BJA

British Journal of Anaesthesia, 125 (5): 826–834 (2020)

doi: 10.1016/j.bja.2020.05.038
Advance Access Publication Date: 15 July 2020
Respiration and the Airway

Evaluation of an enhanced pulse oximeter auditory display: a simulation study

Estrella Paterson[1,*], Penelope M. Sanderson[1,2,3], Isaac S. Salisbury[1], Felicity P. Burgmann[1], Ismail Mohamed[1], Robert G. Loeb[1,4] and Neil A. B. Paterson[3,5]

[1]School of Psychology, The University of Queensland, St Lucia, Queensland, Australia, [2]School of Information Technology and Electrical Engineering, The University of Queensland, St Lucia, Queensland, Australia, [3]School of Clinical Medicine, The University of Queensland, St Lucia, Queensland, Australia, [4]Department of Anesthesiology, University of Florida College of Medicine, Gainesville, FL, USA and [5]Anaesthesia and Pain Management Services, Queensland Children's Hospital, South Brisbane, Queensland, Australia

*Corresponding author. E-mail: estrella.paterson@uqconnect.edu.au

Preliminary results of subjective judgments from the study were presented at the 25th International Conference on Auditory Display (ICAD2019), 23–27 June 2019, Newcastle-upon-Tyne, UK.

Abstract

Background: We compared anaesthetists' ability to identify haemoglobin oxygen saturation (SpO_2) levels using two auditory displays: one based on a standard pulse oximeter display (varying pitch plus alarm) and the other enhanced with additional sound properties (varying pitch plus tremolo and acoustic brightness) to differentiate SpO_2 ranges.
Methods: In a counter-balanced crossover study in a simulator, 20 experienced anaesthetists supervised a junior colleague (an actor) managing two airway surgery scenarios: once while using the enhanced auditory display and once while using a standard auditory display. Participants were distracted with other tasks such as paperwork and workplace interruptions, but were required to identify when SpO_2 transitioned between pre-set ranges (target, low, critical) and when other vital signs transitioned out of a target range. They also identified the range once a transition had occurred. Visual displays were available for all monitored vital signs, but the numerical value for SpO_2 was excluded.
Results: Participants were more accurate and faster at detecting transitions to and from the target SpO_2 range when using the enhanced display (100.0%, 3.3 s) than when using the standard display plus alarm (73.2%, 27.4 s) ($P<0.001$ and $P=0.004$, respectively). They were also more accurate at identifying the SpO_2 range once a transition had occurred when using the enhanced display (100.0%) than when using the standard display plus alarm (57.1%; $P<0.001$).
Conclusions: The enhanced auditory display helps anaesthetists judge SpO_2 levels more effectively than current auditory displays and may facilitate 'eyes-free' monitoring.

Keywords: auditory perception; data display; multitasking; operating theatre; oxygen saturation; patient monitoring; patient safety; pulse oximetry; sonification

Received: 23 January 2020; **Accepted:** 21 May 2020

Editor's key points

- The detection of perioperative changes in SpO_2 is usually based on simple auditory variations in tone rate and pitch and an alarm.
- In a demanding clinical setting, tone rate and pitch might be difficult to judge.
- This simulation study compared a standard simple auditory display plus an alarm with an enhanced display for the detection of changes in SpO_2.
- Under simulated tasks, experienced anaesthetists were more accurate and faster at identifying auditory deviations in SpO_2 when using the enhanced display.

Hypoxaemia is associated with cardiac arrest and arrhythmias, postoperative infection, cognitive impairment, cerebral ischaemia, organ failure, and mortality.[1,2] Despite the availability of continuous oxygen saturation (SpO_2) monitoring, intraoperative hypoxaemia still occurs. An analysis of 95 407 anaesthesia records revealed that 6.8% of patients suffered a hypoxaemic event (SpO_2 <90%) and 3.5% of patients suffered a serious hypoxaemic event (SpO_2 <85%) lasting 2 or more consecutive minutes during surgery.[3] More recent research investigating pre-oxygenation of 2398 patients before induction revealed that hypoxaemia (SpO_2 < 95%) at any time during surgery occurred in 6.6% of patients.[4] Anaesthetists do not constantly view visual displays of vital signs because they are engaged in other tasks that require visual attention. Although some hypoxaemia seems unavoidable, an auditory display that supports earlier and more consistent detection of desaturation might promote faster intervention and potentially decrease the incidence and severity of harm from hypoxaemia.

