

Sonification Supports Eyes-Free Respiratory Monitoring and Task Time-Sharing

Marcus Watson and Penelope Sanderson, Swinburne University of Technology, Hawthorn, Australia

Three experiments explored the effectiveness of continuous auditory displays, or *sonifications*, for conveying information about a simulated anesthetized patient's respiration. Experiment 1 established an effective respiratory sonification. Experiment 2 showed an effect of expertise in the use of respiratory sonification and revealed that some apparent differences in sonification effectiveness could be accounted for by response bias. Experiment 3 showed that sonification helps anesthesiologists to maintain high levels of awareness of the simulated patient's state while performing other tasks more effectively than when relying upon visual monitoring of the simulated patient state. Overall, sonification of patient physiology beyond traditional pulse oximetry appears to be a viable and useful adjunct to visual monitors. Actual and potential applications of this research include monitoring in a wide variety of busy critical care contexts.

INTRODUCTION

An important goal in interface design is to provide the human operator with the right information at the right time in the right format. Some work environments stand out because of the apparent difficulty of achieving this goal. For example, many researchers have pointed out that the anesthesiologist is poorly supported by the design of the anesthesia machine and the associated patient monitoring systems (Cook & Woods, 1996; Seagull & Sanderson, 2001; Watson, Sanderson, & Russell, 2004). The commonly noted frustration with auditory alarm systems is a symptom of the wrong information (critical boundaries rather than trends) arriving at the wrong time (when many other things may also be going wrong) and in the wrong format (an auditory format that must be silenced for work to proceed effectively).

In this paper we investigate whether auditory displays that provide continuous background information about system status might be useful supplements or sometimes even useful substitutes for visual displays. Although we explore this issue for the domain of anesthesia, our find-

ings are intended to generalize to other work domains as well. First, we briefly review problems with the way information about patient status is typically presented to the anesthesiologist. Second, we outline an alternative approach that relies on the continuous presentation of information so that it can be processed inside or outside focal awareness. Finally, we outline the results of prior studies before presenting three experiments that test our approach.

Information Presentation Challenges

As Woods (1995) has pointed out, most auditory displays found in aircraft, control rooms, operating rooms, and the like are alarm-based auditory displays, which are inadequate for directing visual attention. Several studies have demonstrated that most auditory alarms are ignored, are considered to be a nuisance, or serve simply as a reminder of a previously known state of affairs (Seagull & Sanderson, 2001; Watson et al., 2004; Xiao, Mackenzie, Seagull, & Jaber, 2000). Interface design in human factors has focused on visual displays, with only sparse research on other sensory modalities (Sanderson, Anderson, & Watson, 2000; Sarter, 2000). Most

recent research on auditory displays has focused on 3-D localization of auditory objects (Nelson et al., 1998) or the design of discrete auditory icons, or *earcons* (Gaver, 1997), which are not necessarily always appropriate for monitoring.

Supporting Peripheral Awareness

The eyes-free, inherently temporal nature of auditory displays lets them convey information in manner quite different from that of visual displays. Woods (1995) has suggested that auditory displays help operators stay in key control loops “preattentively,” so that the operator maintains peripheral awareness of status and trends while conducting other tasks. Sarter (2000) has pointed out that haptic displays may have the same effect, and Patterson, Watts-Perotti, and Woods (1999) have demonstrated a similar phenomenon with National Aeronautics and Space Administration voice loops. Changes often enter awareness so that an alarm or other device may not be needed to force attention to reorient only once a parameter reaches a critical set point.

Examples of continuous auditory information in use in the operating room already have the potential to work preattentively. For example, the esophageal stethoscope amplifies the patient’s natural heart and respiration sounds. However, the use of the esophageal stethoscope is limited to patients who are already anesthetized, and therefore it provides no auditory support during the high-workload periods of anesthesia induction and emergence. Other factors, such as mechanical ventilation or movement of the patient, may affect the sound produced by an esophageal stethoscope. Moreover, the esophageal stethoscope in its present form does not convey abstract information, such as the concentration of gas in the blood or airway.

A more suitable kind of auditory display for achieving preattentive awareness may be a *sonification*. A sonification is a continuous auditory display that transforms sensed or calculated relations in data into relations in sound for purposes of display (Barrass & Kramer, 1999; Kramer, 1994). A major advantage of sonification over discrete displays such as auditory icons, earcons, and alarms (Gaver, 1997) is that it can provide background information about changed states without a major disruption of attentional focus.

A very successful sonification in clinical anesthesia is the pulse oximetry sonification. The rate of a continuous series of beeps is mapped to heart rate, and the pitch of the beeps is mapped to oxygen saturation in arterial blood. The Australian Incident Monitoring Study (AIMS; Webb et al., 1993) found that pulse oximetry detected the highest proportion (27%) of evolving monitor-based incidents in the AIMS database. Clinicians’ current dependence on set-point alarms to orient attention when needed might be reduced if an effective sonification could be developed for key physiological parameters (Watson et al., 2004).

Effectiveness of Physiological Sonification

Investigators have been trying to extend sonification for patient monitoring, beyond that for heart rate (HR) and oxygen (O₂) used in pulse oximetry, to a wider range of parameters (U.S. Patent No. US5730140, 1998; Fitch & Kramer, 1994; Loeb & Fitch, 2002; Seagull, Wickens, & Loeb, 2001; Watson & Sanderson, 2001). Sonifications have been developed that add blood pressure (BP), respiration rate (RR), end-tidal carbon dioxide (ETCO₂), tidal volume (V_T), and even temperature and pupillary reflex to the HR and O₂ sonification found in pulse oximetry. Such sonifications have the potential to help anesthesiologists identify evolving changes before alarm limits are met, therefore decreasing the total number of alarms occurring in the operating room.

In an early attempt at a comprehensive sonification of patient physiology, Fitch and Kramer (1994) used two independent sound streams to carry information about the eight parameters listed in the preceding paragraph. In tests of participants’ ability to identify which of eight clinically significant events had occurred in a simulated patient, Fitch and Kramer found better performance with the sonification alone than with the visual display alone or with the visual display and sonification together. Although this work was pioneering, it had some practical shortcomings. First, the worse performance with the visual display may have been attributable to the lack of numerical readouts rather than to the use of the visual modality. Second, the experiment did not examine participants’ performance when they had to divide their attention,

as is the case in the operating room. Third, participants were asked to detect physiological states they had been trained on, rather than any deviation from normality. Finally, because the sonification was musically complex, significant states emerged better for musically trained than for physiologically trained participants.

A subsequent study by Loeb and Fitch (2002) reversed Fitch and Kramer's (1994) finding that performance is best with sonification alone. They tested anesthesiologists' performance with a simpler sonification and a visual display that included numerical readouts. Anesthesiologists identified eight different clinical events faster and more accurately with the visual display alone, and with the visual display plus sonification together, than with the sonification alone. Because neither study examined how well sonification helped participants to monitor while performing other tasks, however, it is still hard to draw conclusions about how helpful sonification might be in a clinical context.

Dual-task performance was examined by Seagull et al. (2001) using the Loeb and Fitch (2002) sonification. Nonanesthesiologist participants performed a manual tracking task while monitoring visual, sonified, or both visual and sonified (combined) physiological data from a simulated patient. Tracking error was marginally less when patient data were sonified, but participants detected changes in the simulated patient status faster with the visual display than with the sonification. These results suggest a performance trade-off between the tracking and monitoring tasks that varies with modality. Because Seagull et al. did not test anesthesiologists, it is unclear whether the trade-off was attributable to lack of expertise or to modality effects. When participants were using sonification alone for patient monitoring, adding the tracking task had the least effect on their monitoring performance.

Present Research

Clearly, there are some advantages to using sonification for patient monitoring, but many questions remain. First, there are inconsistencies as to whether sonification alone or sonification supported by visual displays leads to the best performance. Second, the role of domain expertise has not been directly tested. Third, it is unclear what the effect of expertise is on the ability

to time-share monitoring tasks with other activities.

Our goal is to explore these issues with a respiratory sonification that might be used alongside conventional pulse oximetry. As noted, the AIMS study established that pulse oximetry detected 27% of monitor-detected evolving incidents (Webb et al., 1993). Capnography (the display of a continuous measurement of CO₂ concentration in exhaled breath, or ET_{CO}₂) came second, detecting 24% of monitor-detected incidents. As Webb et al. (1993) noted, if RR and V_T had been combined with capnography, 39% of the monitor-detected evolving incidents would have been detected by integrated respiratory monitoring alone. Anesthesiologists' reliance on these specific vital signs is also supported by a more recent study into quantification of monitor effectiveness (Findlay, Spittal, & Radcliffe, 1998). Looking at the AIMS data, if pulse oximetry and respiratory monitoring are both used, then more than 90% of the incidents in the AIMS database could have been detected in a timely fashion. This is because the anesthesiologist could combine information about cardiac and respiratory functioning coming from the various sensors to infer important output signals, such as oxygen transport. If such a benefit could be promoted in an effective "eyes-free" manner simply by adding a relatively straightforward auditory display to the breathing circuit, then it seems very worthwhile.

Unlike previous researchers, we did not include blood pressure in our investigations. Although blood pressure is important for clinical monitoring, readings from an invasive arterial line would be required for a BP sonification. Most operations use noninvasive blood pressure, which is sampled only every few minutes and cannot be validly sonified (for further details see Watson et al., 2004).

The aforementioned conclusions were supported by a work domain analysis of anesthesia that specified the subdomain of the anesthetized patient (Watson, Russell, & Sanderson, 2000; Watson & Sanderson, 1998). The work domain analysis underscored the importance of the five parameters already considered for providing information about cellular and intracellular functioning, and it demonstrated how higher-order properties of patient physiology could be inferred

from the interactions between these parameters. We conjectured that sonification of RR, V_T , and $ETCO_2$, in addition to heart rate and oxygenation, might move monitoring from the rule-based level to the skill-based level so that patient status could be monitored in an integrated and less resource-demanding way (Vicente & Rasmussen, 1990), thereby satisfying a key concern of ecological interface design.

