Advanced Auditory Displays and Head-Mounted Displays: Advantages and Disadvantages for Monitoring by the Distracted Anesthesiologist

Penelope M. Sanderson, PhD, FASSA*

Marcus O. Watson, PhD†

Walter John Russell, MBBS, DIC, FANZCA, FRCA‡

Simon Jenkins, MBBS, FANZCA§

David Liu, BEng(Hons)‖

Norris Green, MBBS, FANZCA¶

Kristen Llewelyn, MBBS#

Phil Cole, BEng, BSc**

Vivian Shek, BIT(Hons), BPsy‖

Stas S. Krupenia, PhD††

BACKGROUND: In a full-scale anesthesia simulator study we examined the relative effectiveness of advanced auditory displays for respiratory and blood pressure monitoring and of head-mounted displays (HMDs) as supplements to standard intraoperative monitoring.
METHODS: Participants were 16 residents and attendings. While performing a reading-based distractor task, participants supervised the activities of a resident (an actor) who they were told was junior to them. If participants detected an event that could eventually harm the simulated patient, they told the resident, pressed a button on the computer screen, and/or informed a nearby experimenter. Participants completed four 22-min anesthesia scenarios. Displays were presented in a counterbalanced order that varied across participants and included: (1) Visual (visual monitor with variable-tone pulse oximetry), (2) HMD (Visual plus HMD), (3) Audio (Visual plus auditory displays for respiratory rate, tidal volume, end-tidal CO_2, and noninvasive arterial blood pressure), and (4) Both (Visual plus HMD plus Audio).
RESULTS: Participants detected significantly more events with Audio (mean = 90%, median = 100%, $P < 0.02$) and Both (mean = 92%, median = 100%, $P < 0.05$) but not with HMD (mean = 75%, median = 67%, *ns*) compared with the Visual condition (mean = 52%, median = 50%). For events detected, there was no difference in detection times across display conditions. Participants self-rated monitoring as easier in the HMD, Audio and Both conditions and their responding as faster in the HMD and Both conditions than in the Visual condition.
CONCLUSIONS: Advanced auditory displays help the distracted anesthesiologist maintain peripheral awareness of a simulated patient's status, whereas a HMD does not significantly improve performance. Further studies should test these findings in other intraoperative contexts.
(Anesth Analg 2008;106:1787–97)

The Anesthesia Patient Safety Foundation's alarm summit in 2004[1] emphasized that the variable-pitch pulse oximeter tone and the capnography auditory alarm must be on and audible if they are to prevent patient incidents. Recognizing the unique alerting ability of the auditory modality, several researchers have suggested that the advantages of the continuous variable-tone pulse oximetry signal could be extended to other vital signs.[2–7] Further researchers have noted that lightweight head-mounted displays (HMDs) of vital signs could also keep anesthesiologists continuously informed about the patient's state.[8–13] In the simulator study to be reported, we investigated advantages and disadvantages of advanced auditory displays versus HMDs for helping anesthesiologists maintain awareness of patient status and detect potentially adverse events when distracted from normal visual scanning.

This article has supplementary material on the Web site: www.anesthesia-analgesia.org.

From the *ARC Key Centre for Human Factors and School of Medicine, The University of Queensland, St. Lucia, Queensland; †School of Medicine, The University of Queensland, St Lucia, Queensland, Australia; ‡Department of Anaesthesia, Pain Medicine and Hyperbaric Medicine, Royal Adelaide Hospital, South Australia, and University of Adelaide; §Royal Adelaide Hospital and University of Adelaide, Adelaide, Australia; ‖School of Information Technology and Electrical Engineering, The University of Queensland, St Lucia, Queensland, Australia; ¶Department of Anaesthesia, Princess Alexandra Hospital, Brisbane, Australia and The University of Queensland; #Royal Adelaide Hospital, Adelaide, Australia; **Cognitive Engineering Research Group, The University of Queensland, St Lucia, Australia; and ††School of Psychology, The University of Queensland, St Lucia, Australia.

Accepted for publication February 20, 2008.

Supported by Australian Research Council Discovery Project grant ARC DP0559504 to P. Sanderson, M. Watson, and W. J. Russell, and by University of Adelaide to D. Liu.

Department and Institution to which the work should be attributed School of ITEE, School of Psychology, and School of Medicine, The University of Queensland.

The respiratory sonification used in this study is the subject of US Patent 7070570 (inventors Sanderson and Watson). The blood pressure earcons are the subject of PCT/AU2003/001622 (inventor Watson).

Reprints will not be available from the author.

Address correspondence to Penelope Sanderson, ARC Key Centre for Human Factors The University of Queensland, St Lucia, QLD 4072, Australia. Address e-mail to psanderson@itee.uq.edu.au.

DOI: 10.1213/ane.0b013e31817325cb

Whether a display is visual or auditory may have a strong impact on monitor effectiveness.[14] Most research on patient monitoring ignores the potential for auditory displays to inform rather than to alert or alarm. Auditory displays can reduce competition for the anesthesiologist's visual attention, letting him or her monitor patient vital signs in the background. Continuous sound may move into focal awareness if it signals an unexpected state but recede into peripheral awareness if it signals an expected state.[15,16]

Several research groups have used desktop simulators to investigate patient monitoring with continuous auditory displays (sonification) of heart rate (HR), oxygenation (SpO_2), blood pressure (BP), respiratory rate (RR), tidal volume (V_T), and end-tidal carbon dioxide ($ETCO_2$).[4–6] Results indicate that participants' ability to report patient vital signs is either more accurate[4,5] or equally accurate[6] with numerical visual displays than with sonification alone, but participants can perform time-shared tasks more effectively when monitoring with sonification.[5,6]

In the above research, BP was either mapped to continuous auditory displays[4,5] or was not used.[6] A continuous mapping of noninvasive BP (NIBP) to sound would be misleading because NIBP readings quickly become outdated.[17] In an initial test with nonmedical participants, discrete auditory displays (earcons) signaling systolic and diastolic BP after each NIBP reading led to high accuracy, with the modal error size only one point on a nine-point scale.[18]

Surprisingly, no study has reported whether sonification or earcons draw attention to the start of potentially adverse events sooner than visual displays do, although this happens with auditory alarms.[19] Achieving early but nonintrusive awareness of changes in patient state is a key goal of using sonification rather than auditory alarms. Such results have been reported for HMDs, which keep patient variables in the forward field of view and so reduce the workload of visual scanning. Prior simulator-based research shows that a HMD with a transparent monocular eyepiece reduced the average time anesthesiologists took to detect each of a series of four to six critical events embedded in an anesthesia scenario[20] and was well-accepted.[21] Anesthesiologists using a HMD while handling major incidents switched gaze between visual monitors and a patient less often, spent more time looking towards the patient, and completed small procedures faster.[8–11,17] However, concerns that HMDs may cause unexpected events not to be seen have been noted in other domains and may carry over to the operative environment.[14,22]

Current levels of patient safety in anesthesia are due to the multiple, partially redundant means that anesthesiologists use to maintain awareness of patient physiologic status, such as visual monitors, clinical observation, auditory and visual alarms, equipment checks and the like.[23] Demonstrating prospectively that any one monitor improves patient outcomes is difficult in both large-scale clinical trials[24,25] and in smaller simulator studies. In his well-known Swiss cheese model of accident causation, Reason[26] argues that adverse outcomes usually result from the rare alignment of "holes" in multiple layers of defense. Accordingly, we will test how well each display prevents unsafe situations when combined continuously with a risk factor. The conditions under which we test are therefore deliberately unrepresentative of normative or typical monitoring practices.