Anaesthetists access visual displays of patient monitors only 5–30% of intraoperative time.[5,6] The auditory display of the pulse oximeter allows anaesthetists to monitor patients' SpO_2 levels and heart rate (HR) in peripheral awareness, while engaged in other visually and cognitively demanding tasks.[7] Current pulse oximeter auditory displays map tone rate to HR and tone pitch to SpO_2, although in different ways depending on the make and model type.[8] Current displays are supplemented with auditory alarms that sound when SpO_2 crosses clinical thresholds.

However, anaesthetists sometimes have difficulty judging SpO_2 values and changes in SpO_2 levels. Experienced anaesthetists listening to a Datex AS/3 pulse oximeter (Datex-Ohmeda, Helsinki, Finland) in a quiet room could not reliably judge whether SpO_2 changed until it had changed by 8%.[9] Further research shows that anaesthetists have difficulty identifying changes in SpO_2 from 97% to 96% (in both directions) and difficulty identifying absolute SpO_2 levels using a pitch-only display.[10] Moreover, as visual attention load increases and noise level rises, anaesthetists become less able to identify a change from 99% to 98% saturation when listening to a Philips patient monitor (Model MP70; Philips Electronics, Amsterdam, The Netherlands).[11]

Our group has added the sound dimensions of tremolo and acoustic brightness to the varying pitch display to distinguish SpO_2 ranges for adult patients[10,12,13] and neonates.[14–16] In laboratory evaluations, clinician and non-clinician participants identified SpO_2 parameters more accurately when using an enhanced display than when using a variable-pitch only display. In the present study, undertaken in the more realistic environment of a simulated operating theatre, we tested the hypothesis that anaesthetists would identify SpO_2 parameters more accurately when using the enhanced auditory display than when using the standard auditory display with an alarm.

Methods

This study received ethical approval from the Children's Health Queensland Human Research Ethics Committee (HREC/18/QRCH/93) and The University of Queensland (Clearance Number: 2018000599). It was pre-registered with the Open Science Framework on December 19, 2018 (osf.io/5w769).

Participants

Twenty consultant anaesthetists working at a tertiary paediatric hospital were recruited for the study. All participants reported normal hearing; no participants were excluded on this basis. Those who completed the study received a small gift (less than $10 Australian) and could apply for continuing professional development points from the Australian and New Zealand College of Anaesthetists.

Design

In a crossover design, participants monitored SpO_2 using the enhanced auditory display during one scenario and the standard auditory display plus alarm during the other scenario. To control for sequence and practice effects, participants were randomly assigned to orders of presentation of the two displays by randomly sorting an equal number of presentation orders using MS Excel's RAND() function (Microsoft Corp., Redmond, WA, USA).

Study context

Participants sat at a desk near the operating field and supervised a junior colleague providing anaesthesia care for a surgical case.[17] They were told that their colleague was being video-recorded for assessment purposes. The junior colleague role was played by a consultant anaesthetist (author NP) and confederates acted as anaesthetic nurse, surgeon, and scrub nurse. The simulator was arranged as an operating theatre with a paediatric patient manikin (Laerdal, Stavanger, Norway) on the operating table, anaesthetic machine, patient monitor (Phillips IntelliVue MX450, Amsterdam, The Netherlands), drug trolley, and scrub table. A laptop was positioned on the desk for the participant to perform a simulated patient categorisation task (Fig. 1).

