

Penelope Sanderson · Jennifer Crawford · Annyck Savill
Marcus Watson · W. John Russell

Visual and auditory attention in patient monitoring: a formative analysis

Received: 19 February 2003 / Accepted: 19 January 2004 / Published online: 20 April 2004
© Springer-Verlag London Limited 2004

Abstract We investigate the effectiveness of sonification (continuous auditory display) for supporting patient monitoring while reducing visual attentional workload in the operating theatre. Non-anaesthetist participants performed a simple continuous arithmetic task while monitoring the status of a simulated anaesthetised patient, reporting the status of vital signs when asked. Patient data were available either on a monitoring screen behind the participant, or were partially or completely sonified. Video captured when, how often and for how long participants turned to look at the screen. Participants gave the most accurate responses with visual displays, the fastest responses with sonification and the slowest responses when sonification was added to visual displays. A formative analysis identifying the constraints under which participants timeshare the arithmetic and monitoring tasks provided a context for interpreting the video data. It is evident from the pattern of their visual attention that participants are sensitive to events with different but overlapping temporal rhythms.

Keywords Anaesthesia · Formative analysis · Sonification · Video analysis

1 Introduction

Over the last decade, several researchers have suggested that *sonification*—the representation of relations in data by relations in sounds—might help anaesthetists monitor patients in the operating room (Fitch and Kramer

1994; Watson et al. 1999; Loeb and Fitch 2000, 2002; Watson and Sanderson 2004; Seagull et al. 2001). A sonification of basic patient vital signs could let the anaesthetist engage in “eyes-free” patient monitoring while doing other tasks, so extending the benefits of the pulse oximetry sonification that is already in the operating room. In this study we examine the “eyes-free” benefit of sonification by supplementing a comparison of different modalities of presentation for patient vital signs with a video analysis of participant visual attention.

Researchers have developed patient monitoring sonifications using various combinations of heart rate (HR), oxygenation (O₂), blood pressure (BP), respiration rate (RR), tidal volume (VT) and end-tidal carbon dioxide (ETCO₂), amongst others. Fitch and Kramer (1994) and Loeb and Fitch (2000, 2002) have demonstrated that listeners can detect and identify anaesthesia problems with sonifications of six or eight vital signs. Effectively designed sonification could even lessen the dependence on traditional alarms (Seagull and Sanderson 2001; Woods 1995; Watson et al. 2004).

Tasks often compete for the anaesthetist’s attention in the operating room (OR). To date, Watson and Sanderson (2004) and Seagull et al. (2001) are the only researchers to have examined patient monitoring performance with sonification under dual task conditions. Seagull et al. (2001) found that non-medically qualified participants performed a visually presented manual tracking task better when they were monitoring with a sonification than with a visual display or combined visual/sonification display. Regardless of whether they were also tracking, however, the participants could detect changes in patient vital signs faster with the visual display than with the sonification. In contrast, Watson and Sanderson (2004) demonstrated that sonification let anaesthetists retain a high patient monitoring accuracy with a slight performance benefit for a timeshared arithmetic task. A non-anaesthetist group, however, appeared to trade off performance between patient monitoring and timeshared tasks, much as Seagull et al.’s (2001) participants had. These results presumably

P. Sanderson (✉) · J. Crawford · A. Savill · M. Watson
ARC Key Centre for Human Factors and Applied Cognitive
Psychology, The University of Queensland,
St Lucia, QLD, 4072, Australia
E-mail: psanderson@humanfactors.uq.edu.au

W. J. Russell
Department of Anaesthesia and Intensive Care,
The University of Adelaide, Adelaide, SA,
5005, Australia

reflect domain expertise, and also perhaps a well-developed timesharing ability in anaesthetists (Gopher 1993).

The present study examines three unresolved issues with the aid of video data. Firstly, in both the Seagull et al. (2001) and Watson and Sanderson (2004) studies the patient monitoring information and the timeshared task were presented on the same display monitor. This is not how tasks are laid out in the OR. During induction and emergence anaesthetists pay relatively little attention to the patient monitoring screen (Loeb 1993, 1994) and when they do they often must peer sideways or backwards at it from where they are standing. In the present study, therefore, participants performed an arithmetic task on a screen in front of them but had to turn their heads to view the patient monitoring screen (see Fig. 1).

A second issue is that in the Watson and Sanderson (2004) studies, participants touched the screen to see patient vital signs. New arithmetic tasks arrived at 10 second intervals, which gave participants ample time to investigate all vital signs before the next arithmetic task arrived. This was a very conservative test of sonification against the visual display. Therefore, we reduced the arrival time of the arithmetic task to 5.0 seconds in one condition and 2.5 seconds in a further condition in an attempt to build a greater reliance on the sonifications. We hypothesised that under these conditions a sonification of both pulse oximetry and respiratory parameters would lead to patient monitoring performance at least as good as—and possibly better than—when visual monitoring is required.

A third issue arising from the Watson and Sanderson (2004) studies was whether monitoring both the pulse oximetry and respiratory sonifications might lead to worse performance than when monitoring one sonification only. This was examined for pulse oximetry

parameters HR and O_2 by comparing the participants' performance at making judgments about HR and O_2 when only HR and O_2 were sonified with participant's performance when all parameters were sonified. It was also examined for respiratory parameters RR, VT and $ETCO_2$ by comparing participants' performance at making judgements about RR, VT and $ETCO_2$ when only they were sonified with performance when all parameters were sonified.

Video analysis We conjectured that headturning could be a powerful measure of the need for visual information under the conditions described above, without the expense and complication of eyetracking data. As Fig. 1 shows, the physical arrangement of the experiment made it very obvious when the participant was attending to the patient monitoring screen. The fact that we included conditions with pulse oximetry sonification and respiratory parameters presented visually only, and vice versa, let us discriminate situations in which visual pulse oximetry information was sought from situations in which visual respiratory information was sought.

As is well known, video data are very rich. Some of the philosophy and methods of exploratory sequential data analysis (ESDA: Sanderson and Fisher 1994, 1997) can help us extract the most information from video data. Deciding upon an appropriate frequency band or bands at which to seek regularities in the data is an important ESDA activity. Finding an appropriate visualisation that allows regularities (the "smooth") and deviations from regularities (the "rough") to be easily seen at such frequency bands is an ESDA technique inherited from exploratory data analysis (EDA: Tukey 1977). To help us find the right frequency band and the right visualisation, we performed a *formative* analysis of the constraints operating in the experiment, and analysed with respect to those. By a formative analysis, we refer to an analysis that identifies requirements that must be met for successful performance and that will constrain behaviour but not prescribe behaviour (Sanderson 1998; Vicente 1999). Because our experiment imposed

Fig. 1 A typical room layout for the experiments reported. In the first picture, the participant attends to the arithmetic task. In the second picture the participant turns to look at the patient monitoring screen. Images are representative of the video from which headturning was analysed



multiple competing demands on participants, who then had to find an acceptable way of managing those demands, we felt a formative analysis would be useful. Our particular focus was on how the periodicities built into the experiment might be reflected in periodicities in participant activity. In this way a formative analysis can show the level at which apparently quite disparate behaviour is similar.

Periodicities and entrainment There is temporal patterning in the anaesthetist's work. Apart from the anaesthesia phases of induction, maintenance and emergence, there are expectations about the timeframe of response of the human body to surgical and pharmacological interventions (Sougné et al. 1993). Rhythms of monitoring may be guided by the rate at which information about a vital sign becomes unacceptably uncertain and must be resampled (Moray 1986). Rhythms of charting may be constrained by professional and organisational requirements. There will also be interruptions to rhythmic activity. The anaesthetist may take phone calls, answer pagers, request further supplies and respond to surgical contingencies, after which background patterns of monitoring can be re-established.

As an example of entrainment, during an exploratory observational study of anaesthetists' activity Seagull (1996) recorded the time elapsed between anaesthetists' episodes of writing on the patient's intraoperative chart. (A record of patient vital signs at 5 minute intervals must be maintained in the OR, which for manual charting normally requires the anaesthetist's attention every five minutes.) Seagull's (1996) inter-charting interval data suggested that apart from near-contiguous writing episodes, for long cases such as cardiac and cranial surgery the most frequent inter-charting interval was indeed five minutes. However, there was a distinct further peak at the inter-charting interval of nine minutes and some evidence of a further small peak at 15 minutes for cranial cases. These results are based on small samples only, but they suggested that there may have been periodic events in the work context, such as the schedule for inflation of the NIBP cuff, that entrained monitoring and charting activities to arrive at time intervals that were approximate multiples of five minutes.

