

Multimodal displays for anaesthesia sonification: timesharing, workload, and expertise

Jennifer Crawford¹, Marcus Watson¹, Oliver Burmeister², and Penelope Sanderson¹
¹ARC Key Centre for Human Factors and Applied Cognitive Psychology
The University of Queensland, Australia
²SCHIL, Swinburne University of Technology, Australia
{jcrawford, mwatson, psanderson}@humanfactors.uq.edu.au, oliver@it.swin.edu.au

Keywords: Sonification, auditory displays, timesharing, multimodality, anaesthesia

Abstract

Physiological monitoring is necessary in health care contexts where a patient is anaesthetised or heavily sedated. Our goal is to determine the safest format for keeping a health care practitioner informed about the patient's state, taking into account other tasks that need to be performed. We report results of a study that compares visual, auditory, and mixed modality displays for monitoring an anaesthetised patient while carrying out another task, reflecting real-world healthcare settings. Results of this study in the context of other studies in our laboratory suggest task trade-offs that reflect participants' professional backgrounds, but are nonetheless encouraging for the development of multimodal displays.

1. Introduction

Auditory alarms in the operating theatre can be a nuisance, distracting anaesthetists and other operating theatre staff from appropriate courses of action. Anaesthetists find it difficult to easily discount alarms when they are ambiguous or poorly specified (Woods, 1995; Watson, Russell & Sanderson, 1999; Seagull & Sanderson, 2001). Alarms signal genuine problems as well as changes that are expected, causing confusion as to their exact nature and whether they warrant attention. Auditory displays such as pulse oximetry (the “beep” of the heart monitor) allow the practitioner to monitor a patient's physiology constantly without the need to stop whatever they are doing to look at a screen. With pulse oximetry a practitioner is able to monitor the trend of the patient's state and adopt corrective measures prior to critical events (Watson, Sanderson & Russell, 2000).

Over the last decade, several researchers have examined the possibility that *sonification* might help anaesthetists monitor patients in the operating room (Fitch & Kramer, 1994; Watson, Russell & Sanderson, 1999; Watson, Sanderson & Russell, 2000; Loeb & Fitch, 2000; Watson & Sanderson 2001a, b; Seagull, Wickens & Loeb 2001). Sonification is the representation of data relations through the relations between sounds. Recently, researchers have tried adding a second sonification (in addition to pulse oximetry) to help anaesthetists monitor further variables—in particular, respiratory variables. For example, Watson and Sanderson's (2001a, b) studies demonstrated that anaesthetists can monitor five patient physiological parameters with two parallel sonifications just as well as they can with a visual display showing five variables. They can do this with only a minimal level of prior familiarisation. These studies also demonstrated that sonification allowed anaesthetists' monitoring performance to be sustained at high levels while they performed other tasks. Watson et al. did note, however, that a non-anaesthetist group appeared to trade off performance between two simultaneous tasks.

Further studies have shown differences in how well participants can monitor simulated patients using different combinations of auditory and visual displays (Fitch & Kramer, 1994; Loeb & Fitch, 2000; Seagull, et al., 2001). Only the Watson and Sanderson (2001a, b) and the Seagull et al. studies examined participants' performance under dual task conditions. Using non-medically qualified participants, Seagull et al. found that participants could do a visual tracking task better when they were monitoring six patient physiological parameters with sonifications than with a visual display. The participants, however, detected changes in physiological parameters faster with the *visual* display than with the sonification. Although both Seagull et al. and Watson and Sanderson (2001a, b) used small subject numbers, both sets of results suggest that with dual tasks there are tradeoffs between tasks and modalities that are not yet well understood.

The above studies raise many issues. A first issue is that in both the Seagull et al (2001) and Watson and Sanderson (2001a,b) studies, the two tasks were presented on the same display monitor. This is not typical of the way time-shared tasks are spatially laid out in the operating room (OR). In the present study, therefore, a primary task (distractor task) was placed in front of the participant and a secondary task (patient monitoring) was presented on a screen behind them. Participants had to turn their heads to view the visual output, as is typical of much work in the OR. We hypothesized that a sonification of both cardiovascular and respiratory parameters (S for cardiovascular plus S for respiratory, making the "SS" condition) would lead to performance on the primary and the secondary tasks that would be at least as good and perhaps better as when just one set of parameters was being sonified (for example, the cardiovascular or respiratory parameters alone).

