Watson, M., Sanderson, P., Woodall, J., & Russell, W. J. (2003). Operating theatre patient monitoring: The effects of self paced distracter tasks and experimental control on sonification evaluations. *Proceedings of the 2003 Annual Conference of the Computer-Human Interaction Special Interest Group (CHISIG) of the Ergonomics Society of Australia (OzCHI2003)*. St Lucia, Qld, 26-28 November.

Operating Theatre Patient Monitoring: The Effects of Self Paced Distracter Tasks and Experimental Control on Sonification Evaluations

M. Watson[1], P. M. Sanderson[1], J. Woodall[1], and W. J. Russell[2]

[1]*The ARC Key Centre for Human Factors and Applied Cognitive Psychology*
The University of Queensland, St Lucia, Australia, 4072
[2]*University of Adelaide, South Australia*

mwatson@humanfactors.uq.edu.au, psanderson@humanfactors.uq.edu.au, john.russell@adelaide.edu.au

Abstract

Three experiments were conducted to explore the effectiveness of continuous auditory displays, or sonifications, to convey information about simulated patient physiological state in dual task trials. Participants' patient monitoring performance under different dual task conditions was investigated to determine if anaesthetic training was required to successfully use the sonification. A patient monitoring experiment using an arithmetic distracter task established that at high workloads, participants performed better when supported by the sonification than by visual displays. A second experiment using a manual labelling manipulation distracter task, that was less demanding on visual attention, demonstrated that participants could effectively use the combined sonification and visual patient monitors. A third experiment using the labelling task demonstrated the potential for non-anaesthetists to achieve better patient monitoring with sonification through greater exposure to the sonification and visual monitors. The results of the three experiments are used to highlight how the design of distracter tasks in divided attention experiment on process monitoring may have an impact on substantive conclusions about display modality effects.

1. Introduction

In the operating theatre (OT), where the problems of vigilance, workload and auditory alarms are well recognised, several attempts have been made to investigate the use of sonification to improve patient monitoring (Watson & Sanderson 2001a, b; Seagull, Wickens & Loeb, 2001; Fitch & Kramer, 1994; Loeb & Fitch, 2000). We demonstrate in this work how the design of laboratory-based experiments can influence the outcomes of monitoring evaluations.

Most current OT patient monitoring systems have been developed with the goal of providing better visual displays; however, these displays are only useful to anaesthetists if they are looking at the monitor. The need for better auditory displays in the OT has been recognised for some time (Woods, 1995) but there has been little progress in developing new auditory systems. Even in the area of medical monitoring standards, auditory displays lag well behind visual displays. Even the most successful auditory display in the OT, the pulse oximetry sonification, does not have a standard associated with how oxygenation is represented in pitch. We propose that the cause of this difference between auditory display and visual display development lies in both the limited understanding about auditory perception and the lack of effective techniques for evaluating auditory displays.

In prior attempts to extend OT patient monitoring sonifications beyond pulse oximetry, a variety of methods have been used to evaluate new sonifications. All of the approaches so far have examined participants' patient monitoring performance under visual, sonification and combined visual and sonification conditions (Watson & Sanderson, 2001a, b; Seagull, Wickens & Loeb, 2001; Fitch & Kramer, 1994; Loeb & Fitch, 2000). Fitch and Kramer evaluated the ability of non-anaesthetists to identify six different physiological events and found that participants performed best when they used the sonification alone to detect the event.

However when an improved version of the sonification was later evaluated by Loeb and Fitch using anaesthetists to detect eight physiological events, they found that the anaesthetists performed worse in the sonification alone condition. In both cases, the evaluations may have have underestimated the effectiveness of the sonifications. Problems with the evaluations included the type of visual display selected, the limited number of physiological conditions investigated and evaluating the displays under unrealistic OT conditions where the participant could focus solely on the monitoring task.

Some experiments examining patient monitoring sonifications have used distracter tasks and have found a variety of results for the three conditions (Watson & Sanderson, 2001a, b; Seagull, Wickens & Loeb, 2001; Crawford, Watson, Burmeister, & Sanderson, 2002; and Sanderson, Crawford, Savill, Watson, & Russell, 2003). Watson and Sanderson and Crawford et al. used an arithmetic task whereas Seagull et al. used a visual tracking task. Watson and Sanderson, Sanderson et al., Crawford et al., and Seagull et al. found that non-anaesthetists performed better at the patient monitoring task when the visual display was available; however, participants performed better at the distracter task when monitoring the patient with the sonification. The cause of the apparent trade-off between patient monitoring performance and the distracter task could be due to several factors including the lack of medical experience in the participant populations, the minimal amount of training with the sonification, or the nature of the distracter tasks. In the Watson and Sanderson experiment, anaesthetists also participated in the tests and they performed better at both the patient monitoring and distracter task than the non-anaesthetists.

Watson and Sanderson (2001a, b) found that anaesthetists performed equally well at the patient monitoring in all three conditions. The anaesthetists also performed better at the arithmetic distracter task under the sonification condition alone. These findings indicate that this particular sonification of patient physiology appears viable; however, many questions relating to workload, different distracter tasks, OT fidelity and expertise still need to be answered before the design principles used to develop this work can be extended beyond this particular sonification.

