

Anesthesia Alarms in Context: An Observational Study

F. Jacob Seagull, University of Illinois at Urbana-Champaign, Urbana, Illinois, and Penelope M. Sanderson, Swinburne University of Technology, Melbourne, Australia

This paper surveys current work on the design of alarms for anesthesia environments and notes some of the problems arising from the need to interpret alarms in context. Anesthetists' responses to audible alarms in the operating room were observed across four types of surgical procedure (laparoscopic, arthroscopic, cardiac, and intracranial) and across three phases of a procedure (induction, maintenance, and emergence). Alarms were classified as (a) requiring a corrective response, (b) being the intended result of a decision, (c) being ignored as a nuisance alarm, or (d) functioning as a reminder. Results revealed strong effects of the type of procedure and phase of procedure on the number and rate of audible alarms. Some alarms were relatively confined to specific phases; others were seen across phases, and responses differed according to phase. These results were interpreted in light of their significance for the development of effective alarm systems. Actual or potential applications of this research include the design of alarm systems that are more informative and more sensitive to operative context than are current systems.

INTRODUCTION

Health care is increasingly drawing the attention of cognitive engineers and human-computer interaction researchers. Information technology is being introduced into a wide variety of medical devices and systems, such as patient physiological monitoring systems, anesthesia machines, respirators, home health care devices, and hospital and medical information systems. At the same time, such medical devices and systems are being used by a wider population, necessitating better interface design. These trends have led to a concern for human error in medicine (Bogner, 1994) as well as for developing guidelines and standards for the design of medical devices and systems.

Safe and effective human performance with medical devices and systems is essential. Yet the human operators of such systems are challenged not only by increasing device complexity but also by growth in the number of devices that have to be coordinated in a single work

environment. This is particularly so in critical care environments such as the operating room (OR), the intensive care unit (ICU), and the emergency room (ER). Anesthetists are challenged by the fact that anesthetic drugs now have a more rapid onset, higher potency, and faster emergence, making swifter responses to changing and potentially life-threatening situations very important (Cook & Woods, 1996; Schreiber & Schreiber, 1989). Alarms play an important role in ensuring swift and effective responses from medical and nursing personnel.

Alarms in Critical Care Environments

The problem of how to design effective alarms in critical care environments has been of particular interest for the past decade. Researchers have applied findings from studies in aviation and nuclear power plant environments to critical care (Cook & Woods, 1996). However, anesthesia and critical care environments pose distinct problems. Practitioners are regularly confronted with arrays of technologies that were

introduced in a piecemeal fashion and were not necessarily designed to operate in conjunction with one another. Moreover, the most important part of the sensed environment – the patient – is not a fixed engineered system but instead has unique physiological characteristics that need to be taken into account when determining what is normal and abnormal. Alarms can reflect the state of the patient (such as unacceptable high heart rate) or the state of the equipment itself (such as loss of signal from electrodes or failure of a breathing circuit), which sometimes may have important consequences.

Researchers in the critical care area have focused on several aspects of the alarm problem: (a) the acoustic properties of alarms, such as audibility, discriminability, and identifiability; (b) the intrinsic meaningfulness of alarms; (c) the use of alarms as sources of information in context; and (d) the organization and management of alarms in context. These aspects will be briefly outlined in the next few paragraphs as they provide a background to the present study.

Some researchers have focused on the audibility of alarms in critical care environments (Momtahan, Hetu, & Tansley, 1993), a problem shared with other domains. Patterson's (1989) approach to designing discriminable auditory alarm sets for aviation has been generalized to critical care environments (Patterson, Edworthy, Shailer, Lower, & Wheeler, 1986) and has been proposed as a standard for them. However, Weinger (1991) has raised objections to the adequacy of an approach based solely on Patterson's criteria. Further work has investigated how accurately individual OR alarms are recognized within a full set of alarms (Loeb, Jones, Leonard, & Behrman, 1992) and how readily they are identified in their own right (Block, 1988). Although audibility and discriminability are necessary, it is unclear whether laboratory-based recognition accuracy tests generalize to the field.

Other researchers have focused on the intrinsic meaningfulness of alarms. Edworthy, Loxley, and Dennis (1991) sought ways to map *urgency onto acoustic parameters*. The work of Gaver (1993) on ecological acoustics and of Blattner, Sumikawa, and Greenberg (1989) on

"earcons" has been particularly influential. Edworthy and Stanton (1995) have built on these foundations to develop a set of alarms for a coronary care unit. However, Woods (1995) argued that creating intrinsically meaningful alarm sounds may be relatively unimportant in highly professional work environments, where all practitioners are highly skilled.

Further studies focus on alarms in their context of interpretation, explaining why alarms are heard and understood but often inappropriately ignored. Kesting, Miller, and Lockhart (1988) found that in the typical OR, an alarm sounds every 4.5 min and that 75% are false alarms. Recognizing that too-frequent false alarms desensitize operators and reduce the probability of attention or response to the alarm, Breznitz (1984) has termed this the "cry wolf" effect. The cry wolf effect has been empirically corroborated in a simple context-free laboratory study (Bliss, Gilson, & Deaton, 1995) in which participants learned to match their probability of response to an alarm, with the probability that the alarm would give true information. However, these results ignore any domain context.

Stanton (1994) noted that most research on human response to alarms focuses on observing and accepting or correcting. However, he identified further responses (analyzing, investigating, and monitoring), in which an operator often cannot tell what the alarm means without reference to its background context. The background context may represent normal system functioning, an attempt to diagnose an unexpected finding, or execution of an action plan based on a diagnosis (Woods, 1995).