Anaesthetists monitor vital signs other than SpO_2 during procedures. Therefore, participants also identified HR, blood pressure (BP), and end-tidal carbon dioxide ($ETCO_2$). A visual display of patient vital signs was positioned in the participant's view showing waveforms plus numerical values for HR, BP, and $ETCO_2$ but only waveform for SpO_2 (Fig. 1, Monitor A). A similar visual display of patient vital signs that included the numerical values for SpO_2 was available to the operating team but out of the participant's sight (Fig. 1, Monitor B). The auditory display for SpO_2 was either the standard display or the enhanced display, and the auditory display for the other vital signs was a threshold alarm. Scenarios were deterministic

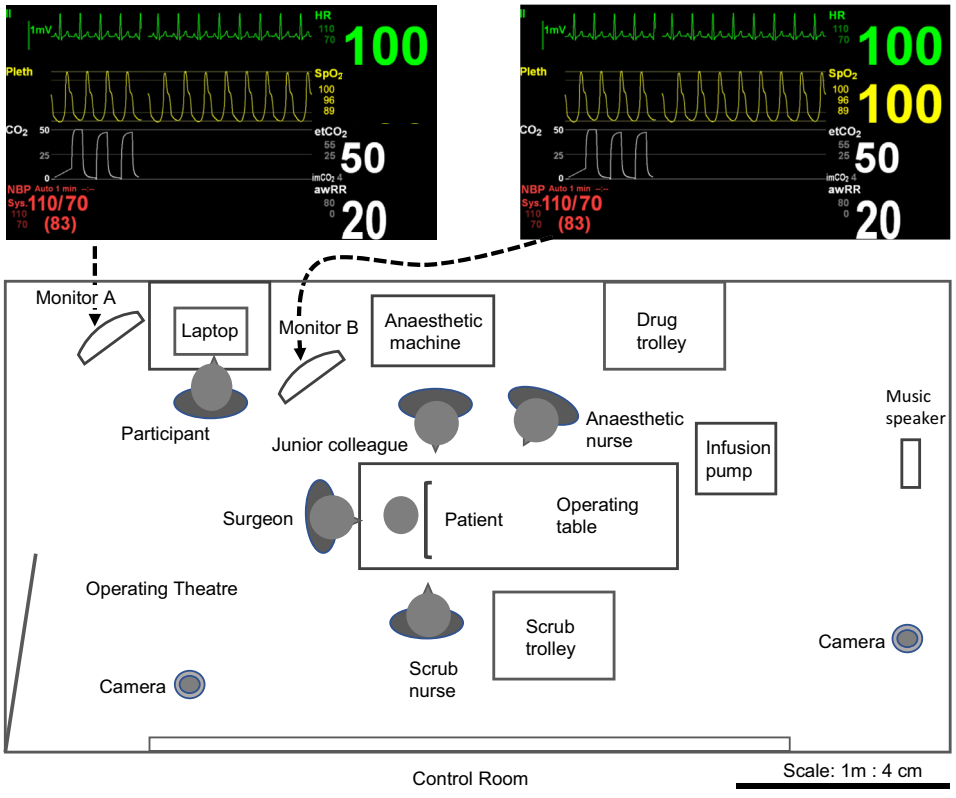

Fig 1. Layout of operating theatre simulator showing locations of participant, actors, and equipment.

with fixed timing of events. The simulator coordinator (author EP) maintained the actors' timing with instructions via earpieces connected to walkie-talkies.

Apparatus and stimuli

Vital sign ranges

SpO_2 ranges were classified as target (100–97% SpO_2), low (96–90% SpO_2), and critical (89–80% SpO_2). Normal range for HR was 71–109 beats min^{-1}, low was <71 beats min^{-1}, and high was >109 beats min^{-1}. Normal range for mean BP was 71–109 mm Hg, low was <71 mm Hg, and high was >109 mm Hg. For $ETCO_2$ the normal range was ≤54 mm Hg and high was >54 mm Hg.