A second issue is that in the Watson and Sanderson (2001a, b) studies new arithmetic tasks arrived at 10-second intervals, which gave participants ample time to use the visual touchscreen to query all five physiological parameters before the next arithmetic task arrived. This was a very conservative test of the sonification against the visual display. Therefore we changed the arrival time of arithmetic task from 10 seconds to five seconds in an attempt to build greater reliance on the sonifications.

A third issue arising from the Watson and Sanderson (2001a, b) studies was whether monitoring two sonifications simultaneously—pulse oximetry and the respiratory sonification—would lead to worse performance for those sonifications than when each was the only set of physiological parameters sonified. This was examined for cardiovascular parameters heart rate (HR) and oxygenation (O₂) by comparing participants' performance at making judgments about HR and O₂ when only HR and O₂ were sonified as opposed to when all parameters were sonified. It was also examined for respiratory parameters respiration rate (RR), tidal volume (VT) and end-tidal carbon dioxide (CO₂) by comparing participants' performance at making judgments about RR, V_t, and CO₂ when only they were sonified as opposed to when all parameters were sonified. Investigating these issues required a between-subjects design to ensure results were not contaminated by carry over effects.

2. Method

Participants

Participants were 24 members of the general population. Twelve were from Swinburne University of Technology and 12 from the University of Queensland. Each between-subjects condition included an equal number of participants from Swinburne and from University of Queensland. All participants gave informed consent and were paid \$10 per hour for their time.

Apparatus and layout

Participants performed the primary task on a laptop computer. They were required to judge whether simple arithmetic expressions were true or false (see Figure 1) and to register their response with a keypress. Every five seconds participants saw updated performance feedback on an X-Y graph on the laptop monitor above the arithmetic task. At the same time, they performed the secondary physiological monitoring task using the Arbiter interface (Watson & Sanderson, 2001a, b) that was located 180° behind them (see Figure 2).

Design

The patient physiological monitoring task was presented in one of four modality combinations, varied between-subjects:

- Sonification (identical to previous studies by Watson & Sanderson, 2001a, b) of pulse oximetry and respiratory parameters (SS condition). No visual patient information appeared on the Arbiter computer screen.
- Visual display of pulse oximetry and respiratory parameters (VV condition). Participants always had to turn their heads to see the patient's physiological status.
- Visual display of pulse oximetry plus visualization of respiratory parameters (SV condition). Participants had to turn their heads to see the respiratory status.
- Visual display of pulse oximetry plus sonification of respiratory parameters (VS condition). Participants had to turn their heads to see cardiovascular status.

The independent variable was presentation modality of the physiological displays: SS, SV, VS or VV. The dependent variables were accuracy and

speed of responding in the primary and secondary tasks.

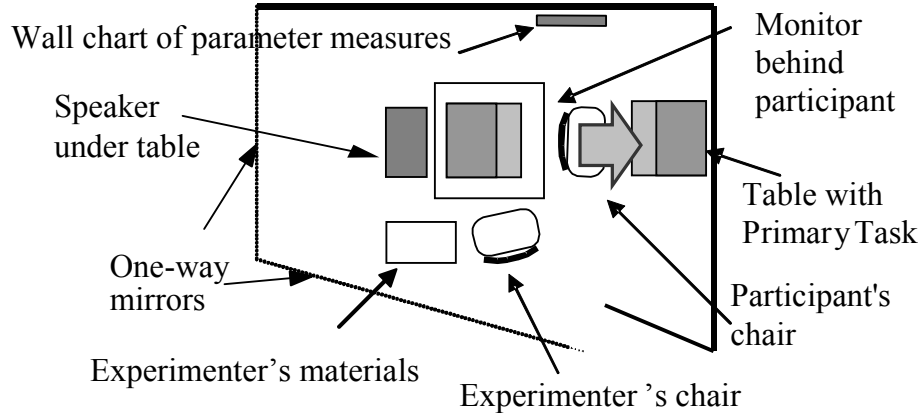


Figure 1. Experimental room-layout.