Anaesthetists often deliberately exploit properties of their physical world ("tailoring") to help them maintain vigilance and remember planned actions without undue workload (Cook and Woods 1996; Watson et al. 2004). However, we have found that anaesthetists may sometimes be unaware of environmental factors that are entraining their attention and activity (Watson 2002). In one case, a senior anaesthetist claimed he did not look regularly at the visual patient monitoring system, preferring instead to watch the patient. However, our video records revealed that this anaesthetist quickly checked the patient monitor approximately every two minutes.

Although the timeshared tasks in our study are artificial, as the abovementioned field examples indicate it is normal for the anaesthetist to be performing tasks that compete with patient monitoring and for the competing tasks to have a periodic structure.

Goal of the study The goal of this study was to examine patient monitoring performance with different levels of sonification support under more challenging dual task conditions than we had observed in the past. By performing a video analysis of headturning we hoped to determine how much sonification guided visual attention but also reduced the need for visual support during normal monitoring. We also wished to learn whether headturning would become entrained to periodicities in the patient monitoring environment, as constructed in this experiment.

2 Conduct of the study

2.1 Participants

The study was run at the University of Queensland Usability Laboratory (UQUL). Participants were 40 students, either senior undergraduate students or postgraduate students at the university, who had no experience of clinical monitoring and no medical or nursing background. They were paid for their participation.

2.2 Apparatus

The patient monitoring task was presented with the Arbiter simulation, which is a flexible display front end for the Body Anesthesia Simulator. Arbiter ran on a Dell workstation, with a 19-inch LCD display. The arithmetic task was programmed in Microsoft Visual Basic and was presented on a laptop. The Arbiter simulation communicated with the Microsoft Visual Basic application to stop and start the arithmetic task as well as the simulation when the participant was prompted for his or her awareness of patient state (see Fig. 2 for the two displays).

The visual display showed HR at the top of the screen and O₂ at the bottom. RR, VT and ETCO₂ were shown in a quite tight cluster between HR and O₂. Waveforms were shown for O₂ and ETCO₂. The auditory display involved two sound streams: (1) the conventional pulse oximetry "beep" that gave HR with its rate and O₂ with its pitch, and (2) the respiratory sonification described in Watson and Sanderson (2004). In the latter, RR was displayed as an inspired:expired pure tone sound. The inspired tone is the upper tone of a musical third interval, and the expired tone is the lower tone of the interval. The relative pitch of the tone pair denoted ETCO₂ concentration. The rise and fall of sound intensity indicated tidal flow so that the integration of sound volume over time denoted tidal volume.

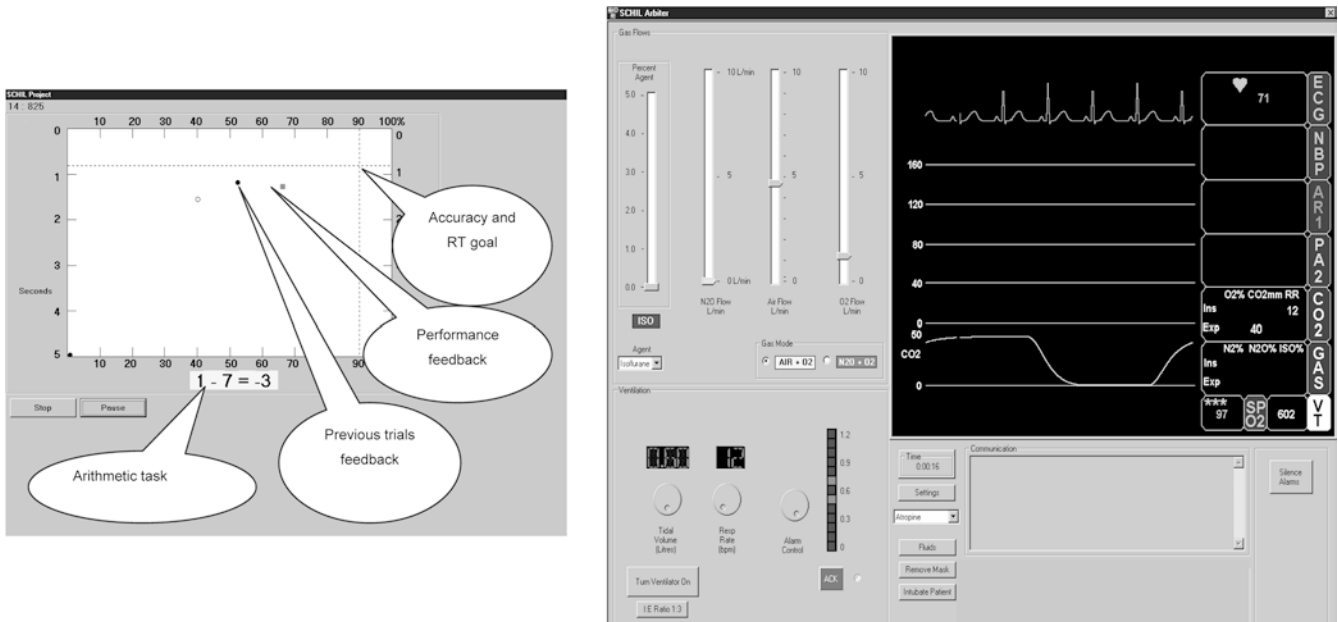


Fig. 2 At the left is the arithmetic task (callouts are not seen in the actual experiment) and at the right is the Arbiter simulation screen

Experimental sessions were videotaped. The MacSHAPA video analysis tool (Sanderson et al. 1994) was subsequently used to identify and timecode the start and end of headturns and the latency of verbal responses to probes. Outputs from MacSHAPA were integrated in an Excel spreadsheet for aggregate analysis.

2.3 Design of the experiment

The participants' patient monitoring performance was studied on a between-subjects basis, under different levels of timeshared task workload and different display modalities. The five display conditions were as follows:

- Sonification (SS condition). Participants heard sonifications of pulse oximetry and respiratory parameters (the SS condition). No patient information appeared on the Arbiter computer screen.
- Visual display of the pulse oximetry and respiratory parameters (the VV condition). Participants had to turn their heads to see the patient's physiological status.
- Visual display of the pulse oximetry plus visualisation of respiratory parameters (SV condition). Participants had to turn their heads to see respiratory status.
- Visual display of pulse oximetry plus the sonification of respiratory parameters (the VS condition). Participants had to turn their heads to see pulse oximetry status.
- Both visual displays and sonification (BB condition). Participants could attend to either or both modalities.

The arithmetic task arrived at either 2.5 second intervals (fast) or 5.0 second intervals (slow).

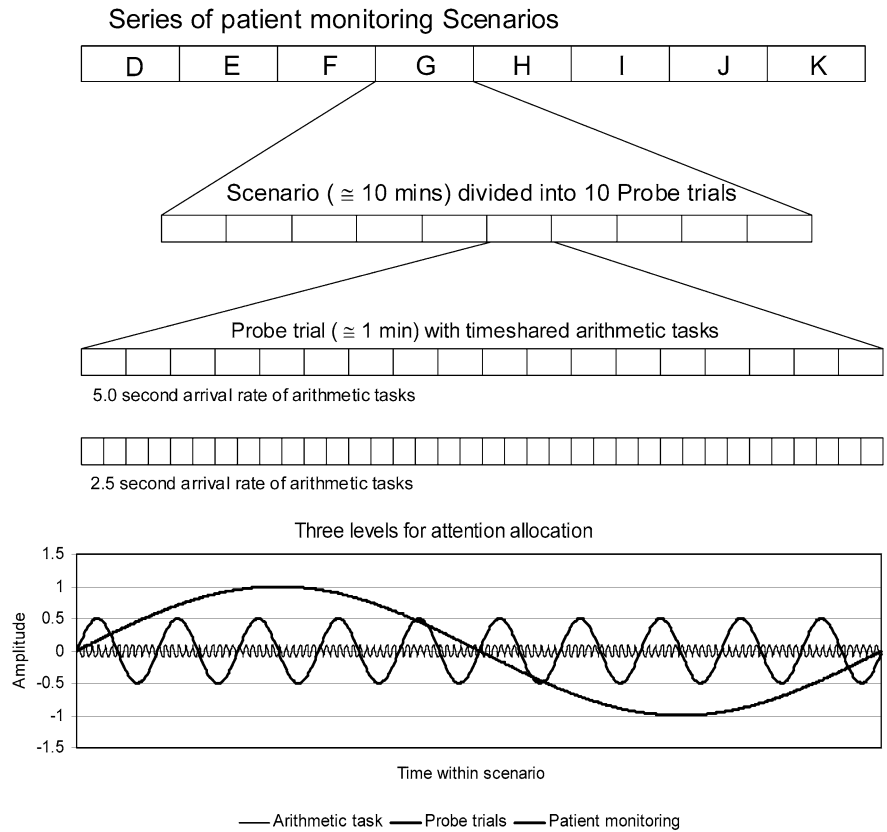
In the main part of the experiment, participants monitored a series of simulated anaesthesia scenarios (D through K) that were variants of those developed by Watson (2002). The scenarios included blood loss, breathing against the ventilator, the ventilator not being turned on and so on. All participants experienced all scenarios, although the order of presentation of the DEF and GHI clusters of scenarios was counterbalanced across participants. Scenarios J and K were control conditions mixed in with the main clusters, in which there was very little change to the patient and no major events. In the present experiment participants could not perform interventions on the simulated patient.