Human operators sometimes have problems organizing and managing alarms; moreover, alarms sometimes cannot keep up with the pace of activity. Moll van Charante, Cook, Woods, Yue, & Howie (1993) found that critical care teams would start one patient intervention, switch to another kind of intervention on the same patient, but then get an alarm signaling that the first intervention was unsuccessful. Visual and auditory interfaces should help anesthetists when things go wrong, but most alarms occur and have to be dealt with in busy situations in which the alarms increase the operator's workload rather than decrease it (Cook & Woods, 1996). To help alleviate these

problems, researchers have started to develop intelligent alarm management for critical care environments, with the goal of reducing the number of unnecessary, cascading, or false alarms (Orr & Westenskow, 1994). In the present paper we examine examples of such situations and consider possible approaches to intelligent alarm management.

Goals of the Study

If alarms in critical care and anesthesia environments are to be more informative and less distracting, it is probably more important to know about how alarms are interpreted in context than simply about their audibility or intrinsic meaningfulness. The present report focuses on anesthesiologists' responses to alarms across different phases of surgery in different kinds of surgical procedures. The induction, maintenance, and emergence phases of anesthesia involve different clinical issues and different uses of anesthesia equipment. Surgical procedures differ in invasiveness, level of life support used, and the potential problems encountered. The goal is to seek evidence for different uses of alarms in different contexts in order to determine whether contextually based intelligent alarm systems would be helpful or not. Four classes of questions were posed:

1. Do the frequency and kind of alarm change across phases of a surgical procedure, and is this affected further by the type of surgical procedure?
2. Do anesthesiologists' responses to alarms differ across the induction, maintenance, and emergence phases of surgical procedures?
3. Do responses to alarms differ across different kinds of procedures?
4. Do responses to the same alarm differ under different conditions?

METHOD

Participants

Participants were highly trained, certified registered nurse anesthetists (CRNAs) working independently under the direction of medically qualified anesthesiologists. Informed written consent was obtained from all CRNAs participating and from all patients whose procedures were

observed, usually on the day prior to surgery. Anonymity of all participants was ensured.

Selection of Cases to Observe and Sampling Strategy

Twenty-four surgical cases were observed by the first author over a four-week period. After pilot investigation and consultation with anesthesia professionals, we chose four types of procedures to sample the degree to which vital organs or systems were involved and the amount of control the CRNA assumes over the patient's vital functions: (a) laparoscopic ("closed") abdominal surgery, including hernia, gall bladder, and gynecological (e.g., tubal ligation) procedures; (b) arthroscopic ("closed") knee surgery, excluding knee replacement; (c) cardiac bypass ("open heart") surgery, including coronary artery bypass grafting (CABG) and valve replacements; and (d) intracranial ("brain") surgery, including cerebral aneurysm, biopsy of neuromas, and duraplasty.

Laparoscopy is minimally invasive, but the abdominal laparoscopic procedures we observed involved vital organs. The patients were intubated (breathing tube inserted down the trachea) to ensure adequate pressure of air to the lungs while pressurized air was being introduced into the body cavity. Arthroscopy is also minimally invasive, and because it does not directly involve the patient's vital organs, no intubation was used. Cardiac and intracranial procedures are both very invasive and involve vital organs: the brain in intracranial procedures and the heart and the lungs in cardiac procedures. Cases were selected on the basis of their availability during the period of observation and their fit to the chosen categories of procedure. Six cases from each type of surgical procedure were observed, making 24 in all.

Equipment Used

The anesthesia machines we observed were manufactured by Ohmeda (Tewksbury, MA) principally the 7850 model but in some cases the 7000 or 7810 ventilators with CO₂, O₂, airway pressure, pulse oximetry, and noninvasive blood pressure monitors. The patient-monitoring system used for ECG, invasive pressures, and other vital signs was the Marquette Tramscope 12e (Milwaukee, WI). For intravenous delivery

of drugs, the Baxter Bard Infus OR infusion pump (Deerfield, IL) and the Abbot Omnilflow 4000 plus (Abbott Park, IL) were used. In the cardiac cases the Abbott Critical Care System Qvue Continuous Cardiac Output Computer was used.

Data Collection

At the time the study was performed, no video or audio recording was authorized. A system of data collection was developed that included (a) process tracing, consisting of a written log of events and activities during each procedure; (b) general field notes from discussions with anesthesiologists during surgical procedures; (c) postoperative review with the anesthesiologist; and (d) collection of perioperative records, documents, and charts.

These sources were selected to provide a rich and diverse set of data and therefore a comprehensive perspective on the procedures observed. During each procedure the observer was stationed near the anesthesia workstation and took notes on forms developed specifically for the observation. The notes captured the following, plus the time at which each happened: (a) principal surgical events and transitions between phases of anesthesia (e.g., start of emergence); (b) primary goal-directed activity and information-seeking by the anesthesiologist (e.g., check tubing); (c) major equipment states (e.g., alarms and warnings) and any actions initiated (e.g., drug delivery); (d) charting, recording, and calculations performed (e.g., estimates of doses and times); and (e) calibration and coordination of equipment (e.g., change alarm limits).

Field notes contained records of intraoperative conversations with the anesthesiologist that documented the anesthesiologist's rationale for actions, explanations of procedures, and opinions. They were also used more generally to record relevant activities or events outside the strict scope of the process tracing. Field notes were recorded alongside the process-tracing log sheet and were included in the transcript of the procedures. After the surgery, the handwritten process trace was typed up along with the field notes. A copy was submitted to the anesthesiologist for review and correction within 24 hr of the case.