1.1. Problems with prior distracter tasks

A problem with the Watson and Sanderson (2001a, b) design was the low workload associated with the ten second arrival rate of the arithmetic distracter task. Sonification may only have an advantage at high workloads or during vigilance tasks that involve low workloads (Kramer et al., 1997). Given that the participants only required one to two seconds per arithmetic task to determine the true/false answer, there was ample time to conduct visual monitoring of the physiological parameters.

1.1.1. Tempo of distracter tasks

Sanderson et al (2003) investigated whether non-anaesthetist patient monitoring performance would drop in the visual condition if the time allowed for the arithmetic task was decreased (see also detailed comparison with Watson & Sanderson results in Crawford et al., 2002). The arithmetic task arrival rates of 5 and 2.5 seconds were examined in a between-subjects design and neither of the rates decreased participants' patient monitoring performance in the visual condition to the level observed for physiological monitoring with the sonification (Sanderson et al., 2003).

One possible reason for these findings is that the arrival rate of the arithmetic task still allowed participants to adopt effective patient monitoring strategies in the visual condition. Hypothetically, if the arrival rate of the arithmetic task had been continually reduced, then eventually it could be expected that the participant's ability to monitor the patient using the visual display would be adversely affected. An alternative hypothesis is the arrival rate of the arithmetic task prevented the participants from developing effective sonification patient monitoring strategies.

The reason for the selection of the arithmetic task was that anaesthetists are required to do regular drug calculations and matching tasks such as blood type. These tasks may require the anaesthetist's visual attention for short periods of time, preventing them from continually monitoring the visual display. There are examples of anaesthetists adopting regular monitoring strategies of the visual display (Seagull & Sanderson, 2001); however, these monitoring strategies do not match the types observed in the Watson and Sanderson (2001a, b) or the Sanderson et al., (2003) studies. In the laboratory environment where the arithmetic task is forced paced at an interval set by the experimenter, participants displayed hyper-vigilant visual monitoring with as many as 30 looks at the screen per minute. These strategies do not reflect normal OT behaviour and therefore participants' performance in the visual conditions is likely to be an overestimation of actual expected OT performance. The use of a forced-paced task may prevent participants from developing effective patient monitoring strategies when using the sonification because the arithmetic arrival rate of 5 or 2.5 seconds is less than the length of a normal breath. This may make it difficult for some individuals to effectively divide their attention between the two tasks. In the OT, anaesthetists may have greater latitude in how they allocate their attention between tasks.

1.1.2. Types of distracter tasks

Another issue that needs further examination is the type of distracter tasks that have been used when evaluating physiological sonifications. So far only the simple arithmetic task and a visual tracking have been evaluated as the distracter task. Both tasks are limited at reflecting the types of tasks performed by anaesthetists in the OT. In this work, we examine two types of distracter tasks: the cognitively focused arithmetic task already used in the prior studies and a task that emphasised manual dexterity. Rather than using a well-known manual task such as the Purdue Pegboard, where there are problems with the short duration of the task and different performance between the sexes (Tiffin & Asher, 1948; Bass & Stucki, 1951) we have focused on developing a task representative of OT anaesthesia tasks.

The task is a labelling task which involves similar activities to blood type matching and syringe labelling found in the OT. The labelling task required the participant to peel off a numbered adhesive label from a sheet, and then place that label inside the borders of a corresponding box on an answer sheet. If the number on the label and the number on the answer box do not match, the participant is required to mark this mistake by placing the label over the box number as opposed to inside the box. The participants were scored on the number of labels correctly applied within a time period.

1.2. Hypotheses

Three experiments were run to test if allowing participants to pace their answering of the distracter task would affect their ability to monitor the five patient physiological parameters. The purpose of the experiments was to investigate whether changes in the tempo and type of distracter task would affect participants' patient monitoring performance using the sonification and visual display. We also tested if the participant's patient monitoring performance is constrained by the type of task or the OT practice of placing the patient monitoring displays behind the anaesthetists. To examine this issue, in Experiment 3 participants were not prevented from looking at the visual monitor (in conditions that it was available) during probes for the patient parameter state.

Based upon the findings of Watson and Sanderson (2001) and Sanderson et al., (2003), we proposed the following hypotheses:

1. Participants will monitor the patient state more effectively when the visual display is available. This is because the visual display allows participants to confirm their belief about parameter states. When participants must rely on the sonification alone and they are not given the initial starting values for each parameter, there is the potential for auditory drift.
2. Participants will achieve more correct responses at the distracter task when patient parameters are presented by the sonification alone. This is based on the findings of Sanderson et al., (2003) where the time to complete an arithmetic problem is far less than the time to turn and look at the visual monitor.
3. There will be no effect of type of task on patient monitoring performance using the sonification alone.
4. Participants will perform better at judging the absolute state of patient parameters when they can look at the visual monitor when probed for parameter states.