The risk factor we manipulate is demand for the anesthesiologist's attention that makes it difficult to scan visual monitors. Anesthesiologists occasionally experience strong demands on visual attention, as when inserting lines, performing blocks, or reading patient notes. To achieve standardized and reliable measurements to compare performance across displays, in our experiment we maintain the challenge to visual attention consistently within and across scenarios. We do not test the displays with undistracted anesthesiologists as a control condition because anesthesiologists' usual defensive monitoring strategies would be active and the unique contribution of the displays we wish to test could not be evaluated.

Based on the research reviewed above, we conjecture that further support for monitoring in the form of advanced auditory displays or HMDs may improve the distracted anesthesiologist's vigilance. In the following study, first we hypothesized that distracted anesthesiologists will detect a greater number of clinically significant events when visual monitors and variable-tone pulse oximetry are supplemented by further auditory displays and/or a HMD. Second, we hypothesized that distracted anesthesiologists will detect clinically significant events faster when monitoring is supplemented by auditory displays and/or a HMD.

METHODS

Participants

The study received Human Research Ethics Committee clearances from The University of Queensland and from Royal Adelaide Hospital. Participants were 16 anesthesiologists (seven consultants and nine residents) recruited from Royal Adelaide Hospital (14 males and two females). Participants had between 3.5 and 36 yrs of anesthesia experience after basic medical training (average 12 yrs). Participation was voluntary and participants were rewarded with a gift worth around AUD$50 that they chose from a small selection.

Design

Each participant experienced all four display conditions in a repeated measures design:

- Visual: Visual monitor plus variable-tone pulse oximetry, with visual and auditory alarms suppressed (condition serves as baseline for comparisons).

Table 1. Counterbalancing Table Used for Experiment

	Position in Sequence for Experiment					Position in Sequence for Experiment			
Order	1st	2nd	3rd	4th	Order	1st	2nd	3rd	4th
1	Vis-A	HMD-B	Aud-D	Both-C	9	Vis-C	HMD-D	Aud-B	Both-A
2	HMD-B	Both-C	Vis-A	Aud-D	10	HMD-D	Both-A	Vis-C	Aud-B
3	Both-C	Aud-D	HMD-B	Vis-A	11	Both-A	Aud-B	HMD-D	Vis-C
4	Aud-D	Vis-A	Both-C	HMD-B	12	Aud-B	Vis-C	Both-A	HMD-D
5	Vis-B	HMC-C	Aud-A	Both-D	13	Vis-D	HMD-A	Aud-C	Both-B
6	HMD-C	Both-D	Vis-B	Aud-A	14	HMD-A	Both-B	Vis-D	Aud-C
7	Both-D	Aud-A	HMD-C	Vis-B	15	Both-A	Aud-C	HMD-A	Vis-D
8	Aud-A	Vis-B	Both-D	HMD-C	16	Aud-C	Vis-D	Both-B	HMD-A

A, B, C, D refer to Scenarios A, B, C, and D which are detailed in Table 2.

Vis = visual condition; HMD = HMD condition; Aud = audio condition; Both = both condition.

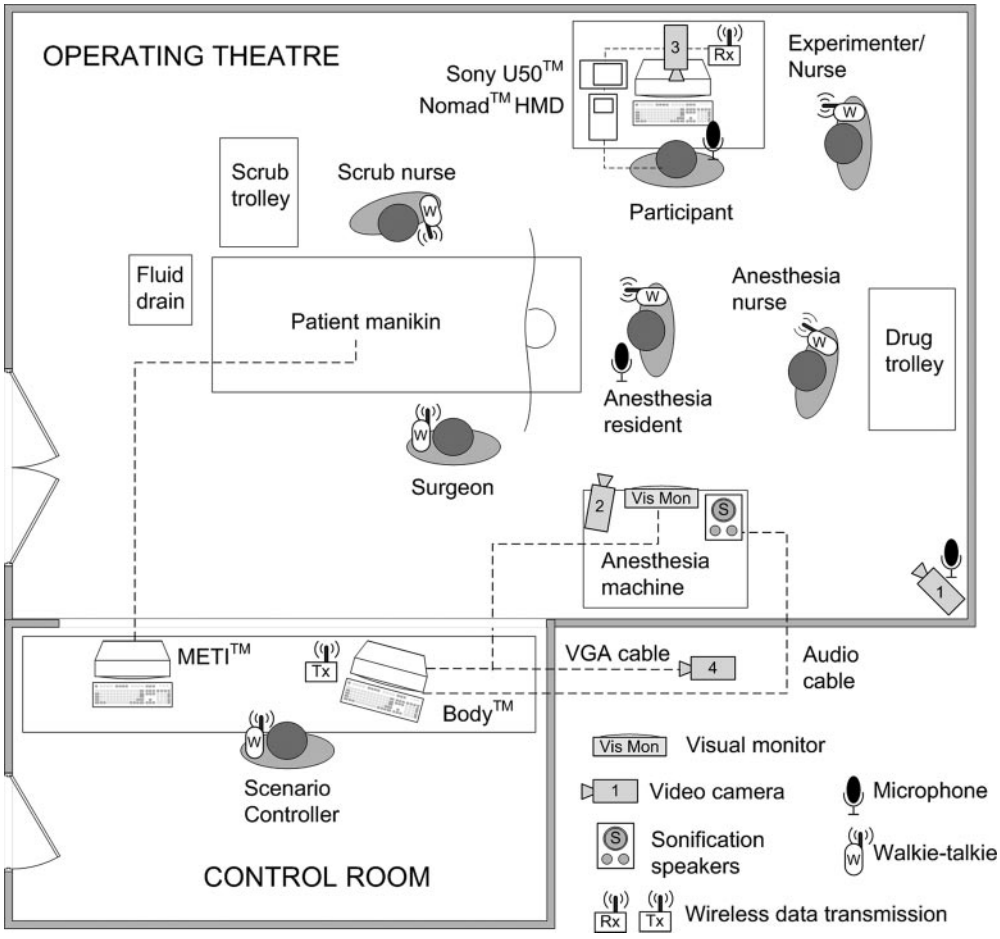

Figure 1. Layout of the operating room simulator, showing representative locations of equipment and actors.