Pulse oximeter auditory displays

The underlying auditory dimensions of the enhanced and standard displays were the same as those used in previous research (Table 1).[13] For both auditory displays, the frequency (experienced as pitch) of pulse tones was mapped to percentage saturation using a logarithmic mapping, with each 1%

Table 1 Sound properties of auditory displays for SpO_2 monitoring.

Sonification and range	SpO_2 values (%)	Sound effects	Sound properties
Standard			
Target	100–97	Variable pitch	Pure sine wave
Low	96–90	Variable pitch	Pure sine wave
Critical	89–80	Variable pitch plus IEC-60601-1-8 Medium-General alarm	Pure sine wave + alarm (sounds at transition from Low to critical and every 2 min while SpO_2 critical)
Enhanced			
Target	100–97	Variable pitch	Pure sine wave
Low	96–90	Variable pitch + tremolo	Sine wave + 4 cycles tremolo per tone (90% wet)
Critical	89–80	Variable pitch + tremolo + brightness	Sine wave + 4 cycles of tremolo per tone (90% wet) + 3rd, 5th, 7th harmonics

increase in SpO_2 mapped to a 1.84% increment in frequency. The pulse tones started at 655 Hz for 80% SpO_2 and ended at 950 Hz for 100% SpO_2. Each tone was 150 ms, with a 10 ms fade in and 10 ms fade out.

In the enhanced display, pulse tones below 97% SpO_2 were modulated with four cycles of tremolo to produce a vibrating effect, and the first three odd harmonics of the fundamental tone were added to pulse tones below 89% SpO_2 to produce a 'bright' effect.

Auditory alarm

In the standard display condition, but not in the enhanced display condition, an alarm (IEC-Medium-General alarm [IEC-60601-1-8])[18] sounded when SpO_2 decreased to less than 90% and every 2 min while it remained in the critical range. The same alarm sounded in both display conditions when HR, BP, or $ETCO_2$ went out of the normal range and every 2 min while they remained out of normal range.

Scenarios

Two scenarios and relevant patient case notes were designed by subject matter experts (authors RL and NP; see Supplementary materials 1 and 2). One scenario was a laryngeal papilloma resection and the other was bronchoscopy for a foreign body removal; each lasted 20 min. Airway cases were chosen because they allow SpO_2 to vary across a range of values. Table 2 shows the vital sign transitions for each scenario. Based on the scenario design, vital sign levels for SpO_2, HR, BP, and $ETCO_2$ were entered into the simulator's software (LLEAP Version 6.4.1; Laerdal), and the monitor screens were then recorded as video files that were played during the experiment (see Supplementary materials 3–7). During each scenario, participants were interrupted twice: once via telephone (patient information request) and once directly (social request).

Table 2 Vital sign transitions for SpO_2, heart rate, blood pressure, and $ETCO_2$ for Scenarios A and B.

Vital sign	Scenario A	Scenario B
SpO_2		
Total transitions	10	7
Target-low	4	2
Low-target	3	2
Low-critical	2	2
Critical-low	1	1
Heart rate		
Total transitions	4	2
Normal–high	2	1
Normal–low	2	1
Blood pressure		
Total transitions	2	2
Normal–high	1	1
Normal–low	1	1
$ETCO_2$		
Total transitions	2	2
Normal–high	2	2

Distractor task

Throughout the scenarios, participants performed a patient categorisation task that emulated a departmental audit of cases performed by individual practitioners. They were instructed to enter case details (patient name, date of birth, procedure, service provider) from a hard copy list into a categorised spreadsheet presented on the laptop.

Background music

Vocal pop music was played through a portable speaker (JBL, Los Angeles, CA, USA) at a volume that did not mask the auditory displays or alarms (see Supplementary materials 8). The average combined sound pressure level at the participant's desk was 66.4 dB(A) (Decibel X application, version 8.1.3; Skypaw Co. Ltd, Hanoi, Vietnam), which falls within the range of average noise levels of 51–77 dB(A) measured in operating theatres.[19]

Questionnaires

At the beginning of the experiment, participants provided information about their age, sex, musical training, and hearing ability. After each scenario, they completed a questionnaire using 9-point Likert items probing their attitudes to the auditory display just experienced and to the categorisation task. At the end of the experiment, participants completed a qualitative questionnaire probing their strategies for monitoring SpO_2 parameters and their opinions of the displays.