We report three experiments, which proceed from the design of a respiratory sonification to a test of its use in a simulated dual-task patient monitoring context. Experiment 1 compares three potential respiratory sonifications distinguished by their relative demands on memory and attention. Experiment 2 compares the monitoring performance of anesthesiologists and nonanesthesiologists with the best sonification from Experiment 1. Experiment 3 tests how well the respiratory sonification helps anesthesiologist and nonanesthesiologist participants to perform time-shared monitoring. All three experiments used a laboratory-based anesthesia simulator (Arbiter) and a relatively artificial task environment because before we performed more valid clinical tests, our initial focus was the perceptual adequacy of the sonifications.

EXPERIMENT 1

Goal

The goal of Experiment 1 was to evaluate three candidate respiratory sonifications to see if one was clearly better than the others and whether any was as effective for supporting monitoring of respiratory parameters as pulse oximetry is for supporting the monitoring of heart rate and oxygen saturation. Each respiratory sonification captured RR, V_T , and $ETCO_2$. Three candidate respiratory sonifications were used: varying, even, and short. The varying sonification was developed from the work domain analysis and the review of the AIMS findings. The even and short sonifications, which were modifications of the varying sonification, took into account memory and attentional demands of auditory perception (Figure 1).

The varying sonification worked from the volume flow meter in the simulated anesthesia circuit to provide a moment-by-moment sonification of accumulated volume. V_T could be in-

ferred from the maximum sound intensity. The even and short sonifications were both attempts to remove the need to wait until the maximum sound intensity to infer V_T . However, because V_T is not known until the end of a breath, the even and short sonifications displayed V_T from the previous breath, whereas the varying sonification displayed V_T from the current breath. The short sonification removed the need to attend to an entire breath to extract information about respiratory rate. The short sonification initialized at the start of a breath and compressed the sonification into around one quarter of the previous breath duration. Participants could integrate the time between sounds to estimate RR without having to pay continuous attention to the sonification. Overall, these manipulations made the even and short sonifications closer to the Loeb and Fitch (2002) and Seagull et al. (2001) sonifications.

The experimental hypotheses were as follows (hypotheses are numbered with the experiment number and the hypothesis number within an experiment):

H1.1. When the pulse oximetry and respiratory sonification are played at the same time, at least one respiratory sonification will support performance at identifying changes in respiratory parameters as well as pulse oximetry does for heart rate and oxygenation parameters.

H1.2. The even and short sonifications will support better performance at identifying changes in V_T than will the varying sonification when pulse oximetry and respiratory sonification are played at the same time. Even and short sonifications will map V_T onto a constant rather than a varying sound intensity, making it easier to compare V_T across breaths.

H1.3. The short sonification will support better performance at identifying changes in RR than will the varying and even sonifications when pulse oximetry and respiratory sonification are played at the same time. RR will be conveyed over a shorter period of time, rather than requiring the participant to attend through to the end of the breath to identify its length.

Method

Participants. This study was conducted with 23 paid participants (7 men and 16 women) from the general public. They were between 19

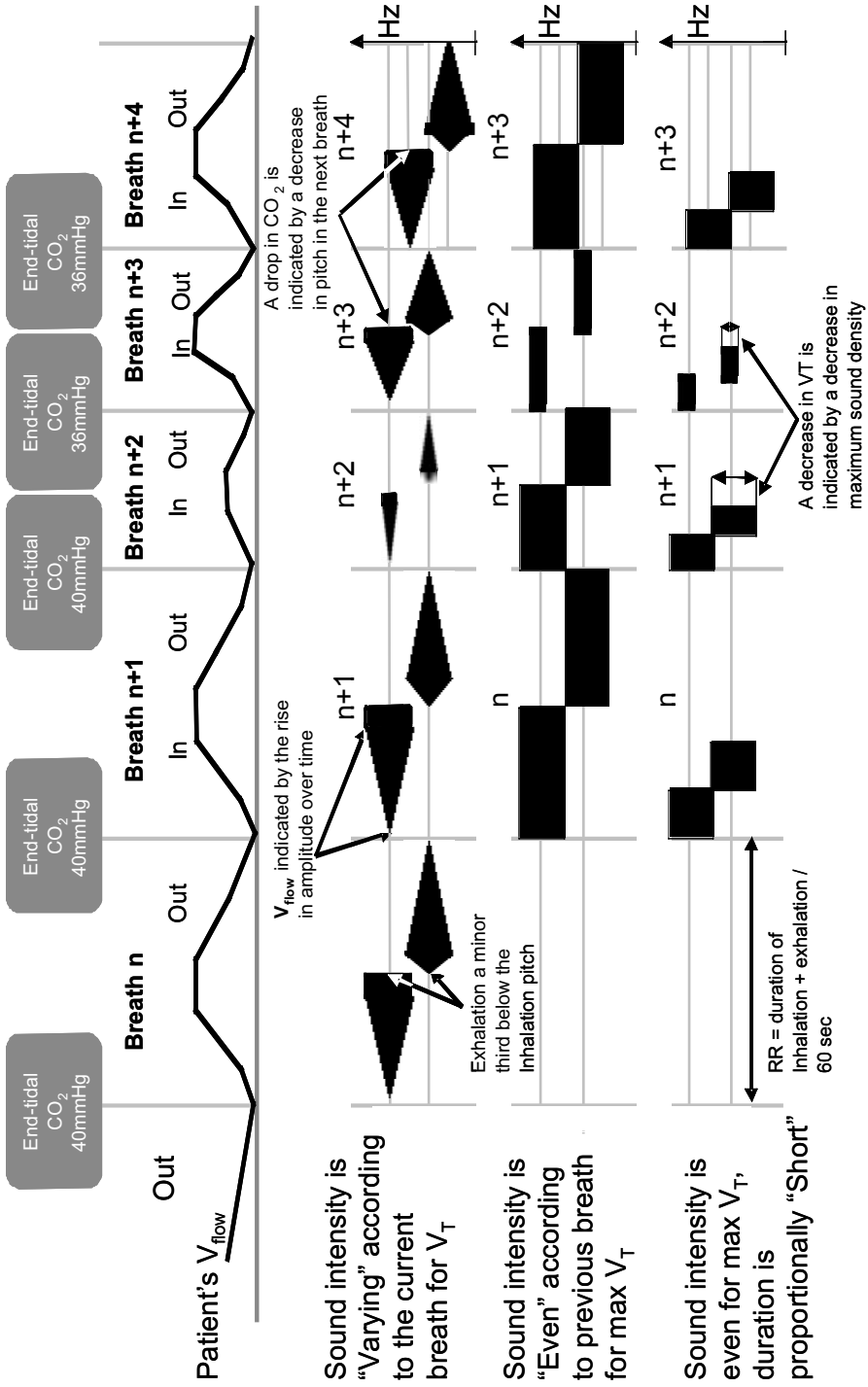


Figure 1. Design of sonifications for Experiment 1. The differences among the three sonifications are plotted for five breaths. The first two breaths (n and n+1) are of equal duration and similar volume. Breaths n+2 and n+3 are shorter, and n+2 is a shallow breath. Breath n+4 increases in duration while maintaining the same tidal volume. The amplitude of each sound is represented by the height of each object, the pitch difference between the inhalation and exhalation is represented by the upper and lower lines for each sonification, and the duration of the inhalation and exhalation is represented by the length of each object. Frequency is represented by the relative vertical position of the object.

and 55 years of age and had 2 or more years of tertiary education.

Stimuli and apparatus. Experiments were run with the Arbiter anesthesia display simulator, which incorporates the Advanced Simulation Corporation's Body99.dll™ anesthesia simulator. Experiment 1 used 12 anesthetic scenarios created with Arbiter to simulate a variety of physiological events and mechanical changes in the anesthetic system that would affect the simulated patient's HR, O₂, RR, V_T, and/or ETCO₂. Each scenario was 4.5 to 5 min long and included events such as a patient spontaneously breathing against the ventilator, a morphine overdose on a spontaneously breathing patient, a right main stem endobronchial intubation on a mechanically ventilated patient, N₂O and O₂ pipes swapped over on a spontaneously breathing patient, and a laryngeal cuff leak on a mechanically ventilated patient. Each anesthesia scenario produced values in both the normal and abnormal range for five physiological parameters: HR, O₂, RR, V_T, and ETCO₂.

All three sonifications used a pure tone and mapped inhalation and exhalation to the upper and lower note of a musical third interval. Previous investigations had revealed that subject-matter experts considered the use of a breath-like tone inadvisable (Watson, Sanderson, & Russell, 2000). RR was represented by a direct temporal mapping of inhalation and exhalation, as sensed in the simulation. V_T was represented by sound intensity, and ETCO₂ was represented by a frequency modulation (pitch change) of the inhalation and exhalation.

The pulse oximetry and respiratory sonifications were presented to participants using pre-recorded audio files generated with Arbiter. All files were played using a fixed amplification (amplitude setting on the Yamaha GX-500 Mini Component System was set to 28/64) to aid standardization. Background noise was measured at between 47.7 and 48.8 dB(A). Participants sat at a table facing the loudspeaker, which was under the table.

Experimental design. Sonifications were tested using a within-subjects design in which the order of presentation of sonifications and scenarios was counterbalanced. Scenarios were grouped into four clusters of three scenarios. The clusters were designed to give a range of

scenarios that would focus on changes evident in respiratory parameters. Participants experienced one cluster for familiarization and one cluster for each of the three respiratory sonifications during evaluation. Each scenario unfolded continuously but was interrupted every 25 to 35 s (9 or 10 times per scenario) for a report of patient status. The lengths of reporting periods were varied to prevent participants from counting the number of breaths or heartbeats to determine RR or HR.

Procedure. The experiment involved three phases: introduction, familiarization, and evaluation. In the 10-min introduction, participants were presented with 30-s samples of the varying, even, and short sonifications in order to demonstrate the performance of the sonifications in the normal respiratory range alongside pulse oximetry. The participants received feedback from the experimenter on what was happening to HR, O₂, RR, V_T, and ETCO₂. In the familiarization phase, the experimenter walked the participant through three anesthesia scenarios and described what was happening to the simulated patient. After the familiarization phase, performance feedback was withheld until after participants had completed the evaluation phase of the experiment. During the evaluation, at the end of each reporting period participants indicated whether each parameter was high, normal, or low (abnormality judgment) and whether the parameter was increasing, steady, or decreasing (direction judgment). They made judgments on the basis of the most salient change they heard. Participants circled their responses on specially prepared paper answer sheets, and the experimenter wrote down any comments they made.