2. METHOD

2.1. Participants

There were 36 participants recruited through The University of Queensland post-graduate lists. None of the participants had prior anaesthesia or medical experience. Each participant was paid $50 for their involvement.

2.2. Design of experiments

We used a dual task paradigm with either the arithmetic or the labelling distracter task and three patient monitoring conditions in a within-subjects design. The first experiment used a self-paced arithmetic distracter. The second experiment replaced the arithmetic task with a labelling task that involves a high manual component. In both the first and second experiments participants were not permitted to look at the visual patient monitor when they were required to report patient parameter states. The third experiment used the labelling task of the second experiment; however, participants were permitted to look at the visual monitor when they were required to report patient parameter states.

2.3. Stimuli and Apparatus

2.3.1. Patient monitoring task

The patient monitoring used anaesthesia scenarios produced from an anaesthesia simulation (the Body™ simulation from Advanced Simulation Corporation). These scenarios were based upon those described in Watson and Sanderson (2001a, b). The Arbiter experimental environment provided the visual and auditory interface for Body™ on a Dell workstation, with a 19-inch LCD display (see Figure 1). The visual interface provided information on the simulated patient's heart rate (HR), oxygenation (O_2), respiration rate (RR), tidal volume (V_T), and end tidal carbon dioxide (CO_2). The V_T on the visual interface was modified so that it rounded the volume down to the nearest 100mL. The respiration sonification used the pure tone and mapped inhalation and exhalation to the upper and lower note of a musical third. RR was represented by a direct temporal mapping of inhalation and exhalation, V_T was represented by sound intensity, and CO_2 was represented by a frequency modulation of the inhalation/exhalation minor third. The flow-rate of gas in the breathing circuit was mapped to sound intensity which indicated V_T.

The three patient monitoring conditions were as follows:

1. Sonification of physiological parameters (S condition). Participants heard sonifications of the five parameters. HR and O_2 were displayed using a pulse oximetry sound stream while RR, V_T and CO_2 were displayed in a second sound stream. No patient information appeared on the Arbiter computer screen.
2. Visual display of physiological parameters (V condition). Participants had to turn their heads to see the patient's physiological status on the monitor 180^o behind them. This simulated the common placement of patient monitoring displays relative to the anaesthetist and patient in the OT.
3. Both visual and sonification displays (B condition). Participants could attend to either or both modalities.

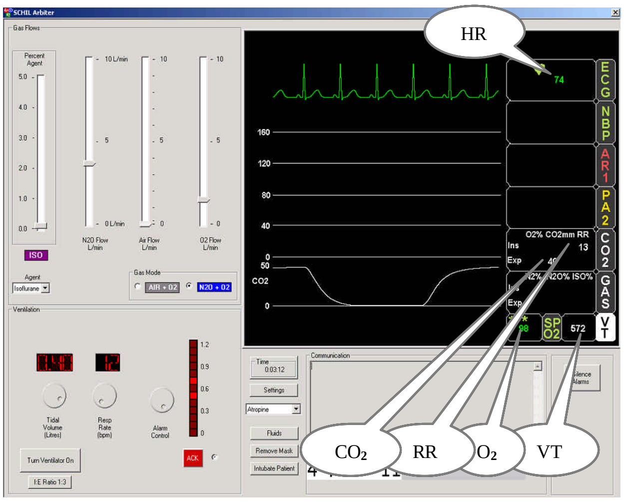

Figure 1. Arbiter patient monitoring display

2.3.2. Arithmetic Distracter Task

The arithmetic task was programmed in Microsoft Visual Basic and was presented on a laptop with a 15-inch screen. This task was based upon the one used in Sanderson et al. (2003; see also Crawford et al., 2002); however, the task was self-paced rather than the forced-paced procedure with new arithmetic task arriving at 5.0 or 2.5 second intervals that had been reported in Crawford et al. Participants were required to make true/false judgments about simple arithmetic expressions that appeared on the computer screen in front of them as quickly and accurately as possible (see left picture Figure 2). The true/false keystroke response was collected automatically and the participants received a running accuracy and reaction time readying on the graph in front of them.

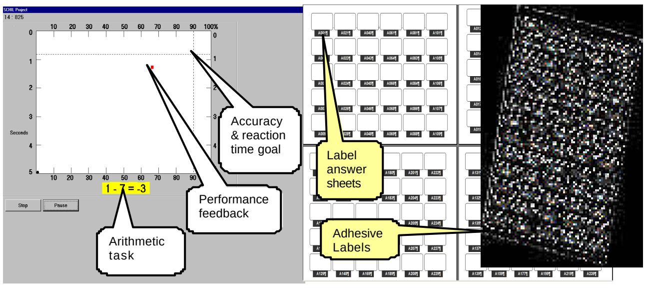

Figure 2. Arithmetic and manual distracter tasks

2.3.3. Labelling Distracter Task

The labelling task used a list of four-digit numbers on adhesive labels, up to several A4 pages long (see right picture in Figure 2). Each A4 page contained 65 labels 21.2 x 38.1mm in size. The answer sheet contained columns of labelled boxes 25.3 x 41.2mm on which correctly matched labels were placed. The task involved a pattern matching task and a placement accuracy task. Ten percent of the labels were incorrect matches to the answer sheet, with the error indicated by one of the last two digits being different. Labels that matched the answer sheet box were scored as correct if none of the labels overlapped the box's border. Incorrect matches were scored as correct if the label was placed over the box's border. Data were collected manually from the labelling task.

