linked to any coordination failure
- Coordination failure was not related to user image mismatch
- Coordination was described as satisfactory or better.

4.5.3. Validity

This case study is both exploratory and explanatory (Table 2).

Table 2. Types of questions asked within this study

Exploratory questions	Explanatory questions
• **Does user image mismatch exist?** • **What are the types of coordination failures that anesthesiologists describe?**	• **What role does user image mismatch play in coordination failure between anesthesiologists and anesthesia alarms?**

Therefore, measures to ensure construct, internal and external validity were required (Yin, 2009). These efforts included:

- Data triangulation through the use of multiple sources of evidence (Figure 7)

(Level 1) questions. If the Level 1 and 2 questions are similar, then the interviewees will be less able to provide an unbiased response.

- Triangulation of viewpoints (user image versus coordination failure)
- Use of a theoretical framework
- Establishment of a chain of evidence
- Seeking of rival explanations
- Theme searching and coding of themes
- Review of report draft by key informants

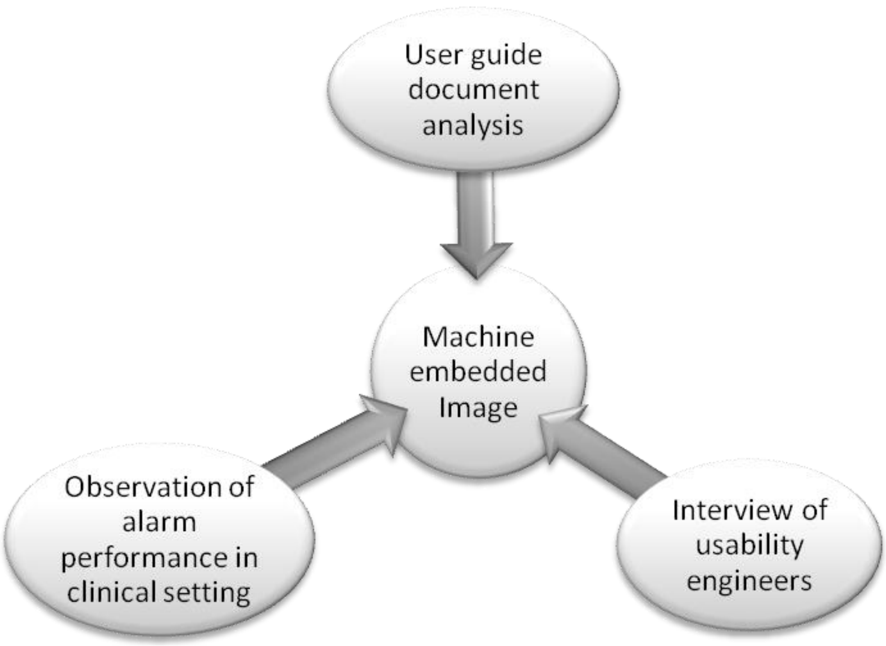


Figure 7. Data triangulation in the development of machine-embedded image

4.5.4. Reliability

Reliability is difficult to ensure in pilot studies like this. Here, efforts to achieve reliability have focused on a detailed description of the method, careful data collection and the provision of adequate argument and evidence paths for the reader to follow.

4.6. Research Ethics

The research proposal was reviewed with the Chair of the Research Ethics Board at Hamilton Health Sciences, the institution from which the interview participants were drawn. The chair was satisfied that no further review was required. No patients were involved in the investigation and the study did not take place during patient care.

Participation by anesthesiologists was voluntary. Informed consent was obtained. Interviewees' participation was held in strict confidence. Interviewees are referenced by non-identifying initials.

There is potential goal conflict in that the investigator will define the limitations and/or shortcomings of the anesthetic equipment used in her place of work. The support of the hospital Anesthesiology department chief was sought and received. Of equal importance, the investigator reached out to GE for collaboration, in doing so, emphasizing that the project was positive in nature, promoting a better understanding of how the alarms can aid the anesthesiologist toward their joint goal of patient safety.

5. Results

Chapter 5 of the Participant Guide for the GE Carescape B850 monitor was analyzed by the investigator. Three specific alarm design features were analyzed and a resultant machine-embedded image was drawn from each design feature. A series of interviews were undertaken by email with the principal alarm engineer for GE in Finland, Borje Rantala. Delays in establishing contact precluded the inclusion of this data in the construction of the machine-embedded image. The GE interviews are summarized in Appendix D. User-described images were drawn from interview analysis. As outlined in the Methods section, user images (machine-embedded and user-described) are, by necessity, hypothetical and represent first order generation concepts.

Fourteen anesthesiologists (eight male) were interviewed individually and in person, between December 05, 2011 and February 13, 2012. The mean interview duration was 48 minutes. All interviewees provided written consent to having the interview recorded digitally, and to being quoted anonymously in the thesis report and any future publications. The researcher took hand-written notes during the interviews and later transcribed the recordings. The interviewees included: Two individuals with in-depth knowledge of the GE anesthesia product due to having researched the product prior to purchase; one mechanical engineer; one individual with advanced software design skills; and one individual who teaches a yearly curriculum session on anesthetic machines to anesthesiology residents.

The results, analysis and discussion were reviewed and endorsed by three key informants, as a means of ensuring validity.

Alarm behaviour #1: Variable effects of the pause button

An alarm for a given physiologic parameter (e.g. HR) can be in one of three states:

i. Inactive: The parameter is currently within normal range.

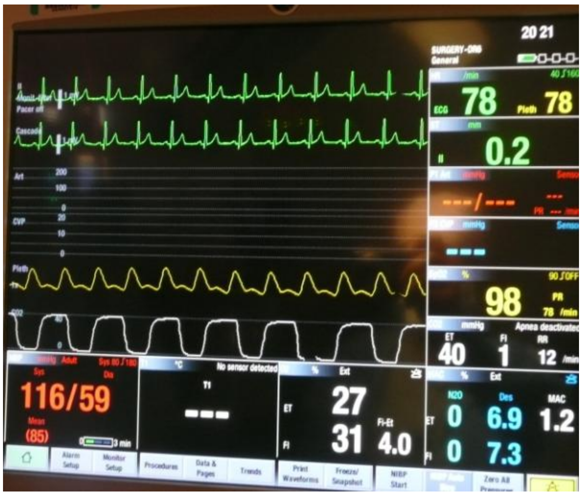
ii. Active: The parameter is currently outside the normal range.

iii. Latched: The parameter *was* outside the normal range, but has now returned to normal. The alarm is "latched" in the active mode until the user presses the pause button. Latching applies to a subset of monitored variables.

The pause button (Figures 8 & 9),often referred to as "the silence button" by interviewees, exhibits different behaviours depending on the current state of the alarm and how many times it is pressed. The single button, therefore, has six possible (distinct) effects depending on the context in which it is pressed and on whether it is pressed once or twice (Table 3).

Table 3. Possible effects of pressing alarm pause button

Alarm Status	Press pause button ONCE	Press pause TWICE
Alarm is inactive	Starts a 2 minute silence period for all new active or latched alarms that occur during the countdown time. A countdown clock displays (Figure 10).	Reverts all the alarms to active mode.
Alarm is active	Starts a 2 minute pause of all active alarms. ***No*** countdown clock displays. A small bell with an "X" through it is displayed (Figure 11).	Starts a 2 minute pause period for all active, latched and new alarms (except breakthrough alarms) that occur during the 2 minute period. A countdown clock displays (Figure 10).
Alarm is latched	Removes the latched alarm	The second press will have occurred during the system being in "alarms inactive mode" and will behave accordingly.



Pause button located on bottom right hand corner of touch-screen

Figure 8. Monitor screen of the Carescape B850

Figure 9. Pause button

Figure 10. Effect of pushing the pause button twice during an active alarm condition (as in the case above) or once when no alarm conditions are active

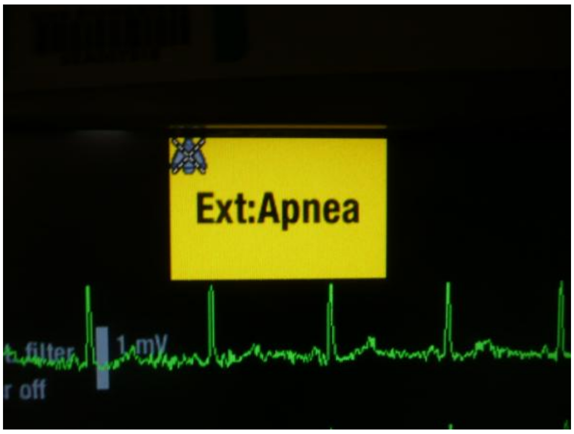


Figure 11. Effect of pushing pause button once during active alarm condition

Machine-embedded user image #1:

The user will be sensitive to the role of both context and repetition in determining the impact of a given physical manoeuvre (pressing the pause button).