Figure 3 shows the structure of the experiment, but without the counterbalancing of scenarios. Each of the eight scenarios runs for around 10 minutes. Each scenario is divided into periods of 60 ± 5 seconds at which point a voice asks participants to report the status and direction of any change in one of the five physiological variables being monitored (the "probe").

2.4 Procedure

Participants were positioned as shown in Fig. 1. In front of them was a laptop computer on which they performed the arithmetic task. They were required to judge whether simple arithmetic expressions (such as $1+5=7$ or $2-8=-6$) were true or false and to register their response with a keypress. After every arithmetic trial, updated performance feedback was available on a speed-accuracy graph on the laptop screen. At the same time, participants performed the patient monitoring task using the Arbiter interface (Watson and Sanderson 2004) located 180° behind them (see Fig. 1). They gave their answers verbally. Participants received thorough

Fig. 3 Upper graph shows hierarchical nesting of events within the experimental design. Lower graph shows a stylised representation of the periodicities of factors competing for attention in the experimental design



training in both the patient monitoring task and the arithmetic task, and both together, for about an hour in all conditions before proceeding to the main part of the experiment.

Monitoring was broken up by rest pauses where needed, and also by the need to complete pen-and-paper questionnaires about monitoring confidence and workload. The entire session was videotaped. The experimenter sat with the participant in order to set up each scenario and was present throughout the experiment to note the participant's answers to probe questions.

3 Analysis of patient monitoring

In this section, we present the basic findings relating to the effectiveness of respiratory sonification. We then inspect the video data more closely, propose a formative model based on the constraints of the experimental demands and analyse headturning behaviour in this context.

3.1 An analysis of participant data

Figures 4 and 5 provide overviews of the experimental results, integrating either the accuracy (Fig. 4) or speed (Fig. 5) with which participants answered questions about the status of patient vital signs, their reaction time at the arithmetic task and the amount of time they spent with their head turned away from the arithmetic task

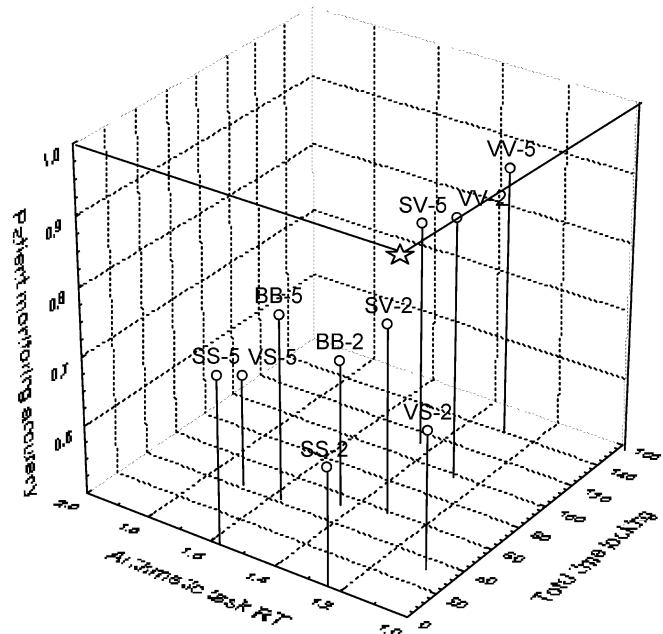


Fig. 4 Basic pattern of the experimental results covering accuracy of patient monitoring (Y axis), arithmetic task speed (X axis) and the total amount of time per trial looking at the patient monitoring screen (Z axis). Ideal performance is at the near upper corner marked by a star. Clearly, no condition in the present experiment approaches it

towards the patient monitoring screen. No condition approaches the ideal (marked by a star on the upper near vertex) when patient monitoring accuracy was

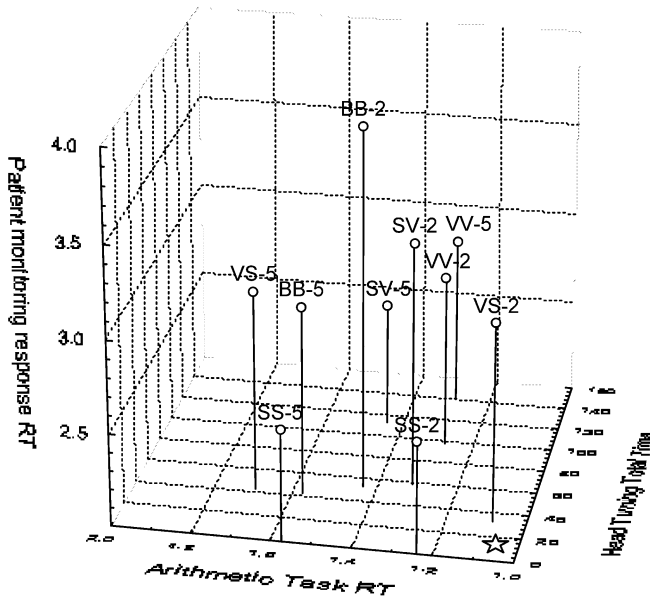


Fig. 5 Basic pattern of experimental results covering speed of response to probes about patient status (Y axis), arithmetic task speed (X axis) and the total amount of time per trial looking at the patient monitoring screen (Z axis). Ideal performance is at the lower right corner marked by a star and some conditions approach it

considered, whereas the SS condition approached the ideal when patient monitoring speed was considered (star on lower near vertex). Respiratory sonification eased visual monitoring demands and led to faster arithmetic task performance and faster responses to probes about patient vital signs, but the accuracy of those responses tended to be worse. We will have a closer look at these findings.

Patient monitoring performance Participant's accuracy at reporting the status of patient vital signs ranged from greater than 95% to less than 50% and most responses were between 2.5 and 3.0 seconds long. Overall, participants presented with sonified respiratory information (BB, SS, and VS) were least accurate and those presented with visual respiratory information (VV, SV) were most accurate, $F(4, 30)=9.8$, $MSE=0.054$, $p<0.001$. Experimenter recordings backed by video analysis indicated that participants in the SS conditions were fastest to respond to probes and those in the BB conditions slowest, $F(4, 30)=2.74$, $MSE=2.92$, $p<0.05$. Moreover, responding to probes was marginally slower during a fast rather than a slow presentation of the arithmetic tasks, even though the arithmetic task was suspended while participants dealt with the probe, $F(1, 30)=3.39$, $MSE=2.92$, $p=0.075$. There may be a response bottleneck for BB participants after time-pressured performance.

Participants responded more slowly, $F(4, 120)=13.98$, $MSE=0.45$, $p<0.001$, and less accurately, $F(4, 120)=24.86$, $MSE=0.026$, $p<0.001$, to questions about respiratory parameters RR, VT, and $ETCO_2$ than to

pulse oximetry parameters HR and O_2 . When respiratory parameters were sonified, accuracy was lowest, $F(16, 120)=3.02$, $MSE=0.026$, $p<0.001$ —especially for the GHI scenarios which had strong respiratory involvement, $F(16, 120)=3.02$, $MSE=0.019$, $p<0.001$ —and responses were longest, $F(16, 120)=5.61$, $MSE=0.026$, $p<0.001$. The reasons will be discussed shortly.

Arithmetic task performance We had told participants that the arithmetic task was their primary responsibility. Most participants were able to keep their accuracy around or above 90% so there were no significant differences in accuracy across conditions. Participants responded faster when the arithmetic task arrived at a faster rate, $F(1, 30)=9.19$, $MSE=0.131$, $p<0.01$ (see Figs. 4 and 5).