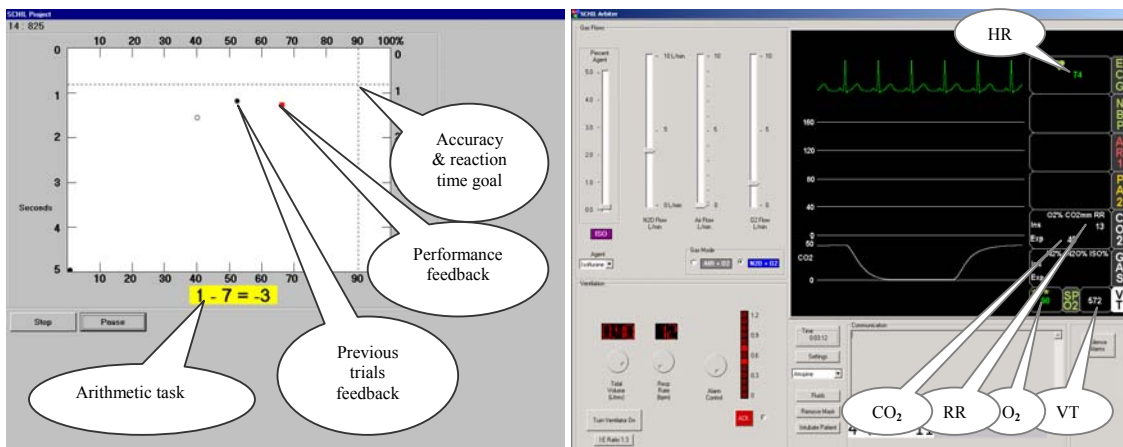


Figure 2. Left: Primary (arithmetic task). Right: Secondary task visual display (the tradition anaesthetic monitor).

Participants monitored plausible operating room scenarios (the same 10-minute anaesthesia scenarios as in Watson and Sanderson 2001a,b) plus a group of control scenarios with no abnormal events. Scenarios were grouped into clusters around anaesthesia themes. In the sonification conditions the same volume setting on the speakers was used so that each participant heard the same stimuli. The brightness and colour contrast were held constant in the visual conditions.

Approximately every minute, participants were prompted with the name of one of the five physiological parameters (the “probe”), whereupon they gave a verbal report about any recent abnormality and the direction of the last change on the probed parameter. Probes were evenly distributed across physiological parameters, so there was no in-built bias in questioning towards any parameter. During the trial, the arithmetic task changed every five seconds. The five-second pacing of the arithmetic tasks gave participants using visual displays enough time to look back at the patient monitor. Participants had less time to survey all parameters in their visual format than in the Watson and Sanderson experiment.

Differences between the Watson and Sanderson (2001a,b) and the present procedures are outlined in Table 1. In all the Watson and Sanderson conditions involving visual displays, participants had to touch a screen to see the physiological parameter requested (the so-called “withholding” technique). The information disappeared after five seconds. In the present experiment, participants turned around to see the visual displays. The present experiment also added feedback on the patient monitoring task during familiarisation sessions. There was also continuous feedback on the primary task so the participants would maintain their accuracy and allocate attention in a more effective way (Schneider & Fisk, 1982; Gopher, 1993).

Table 1. Differences between Watson and Sanderson (2001) and present experiment

Exp	Design	Training with feedback	Condition					Primary task feedback	Time for Task	Visual display
			SS	SV	VS	VV	BB			
Watson & Sanderson	Within - SS	No	Yes	--	--	Yes	Yes	No	10 sec	Withholding
Crawford	Between-SS	Yes	Yes	Yes	Yes	Yes	--	Yes	5 sec	Head turning

3. Results

Figure 3 maps the primary and secondary task performance as fast, accurate (“good”) performance in the top right hand corner; slow and inaccurate (“bad”) in the bottom left corner. Participants from the present experiment are referred to as “C”, and those from the Watson and Sanderson (2001a, b) as “IT” for a group of IT postgraduates and “A” for a group of anaesthetists.

Primary task performance. For the primary task (arithmetic) we found no difference in accuracy and reaction time between the four conditions. As intended, the performance feedback and the instructions served to equalise primary task accuracy across conditions more than in the Watson and Sanderson (2001a, b) studies. Because no reaction time scores were taken for the primary task in the Watson and Sanderson experiments, no direct comparisons can be made for this measure.

Secondary task performance. For the secondary task (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). For absolute judgments, we find a significant effect of Modality, $F(3,8)=4.46$, $MSe=0.057$, $p<0.05$. Our results suggest that performance was best overall in the VV rather than the SS condition, the rankings from most to least accurate being VV, SV, SS, and VS. There are also significant effects of Parameter, $F(4,80)=7.11$, $MSe=0.04$, $p<0.01$, with changes in VT being less accurately detected in all conditions. The most notable feature of the results is a three-way interaction of Modality, Scenario cluster, and Parameter, $F(12, 80)=2.56$, $MSe=0.02$, $p<0.01$. Participants find it more difficult to detect changes in respiratory parameters in the SS and VS modalities than in other modalities, but exactly which respiratory parameters are worse varies across anaesthesia scenarios. For directional judgments there is no effect

of Modality. Even more than for absolute judgments, participants' ability to make directional judgments varies according to the anaesthesia scenario used. There are no other interpretable effects that include Modality.