Interview Analysis:

None of the interviewees was able to accurately describe the behavior of the pause button. Most found the behavior of the button perplexing and inconsistent. In expressing their difficulties with the pause button, interviewees identified:

- The brevity and unpredictability of the duration of the pause

- The inconsistency of effect when pushing the button once or twice

Some individuals reported pushing the button once while others reported pushing it twice or even thrice in attempting to gain a period of alarm silence. Interviewees did not understand the impact of hitting the pause button once versus twice in any of the three possible alarms contexts that might present. Some believed that the second hit was to provide a longer period of silence. Many identified that the second push "sometimes" seemed to undo the effect of the first push, contributing to the sense of inconsistency with button performance.[11]

None of the interviewees understood the difference between silencing the current (active) alarm with one push, versus silencing current and new alarms with the second push. The occasionally frantic button-pushing reflected efforts to gain the expected period of silence for the given abnormal parameter. Specifically, by trying to gain extra time to manage the current alarm problem, users were unknowingly disabling new alarm conditions for two minutes:

> *"I hit it once. Actually, I sometimes hit it five times. You get frustrated. I hit it until I get the timer. I watch it go down, and I think, okay, I've got two minutes, and then it comes back. There have been whole cases where I've watched it alarming at 5 seconds, and thinking okay, why isn't this working?" (H.G.)*

> *"I've done both [hit it once and twice]…I've tried everything. I don't know whether it makes a difference. Actually, I've pressed it twice but sometimes it resets if you press it twice. I don't know if that's true or if it's me pressing it too many times." (S.M.)*

> *"I usually hit it twice, I guess. But I think if you hit it twice you only get two minutes anyway, right? I hit it twice, just to make sure it's hit." (G. B.)*

> *"I honestly feel that I'm not getting the same response every time I hit it. I find that sometimes it silences it for a couple of minutes, sometimes I hit it twice and it seems to say, "OK, go back to the normal" and*

---

[11] Possible explanations for this observation, which was described by most interviewees, are suggested below.
i) By the time the user actually pushes the pause button, the alarm condition has already resolved. In this case, the second push actually re-activates the alarm, whereupon a frequently-recurring condition immediately alarms.
ii) Due to a lag in the touch-screen response, extra pushes are applied when the user assumes the first was not sensed. For example, if three pushes are applied when only two were intended, the third push returns the alarm to the active state.
iii) The pause period was successfully activated, but a "breakthrough" alarm condition occurs (and therefore alarms) during the two-minute pause period. Breakthrough conditions are explained further in Alarm behaviour #2.

it gives me the exact same alarm that I was trying to silence. I think hitting it twice is not the answer. Or whether it's three times." (W.T.)

User-described Image #1:

The user is unable to navigate the function of a single button that is sensitive to both context and repetition in determining the impact of a press.

Conclusion regarding User Image #1:

User image mismatch exists in the design of the pause button, whose effect varies based on both context and repetition. Users expressed frustration and confusion regarding the function of this button. Its precise behavior and impact was not understood by any of the interviewees despite their high level of technical sophistication. Many operated with frank misconceptions regarding its function.

A deeper level of mismatch was revealed through the analysis of the interview material. The binary evaluation (normal or abnormal) that the machine applies to each parameter fosters many of the misperceptions that arise from the use of the silence button.

For example, the design of the button is such that one press would give two minutes of silence for the currently active alarm condition but would allow other parameter abnormalities to alarm if they occurred during this two minute period.[12] If the original alarmed parameter briefly "normalizes" before becoming once again abnormal[13], the machine is programmed to present this "new alarm condition" to the anesthesiologist.

The anesthesiologist, by contrast, does not view the fluctuations around a threshold as distinct events, and may intuitively adapt by pushing the button twice to achieve the 2 minute silence for the original condition. This adaptation may have unintended consequences. In

---

[12] This appears to be a sensible design: An anesthesiologist might appreciate 2 minutes of silence while treating low blood pressure, but might still want to be alerted to a new arrhythmia that occurs in that 2 minute interval.
[13] Physiologic parameters are, by nature, fluctuating.

attempting to get a full two minutes' silence on the initial alarm condition, he unknowingly forfeits the option of being alerted to legitimately new conditions. As the interview data revealed, users are not aware of which alarms they are disabling (existing and new) with this manoeuvre.

In summary, the user image mismatch in item #1 relates to the opacity of a button design (variable effect depending on context and number of pushes) but also exposes a more fundamental user mismatch caused by the machine's binary determination of normality, whereby the machine sees "new conditions" and the anesthesiologist does not. This issue will be further explored in alarm behaviour #2.

Alarm behaviour #2: Alarm repetition

If a monitored variable strays outside the preset normal range, the alarm will be activated. The anesthesiologist can press the pause button, which will silence that condition for two minutes, whereupon the alarm will recur. There are several scenarios where one would experience a shortened period of silence *after pressing the button once.*

i) If a second alarm condition occurs during the two-minute pause, the alarm will sound.

ii) If a patient is in a physiologic crisis, with one or more variables remaining abnormal over an extended period of time, the alarms will continuously reactivate after each two-minute suspension expires in a staggered fashion.

iii) If the parameter normalizes, even briefly, during the two minute pause, the alarm (and its suspension) will reset. If the parameter again strays outside the normal limits, the alarm will reactivate. The effective alarm suspension time could be very brief, depending on how closely the variable is hovering near the threshold.

Even if the anesthesiologist has pressed the button twice, the occurrence of a "breakthrough condition" will cause the alarm to sound. Parameters designated as breakthrough conditions are programmed to sound an audio alarm even if a two minute silence (for existing and new conditions) is in effect. The designation of breakthrough conditions is not modifiable by the user.

There is no mechanism to gain a longer, though still finite period of silence. Furthermore, there is no way to guarantee a "hard" two minutes of silence with either a single or a double press for the reasons described above. Some (low or medium-priority) alarms can be terminated permanently or prevented in the first place by selecting the more general "Audio Off" function which is buried several screens deep, and would result in the non-selective silencing of all future alarm conditions within that category. Some alarms can be terminated by widening the pre-set threshold limits to levels that are so extreme that for practical purposes, the alarm is inactivated.

Machine-Embedded User Image #2:

Frequent repetition of alarm conditions is required to direct the user's attention to key abnormalities and can do so without negatively impacting on workflow or the cognitive functions required to treat those condition(s).

Interview Analysis:

Interviewees expressed the view that alarms are essential to highlight relevant clinical information that may go unnoticed. Furthermore, interviewees overwhelmingly endorsed the value of a physiologic alarm recurring if a parameter had not normalized after a given period of time.

All interviewees maintained that the behavior of alarms (even those that highlighted real problems) should be designed to accommodate the anesthesiologist's workflow, both cognitive and technical. Almost all subjects expressed the view that they should be able to readily direct the behavior of the audio alarms once triggered, such as being able to select either a short suspension (two minutes) or a longer suspension (five minutes) according to the circumstances. Many also expressed the opinion that the anesthesiologist should be able to permanently acknowledge certain types of alarms (technical alarms, as well as those that were frankly spurious), such that they don't recur:

> *"As the anesthesia provider, I feel myself that I'm the primary monitor and I would like the ability to decide what's necessary and what's not in that context. And if I don't want to be reminded about something again, I would like to have that ability." (S.S.)*

Nearly all of the subjects reported the behavior of audio alarms to be intrusive and distracting to their own ability to manage the patients as well as to surgeons and nurses[14]. This applied to true as well as false alarms, although interviewees acknowledged that the behavior of alarms would be more tolerable but for their frequency. Regarding the impact of frequent audio alarms whose immediate recurrence they were unable to readily control, subjects describe the following:

- Time spent "managing" the machine rather than directing attention more appropriately to the patient.
- Alarm fatigue manifested in all team members
- Anxiety created in nurses and surgeons
- Erosion of confidence from other team members

"You want to be focused on your patient and ... have open communication with everyone, there's a safety check going on, there might be questions about the patient….and you are no longer participating in that, you are focused on shutting alarms off. It distracts me from things in the environment that I would normally be doing and listening for…sometimes you pick up on a little detail. I don't feel like I'm engaged. I'm engaged with the machine, I'm not engaged with the environment" (F.I.)

"It's irritating because it's a critical time for the patient and it's distracting and you hear it so much that you almost become immune to it so you don't look at the machine, you just think it's nonsense or noise at a time when you maybe should be paying attention to the alarm. I don't get excited when I hear an alarm off this machine….It doesn't mean anything, It's not significant. It's just something that has to be turned off." (H.G.)

[14] The concept of "user" extends to all those who are exposed to and respond to the behaviour of the alarm.

"I find it also affects the surgeon… They become nervous, like "what's going on up there", and I feel maybe they're even losing confidence in me, like, 'Why are you alarming?' So then you have to have that added reassurance to them…and look after the patient." (N.L.)

"Even when I appreciate the alarm, that it's alerted me to something important, I find that it does it in such as that it divides my attention more than it should…I have to treat the patient and the alarm. I want to focus on the patient, fix the problem, the real problem, thanks for telling me. It's too loud and too persistent. It just fills your environment...Real alarms matter, but I want to treat the patient, not the alarm." (S.S.)

"An alarm points out an abnormality, having your attention drawn to an abnormality should make things safer, therefore alarms should make things safer. But if it's alarming that much, you just start tuning it out. I have tuned it out and missed things…A few times, I've looked up and actually seen what it's been trying to tell me. I think if it was a single tone, my attention span would be better for that. Rather than multiple, musical tones." (P.O.)

"I've actually told a nurse, in the middle of doing something, like the art line, you've lost the art line, you're trying to fix it and the machine's alarming, 'Just hit the button and every time it comes up just hit the button'. It's not really what you should be using a nurse for but it's just irritating for the whole room all around." (H.G.)

One interviewee had a different focus, viewing the perceived need to pause an alarm as a distraction from a more fundamental problem:

"There are times when I wish I could disable it for longer, but I see those times as a failure of the alarm system in general. Like why would I have to [disable the alarms]? Well, the only reason I have to is because some other part of the alarm system isn't doing what it's supposed to be doing…. If you want to

disable an alarm for an extended period of time, and I count five minutes as an extended period of time, it must mean that the alarm system is unsatisfactory in some way: It's intrusive; it's alarming for the wrong reason, the wrong parameter. If you want to disable it, it means you're not valuing the information it is giving you….I wish the alarm system was more tuned, to what we do. And I wish I would never want to disable it for five minutes." (B.L.)

As discussed earlier, an alarm based on simple thresholds is bound to a binary determination of normality that leads to a disconnect between how man and machine view patient physiology and accounts for many of the instances where users were describing shortened durations of silence. As an example, where the BP threshold is set at 80 mmHg:

i.	12:00.00	BP 81 mmHg	no alarm (BP is "normal")
ii.	12:00.01	BP 79 mmHg	alarm activates (BP is "abnormal")
iii.	12:00.03		anesthesiologist pauses 2 minutes
iv.	12:00.04	BP 81 mmHg	alarm (and pause) resets (BP is "normal")
v.	12:00.06	BP 79 mmHg	alarm activates (BP is "abnormal")

vi. i-v repeats, multiple times per minute, as BP hovers around the "threshold". This example is more realistic than it might appear: The fluctuation of blood pressure with the respiratory cycle is a normal physiologic effect. A similar cycle could be created by an intermittently-noisy ECG signal.