At the end of each cluster, the participants reported their confidence in the accuracy of their judgments and the workload that they had experienced when monitoring the patient with the sonifications, using 7-point scales on the answer sheet. At the very end of the experiment, the participants reported which sonification they found easiest to use for each of the respiratory parameters and which they found easiest to use as a whole.

Results

Judgments. Abnormality and direction judgments were analyzed in within-subjects analyses

of variance (ANOVAs) with two within-subjects parameters: sonification (with the levels varying, even, and short) and parameter (with the levels HR, O₂, RR, V_T, and ETCO₂). For abnormality judgments there was a significant effect of sonification, $F(2, 44) = 7.955$, $MSE = 177.31$, $p < .01$, and of parameter, $F(4, 88) = 37.91$, $MSE = 172.88$, $p < .001$, but no interaction between sonification and parameter, $F(8, 176) = 1.661$, $MSE = 142.24$, ns . Results indicate better performance overall for the varying sonification, especially for the V_T and ETCO₂ parameters (see Figure 2).

For direction judgments (see Figure 3), no significant effect was found for sonification, $F(2, 44) = 0.791$, $MSE = 93.12$, ns ; however, there was a significant effect of parameter, $F(4, 88) = 16.998$, $MSE = 177.35$, $p < .001$. There was no significant interaction between sonification and parameter, $F(8, 176) = 0.263$, $MSE = 127.71$, ns .

Preferences. Subjective preferences indicated a slight preference for the varying sonification. In an ANOVA similar to the one described earlier, sonification showed a marginally significant effect, $F(2, 44) = 2.717$, $MSE = 1.127$, $.10 > p > .05$, with varying sonification always producing

slightly higher confidence. Parameter was significant, $F(4, 88) = 11.06$, $MSE = 2.319$, $p < .001$, indicating greater confidence overall in judgments about changes in HR. There was no interaction between sonification and parameter, $F(8, 176) = 0.624$, $MSE = 0.623$, ns . When participants were asked at the end of the experiment which respiratory sonification they preferred for each respiratory parameter (RR, V_T, and ETCO₂), the varying sonification received the highest number of preferences for all three parameters. There were no significant results for workload. Taken together, the results suggest that the varying sonification not only supported better performance but also was preferred by participants.

Performance against chance and event base rates. One potential concern was that even with the varying sonification, performance seemed to be not as good for the respiratory parameters as for heart rate and oxygenation. There are two reasons, however, not to draw such a conclusion. First, because of our focus on respiratory parameters, our scenarios included many more respiratory events than heart rate and oxygenation events. Participants may have become

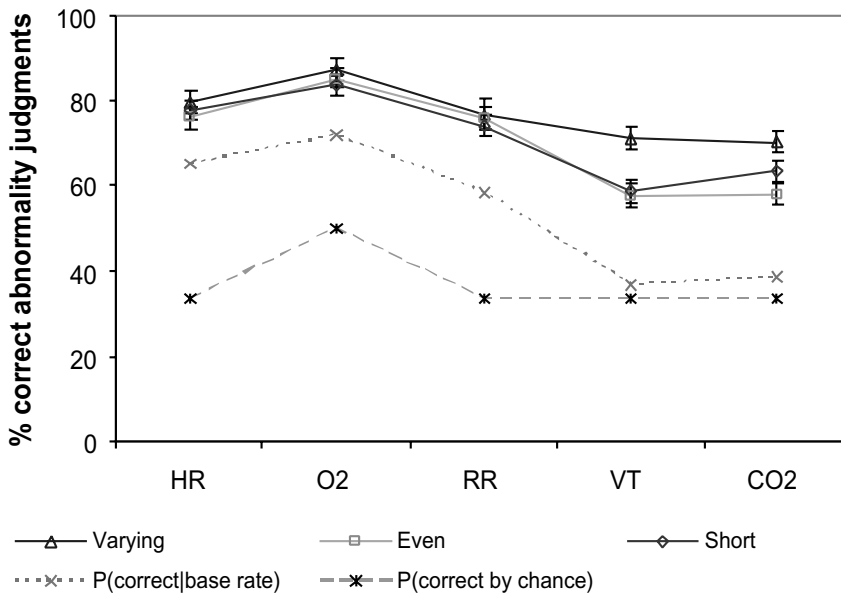


Figure 2. Experiment 1: Percentage correct abnormality judgments. Dotted line for $p(\text{correct}|\text{base rate})$ gives the probability of being correct given the base rate probability of abnormalities in the parameter in question in the experiment. Dashed line for $p(\text{correct}|\text{chance})$ gives the probability of being correct given the number of answer alternatives.

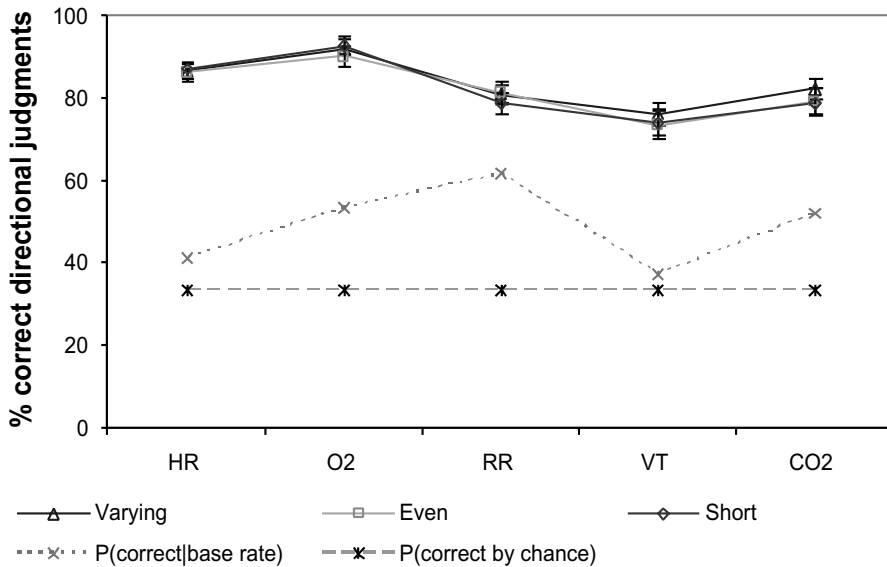


Figure 3. Experiment 1: Percentage correct direction judgments. See Figure 2 caption for explanation of dotted and dashed lines.

sensitive to the base rate probability of events on different parameters. The percentage of changes to be reported ranged from 17% to 52% across parameters for abnormality judgments and from 22% to 51% for direction judgments. For infrequently changing parameters, this bias may have led to fewer false positive responses. The data were not suitable for a signal detection analysis, but we were able to test whether performance was better than would be expected by chance, as we will describe. Second, the ways that HR and O₂ became abnormal led to a higher probability of correct responding. The only abnormal state for O₂ is low, given that its normal value is 100% or thereabouts, and in the scenarios chosen O₂ moved only downward. Finally, a few HR changes were quite dramatic, making them easier to judge.

To judge the influence of those two factors on performance, we looked to see if participants were detecting changes in each physiological parameter (a) better than would be expected if they randomly chose a response, $p(\text{correct}|\text{chance})$, and (b) better than would be expected by chance if they simply matched the event probability for that parameter, $p(\text{correct}|\text{base rate})$. In Figures 2 and 3, values for $p(\text{correct}|\text{chance})$ and $p(\text{correct}|\text{base rate})$ are shown as dashed and dotted lines, respectively. Judgments

of abnormality are an approximately constant amount better than $p(\text{correct}|\text{base rate})$, and for the varying sonification for V_T and ET_{CO}₂ they are disproportionately better than base rate, as compared with pulse oximetry. This is supported by an analysis of individual results. We counted the number of participants in each condition who judged abnormality and direction better than chance (see Table 1). The varying sonification supports not only better performance overall but also better performance for respiratory than for pulse oximetry parameters.

Discussion

Experiment 1 demonstrated that the varying sonification is a workable sonification of respiratory parameters. Our concern about possible attentional and memory demands of the varying sonification was apparently not warranted. It was clear, however, that the range of response options and event base rates affected the probability of making correct judgments. This factor was taken into account in the analysis and in future experiments.

We had hypothesized that at least one respiratory sonification would support performance at identifying changes in respiratory parameters as effectively as pulse oximetry does for heart rate and oxygenation (Hypothesis H1.1).

TABLE 1: Participants in Experiment 1 Whose Judgments Were Significantly Better Than Chance or Event Baseline for the Three Conditions (Maximum $N = 23$)

Measure	Sonification	HR	O ₂	RR	V _T	ETCO ₂
Abnormality judgment	Varying	15*	15*	17**	13	18**
	Even	10	13	13	5	8
	Short	12	11	14	6	7
Direction judgment	Varying	16**	21**	16**	18**	19**
	Even	20**	20**	17**	13	17**
	Short	20**	22**	14*	17**	16**

Note. Measures were compared with $p(\text{chance})$ and $p(\text{correct/base rate})$.

**A significant number of participants better than the comparison, $p < .05$. *Marginal results, $.10 > p > .05$. No asterisks = nonsignificant results.

Performance with the varying sonification comes closest to this and is supported by participants' preferences. The varying sonification also has the advantage that it sonifies the current rather than the previous breath. The varying sonification was therefore deemed acceptable for further experimentation, even though further refinement of the sonification might lead to even better results. The superiority of the varying sonification led to the disconfirmation of Hypotheses H1.2 and H1.3, which had predicted better performance with the other sonifications for judgments about V_T and RR. This suggests that using sound intensity to represent the accumulation of V_T leads to successful interpretation by participants and that there is no advantage to compressing the sound in time.

An important shortcoming of Experiment 1 is that all participants were nonanesthesiologists with no physiological or medical training. In Experiment 2 we examined whether anesthesiologists would work as effectively with the varying sonification as the nonanesthesiologists did.

