MONITORING THE ANESTHETIZED PATIENT: AN ANALYSIS OF CONFUSIONS IN VITAL SIGN REPORTS

Jennifer Crawford, Annyck Savill and Penelope Sanderson
ARC Key Centre for Human Factors and Applied Cognitive Psychology
The University of Queensland, St Lucia, Australia
crawford@itee.uq.edu.au

In the domain of anesthesia, auditory displays for respiratory monitoring have been suggested that could accompany the long-standing pulse oximetry system. However, research results to date suggest that participants' ability to identify physiological changes with respiratory sonification is not as great as it is for pulse oximetry. The reasons for this have not been established. Some possibilities are that respiratory parameters may be inherently more dynamic and may have greater intercorrelations, leading to greater confusability. In addition, there may be biasing factors specific to the various studies conducted that have led to slightly worse performance with respiratory sonification. In this paper we work through these possibilities and conclude that some are attributable to experimental design while others may be inevitable. The most important area of investigation in the future is to see whether the results are relevant to monitoring a synthesis of patient state.

INTRODUCTION

Sonification is the representation of data relations through relations in sound (Kramer, 1994). It allows eyes-free monitoring for information such as physiological parameters. It could therefore be very useful in such domains as anaesthesia where the practitioner's eyes may be relieved for other tasks (Watson & Sanderson, 2001, Sanderson, Anderson & Watson, 2000). The potential advantage of sonification over alarm systems is its ability not only to alert, but also to inform anaesthetists of patient state.

Researchers are developing sonifications for the anesthesia environment that extend the advantages of the pulse oximetry system (for heart rate, HR, and blood oxygenation, O2) to respiratory parameters. However, it seems that under some conditions participants do not identify physiological changes as accurately with some respiratory sonifications being developed as with the more familiar pulse oximetry (Loeb & Fitch, 2002). The focus of this paper will be to briefly discuss why respiratory sonification would be a useful tool and then to identify and evaluate the problems that have become apparent.

Pulse oximetry has been used for many years and has demonstrated its ability to support a high level of incident detection (Runciman, Webb, Barker & Currie, 1993). Data from the Australian Incident Monitoring study (Runciman, et al., 1993) indicate that if respiratory parameters (end-tidal carbon dioxide, ETCO2, respiration rate, RR, and tidal volume, Vt) could be sonified, then these parameters, together with pulse oximetry, would capture over 90% of evolving incidents without the need to look at the patient monitoring screen.

Several studies have tested the effectiveness of prototype respiratory sonifications alongside the pulse oximetry (Crawford, Watson, Burmeister & Sanderson, 2002; Fitch & Kramer, 1994; Loeb & Fitch, 2000, 2002; Watson & Sanderson, 2001). Three respiratory parameters—RR, Vt and ETCO2—are integrated into a single sound stream that plays alongside pulse oximetry. Levels of performance with respiratory sonification have been encouraging, but to date have not equaled the accuracy of pulse oximetry (Crawford, et al, 2002; Loeb & Fitch, 2002; Watson & Sanderson, 2001; 2002). Researchers have noted that participants sometimes experience confusion in identifying sonified respiratory parameters (Loeb & Fitch, 2002; Watson & Sanderson, 2001; Crawford et al., 2002). The reasons behind these problems need to be understood in order to develop a functional respiratory sonification to aid anaesthetists.

In summary, the overall goal of this research was to assess whether sonification helps participants identify the status of anesthetised patients when compared to visual displays. The specific goal of this paper was to examine the errors participants made when identifying the respiratory status of simulated patients with respiratory sonification.

METHOD

Participants

Participants were 40 postgraduate students from the University of Queensland with no clinical monitoring or medical experience. The purpose of using non-anesthetist participants was to help us identify the perceptual discriminability of our sonifications and the coherence of the experimental arrangements before testing anesthetists. The participants were paid for their participation.

Apparatus

The visual displays and sonifications were presented using the Arbiter simulation, which is a flexible display front end for the Body Anesthesia Simulator™. The arithmetic task was presented on a laptop. There were two sound streams for the auditory display: the first represented traditional pulse oximetry, whereas the second included the parameters RR, Vt and ETCO2. RR was represented directly as the rate of inhalation and exhalation, Vt in volume (loudness) of sound and ETCO2 by pitch (see Watson and Sanderson, 2001, for further information).

Design and Procedure

Design. The experiment was a between-subjects design. The between subjects factors were display modality (Display) and rate of presentation of a simultaneous arithmetic task (Rate). The participants were randomly assigned to conditions.