The alarm behaves as though multiple separate adverse events occurred and resolved within that brief time frame, an interpretation that is not in keeping with that of the anesthesiologist. Not surprisingly, interviewees were unanimously dissatisfied with the resetting of the alarm pause with the "normalization" of the parameter. Some interviewees understood the cause of the shortened pause period, and all experienced its impact on a regular basis:

"Most of time I'm pressing it because it's a false alarm and I just want to stop getting bothered by the alarm and it keeps coming back. I feel like it's that "LOST" episode where you have to keep entering that number so the island doesn't blow up. Sort of a make-work program. My perception, it may not be reality, is that sometimes the alarm comes back in less than two minutes." (S.S.)

"It's not helpful to have to repeatedly silence the monitor for a condition that tends to recur once acknowledged. Like low BP that you've already dealt with and it's on its way up. In situations like where there is a respiratory variation around the threshold. I end up repeatedly silencing, I don't go in and change parameters because usually if that's happening, I'm too busy doing other things." (D.L.)

"The alarm goes off… And that's fine, you can silence it but the silencing doesn't last very long and it comes back with a vengeance. And so you're being distracted by this annoying false alarm and you're almost thinking that the machine thinks it knows better than you. 'Here, pay attention to me, do this, fix this' and you've looked at it, you've processed it, you've decided what you need to do, if anything, and the machine keeps bugging you. And so you silence again, and it comes back." (S.S.)

Some developed rationales to explain why they did not obtain the two-minute silence that they believed the button would deliver:

"I'm not sure I've ever succeeded in getting it to be quiet for 2 minutes. It may be that my sense of time is distorted……although it seems time tends to drag more when things aren't going well. Perhaps it's just the frequency with which the machine alarms. It's like, 'Geez, didn't I just silence that?'" (D.L.)

"Sometimes I think the touch-screen has malfunctioned, like in not recognizing the touch." (E.B.)

One interviewee expressed the opinion that the inconsistent duration of alarm silence in itself is detrimental:

Usually by two minutes [the problem can be corrected]. If at two minutes, if it was alarming again, because of the problem, that wasn't a bad thing because it put a timer on me. You know, like, I should have this fixed in two minutes. Whereas now, I don't even have that sense of how long it's taking me to fix things….because in situations like that, your sense of time is skewed." (P.O.)

Finally, several individuals commented that the machine "reacts" too quickly to abnormalities:

"This machine seems to alarm very quickly for everything. No one is going to die in 5 seconds. It seems to not give you enough time, a period of time, where you can do your work without setting off fires and bells and whistles." (H.G.)

"The machine is just sensing a change….it seems to me, though, that it's sensing it very quickly. This machine seems to jump right on it. Like no breathing for 5 seconds and it seems to call it apnea." (W.T.)

Self-Described User Image #2:

The user finds the frequent repetition of alarms distracting to clinical management of the patient. Specifically, the user finds that the alarms demand "management" where focus should be directed toward the patient; promotes alarm fatigue; and creates anxiety in and erosion of confidence from other team members. Dissatisfaction with the alarm repetition and lack of directability applies to valid alarm conditions but is amplified by the frequency of false positives.

Conclusions Regarding User Image #2:

User image mismatch exists as a result of the frequency, persistence and lack of directability of alarm repetition. Though the cluster of effects (summarized above) was consistently described, analysis revealed a variety of underlying causes at play, depending on the specific situation. In some cases, the mismatch results from and exposes a discrepant approach to the analysis of physiologic parameters: Binary (machine) versus continuous (human). Other times, the mismatch reflects a discrepancy in the time-tolerance for out-of range variables between man and machine. Most commonly, the mismatch appears to reflect an erroneous view of the anesthesiologist's workflow in response to real physiologic abnormalities.

Alarm behaviour #3: Display and prioritization of alarms

Data is objective. Information, by contrast, is "interpreted data" and emerges from the understanding of the interpreter.   While the screen displays continuous data of various parameters in fixed locations (Figure 8), interpretations of abnormalities appear at the top of the screen in the message area. Up to five alarm messages may be visually displayed in the message area, ordered horizontally from left to right. Abnormal variables have an intrinsic, preset prioritization (low, medium or high). The highest priority and newest alarm conditions are displayed at the far left. A subset of abnormal conditions will escalate in priority over time if the abnormal alarm condition has not been resolved and the alarm pause button has not been pressed. In general, the alarms conditions are "ordered" in priority from left to right. Each of the (up to five) alarm conditions may switch positions in the message area as relative priorities change.   In addition to position, distinct colours and auditory melodies are assigned to each specific priority of alarm condition (Figure 12).

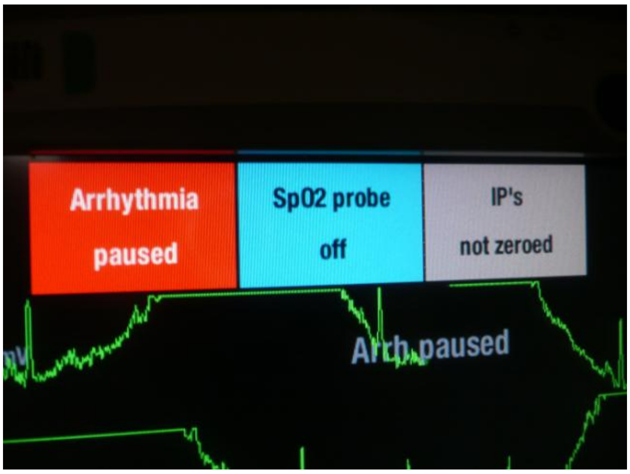


Figure 12. The monitor screen message area

Machine-embedded user image #3:

The user will exploit decision-support provided by the itemization and prioritization of abnormal conditions in the message area of the monitor.

Interview Analysis:

Interviewees uniformly described using the sound of an alarm as an auditory trigger to scan the monitors for physiologic parameters (both numeric and waveform). The message area at the top of the screen was often described as being viewed only if a scan of the monitored variables did not reveal the nature of the problem. The importance of adjudicating a presumed abnormality based on examination of the patient, assessment of surgical events, and waveform interpretation of monitor data was emphasized by all subjects:

> *"I must say, I don't look at those [the message area]. I'm usually looking at the parameters because you can see those in a jiffy, and see exactly what is happening. If I look at the monitor and I don't see anything obvious, I'll look at the alarm and see what it is telling me."* (G.B.)

"It's not that I bypass the information display, but it doesn't have any…I don't put much validity in it. It's more of a flag to look at the monitors. I must say I don't get alarmed by the blinking lights and bells. I look directly at the monitor waveforms. I have to make the decision myself". (E.B.)

"I don't often look at the alarms first, I'm looking at the patient. I…put my hand on the patient, check the airway, make sure we're connected, make sure I don't hear or see any leaks that are obvious, and I'll scan the monitors and just have a look and see what's going on... think about what's going on with the patient, what's going on with the surgeon, because sometimes the alarm's going off because of something that's just happened in the field, and I haven't been communicated that information, the first time I'm hearing that there's suddenly four litres of blood on the floor, the surgeon is silent....So if everything's okay in the environment, then I focus back in on the machine. It gives me some perspective, it allows my brain to buy time, so if I examine the patient and the monitor, see that their sats [oxygen saturations] are good and they have a good pulse, I know I have time to go deal with the machine." (F.I.)

None of the interviewees was aware that the monitor actively prioritized alarms, or that they were prioritized from left to right on the screen. Some interviewees were aware that alarm messages were changing positions but did not attach significance to it:

"I wasn't aware that it did that. Well, it shouldn't be in the business of doing that. That would just be more information coming at me that is not helpful, more things for me to sort out. I need to look at the actual data." (F.I.)

"Does it really? I wouldn't have even noticed. It's not something I would look at. It's probably because if all three were going off in a case, I'm not even looking at the monitor. I've never noticed that there's any order to it, that one's on the left… What alerts me more than anything is the auditory cue, the noise. The fact that I've never noticed… probably means it's not, doesn't really stick in my mind that much." (W.T.)

"I didn't know the machine prioritized alarms along the top. I'm sure I've been in situations where I've had multiple alarms and that wouldn't be my first approach, to look to the machine to let me know what was going on, I would look to the actual readings… and go from there. I wouldn't be looking at the alarms, which one was yellow, which was orange, the specific levels of alert. I need more specific information for that context to make a clinical decision which includes of course examining the patient and talking to the surgical team to look what's going on from the surgery point of view. Far beyond the monitors. So, thinking about it, I'm not going to treat the monitors or the alarms, I'm just focused on the patient and the whole context of everything. I think I learn to tune it out, visually. I no longer notice it because it's not giving me any useful information… I'm not going to waste my time analyzing the colour and the sequence of alarms. I'm going to look at all the monitors at once, what's wrong, what most disturbing. I need to look myself. I'm not going to trust the sequence of the alarms that it's displaying to me to know what's most serious, I'm going to look at the individual numbers myself." (S.S.)

Discussion of how interviewees prioritize multiple physiologic abnormalities revealed themes that clarify why the machine-generated prioritizations were not found to be relevant by the anesthesiologists:

i) Relationships matter: Interviewees were uniform in describing that they do not process multiple abnormalities as isolated, single entities. They described searching for a single explanation to explain clustered abnormalities. Interviewees reiterated the vital role of using the parameter information along with clinical examination of the patient in holistic fashion to get an overall picture of the patient's status.

ii) Context matters: The degree of priority that an abnormality reflects cannot be excised from the clinical setting in which it occurs. For example, a blood pressure of 140/80 mmHg, comfortably within the pre-set limits of "normality", may be too high for a patient being anesthetized for surgery to repair an aortic dissection.

iii) Magnitude matters: A high-priority parameter that is minimally outside normal range may be less significant than a low-priority parameter that is markedly out of range.

iv) Pragmatism matters: Anesthesiologists described treating what was easiest to fix first, to allow time to focus on the more complex issues.

v) Raw data matters: Alarm thresholds are based on numeric values which are derived from an analog (waveform) signal. The anesthesiologist adjudicates the validity of the numeric value by evaluating the accompanying waveform.