Contrary to expectation, secondary task performance is slightly better overall than for the roughly comparable IT postgraduate group of Watson and Sanderson (2001a, b) (see second and third panels on Figure 3). This suggests that even with a between-subjects design, a faster primary task event rate (5" rather than 10"), and the need to turn around to view visual displays, the present participants managed to maintain a better, rather than worse, picture of patient status. It is interesting that in ongoing work with a high-workload 2.5" primary task event rate, Savill (2002) is finding a similar level of accuracy to the secondary task performance accompanied by faster primary task reaction time—not worse performance as we might have expected.

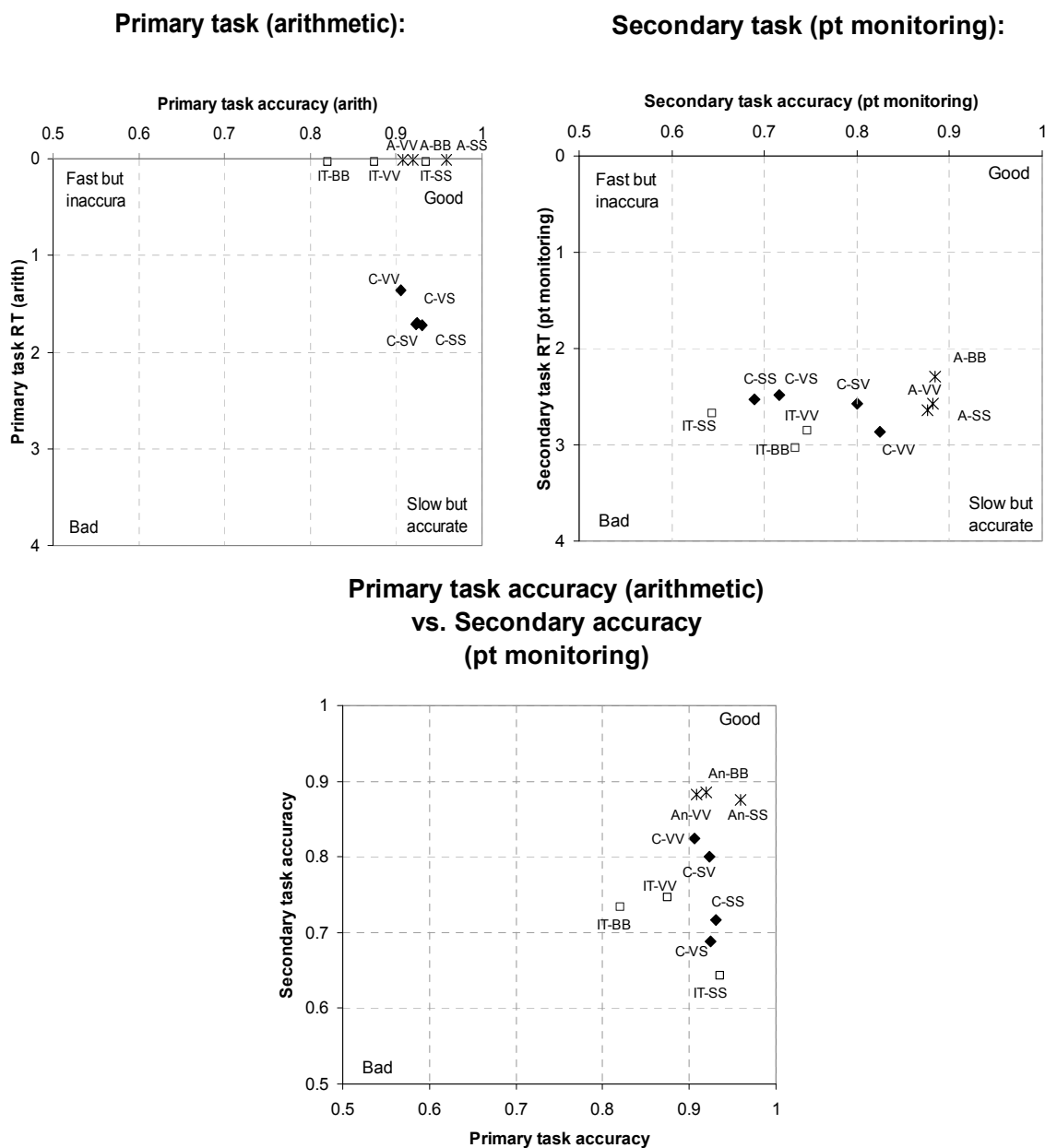


Figure 3. Primary and secondary task performances

The above results were obtained with a coding scheme that was based on responses that were correct and easy to for participants to discern when they had visual support in the experiment. The coding scheme therefore might appear to have been unreasonably biased against finding superiority for sonification. We have performed further analyses with an alternative scoring scheme that is more closely aligned with more nuanced changes in the sonified parameters, but the results are very similar to those presented above. We chose the present coding scheme in order to perform a direct comparison in this paper with the Watson and Sanderson (2001a, b) study.

Effect of two vs one sonification. We also wished to test whether two sonifications compromised the effectiveness of each sonification alone. We therefore compared secondary task performance on the cardiovascular (HR and O₂) parameters in the SS vs SV conditions, and on the respiratory parameters (RR, VT and CO₂) in the SS vs VS conditions. There was no evidence of differences in either case. This indicates that the presence of a parallel respiratory sonification does not compromise detection of cardiovascular status via pulse oximetry, and that the presence of a parallel pulse oximetry sonification does not compromise detection of respiratory status via a respiratory sonification.

Spontaneous comments. Participants were encouraged to provide comments about the status of the physiological parameters they were monitoring at any point during the experiment. Results indicate that when participants are supported by any degree of sonification they are much better able to provide relevant comments than when they are supported by the visual display alone, $F(3,20) = 4.60$, $MSe = 1029.5$, $p < 0.05$. The average number of comments for participants in the different conditions were as follows: SS=70; SV=78; VS=73; VV=17. In the VV condition, participants must infer parameter status from many successive numerical readouts, holding previous values in memory and possibly in the phonological loop of working memory. Because they are also dealing with numbers in the primary task, VV participants may have found it confusing to verbalise numerical information as it may have interfered with the primary task. In the SS, VS, and SV conditions there is a degree of sonification support so that some monitoring can be performed acoustically without remembering figures. Participants make many more comments under these conditions. However note that the SS condition did not produce more verbalisation than the SV and VS conditions.