Procedure

At the beginning of each session and in a room separate from the simulator, the participant was trained to (1) identify SpO_2 values and ranges using one of the auditory displays; (2) identify HR, BP, and $ETCO_2$ values and ranges using visual display and auditory alarms; and (3) perform the distractor task. Participants were instructed that during scenarios they were to verbally announce if vital signs changed range. For parameters other than oximetry, we instructed participants to identify transitions (normal to either low or high ranges) indicated by the auditory alarm.

Charts showing vital sign ranges and corresponding responses were available in the training and simulator rooms. The participant was provided with case notes for the patient in the scenario that followed and then they were taken to the simulator, introduced to the operating team, and the operating theatre set-up was explained.

During scenarios, participants stated when SpO_2 transitioned from one range to another. Specifically, they noted when SpO_2 moved from target to low or *vice versa* (SpO_2 target transitions), and when it moved from low to critical or *vice versa* (SpO_2 critical transitions). The measure of all SpO_2 range transitions was the combination of the above. Participants also stated when any other vital sign left its target range (transitions out of target range for other vital signs). Participants were also asked to identify the range SpO_2 moved to once an SpO_2 range transition had occurred (SpO_2 range identifications) and the range other vital signs moved to when they left the target range (range identifications for other vital signs).

Throughout the scenarios, participants worked on the distractor task, responded to interruptions, and interacted with the junior colleague. After the first scenario, the

participant went back to the training room and filled out a questionnaire about the scenario. They were then trained on the second auditory display, went to the simulator to complete the second scenario, and filled out questionnaires in the training room.

Study outcomes

The primary outcome measure was participants' accuracy at detecting SpO_2 target transitions. Secondary outcomes included (1) accuracy at detecting SpO_2 critical transitions; (2) accuracy at detecting all SpO_2 range transitions; (3) accuracy at detecting transitions out of target ranges for other vital signs; (4) latency to detect SpO_2 target transitions; (5) latency to detect SpO_2 critical transitions; (6) latency to detect all SpO_2 range transitions; (7) latency to detect transitions out of target range for other vital signs; (8) accuracy of SpO_2 range identifications; (9) accuracy of range identifications for other vital signs; (10) distractor task performance; and (11) subjective judgments about auditory displays and tasks. Secondary outcome measures were exploratory in nature; therefore, statistical corrections for multiple comparisons were not made.

For each vital sign, transition detections were counted if they were correctly identified before the next change in range for that vital sign occurred (true positive). If the participant failed to identify a transition that had occurred, this was recorded as a false negative. If the participant identified a transition correctly but then said another transition had occurred before the next actual transition, this was recorded as a false positive.

Each participant's accuracy at *detecting* range transitions was defined as the number of correct identifications of a transition of a certain type (true positives) divided by the total number of transitions of that type, and converted to a percentage. Accuracies were then averaged across all participants. The calculation was performed for the scenario with the standard display and for the scenario with the enhanced display. Accuracies for *identifying* the range once a transition had occurred were calculated in a similar manner. Latency is defined for each participant as the average interval between each transition's occurrence and the participant's identification, for all identified transitions, expressed in seconds.

Sample size

Using G*Power,[20] we performed a power analysis using the *t*-test family for dependent means. Estimated standard deviations came from a prior simulator study[17] and estimated effect size (d=1.77) came from a prior laboratory study that measured SpO_2 target transition detection accuracy.[13] A two-tailed within-subjects test (difference between two dependent means) with power=0.8 and alpha=0.05, revealed that a sample size of 18 participants would show an effect if it existed. We tested 20 participants.