"They're usually all interrelated. They are usually not 2 or 3 separate things, they are usually all one root cause thing that you have to get a hold of." (H.G.)

"Basically, I scan all the monitors and try to put it into context of what's the primary problem. Because often there's a primary problem and the other ones are secondary. So in the example where you've got the hypotension alarm, the hypoxia alarm, and an arrhythmia alarm going off, it's often due to hypotension, the person's not perfusing, they're not reading their oximeter…it's unlikely that all the wheels fall off the patient at once. It's normally a central problem that has secondary consequences. I would try to find out what exactly is going on with my patient and tackle that first." (S.S.)

"First of all, I start off by ignoring all the alarms. Then I resuscitate based on my own priorities. And what I can fix quickly. It's always airway, breathing, circulation, probably in that order, but I might treat circulation first if I know treating the airway is going to take up both hands…But I do want the alarm off once I've acknowledged it so I can concentrate. Especially, the more problems the patient is having, the more I need quiet. The more alarms are going off, I can't, I'm distracted by the alarms instead of trying to integrate the information which is often complicated." (P.O.)

Some users noted incongruity between the machine's allocation of priority and their own clinical experience. For example, ventricular fibrillation (a lethal cardiac arrhythmia) is a high-priority alarm condition that would be reported many times during each case (erroneously, due to electro-cautery or motion artefact) while true ventricular fibrillation would be an exceedingly rare event and would not occur without abnormalities in almost all other parameters. Paradoxically, therefore, some of the "high priority" alarms that the monitor reports are often those that are least interesting to the anesthesiologist:

"There's like boxes that will come up for everything that is wrong. But a lot of times, you look through them, and you're eliminating them because they are of no consequence. They'll be a whole bunch of boxes at the top, and the little blue light going off." (P.O.)

Apart from not being found relevant, the alarm presentation was described by some as overwhelming and confusing:

"It [the problem] doesn't come right at me. There's multiple screens. There's clutter of information. Things are flashing and you're wondering, is that why it's alarming? The little red light is flashing at the top, and it's beeping and someone's asking you why the alarms are going off, if there's a problem. It's tough to figure out what it is. It's easy to find the button and turn it off, it's just the distraction of 'Is it a real alarm, is it important, do I need to do something right now?' Then once I've turned it off, you have to figure out what you've turned off. Right, because you don't exactly know what's going on so you're trying to figure out what's beeping, turn the distraction off so that you can refocus because the beep is information, right?...Again, you're trying to reassure yourself that the patient is fine, so you're checking your sats [oxygen saturations], ECG, BP, waveforms, make sure you've got end-tidal CO2, make sure the patient's ALIVE, and so finding all that relevant information quickly and reassuring yourself so you take the emotion out, because the minute you've got emotion into your decision, and your thought processes, it's much more difficult to find those solutions." (F.I.)

"Not giving me the two minutes, makes me panicked. Even though I'm not panicked. Even though if you asked me if I was worried about this blood pressure, I would say no, the fact that it's continually alarming makes me panicked, even thought I know rationally it's a normal situation." (P.O.)

"There was a time I went to put the patient to sleep and I couldn't bag the patient [achieve ventilation through a bag-mask apparatus]. There was a [circuit] disconnect. Because the alarms are so confusing and they come on so often, the alarm was not helpful to me in trying to diagnose the problem… I find it's almost overwhelming, the amount of alarms that go off." (S.M.)

"There's multiple things that can appear at the top of the screen in terms of the alarms that are going off. I don't necessarily look at the alarm, I look at the patient and my vital signs or whatever, the ventilation parameters and all that kind of stuff before I would address the alarms. So if it's something important, I would address that first by looking at the patient versus the alarm, the different alarms. Sometimes there are so many alarms, it doesn't help guide you to what's actually wrong, what's going on with the patient." (S.M.)

Several anesthesiologists felt that more subtle auditory cues would have a bigger impact:

"I think most anesthetists are attuned to noise in general, but probably the more subtle noises have a bigger impact on us… I'm listening to the ECG for rhythm abnormalities and rate. We all do this. Even when we're not listening, we're listening. We hear an increase in heart rate by, like 5 and we're sensitive to that. I don't need this going off every second because of artifact, and you can't even turn that one off." (P.O.)

"I have to say that high PEEP [positive end-expiratory pressure alarm, indicating a kinked breathing circuit] and the time when it alarmed about the [anesthetic gas] canister being empty...those were the only

two times that machine has ever given me useful information and both of them were visual not audible. But I was surprised that the two [pieces of] useful information… that I acted on were tiny little things that I just happened to see and thought, 'Ohhh'….and there was no big alarms and bells and whistles" (H.G.)

"One of my main concerns about the alarm setup is that it alarms so frequently, and there does not seem to be a distinction in the severity of the alarm tone related to the condition that it's distracting and fatiguing and I think it has a cry-wolf effect on the whole room. Machines that are quieter, when they start alarming, people pay attention, nurses will ask you, 'Is everything okay, do you need some help?'" (D.L.)

User-described Image #3:

Users view abnormalities holistically and seek to find a single explanation for a cluster of abnormalities. Physiologic measurements are interpreted through waveforms and within the context of the clinical situation (e.g. surgical events) and the physical examination of the patient. Parameters are not seen as normal or abnormal, rather as existing along a continuum that is grounded to a scale dictated by both patient and surgical factors.

The user does not exploit the decision-support provided by the itemization, interpretation and prioritization of alarm conditions in the message area. The monitor's method of alarm prioritization is not found to be relevant to the interviewees. Furthermore, the users describe the alarm display as confusing and sometimes overwhelming and endorse a "less is more" philosophy, particularly for auditory alarms.

Conclusions Regarding User Image #3:

The display and prioritization of alarm conditions creates user image mismatch on several different planes.

i) The alarm's algorithm for prioritizing alarm conditions does not reflect some of the key factors that influence how the anesthesiologist ascribes importance to an out-of-range variable. Factors that are not considered include the magnitude of the abnormality, it's co-existence with other abnormalities, individual patient factors and the immediate surgical context in which the abnormality is occurring.

ii) The interpretation of data presented in the message area was not trusted or utilized by anesthesiologists, who seek data directly through waveforms and numeric figures.

iii) The concept of individually listing abnormalities in an order does not reflect how the anesthesiologist cognitively assesses abnormalities (as symptoms of a single problem).

iv) What appears to be an attempt at decision-support was seen as contributing to contribute to visual and auditory clutter and was not utilized by the anesthesiologist.

The analysis of interview data suggests that the design of alarms (including those that incorporate features of decision-support) should be grounded by an understanding of how anesthesiologists make decisions.

6. Analysis

The fundamental finding of this work is that mismatch exists between the image of the user that is embedded within the design of the Carescape B850 anesthesia monitor alarm and the corresponding self-image described by the user. The mismatches were interpreted as practical problems[15] that were rooted in more fundamental discrepancies, including the following:

i. The machine adjudicates abnormality in a binary fashion while the anesthesiologist views parameters along a continuum.
ii. The machine interprets abnormalities in isolation while the anesthesiologist seeks to understand clustered abnormalities in terms of a single, underlying cause.
iii. The anesthesiologist interprets numeric data in the context of analog waveforms, clinical setting and individual patient factors.
iv. The anesthesiologist considers a single abnormal parameter based on the magnitude of the abnormality, its co-existence with other abnormalities and the immediate clinical context.

One aim of the study was to examine whether user image mismatch was observable as coordination failure (Klein et al., 2004). The requirements for successful coordination in a cognitive system are interpredictability, directability and common ground. Deficiencies in each of those requirements were identified through the evaluation of the user image mismatches.

Interpredictability

In the discussion of alarm behaviour #1 ("Variable effect of the pause button"), and #2 ("Alarm repetition"), unpredictability of the machine was commonly identified by interview subjects. The lack of predictability of the duration of silence achieved by pushing the

---

[15] User image mismatches were described on a pragmatic level: "The button function is opaque"; "The alarms are repeated too frequently"; "The alarm messages are not used or trusted by the user."

pause button had many adverse effects, including the anesthesiologists' inadvertent (and unrecognized) silencing of new, incoming alarms.

Lack of predictability was a *result* (not a cause) of the user image mismatches described. Firstly, the design of the pause button resulted in six possible effects of the button, dependent on context and number of pushes. This opacity of design resulted in what was, from the perspective of the human operator, unpredictable behaviour. Secondly, the machine's binary determination of abnormality caused the alarm silence period to reset frequently as a parameter would hover near the threshold. Because the anesthesiologist did not perceive separate new abnormal events in the way that the machine was, the behaviour of the machine was perceived as random and unpredictable.

"Interpredictability" requires that each actor must be predictable to the other. We must consider whether a lack of predictability of the anesthesiologist to the machine plays a role in user image mismatch. For example, alarm repetition might be unnecessary if the machine was able to "understand" that the anesthesiologist was attending to the (still abnormal) alarm condition. Though interviewees emphasized the frustrations of working to correct a problem while the alarm loudly repeated its alerts, many also described the value of alarms that alerted them to conditions occurring while they were distracted with other tasks, such as checking laboratory results on the computer or answering a telephone call from the post-anesthetic care unit (PACU). Currently, the machine has no method of forming an accurate expectation of how the human will behave in response to an alarm and thus repeats alarms with frequency until the parameter has normalized.[16] Coordination failure, due to the perceived unpredictability of the human operator, along with the absence of a method of communicating stance[17] can be seen to

[16] Recall that according to Seagull and Sanderson (2001), the most common response to an alarm, across all phases of anesthesia was "ignoring".

[17] Some might argue that "lack of awareness of stance" is really a subtype of "unpredictability". This reflects the artificiality of separate headings and the fluidity of the factors at play. There are instances where the human operator moves through a sequence of steps or tasks in a predictable manner. During some steps he is interruptible and open to information; during others, he is not. His taskflow is more or less predictable. What is lacking is any method of

play a causative role in the development of the user image mismatch described in the discussion of Alarm behaviour #2 ("Alarm repetition"). Human unpredictability causes the mismatch; that mismatch then causes (perceived) machine unpredictability.