2.4. Procedure

Participants were positioned with the anaesthetic monitor behind them and the distracter task in front of them (see Figure 3). Participants performed the distracter task while monitoring the patient. Participants were first trained how to conduct the distracter task and then learned how to use the visual and sonification displays using a series of six training scenarios. After participants completed a baseline trial of the distracter task to determine their arithmetic ability, they performed the patient monitoring together with the distracter task for each of the three monitoring conditions. Once the participants had completed all the patient monitoring trials, they completed a post experiment baseline trial on the distracter task. The distracter tasks were self-paced, with the participants instructed to complete as many labels as possible. At the end of each scenario participants were given feedback about their accuracy compared to the first baseline.

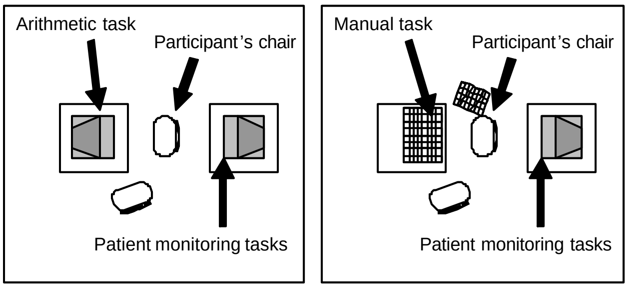

Figure 3. Room layout for the arithmetic and manual distracter tasks

Participants monitored a series of simulated realistic anaesthesia scenarios (labelled A through I) that were variants of those developed in Watson (2002). All participants experienced all scenarios, although the order of presentation of scenarios was counterbalanced across participants for each of the three experiments. Baseline measures for the distracter task were taken without the patient monitoring task pre- and post-condition evaluation. Each of the nine scenarios ran for approximately 10 minutes. Each scenario was divided into trials of approximately 1 minute; with the display pausing for the last five seconds of each trial, where a voice prompt asked the participants to report status and direction of any change of one of the five physiological variables being monitored. Participants responded verbally to the voice prompt and the experimenter recorded the response. The probes were balanced within each monitoring condition such that each of the five physiological parameters was probed an equal number of times.

12 participants performed the arithmetic task and two groups of 12 participants performed the labelling task. One of the two groups conducting the labelling distracter task was allowed to view the visual monitor in the visual and combined conditions during the probe for patient parameters.

3. Results

3.1. Arithmetic task experiment

Figure 4 maps the patient monitoring and arithmetic distracter task performance as fast, accurate ("Good") performance in the top right hand corner; slow and inaccurate ("Bad") in the bottom left corner. Figure 4 is also used to compare participants from the present experiment and those from the Watson and Sanderson (2001a, b) as "IT" for a group of IT postgraduates and "An" for a group of anaesthetists.

For the patient monitoring we examine participants' ability to discriminate high, normal or low states of physiological parameters (absolute judgments), and their ability to discriminate the direction of change (direction judgments). Within-subject ANOVAs were conducted for the participants' absolute and directional judgments.

For absolute judgments, we found no significant effect of Modality, $F(2,22) = 0.74$, $MSe = 375.0$, ns. There were significant effects of Parameter, $F(4,44) = 4.38$, $MSe = 1243.8$, $p<0.01$, with changes in V_T being less accurately detected in all conditions. There was also a two-way interaction of Modality, and Parameter, $F(8, 88) = 2.11$, $MSe = 776.2$, $p<0.05$. Participants find it more difficult to detect changes in V_T than either HR or O_2 in the B condition. For directional judgments we found no effect of Modality or Parameter.

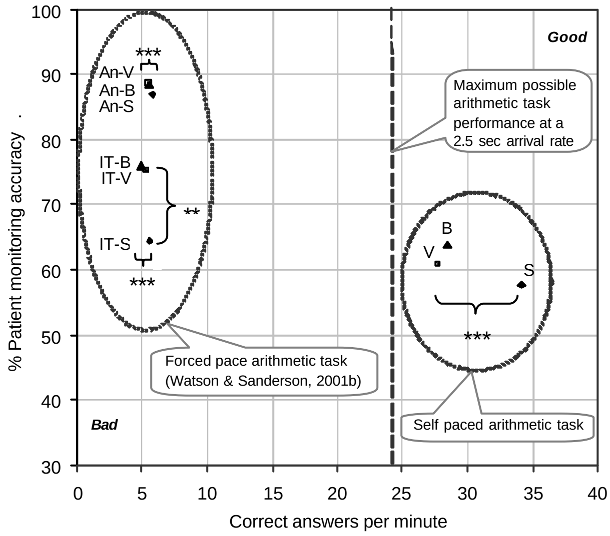

Figure 4. Patient absolute judgments and arithmetic distracter task performance pooled across parameters. S = sonification, V = visual and B = sonification and visual patient monitoring. An = anaesthetist, IT = Information Technology Students from Watson & Sanderson (2001a). ** = p<0.01, *** = P<0.001