- HMD: As for Visual, plus HMD.
- Audio: As for Visual, plus respiratory sonification and BP earcons.
- Both: As for Visual, plus HMD, respiratory sonification, and BP earcons.

The order of display conditions was counterbalanced using a Latin Squares design. Scenarios were mapped to create 16 display orders so that each display condition was tested with all four scenarios (Table 1). As a result, across the experiment, participants' performance for each display condition was observed in Scenarios A, B, C, and D, and in each of the four serial positions, so there was no confounding of display condition with scenario or with serial position.

Apparatus and Stimuli

Simulator and Software

The study was conducted using a full-scale simulator at Royal Adelaide Hospital. The participant sat at a computer on which a distractor task was presented with his or her back to the resident (Fig. 1). The resident was played by one of two actors who were practicing anesthesiologists. The surgeon, anesthetic nurse, scrub nurse, and scout nurse were

all played by actors trained for these roles. The scout nurse was introduced as a support person for the participant. She stood near the participant to aid recording of patient events, to answer any questions, and to take sound level readings.

To achieve a realistic operating room environment with tight control over the timing of scenario events, a METI ECS™ simulator was used alongside the Body 2003™ dynamic link library simulator. The patient's vital signs were initially generated using the Body™ simulator. Custom software manipulated the Body™ outputs to meet the scenario profiles and the result was displayed on a custom-designed patient monitor. Finally, matching scenarios were developed on the METI™ to complement the above-mentioned scenarios with physical changes on the manikin.[27] The simulated patient's vital sign information was broadcast over a TCP/IP protocol to a METI Waveform Display™ style standard visual patient monitor; auditory displays for variable-tone pulse oximetry, respiratory sonification, and BP earcons (Csound 5 scripts); the HMD, driven by a Sony U50™ palmtop PD; and the abstract classification distractor task.

Scenarios

Five 22-min scenarios were developed: one for training and four to use in the experiment itself (Table 2). Each scenario included three principal events (1, 2a, and 3a) which ran for between 1:08 and 2:17 (min:s) and which were usually first detectable from respiratory vital signs. Each event started when first indications of it appeared in the monitoring displays (Table 2) and ended either when the resident resolved the event by taking action or the end of the scenario was reached. In each scenario, the registrar made an unsuccessful attempt at resolution resulting in a further event, 2b or 3b.

Scenarios were completely deterministic and were closely controlled from the simulator control room (Fig. 1). Actors wore lightweight walkie-talkies with earpieces and their interventions with the manikin and simulated operating room equipment were cued by a scenario controller. Only the resident made interventions to the simulated patient, and did so strictly according to the scenario script. No participant indicated awareness that the scenarios were deterministic and noninteractive. Visual and auditory alarms were suppressed on all equipment so that the relative effectiveness of the auditory displays and the HMD could be fully assessed.

Displays

The respiratory sonification combines information about RR, I:E ratio, V_T, and $ETCO_2$ into one sound stream.[6,7] Flow of gas is represented by relatively pure rather than breath-like tones. As the patient's chest rises with inspiration and V_T accumulates, the sound pressure level (SPL) of the sonification increases to a maximum (at V_T of 500–600 mL, maximum SPL is usually around 55 dB(A)). As the patient's chest falls with an expiration, SPL decreases to zero. The inspiration sound is mapped to a musical pitch representing the most recent $ETCO_2$ level (when normal, between 349 and 358 Hz, or around F4) and the expiration to a pitch a musical minor third interval below the inspiration pitch (at approximately 0.841 of the first pitch, around D4 for normal). As $ETCO_2$ increases or decreases, the inspiration and expiration pitch together increase or decrease (maximum inspiration was 566 Hz for $ETCO_2$ of 69 mm Hg and minimum inspiration was 166 Hz for $ETCO_2$ of 0 mm Hg).

The NIBP earcons sound just after a NIBP reading has been taken and are scaled in pitch and duration to represent nine levels of systolic and diastolic BP.[18,28] Two initial tones (at 503 Hz, just above B4) provide a normal reference level (beacon) for systolic and diastolic BP. Two subsequent tones indicate systolic and diastolic BP by playing at one of four higher pitches (905 Hz and higher) and slightly shorter duration if hypertensive and at one of four lower pitches (304 Hz and lower) and slightly longer duration if hypotensive (Video clips 1 and 2; please see video clips available at http://www.anesthesia-analgesia.org).

The HMD was a Microvision Nomad™ with a transparent monocle. In the lower right part of the monocle there was a display of HR, SpO_2, BP, RR, V_T, and $ETCO_2$. To enhance equivalence with the auditory information for this experiment, no waveforms were provided. Figure 2 approximates what was seen through the HMD from the participant's location.

Distractor Task

To test display effectiveness when the anesthesiologist is distracted, we developed a distractor task (Fig. 3). Abstracts of around 200 words from *Anesthesia & Analgesia* were displayed on a laptop screen. Participants classified each abstract according to (1) the kind of anesthesia application it covered, (2) the evidence class of the research, (3) the 2-year period of publication of the paper, and (4) the likely impact of the paper on anesthesia practice. Participants had up to 40 s to read each abstract and to enter their four classifications by clicking the corresponding buttons on a computer screen, but within the 40 s the task was self-paced. At 35 s, a bell rang to remind the participant to complete the classification task before the next abstract appeared.

Auditory distraction was not formally controlled, but consisted of normal conversation between operating room actors on a set of topics unrelated to the patient. In addition, certain surgical sounds such as diathermy were emulated. Visual, auditory, and cognitive distraction were therefore present throughout each scenario.