EXPERIMENT 2

Goals

There were two main goals in Experiment 2. First, we wished to compare the physiological monitoring performance of anesthesiologists and nonanesthesiologists with the varying sonification, which supported the best performance in Experiment 1. The questions were whether anesthesiologists would perform better overall than nonanesthesiologists and whether they would perform as well with the respiratory sonification as with pulse oximetry, with which they

are clearly very familiar. Second, we wished to investigate whether the proportion of abnormal signals in each physiological parameter was affecting monitoring accuracy. Therefore we chose scenarios with event base rates that were more similar across parameters than was the case in Experiment 1. Specifically, we wished to see if the varying sonification would support monitoring performance at the same level seen for pulse oximetry once we made a correction for the bias in event base rates in the design of the scenarios used in Experiment 1. Our hypotheses were therefore as follows.

H2.1. The varying sonification will support performance at identifying changes in respiratory parameters as effectively as pulse oximetry does for heart rate and oxygenation when both sonifications are played at the same time.

H2.2. Anesthesiologists will detect abnormalities and directional changes better than will participants with no physiological training.

H2.3. Participants' ability to detect abnormalities and directional changes will be lower overall than those observed in Experiment 1 because of the reduced predictability in the scenarios used for Experiment 2.

Method

Participants. Experiment 2 used 11 anesthesiologists and 10 information technology (IT) postgraduates with no anesthesia experience. The anesthesiologists were between 29 and 62 years of age and had 4.5 to 33 years' postgraduate medical experience (4 women, 7 men). The IT postgraduates were between 23 and 44 years of age (2 women, 8 men). All had prior experience studying and working in technology-related

areas and were considered capable of understanding the idea of mapping signals to auditory parameters.

Stimuli and apparatus. The experiment was presented to the participants much as for Experiment 1. Background noise was 45.9 to 47.2 dB(A) for the IT postgraduates and between 47.8 and 57.5 dB(A) for the anesthesiologists. Modifications were made to some of the anesthesia scenarios to see if the probability of abnormal states or the predictability of parameters was affecting participants' performance. All the scenarios remained physiologically conceivable but would be unlikely to occur in the operating room. Because many physiological events affect respiratory parameters long before they affect oxygen saturation, there was still a bias toward more abnormal values for V_T and $ETCO_2$ than for HR and O_2 .

Experimental design. There were several changes in the experimental design, compared with Experiment 1. First, a new independent variable was the participant groups: anesthesiologists and IT postgraduates. Second, only the varying sonification was tested. Third, the number of scenarios used for the evaluation was reduced from nine to six. Fourth, the total number of abnormal changes in any one parameter was adjusted to fall into a narrower range than

in Experiment 1. The percentage of abnormal changes fell between 33% and 50% across the five parameters, and the percentage of directional changes fell between 33% and 54% across the five parameters.

Procedure. Participants were guided through short introduction (10 min) and familiarization (20 min) sessions, followed by the evaluation phase, which had two procedural changes as compared with Experiment 1. First, when giving responses, participants were asked to indicate the last directional change they heard, rather than the most salient directional change, so as to reduce ambiguities in scoring. Second, participants answered the confidence and workload questionnaires at the end of each scenario, rather than at the end of each cluster, so that we could better detect any trends with exposure to the task.

Results

Judgments. Abnormality judgments were analyzed in a between/within-subjects ANOVA, with the between-subjects factor of group (with the levels of anesthesiologists and IT postgraduates) and the within-subjects factor of parameter; see Figure 4 for results. There was a significant effect of group, $F(1, 19) = 20.604$, $MSE = 0.013$, $p < .001$, with anesthesiologists

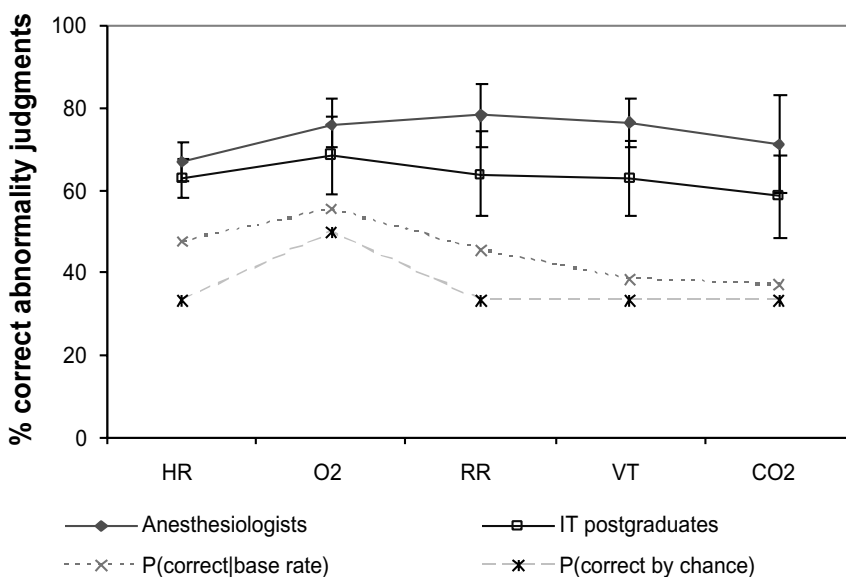


Figure 4. Experiment 2: Percentage correct abnormality judgments. See Figure 2 caption for explanation of dotted and dashed lines.

performing better overall. There was also a significant effect of parameter, $F(4, 76) = 4.786$, $MSE = 0.005$, $p < .01$, with HR judged least accurately and O₂ most accurately. There was no interaction between group and parameter, $F(4, 76) = 1.923$, $MSE = 0.005$, ns .

Direction judgments were analyzed in a similar between/within-subjects ANOVA, with the between-subjects factor of group and the within-subject factor of parameter; see Figure 5 for results. There was a significant effect of group, $F(1, 19) = 10.341$, $MSE = 0.011$, $p < .01$, with anesthesiologists performing better, and of parameter, $F(4, 76) = 22.09$, $MSE = 0.005$, $p < .001$, with HR and V_T showing the least accurate judgments. There was no interaction between group and parameter, $F(4, 76) = 1.863$, $MSE = 0.005$, ns . As expected, the adjustment to the scenarios that was made in order to achieve a more restricted range for the percentage of events to report across physiological parameters reduced the advantage for pulse oximetry parameters over respiratory parameters.

Performance against chance and event base rates. The results were compared against $p(\text{correct}|\text{chance})$ and $p(\text{correct}|\text{base rate})$, using the procedure described for Experiment 1. Dashed and dotted lines in the lower parts of Figures 4 and 5 show the profiles of expected correct

responses if responding had followed chance or base rate alone. Abnormality judgments suggest that anesthesiologists performed disproportionately better on the respiratory parameters. The direction judgments follow the $p(\text{correct}|\text{base rate})$ curve quite closely. Table 2 presents the number of participants in each condition who performed significantly better than $p(\text{correct}|\text{chance})$ or $p(\text{correct}|\text{base rate})$ for each type of judgment. Results indicate that even with the base rate probabilities distributed more evenly across physiological parameters, the varying sonification supports judgments about the respiratory parameters as well as, if not better than, pulse oximetry does for heart rate and oxygenation. In addition, a greater proportion of anesthesiologists than IT postgraduates showed results that are better than chance.

Preferences. Participants' confidence in their abnormality and direction judgments were analyzed in a between/within-subjects ANOVA, with the between-subjects factor of group and the within-subjects factors of parameter and practice (from the first through the sixth presented scenario). Overall levels of confidence ranged from 3.6 to 5.9 on a 7-point scale. The only significant effects were for parameter, $F(4, 76) = 8.656$, $MSE = 3.892$, $p < .001$, and for the Group \times Practice interaction, $F(5, 95) = 2.552$,

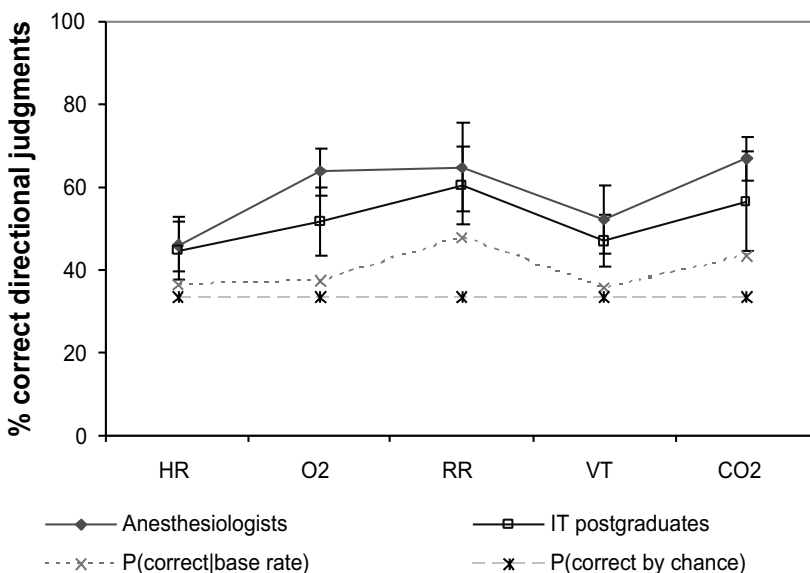


Figure 5. Experiment 2: Percentage correct direction judgments. See Figure 2 caption for explanation of dotted and dashed lines.

TABLE 2: Participants in Experiment 2 Whose Judgments Were Significantly Better Than Chance or Event Baseline

Measure	Group		HR	O ₂	RR	V _T	ETCO ₂
Abnormality judgment	Anesthesiologists	<i>p</i> (c)	11**	10**	11**	11**	11**
	IT postgraduates	<i>p</i> (c)	10**	4	9**	7	8**
	Anesthesiologists	<i>p</i> (clb)	7	8*	11**	11**	10**
	IT postgraduates	<i>p</i> (clb)	5	3	7	7	7
Direction judgment	Anesthesiologists	<i>p</i> (c)	5	11**	11**	7	11**
	IT postgraduates	<i>p</i> (c)	4	6	9**	3	7**
	Anesthesiologists	<i>p</i> (clb)	1	9**	6	7	9**
	IT postgraduates	<i>p</i> (clb)	3	3	5	3	5

Note. Measures were compared with *p*(chance) — *p*(c) — and with *p*(correct/base rate) — *p*(clb). For anesthesiologists *N* = 11; for IT postgraduates *N* = 10.