In addition, a number of official documents were collected for each case. These included

the preoperative evaluation record, the intraoperative anesthesia record, and results of any tests taken during the procedure.

RESULTS AND DISCUSSION

Summary and Characterization of Cases Observed

Table 1 shows the configuration, anesthetic management, and time per phase for the four kinds of procedures observed. As discussed, procedures were selected for differences in involvement of physiological and anatomical systems, degree of invasiveness, and requirements for life support. Table 1 presents dimensions that further distinguish the four kinds of procedure.

Number of cases. Two cases were removed post hoc from the analysis, leaving 22 cases. One was an arthroscopic surgery in which the anesthesiologist changed the configuration of the technology in conscious reaction to the presence of observers, which led to an unusually high rate of alarms. The second case was a cardiac surgery observed near the start of observations in which insufficient data had been recorded to perform the analyses needed.

Intubation. Intubation was used when surgery affected or could affect vital organs and so was present for cardiac, intracranial, and laparoscopic cases but absent in arthroscopic cases. In some cases a laryngeal mask airway was used, which was less invasive than intubation.

Arterial line. Sometimes an internal (invasive) blood-pressure measurement, such as an arterial line or Swan-Ganz catheter, was used for patient monitoring. Invasive lines were used in cardiac and intracranial cases because of the sensitivity of both procedures to blood pressure and oxygenation.

Inhaled agents. In some cases, total intravenous anesthetic techniques were used, involving an intravenous pump and no inhaled (gas) agents. All cases used IV agents, but none of the cardiac cases involved inhaled agents.

BP/HR concern. Heart rate (HR) and blood pressure (BP) were sometimes sources of concern, requiring an administration of vasoactive drugs or a change in the level of anesthetic agent. BP/HR concern was noted only if the CRNA indicated that it rose above a general background level.

TABLE 1: Summary of Procedures Observed, Patient Configuration, Special Notes, and Time for the Induction, Maintenance, and Emergence Phases of Each Kind of Procedure Observed

| Procedure | Cases (n) | Patient Configuration | | | Special Notes | | Times | | |
|-------------|-----------|-----------------------|----------------|----------------|-------------------|-------------------|----------------------|----------------------|----------------------|
| | | Intubation | Arterial Line | Inhaled Agents | BP/HR Concern | Health Concern | Induction | Maintenance | Emergence |
| Laparoscopy | 6 | Y | N | Y | -- | Mild (3) | 0:05:00 [0:00:36] | 0:38:00 [0:06:12] | 0:11:00 [0:03:48] |
| Arthroscopy | 5 | N (4) LMA (1) | N | Y | Y (1) Mild (2) | Mild (4) | 0:12:00 [0:00:48] | 0:48:00 [0:06:42] | 0:11:00 [0:01:18] |
| Cardiac | 5 | Y | Y | N | Y (1) Mild (2) | Y (5) | 0:29:00 [0:04:36] | 4:41:00 [0:42:18] | 0:13:00 [0:01:24] |
| Cranial | 6 | Y | Y (5) N (1) | Y | Y (5) | Y (1) Mild (2) | 0:07:00 [0:01:48] | 3:42:00 [0:19:54] | 0:17:00 [0:00:48] |

Note: Time is given in hr: min: s. The time in brackets is the standard error of the mean. LMA = laryngeal mask airway.

Health concerns. A patient requiring surgery may be unhealthy in a way that is unrelated to the need for surgery but still capable of complicating the surgery. An example is a chronic smoker with lung problems who is having knee surgery; although the two conditions are unrelated, concern over breathing difficulties creates additional complications.

Time in phase of procedure. Table 1 shows the average time in induction (includes preparation of the patient), maintenance, and emergence for each type of procedure. These results reflect practice in the particular operating suite observed, as well as the nature of the procedures themselves. Induction and emergence include only the time that the patient was in the OR.

Considerable preparation was done before intracranial patients were brought to the OR, whereas for cardiac patients, most preparation was done in the OR. Cardiac patients did most of their emergence in the ICU rather than the OR. Cardiac and intracranial cases were considerably longer than the others, with most of the anesthetist's time taken up in maintenance. A two-way analysis of variance (ANOVA) showed significant effects of kind of surgical procedure, $F(3, 18) = 35.9, p < .0001$; phase of procedure, $F(2, 18) = 137.4, p < .0001$; and their interaction, $F(6, 36) = 25.6, p < .0001$.

Even in this relatively small sample, then, characteristic patterns of anesthesia practice and patient configuration are readily apparent, despite variations caused by practitioner preferences and each patient's individual needs. There is consistency among the cases within a procedure type and significant differences among procedure types.

Coding of Alarms

During each case, we noted audible alarms from all equipment used by the anesthetist. Alarms were usually noted from transport monitors as the patient was brought into and out of the OR, but statistical analyses were performed only for audible alarms that occurred within the OR. Some alarms on the OR equipment were not recorded, either because they were transient and disappeared before they could be recorded or because they oc-

curred when alarms had been silenced temporarily by the anesthetist. Our focus was on audible alarms. Therefore, these data have been subjected to various sources of random and systematic variability, but none are so strong, we believe, as to call into question the general picture that emerges.