3.2 Video analysis of visual attention

In this section, we outline how the video data were analysed and related to the purpose of the experiment. We sought systematic results at three levels to answer the following questions.

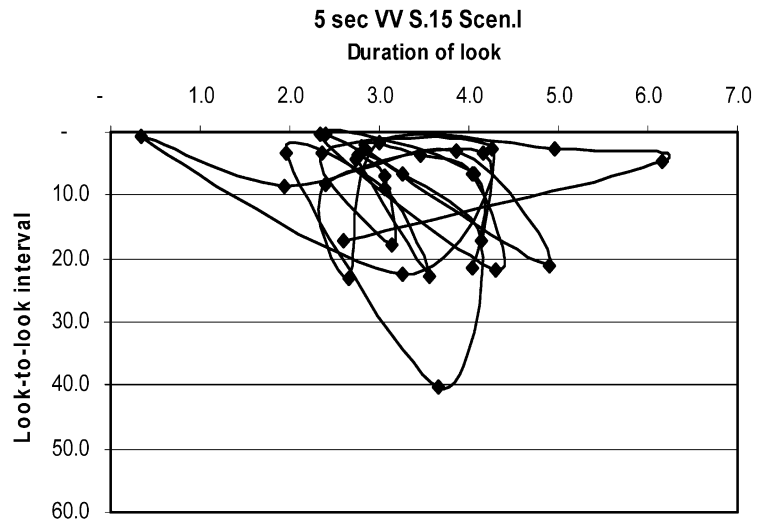
Arithmetic distractor task Do different arrival rates for arithmetic tasks successfully manipulate visual attention towards and away from the patient monitoring screen?

Probes during scenarios Do participants change their visual search pattern across the period between probes, and does this interact with modality?

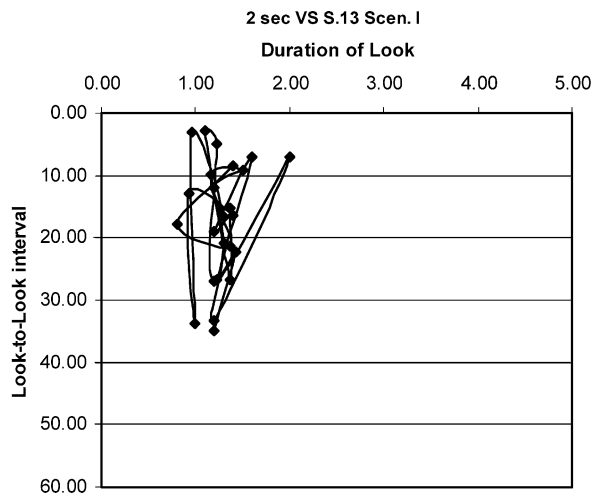
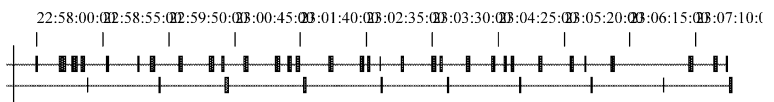
Physiological changes How sensitive are participants to changes in patient physiology during monitoring and do they change their monitoring strategy in response to what is seen and heard? Does this change with modality?

Figure 6 shows some relatively low-level data from the video analysis in which the strategies of two participants are contrasted. In each case, the MacSHAPA timeline shows the timing and duration of participants' responses to probes in the lower line and their looks at the monitor ("headturning") in the upper line. The accompanying state transition diagrams plot the durations of successive headturns against the time interval prior to each look. For example, the MacSHAPA timeline shows that Participant 15 (VV 5.0 second condition) made a lot of headturns towards the end of the probe interval. The upper state transition diagram indicates that he kept most responses from 2.0 to 4.0 seconds in length and most look-to-look intervals at either around 20 seconds or fewer than 10 seconds. Participant 13 (VS 2.5-second condition) had headturn durations that were centred around 1 second, and his look-to-look intervals ranged evenly from around 5–35 seconds. Participants showed a wide variety of patterns, yet were generally consistent with respect to

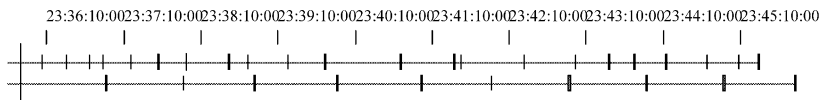
Fig. 6 A state transition graph and MacSHAPA timeline graph from two participants (*upper* participant 15, VV, 5.0-s [slow] arithmetic task; *lower* participant 13, VS, 2.5-s [fast] arithmetic task). The *timeline* graph shows a sequential pattern of headturns, where each node shows the time elapsed before each look and the duration of each look (*lower line* shows probes; *upper line* shows headturns)



codes Timeline



Codes Timeline



their own pattern. To fully understand these results, we need a formative analysis of the constraints underlying behaviour at each level of the task, which is presented in the next sections.

3.3 Impact of the arithmetic task on visual attention

The arithmetic task is the highest frequency event shown in the diagram at the bottom of Fig. 3. Participants'

ability to report the status of patient vital signs will improve the longer they are able to attend to an information source until they have assimilated all the data necessary to successfully answer the probe. When vital signs are only presented visually, such episodes will map directly onto overt looks at the screen.

Formative model Figure 7a shows how the frequency and duration of looks combine to produce total time looking. The parts of this surface on which participants

will travel will depend on how long they need to look to get the information they need, given the display tools at their disposal, and how they wish to schedule looks to track trends and avoid missing sudden changes. When the data are presented partially or totally through a sonification (SS, SV, or VS) attending may be covert, which may become evident in a drop in the relative lack of overt looks at the screen and less total time looking.

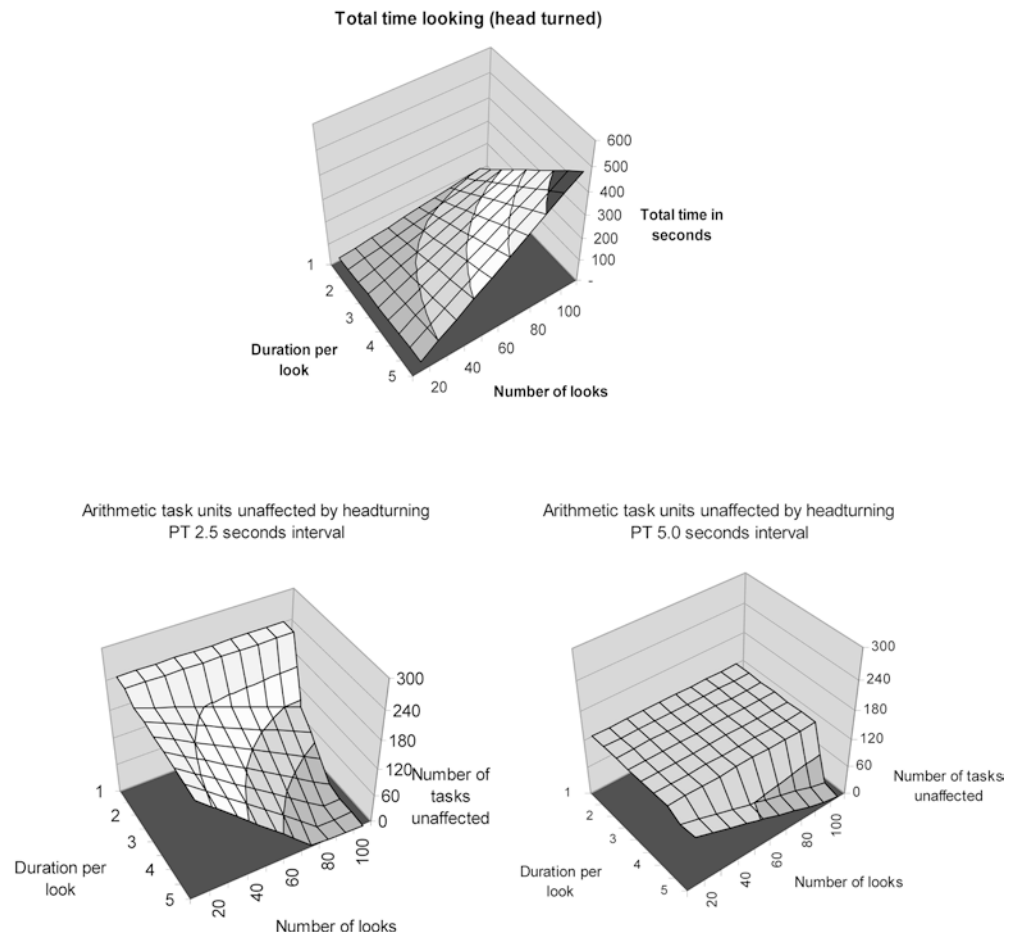
In our experiment the participant must maintain good arithmetic task performance while monitoring the patient. If the participant looks at the patient monitoring screen too long, he or she misses responding to an arithmetic task. One solution is to perform the arithmetic task and then, in the time left before the next arithmetic task arrives, to take a quick look at the patient monitoring screen. A second solution is to look at the patient monitoring screen only as often as is required to report on any abnormality or direction of change. The participant would perform the arithmetic task as well as possible for several trials, and would then take a somewhat longer look at the screen.