Participants were required to monitor the state of the simulated patient using one of the following unique display combinations for pulse oximetry (HR and O2) and respiratory (RR, Vt, and ETCO2) parameters, as shown below:

SS: pulse oximetry and respiratory parameters were all sonified
SV: pulse oximetry parameters were sonified and respiratory were visual
VS: pulse oximetry parameters were visual and respiratory were sonified
VV: all parameters were presented visually
BB: all parameters were presented both visually and through sonification.

Patient monitoring. For the patient monitoring task, participants monitored eight simulated anesthesia scenarios of approximately nine minutes duration, which were variants of those developed in Watson (2002). Participants monitored the patient physiological parameters that were displayed on screen or through speakers 180 degrees behind them. Because of the deliberate position of the patient monitoring equipment behind the participants, they had to turn to see visual information (see Figure 1).

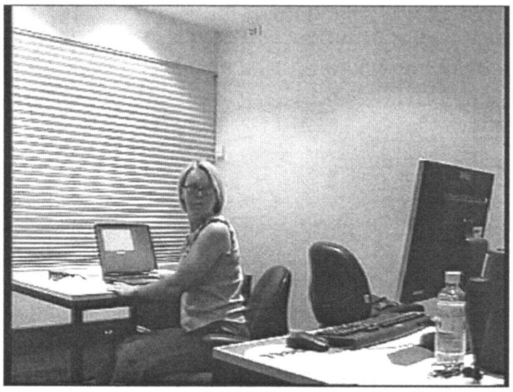

Figure 1. Participant turns from the arithmetic distractor task to examine the patient monitoring system behind them.

Approximately every minute the simulation would pause and a voice would name a parameter (probe). Participants were required to respond verbally whether that parameter was High, Low or Normal (Abnormality judgment) and whether it was Increasing, Steady or Decreasing (Directional judgment). Because sonification is intended to be informative about the patient's state, rather than simply providing normal/abnormal information, we tested participant's situational awareness with this finer-grained discrimination than asking them simply to report whether the patient was in a normal or abnormal state.

Arithmetic distractor task. Participants gave true/false answers to an arithmetic task using a key press while monitoring the patient. The arithmetic task was used to simulate drug dosage calculations that anaesthetists have to perform during an operation. Participants were told that the arithmetic task had primary importance and that they should answer it as quickly and accurately as possible, while maintaining awareness of the patient's state.

Half the participants in each Display condition received a new arithmetic task at a slow rate (Rate of every 5.0 seconds) and half at a fast rate (Rate of every 2.5 seconds). A graph of their speed and accuracy to this task was provided as feedback.

Prior to the experiment, participants trained for 40 minutes within their assigned Display and Rate condition, becoming familiar with the typical physiological events of the configuration they were testing.

RESULTS AND DISCUSSION

To put the present paper in context we will survey the basic findings of this study before examining patient monitoring accuracy in greater detail.

General performance results

Several ANOVAs were performed to assess the effect of Display and Rate on patient monitoring speed and accuracy, and on arithmetic distractor task speed and accuracy. Results for patient monitoring accuracy, arithmetic task RT and total time with head turned (not examined here) are shown in Figure 2. In what follows we highlight the main findings.

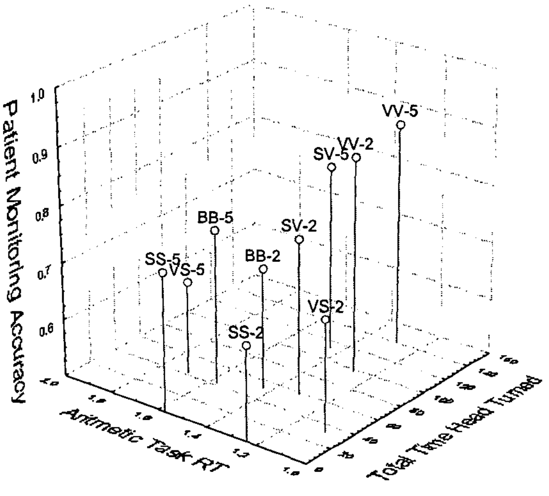

Figure 2: Results for patient monitoring, arithmetic distractor task, and time with head turned.

Participants presented with visual respiratory information (VV, SV) were most accurate and those with sonified respiratory information (BB, SS, and VS) were least accurate, $F(4, 30) = 9.8$, $MSE = 0.054$, $p < 0.001$.

However, participants identified patient status more quickly when the information was sonified than when it was in the visual modality, $F(4, 30) = 2.74$, $MSE=2.92$, $p < 0.05$.