**A significant number of participants better than the comparison, *p* < .05. *Marginal results, .10 > *p* > .05. No asterisks = nonsignificant results.

MSE = 0.892, *p* < .05. Overall, participants felt most confident judging O₂, but there was no difference in confidence for the other parameters. The IT postgraduates' judgments of confidence tended to vary more across successive scenarios than did the anesthesiologists'.

A similar ANOVA for rated workload indicated a marginally significant effect for group, *F*(1, 18) = 3.663, *MSE* = 4.968, .10 > *p* > .05, with anesthesiologists rating their workload higher. There was also a marginally significant effect of practice, *F*(5, 90) = 2.215, *MSE* = 0.343, .10 > *p* > .05, showing a tendency to rate workload higher later in the experiment. No other effects were significant.

Discussion

Experiment 2 provided a further opportunity to test the hypothesis that the varying sonification supports performance at identifying changes in respiratory parameters as effectively as pulse oximetry does for heart rate and oxygenation (Hypothesis H2.1). The probabilities with which changes happened across the five physiological parameters were closer to one another than in Experiment 1. The results indicate that evaluations of how well parameters have been mapped to sound dimensions are highly conditioned by the probability of correct responding by chance and by the event base rate probability.

Overall, anesthesiologists made more accurate abnormality and direction judgments than IT postgraduates did, supporting Hypothesis H2.2. However, overall performance for the

judgments was worse in Experiment 2 than in Experiment 1. For the purpose of control, Experiment 2 used scenarios that were less physiologically probable than in Experiment 1. There was less certainty about which parameters would show abnormalities because all had a more equal chance of doing so. Thus overall judgment performance was worse in Experiment 2 than in Experiment 1, which supported Hypothesis H2.3.

The absolute levels of performance may raise concerns about whether performance with full sonification would be acceptable if transferred to the operating room. The conditions under which participants performed in Experiment 1 and particularly Experiment 2 were extreme. Not only were the anesthesia scenarios unusual, but also the experimental demands were high. In addition, participants were given no feedback about their performance during the experiment and had no visual information as backup. Validation in situations much closer to those encountered in the operating room are necessary before the clinical suitability of the respiratory sonification can be determined.

Experiments 1 and 2 examined performance with sonification alone and did not compare performance with sonification alone with performance supported by a visual display. Moreover, Experiments 1 and 2 tested the respiratory sonification under focal awareness, rather than when attention was divided over more than one task. Experiment 3 therefore compares the performance of anesthesiologists and nonanesthesiologists under divided-attention conditions with displays in different modalities.

EXPERIMENT 3

Goals

Our goal in Experiment 3 was to see how well participants could monitor simulated patient physiological parameters under sonification, visual, or sonification plus visual conditions while performing another task. Experiment 3 therefore allowed us to combine aspects of the Loeb and Fitch (2002) study with the Seagull et al. (2001) study and, possibly, to clarify discrepancies between them.

In the ideal case, a sonification should help anesthesiologists perform another task better than with a visual display while maintaining equal if not better performance on patient monitoring. The additional task we used was an arithmetic true-false task. The arithmetic task captured some characteristics of anesthesiologists' tasks, such as drug calculations and drug/fluid selection. Although drug calculations often involve multiplication rather than addition or subtraction, the cognitive demands are similar.

We wished to see whether patient information could be accurately obtained from the sonification while the participant attended to the arithmetic task. We also wished to see whether performance on both tasks was better when sonification was available and whether anesthesiologists showed a selective performance advantage with sonification because of their domain knowledge. Our hypotheses were therefore as follows.

H3.1. Participants will achieve higher abnormality and direction judgment scores in the combined sonification plus visual condition than in the visual condition or the sonification condition. Sonification will direct attention to changes, whereas the visual display will allow for occasional calibration of how the sonification is being interpreted.

H3.2. Participants will report greater confidence in their ability to perform the patient monitoring task with the combined sonification plus visual condition, as compared with the sonification condition or the visual condition, because the visual display will let participants confirm what they hear in the sonification.

H3.3. Participants will rate workload higher for the sonification plus visual condition than for the visual condition and will rate workload high-

er for both of these than for the sonification condition. The sonification plus visual condition has more sources of information to attend to in more modalities, whereas the sonification condition allows participants to share workload between the two modalities.

H3.4. Anesthesiologists will be better at categorizing the clinical events for each scenario and for all three monitoring conditions, as compared with IT postgraduates, because they have prior experience with patient physiology.

Method

Participants. Experiment 3 was conducted using the same 21 participants who took part in Experiment 2 (11 anesthesiologists and 10 IT postgraduates).

Stimuli and apparatus. The Arbiter visual display was presented on a 21-inch (53-cm) MicroTouch™-enabled touch screen (1280 × 1024 resolution; see Figure 6). All numerical parameters were presented in the same font and font size to prevent participants favoring one over the other. Pulse oximetry parameters were green and respiratory parameters were white, so that participants using both the visual display and the sonifications could associate the colors with the different sonifications. The ranges for the different parameters were written on paper and placed over the left side of the visual monitor. The pulse oximetry and respiratory sonifications were produced in real time using Arbiter. Background noise levels were the same as in Experiment 2.

Experiment 3 used the “withholding” technique in the visual conditions, requiring participants to touch the part of the video screen associated with a physiological parameter to see a readout of its value. The readout lasted for 5 s or until another parameter was touched, at which point it disappeared. Waveforms for electrocardiography and capnography were also displayed during the HR and ET CO_2 readouts, respectively.

The nine anesthesia scenarios used in the evaluation simulated a variety of realistic physiological events and mechanical events in the anesthesia equipment that induced changes in the simulated patient's cardiovascular and respiratory systems. Each scenario was approximately 9 min long. The number of changes seen in a

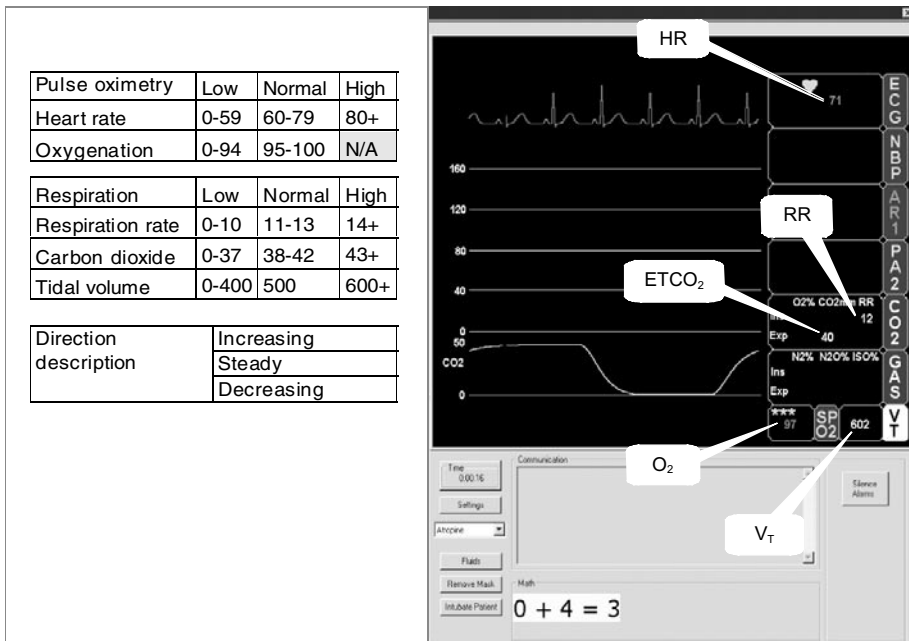


Figure 6. Arbitrator monitor setup for Experiment 3. The left panel shows the five parameters and two visual plots without the withholding method. The right panel shows what the participants viewed during Experiment 3, with the range sheet concealing the gas and ventilator settings on the left of the main monitor.

physiological parameter across all the scenarios is in the following order, from most to least frequent: $V_T > ETCO_2 > RR > O_2 > HR$.

The distractor task required participants to make true/false judgments of arithmetic expressions displayed at the bottom center of the screen (see Figure 6). Values to be added or subtracted were random numbers between 0 and 9, and answers fell between -9 and 18. New arithmetic expressions appeared every 10 s, and the background color toggled between white and yellow as a new problem was presented. Responses to the arithmetic task and screen touches to request physiological information were recorded automatically.

Experimental design. Experiment 2 and Experiment 3 were conducted in series, so that Experiment 2 trained participants with the respiratory sonification that was used in Experiment 3. Group was a between-subjects variable and modality a within-subjects variable. The three modalities were sonification alone for all parameters (S), visual display alone for all parameters (V), and both sonification and visual display for all parameters (B). The orders of

cluster presentation and modality were counter-balanced across participants to avoid confounding of cluster, modality, and order.

The update rate of the arithmetic task gave participants enough time to query all five parameters in the V and B conditions, if they wished, before the next arithmetic expression appeared. Although sonification is intended to provide information in situations in which visual information is unavailable or inconvenient, our experiment used the best possible conditions for the visual display to succeed. If S leads to superior performance, then it would be under conditions in which it would be possible to do the task visually to the same level of performance. In contrast, if S leads to the same level of performance as for V, then at least the sonification does not support worse performance than a visual display does. Finally, if S leads to worse performance than does V, then in later experiments we could increase the workload of the arithmetic task to see if the modality effect reverses or at least diminishes.

Procedure. Experiment 3 had two phases: familiarization with the visual display, which

took around 10 min; and evaluation, which took around 130 min.

Participants were told that the arithmetic task was the primary task and that they should aim to achieve 100% correct responses on it while still performing the physiological monitoring. Every 45 to 60 s, the scenario was halted and a recorded voice spoke the name of a physiological parameter (a longer interprobe interval was used than before in order to give participants a longer period over which to judge RR). The participant was asked to vocalize, without looking at the visual display (if available), (a) whether the probed parameter was high, normal, or low and (b) whether it was increasing, decreasing, or steady. Parameters probed were quasi-randomized to prevent participants from guessing what the next parameter would be.