Directability

Lack of directability was a major theme in the discussion of alarm behaviour #2 ("Alarm repetition"). Interviewees attributed several negative effects to their inability to meaningfully control the behaviour of alarms, including inappropriate diversion of attention and workload, alarm fatigue, and anxiety in and erosion of confidence from other team members. Interviewees described feeling that the alarm seemed to "think it knew better than" them. The majority of interviewees felt that there needed to be a simple mechanism to more definitively control the alarm: To achieve a consistent duration of silence; to select a longer period of silence; to silence a given alarm permanently. Many complained of high coordination costs, having to navigate several screens to permanently disable a given parameter:

> *" I understand the importance of alarms, it's just like going to find where to adjust them and deal with them, and I'm more than willing to do that, but I just find the interface is more complicated that it needs to be, especially in our critical situations that we deal with, it should be a one, no more than 2 step process to get where you want…. And it should just not take so long to get to where you need to go, and as many steps." (N.L.)*

Though interviewees focused on their need to better direct the behaviour of the machine, the machine also seeks to direct the human, particularly through changing conditions, or when the anesthesiologist's attention is diverted. Unfortunately, the less the machine's direction is felt to be relevant, the more the human deflects it, either actively (silencing alarms) or

---

communicating (to the machine) which step he is on. In this sense, the obstacle is communication (to achieve common ground) rather than some inherent unpredictability of human action.

passively (alarm fatigue). Improved ability of the machine to take into account the human stance would allow it to be more selective in its efforts to attract the human attention which would ultimately allow it to do so more effectively. For example, if the machine knew that the human was actively working on the problem (and was not otherwise preoccupied elsewhere), it could suspend or attenuate auditory alerts to allow managerial cognitive functions to proceed without distraction.

In summary, lack of directability was inextricably linked to the user image mismatch described through "Alarm behaviour #2". As with interpredictability, directability is interpreted bidirectionally within the man-machine relationship.

Common Ground

Lack of common ground has been identified as the most frequent cause of breakdown in joint cognitive activity (Klein et al., 2004) and plays a contributing role in each of the three user image mismatches identified. The user image mismatch defined in alarm behaviour #1 ("Variable effect of the pause button") reveals a simplistic loss of common ground whereby man and machine lack a mutual understanding of the action of a commonly-used button.

The analysis of alarm behaviour #2 ("Alarm repetition") revealed a fundamental loss of common ground, with man and machine operating on discrepant definitions of what constitutes a "new condition". On a practical level, interviewees expressed the view that the machine did not take into account the anesthesiologist's workflow when dealing with an abnormality; that the monitor seemed to assume that the anesthesiologist was not attending the patient; that the monitor expected an unrealistically rapid correction of abnormalities.

Finally, the user image mismatch related to alarm behaviour #3 ("Display and prioritization of alarms"), was also rooted in a lack of common ground, with the machine's display and prioritization of alarm conditions overlooking the fundamental principles that anesthesiologists use to assess the importance of outlying parameters.

Klein et al. (2004) identified what they felt to be the most "salient" causes of loss of common ground in joint cognitive activity, causes which are readily observable in the interview data from the current study:

i) Discrepant access to data:

The anesthesiologist responds to numeric and analog data interpreted through magnitude, trends, patient factors and surgical events. The machine (alarm) responds to numeric data viewed within simple thresholds. Indeed, its limited contextual field was often cited as the basis for the lack of relevance the anesthesiologist attached to the alarm's messages.

ii) Lack of clarity regarding the joint goal:

The machine's goal is to sound an alarm for any out of range parameter and repeating the alarm at frequent intervals until the parameter "normalizes". The anesthesiologist's primary goal is to maintain physiologic homeostasis. That primary goal is supported by secondary goals which include: Communicating with other team members; being aware of important surgical events; maintaining a calm atmosphere; using executive functions to problem-solve complex situations; and performing the "technical" aspects of care, such as correcting equipment problems and administering drugs, fluids or blood products.

Though each team member may have their own local goal(s), those goals must be relaxed when they hinder the joint goal. Many interviewees described instances where the machine's 'local goal' impeded what should have been the overall goal of the system. One such example was that of the anesthesiologist who described assigning a nurse to the task of repeatedly pressing the pause button while he/she was correcting a technical problem with an arterial line.

iii) Lack of awareness of differing stances between workers (workload, competing priorities):

If an abnormal condition is sustained, the machine has no way of knowing if the anesthesiologist is working on the problem (that has yet to be corrected), is working on another, higher priority situation that requires his attention, or is distracted with a lower priority task. Many interviewees described frustration with alarms occurring during emergence, a period of time where the anesthesiologist is closely attending the patient and where the parameter expectations predictably change:

> *"At emergence, you don't really need an alarm. Your attention is so close to the patient, anyway. You're watching their ventilation, their saturation, their tidal volume, you're trying to determine when to take out their endotracheal tube, like, you're looking at your monitor non-stop….I don't need an alarm to tell me what's abnormal. I feel like at emergence, it's like it doesn't know that I know what I'm doing." (P.O.)*

iv) Confusion over who knows what:

The variable effect of the pause button caused resulted in the anesthesiologists' inadvertently disabling of new, incoming alarms. The anesthesiologist, therefore, would be unaware that he would be missing alarms about any new, abnormal parameter. The machine, on the other hand, interprets an intentional disabling of the alarms on those new abnormal conditions.

Oversimplification

We have seen, therefore, how a breakdown in each of the three requirements for coordination in joint cognitive systems can manifest in the interaction between anesthesiologist and anesthesia monitor alarm. In addition to these previously published factors (Klein et al., 2004), a further factor is proposed as a contributing to coordination failure in joint cognitive systems, that of "oversimplification". The analysis of alarm behaviour revealed instances where the machine oversimplified complex processes.   One example was the algorithm for display and prioritization of alarm conditions. The machine did not (or was unable to) consider key factors that would affect the criticality of a given abnormal parameter. By operating instead on a very simple, fixed algorithm, the end result was, paradoxically, a display that was found to be confusing, distracting and ultimately irrelevant.

Another example of oversimplification was the method of communication between man and machine, which was extremely limited and minimized the importance of a two-way dialogue. For example, the only messages that the anesthesiologist could communicate to the machine were through the use of the pause button:

i) "I would like 2 minutes of silence for this currently-alarming condition."

OR

ii) "I would like 2 minutes of silence for the existing and any new incoming alarming conditions."

As we have previously discussed, the inability to generate a richer dialogue played a role in the coordination failure between man and machine. In a less simplified design, the anesthesiologist could choose from a range of responses to communicate to the machine in the setting of an alarming parameter that would then dictate future alarm behaviour for that parameter. The oversimplification of dialogue between man and machine increases an already paradoxically complex (and again, distracting and confusing) interaction. The need for a more sophisticated, two-way exchange between man and machine was described by one of the interview subjects:

> *"If the machine tells you there's a problem, and you attend to the problem, maybe you should tell the machine that you've fixed the problem. And then it would stop. Otherwise the machine just keeps monitoring, "Is the problem still there? Is the problem still there? Is the problem still there?"…The interaction we have with the machine…The machine is monitoring certain parameters and defining whether there is a problem based on those parameters… what I'm talking about is a higher form of interaction. Say the oxygen saturation is 88%. The machine tells you that it's a low sat....So the anesthesiologist will go about doing the things they usually do…and perhaps ultimately just increase the FIO2 [fraction of inspired oxygen]. But if you just increase the FIO2 and the saturation goes back above the threshold, is that okay? As far as the machine's concerned, it's okay but is your patient okay?*

Is there something still going wrong with your patient? Quite possibly. It's a simplistic relationship, BOTH ways, it's simplistic both ways. It could function more as a partner in diagnosis and the amount of computer power in there is totally capable. It's absolutely under-utilized." *(B.L.)*

A third example of oversimplification is the fact that the monitor applies the same parameters and expectations to all phases of anesthesia. One result is frequent alarms that occur during the emergence phase (where many parameters are expected to change). Anesthesiologists most frequently ignore alarms during this critical phase, likely because they are already closely engaged with the patient at this time and do not require the machine to attract their attention, and/or because the "abnormalities" causing the alarms are expected during this phase of anesthesia. The "oversimplification" of alarm algorithm (applying one set of parameters across all phases of anesthesia) paradoxically creates a more cluttered response pattern. The current research suggests that, just as a lack of interpredictability, directability or common ground can contribute to coordination failure, so can the tendency of one element of the joint cognitive system to oversimplify (or ignore) the inherent complexity of their joint activity.

In the earlier discussion, we have seen how user image mismatch can contribute to coordination failure within the joint cognitive system comprising anesthesiologist and machine. The analysis also suggests that user image mismatch obstructs coordination within the wider healthcare team. For example, interviewees described intrusive alarm behaviour distracting from team discussions or subtle clinical signs that might otherwise be picked up. This could create a loss of common ground, with the anesthesiologist missing out on information that others would assume had been shared. Some interviewees described how the repetitive alarms eroded confidence from other team members, creating the sense that the anesthesiologist had lost control of the situation, becoming unreliable, even unpredictable. These points are captured in the comments below:

"It makes the room noise louder. Everyone speaks louder when those alarms are going off. Because we're talking over them. So it's hard to quiet the room. And they [the surgeons] will become, like, hostile, because I can't get my machine to stop alarming. And I can't blame them. Because sometimes they're doing very intricate work." (P.O.)

In summary, coordination failure was interpreted to be an integral aspect of each of the three user image mismatches studied. Moreover, breakdowns in each of the three requirements for coordination in joint cognitive systems (interpredictability, directability and common ground) were contributing factors. An additional factor, "oversimplification" has been proposed. Coordination failure was interpreted as having a causative role in user image mismatch in some cases. At other times, coordination failure appeared to be a consequence of user image mismatch. In this latter instance, the resulting coordination failure was observed within the joint cognitive system of anesthesiologist and alarm, but also between the members of the larger operating room team.