A response surface gives an overview of the tradeoffs (see Figs. 7b and 6c). Figure 7b shows the case where arithmetic tasks arrive every 2.5 seconds—240 tasks in a 10-minute scenario. If the participant is to have the best chance of scoring 100% correct, he or she must answer

every arithmetic task. Therefore his or her looks at the patient monitoring screen must always be quite short to leave time to answer the arithmetic task. If we assume that it takes about 1.0 seconds to make a true/false judgement about a simple arithmetic expression, then there are 1.5 seconds left to turn around and take a quick look at the patient monitoring screen. If the participant achieves this, then the number of arithmetic trials completed will remain at 240 and only inherent error will reduce arithmetic task accuracy from 100%. We would normally expect relatively accurate, fast performance with this strategy.

If participants lengthen their looks at the patient monitoring screen, then they may start to miss arithmetic tasks. If looks are really long, then possibly two or even three arithmetic tasks in a row are missed. If there are many long looks, then the potential for high accuracy on the arithmetic task drops off drastically. This “cliff” can be seen in Fig. 7b. Long looks at the patient monitoring screen only pay off if there are very few of those looks, as can be seen with the return of the potential for high accuracy on the opposite side of the response surface. Anything in between must reduce the opportunity for a high score. Figure 7c shows the equivalent response surface for participants for whom the arithmetic task arrives every 5.0 seconds. Clearly the constraints are less

Fig. 7 The upper response surface shows how the duration per look and the number of looks combine to produce the total time looking at the patient monitoring screen. Two lower response surfaces show how the number and the duration of looks at the patient monitoring screen combine to affect performance on the arithmetic task. The *left figure* shows results for the arithmetic task arrival rate of 2.5 seconds (fast); the *right figure* for 5.0 seconds (slow)



stringent. However, once looks at the patient monitoring system threaten the accuracy of arithmetic tasks, a single failure leads to the same decrease in arithmetic tasks affected as for the faster rate.

Putting Fig. 7a together with either Figs. 7b or 6c, then, it is clear that participants attuned to the needs of the arithmetic task should avoid the region in Fig. 7a where the amount of time spent looking at the patient monitoring screen places a participant well down the “cliff” of effective arithmetic task performance in Figs. 7b or 6c. Figure 7 is therefore a formative model of performance.

Results Figure 8 shows the results for individual participants in the four conditions in which there was headturning data (participants in the SS condition never experienced a visual display and so did not turn their heads). Dotted lines indicate important general aspects of the data with respect to the constraints of the formative model described above. Firstly, average headturn duration seldom went below one second (it averaged 1.3 seconds for the 2.5 second arithmetic task and 1.6 seconds for the 5.0 second arithmetic task). Secondly, there were usually at least 10 headturns per scenario, representing at least one headturn per probe. Thirdly, there is a void region in the centre where the participants’ pattern of responding does not correspond to the areas identified in Fig. 7 in which suboptimal performance would be expected. Fourthly, patterns of responding cluster around the two principal ridges shown in Fig. 7b, with the participants using the slower arithmetic task able to migrate slightly closer to the void region.

We can now examine how data are clustered according to experimental group. Firstly, participants in conditions in which some of the patient parameters were sonified generally spent far less time looking at the patient monitoring screen compared with the VV condition, $F(3, 24)=9.6$, $MSE=2308.2$, $p<0.001$. The reduction was particularly strong when respiratory parameters were sonified (see also Figs. 4 and 5). All cases to the right of the diagonal line in Fig. 8 include visually presented respiratory parameters; no cases include respiratory sonification. This line falls approximately along a band of Fig. 8 where total time looking per scenario is around 70–80 seconds. Sonification of RR, VT and ETCO₂ would have had the greatest effect in reducing the total time needed to look, firstly because three rather than two (HR and O₂) parameters are sonified and secondly because many of our scenarios involved complex respiratory events. VV participants, and SV participants doing the fast arithmetic task, made many more looks, $F(3, 24)=5.05$, $MSE=2656.1$, $p<0.01$, but BB participants made longer looks, $F(3, 23)=4.32$, $MSE=2.96$, $p<0.05$.

Secondly, participants in the VV conditions settled on strategies that either involved many looks of moderate length at the patient monitoring screen, or very few looks which were somewhat longer. The two VV participants closest to the void in the centre of the figure, participants 15 and 36, concentrated on arithmetic task performance after a probe, and ramped up their looking only as the next probe approached. This way they were able to quickly give an accurate patient status report without incurring the cost of spending too much time

Fig. 8 The average duration of headturns and the frequency of headturns within a scenario for all participants in the VV, VS, SV, and BB conditions

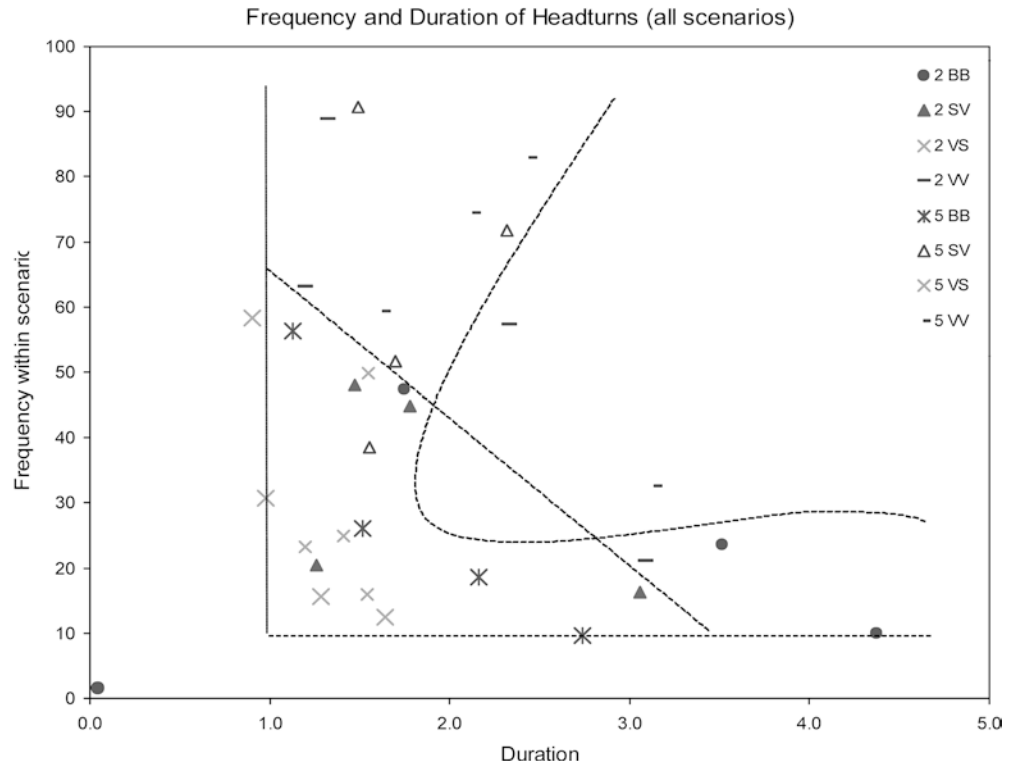
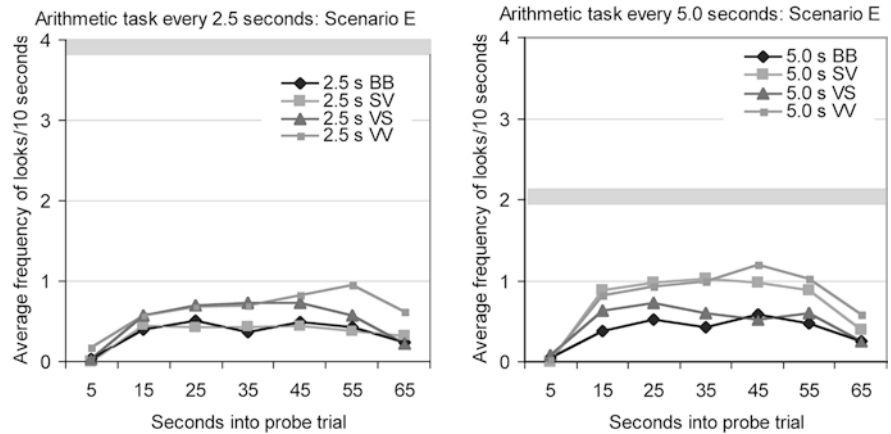