For the arithmetic task we examined the number of correct answers per minute and participants' baseline arithmetic task performance pre- and post-experiment. A within-subject ANOVA was conducted for Modality and Serial Position (practice effect within each modality) for the arithmetic task. We found a significant effect of Modality, $F(2,22) = 9.94$, $MSe = 458.9$, $p<0.001$. There were significant effects of Serial Position, $F(2,22) = 4.98$, $MSe = 98.2$, $p<0.05$, with participants improving between the first and last scenario in each Modality. There was no interaction of Modality and Serial Position $F(4,44) = 1.22$, $MSe = 24.03$, ns.

A within-subject ANOVA for Order of the baseline task found that there was a significant effect of Order, $F(1,11) = 28.3$, $MSe = 92.2$, $p<0.001$. Participants completed more answers correctly in the pre-experiment baseline (mean = 42.1 SEM = 3.25) than the post-experiment baseline (mean = 38.2 SEM = 3.39).

3.2. Labelling 'no looking' task experiment

Figure 5 maps the patient monitoring and labelling distracter task performance as fast, accurate ("Good") performance in the top right hand corner; slow and inaccurate ("Bad") in the bottom left corner. Figure 5 is also used to compare participants from the 'no looking' Experiment 2 and the 'looking' Experiment 3.

Patient monitoring was analysed for abnormality and directional judgments as in the arithmetic distracter task experiment. Within-subject ANOVAs were conducted for the participants' absolute and directional judgments.

For absolute judgments, we found a significant effect of Modality, $F(2,22) = 4.53$, $MSe = 1594.1$, $p<0.05$. Participants were better able to monitor the patient in the B condition than in the S condition. There were significant effects of Parameter, $F(4,44) = 4.62$, $MSe = 1243.8$, $p<0.01$, with changes in V_T being less accurately detected in all conditions. There was no interaction of Modality and Parameter, $F(8, 88) = 0.70$, $MSe = 268.9$, $p<0.05$. For directional judgments we found no effect of Modality or Parameter.

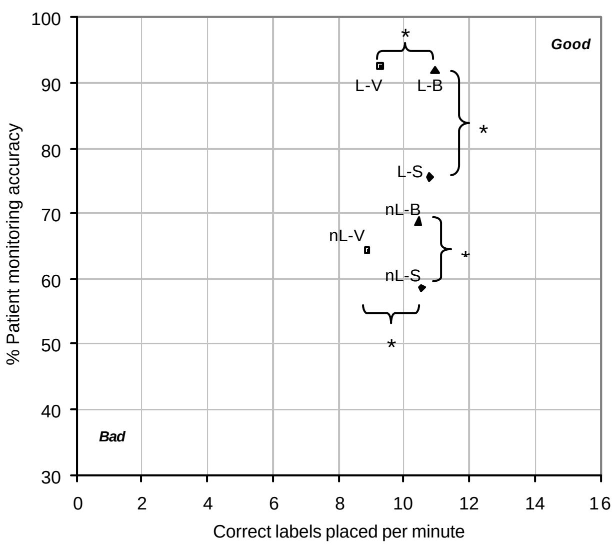

Figure 5. Patient absolute judgments and labelling distracter task performance pooled across parameter. S = sonification, V = visual, B = sonification and visual patient monitoring, and L = looking and nL = no looking allowed during patient probes. * = p<0.05, *** = p<0.001

We analysed the labelling 'no looking' task for the number of correct answers per minute and participants' baseline labelling task performance pre- and post-experiment. A within-subject ANOVA was conducted for Modality and Serial Position (practice effect within modality) for the arithmetic task. We found a significant effect of Modality, $F(2,22) = 4.98$, $MSe = 26.88$, $p<0.05$ with participants performing better in the S and B conditions than the V condition. There were significant effects of Serial Position, $F(2,22) = 44.7$, $MSe = 35.33$, $p<0.001$, with participants improving between the first and last scenario in each Modality. There was a two-way interaction between Modality and Serial Position, $F(4,44) = 14.3$, $MSe = 7.09$, $p<0.001$.

A within-subject ANOVA for Order of the baseline task found that there was a significant effect of Order, $F(1,11) = 27.9$, $MSe = 51.0$, $p<0.001$. Participants completed more answers correctly in the post-experiment baseline (mean = 12.48 SEM = 0.48) than the pre-experiment baseline (mean = 9.57 SEM = 0.50). Participants improved their performance on the labelling task with continued exposure.

3.3. Labelling 'looking' task experiment

Patient monitoring was analysed for absolute and directional judgments as in the arithmetic distracter task experiment. Within-subject ANOVAs were conducted for the participants' absolute and directional judgments.