Table 2 shows the distribution of audible alarms across the 22 cases observed. The first column gives a letter code for each CRNA. Most of the cases represented a different combination of CRNA and surgeon. However, because of the limited period in which we had permission to observe and the specialization of CRNAs, CRNAs were not randomly allocated to procedures. Multiple procedures with the same CRNA are seen, especially for the laparoscopic and arthroscopic cases. Because of these unavoidable constraints of the field environment, the statistics we present should be considered suggestive rather than definitive.

Alarms do not function simply to warn of problems but instead are used as tools with widely varying functions depending on type and phase of procedure. Based on pilot work, we classified anesthetists' responses to alarms into four categories:

1. "Correction" or "Change" (C): Act to correct an unexpected event signaled by valid alarm. This is the "canonical" design expectation for alarms and involved correcting an unforeseen problem by administering drugs, changing the patient's configuration, or changing equipment settings.
2. "Intended" or "Expected" (E): No action needed to a valid alarm. The measurement of the physiology was valid and out of acceptable normal range, but the abnormal state was being tolerated by the anesthetist. Examples include hyperventilating a patient before intubation or letting CO_2 levels rise to encourage independent breathing by the patient during emergence.
3. "Ignore" or "Nuisance" (N): No action needed to a nuisance alarm. This includes recognizing that an artifact was sensed by the technology, such as movement of the sensor, surgical activity, or transient physiological change, as when a breathing tube is briefly but purposefully disconnected.
4. "Reminder" (R): Act to initiate an expected action on valid alarm. Here the alarm is used as a reminder to carry out a needed task such as manually ventilating the patient.

TABLE 2: Details of Audible Alarms within Each Procedure

| CRNA(s) | Induction | | Maintenance | | Emergence | |
|-------------|---------------------|------|--|------|---|------|
| | Responses | Runs | Responses | Runs | Responses | Runs |
| Laparoscopy | | | | | | |
| A | R ² N | 2 | n | 0 | NC | 2 |
| A | n | 0 | n | 0 | N ² | 1 |
| B | N | 1 | n | 0 | n | 0 |
| C | n | 0 | n | 0 | n | 0 |
| B | EN ² | 2 | n | 0 | R ² NR ² C ² N | 5 |
| C | NC | 2 | C | 1 | N | 1 |
| Arthroscopy | | | | | | |
| D | n | 0 | C | 1 | N ² | 1 |
| E | n | 0 | RC | 2 | R ² | 1 |
| E | n | 0 | CR ⁵ E ³ CCER ⁶ | 7 | R ² | 1 |
| D | NR | 2 | R ³ | 1 | N | 1 |
| E | n | 0 | CR | 2 | RCN | 3 |
| Cardiac | | | | | | |
| S | n | 0 | C | 1 | n | 0 |
| C | n | 0 | C | 1 | N | 1 |
| I | CEN | 3 | CNC | 3 | N | 1 |
| J | n | 0 | CNNCNC | 6 | NNN ² C | 4 |
| K | N ³ CNNC | 5 | NNCN | 4 | NNN ² N | 4 |
| Cranial | | | | | | |
| F | NCN ² | 3 | C ² | 1 | N | 1 |
| C | E | 1 | CNNC | 4 | N | 1 |
| A | EN | 2 | C | 1 | NN | 2 |
| B/G | ECNCCN | 6 | NCCC | 4 | NCNN | 4 |
| C | RCN | 3 | NCCNCN | 6 | CRNN | 4 |
| H | N ² ECN | 4 | CNNNCN | 6 | NNN ⁵ R ³ C | 5 |

Note. Letters in the "CRNA" column are codes for the CRNAs observed. Letters in the "Responses" columns stand for the CRNA's response to each audible alarm. Numbers in the "Runs" columns indicate how many runs were seen for the phase indicated. C = corrected; E = intended or expected; N = ignored nuisance alarm; R = reminder, and n = no audible alarms recorded. Superscripts indicate a run of similar responses to the same alarm.

In Table 2 we have classified audible alarms according to the kind of responses made by CRNAs. As the table indicates, the number of alarms is not normally distributed across cases within procedure phases. Many audible alarms were repetitions of an initial alarm. For some analyses it makes sense to count individual alarms, whereas for others it makes sense to count runs of a similar kind of alarm. A series of three apnea (lack of breathing) alarms that CRNAs use as reminders to ventilate a patient manually are treated not as separate events, RRR, but as a single episode: R³. Similarly, a corrective action to a low-HR alarm followed by two corrective actions to a low-BP alarm are considered two separate incidents and so are coded as CC² rather than CCC. In this way

we "smooth" the observations to reflect episodes in which the CRNA deals with a persistent problem.

Using this approach, we reduced the initial 165 audible alarms to 132 episodes or "runs" of audible alarms. Table 3 shows that audible alarms that are expected or that lead to corrections generally do not arrive in runs, whereas alarms that are ignored or act as reminders are more likely to arrive in runs, $\chi^2_{(1)} = 25.26, p < .001$. The rightmost field of Table 3 shows the number of runs of the same audible alarm with the same CRNA response. Almost half the audible alarms or episodes are ignored (48%), 33% are corrected, 12% act as reminders, and 6% are expected by the CRNA.