Participants responded more slowly, $F(4, 120) = 20.22$, $MSE = 0.236$, $p < 0.001$, and less accurately, $F(4, 120) = 24.86$, $MSE = 0.026$, $p < 0.001$, to questions about the respiratory parameters (RR, VT and ETCO2) than to questions about the pulse oximetry parameters (HR and O2). Responding was least accurate of all $F(16, 120) = 5.61$, $MSE = 0.026$, $p < 0.001$ when the respiratory parameters were sonified.

There were no significant differences for the arithmetic accuracy across the various modalities, $F(4, 30) = 0.32$, $MSE = 0.006$, ns. However participants responded more quickly to the arithmetic questions in the fast arithmetic task condition, $F(1, 30) = 9.19$, $MSE = 0.131$, $p < 0.01$.

Clearly, participants had some trouble responding to probes asking about respiratory status, particularly when the respiratory parameters were sonified. Although such effects may be attenuated for trained anesthetists (Watson & Sanderson, 2001) and although we are focusing on low-level data reports of individual parameters rather than overall clinical trends of the patient state (Bennett, Payne, Calcaterra, & Nittoli, 2000), this finding bears further investigation.

Frequencies

		Responses			
Stimuli		H	N	L	ALL
HR	H	201	22	17	240
	N	11	224	5	240
	L	0	0	0	0
	ALL	212	246	22	
O2	H	0	0	0	0
	N	1	235	4	0
	L	4	78	156	240
	ALL	5	313	160	
RR	H	61	49	8	478
	N	11	218	10	0
	L	3	10	106	118
	ALL	75	277	124	
VT	H	26	13	0	119
	N	37	111	12	476
	L	25	130	124	0
	ALL	88	254	136	
CO2	H	72	39	8	160
	N	26	193	21	279
	L	9	46	65	478
	ALL	107	278	94	

Proportions

		Responses		
Stimuli		H	N	L
HR	H	0.84	0.09	0.07
	N	0.05	0.93	0.02
	L	0.00	0.00	0.00
	ALL	0.88	1.03	0.09
O2	H	0.00	0.00	0.00
	N	0.00	0.98	0.02
	L	0.02	0.33	0.66
	ALL	0.02	1.31	0.67
RR	H	0.52	0.42	0.07
	N	0.05	0.91	0.04
	L	0.03	0.08	0.89
	ALL	0.59	1.41	1.00
VT	H	0.67	0.33	0.00
	N	0.23	0.69	0.08
	L	0.09	0.47	0.44
	ALL	0.99	1.49	0.52
CO2	H	0.61	0.33	0.07
	N	0.11	0.80	0.09
	L	0.08	0.38	0.54
	ALL	0.79	1.52	0.70

Table 1. Comparison of correct responses (rows) and participant responses (columns) to parameter states. Frequencies at left, probabilities at right. Cells with dark backgrounds indicate unusually high levels of errors. CO2 is end-tidal CO2, or ETCO2.

Identification of patient status

In this section we look at the participants' judgments of patient status more closely in order to get a clearer idea of why accuracy was lower for the respiratory variables. Six possible causes were identified, which we will consider in turn in an *exploratory* rather than a confirmatory analysis. Sometimes it is impossible to attribute the findings to just one cause because there may be more than one contributing to the result.

1. Negative ("Normal") response bias. For all parameters, participants had a tendency towards responding "Normal" that was approximately equivalent for pulse oximetry and respiratory parameters when combined. However, as Table 1 shows, the tendency was driven largely by O2 in the pulse oximetry parameters and Vt in the respiratory parameters. The tendency was already present when O2 or Vt was displayed visually, but was stronger when it was sonified. It is not clear whether participants genuinely thought that a parameter was within the normal range, or whether they "guessed" normal when they weren't sure. Overall, the bias was not stronger for respiratory parameters. Therefore, the results to date that have suggested respiratory performance is worse than pulse oximetry is not explained by the tendency to report normal.

2. More possible respiratory states. Because normal oxygenation is 100%, O2 has only two states: normal and low. For the current experiment, though, it happened that HR never went into the low range and only had two states: normal and high. The respiratory parameters had three states High, Normal and Low. Perhaps worse performance with respiratory parameters was due to greater response uncertainty. As Table 2 shows, if we examine cases where Normal was the *correct* response, and see if there were proportionally more wrong answers with the respiratory than the pulse oximetry parameters, the greater response uncertainty will have led to more errors. As Table 2 shows, participants were less likely to say "Normal" correctly to probes about respiratory status than about pulse oximetry status.

Frequencies

	Responses			
	H	N	L	
CV	12	459	9	480
RE	74	522	43	639

Proportions

	Responses		
	H	N	L
CV	0.03	0.96	0.02
RE	0.12	0.82	0.07

Table 2: For questions about pulse oximetry (CV for cardiovascular) and respiratory (RE) parameters when the true state is normal, the frequency or proportion of participant responses in the high, normal or low categories.