A questionnaire at the end of each scenario asked participants to recall information about the scenario they had just monitored, how confident they felt about their responses for the different parameters, and to rate the workload involved in monitoring the simulated patient while performing the arithmetic task. At the end of the experiment, participants stated which condition they found easiest to monitor and which condition had the highest workload.

Results

Arithmetic task. Accuracy of arithmetic task performance was analyzed in a between/within-

subjects ANOVA, with the between-subjects factor of group (anesthesiologists and IT postgraduates) and the within-subjects factor of modality (S, V, and B conditions). Results are shown in Figure 7. There was a significant difference for group, $F(1, 19) = 9.540$, $MSE = 0.006$, $p < .01$, with the anesthesiologists performing better than the IT postgraduates in all conditions. There was also a significant effect of modality, $F(2, 38) = 10.054$, $MSE = 0.002$, $p < .001$, showing best performance for the S condition, followed by the V and then the B condition. The Group \times Modality interaction was not significant.

Patient monitoring. Participants' accuracy at making abnormality and direction judgments was examined under the S, V, and B conditions. Results for abnormality and direction judgments, collapsed over physiological parameters, are given on the y axes of Figure 7.

Abnormality judgments were analyzed in a between/within-subjects ANOVA, with the between-subject factor of group and within-subjects factors of modality and parameter (HR, O₂, RR, V_T, and ETCO₂). There was a significant difference for group, $F(1, 19) = 37.45$, $MSE = 1.98$, $p < .001$, with anesthesiologists showing better performance. There was also a significant effect for modality, $F(2, 38) = 4.43$, $MSE = 1.2$, $p < .02$, with B showing the best performance, followed by V, and S showing the worst performance. In addition, there was a marginally

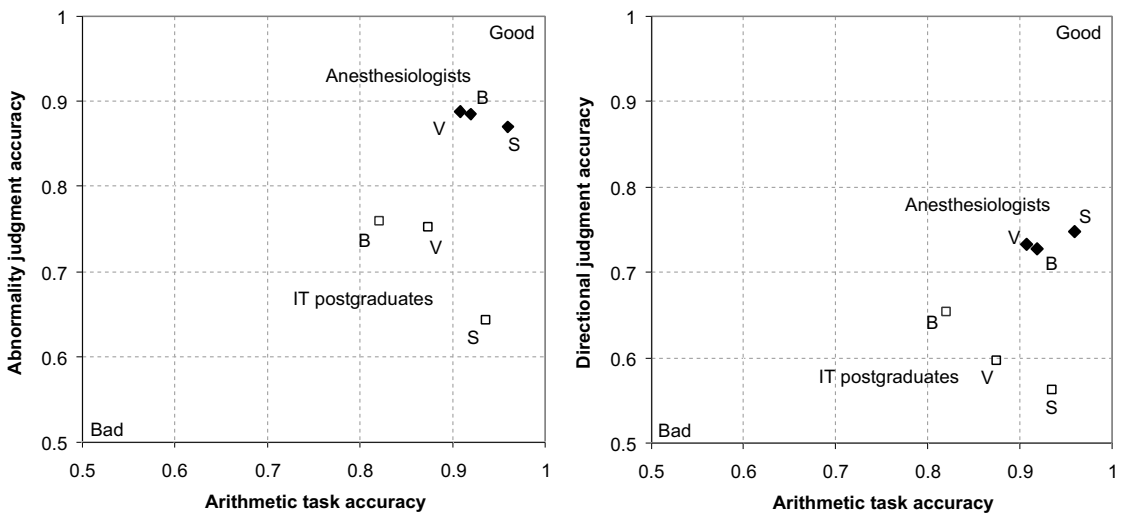


Figure 7. Experiment 3: Results for arithmetic and patient monitoring accuracy.

significant interaction between group and modality, $F(2, 38) = 2.46$, $MSE = 1.2$, $.10 > p > .05$, which indicated that there were no real performance differences among modalities for anesthesiologists but somewhat worse performance with the S condition for IT postgraduates (see Figure 8). Finally, there was a significant effect for parameter, $F(4, 76) = 8.08$, $MSE = 0.75$, $p < .001$, with HR showing best performance and V_T worst, regardless of modality. The marginally significant Group \times Parameter interaction, $F(4, 76) = 2.34$, $MSE = 0.75$, $.10 > p > .05$, indicated that differences in judgment accuracy between parameters were somewhat more marked for IT postgraduates as compared with anesthesiologists. No other effects were significant.

In an ANOVA similar to the one just described, direction judgments showed a significant main effect of group, $F(1, 19) = 34.03$, $MSE = 1.7$, $p < .001$, with anesthesiologists performing better (see Figures 7 and 9). There was also a significant effect of parameter, $F(4, 76) = 6.83$, $MSE = 1.21$, $p < .001$, with V_T

slightly less accurate in all modalities, not in only S. However, the effect of modality was not significant, and no other effects were significant.

Participants were asked to describe the clinical event they monitored in each scenario, and the results were analyzed in a between/within-subjects ANOVA, with the factors of group and modality. Answers were scored using a scheme determined independently with an anesthesia subject-matter expert. Not surprisingly, there was a significant effect of group, $F(1, 19) = 55.846$, $MSE = 1.166$, $p < .001$, with anesthesiologists performing far better (averaging 2.3 out of 3 points) than IT postgraduates (averaging 0.27 out of 3 points). No other effects were significant. This result confirms that anesthesiologists were not only better at discerning changes in physiological parameters but also could reasonably accurately identify the underlying clinical event.

The results suggest that anesthesiologists maintained patient monitoring performance at a high level, regardless of modality, but showed better performance on the arithmetic task with

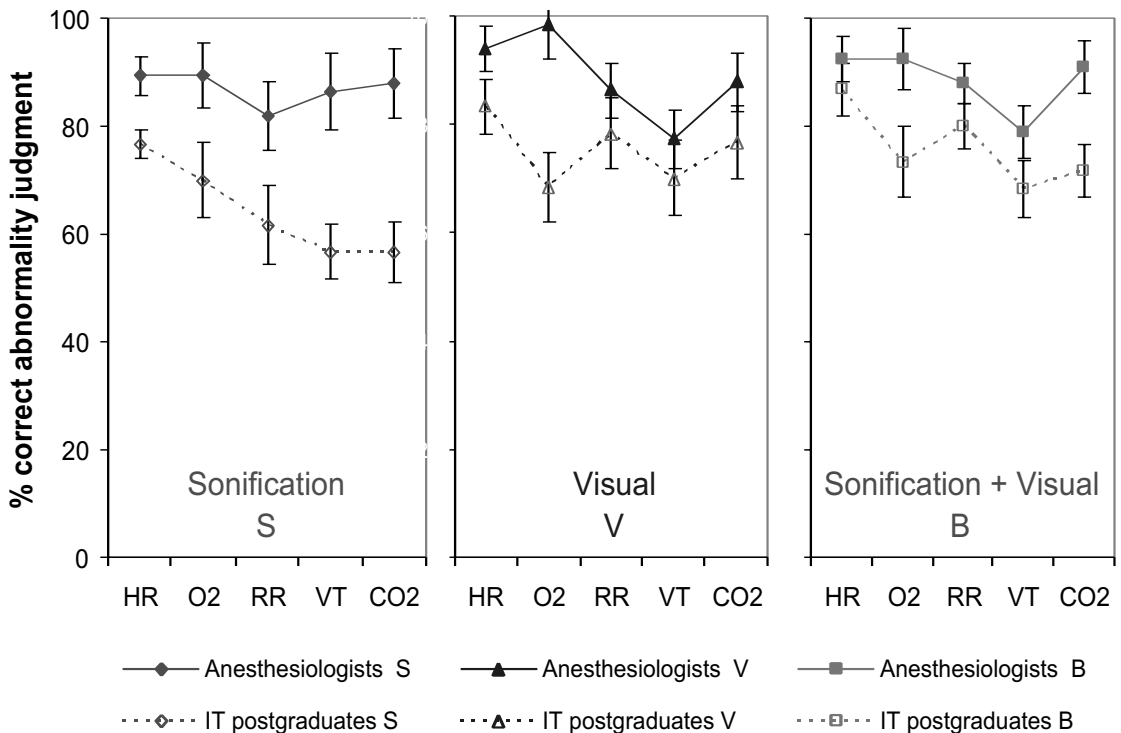


Figure 8. Experiment 3: Percentage correct abnormality judgments for the three experimental conditions.

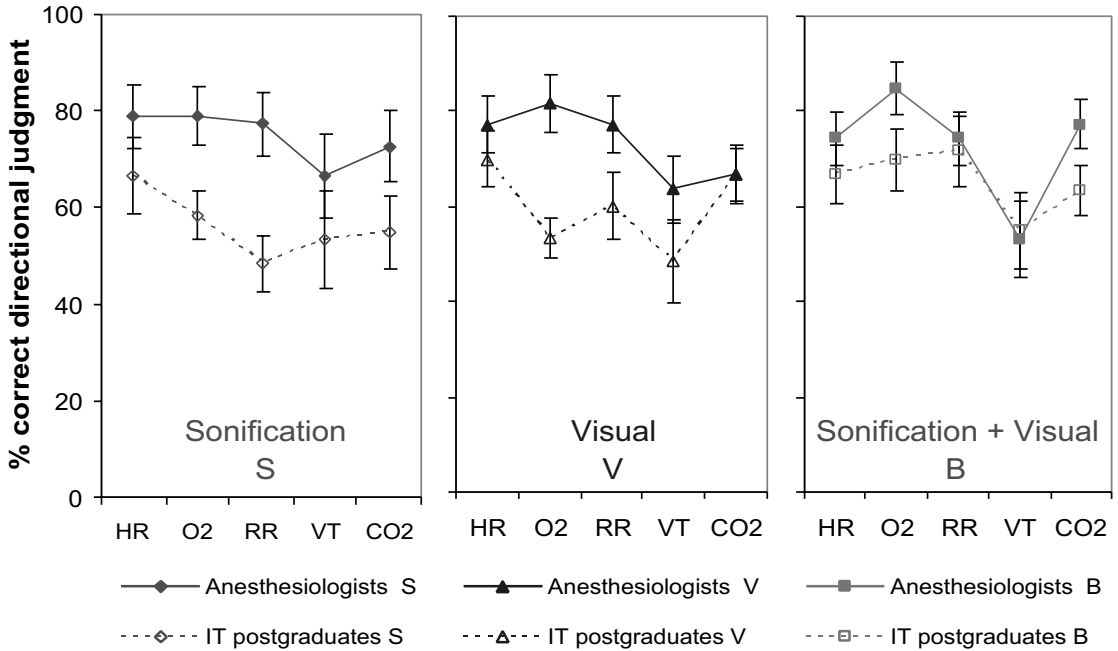


Figure 9. Experiment 3: Percentage correct direction judgments for the three experimental conditions.

sonification alone. In contrast, the IT postgraduates tended to let monitoring performance drop with sonification, trading off quality of monitoring for better arithmetic task performance.