TABLE 3: Tabulation of CRNA Responses to Alarms across Different Phases, Collapsed over Kinds of Procedures

| Responses | All Alarms ^a | | | | Run Lengths ^b | | Same-Code Runs ^c | | | |
|-----------|-------------------------|-----|-----|-----|--------------------------|-----|-----------------------------|-----|-----|-----|
| | I | M | E | | 1 | > 1 | I | M | E | |
| Corrected | 10 | 28 | 8 | 28% | 42 | 2 | 10 | 27 | 7 | 33% |
| Intended | 6 | 4 | 0 | 6% | 7 | 1 | 6 | 2 | 0 | 6% |
| Ignored | 22 | 17 | 37 | 46% | 56 | 8 | 18 | 17 | 29 | 48% |
| Reminder | 4 | 16 | 13 | 20% | 7 | 9 | 3 | 6 | 7 | 12% |
| | 25% | 39% | 35% | 165 | 85% | 15% | 28% | 39% | 33% | 132 |

Note. I = induction; M = maintenance; E = emergence.

^a Results for all 165 alarms heard. ^b Number of runs of the same response to a given alarm of length = 1 versus number of runs of length > 1. ^c Results for all 132 runs of the same response to a given alarm.

Answers to Four General Questions

In this section we show how the data answered the four general questions about audible alarms that were posed in the introduction.

Do the frequency and kind of alarm change across phases of a surgical procedure, and is this affected further by the type of surgical procedure? The answer to both parts is a qualified "yes." As Table 1 shows, the phases of different procedures had different durations, making direct comparisons of alarm frequencies across phases and procedures difficult. The data were skewed by instances of cascades of audible alarms that reflect a CRNA's preferred use of alarm settings. Table 2 shows that for laparoscopies and arthroscopies, there were generally few audible alarms until a cascade of events occurred. Not even smoothing the data by col-

lapsing over runs removed these cascades, as they usually reflected genuinely complex situations. Our data suggest that the CRNA may actively engineer a cognitive and acoustic environment for the task at hand that sometimes leads to many alarms, but sometimes to very few, depending on need.

Table 4 shows the actual alarm episodes observed in the study, how they were distributed across phases, and how anesthetists responded to them. By far the largest category of alarm episodes was apnea alarms (31.8%), mainly because apnea alarms could be triggered many ways by events detected in the breathing circuit. Apnea, SpO₂ (oxygenation of the blood), and ECG lead failure ("Leads Fail" or failure of signal from an electrode) together accounted for 51.5% of all audible alarms.

Because of the sparsity of data, it was not

TABLE 4: Sources of Errors and CRNA Responses, as Distributed across Induction, Maintenance, and Emergence Phases of the Procedures Observed

| Category | Induction | | | | Maintenance | | | | Emergence | | | | Total | % |
|------------------------|-----------|---|----|---|-------------|---|----|---|-----------|---|----|---|-------|------|
| | C | E | N | R | C | E | N | R | C | E | N | R | | |
| Apn ea | 0 | 2 | 11 | 3 | 2 | 0 | 3 | 5 | 3 | 0 | 6 | 7 | 42 | 31.8 |
| SpO ₂ | 1 | 0 | 1 | 0 | 4 | 0 | 4 | 0 | 1 | 0 | 3 | 0 | 14 | 10.6 |
| Leads fail | 0 | 0 | 0 | 0 | 3 | 0 | 1 | 0 | 0 | 0 | 8 | 0 | 12 | 9.1 |
| Low ETCO ₂ | 0 | 3 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 4 | 3.0 |
| BP | 0 | 0 | 0 | 0 | 6 | 0 | 2 | 0 | 0 | 0 | 2 | 0 | 10 | 7.6 |
| HR | 0 | 0 | 0 | 0 | 4 | 2 | 1 | 0 | 0 | 0 | 0 | 0 | 7 | 5.3 |
| High ETCO ₂ | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 2 | 1.5 |
| Vent | 2 | 0 | 1 | 0 | 1 | 0 | 2 | 1 | 0 | 0 | 2 | 0 | 9 | 6.8 |
| High agent | 3 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 5 | 3.8 |
| Calibration | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0.8 |
| SvO ₂ | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 1.5 |
| Miscellaneous | 2 | 0 | 4 | 0 | 0 | 0 | 3 | 0 | 1 | 0 | 6 | 0 | 16 | 12.1 |
| Pump | 1 | 0 | 0 | 0 | 4 | 0 | 1 | 0 | 0 | 0 | 2 | 0 | 8 | 6.1 |
| Total | 10 | 6 | 18 | 3 | 27 | 2 | 17 | 6 | 7 | 0 | 29 | 7 | 132 | 100 |

TABLE 5: Tabulation of the 132 Alarm Episodes into High versus Low Frequency Alarm Categories by Corrective versus Noncorrective Responses

| Responses | Corrective | Noncorrective | Percentage |
|----------------|------------|---------------|------------|
| Alarm Types | | | |
| High frequency | 14 | 54 | [52%] |
| Low frequency | 30 | 34 | [48%] |
| Total | [33%] | [67%] | 132 |

Note. High frequency alarm categories include apnea, SpO₂, and ECG leads failure, and account for just over 50% of all audible alarms (see Table 4). Low frequency audible alarm categories are all others.

possible to test for significant differences among alarm categories in how responses were distributed between phases. Despite the impression given by the raw data, the three most frequent types of audible alarms were not more likely to be ignored than the less frequent types, $\chi^2_{(1)} = 1.97$, *ns*. CRNAs may have silenced alarms at times they suspected they would only ignore them. However, Table 5 shows that the less frequent types of audible alarm were more likely to lead to corrections (C) than were the three most frequent audible alarm types, $\chi^2_{(1)} = 11.05$, $p < .001$.