3. More changes in respiratory parameters. Worse responding to respiratory parameters may be because they changed more often, making tracking more difficult. We looked at the number of times each parameter changed across the whole experiment. As Table 3 shows, the HR and O2 parameters had fewest changes whereas respiratory parameters experienced the most changes, when taken individually and all together. There is therefore more volatility in the respiratory parameters.

Parameter	HR	O2	RR	VT	CO2
No. of changes	57	105	244	171	307

Table 3. Number of changes per parameter for entire experiment. CO2 is end-tidal CO2, or ETCO2.

4. Rushed responding in fast arithmetic condition. It may be that sonified respiratory parameters were particularly hard to identify when there was time pressure in the fast arithmetic conditions. Normal breath lengths are around 5 seconds (12 breaths per minute is a typical normal rate) whereas the arithmetic task needed to be attended to every 5.0 or 2.5 seconds. Perhaps it was harder to monitor respiration under the fast arithmetic rate, especially when the respiratory parameters were sonified?

Table 4 shows the number of errors in identification for each of the five parameters in the fast and slow arithmetic condition, dividing the results into cases in which the respiratory parameters were sonified (SS and VS) at left, or in visual format (VV and SV) at right. When the respiratory parameters are sonified there seems to be no tendency for RR and ETCO2 to show more errors with the fast arithmetic task, but for Vt there were slightly more errors in the fast arithmetic task. When the respiratory parameters are in visual form, the trend for more Vt errors with the fast arithmetic task is stronger. Overall, then, we cannot say that the arithmetic task rate is having more effect on respiratory parameters when they are sonified than when they are displayed visually—it is rather the other way around. (BB was not included in this analysis as it contained both visual and acoustic displays which could not be isolated for a direct comparison.)

Resp sonified				Resp in visual format			
	fast	slow	diff		fast	slow	diff
HR	12	6	*6*	HR	16	14	2
O2	16	11	*5*	O2	12	17	-5
RR	27	32	*-5*	RR	9	4	5
VT	57	47	*10*	VT	43	20	*23*
CO2	43	44	*-1*	CO2	22	12	*10*
	155	140			102	67	

Table 4: Number of errors for each parameter divided into sonified respiratory parameters (left) and visual respiratory parameters (right) and further divided into fast and slow arithmetic task rate. CO2 is end-tidal CO2, or ETCO2.

5. Response to the wrong parameters. Some errors may have happened if participants were asked about one parameter but gave a response that was correct for another parameter. This could be due to systematic confusion or a temporary slip. Given the greater complexity of the respiratory parameters, such errors could have been more prevalent with respiratory than with pulse oximetry parameters. It is difficult to determine the level of resulting error because some of the other explanations given above may be involved. For example, when the participant answers "normal", it may be due to general bias towards saying normal or because other parameters are within the normal range. However, if a probed parameter is high, but the participant says "low" and other parameters are low, the participant may have responded to the wrong parameter or may be influenced by the state of the other parameters.

An investigation of the extreme respiratory answers was conducted to see if there were any "lures" (other parameters that may have influenced the response) (see Table 5). We found that when Vt was low and participants labeled it high, another parameter was high for all cases. These included HR, RR and ETCO2. When ETCO2 was low and participants said high, either HR or RR was high. It was found that when ETCO2 was high, participants said low, and for all cases Vt was low. This would suggest that Vt, HR and RR parameters may be mistaken for ETCO2 and that ETCO2, RR and HR may be mistaken for Vt. Examples of participants comments were: "I was trying to distinguish ETCO2 (pitch) from Vt so if asked I would know. Both are not clear to me." "It is hard to identify Vt and ETCO2."

Probe	State	Response	Cases	Lures	Prop
Vt	L	"H"	26	26	1.00
CO2	H	"L"	8	8	1.00
CO2	L	"H"	9	9	1.00
RR	H	"L"	8	6	0.75
RR	L	"H"	3	0	0.00

Table 5. Comparison of correct response (State) and participant response (Response) and possible number of lures for each. CO2 is end-tidal CO2, or ETCO2.

6. Probes may cut off parameter changes. Probes sometimes appeared very soon after a parameter change, so that the new state may have been missed. This happened for one out of 34 pulse oximetry (3%) and for four out of 46 possible respiratory probes (9%). Therefore, proportionately more probes of respiratory status than of cardiovascular status (through the pulse oximetry system) were compromised by this problem.