Questionnaires

A background questionnaire asked participants about their medical training, anesthesia experience, and musical training. Postscenario questionnaires asked participants how easy it was to monitor the

Table 2. Scenarios Used. Each Scenario Lasts for 22 Minutes

Scenario	Procedure	Event 1	Event 2a	Event 2b	Event 3a	Event 3b
Training	Standard induction scenario (GA and ETT)	Intubation Starts 01:30 W of Opp 1:35 RR 10 ↓ 0 V_T 700 ↓ 0 $ETCO_2$ 40 ↓ 0 BP 112/66 ⇗ 121/76 HR 87 ↘ 77 SpO_2 99	Bleeding (surgeon brings under control) Starts 16:15 W of Opp 4:45 RR 12 V_T 600 $ETCO_2$ 40 ↓ 27 BP 119/73 ⇘ 114/63 ⇘ 109/74 ⇘ 97/70 HR 70 ↑ 119 ↘ 73 SpO_2 99	—	—	—
Scenario A	Laparoscopic cholecystectomy (GA and ETT)	Esophageal intubation (AR reintubates patient) Starts 06:28 W of Opp 1:42 RR 0 ↑ 13 V_T 0 ↑ 500 (ins ↑ but exp ↓) $ETCO_2$ 0 ↑ 5 ↘ 0 BP 165/106 ↓ 141/80 HR 73 ↘ 61 SpO_2 99	CO_2 line obstruction via AR foot (comes good but AR assumes loose filter) Starts 13:52 W of Opp 2:09 RR 12v V_T 650 $ETCO_2$ 44 ↓ 0 BP 116/71 = 117/72 HR 74 SpO_2 98	CO_2 line obstructed by AR foot again (AR removes foot from line) Starts 16:35 W of Opp 1:16	Ventilator overpressure (AR turns down ventilator) Starts 17:51 W of Opp 1:35 RR 13 V_T 600 ↑ 1250 ↑ 1300 $ETCO_2$ 41 ↘ 37 BP 116/73 = 117/72 HR 75 SpO_2 97	—
Scenario B	Removal of skin lesion on nose (GA and LMA)	Surgical stimulus (AR gives fentanyl) Starts 07:29 W of Opp 1:44 RR 13 ↗ 19 V_T 600 ↗ 1000 $ETCO_2$ 44 ↘ 40 BP 144/82 ⇗ 182/122 HR 67 ↗ 111 SpO_2 99	Drug overdose (AR assists ventilation) Starts 11:57 W of Opp 1:45 RR 11 ↘ 5 ↓ 0 V_T 500 ↘ 200 ↓ 0 $ETCO_2$ 32 ↘ 27 ↓ 0 BP 124/83 ⇘ 123/76 ⇘ 117/70 HR 84 ↘ 79 SpO_2 99	—	Y-piece disconnect (AR reconnects it) Starts 16:48 W of Opp 1:50 RR 13 V_T 600 ↓ 0 $ETCO_2$ 47 ↓ 0 BP 116/69 = 117/68 = 119/69 HR 74 SpO_2 99 ↘ 98	Y-piece leaking (AR increases V_T) Starts 18:39 W of Opp 3:06
Scenario C	Laparoscopic cholecystectomy (GA and ETT)	Ventilator not switched on (AR switches it on) Starts approx 04:30 W of Opp 1:59 RR stays at 0 V_T stays at 0 $ETCO_2$ stays at 0 BP 141/93 ⇘ 134/81 ⇗ 147/86 HR ↘ (80 to 75) ↘ 63 SpO_2 99	Pt breathing against ventilator (AR gives muscle relaxant and turns up volatile) Starts 10:31 W of Opp 1:29 RR 15 ↓ (10 → 19) V_T 600 ↓ (300 → 800) $ETCO_2$ 41 ↓ (44 → 26) BP 123/71 = 119/73 HR 70 ↗ 72 SpO_2 99	—	Gas embolism (AR treats as fluid loss) Starts 18:39 W of Opp 1:08 RR 15 V_T 600 $ETCO_2$ 38 ↘ 23 BP 123/72 ⇘ 74/26 HR 72 ↗ 76 SpO_2 99 ↓ – (SpO_2 unrecordable)	Gas embolism (management by AR runs to end of scenario) Starts 18:39 W of Opp 1:55
Scenario D	Ankle repair (GA with LMA)	Airway obstruction (AR repositions LMA) Starts 04:48 W of Opp 1:52 RR 12 V_T 600 ↓ (300 → 400) $ETCO_2$ 43 ↓ 30 BP 130/79 ⇗ 148/86 ⇗ 154/91 HR 70 ↘ 64 SpO_2 99	Possible excessive gas (AR turns down sevoflurane) Starts 08:38 W of Opp 2:17 RR 13 ↘ 10 V_T 550 ↓ (300 → 500) $ETCO_2$ 45 ↘ 31 BP 130/79 ⇘ 123/71 HR 70 ↗ 95 SpO_2 99	Still excessive gas (AR turns down sevoflurane more) Starts 11:30 W of Opp 1:25	CO_2 line disconnection (AR reconnects it) Starts 18:48 W of Opp 2:13 RR 12 V_T 500 $ETCO_2$ 41 ↓ 0 BP 129/79 = 127/77 = 127/77 HR 74 ↘72 SpO_2 99	—

Times in each cell are the starting time of the event and event duration or maximum time (mins:secs) allowed for detection before the resident resolved the event.

GA = general anesthesia; ETT = endotracheal tube; LMA = laryngeal mask airway; AR = resident; W of Opp = window of opportunity to respond.

↑ ↓ abrupt increase or decrease within a 10 s period, ↗ ↘ gradual increase or decrease lasting more than 10 seconds, ⇗ ⇘ abrupt increase or decrease after a delay of more than 10 sec, (x → y) fluctuating between values x and y. BP readings are taken from NIBP cuff every minute.

patient in the condition they had just experienced, how quickly they believed they detected abnormal changes and returns to normality, how easy it was to do the distractor task, how much the distractor task interfered with patient monitoring and *vice versa*.

A postexperiment questionnaire asked participants to compare their experience of the four display conditions, ranking ease of monitoring with the four displays, preference for monitoring, and satisfaction with amount of training received for the HMD, the respiratory sonification, and the BP earcons. Participants also provided open-ended comments about what they liked and disliked about the HMD, the respiratory sonification, and the BP earcons. Further personality questionnaires were administered, whose results are not reported here.

Procedure

The experiment ran for around 4 h and had four phases: consent and orientation; training; scenarios; and final questionnaires.

Consent and Orientation

After reading an information sheet about the goals of the study, participants gave written in-

Figure 2. Representation of the information displayed on the Microvision Nomad™ HMD monocle, taken from the perspective of the participant. Information appeared as luminous red characters in the right visual field.

formed consent to participate. They completed the background questionnaire. Then they were oriented to the simulated operating room and the functioning of its equipment for the purposes of the study.

Training

In a briefing room, the experimenter trained each participant on the tasks and displays they would experience. First, a PowerPoint™ presentation introduced the respiratory sonification, the BP earcons, and the information that would be seen on the HMD. The mapping of sound to vital signs and the range of values used were explained. The above introduction lasted around 20 min. Second, participants watched and listened to nine 2-min video clips of the visual patient monitor during patient monitoring, supplemented by variable-tone pulse oximetry, respiratory sonification, and BP earcons. The first two clips showed stable vital signs during spontaneous and mechanical ventilation. The remaining seven clips showed changes in vital signs associated with commonly recognized anesthesia incidents. Third, participants learned about the abstract classification task software and practiced classifying abstracts with the software for 5 min.