For example, the BP alarm and the infusion pump alarm almost always led to corrective action during maintenance. Although the data for infrequent alarms are sparse, some types of alarms tend to appear in specific phases, such as BP during maintenance, when it is usually corrected, and lead failures during emergence, when they are ignored. In cardiac cases in particular, abnormal readings for BP had different meanings under different conditions. During bypass, BP readings might be clearly abnormal, such as 22/34, but did not provoke a reaction from the CRNA because normal BP would not be expected during bypass.

Do responses to alarms differ across induction, maintenance, and emergence phases? The answer is a definite "yes." The rightmost field of Table 3 shows the number of runs of the same audible alarm with the same CRNA response, by phases. During induction most alarms are ignored (N), during maintenance most alarms lead to corrections (C), and during emergence most alarms are again ignored (N). A chi-squared analysis showed a significant dependency between phase and response, suggesting that alarms indicate intended events (E) more often than expected during induction, and corrections (C) less often than expected

during emergence, $\chi^2_{(6)} = 26.9$, $p < .001$. Manual ventilation throughout arthroscopy – and in other procedures after extubation in emergence – led to apnea alarms acting as reminders to the CRNA to ventilate the patient if attention was temporarily drawn elsewhere. The reminders were often interspersed with other activities in cascades of events, as shown in the following examples from Table 2.

1. The sequence R²NR²C²N during emergence for the fifth laparoscopy involved a series of apnea alarms acting as reminders (R) to perform manual ventilation as well as high ETCO₂ (end-tidal carbon dioxide) alarms sounding in the process of getting the patient's spontaneous breathing to return.
2. The sequence CR⁵E³CCER⁶ during maintenance for the third arthroscopy involved apnea alarms acting as reminders to perform manual ventilation and HR intentionally (E) being allowed to fall low before the onset of surgery, which increases HR.
3. The sequence NNN⁵R³C during emergence in the sixth cranial case represented a series of ignored (N) apnea alarms while the patient was being suctioned, followed by apnea alarms acting as reminders to ventilate the patient manually.

Thus apnea alarms, the most common in our data set, can prompt different responses depending partly on the phase of an operation. As Table 4 shows, there was an overall tendency for apnea alarms to be ignored during induction, to act as reminders during maintenance, and either to be ignored or to act as reminders during emergence, $\chi^2_{(6)} = 10.98$, $.1 > p > .05$. Other notable features in Table 4 are the large proportion of corrective responses to blood pressure, heart rate, and infusion pump alarms during maintenance (these come almost exclusively

from cardiac and intracranial cases), and the almost complete absence of BP and HR alarms from the induction and emergence phases in this data set.

Do responses to alarms differ across different kinds of procedures? A qualitative assessment of our data suggests that the answer to this question is sometimes yes, but the data are too sparse for a full quantitative analysis. Table 2 shows that the distribution of responses across phases is similar for the cardiac and intracranial procedures. Both procedures show proportionally more corrections, more ignored alarms, and fewer alarms acting as reminders than for the laparoscopic and arthroscopic procedures. The arthroscopic procedure differs from the other three, with many more alarms acting as reminders and fewer alarms being ignored than for the other three procedures. This is because of the greater use of manual ventilation during the maintenance phase of arthroscopic procedures.

The apnea alarms acted as a reminder to ventilate the patient. Although this finding might appear to be unduly influenced by CNRA E's data, arthroscopies were the only procedures in which manual ventilation rather than intubation was used and in which apnea alarms were used by both CRNAs as reminders to perform manual ventilation. The arthroscopic procedures also had a slightly smaller proportion of alarms requiring the CRNA to take corrective action, reflecting the relatively noninvasive nature of the procedure and the fact that the arthroscopic patients were most likely to be healthy young adults. Thus we see a consistent pattern of alarms within procedure type that can be attributed to characteristics of the procedure.

Our field notes suggest that CRNAs were more likely to silence alarms completely during cardiac procedures than during other procedures because the noise annoyed the cardiac surgeons. In addition, while the patient was on cardiac bypass during part of the maintenance phase, oxygenation and circulation were controlled by the bypass machine. Alarms were irrelevant and thus were suppressed.

Do responses to the same alarm differ under different conditions? Our data suggest that for some alarms at least, the answer is yes. Our statistical analyses suggest that an apnea alarm will

require different responses at different phases of a surgical procedure (see Table 4). It is remarkable that apnea alarms led to so few corrections, given their total number. Our field notes indicate that during induction, an apnea (or low ETCO_2) alarm may result from deliberate hyperventilation, which is an intended state just before intubation. The apnea alarm is normally intended to tell the anesthetist there is a problem with the patient, but in this case it is used as a sign that all is well with the patient and that intubation can start. However, if the apnea alarm occurs during maintenance, a problem would quickly be suspected and corrective measures taken.

During emergence and before extubation, the anesthetist might tolerate ETCO_2 levels a little higher than usual in the belief that they will stimulate the return of the patient's own unassisted breathing. Finally, during extubation or in the event that the patient is moved, the apnea alarm will sound when the patient is purposely disconnected from the breathing apparatus. As another example, anesthetists usually tolerate heart rates and blood pressure levels at the lower alarm level just before surgery starts because the surgery itself is a stimulus that will elevate both HR and BP, even though the patient is anesthetized.