7. Discussion

In "Behind Human Error" (Woods, Dekker, Cook, Johannesen & Garter, 2010), "clumsy technology" in the operating room is described through the examples of a centralized data display for cardiac anesthesia and a drug infusion pump. The authors describe the features of "clumsy technology" as follows:

- *Instead of freeing up resources, clumsy technology creates new kinds of cognitive work.*
- *Instead of offloading tasks, autonomous but silent machine agents create the need for team play, coordination and intent communication with people, demands which are difficult for automated systems to meet.*
- *Instead of focusing user attention, clumsy technology diverts attention away from the job to the interface.*
- *Instead of aiding users, generic flexibilities create new demands and error type.*
- *Instead of reducing knowledge requirements, clumsy technology demands new knowledge and more difficult judgments.*

Each of the features above has been described by the informants in the current work. Are anesthesia alarms merely another example of clumsy technology? The crux lies in how the role of the alarm is viewed: Is the anesthesia alarm a mechanical device, created to relieve the anesthesiologist from the role of monitoring data parameters? Or, is the alarm an independent actor fulfilling its own role in the joint cognitive system?

As modeled mathematically (Sorkin & Woods, 1985) and described earlier in Section 2.2.2., the machine is an independent actor, albeit one whose actions are governed by (human-designed) software. The human makes his own direct determinations while at the same time being exposed to and influenced by the determinations of the machine. The machine is not "offloading" a task of the anesthesiologist but rather fulfilling a specific role (sustained data scanning) to which it is better suited.

Therefore, the second feature of clumsy technology, that "instead of offloading tasks, autonomous but silent machine agents create the need for team play, coordination and intent communication with people" (Woods et al., 2010) cannot be raised as a criticism in the discussion of anesthesia alarms. As an independent actor in a joint cognitive system, the anesthesia alarm absolutely must strive for team play, coordination and communication in order to fulfill its role successfully. It is the challenges in achieving this coordination (not the regret that it is required) that forms the foundation of this current work.

In this study, the theoretical concept of "user image" was operationalized in the setting of anesthesia monitor alarms. User image mismatch was interpreted as arising from discrepancies between fundamental characteristics of the human operators and the design of the machine and observed to play a role in instances of coordination failure, both as cause and effect. Through interviewees' descriptions, we have seen how user image discrepancies play out in the clinical arena, in coordination between man and machine as well as between other members of the operating room team.

It is suggested that successful coordination in joint cognitive systems requires consideration of fundamental user characteristics in the design of the system. When addressing practical difficulties between man and machine in the applied setting, seeking to align the machine's design with what is understood about how the human operator thinks and works may improve coordination between man and machine.

The most important limitation of this study (which is also, paradoxically, its central strength) is that it focuses on the human's perceptions, insights and opinions rather than measures of performance. We must be careful not to draw conclusions about what sort of alarm design would be better, based on the interview data. The interview data points to coordination breakdown and user image mismatches. We must resist concluding that eliminating the "offending" alarm feature would result in an improved overall performance of that joint cognitive system. Any change exposes the risk of consequences unanticipated by both engineers and users (Dekker, 2005; Woods et al., 2010). For example, that users are perturbed by repetitive alarms does not mean that eliminating that feature would make for a safer system. By focusing on improving measures of “coordination”, we can avoid the pitfall of proposing the elimination of annoying alarm features. One such example is the suggestion for a richer dialogue between man and machine in order that the machine can divert the anesthesiologist’s attention more selectively and therefore more effectively. Even with an emphasis on building coordination, any design changes would have to be carefully tested in simulated and clinical settings.

Another, more elusive pitfall of this work is summarized by the phrase, “What you look for you will find”. Do the relationships described (between user image and coordination) fit a little bit too “neatly”? Could they represent merely a folk model? Dekker and Hollnagel (2004) have cautioned about the susceptibility of the field of Human Factors to the use of folk models. The characteristics of folk models, paraphrased from their 2004 work, include:

- Substitution of one label for another rather than decomposing a large construct into more measurable specifics
- Immunity to falsification (and therefore resistance to the most important scientific quality check)
- Overgeneralization to situations they were never meant to speak about

In their discussion of folk models, Dekker and Hollnagel (2004) also raise a concern about the focus on “uncertain states of the mind” rather than “performance measures”, as addressed above. Future research, particularly if applied to other joint cognitive systems, will help to clarify this important concern.

The power of these categories (regardless of whether they represent a folk model) is that they support us in redirecting the focus of attention from items such as false alarms and context sensitivity to concepts that exist on the “systems” level. Whether or not the categories outlined in previous work on coordination find a lasting place in the discourse on medical alarms should not impact on the goal of shifting the focus from micro to macro elements in the interaction between anesthesiologist and anesthesia alarms.

A notable contradiction was observed in the interview data. The alarms were described as causing anxiety and unwanted attention from other team members. At other times, interviewees blamed the ubiquity of alarms for causing other team members (e.g. nurses) not to notice the alarms or ask if help was required. It is possible that this apparent contradiction results from the variable impact of alarms on different individuals, or under different circumstances. It also possibly reflects the effect of duration of exposure to the alarm, with the frequent alarms initially causing anxiety which over time evolved into apathy.

One could argue that inadequate education is at the heart of some of the user image mismatches described. While improved education and training might have improved the understanding of the alarm function, design must incorporate realities of the "real world". Physician-training in the use of equipment does not approach the rigour seen in the aviation industry. In addition, because individual anesthesia machines within a fleet do not become disabled simultaneously, they are rarely replaced en masse. As a result, anesthesiologists often use two or more different models of equipment throughout their average workweek. The reality that many hospital corporations follow a multi-campus model can foster the same result. Furthermore, the use of "locum" anesthesiologists, who work on a casual basis across a wide number of anesthesia departments, is common practice. For all of these reasons, anesthesia machine and monitor equipment must be as intuitive and transparent[18] as possible.

In summary, the difficulties with anesthesia alarms are not characteristic of "clumsy technology". Viewing the anesthesia alarm as an independent actor, we see how coordination and team play is a requirement for, not a detriment to the successful interaction between man and

[18] Transparency in function may suffice when intuitiveness is not achievable. For example, a snooze button on an alarm clock may not be "intuitive" to the uninitiated. However, after a single use, its function will become obvious (due to its transparency). The pause button on the Carescape B850 was not transparent. Its function did not become obvious through trial and error.

machine in the anesthesia setting. Concepts such as "user image mismatch" and "coordination" can be used as a scaffold on which to elevate the focus from specific aspects of the individual elements to the interaction between those elements in a joint cognitive system. Despite important caveats, this work bridges what might otherwise be seen as a gap between theoretical concepts (fundamental cognitive traits of the user) and the practical realities of how man and machine coordinate their work. Further work, particularly in other applied fields, will be required to determine what extent this research's finding can be found elsewhere.

7. Conclusion

The goal of this project was to explore the existence of mismatch between the machine-embedded and user-described images of the anesthesiologist and to assess whether the impact of any such mismatch was observable within the framework of coordination failure. Three specific alarm functions of the Carescape B850 Anesthesia monitor were studied. User image mismatches were constructed, arising from incongruity between alarm design and fundamental characteristics of the human operator. Characteristics and behaviours of the human operator that were germane to the observed instances of user image mismatches included:

- The interpretation of numeric data in a continuous (non-binary) fashion
- The interpretation of an abnormal parameter in the context of other data parameters as well as the larger clinical picture
- The interpretation of numeric data through their (analog) waveforms
- The interpretation of individual abnormalities symptoms of a single problem
- The reliance on primary data rather than interpreted information

The construct of user image mismatch was observable within the framework of coordination failure. Coordination breakdowns occurred between anesthesiologist and anesthesia alarm working together in a joint cognitive system, as well between anesthesiologist and other members of the operating room team. The user image mismatches were alternately interpreted as both symptom of and contributor to the coordination failures described by interviewees.

A lack of common ground was the most frequent theme in these descriptions, though the need for improved interpredictability and directability were also described. Coordination failure manifested as:

- Incongruent time-tolerances (between man and machine) for “abnormal” parameters

- The human operator’s inability to convey stance, which resulted in the human appearing “unpredictable” to the machine. Awareness of stance would allow the machine to direct the human’s attention more selectively, and therefore, more effectively
- A lack of understanding of key machine functions by the human operators, due in part to a lack of intuitive and transparent functional design

A further theme, “oversimplification”, is proposed. The limited way in which the human was able to communicate to the machine was notable, and it is suggested that allowing a richer two-way dialogue could improve these components of coordination.

In many ways, the distinction between user image mismatch and coordination was an artificial one. More accurately, the use of the framework of coordination provided an alternate lens through which to view the interactions in a joint cognitive system. Though improving the performance of the individual elements of a joint cognitive system (context sensitivity, false alarms) is a worthwhile goal, the current work underlines the importance of focusing on the macro elements in the relationship between man and machine.

8. Acknowledgements

The investigator is grateful to Johan Bergström who provided inspiration and insightful guidance and to Jim Nyce for his support and encouragement. The collaboration of Borje Rantala at General Electric in Finland is gratefully acknowledged as are the efforts of Patrik Nyström, who first suggested, then ultimately facilitated that collaboration. Finally, the investigator is sincerely appreciative of the fourteen anesthesiologists who made this project possible by sharing their time and experience.

Appendix A: Interview Tools (In-depth and Focused)

In-depth Interview Questions

1) Can you describe the situation(s) where you feel that you and the alarm are most out of step?

a) What aspects of the alarm behaviour account for that?

b) Could you understand why the alarm was behaving the way it was? How was it seeing the situation differently than you were?

2) Has there ever been a time that you felt that the alarm hindered your ability to provide quality care?

a) How was it getting in the way of care?

3) In what types of circumstances do you appreciate the alarms? When do they most contribute to your ability to provide safe care?

4) Do you any advice for alarm designers?

5) Are there any other comments you would like to share with me?

Focused Interview Questions

1) The Pause button:

a) Do you find it easy to use the pause button when you want to silence the alarms?

b) Do you find that that function behaves as you expect it to? Are you able to get the period of "quiet" that you are expecting when you press the button?