The experimenter and participant then moved to the simulated operating room for a 22-min training

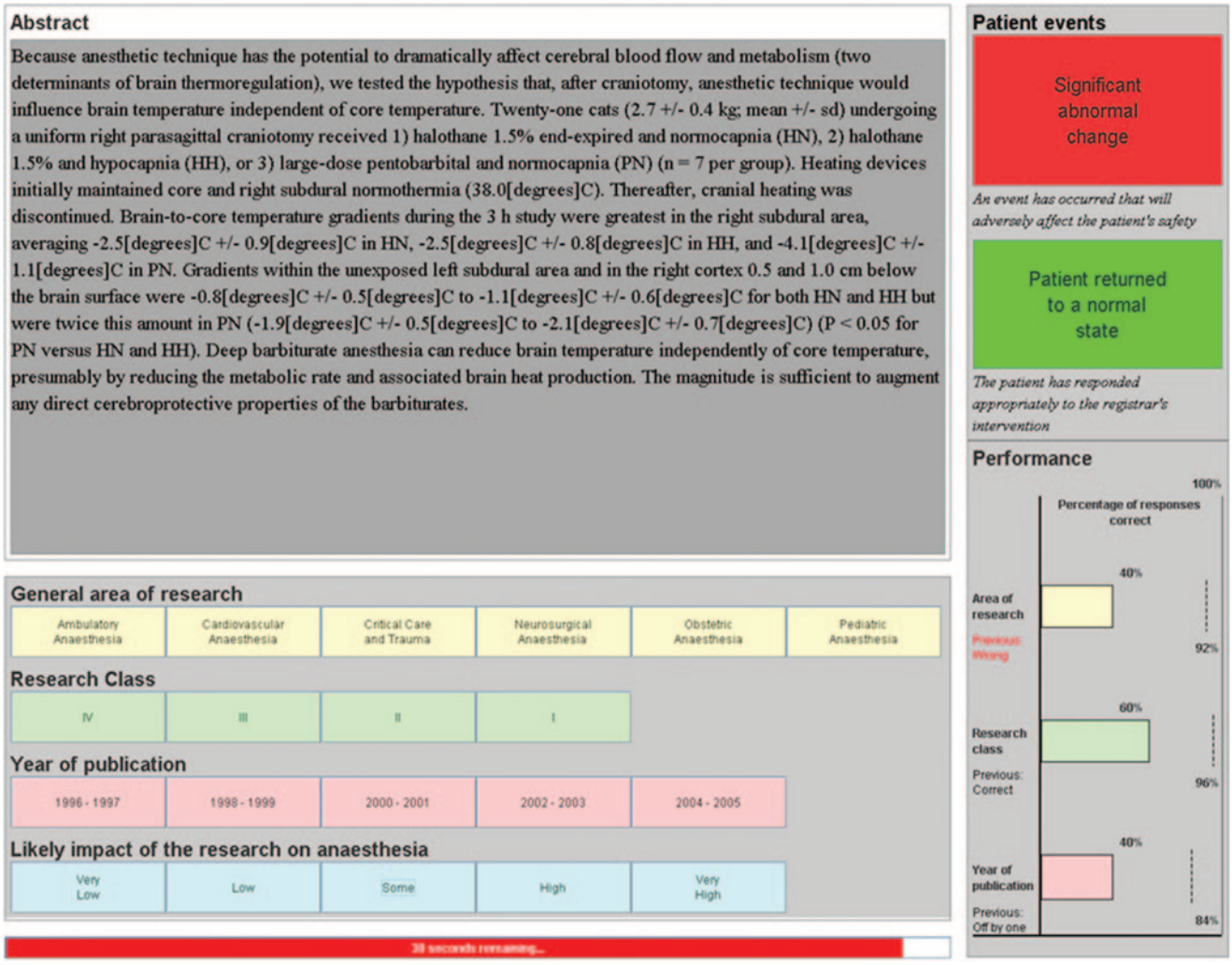

Figure 3. Abstract classification distractor task interface. Abstract is top left, the four classification category responses buttons are bottom left, performance feedback to participant is bottom right, and patient response buttons are top right.

scenario (Table 2) using the respiratory sonification, BP earcons, and HMD in an approximation to the Both condition. The participants sat at the location indicated in Figure 1 with the anesthesia machine and visual monitor behind them, and performed the abstract classification task. Because there were no actors in the simulated operating room during training, the experimenter read out the events as they occurred to provide context for the vital signs. Half-way through the training scenario the experimenter removed the HMD so that the participant could practice monitoring without the vital signs in forward view.

Scenarios

Participants then completed the four 22-min scenarios. In each scenario, participants used a different display configuration. Instructions were as follows. Participants were told they were supervising a resident while the resident conducted anesthesia. Because the resident might be slow to respond to problems, the participant was asked to note if there was any clinically significant event, but to remain seated. The participant could turn to look at the patient monitor or make a comment or direct the resident any time they wished. However, they were asked not to neglect the abstract classification task and to give each abstract a considered response. The participant could note a clinically significant event or a return to a normal clinical state either by (1) clicking the relevant button on the computer screen, (2) noting it to the resident, or (3) noting it to the experimenter/scout nurse standing nearby.

During each scenario the actors' activities were directed by the controller over walkie-talkies. The actors conversed on a set of standard topics, but comments about the patient were confined to scripted comments cued by the controller and delivered at the same time in each scenario. The experimenter/scout nurse did not participate in general conversation. After each scenario, the participant and experimenter returned to the briefing room where the postscenario questionnaire was completed.

Final Questionnaires

After the final postscenario questionnaire, participants completed the postexperiment questionnaire and further inventories.

Statistics

Timestamped responses by participants to the abstract classification task and their identification of significant changes in patient state were captured by the abstract classification task software and stored in databases. Video data were captured in quad video display format (Fig. 4) onto DVD or hard drive and VHS tape. Digitized video data were analyzed using MacSHAPA[29] to determine event detections and detection times.

Detection and questionnaire responses were tested for significance using the Friedman Rank Test for k correlated samples, followed by tests of the HMD, Audio, and Both conditions against the Visual condition with Wilcoxon Signed Ranks tests with $\alpha = 0.05$, using a Bonferroni correction for multiple tests ($P = 0.0167$). Detection speed and distractor task performance were tested for significance with Statistica™ 7 using repeated-measures ANOVAs followed by two-way Tukey HSD tests with $\alpha = 0.05$. Detection and speed of detection for the unexpected resolution on either the second (2b) or third (3b) event (Table 2) were not included because participants often handled the unexpected resolution as a continuation of the previous event, creating a dependency and making it impossible to score separately. To test the effect of display on event detections, the proportion of the three major patient events detected in each scenario was calculated for each scenario and each participant. For detection times, two

Figure 4. Sample of video data captured in quad format. Within each view, top left is view of resident's activities from video camera 1 in Figure 1, top right is view from anesthesia machine to participant from video camera 2 in Figure 1, bottom left is view of participant from video camera 3 in Figure 1, and bottom right is a capture of the video monitor display at video source 4 in Figure 1. Upper figure: participant concentrates on the abstract classification task. Lower figure: participant checks activity of resident after hearing indications of a gas embolism event.

measures were used. First, the average detection time for the three major patient events in each scenario was calculated for each scenario and each participant, using data from detected events only (detection time for events detected). Second, the maximum event duration was added for undetected events to yield the minimum time in which a detection might have occurred (estimated minimum detection time).