CONCLUSIONS

Our goal in this investigation was to examine the errors participants made when using sonification to identify the respiratory status of patients. Because other researchers have found similar effects (Loeb & Fitch, 2002) we must discriminate the degree to which the errors are due to our experimental arrangements as opposed to inherent properties of respiration in the anesthetized patient that might be the same across different studies.

The following two reasons were eliminated as possible explanations. First, we thought that participants' tendency to respond "normal" might be more pronounced for respiratory variables than for pulse oximetry variables. This was not the case. Second, we thought that responses to respiratory probes might be less accurate in the fast arithmetic condition. This was not the case either.

The following reason was supported as a possible explanation, but it was specific to our experiment. Approximately 3% of pulse oximetry probes were compromised by a state change in the to-be-probed parameter too soon before the probe, whereas approximately 9% of possible respiratory probes were compromised in this manner.

Further reasons were supported as possible explanations but could also be found in other experiments in the clinical context. First, there were more changes in the respiratory parameters. This could be standardized in future tests to see if performance for the pulse oximetry and respiratory parameters is equivalent if the number of changes, or dynamism, is also equivalent. Second, responses to respiratory probes were sometimes the reverse of expected ("H" when L and "L" when H), and appeared to have been captured by changes in an unprobed parameter. We were not able to compare the pulse oximetry and respiratory parameters since responses for pulse oximetry could not conform to the pattern above (states were H and N for HR and L and N for O2). This could be tested by using scenarios in which HR has low as well as high states. It can also be tested with sonifications in which blood pressure—a cardiovascular variable—is sonified. This would also control for another possible cause—the fact that there were fewer cardiovascular states and therefore less possibility for confusion. Finally, it can be tested by comparing performance for pulse oximetry versus respiratory parameters in scenarios in which potential lures do versus do not occur.

In summary, there are reasons to believe that when participants are tested for their ability to identify the status of probed physiological parameters, they may inevitably be less accurate for respiratory than for pulse oximetry parameters and this is only slightly affected by whether respiration is sonified. The degree to which the performance with respiratory sonification is less accurate will depend upon how well other factors specific to an experiment have been controlled. The slightly worse performance with respiratory sonification does not seem to be due to the fact that we used two rather than three variables in the pulse oximetry stream (HR and O2 vs HR, BP, and O2). Loeb and Fitch (2002) conducted an experiment using three parameters in both sound streams (HR, BP, and O2 vs RR, Vt, and ETCO2) and the problems found in the respiratory sonification were still evident.

It is important to note that our experiment focused solely on reports of low-level parameter states rather than a high-level synthesis of patient state. Bennett et al (2000) have shown that the degree of relational judgment and variable integration required to give a response can strongly affect the pattern of results. It may be that when participants—especially knowledgeable participants—are asked to draw higher-level inferences from low-level data, the apparent difference in accuracy between ability to report pulse oximetry and respiratory parameters may disappear. Moreover, unique benefits of sonification for respiration may appear. The very factors that make it hard to disentangle the status of individual respiratory parameters may make it easier to identify integrate patient states. The Loeb and Fitch (2002) dataset could possibly throw light on this, but there is not enough detail in their report to make an initial determination of the result. This is the area of greatest need in future work and we are examining it in a series of studies at present in our laboratory.

REFERENCES

Bennett, K., Payne, M., Calcaterra, J., & Nittoli, B. (2000). An empirical comparison of alternative methodologies for the evaluation of configural displays. *Human Factors*, 42, 287-298.

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

Kramer, G. (1994). Some organizing principles for representing data with sound. In G.Kramer (Ed.). *Auditory display: Sonification, audification and auditory interfaces* (pp.185-221). Sante Fe Institute studies in the sciences of complexity: Proceedings Volume XVIII. Reading, MA: Addison-Wesley.

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.* (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 Orlando. Abstract from anestech.org/publications. File: Annual_2000/Loeb.html

Loeb, R. G., & Fitch, W. T. (2002). A laboratory evaluation of an auditory display designed to enhance intraoperative monitoring. *Anesthesia and Analgesia*, 94, 362-368.

Runciman, W. B., Webb, R. K., Barker, L., & Currie, M. (1993). The Pulse Oximeter: Applications and Limitations – An analysis of 2000 incident report. *Anaesth. Intens. Care*, 21, 543-550.

Sanderson, P., Crawford, J., Savill, A., & Watson, M. (2003). Visual and auditory attention in patient monitoring: A formative analysis.. Manuscript submitted for publication.

Watson, M. (2002). *Sonification and anesthesia: Ecological design and empirical evaluation*. Doctoral dissertation, School of Information Technology, Swinburne University of Technology, Melbourne, Australia.

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.