Fig. 9 The averaged frequency of looks at the patient monitoring screen while performing the 2.5 second (fast) arithmetic task (*left*) and the 5.0 second (slow) task (*right*). Grey bands indicate a frequency associated with the literal rate of the arithmetic task itself. Frequency results suggest a capture at one of two multiples of the arithmetic task arrival rate



looking. In contrast, participants in the BB conditions received both visual and sonification support, so they could choose to perform the task either way. However, BB participants were the least consistent in their choice of strategy and their scattered data points on some occasions reflect inconsistency within as well as between participants. It is notable that one BB participant chose not to look at the patient monitoring screen at all (see data point at the 0,0 coordinate) and behaved as if he were an SS participant.

3.4 Impact of the probe trial on visual attention

The second constraint in the experiment that shaped visual attention was the arrival of the probe approximately every minute. This is the medium frequency cycle shown at the bottom of Fig. 3. Were participants' strategies for looking different at different points in the interval between probes? The results shown in the two top figures of Fig. 9 show the average frequency of headturns for each 10 second interval of the approximately 60 second probe trial. The gray band shows the rate at which the arithmetic task arrives: four times every 10 seconds for the 2.5 second arithmetic task and twice every 10 seconds for the 5.0 second arithmetic task. Results suggest that the participants' looking rate is captured at one of two multiples of the arithmetic task arrival rate. When the arithmetic task arrives every 2.5 seconds, most participants take on average one look every 20 seconds (every eight arithmetic trials). However, VV participants step up the frequency of looks as they anticipate the arrival of the problem until they take one look every 10 seconds (every four arithmetic trials). When the arithmetic task arrives every 5.0 seconds, however, there is time for VV and SV participants to schedule one look every 10 seconds without damaging arithmetic task performance. VS and BB participants still only look once every 20 seconds.

3.5 Impact of the scenario on visual attention

The specific anaesthesia events in a scenario should affect visual monitoring. This is the lowest frequency cycle

shown at the bottom of Fig. 3. The ANOVA results for the frequency of looks show not only a main effect of modality, already reported, but also an interaction of modality with scenario, $F(15, 120) = 2.53$, $MSE = 55.6$, $p < 0.01$. For total time looking, there are significant differences for scenario, $F(5, 120) = 2.48$, $MSE = 258.5$, $p < 0.05$, and for the interaction between scenario and modality, $F(15, 120) = 3.36$, $MSE = 258.5$, $p < 0.001$ (see Fig. 11). The duration of looks is not affected by scenario. A full discussion of visual attention under all scenarios is being prepared for separate publication but here we present results for one distinct scenario to illustrate the method.

Scenario E This scenario, shown in Fig. 10, represents an isoflurane (anaesthesia gas) overdose on a spontaneously breathing patient. Points are plotted in the graph when there is a change in the vital sign represented. The scenario is based upon two real clinical incidents from Watson (2002) where anaesthetists were distracted from the patient just after intubation. The scenario includes two pre-existing errors: a failure to turn on the mechanical ventilator and a failure to reduce the isoflurane level after the patient lapsed into unconsciousness. As Fig. 10 shows, $ETCO_2$, RR, and VT all drop during the first 5 minutes of the scenario. In both clinical observations of such an incident, the anaesthetists detected the lack of ventilation as soon as the pitch of the pulse oximeter (oxygen saturation level) started to drop (see approximately 240 seconds into the event). In both cases a rapid response from the anaesthetist corrected the abnormal state. We have allowed the simulation to continue approximately 5 minutes beyond the point at which the anaesthetists intervened in the real-world cases. Clearly, continuing such a scenario without the possibility for intervention is clinically artificial, but it allowed us to test possible differences in how different modalities supported participants' attentional strategies in an extreme case.

Figure 11 shows the total time participants spent looking at the patient monitoring screen in each 60 second interval of the 10 minute scenario. The top two graphs show results combined across all scenarios for

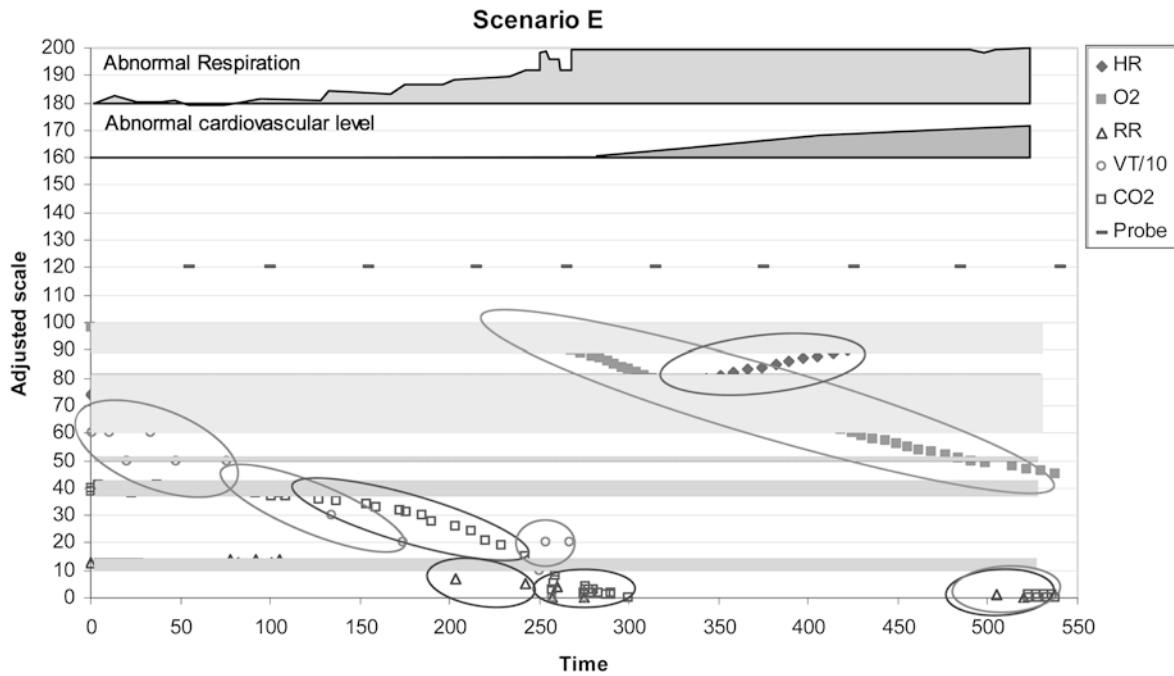


Fig. 10 Scenario E, involving a failure to turn on the ventilator and an isoflurane overdose. The grey bars at top give a rough impression of the degree of deviation from the normal range of respiratory and pulse oximetry parameters, respectively

2.5 seconds (left) and 5.0 seconds (right) arrival rates of the arithmetic task, and reflect the general pattern seen in Figs. 4 and 5. The lower two graphs show results for Scenario E. Visual sampling starts at typical levels for all scenarios, but rapidly diverges when respiration fails around 240 seconds into the scenario. When respiration fails, SV participants see that the three respiratory parameters have fallen to near zero and, since they cannot intervene, feel less need to sample so frequently. This is shown strongly with the 5.0 second arithmetic task and is still present in the 2.5 second arithmetic task. However, VS participants are alerted when the respiratory sonification falls silent—they now have no auditory feedback and so compensate with greater visual attention than before. Again, this is evident for both arithmetic task rates. In further contrast, at the 5.0 second arithmetic rate the BB and the VV participants both show an increase in total time looking when O₂ starts to decrease, whereas at the 2.5 second rate, the BB participants' attention is captured by the respiratory failure. Further analysis will focus on the fine-grained structure of visual attention immediately after the onset of changes in vital signs.