Comparison with other experiments. There was evidence of some tactical trade-offs between primary and secondary tasks for non-anaesthetists in the Watson and Sanderson (2001a, b) experiment. The IT-SS condition showed worse secondary task performance but slightly better primary task performance than the other two IT conditions. In the present study, the C-SS condition also shows worst secondary task performance but only a non-significant increase in primary task accuracy.

4. Conclusion

The first two issues raised in the introduction were whether physiological sonification would lead to more accurate monitoring of physiological parameters and better performance on a further task timeshared with monitoring. Previous results suggested that requiring participants to rely more on the sonification might allow its benefits to become more apparent. In the present study, to our surprise, putting the patient monitoring visual display behind the participant and increasing the arrival rate of arithmetic tasks did not lead to the expected advantage for the SS display. Our observation was that participants were hypervigilant in all displays including the visual modality – probably much more vigilant than an anaesthetist would be in the OR context to which we are trying to generalise. In current work in our laboratory, Savill (2002) has performed another version of the present study but with an even faster 2.5” arrival rate of arithmetic tasks. Preliminary results suggest that there is still no absolute or relative advantage of the sonification with the coding scheme we are using. Despite the extra time pressure, the overall monitoring accuracy of Savill’s participants is as good as in the present study, rather than worse.

The third issue raised in the introduction was whether using two sonifications for monitoring would make it more difficult to discern the information in either sonification alone, so compromising overall performance. The SS results never fall into this pattern with respect to the SV and VS conditions .

The tendency towards worse patient monitoring performance with full sonification might appear to support the position that sonification is ineffective. This conclusion, however, may be premature because it is exclusive to non-anaesthetist populations. Non-anaesthetist populations appear to take advantage of full sonification to shift their attention toward the arithmetic task. Although participants were required to give priority to the arithmetic task, the fact that sonification allows them to very easily shift priority even further towards the primary task can come at the cost of patient monitoring performance. It is notable that anaesthetists perform the arithmetic task slightly better when patient information is sonified, but they do so *without* sacrificing patient monitoring performance. Anaesthetists not only perform patient monitoring more accurately the non-anaesthetist groups do, but they also show no difference in monitoring performance between the three modalities (Watson and Sanderson (2001a, b). Anaesthetists' ability to maintain patient monitoring performance may be due to any or all of the following: (1) standards of professionalism that do not allow them to sacrifice patient monitoring even when an experimental task calls for it, (2) knowledge of patient vital signs and patterns of signs that mean the sonification task together with the arithmetic task poses less overall workload, and (3) better strategic allocation of attention due to long experience dividing attention in a manner similar to the experimental demands.