For the distractor task, correct answers were determined as follows. (1) Abstract classification followed *Anesthesia & Analgesia*'s classification for the six selected categories. (2) Research evidence class was rated by two anesthesiologists prior to the study. Abstracts were used in the study only if there was agreement or near agreement between the two raters on research evidence class. Participant responses were scored as correct if they matched the response of at least one of the raters. (3) Year of publication was determined from journal records and participants were given half points if their responses were in the neighboring period to the correct one. (4) Likely impact was considered too subjective to score, but was used to keep participants distracted.

Ambient Noise Levels

For eight of the 16 participants, readings of ambient SPL were taken before the onset of the first event and after the onset of each event. Readings were taken near the participant by the experimenter/nurse in the corner of the room (Fig. 1). Readings were taken with a hand-held SPL meter set to the A-weighted scale on the slow setting.

RESULTS

Event Detection

Participants' detection of the three principal events of each scenario was analyzed. Preliminary analyses determined that there were no main effects or interactions of participant expertise or musical training on detections or detection speed.

Proportion of Detections

Compared with their performance in the Visual condition ($\bar{X}$ = 52%, Md = 50%, event detection) participants detected significantly more events in the Audio condition ($\bar{X}$ = 90%, Md = 100%, $P = 0.004$) and in the Both condition ($\bar{X}$ = 92%, Md = 100%, $P = 0.011$) but, under the Bonferroni correction, not in the HMD condition ($\bar{X}$ = 75%, Md = 67%, $P = 0.036$). The direction of results was consistent across scenarios A, B, C, and D, with the HMD condition always better than the Visual condition, and the Both and Audio conditions always better than or equal to the HMD condition, and thereby always better than the Visual condition.

Detection Speed

In the first analysis, only the times for events that the participant detected were used (Table 3). No condition was significantly faster than the Visual condition, although the overall ANOVA shows a trend consistent with event detections, $P = 0.09$. In the second analysis, if a participant did not detect an event then the time until the resident resolved the event was used. Compared with performance in the Visual condition (Table 3), the estimated minimum detection time for all events was faster in the Audio condition ($P = 0.003$) and in the Both condition ($P = 0.0005$) but not in the HMD condition ($P = 0.14$).

Distractor Task Performance

The number of abstracts completed and the rate of correct answers are given in Table 3. Accuracy was greater than chance in all display conditions for abstract classification and research evidence class but not for year of publication. Performance did not differ significantly across display conditions for number of completed abstracts, accuracy of categorizing each abstract, or accuracy of estimating year of publication. For research evidence class classification, however, the Tukey HSD test indicated that performance was worse in the Both condition than in the Visual condition ($P = 0.031$).

Questionnaire Findings

Self-Report of Performance

Compared with their ratings for the Visual condition, participants rated it easier to monitor in the HMD condition ($P = 0.002$), the Audio condition ($P = 0.004$) and the Both condition ($P = 0.003$) (Table 3 for means). Participants believed they responded faster to patient events in the HMD condition ($P = 0.011$) and in the Both condition ($P = 0.0002$) than in the Visual condition. However, under the Bonferroni correction, their ratings for their response speed to patient events in the Audio condition just failed to be significantly higher than for the Visual condition ($P = 0.021$). Participants believed they detected the patient returning to normal faster than in the Visual condition only for the Both condition ($P = 0.003$). No further display differences were found for self-reports.

Preferences, Likes and Dislikes

Participants' postexperiment preferences were almost evenly divided across the advanced display conditions (Table 3). Only one participant preferred to monitor with the Visual condition. Participants' written comments on likes and dislikes were classified *post hoc* into categories. Counts were made for each category and the results are shown in Table 4.

Ambient Noise Levels

Average SPL across the Visual and HMD conditions was 57.5 dB(A) whereas across the Audio and Both conditions it was 59.5 dB(A). Minimum sustained SPL for the Visual and HMD conditions was 50 dB(A) and for Audio and Both it was 51.3 dB(A). Maximum sustained SPL for Visual and HMD conditions was 68.3 dB(A) and for Audio and Both it was 72.5 dB(A). Occasional peaks

Table 3. Summary of Basic Findings for the Four Display Conditions, Across all 16 Participants. Percentages in Parameter Column are Percentage Correct with Chance Responding to Distractor Task Questions. All Self-Reports are on 7-Point Likert Scales

Parameter	Visual (Control)	HMD (Visual + HMD)	Audio (Visual + audio)	Both (Visual + HMD + audio)
Monitoring performance				
Percentage of events detected (Median)[a]	50%	67%	100%†	100%*
Mean (SD)[b]	52% (32%)	75% (26%)	90% (20%)	92% (19%)
Detection time for events detected–Mean (SD)	44 s (21)	40 s (19)	33 s (15)	28 s (11)
Estimated minimum detection time–Mean (SD)	68 s (22)	52 s (25)	41 s (19)†	36 s (12)‡
Distractor task–Mean (SD)				
Number of abstracts completed	38.5 (4.6)	37.7 (3.5)	37.8 (5.3)	37.6 (4.0)
Correct anesthesia category classifications (16.67%)§	69% (11%)	66% (11%)	67% (10%)	68% (12%)
Correct research evidence class classifications (37%)§	58% (17%)	55% (14%)	52% (15%)	50% (16%)*
Correct year of publication (36%)	36% (6%)	36% (8%)	34% (9%)	34% (7%)
Self-reports (scale from 1 to 7)–Medians				
Ease of monitoring (7 = easy)[a]	2.75	4.5‡	5.0†	5.0‡
Response speed to detecting events (1 = fast)[a]	5.0	3.0*	2.0	2.0‡
Response speed to detecting resolution (1 = fast)[a]	5.0	2.0	3.0	3.0†
Ease of doing DT (7 = easy)	4.0	3.0	2.0	2.0
DT interference with monitoring (1 = strong)	3.0	3.0	3.0	3.0
Monitoring interference with DT (1 = strong)	3.0	3.0	3.0	3.0
Subjective preference				
Preference for monitoring (postexperiment, of N = 16)	1	5	4	6

DT = distractor task.

[a] Apparent inconsistencies between order of medians across conditions and the presence or absence of significance is caused by the sensitivity of the Wilcoxon signed ranks test to differing ranges of ranks for data: the ranges are not reflected in the median.