The results reported in this section may be constrained in their generality and scope. First, they are based on the practice of CRNAs rather than medically trained anesthesiologists. Second, the data were gathered in the U.S. at one specific hospital. Third, they reflect the particular alarm settings and thresholds used: Limits set closer to normal values would lead to more alarms being ignored. Fourth, we have not judged the appropriateness or speed of reaction to alarms. Fifth, we have not recorded how often CRNAs silenced unwanted alarms, rather than simply ignoring them. Finally, we have not analyzed responses to visual alarms while auditory alarms were silenced or the volume was turned to zero. Further studies supported by audiovisual recording will correct some of these deficiencies (e.g., Seagull, Xiao, & Mackenzie, 1999).

CONCLUSIONS

Our results demonstrate that anesthetists

respond to alarms in different ways, depending on anesthesia and surgical context. Even when responses to audible alarms only are analyzed, as here, ignoring the alarm is still the most likely response, thus strengthening the results of Kesting et al. (1988) and Bliss et al. (1995). A major subjective source of annoyance for the CRNAs was the inability to filter out irrelevant alarm information in any useful way. On the Ohmeda 7850, alarms could be silenced or suppressed for a fixed interval (e.g., 2 min), or the alarm volume could be turned to zero. However, these measures would affect all alarms, not just a subset of alarms selectively. Also on the Ohmeda, the alarm limits for a parameter whose readings the anesthetist wished to ignore could be set to extreme values that would seldom be reached, such as an HR lower limit of 40 beats per min (bpm). Again, this is just a gross modification of the information environment to handle complex temporary situations. Overall, can such findings and observations be useful in thinking about how to reduce problems with alarms?

In their contexts, many alarms do not require action, such as apnea alarms during intubation. If the technology could reliably sense that intubation was taking place, there would be no need to sound the apnea alarm, but intelligent alarm management of this kind is not yet available. Our data suggest that, at the very least, intelligent alarm management behavior might be conditionalized on phase of procedure, kind of procedure, anesthetic interventions, and surgical events. Various approaches to more intelligent alarm management have been proposed, including rule-based reasoning systems (Loeb, Brunner, Westenskow, Feldman, & Pace, 1989), neural networks (Orr & Westenskow, 1994), and redundancy-based sensing systems (Navabi, Mylrea, & Watt, 1989), but to date no approach has proven reliable enough at independently interpreting contexts and events beyond the low-level physical parameters to become an accepted system feature.

However, studies of human interaction with intelligent decision aids indicate that dissociations often occur between the understandings of machine and human agents of the current context (Suchman, 1987). There will always be exceptions to a preprogrammed plan or pat-

tern-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. The anesthetist might then have to program actual or changed conditions or a change of plan into the device exactly when his or her attention would probably be urgently needed elsewhere, making such systems unsafe.

Other less complex solutions may be possible. Our data show that alarms are ignored most frequently during induction and emergence. In an independent study, Watson and Sanderson (1998) demonstrated the potential impact of managing alarms differently across phases with a simple hand simulation based on 42 hr of observational data. In one run, the rate at which audible alarms arrived was adjusted as if the anesthetist had been able to selectively silence low-priority alarms during the emergence and induction phases of the procedures. Results showed that during induction the audible alarm rate would have dropped to about 30% of the unmanaged rate, during maintenance it would have remained about the same, and during emergence it would have dropped to about 10% of the unmanaged rate.

Using a less drastic simulated intervention, Watson, Sanderson, and Russell (2000) showed that if a continuous auditory display could be used to indicate HR, SpO₂, ET/CO₂ and RR, thus obviating the need for intrusive limit-based alarms, then total auditory alarm soundings would be 52% of the baseline number. In both simulations these results were achieved without loss of important warnings for the cases observed, but, naturally, there is no guarantee this would always be the case.

An alternative approach is to let anesthetists create personal, temporary alerts, thus letting users tailor alarm systems to the current context (Watson et al., 2000). Such systems have been developed for industrial process control, in which case they are termed *user-initiated notification* systems (Guerlain & Bullemer, 1996). For example, an anesthetist may set the normal lower HR limit to 60 bpm but just before incision may tolerate HR as low as 50 bpm for the next 60 s, as long as BP remains acceptable. However, the success of user-initiated notifications in the anesthesia environment

would depend largely on how quickly and easily the anesthetist could set them up, see them, monitor them, and dismiss them.

Our data demonstrate that operational contexts exist that could provide a basis for better alarm management. Clearly, effective responses to unusual conditions must reflect some agent's interpretation of patient configuration, physiological state, and interventions performed. Design approaches to context that are simpler than intelligent alarm management systems (such as highly user-configurable information management tools or inherently less intrusive continuous auditory displays) may represent better human engineering, but they should still reflect the contexts uncovered here. Overall, the problem of anesthesia alarms must be viewed as part of the larger issue of how information should be designed and matched to the real-time cognitive and perceptual needs of the anesthetist.

ACKNOWLEDGMENTS

The authors thank the anesthetists, CRNAs, and patients who kindly consented to take part in this study. We are also grateful to Peter Benda, Michael Mullin, and Marcus Watson for their comments on earlier versions of the manuscript.