4 Conclusions

In this study, we combined traditional measures of performance with a formative analysis of visual attention to investigate whether sonification helps participants handle multiple task demands while monitoring a

simulated patient undergoing anaesthesia. Our findings suggest that—if used carefully—sonification can be a useful patient monitoring tool. Video analysis indicated that participants found disparate but effective ways to handle the constraints imposed by extraneous tasks and still maintain adequate levels of awareness of patient status. We will outline the principal findings and then comment on the video analysis.

Visual attention The partial sonification of patient information (VS or SV) and sonification added to visual information (BB) quite markedly reduced how often participants turn around to look at the patient monitoring screen. The greatest reduction in the need to look at the patient monitoring screen was when respiratory information was sonified, but we cannot conclude that this is a generalisable finding. Firstly, the visual display of the three respiratory parameters on the Arbiter screen was rather busy compared with the pulse oximetry information. Secondly, the respiratory sonification carried information about three rather than two variables, and so would reduce the need to look more markedly. Thirdly, our scenarios were biased towards respiratory events, so participants' need to turn their heads would have been greater in the SV than in the VS condition. Other scenarios and visual layouts could have reversed the VS and SV results, particularly if BP information were included in the pulse oximetry sonification. An overall benefit of sonification can be seen when it supplements a visual display, as in the BB condition. Here, the number of looks at the visual display was at a minimum, outside the SS condition itself.

Speed of reporting the patient status Participants monitoring with sonification alone were quickest to answer questions about patient status. This is in contrast

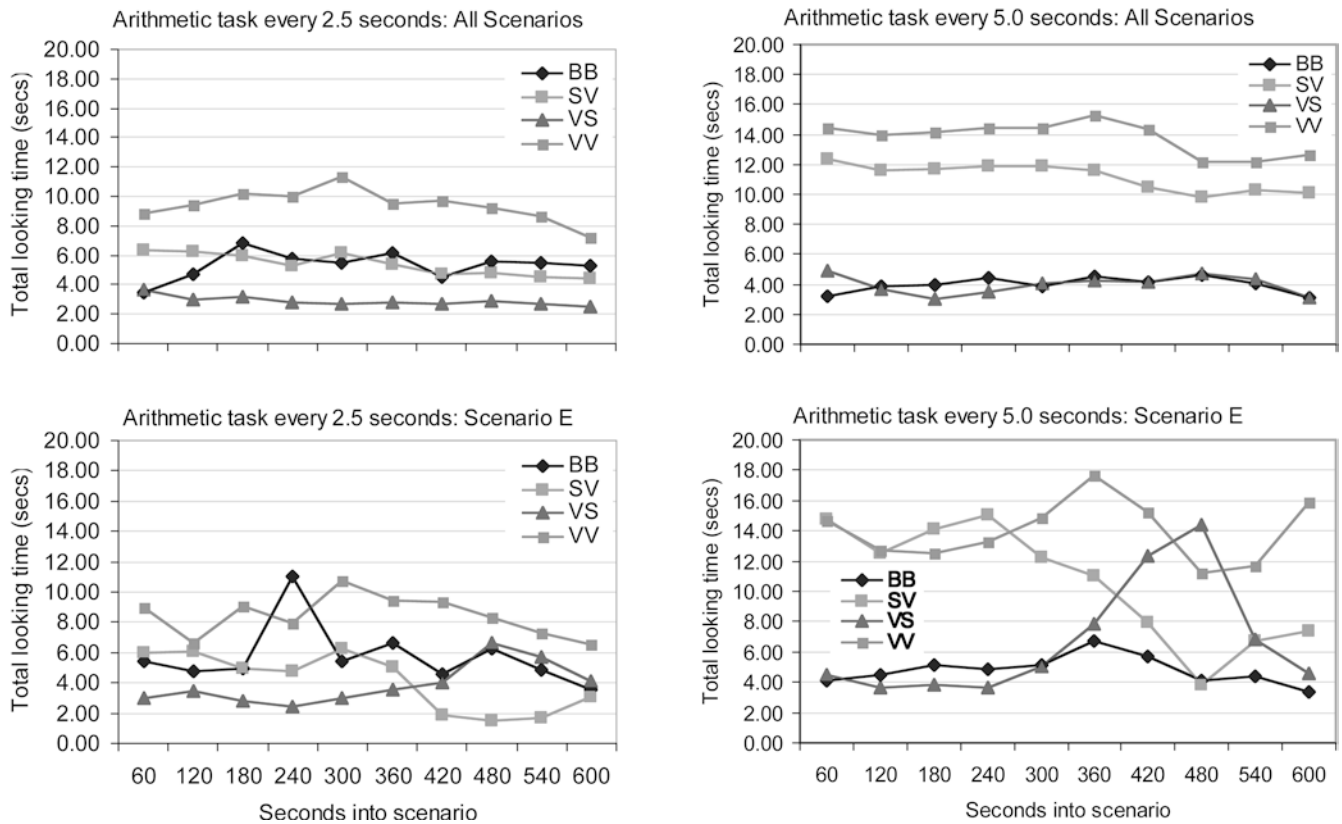


Fig. 11 The amount of time spent looking at the patient monitoring screen in all scenarios (*top*) and in Scenario E (*lower*). At *left* are results for the 2.5 second arithmetic task and at *right* the 5.0 second task. Results show that both modality and scenario influence the time spent looking

to previous work. Loeb and Fitch (2002) did not find any speed advantage for sonification alone—if anything they find the opposite—and Seagull et al. (2001) found slower responses to changes in patient status with sonification alone. However, both those experiments used manual button-press responses whereas our experiment required a vocal response. The exclusively auditory combination of sonified information, vocal probe and vocal response in our SS condition may have led to response facilitation and contributed to faster responding.

Paradoxically, however, participants in the BB condition, which included full sonification, were the slowest responders when handling the fast arithmetic task. BB participants may have experienced a form of response competition on which source of information to rely upon when responding—the visual memory of numbers on the screen, or the echoic memory of sound relations. Choosing the basis on which to respond may be delaying the response. Interhemispheric rivalry may even have contributed to the delay (Kinsbourne and Hicks 1981; Pashler and O'Brien 1993; Ruthruff and Pashler 2001). Similarly, Seagull et al. (2001) found marginally worse tracking in the combined auditory and visual condition,

and their participants reported difficulty “harmonising” the two sources of patient information.

Patient monitoring accuracy Accuracy appeared to be determined by the opportunity participants had to monitor the visual display of the respiratory parameters. Performance was best with the visual only conditions and in the SV condition with the slow arithmetic task. Accuracy was worse for sonification alone. These findings are consistent with other work: patient monitoring accuracy was worst with sonification for Watson and Sanderson's (2004) non-anaesthetists, Seagull et al.'s (2001) non-anaesthetists and Loeb and Fitch's (2002) anaesthetists. However, Seagull et al. (2001) and Loeb and Fitch (2002) found that visual displays and combined visual/sonified displays were equally good, whereas in our study the visual only display (VV) supported better performance than the combined (BB) display. Our BB participants had to turn to see the visual information whereas in previous studies the visual information was in the forward field of view. As a result our BB participants relied more upon the sonification in order to give priority to the arithmetic task. Therefore the accuracy of their responses to probes about patient vital signs suffered.

The less accurate identification of respiratory changes does not necessarily mean that the respiratory sonification was poorly designed. Our scenarios focused on respiratory events, which increased stimulus uncertainty for respiratory parameters. Moreover, the respiratory parameters showed greater interactions and more

complex patterning, such as hard-to-score “fluctuating” episodes, than other parameters. The relative difficulty of identifying respiratory rather than pulse oximetry changes from sonification cannot be entirely due to the fact that three rather than two vital signs were integrated, because Loeb and Fitch (2002) report the same problem when comparing a sonification of HR, O₂ and BP with a sonification of RR, ETCO₂ and VT. Watson and Sanderson’s (2004) anaesthetists did not show worse performances for sonified respiratory variables, but that was in the context of full 10 minute anaesthesia scenarios in which a schema of patient status could develop in a way that could not with Loeb and Fitch’s 2 minute scenarios.

The abovementioned information indicates a shortcoming of assessing the effectiveness of display modalities simply by comparing how accurately participants can identify vital signs. In a study comparing configural visual displays, Bennett et al. (2000) demonstrated that display superiority depends on whether a memory methodology or visual recognition methodology is used and whether low-level data or high-level properties are probed. In our study, we probed working memory for low-level data, whereas our results might be quite different if we probed for high-level properties. In an earlier work, Crawford et al. (2002) found that participants in the sonification condition were much more able to talk spontaneously about patient status than participants in other conditions. We are comparing responses to low-level data vs. high-level properties in currently ongoing work.