Screen queries with withholding technique. We analyzed the touch screen queries to see if the fewest screen queries were in the B condition. Results of a three-way ANOVA showed no effect of group or modality. However, there was a strongly significant effect of parameter, $F(4, 76) = 23.04, MSE = 53.2, p < .0001$, indicating that screen queries ran from least to most in the following order: HR < O₂ < RR < V_T < ETCO₂. In addition, there was a marginal interaction of Modality × Parameter, $F(4, 76) = 2.12, MSE = 36.9, .10 > p > .05$, suggesting a trend toward fewer screen queries for HR and O₂ in the B condition. No other effects were significant.

Preferences. Participants’ confidence in their judgments was tested in a three-way ANOVA similar to the one just described. Results showed a significant effect of modality, $F(2, 38) = 4.499, MSE = 2.733, p < .05$, with B producing the greatest levels of confidence, S the lowest, and V in between. A significant effect for parameter, $F(4, 76) = 6.855, MSE = 0.626, p < .001$, indicat-

ed that participants had the greatest confidence in their O₂ judgments and the least in their RR judgments. No other effects were significant. Workload did not show any fully significant differences across conditions.

Finally, participants were asked which modality they found easiest to use and which imposed the highest workload; results are shown in Table 3. Anesthesiologists preferred the B condition even though many of them felt it imposed the highest workload of the three. The IT postgraduates had a mild preference for the S condition, apparently feeling that any modality that included the visual display led to high workload.

Discussion

Experiment 3 provided some insight into the relative effectiveness of sonification and visual displays for patient monitoring while other tasks are the focus of attention. The experiment also highlighted some important differences in performance between anesthesiologists and a nonanesthesiologist population.

Results for abnormal judgments showed that performance was equally good for the visual-only condition and the combined sonification plus visual condition and worse for the

TABLE 3: Experiment 3 Participants' Reports of Easiest and Highest-Workload Modalities

Question	Group	S	V	B
Easiest monitoring condition	Anesthesiologists	1	3	7
	IT postgraduates	5	2	3
Highest workload condition	Anesthesiologists	5	1	5
	IT postgraduates ^a	0	4	5

Note. For anesthesiologists, $N = 11$; for IT postgraduates, $N = 10$. S = sonification alone for all parameters; V = visual display alone for all parameters; B = both sonification and visual display for all parameters.

^aOne IT postgraduate spread workload ratings across the three categories, which has been discounted.

sonification condition, which was at variance with Hypothesis H3.1. Because the anesthesiologists showed equally good performance in all conditions, however, the selectively poorer sonification results for the IT postgraduates may indicate that they had not yet learned to map the sound to physiological meanings. The screen-querying data suggested that participants were adjusting their querying rate in line with the underlying event rate of the five different parameters: Cardiovascular changes were less frequent than respiratory changes, and heart rate and oxygenation were queried significantly less often than were respiratory parameters.

The combined sonification and visual display gave participants greater confidence than did sonification or a visual display alone, which was consistent with Hypothesis H3.2. Moreover, anesthesiologists rated the combined display as the easiest monitoring condition. Interestingly, the screen query results suggested that adding sonification to the visual display reduced participants' need to query heart rate and oxygenation but had no effect on querying the respiratory parameters. This may represent better calibration to the base rate of change on pulse oximetry parameters with sonification, greater stability of these parameters, or greater willingness to rely on the familiar pulse oximetry tone.

Workload showed no differences among modalities when it was rated during the experiment, but in their final workload ratings both groups rated the combined visual plus sonification condition as imposing the highest workload. Therefore Hypothesis H3.3 was only partially supported. Finally, anesthesiologists were better at providing a clinical assessment of the events in the scenario, supporting Hypothesis H3.4.

Overall, Experiment 3 showed that when a visual display is supplemented by sonification, anesthesiologists can maintain monitoring accuracy. Moreover, anesthesiologists performed significantly better on an arithmetic task when the monitoring information was sonified than they did under any other conditions. In contrast, for IT postgraduates, sonification led to relatively less accurate judgments of abnormality, probably because of their lack of physiological training. At the same time, sonification led to better performance by IT postgraduates on the arithmetic task, probably because no further attention to the patient monitoring task would have improved performance.

The V_T parameter was slightly less effective than the other sonified dimensions in Experiments 1 and 2, but in Experiment 3, in which attention was drawn to the arithmetic task, the disadvantage of V_T was exacerbated in conditions involving visual support (B and V). One reason is that in the scenarios used, V_T showed many small fluctuations. This possibly made V_T more challenging to monitor in any modality than if it had had a few large fluctuations. In contrast, HR had few large changes, and better judgment performance was seen. A further reason is that V_T was the only parameter for which three digits had to be read from the screen, possibly making it harder to monitor visually than other parameters.

Comparison with other studies. Although other studies have used different experimental designs and different dependent measures, our findings are consistent with them and help in part to integrate them. Specifically, the anesthesiologists in Loeb and Fitch's (2002) single-task study detected and identified events faster with

a combined visual and sonification display than with either display alone. In our Experiment 3, the anesthesiologist participants detected changes equally accurately in all modalities, probably because they were at a performance ceiling, but they expressed greatest confidence with the combined visual and sonification display.

The nonanesthesiologists in the Seagull et al. (2001) dual-task study showed a pattern of performance gratifyingly similar to that of our IT postgraduates (see Figure 7). The Seagull et al. participants detected physiological changes fastest with displays providing visual support (like our B and V conditions), but they achieved the best tracking scores with sonification alone, intermediate scores with the visual display alone, and the worst scores with the combined sonification and visual display (like our S, V, and B conditions).

These findings indicate that generalizable conclusions cannot always be drawn from experiments with nonanesthesiologist participants. Anesthesiologists' greater domain knowledge, greater experience at allocating attention to physiological information across modalities, and greater habit of preattentive reliance on sonification will all lead to qualitatively different use of multimodal information. For all participants, distractor tasks are performed better if monitoring is supported with sonification alone because the monitoring and arithmetic tasks are unambiguously supported by different input modalities (Wickens, 1976, 1984; Wickens, Sandry, & Vidulich, 1983). Nonanesthesiologists lack domain expertise, so some visual support is necessary if their monitoring performance is to improve (Seagull et al., 2001, and our Experiment 3). In contrast, for anesthesiologists, sonification alone does not hurt monitoring performance and sometimes may even be necessary for it (Loeb & Fitch, 2002, and our Experiment 3).

Our screen-querying data from Experiment 3 suggest that the expected frequency of change in a parameter may have driven the rate at which participants queried that parameter (Moray, 1986; Wickens & Gopher, 1977). The HR and O₂ parameters had lower event rates, and participants looked at them less frequently than they looked at RR, V_T, and ETCO₂. This interpretation is corroborated by the fact that in the

Seagull et al. (2001) experiment, HR and O₂ had higher event rates and eye-tracking data revealed that participants looked at them more frequently. In addition, when we added sonification to the visual display, queries were reduced principally to the HR and O₂ parameters. This is very similar to the findings by Seagull et al. that participants tended to look at the HR and O₂ parameters less with the combined sonification and visual display than with the visual display alone.

CONCLUSIONS

We have described three experiments that proceeded from a comparison of monitoring performance with three candidate respiratory sonifications, to a test of the most effective sonification with anesthesiologist participants, and finally to an evaluation of the most effective sonification in a simulated dual-task patient monitoring context. Taken together, the results of Experiments 1 and 2 show that with a minimal level of familiarization, participants can monitor respiratory parameters with a respiratory sonification as well as they can monitor heart rate and oxygenation status with pulse oximetry sonification. Experiment 3 shows that when anesthesiologists carry out a distracting task at the same time as monitoring, as is often the case in the operating room, the sonification helps them to time-share. However, instead of boosting monitoring performance, sonification allowed anesthesiologists' monitoring performance to be sustained at high levels while performance at the time-shared task improved. Our findings are consistent with those of other researchers in the area, despite differences in experimental design, procedure, kinds of sonification, and dependent measures.

Overall, our findings suggest that sonification should be viewed not as a substitute for visual displays but, instead, as a supplement to them. The combined visual display and sonification led to more confident judgments and was preferred by anesthesiologists as the easiest condition under which to monitor. Our anesthesiologist participants could distinguish sonified normal and abnormal readings of respiratory parameters to a level of accuracy comparable to their performance with pulse oximetry under both single-task (Experiment 2) and dual-task (Experiment 3) conditions.

Sonification may help anesthesiologists preserve a general awareness of patient state in situations in which they cannot check the visual display or prefer not to (Woods, 1995). Therefore it is possible that sonification could safely reduce the reliance on alarms for eyes-free monitoring in the operating room by providing continuous information (Seagull & Sanderson, 2001; Watson et al., 2004). We are working toward further trials of the respiratory sonification with anesthesiologists in settings closer to the clinical context, such as full-scale anesthesia simulators.