REFERENCES

- Blattner, M. M., Sumikawa, D. A., & Greenberg, R. M. (1989). Earcons and icons: Their structure and common design principles. *Human-Computer Interaction*, 4, 11-44.
- Bliss, J. P., Gilson, R. D., & Deaton, J. E. (1995). Human probability matching behavior in response to alarms of varying reliability. *Ergonomics*, 38, 2300-2313.
- Block, F. E. (1988). Evaluation of users' abilities to recognize musical alarm tones. *Journal of Clinical Monitoring*, 8, 285.
- Bogner, M. S. (1994). *Human error in medicine*. Mahwah, NJ: Erlbaum.
- Breznitz, S. (1984). *Cry wolf: The psychology of false alarms*. Mahwah, NJ: Erlbaum.
- Cook, R. L., & Woods, D. D. (1996). Adapting to new technology in the operating room. *Human Factors*, 38, 593-613.
- Edworthy, J., Loxley, S., & Dennis, I. (1991). Improving auditory warning design: Relationship between warning sound parameters and perceived urgency. *Human Factors*, 33, 205-251.
- Edworthy, J., & Stanton, N. (1993). A user-centered approach to the design and evaluation of auditory warning signals: 1. Methodology. *Ergonomics*, 38, 2262-2281.
- Gaver, W. (1993). How in the world do we hear? Explorations in ecological acoustics. *Ecological Psychology*, 5, 283-313.
- Guérlain, S., & Bullemer, P. (1996). User-initiated notification: A concept for aiding the monitoring of process control operators. In *Proceedings of the Human Factors and Ergonomics Society 40th Annual Meeting* (pp. 283-287). Santa Monica, CA: Human Factors and Ergonomics Society.
- Kesting, I. G., Miller, B. T., & Lockhart, C. H. (1988). Auditory alarms during anesthesia monitoring. *Anesthesiology*, 69, 106-107.
- Loeb, G. J., Brunner, J. X., Westenskow, D. R., Feldman, B., & Pace, N. L. (1989). The Utah anesthesia workstation. *Anesthesiology*, 70, 999-1007.
- Loeb, R. G., Jones, B. R., Leonard, R. A., & Behrman, K. (1992). Recognition accuracy of current operating room alarms. *Anesthesia and Analgesia*, 74, 499-505.
- Moll van Charante, E., Cook, R. L., Woods, D. D., Yue, L., & Howie, M. B. (1993). Human-computer interaction in context: Physician interaction with automatic intravenous controllers in the heart room. In H. G. Stassen (Ed.), *Analysis, design, and evaluation of man-machine systems, 1992* (pp. 263-274). New York: Pergamon.
- Momtahan, K. L., Hetu, R., & Tansley, B. W. (1993). Audibility and identification of auditory alarms in operating rooms and an intensive care unit. *Ergonomics*, 36, 1159-1176.
- Navabi, M. J., Mylrea, K. C., & Watt, R. C. (1989). Detection of false alarms using an integrated anesthesia monitor. In *IEEE Engineering Medical Biological Society 11th Annual International Conference*, (pp. 1774-1775). Los Alamitos, CA: IEEE Computer Society Press.
- Orr, J. A., & Westenskow, D. R. (1994). A breathing circuit alarm system based on neural networks. *Journal of Clinical Monitoring*, 10, 101-109.
- Patterson, R. D. (1989). Guidelines for the design of auditory warning sounds. In *Proceedings of the Institute of Acoustics*, 11, 17-25. Institute of Acoustics: Edinburgh, U.K.
- Patterson, R. D., Edworthy, J., Shailer, M. J., Lower, M. C., & Wheeler, P. D. (1986). *Alarm sounds for medical equipment in intensive care areas and operating theatres* (Report No. AC598). Southampton, England: Institute of Sound and Vibration Research.
- Schreiber, P. I., & Schreiber, J. (1989). Structured alarm systems for the operating room. *Journal of Clinical Monitoring*, 5, 201-204.
- Seagull, J., Xiao, Y., & Mackenzie, C. F. (1999). Monitoring behavior: A pilot study using an ambulatory eye-tracker in surgical operating rooms. In *Proceedings of the Human Factors and Ergonomics Society 43rd Annual Meeting* (pp. 850-854). Santa Monica, CA: Human Factors and Ergonomics Society.
- Stanton, N. A. (1994). Alarm initiated activities. In N. A. Stanton (Ed.), *Human factors in alarm design* (pp. 93-117). London: Taylor & Francis.
- Suchman, L. (1987). *Plans and situated actions*. Cambridge, England: Cambridge University Press.
- Watson, M., & Sanderson, P. (1998). Work domain analysis for the evaluation of human interaction with anesthesia alarm systems. In *Proceedings of the 1998 Australasian Computer Human Interaction Conference* (pp. 228-235). Los Alamitos, CA: IEEE Computer Society.
- Watson, M., Sanderson, P., & Russell, I. W. (2000). Alarm noise and end-user tailoring: The case for continuous auditory displays. In *Proceedings of the Fifth Annual Symposium on Human Interaction with Complex Systems* (pp. 75-80). Urbana, IL: Beckman Institute.
- Weinger, M. B. (1991). Proposed new alarm standards may make a bad situation worse. *Anesthesiology*, 74, 791-792.
- Woods, D. D. (1995). The alarm problem and directed attention in dynamic fault management. *Ergonomics*, 38, 2371-2394.

F. Jacob Seagull received his M.S. in engineering psychology from the Technion-Israel Institute of Technology. He is currently completing a Ph.D. in the Department of Psychology at University of Illinois at Urbana-Champaign, Urbana, Illinois.

Penelope M. Sanderson is a professor of computer science in the School of Information Technology, Swinburne University of Technology, Melbourne, Australia, where she is also director of the Swinburne Computer-Human Interaction Laboratory. She received a Ph.D. from the University of Toronto in 1985.

Date received: June 9, 1998

Date accepted: July 14, 2000