For absolute judgments, we found a significant effect of Modality, $F(2,22) = 4.59$, $MSe = 1210.7$, $p<0.05$. Participants were better able to monitor the patient in the B and V condition than in the S condition. There were no significant effects of Parameter, $F(4,44) = 2.11$, $MSe = 119.9$, ns, or interaction of Modality, and Parameter, $F(8, 88) = 0.77$, $MSe = 189.0$, ns. There were no significant effects found for the directional judgment analysis.

The labelling 'looking' task was analysed using a within-subject ANOVA for Modality and Serial Position (practice effect within modality). For the labelling 'looking' we found a significant effect of Modality, $F(2,20) = 7.70$, $MSe = 37.55$, $p<0.05$ with participants performing better in the S and B conditions than the V condition. There were significant effects of Serial Position, $F(2,20) = 26.87$, $MSe = 46.50$, $p<0.001$, with participants improving between the first and last scenario in each Modality. There was a two-way interaction between Modality and Serial Position, $F(4,40) = 10.83$, $MSe = 8.42$, $p<0.001$.

3.4. Arithmetic vs. Labelling 'no looking'

We analysed participants' performance between the arithmetic and labelling task for absolute and directional patient monitoring judgments. Using a between, within-subject ANOVA for Task, Modality and Parameter, we found no significant effects for Task $F(1,22) = 0.21$, $MSe = 482$, ns. There were significant effects of Modality, $F(2,44) = 4.10$, $MSe = 1755$, $p<0.05$; Parameter, $F(4,88) = 8.27$, $MSe = 2376$, $p<0.001$; and a two-way interaction between Modality and Parameter $F(8,176) = 2.03$, $MSe = 766$, $p<0.001$. There were no other significant interactions for absolute judgments. We found no significant effects for the ANOVA of directional judgments.

The patient monitoring results indicate that participants' patient monitoring responses were similar across the two types of tasks. It is not possible to directly compare the distracter task performance between the arithmetic and labelling task as there is no corresponding measure for workload.

3.5. Labelling 'no looking' vs. 'looking'

We analysed participants' performance between the labelling 'no looking' and 'looking' task for absolute and directional patient monitoring judgments. Using a between, within-subject ANOVA for Task, Modality and Parameter, we found significant effects for Task $F(1,22) = 32.5$, $MSe = 46316$, $p<0.001$. There were significant effects of Modality, $F(2,44) = 7.97$, $MSe = 6225$, $p<0.001$; Parameter, $F(4,88)= 4.41$, $MSe = 1066$, $p<0.01$; and a two-way interaction between Parameter and Task $F(4,88) = 2.83$, $MSe = 685$, $p<0.05$. There were no other significant interactions.

The ANOVA for directional judgment indicates that there are significant effects of Task $F(1,22) = 4.56$, $MSe = 4112$, $p<0.05$. There were no significant effects of Modality, $F(2,44) = 0.49$, $MSe = 363$, ns. There were no significant effects of Parameter, $F(4,88) = 3.45$, $MSe = 821$, $p<0.01$. There were no other significant interactions.

We analysed participants' performance between the labelling 'no looking' and 'looking' using a between, within-subject ANOVA for Task, Modality and Serial Position (practice effect within modality). There was no significant effects for Task $F(1,22) = 0.21$, $MSe = 482$, ns. There were significant effects of Modality, $F(2,44) = 4.10$, $MSe = 3511$, $p<0.05$; Parameter, $F(4,88)= 8.27$, $MSe = 2376$, $p<0.001$; and a two-way interaction between Modality and Parameter $F(8,176) = 2.03$, $MSe = 766$, $p<0.05$. There were no other significant interactions.

Differences between the 'looking' and 'no looking' experiments only emerge in the absolute patient monitoring performance, where participants could benefit by looking at the monitor during probes.

4. Discussion

4.1. Arithmetic task and patient monitoring

Participants performed better at the arithmetic task when monitoring the patient with the sonification alone (see Figure 4) which reflects the findings of Watson and Sanderson (2001a, b). The self-paced arithmetic results support hypothesis 2 that participants would do better at the distracter task when using the sonification. What is different between our findings and the previous study is that there are no modality differences on the participants' patient monitoring performance in the self-paced arithmetic task. These findings fail to support hypothesis 1, that participants would monitor the patient better when the visual display is available.

Assuming that there are no population differences between the postgraduate students participating in this research and the IT postgraduate students who took part in Watson and Sanderson (2001a, b) then changing the distracter task to be self-paced has had a major effect on the participants' ability to monitor the patient when the visual display was available. Participant in the self-paced arithmetic task were no longer able to maintain the high level of patient monitoring performance in the visual and combined sonification and visual conditions. This is evident from the difference between the two groups' performance (IT-V & IT-B vs. V & B) in the visual and combined display condition in Figure 4.

We suggest that the difference between Watson and Sanderson (2001a, b) and our current finding is due to the increase in workload experienced by the participants. The participants in Watson and Sanderson only received six arithmetic questions per minute, so the workload associated with the distracter task was minimal. Even when Sanderson et al (2003) examined higher workloads, these may not have been enough to prevent hyper-vigilant participants from using the visual display successfully. Participants may have had enough time to maintain regular checking of the monitor when the visual display was available. If participants in Sanderson et al (2003) achieved 100% accuracy at the arithmetic task, the maximum number of correct answers they could have achieved at a 2.5 second delivery rate was 24 answers per minute, indicated by the horizontal line in Figure 4. The mean arithmetic task performance in all three conditions exceeded the maximum performance obtainable by the Sanderson et al. participants.