One message that is reinforced by our findings is how important it is to conduct sonification studies with target populations. Our non-anaesthetist populations have been quite similar across studies, which suggests they may have trade-offs and priorities that are exclusive to non-anaesthetists. If we want to generalise about expected performances across modalities, it is important that we keep using anaesthetist participants as well as non-anaesthetists participants. It may be that for a non-anaesthetist population a sonification will be superior only under conditions that make visual displays extremely inconvenient to use. For non-anaesthetists, however, the sonification is seen more readily as a useful and meaningful cue to exploit to ensure patient safety.

5. Design implications

Overall, our findings have direct implications for the design of critical care environments for "eyes free" monitoring of patients. Our results suggest that participants can handle multiple modalities in real-time information displays. Although visual formats still attract the best performance when there is opportunity and motivation to refer to the visual display, as opportunity and motive are removed, so the advantages of auditory displays will emerge.

Which particular combination of modalities is best to use will depend upon the subjective priorities of participants and moment-by-moment cognitive load. As priorities and load change, participants may need to adapt their information seeking strategies, indicating that providing options as to which modality can be used may be the best design strategy. A great deal of practitioner adaptation occurs in complex work environments, practitioners often "tailoring" available technology and information resources to suit their immediate needs (Vicente, Roth, & Mumaw, 2001). In light of this, giving practitioners the flexibility to choose how their attention will be directed and how their cognitive resources will be absorbed may be an effective design philosophy under the right conditions.

6. Acknowledgments

We would like to express our appreciation to Swinburne University of Technology and in particular to John Craick and the Swinburne Computer Human Interaction Laboratory (SCHIL) for the use of their resources for this project. In addition, we would like to acknowledge the support of a Swinburne Research Development Grant and an ARC Discovery Grant DP0209952 to Professor Sanderson and to Dr W. John Russell of Royal Adelaide Hospital. Finally the first author would like to thank Annyck Savill for her contributions to the experiment reported in this paper.

7. References

- Crawford, J., (2002). *Monitoring the Anaesthetized Patient: A Preliminary Study of Multimodal Information*. Unpublished master's thesis, School of Information Technology, Swinburne University of Technology, Melbourne, Australia.
- 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
- Gopher, D. (1993). The skill of attention control: Acquisition and execution of attention strategies. In D. Meyer & S. Kornblum (Eds.). *Attention and Performance XIV: Synergies in Experimental Psychology, Artificial Intelligence, and Cognitive Neuroscience - A Silver Jubilee*. Cambridge, MA: MIT Press.
- 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
- Savill, A. (2002). *High cognitive workload and its effects on cardiac and respiratory monitoring with auditory and visual displays*. Bachelor's honours thesis in progress. School of Psychology, The University of Queensland, St Lucia, QLD, Australia.
- Seagull, F. J., & Sanderson, P. M. (2001). Anaesthesia alarms in surgical context: An observational study. *Human Factors*, **43**(1), 66-77.
- 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.
- Schneider, W., & Fisk, D.A. (1982). Concurrent automatic and controlled visual search: Can processing occur without cost? *Journal of Experimental Psychology: Learning, Memory & Cognition*, **8**, 261-278. BF1.J557
- Vicente, K. J., Roth, E. M., & Mumaw, R. J. (2001). How do operators monitor a complex, dynamic work domain? The impact of control room technology. *International Journal of Human-Computer Studies*, **54**, 831-856.
- Watson, M., Russell, W. J., & Sanderson, P. (1999). Ecological interface design for anaesthesia monitoring. *Proceedings of the 9th Australasian Conference on Computer-Human Interaction OzCHI99*. (pp 78-84).Wagga Wagga, Australia: IEEE Computer Society Press
- 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.
- Watson, M., Sanderson, P., & Russell, W. J. (2000). Alarm noise and end-user tailoring: The case for continuous auditory displays. *Proceedings of the 5th International Conference On Human Interaction With Complex Systems HICS2000*. (pp. 75-79).Urbana-Champaign, IL: U.S. Army Research Laboratory.
- Woods, D. D. (1995). The alarm problem and direct attention in dynamic fault management. *Ergonomics*, **38**, 2371-2393.