[b] Tests for significant differences of the HMD, Audio and Both conditions from the Visual condition have not been reported for means for percentage of events detected: see medians for tests of significance.

Significance levels for difference from Visual condition: * $P < 0.05$, † $P < 0.02$, ‡ $P < 0.01$, after correction for multiple tests via Tukey HSD (means) or Bonferroni correction (medians). § = chance value lies below lower bound of 95% CI for all conditions.

Table 4. Likes and Dislikes About the Displays for the 16 Participants

Likes		Dislikes	
Respiratory sonification		**Respiratory sonification**	
Immediacy of information	63%	Noted no dislikes	31%
Tracks information not on visual display	19%	Concerns with extra noise in OR	56%
Particular sound used	19%	Specific mappings of values to sounds	19%
Blood pressure earcons		**Blood pressure earcons**	
Auditory cue when NIBP taken	50%	Noted no dislikes	38%
Indication of actual BP information	44%	Concerns with transience of earcons	44%
No response	6%	Specific earcon design issues	19%
		Concerns with extra noise in OR	19%
Head-mounted display		**Head-mounted display**	
Easy availability of information	83%	Weight and bulk, and creation of headaches	56%
Unobtrusiveness of information	13%	Visual interference or distraction	31%
Noted no likes	6%	Information was redundant with monitor	13%

Results add to greater than 100% in some categories because participants listed two or more likes or dislikes.

were observed into the 80–93 dB(A) range in all conditions. When compared with monitoring using Visual displays plus audible pulse oximetry, respiratory sonification increased sustained minimum SPL by + 1.3 dB(A), average SPL by + 2.0 dB(A), and sustained maximum SPL by + 4.2 dB(A). Due to the lower statistical power and high variability in readings, a statistical test of these data was not attempted.

DISCUSSION

First, anesthesiologists detected a greater number of clinically significant events when advanced auditory displays were present, but not when the HMD was used. Adding the HMD when the advanced auditory displays were already being used did not lead to further gains. Second, detection time for events

successfully detected was not significantly faster with the advanced auditory displays when using only the events that anesthesiologists successfully detected. If the maximum times allowed for undetected events were included, however, then estimated minimum possible detection time was significantly faster when advanced auditory displays were present.

These results extend previous findings[4–6] by showing that even when used alongside visual displays and variable-tone pulse oximetry, respiratory sonification and BP earcons improve detection of significant events when the anesthesiologist is distracted. In contrast to previous findings[20] the HMD does not significantly increase event detection, probably because it does not signal changes as effectively as auditory displays do when the anesthesiologist is distracted. Rated ease of use and preference were higher for the HMD than performance with it would suggest, possibly because users are unaware of events they missed.

Advanced auditory displays dramatically increased the percentage of detected events from a median of 50% in the Visual condition to a median of 100% in the Audio and Both conditions respectively. Any failure to detect an event produced a detection delay of anywhere from 1:08 to 2:17 (min:s) (the range of event durations in the scenarios). Such delays are equivalent to those found with use of fixed-tone versus variable-tone pulse oximetry (1:37 delay)[30] and they exceed the average delay anesthesiologists showed in detecting an anoxic gas supply in the absence of pulse oximetry and visual capnography (0:53 delay).[31] The significant differences in the frequency of the delays observed indicate that informative audible information can improve the response of the distracted anesthesiologist by minutes rather than seconds.

When detection time for events detected (Table 2) was analyzed, there was only a trend in the direction predicted. Participants sometimes waited to complete the current abstract before noting an emerging patient event and initiating an exchange with the resident. This strategy may have attenuated real differences between display conditions in event detection times.

The abstract classification distractor task results suggest that participants were fully engaged in the task almost equally across display conditions. Improvements in detecting events were therefore not due to differing degrees of prioritization of the distractor task. The slightly better performance for research evidence class classification in the Visual condition compared with the Both condition suggests that identifying research evidence class was a resource-limited task[32] that yielded better results with greater effort. In the Both condition, the displays made participants more aware of the patient so they could not give as much effort to the distractor task as in the Visual condition. Clinically, it is desirable for a display to make it harder to block out awareness of the patient.

The results suggest that, when combined continuously with the risk factor of distraction, advanced auditory displays may draw attention to unsafe situations when the anesthesiologist is otherwise distracted. The results do not indicate that auditory displays lessen the need for visual monitors, clinical observation, alarms, or equipment checks, or that anesthesiologists might permit higher levels of distraction. Such conclusions would lessen the benefits of the displays and open new areas of risk.[26,33]

Tolerance of Advanced Displays

Participants felt adequately prepared to use the advanced displays and saw benefits in all of them. Concerns with the weight and bulk of HMDs will reduce as technology improves. Concerns with the added noise of auditory displays require further investigation. Given the overall variation in SPL noted, these are not large increases. Sound patterning, social effects, and perceived control are usually just as critical to tolerance.[34,35] We are investigating the advantages and disadvantages of optional earpieces that can bring the use of the advanced auditory displays completely under the anesthesiologist's control.

Limitations and Future Research

First, the advanced displays were all unfamiliar and the relatively short exposure may not have produced a fully representative view of their effectiveness in practice. Longer simulator studies and clinical evaluations are needed. Second, although our results demonstrate the effectiveness of advanced auditory displays when anesthesiologists are distracted, the displays should be tested for usefulness and tolerance during normal monitoring. Third, tests with visual and auditory alarms enabled would compare benefits of the continuous informing property of advanced displays with the discrete alerting property of alarms. Fourth, the scenarios are not fully representative of situations in the operating room in which advanced displays would be used. At 22 min each, the scenarios were relatively short and eventful, and they were oriented towards respiratory events. Fifth, performance in the HMD condition might have been significantly better than in the Visual condition with greater statistical power and if waveforms had been used. Sixth, participants' performance may have been affected by awareness of the treatment conditions and their awareness of being observed.

Finally, performance with the HMD may have been better, and performance with auditory displays worse, if the participant had been distracted solely by auditory information. However, preliminary results of a laboratory study using the scenarios and displays used here, but testing with either a visual or an auditory distractor task, suggest no selective benefits of intermodal distraction and no selective costs of intramodal distraction. Further, prior research indicates that auditory distraction in the form of music

imposes no costs on monitoring with advanced auditory displays[36] due to auditory streaming effects exploited in the design of such displays[37] and that music can even improve detection of trends.[36] Our next step is to test these displays in the clinical context.

ACKNOWLEDGMENTS

We acknowledge Queensland Health's Skills Development Centre for access in our preparatory work, with special thanks to Lucas Tomczak, Daniel Host and Andi Thompson. Our grateful thanks to the Royal Adelaide Hospital staff members who participated in the study. We also thank Philippe Lacherez for help with preparation of materials for the abstract classification distractor task and Matt Thompson for help with video analysis.