We have also investigated whether nonanesthesiologists' dependence on visual displays can be reversed by making it more convenient to rely on sonification for patient monitoring. Even when arithmetic tasks arrive much faster and the patient monitoring screen is directly behind participants, visual-only participants keep turning around to read the visual display (Sanderson, Crawford, Savill, & Watson, 2004). However, with self-paced perceptual-motor distractor tasks, a combined sonification and visual display supports patient monitoring as effectively as the visual-only condition, and it supports the distractor task as effectively as the sonification-only condition (Watson, Sanderson, Woodall, & Russell, 2003). Our research indicates that when the effectiveness of displays in different modalities is judged for patient monitoring, it must be judged not only with respect to task performance but also with respect to ecologically valid measures of attention and workload that reflect the structure of the environment.

These findings encourage us that principles of ecological interface design may be extended to sonification design (Sanderson et al., 2000). Sonification has the potential to convey physiological information to anesthesiologists in a way that is meaningful and informative and that conserves attentional demand. In the present studies we restricted our sonification to key physiological parameters available through current sensing technology. However, our anesthesia simulator opens the possibility of testing the effectiveness of novel sonifications of higher-level properties, such as oxygen transport. Even if the underlying sensing technology may not currently exist, better performance with a sonification of higher-order properties may

stimulate an investigation into how such information might be derived, as Reising and Sanderson (2002) have noted.

Sonification could be used in other monitoring environments where operators must share their visual attention among several tasks. Sonification could also support a team environment, even when the members' roles may be quite diverse, as long as team members share common tasks and goals. For example, Watson, Sanderson, and Anderson (2000) have adopted elements of ecological interface design to propose a sonification to support cockpit crews during aircraft landing approach.

In summary, sonification of patient physiology beyond traditional pulse oximetry appears to be a viable and useful adjunct when monitoring patient state. Sonification may help anesthesiologists maintain high levels of awareness of patient state and at the same time do other tasks more effectively than when relying on visual monitoring of patient state. Evidence suggests that sonification adds useful information to visual displays that allows participants to adjust their attentional strategy to achieve higher levels of confidence in their accuracy, despite higher workload. The respiratory sonification needs to be tested in a high-fidelity simulator or clinical environment for the usefulness of these findings to be fully established.

ACKNOWLEDGMENTS

We thank our three anonymous reviewers for their helpful and perceptive comments. We also gratefully acknowledge the collaboration and support of Dr. John Russell of Royal Adelaide Hospital and his colleagues. The respiratory sonification used in this study was originally developed between 1998 and 2001 by Watson and Sanderson when Watson was a Ph.D. student under the supervision of Sanderson at the Swinburne Computer Human Interaction Laboratory, Swinburne University of Technology, Melbourne, Australia. Financial support for this research was provided by an Australian Research Council (ARC) Small Research Grant through Swinburne University of Technology and a Swinburne University Research Development Grant. Support for the write-up of this research was provided by ARC Discovery Grant DP0209952 to Sanderson and Russell.

REFERENCES

- Barras, S., & Kramer, G. (1999). Using sonification. *Multimedia Systems*, 7, 25–31.
- Cook, R. I., & Woods, D. D. (1996). Adapting to new technologies in the operating room. *Human Factors*, 38, 595–615.
- Findlay, G. P., Spittal, M. J., & Radcliffe, J. J. (1998). The recognition of clinical incidents: Quantification of monitor effectiveness. *Anaesthesia*, 53, 589–605.
- Fitch, W. T. (1998). *U.S. Patent No. US5730140*. Washington, DC: U.S. Patent and Trademark Office.
- Fitch, W. T., & Kramer, G. (1994). Sonifying the body electric: Superiority of an auditory over a visual display in a complex, multivariate system. In G. Kramer (Ed.), *Auditory display: Sonification, audification and auditory interfaces* (pp. 307–325). Reading, MA: Addison-Wesley.
- Gaver, W. (1997). Auditory interfaces. In M. Helander, T. K. Landauer, & P. Prabhu (Eds.), *Handbook of human-computer interaction* (2nd ed., pp. 1003–1043). Amsterdam: North-Holland.
- Kramer, G. (1994). Some organizing principles for representing data with sound. In G. Kramer (Ed.), *Auditory display: Sonification, audification and auditory interfaces* (pp. 1–77). Reading, MA: Addison-Wesley.
- Loeb, R. G., & Fitch, W. T. (2002). A laboratory evaluation of an auditory display designed to enhance intraoperative monitoring. *Anesthesia and Analgesia*, 94, 362–368.
- Moray, N. (1986). Monitoring behavior and supervisory control. In K. Boff, L. Kaufman, & J. Thomas (Eds.), *Handbook of perception and human performance* (pp. 40/1–40/51). New York: Wiley Interscience.
- Nelson, W. T., Hettinger, L. J., Cunningham, J. A., Brickman, B. J., Hass, M. W., & McKinley, R. L. (1998). Effect of localized auditory information on visual target detection performance using a helmet-mounted display. *Human Factors*, 40, 452–458.
- Patterson, E. S., Watts-Perotti, J., & Woods, D. D. (1999). Voice loops as coordination aids in space shuttle mission control. *Computer Supported Cooperative Work*, 8, 553–571.
- Reising, D. C., & Sanderson, P. (2002). Work domain analysis and sensors: I. Principles and simple example. *International Journal of Human-Computer Studies*, 56, 569–596.
- Sanderson, P., Anderson, J., & Watson, M. (2000). Extending ecological interface design to auditory displays. In *Proceedings of the 10th Australasian Conference on Computer-Human Interaction* (OZCHI 2000, pp. 259–266). Los Alamitos, CA: IEEE Computer Society.
- Sanderson, P., Crawford, J., Savill, A., & Watson, M. (2004). Visual and auditory attention in patient monitoring: A formative analysis. *Cognition, Technology, and Work*, 63, 172–185.
- Sarter, N. B. (2000). The need for multi-sensory interfaces in support of effective attentional allocation in highly dynamic event-driven domains: The case of cockpit automation. *International Journal of Aviation Psychology*, 10, 231–245.
- Seagull, J. F., & Sanderson, P. M. (2001). Anesthesia alarms in context: An observational study. *Human Factors*, 43, 66–78.
- Seagull, J. F., Wickens, C. D., & Loeb, R. G. (2001). When less is more? Attention and workload in auditory, visual and redundant patient-monitoring conditions. In *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting* (pp. 1395–1399). Santa Monica, CA: Human Factors and Ergonomics Society.
- Vicente, K. J., & Rasmussen, J. (1990). The ecology of human-machine systems: II. Mediating “direct perception” in complex work domains. *Ecological Psychology*, 2, 207–249.
- Watson, M., Russell, W. J., & Sanderson, P. (2000). Anesthesia monitoring, alarm proliferation, and ecological interface design. *Australian Journal of Information Systems*, 7, 109–114.
- Watson, M., & Sanderson, P. (1998). Work domain analysis for the evaluation of human interaction with anaesthesia alarm systems. In *Proceedings of the Australian/New Zealand Conference on Computer-Human Interaction* (OZCHI98, pp. 228–235). Los Alamitos, CA: IEEE Computer Society.
- Watson, M., & Sanderson, P. (2001). Intelligibility of sonification for respiratory monitoring in anaesthesia. In *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting* (pp. 1293–1297). Santa Monica, CA: Human Factors and Ergonomics Society.
- Watson, M., Sanderson, P., & Anderson, J. (2000). Designing auditory displays for team environments [CD-ROM]. In *Proceedings of the 5th Australian Aviation Psychology Symposium* (AAVPA) [CD-ROM]. Sydney: Australian Aviation Psychology Association.
- Watson, M., Sanderson, P., & Russell, W. J. (2000). Alarm noise and end-user tailoring: The case for continuous auditory displays. In *Proceedings of the 5th International Conference on Human Interaction With Complex Systems* (HICS2000, pp. 75–70). Urbana-Champaign, IL: U.S. Army Research Laboratory.
- Watson, M., Sanderson, P., & Russell, W. J. (2004). Tailoring reveals information requirements: The case of anaesthesia alarms. *Interacting with Computers*, 16, 271–293.
- Watson, M., Sanderson, P., Woodall, J., & Russell, W. J. (2003, November). *Operating theatre patient monitoring: The effects of self paced distracter tasks and experimental control on sonification evaluations*. Presented at the 2003 Annual Conference of the Computer-Human Interaction Special Interest Group (CHISIG) of the Ergonomics Society of Australia (OZCHI2003), St. Lucia, Australia.
- Webb, R. K., van de Walt, J., Runciman, W. B., Williamson, J. A., Cockings, J., Russell, W. J., et al. (1995). Which monitor? An analysis of 2000 incident reports. *Anaesthesia and Intensive Care*, 21, 529–542.
- Wickens, C. D. (1976). The effects of divided attention on information processing in tracking. *Journal of Experimental Psychology: Human Perception and Performance*, 2, 1–13.
- Wickens, C. D. (1984). Processing resources in attention. In R. Parasuraman & R. Davies (Eds.), *Varieties of attention* (pp. 63–102). New York: Academic.
- Wickens, C. D., & Gopher, D. (1977). Control theory measures of tracking as indices of attention allocation strategies. *Human Factors*, 19, 249–366.
- Wickens, C. D., Sandry, D. L., & Vidulich, M. (1983). Compatibility and resource competition between modalities of input, central processing, and output. *Human Factors*, 25, 227–248.
- Woods, D. D. (1995). The alarm problem and direct attention in dynamic fault management. *Ergonomics*, 38, 2371–2393.
- Xiao, Y., Mackenzie, C. F., Seagull, J. F., & Jaber, M. (2000). Managing the monitors: An analysis of alarm silencing activities during an anesthetic procedure. In *Proceedings of the XIVth Triennial Congress of the International Ergonomics Association and 44th Annual Meeting of the Human Factors and Ergonomics Society* (pp. 4.250–4.253). Santa Monica, CA: Human Factors and Ergonomics Society.

Marcus Watson is a research fellow at the Australian Research Council Key Centre for Human Factors and Applied Cognitive Psychology at the University of Queensland, St. Lucia, Australia. He received his Ph.D. in information technology from Swinburne University of Technology in 2002.

Penelope M. Sanderson is a professor in the Australian Research Council Key Centre for Human Factors and Applied Cognitive Psychology at the University of Queensland, St. Lucia, Australia. She received her Ph.D. in psychology in 1985 at the University of Toronto.

Date received: October 29, 2002

Date accepted: December 30, 2003