Coordination of activity and workload The video analysis of headturning provided us with quite a rich understanding of how timesharing was managed under the different display and arithmetic conditions. Through careful experimental design and formative modelling, this understanding was achieved without using eyetracking. MacSHAPA timeline graphs and state transition diagrams of headturn durations and interlook intervals showed strong differences between modalities and sometimes also between participants within modalities.

By performing a formative analysis of the constraints to be satisfied in the task, we were able to characterise the strategies that individual participants developed for looking at the patient monitoring screen. Their strategies placed them onto regions of a response surface that maximised their opportunities to answer the arithmetic questions with the speed and accuracy required. Across each approximate 60 second probe interval, participants ration their looks at the patient monitoring screen to about one every 10 or 20 seconds, depending on how much sonification support they already have. At the same time, the performance of participants relying on the visual display, in particular, is entrained to the approximate 60 second period of the probe—they increase the frequency of looks at the patient monitoring screen when they judge the probe to be imminent. Despite entrainment to these higher frequency-band tasks

(Sanderson and Fisher 1994) participants’ attentional strategies still showed interpretable coupling with the events in the anaesthesia scenario they were monitoring, as the results for Scenario E show. Analysis is under way on the moment-by-moment reaction of participants in different conditions to scenario events under visual and/or sonified conditions. Although the present study provides just a very simple analysis of the dynamic organisation of visual attention over time, it suggests it might be worth applying techniques from dynamical systems theory (Newton 1995).

Overall, the combination of (1) a flexible experimental environment for generating and displaying interactive or non-interactive anaesthesia scenarios (Arbiter), (2) video analysis tools (MacSHAPA) and (3) a formative framework for visualising and interpreting results, provides a powerful basis for understanding what happens to visual attention and patient monitoring performance as situational constraints change. For example, we can track the effect on visual attention and patient monitoring of such factors as domain expertise, the self-pacing rather than forced-pacing of timeshared tasks, the effects of extraneous noise and different methods of eliciting awareness of patient status, as well as further forms of patient information display. Moreover, when Arbiter is run in interactive mode, we can track how participant interventions may reset rhythms of auditory and visual attention. With this conceptual framework, and with our growing knowledge of the relative sensitivity of behaviour to well-understood manipulations, we are better able to approach the question of how auditory and visual attention can best be supported in complex real-world contexts. In such contexts the analyst’s role is as much to discover the factors entraining visual and auditory attention and contributing to successful performance as it is to test hypotheses about such factors.

Acknowledgements The respiratory sonification used in this study was originally developed between 1998 and 2001 by Watson and Sanderson when Watson was a PhD student under the supervision of Sanderson at the Swinburne Computer Human Interaction Laboratory (SCHIL), Swinburne University of Technology, Melbourne, Australia. We acknowledge support under ARC Discovery Grant DP0209952 to Professor P. Sanderson and Dr. W. John Russell.

References

- Bennett K, Payne M, Calcaterra J, Nittoli B (2000) An empirical comparison of alternative methodologies for the evaluation of configural displays. *Hum Factors* 42:287–298
- Cook RI, Woods DD (1996) Adapting to new technologies in the operating room. *Hum Factors* 38:593–613
- Crawford J, Watson M, Burmeister O, Sanderson P (2002) Multimodal displays for anaesthesia sonification: timesharing, workload, and expertise. In: Proceedings of the Joint ESA/CHISIG Conference on Human Factors (HF2002), Melbourne, Australia, 27–29 November 2002
- Fitch T, Kramer G (1994) Sonifying the body electric: superiority of an auditory over a visual display in a complex, multi-variate system. In: Kramer G (ed) *Auditory display: sonification,*

- audification and auditory interfaces. In: Proceedings of the International Conference on Auditory Displays (ICAD94), Addison-Wesley, Reading, MA, pp 307–326
- Gopher D (1993) The skill of attention control: acquisition and execution of attention strategies. In: Meyer D, Kornblum S (eds) *Attention and performance XIV: synergies in experimental psychology, artificial intelligence, and cognitive neuroscience—a silver jubilee*, MIT Press, Cambridge, MA
- Kinsbourne M, Hicks R (1981) Functional cerebral space: a model for overflow, transfer, and interference effects in human performance. In: Requin J (ed) *Attention and performance VII*, Lawrence Erlbaum, Hillsdale, NJ, pp 345–362
- Loeb RG (1993) A measure of intraoperative attention to monitor displays. *Anaesth Anal* 76:337–341
- Loeb RG (1994) Monitor surveillance and vigilance of anaesthesia residents. *Anesthesiology* 80:527–533
- Loeb RG, Fitch WT (2000) Laboratory evaluation of an auditory display designed to enhance intra-operative monitoring. In: Proceedings of the Society for Technology in Anaesthesia, Orlando, FL, 13–15 January 2000. Abstract from http://anestech.org/publications/Annual_2000/Loeb.html.
- Loeb RG, Fitch WT (2002) A laboratory evaluation of an auditory display designed to enhance intraoperative monitoring. *Anesth Anal* 94:362–368
- Moray N (1986) Monitoring behaviour and supervisory control. In: Boff K, Kaufman L, Thomas J (eds) *Handbook of perception and human performance*, vol 2, Wiley Interscience, New York
- Newton D (1995) The dynamics of action and interaction. In: Smith L, Thelen E (eds) *A dynamic systems approach to development: applications*, MIT Press, Cambridge, MA
- Pashler H, O'Brien S (1993) Dual-task interference and the cerebral hemispheres. *J Exp Psychol Hum Percept Perform* 19:315–330
- Ruthruff E, Pashler HE (2001) Perceptual and central interference in dual-task performance. In: Shapiro K (ed) *Temporal constraints on human information processing*, Oxford University Press, Oxford, UK
- Sanderson PM (1998) Cognitive work analysis and the analysis, design, and evaluation of human–computer interactive systems. In: Proceedings of the Australian/New Zealand conference on computer–human interaction (OzCHI98). IEEE Computer Society, Los Alamitos, CA, pp 220–227
- Sanderson PM, Fisher C (1994) Exploratory sequential data analysis: foundations. *Hum Comput Interact* 9(3):251–317
- Sanderson PM, Fisher C (1997) Exploratory sequential data analysis: exploring observational data in human factors. In: Salvendy G (ed) *Handbook of human factors and ergonomics*, 2nd edn. Wiley, New York, pp 1472–1476
- Sanderson PM, Scott JJP, Johnston T, Mainzer J, Watanabe LM, James JM (1994) MacSHAPA and the enterprise of exploratory sequential data analysis (ESDA). *Int J Hum Comput Stud* 41(5):633–681
- Seagull FJ (1996) Breathe deeply and count backwards from ten: a cognitive work analysis of anaesthesia practice. Unpublished report of fieldwork for PSY493, Department of Psychology, University of Illinois at Urbana-Champaign
- Seagull FJ, Sanderson PM (2001) Anaesthesia alarms in surgical context: an observational study. *Hum Factors* 43(1):66–77
- Seagull FJ, Wickens CD, Loeb RG (2001) When is less more? Attention and workload in auditory, visual and redundant patient-monitoring conditions. In: Proceedings of the 45th Annual Meeting of the Human Factors and ergonomics society, San Diego, CA, October 2001
- Sougné J, Nyssen AS, De Keyser V (1993) Temporal reasoning and reasoning theories: a case study in anaesthesiology. *Psychol Belg* 33:311–328
- Tukey J (1977) *Exploratory data analysis*. Addison Wesley, Boston, MA
- Vicente KJ (1999) *Cognitive work analysis: towards safe, productive, and healthy computer-based work*. Lawrence Erlbaum Associates, Mahwah, MA
- Watson M (2002) *Sonification and anaesthesia: ecological design and empirical evaluation*. Dissertation, Swinburne University of Technology
- Watson M, Sanderson P (2004) Sonification helps eyes-free respiratory monitoring and task timesharing. *Hum Factors* 46(2)
- Watson M, Russell WJ, Sanderson P (1999) Ecological interface design for anaesthesia monitoring. In: Proceedings of the 9th Australasian conference on Computer–Human Interaction OzCHI99. IEEE Computer Society Press, Wagga Wagga, Australia, pp 78–84
- Watson M, Sanderson P, Russell WJ (2004) Tailoring reveals information requirements: the case of anaesthesia alarms. *Interact Comput* 16:271–293
- Woods DD (1995) The alarm problem and direct attention in dynamic fault management. *Ergonomics* 38:2371–2393