2) Acknowledging Abnormal Conditions:

a) When you press the pause button, is it usually because:

- there is no clinically abnormal condition OR
- you are aware of the abnormal condition and are actively working on it?

b) Do you find you are prone to forgetting about existing (active) abnormal conditions, especially if distracted? It is helpful to be reminded about them? If so, which conditions?

c) Do you find the alarms hinder your ability to concentrate on treating the problem?

d) Are you ever required to disable the audio on all the alarms? If so, what prompts you to do this?

- Too frequent alarms?
- High probability of false alarms?

- Inability to acknowledge alarms?

<u>3: Display and Prioritization of alarm conditions</u>

a) When a patient has multiple alarms going off at once, how do you decide which "abnormality" to deal with first?

b) Is it helpful to have the machine prioritize alarm conditions for you and to display them according to priority?

c) What advice would you give to alarm designers regarding how alarm conditions should be prioritized?

Appendix B: Request for participant letter

Dear Colleagues,

I am in my second year of a two-year Master's (M.Sc.) program in Human Factors and Systems Safety, at Lund University (Sweden). My chosen thesis project is on anesthesia alarms.

I will be applying a cognitive systems engineering approach to anesthesia alarms in an effort to better understand what at times seems to be a lack of coordination between the alarms and the anesthesiologist in what should be their joint goal of alerting to important changes in patient physiology.

I will be focusing on the AISYS anesthesia machine with the Carescape B850 monitor, what many of us call "the new machine", installed in some of the operating rooms at the JH and MUMC sites. Part of the project will involve analysis of the alarm algorithm and I'm fortunate to have the collaboration of the senior usability engineer from GE on this front. The project has the support of Dr. Kolesar and has been vetted by the chair of the R.E.B. at the H.H.S.

A very important part of the project will involve interviewing anesthesiologists regarding aspects of their interactions with the alarms during daily practice. This is the reason for my email.

- I am seeking 10-15 individuals who would be willing to be interview subjects.
- Subjects must have some experience with the new machines, although no detailed understanding of their function would be required. The questions would revolve around aspects of practice rather than aspects of alarm function per se.
- Subjects must be willing to undertake an interview (conducted by me) at a date, time and location of the subject's convenience. The interview would take approximately 30-40 minutes. It is possible that a follow-up interview will be requested.
- Subjects will be given a consent form to read and sign prior to the interview. I will be asking to record the interview for logistic reasons although this is not mandatory.
- Interviews will be conducted from late November to February.
- There is no remuneration for your participation but I will bring the beverage of choice.

My project is to probe the mismatch between man and machine, **not** to determine if there *is* a mismatch. For that reason, I welcome a biased selection of respondents, those who have experienced in practice a sense that the alarms get in the way of their ability to do their job as anesthesiologists. This is your chance to get it off your chest and perhaps contribute towards improvements in the future.

If you think you might be interested in participating, please respond to me by email (at this address). I will provide further information and answer any questions that you may have.

Thanks in advance for considering this request.

Sincerely,
Karen Raymer

Appendix C: Consent for Participation in Interview Research

1. I volunteer to participate in a research project conducted by Dr. Raymer from Lund and McMaster Universities. I understand that the project is designed to gather information about the interaction of anesthesiologists and anesthesia alarms. I will be one of approximately 15 people being interviewed for this research.

2. My participation in this project is voluntary. I understand that I will not be paid for my participation. I may withdraw and discontinue participation at any time without penalty. If I decline to participate or withdraw from the study, that information would be confidential.

3. I understand that most interviewees will find the discussion interesting and thought-provoking. If, however, I feel uncomfortable in any way during the interview session, I have the right to decline to answer any question or to end the interview.

4. Participation involves being interviewed by Dr. Raymer. The interview will last approximately 30-45 minutes. It is possible that Dr. Raymer may request a follow-up interview. Notes will be written during the interview. An audio tape of the interview will be made and transcribed. If you don't wish to be taped, Dr. Raymer will take notes. This may extend the duration of the interview.

5. I understand that the results of this study (including material arising from interviews) will be published as a thesis; results may also be published as articles in scientific journal(s). I understand that the researcher will not identify me by name in any reports and that my confidentiality as a participant in this study will remain secure.

6. There are no other individuals who will be present at the interview or have access to raw notes or transcripts.

7. I understand that this study was presented to the chair of the research ethics board (R.E.B.) at H.H.S. and it was determined that no review was required.

8. I have read and understand the explanation provided to me. I have had all my questions answered to my satisfaction, and I voluntarily agree to participate in this study.

9. I have been given a copy of this consent form.

----------------------	--------------------------	-----------------------------
My Printed Name	My Signature	Date

-------------------------------	---------------------------------
Signature of Investigator	Date

For more information, please contact:
Dr. Karen Raymer: karenraymer@cogeco.ca
905-308-0352

Appendix D: General Electric Interviews

The planned methodology was to develop the machine-embedded images from analysis of the user guide, expert observation in the clinical setting and interview(s) with a GE engineer. Attempts to establish contact with the relevant individual at GE began in November, 2011. Though contact and interest in collaboration was readily established with the Madison, WI group, it was quickly ascertained that they were responsible for the anesthetic gas delivery unit rather than the monitor.

The anesthesia monitor is designed and manufactured at the Finland location and the two departments were sufficiently separate that the individuals in Madison were not able to refer the investigator to their Finland counterparts.

A Finnish colleague very quickly accessed the relevant name and email address for the Anesthesia monitor usability engineer in Finland. Multiple emails sent from November, 2011 to February, 2012 were unanswered. In February, attempts were made to establish contact in Finland through Canadian GE employees from the sales/managerial side. Simultaneously, the Finnish colleague requested contact directly. As a result, in March 2012, the investigator received an email indicating the willing participation of the principal alarm engineer. Unfortunately, input from the alarm engineer was not available in time to inform the construction of the machine-embedded images.

Exchanges occurred by email over a two-week period. The dedication of the alarm engineer was readily apparent. All questions were answered in a timely manner though responses were often quite general and almost always limited in scope. There was no expressed interested in a discussion of the findings of this project, despite a once-repeated offer to do so.

When asked which specific aspects (if any) of the Anesthesia alarm design presented particular challenges or concerns for the design team, he responded that, "Actually anesthesia is fairly easy as alarms go… [Since] caregivers are always present, simple solutions, like switching complex alarms to low priority or having the audio entirely off, is feasible." Contrasts to the ICU setting were mentioned, such as the faster pace and more extensive artifactual influences. It was also explained that the anesthesia monitor is able to utilize contextual information from the anesthesia machine (ventilator, gas delivery unit) such that thresholds can be modified according to the varied expectations of different phases of anesthesia such as induction or emergence. This response was surprising. As a simple logic exercise they are simple; their interaction with and impact on the human operator is anything but. The literature as well as recent attention paid by prominent organizations (AAMI, 2011; ECRI, 2011; Weil, 2009) would seem to support that Anesthesia alarms are complex and sub-optimally tuned. That the caregiver is "always present" is an ideal that is not always achieved. Moreover, as the participants in this study explained, they may be distracted by other tasks even when physically present. Manually lowering the responsiveness of complex alarms or disabling the audio on alarms are "solutions" that may expose other risks.

The engineer explained that in his view, "the most interesting challenge is to integrate the alarms of the devices at the bedside, using the context knowledge benefits but avoiding the pitfalls. A typical pitfall is the request for a single 'silence alarms' key, which has the very real risk of reflex-style silencing alarms that would require deeper looking-into." 'Context-knowledge benefits' refers to the ability to adjust alarm behaviour according to the specific situation. The examples given apply to the ICU setting: Suspending ventilator-related alarms during suctioning of the airway or lowering blood pressure thresholds during the infusion of nitroprusside (a blood pressure-lowering medication). Woods (2002) has cautioned that attempts to offload context sensitivity to the machine are doomed to brittle and limited results, as discussed earlier. It is not apparent how reflexivity is a pitfall of context-knowledge benefits. The concept of reflexivity was referenced earlier with respect to a rail industry tragedy (Whittingham, 2004).

Regarding the process through which the user image is developed and what goals and conflicts affect that process, the reply was simply that "user requirements have been built over the years through:

1. Anesthesiologists and nurses employed by the company as clinical advisers.
2. Engineering presence in the OR during clinical trials
3. Customer Advisory Board with senior clinicians"

When asked specifically about the repetitiveness of alarms as discussed in Alarm behaviour #2, despite a rather detailed outline of the problems studied in this work, the reply was that, "We are working on how to reduce what we call 'clinically irrelevant alarms'". When asked about the factors considered in the design of the pause button, as pertaining to the issues discussed in "Alarm behaviour #1", the reply was that, "The design was a compromise between Datex and Marquette legacies + evolving IEC standards requirements. Not simple."

The most interesting reply came in response to the question regarding the "Alarm behaviour #3", the prioritization and display of alarm conditions. "The idea with the message field is to have messages and alarms displayed ordered in priority and time order; highest and newest to the left. Of 16 alarms in the buffer, 5 are displayed. The logic is that in the case of an emergency, usually several alarms fire in sequence, and this way the caregiver can follow the evolution of the emergency."

Would an anesthesiologist look to the "order" of appearance of abnormalities in the sense-making process of an emergency? Such an approach can be readily imagined to be useful in a nuclear plant, space shuttle or any other mechanical process, where one abnormality triggers another. In a car, the coolant light might activate (indicating low coolant fluid), with the engine temperature light going off some time later. Understanding the order of events here helps to put together the cause and effect. This sort of mechanistic view doesn't lend itself to human physiology. When multiple abnormalities are occurring, they usually all symptoms of an underlying cause (which itself is rarely monitored directly), rather than one causing the next in sequence. Accordingly, one can't draw conclusions based on the ordering of abnormalities, especially when they are usually happening within a very tight timeframe.

There is another equally important reason to question the relevance of the chronological ordering of alarm conditions. The message area is ordering conditions based on when they reached the alarm threshold. Clinicians are interested in trends, when condition changes from baseline. The moment that the parameter reaches an arbitrary threshold is just that: Arbitrary. Placing undue importance on the moment of an alarm event again underscores the binary approach to physiology that has previously been discussed.