REFERENCES

1. APSF. Clinical Alarms. Anesthesia Patient Safety Foundation. Retrieved from http://www.apsf.org/initiatives/clinical_alarms/ on on Anesthesia Patient Safety Foundation website, 30 January, 2007
2. Fitch WT. Sonification system using synthesized realistic bodily sounds modified by other medically important variables for physiologic monitoring US Patent 5730140. USA, 1998
3. Fitch WT, Kramer G. Sonifying the body electric: superiority of an auditory over a visual display in a complex, multivariate system. In: Kramer G, ed. Auditory display: sonification, audification, and auditory interfaces reading. MA: Addison-Wesley, 1994
4. Loeb RG, Fitch WT. A laboratory evaluation of an auditory display designed to enhance intraoperative monitoring. Anesth Analg 2002;94:362–8
5. Seagull FJ, Wickens C, Loeb RG. When is less more? Attention and workload in auditory, visual and redundant patient-monitoring conditions. Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting: Santa Monica, CA: HFES, 2001:1395–9
6. Watson M, Sanderson P. Sonification helps eyes-free respiratory monitoring and task timesharing. Hum Factors 2004;46:497–517
7. Watson M, Sanderson P. Designing for attention with sound: challenges and extensions to ecological interface design. Hum Factors 2007;49:331–46
8. Platt MJ. Correspondence: heads up display. Brit J Anaesth 2004;92:602–3
9. Ormerod DF, Ross B, Naluai-Cecchini A. Use of a see-through head-worn display of patient monitoring data to enhance anesthesiologists' response to abnormal clinical events. Proceedings of the 6th International Symposium on Wearable Computers (ISWC'02). Los Alamitos, CA: IEEE Computer Society, 2002:131–2
10. Ormerod DF, Ross B, Naluai-Cecchini A. Use of an augmented reality display of patient monitoring data to enhance anesthesiologists' response to abnormal clinical events. Stud Health Tech Informat 2003;94:248–50
11. Ross B, Ormerod DF, Hyde JP, Fine M. Use of a head-mounted display of patient monitoring data to enhance anesthesiologists' response to abnormal clinical events. 2003 International Meeting on Medical Simulation: Society for Technology in Anesthesia, 2002
12. Hudspith S. Controlling and monitoring induced unconsciousness: ergonomics and design in anaesthesia. Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting. Santa Monica, CA: HFES, 1990
13. Block FE, Yablock DO, McDonald JS. Clinical evaluation of the head-up display of anesthesia data. Preliminary communication. Int J Clin Monit Comp 1995;12:21–4
14. Sanderson P, Watson M, Russell WJ. Advanced patient monitoring displays: tools for continuous informing. Anesth Analg 2005;101:161–8
15. Woods DD. The alarm problem and directed attention in dynamic fault management. Ergonomics 1995;38:2371–93
16. Sarter NB. The need for multisensory interfaces in support of effective attention allocation in highly dynamic event-driven domains: the case of cockpit automation. Int J Aviat Psychol 2000;10:231–45
17. Watson M, Sanderson P, Russell WJ. Tailoring reveals information requirements: the case of anaesthesia alarms. Interact Comput 2004;16:271–93
18. Watson M, Gill T. Earcon for intermittent information in monitoring environments. Proceedings of the 2004 Conference of the Computer-Human Interaction Special Interest Group of the Human Factors and Ergonomics Society of Australia. Wollongong, NSW: IEEE Computer, 2004
19. Morris RW, Montano SR. Response times to visual and auditory alarms during anaesthesia. Anaesth Intens Care 1996;24:682–4
20. Via DK, Kyle RR, Kaye RD, Shields CH, Dymond MJ, Damiano LA, Mongan PD. A head mounted display of anesthesia monitoring data improves time to recognition of crisis events in simulated crisis scenarios. Society for Technology in Anesthesia (STA2003), 2003
21. Via DK, Kyle RR, Geiger PG, Mongan PD. A head-mounted display of anesthesia monitoring data is of value and would be used by a majority of anesthesia providers. Anesth Analg 2002;95:S132
22. Crawford J, Neal A. A review of the perceptual and cognitive issues associated with the use of head-up displays in commercial aviation. Int J Aviat Psych 2006;16:113–33
23. Brodsky JB. What intraoperative monitoring makes sense? Chest 1999;115:S101–S105
24. Pedersen T, Moller AM, Pedersen BD. Pulse oximetry for perioperative monitoring: systematic review of randomized, controlled trials. Anesth Analg 2003;96:426–31
25. Runciman WB. Qualitative versus quantitative research–balancing cost, yield and feasibility. Qual Saf Health Care 2002;11:146–7
26. Reason J. Human error: models and management. Brit Med J 2000;320:768–70
27. Liu D, Jenkins S, Watson M, Sanderson P, Russell WJ. Extending simulators to improve support for patient monitoring display research. Anesth Analg 2007;104:3S:S4
28. Watson M. Scalable earcons: bridging the gap between intermittent and continuous auditory displays. Proceedings of the International Conference on Auditory Display (ICAD). Queen Mary University of London, 2006
29. Sanderson P, Scott J, Johnston T, Mainzer J, Watanabe L, James J. MacSHAPA and the enterprise of Exploratory Sequential Data Analysis (ESDA). Int J Hum-Comput St 1994;41:633–81
30. Craven RM, McIndoe AK. Auditory monitoring–how much information do we register? Brit J Anaesth 1999;83:747–9
31. Lampotang S, Gravenstein JS, Euliano TY, van Meurs WL, Good ML, Kubilis P, Westhorpe R. Influence of pulse oximetry and capnography on time to diagnosis of critical incidents in anesthesia: a pilot study using a full-scale patient simulator. J Clin Monitor Comp 1998;14:313–21
32. Norman DA, Bobrow DG. On data-limited and resource-limited processes. Cog Psych 1975;7:44–64
33. Cook R, Rasmussen J. "Going solid": a model of system dynamics and consequences for patient safety. Qual Saf Health Care 2005;14:130–4
34. Kjellberg A, Landstrom U, Tesarz M, Soderberg L, Akerlund E. The effects of nonphysical noise characteristics, ongoing task and noise sensitivity on annoyance and distraction due to noise at work. J Env Psych 1996;16:123–36
35. Hodge B, Thompson JF. Noise pollution in the operating theatre. Lancet 1990;335:891–4
36. Sanderson P, Tosh N, Philp S, Rudie J, Watson M, Russell WJ. Effects of ambient music on simulated anaesthesia monitoring with visual and auditory displays. Anaesth 2005;60:1073–80
37. Bregman AS. Auditory scene analysis. Cambridge, MA: MIT Press, 1990