Another factor indicating that the workload of the distracter task was high is the decrease in performance of the post experiment baseline measure of the distracter task. In our previous work with the forced-paced arithmetic task, there was no difference in participants' performance between the pre- and post-experiment baselines. The decrease in baseline indicates the participants were getting tired towards the end of the experiment.

The arithmetic task results suggest that a sonification may help the overall performance of people conducting high workload tasks. These results reflect those found for anaesthetists at the low workload arithmetic tasks by Watson and Sanderson (2001a, b), as shown in Figure 4. The results of the arithmetic task are further supported by the findings of the labelling task.

4.2. Labelling task and patient monitoring

Participants' performance at the labelling task was better when the sonification was available (nL-S and nL-B in Figure 5). This is different from the finding of the arithmetic results where only the sonification alone supported better performance at the distracter task. These results do not support hypothesis 2 and indicate the type of task has an effect on participants ability to effectively combine information form the visual and auditory modalities.

The difference in visual requirements between the arithmetic and labelling task may have allowed participants to effectively combine the visual monitor with the sonification when they were both available. Participants could use the changes they heard in the sonification to guide when they looked at the visual display. This allowed for better performance at both the patient monitoring and the distracter task when both displays were available. To establish wether this situation is the actual occurring we will need to examine the head turning data captured during the study.

The result may be an underestimation of participants' ability to conduct the labelling task. Unlike the arithmetic task were participants' results showed signs of tiring, participants' labelling results indicated a practice effect between the pre- and post-baseline measurements. It was also notable that participants performed better in the second and third scenario than the first scenario in the combined visual and sonification condition in both the 'no looking' and 'looking' labelling experiments. It is unlikely that this distinction is due to the difference between scenarios because of the counter-balanced design of the experiment. Instead, it appears to be due to participants learning how to effectively manage the two displays together and is evident in the interaction between the Modality and the Serial Position. Therefore, with proper training, the sonification may prove to be a very good adjunct to visual monitors.

As the combined visual and sonification display is the proposed monitoring condition for the OT, these results add further weight to the value of a respiratory sonification for the OT. The value of sonifications will be dependent upon the types of tasks an anaesthetist is performing. Under conditions when there is enough time to regularly check the visual monitor, the sonification may only be useful in alerting the anaesthetist to changes in the patient state. However, when workloads become high, sonifications may be an effective way to free up the visual attention of anaesthetists for other tasks.

4.3. Task effect on patient monitoring

There was no difference in the patient monitoring performance when comparing the arithmetic and labelling tasks. These findings support hypothesis 3 and suggest that it is not the nature of the task, but instead the requirement to divide attention between the distracter tasks and the visual monitor behind the participants that accounts for participants' distracter performances in the different monitoring conditions. The positioning of the monitor behind the participant is a common practice in the OT due to patient accessibility requirements of other OT staff. We therefore argue that the tasks used in these experiments are relevant to patient monitoring in the OT.

Many tasks performed by anaesthetists in the OT will allow them to look at the monitor at regular intervals, even if at a less frequent rate than observed in the laboratory experiments. The labelling experiment shows the potential for people to better manage their attention between tasks when using the sonification and visual display together. In some cases, however, such as the placement of a central venous pressure line (CVP), anaesthetists are very limited in their ability to look at the monitor. The placement of a CVP task may have similar or higher visual demands than the arithmetic task and hence, the anaesthetist may perform better relying on the sonification alone.

4.4. Looking during the probes

The labelling 'looking' experiment reflects the best possible patient monitoring performance that could be expected by participants. The majority of participants chose to look at the visual display during patent monitor probes. We are in the process of analysing the head turning data in the same manner as Sanderson et al. (2003) in order to determine the effect of the 'looking' condition. The conclusions are therefore limited to identifying likely interactions for future investigations.

When comparing participants' performance between the two labelling experiments, 'no looking' and 'looking' during the probe (see Figure 5); it is important to consider that participants were required to make a directional as well as abnormality judgment of the patient's state. In the visual conditions, participants therefore maintained some head turning to look at the visual display throughout the trials. As expected, allowing participants to look at the monitor during the probes did not improve performance on the labelling task (see Figure 5). Participants were better at judging the patient state when they were able to look at the monitor during the probes which supports hypothesis 4. This in itself is not surprising; however, the fact that participants in the looking experiment also performed better at the sonification indicates some sort of training has occurred.

One possible reason is that participants who experienced the combined sonification and visual display prior to the sonification alone were able to more effectively learn the range in the sound dimensions representing the physiological parameters. Further work examining the monitoring strategies and head turning of the participants is required before the improvement in patient monitoring using the sonification alone can be explained. Such measures have proven highly informative in prior work (see results of video analysis of headturning in Sanderson et al., 2003) If the improvement in patient monitoring performance can be attributed to the training during the combined sonification and visual display, then it will be important to include more extensive training in future sonification evaluations.