The description of prioritizing both high priority conditions and most recently-activated conditions is unclear. How are two variables plotted on the same axis? Finally, the fact that vast majority of alarm conditions are clinically irrelevant limits the utility of any micro-managed presentation of those alarms, such as that which manifests in the message area.

In summary, the timeline of the collaboration with GE was such that the planned influence on machine-embedded image development could not occur. The engineer highlighted the tension between clinically irrelevant alarms and the risk of making it "too easy" for the clinician to disable alarms. That tension is highly interlinked: The more false alarms there are, the more "practiced" the clinician would be in disabling them. The most interesting challenge was felt to be incorporating context sensitivity without allowing the inappropriate limiting of alarm response. Monitor and alarm design seemed to be heavily influenced by previous model designs. The emerging IEC standards add additional complexity to design considerations. The impact of a "binary" view of physiologic parameters on alarm design was again seen through the description of the design rationale of the message area.

References

Association for the Advancement of Medical Instrumentation (AAMI). (2011). *A siren call to action. Priority issues from the 2011 medical device alarms summit* [online]. Available from http://www.aami.org/alarms/Materials/2011_Alarms_Summit_publication.pdf [Accessed 24 February 2012].

Baig, M.M, Gholamhosseini, H.G., Kouazani, A., & Harrison, M.J. (2011). Anaesthesia monitoring using fuzzy logic. *Journal of Clinical Monitoring and Computing*, 25, 339-347.

Bitan, Y., & Meyer, J. (2007). Self-initiated and respondent actions in a simulated control task. *Ergonomics,* 50 (5), 763-788.

Blaxter, L., Hughes, C., & Tight, M. (2006). *How to Research (3rd edition).* Berskshire, England: Open University Press, McGraw Hill House.

Bliss, J.P., & Dunn, M.C. (2000). Behavioural implication of alarm mistrust as a function of task workload. *Ergonomics,* 43(9), 1283-1300.

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

Bustamante, E.A., Bliss, J.P., & Anderson, B.L. (2007). Effects of varying the threshold of alarm systems and workload on human performance. *Ergonomics,* 50(7), 1127-1147.

Cooper, J.B., Newbower, R.S., Long, C.D., & McPeek, B. (2002). Preventable anesthesia mishaps: A study of human factors. *Quality and Safety in Health Care,* 11(3), 277-282.

Dekker, S.W.A. (2005). *Ten Questions about Human Error.* New York, NY: CRC Press.

Dekker, S.W.A., & Hollnagel, E. (2004). Human factors and folk models. *Cognition, Technology & Work,* 6, 79-86.

Dixon, S.R., Wickens, D., & McCarley, J.S. (2007). On the independence of compliance and reliance: Are automation false alarms worse than misses? *Human Factors,* 49(4), 564-572.

ECRI Institute. (2011). Top 10 Health Technology Hazards for 2012. *Health Devices*, 40 (11).

Edworthy, J., & Hellier, E. (2006). Alarms and human behaviour: implications for medical alarms. *British Journal of Anaesthesia*, 97, 12-17.

Ferris, T.K., & Sarter, N. (2011). Continuously informing vibrotactile displays in support of attention management and multitasking in anesthesiology. *Human Factors*, 53(6), 600-611.

Gorges, M., Markewitz, B.A., & Westenskow, D. R. (2009). Improving alarm performance in the medical intensive care unit using delays and clinical context. *Anesth Analg*, 108, 1546-1552.

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.

Hagenouw, R.R.P.M. (2007). Should we be alarmed by our alarms? *Current Opinion in Anesthesiology*, 20, 590-594.

Healey, A.N., Sevdalis N., & Vincent, C.A. (2006). Measuring intra-operative interference from distraction and interruption observed in the operating theatre. *Ergonomics*, 49 (5-6), 589-604.

Heeger, D. (1997). Signal detection theory. *Journal of Physiology London*, 12 (3), 1-10.

Heeger, D. (2007). *Signal detection theory.* Retrieved Sept. 9, 2011, from David Heeger's website: http://www.cns.nyu.edu/~david/handouts/sdt/sdt.html

Hollnagel, E., & Woods, D. D. (1983). Cognitive systems engineering: New wine in new bottles. *International Journal of Man-Machine Studies*, 18, 583-600.

Imhoff, M., & Kuhls, S. (2006). Alarm Algorithms in critical care monitoring. *Anesthesia and Analgesia,* 102, 1525-1537.

Imhoff, M., Kuhls, S., Gather, U., & Fried, R. (2009). Smart alarms from medical devices in the OR and ICU. *Best Practice & Research Clinical Anaesthesiology,* 23, 39-50.

Kestin, I.G., Miller, B.R., & Lockhard, H. (1988). Auditory alarms during anaesthesia monitoring. *Anesthesiology,* 69, 106-109.

Kiefer, N., & Hoeft, A. (2010). Display of information in the operating room. *Current Opinion in Anesthesiology*, 23, 772-777.

Klein, G., Feltovich, P.J., Bradshaw J.M., & Woods, D.D. (2004). Common ground and coordination in joint activity. In W. Rouse & K. Boff (Eds.), *Organizational simulation*. (pp. 139-84). Hoboken, NJ: John Wiley & Sons.

Kowalcyzk, L. (2011, February 13). Alarm fatigue blamed in hospital deaths. *The Boston Globe*. Retrieved October 4, 2011, from http://www.boston.com/lifestyle/health/articles/2011/02/13/patient_alarms_often_unheard_unheeded/?page=full

Masalonis, A., & Parasuraman, R. (2003). Fuzzy signal detection theory: analysis of human and machine performance in air traffic control, and analytic consideration. *Ergonomics*, 46(11), 1045-1075.

Meyer, J. (2001). Effects of warning validity and proximity on responses to warnings. *Human Factors*, 43(4), 563-572.

Mondor, T., & Finley, G. (2003). The perceived urgency of auditory warning alarms used in the hospital operating room is inappropriate. *Canadian Journal of Anesthesia*, 50, 221-228.

Nyssen, A.S., & Blavier, A. (2006). Error detection: A study in anaesthesia. *Ergonomics,* 49 (5-6), 517-525.

O'Carroll, T.M. (1986). Survey of alarms in an intensive therapy unit. *Anaesthesia*, 41, 742-744.

Parasuraman, R., Hancock, P.A., & Olofinboba, O. (1997) Alarm effectiveness in driver-centered collision-warning systems. *Ergonomics,* 40, 390-399.

Phipps, D.L., Parker, D., Pals, E., Meakin, H., Nsoedo, C. & Beatty, P. (2008). Identifying violation-provoking conditions in a healthcare setting. *Ergonomics*. 51 (11), 1625-1642.

Sanderson, P.M., Wee, A., & Lacherez, P. (2006). Learnability and discriminability of melodic medical equipment alarms. *Anaesthesia*, 61, 142-147.

Schmid, F., Goepfert, M.S., Duhnt, D., Eichhorn, V., Diedrichs, S., Reichenspurner, H., et al. (2011). The wolf is crying in the operating room: Patient monitor and Anesthesia workstation alarming patterns during cardiac surgery. *Anesthesia and Analgesia*, 112 (1), 78-83.

Seagull, F.J., & Sanderson, P.M. (2001). Anesthesia alarms in context: An observational study. *Human Factors*, 43, 66-78.

Siebig, S., Kuhls, S., Imhoff, M., Gather, U., Scholmerich, J., & Wrede, C.E. (2010). Intensive care unit alarms: how many do we need? *Critical Care Medicine*, 38, 451-456.

Smith, A.F., Goodwin, D., Mort, M., & Pope, C. (2003). Expertise in practice: An ethnographic study exploring acquisition and use of knowledge in anaesthesia. *British Journal of Anaesthesia,* 91, 319-328.

Sorkin, R.D., & Woods, D.D. (1985). Systems with human monitors: a signal detection analysis. *Human- Computer Interaction,* 1, 49-75.

Solet, J.M. & Barach, P.R. (2012). Managing alarm fatigue in cardiac care. *Progress in Pediatric Cardiology.* 33, 85-90.

Stoelting, R.K., & Miller, R.D. (2006). *Basics of Anesthesia*. New York, NY: Churchill Livingstone.

Takla, G., Petre J.H., & Doyle, D.J. (2006). The problem of artifacts in patient monitor data during surgery: a clinical and methodological review. *Anesthesia and Analgesia,* 103, 1196-1204.

Wee, A.N., & Sanderson, P.M. (2008). Are melodic medical equipment alarms easily learned? *Anesthesia and Analgesia*, 106, 501-508.

Weil, K.M. (2009, September 18). *Alarming Monitor Problems*. Retrieved January 10, 2012 from The Food and Drug Administration's website: http://www.fda.gov/MedicalDevices/Safety/AlertsandNotices/TipsandArticlesonDeviceSafety/ucm222022.htm

Whittingham, R.B. (2004). *The blame machine: Why human error causes accidents*. Burlington, MA: Elsevier.

Woods, D.D. (1985). The alarm problem and directed attention in dynamic fault management. *Ergonomics*, 38(11), 2371-2393.

Woods, D., Dekker, S.D.A., Cook, R., Johannesen, L., & Garter., N. (2007). *Behind Human Error (2nd Edition).* Surrey, England: Ashgate Publishing.

Woods, D. D., Patterson, E. S., & Roth, E. M. (2002). Can we ever escape from data overload? A cognitive systems diagnosis. *Cognition, Technology & Work*, 14, 22-36.

Wright, M.C., Dawodu, M., Segall, N., Taekman, J.M., & Mark, J.B. (2011) Anesthesia crisis alerting: A qualitative study toward understanding the complexity of crisis identification. *Anesthesiology*, A741.

Xu, X., Wickens, C., & Rantanen, E.M. (2007). Effects of conflict alerting system reliability and task difficulty on pilots' conflict detection with cockpit display of traffic information. *Ergonomics*, 50 (1), 112-130.

Yin, R. K. (2009). *Case study research: Design and methods (4th edition).* Thousand Oaks, CA: Sage Publications.