4.5. Selection of task and experimental design

The results of these three experiments and the prior work (Watson & Sanderson, 2001a, b; Crawford et al., 2002) illustrate the need to carefully consider how the design of a task will influence the types of strategies available to participants. Changing the constraints of the task from a forced-paced to a self-paced task had a large impact on participants' performance and further supports the need for multiple studies to be conducted when evaluating new displays. Balancing fidelity with control is difficult in any experimental design; however, factors driving participants' performance such as a forced-paced task may reflect poorly on attention sharing strategies that are likely to exist in the OT. If the available strategies are potentially unrealistic and highly constrained due to the task, this might lead to poor integration of information across modalities.

Other experimental constraints may also affect the apparent effectiveness of a display. The patient monitoring probe used in this work may have adversely affected participants' performance with the respiratory sonification. In stopping the scenarios approximately every minute to probe for patient state information, the rhythm of the respiratory sonification may be lost. This is especially true when the patient's respiratory rate is low. Participants' directional judgments may have been better if they were given longer to listen to the changes.

Further work will need to examine the respiratory sonification under more realistic OT conditions and with anaesthetists. Watson and Sanderson (2001a, b) have already demonstrated that under low distracter task workloads, anaesthetists perform better overall with the sonification. Whether those findings would hold for high workload is yet to be tested. Other potential advantages of the respiratory sonification are that it might convey higher order information and support anaesthetists during long operations where vigilance might be much lower than exhibited in the laboratory experiments. High fidelity simulator trials with anaesthetists must be run to further examine the viability of a respiratory sonification.

It is possible that the present results will generalise to other domains. However, there may be further interactions of visual and auditory displays with evaluation tasks. We cannot yet make generalisations to other domains or tasks on the basis of the current findings. Encouragingly, the community at work on these problems in the anaesthesia environment seems to be accumulating convergent evidence on when visual displays vs sonifications will provide effective support for anaesthesia staff.

5. Conclusion

Participants with minimal training were able to use the combined pulse oximetry and respiratory sonifications to timeshare patient monitoring with a variety of tasks. The three experiments highlight the need to carefully consider the design of the distracter in dual task experiments. Tasks that do not reflect the constraints of the work environment for which the displays are intended for may prevent participants developing appropriate strategies to conduct all tasks. Unrepresentative task selection may lead to unrepresentative assessments of new displays.

6. Acknowledgments

Financial support for this research was provided by ARC Large Research Grant DP0209952 to Prof P. Sanderson and Dr W. J. Russell. The respiratory sonification used in this work was initiated when Watson and Sanderson were at Swinburne University of Technology.

7. References

Bass, B.M. & Stucki, R.E. (1951). A note on a modified Purdue pegboard. *Journal of Applied Psychology, 35*, 312-313.

Crawford, J. Watson, M. Burmeister, O. & Sanderson, P. (2002). Multimodal displays for anaesthesia sonification: Timesharing, workload, and expertise. *Proceedings of the ESA/CHISIG Conference on Human Factors* (HF2002). Melbourne, Australia, November 27-29.

Fitch, T. & Kramer, G. (1994) Sonifying the body electric: Superiority of an auditory over a visual display in a complex, multi-variate system. In G. Kramer (Ed), Auditory display: Sonification, audification and auditory interfaces. *Proceedings of the International Conference on Auditory Displays ICAD94* (pp. 307-326). Reading, MA: Addison-Wesley

Loeb, R. G., & Fitch, W. T. (2000). Laboratory Evaluation of an Auditory Display Designed to Enhance Intra-Operative Monitoring. *The Society for Technology in Anaesthesia Annual Meeting* 13-15 January 2000 Orlando. Abstract from anestech.org/publications. File: Annual_2000/Loeb.html

Sanderson, P., Crawford, J., Savill, A., Watson, M., & Russell, W. J. (2003). Visual and auditory attention in patient monitoring: A formative analysis. Paper accepted pending revisions for *Cognition, Technology, & Work*..

Seagull, F. J., Wickens, C.D., & Loeb, R.G. (2001). When is less more? Attention and workload in Auditory, Visual and redundant Patient-Monitoring Conditions. *Proceedings of the Human Factors and Ergonomics Society 45th Annual meeting*, 45(1), 1395-1399.

Tiffin, J. & Asher, E.J. (1948) The Purdue pegboard: Norms and studies of reliability and validity. *Journal of Applied Psychology*, 32, 234-247.

Watson, M., & Sanderson, P. (2001a). Intelligibility of Sonification for Respiratory Monitoring in Anaesthesia.

Proceedings of the Human Factors and Ergonomics Society 45th Annual meeting, 45(1), 1293-1297.

Watson, M., & Sanderson, P. (2001b). Respiratory Sonification helps Anaesthetist Timeshare Patient Monitoring with other Tasks. *Proceedings of the 10th Australasian Conference on Computer-Human Interaction OzCHI99*. (pp 175-180). Perth, Australia: IEEE Computer Society Press.