Data analysis

Each of the simulation sessions was audio–video recorded. Recordings were coded for verbal responses using Datavyu (Datavyu; 2014, Version 1.2.2, http://datavyu.org); accuracy and latency measurements were extracted from these coded files. A second investigator coded eight scenario recordings to assess inter-rater reliability (IRR). Cohen's kappa was used to determine IRR for the primary outcome of accuracy to detect target transitions.[21] Qualitative questionnaire responses about the displays were analysed using Linguistic Inquiry and Word Count software (LIWC [2015]; http://liwc.wpengine.com).[22]

All statistical analyses were performed using SPSS (version 25; IBM, Chicago IL, USA). We evaluated normality of residuals using Q–Q plots. If the data did not conform to the assumptions of the general linear model, data were analysed using non-parametric tests. If the distribution of differences between paired observations was normal and symmetrical, we used the Wilcoxon signed rank test. If the distribution of differences between paired observations was non-normal or non-symmetrical we used the sign test. We used Poisson regression to analyse the number of data entries made by participants for the patient categorisation task in each display condition.

Results

Twenty anaesthetists participated in the experiment (mean age=43 yr, range=34–53 yr, 40% female, with 55% having 1 yr or more of formal music training). Results for accuracy and latency of responding to range transitions and accuracy of range identification are shown in Table 3. Results for subjective judgements are shown in Table 4.

Primary outcome

Participants detected SpO_2 target transitions more accurately when using the enhanced display (100.0% [100.0, 100.0]) than when using the standard display plus alarm (73.2% [50.0, 75.0]), (P<0.001). The IRR was 1.0, indicating perfect agreement.

Secondary outcomes

Vital sign parameters

Participants detected all SpO_2 transitions taken together more accurately with the enhanced display than with the standard display plus alarm (P<0.001). Participants were slightly better at detecting critical transitions with the enhanced display than the standard display (P=0.041). There was no difference between display conditions for participants' accuracy at detecting transitions out of target ranges for the other vital signs (P=1).

Participants were faster at detecting SpO_2 target transitions, SpO_2 critical transitions, and all SpO_2 transitions with the enhanced display than with the standard display plus alarm (P=0.004, P<0.001, and P<0.001, respectively). There was no difference across display conditions in how quickly participants detected transitions out of the target ranges for the other vital signs (P=0.550).

Participants identified SpO_2 range once a transition had occurred more accurately with the enhanced display than with the standard display plus alarm (P<0.001). There was no difference across display conditions in participants' accuracy at identifying the range of other vital signs after a transition out of target range had occurred (P=1).

Distractor task performance

The display condition had a significant association with the number of entries participants made in the patient categorisation task (b=0.207, standard error (SE)=0.037, P<0.001). Participants made 23.0% more entries when in the enhanced

Table 3 Accuracies and latencies of range transitions and range identifications for SpO_2 and other vital signs (heart rate, blood pressure and $ETCO_2$). Accuracies are percentage correct; latencies are expressed in seconds. Values are median [95% confidence interval]. Target transitions refer to transitions between target and low in both directions. Critical transitions refer to transitions between low and critical in both directions. Range transitions refer to transitions between any two ranges. †Sign test. ‡Wilcoxon signed rank test.

Outcome	Standard display	Enhanced display	W values	Z values	P values
Primary outcome					
Accuracy at detecting SpO_2 target transitions†	73.2 [50.0, 75.0]	100.0 [100.0, 100.0]	–	3.88	<0.001
Secondary outcomes					
Accuracy at detecting SpO_2 critical transitions†	100.0 [66.7, 100.0]	100.0 [100.0, 100.0]	–	2.04	0.041
Accuracy at detecting all SpO_2 range transitions†	75.7 [60.0, 85.7]	100.0 [100.0, 100.0]	–	4.13	<0.001
Accuracy at detecting out of target range transitions for other vital signs†	100.0 [100.0, 100.0]	100.0 [100.0, 100.0]	–	0.00	1.0
Latency to detect SpO_2 target transitions†	27.4 [11.6, 41.2]	3.3 [1.9, 5.7]	–	−2.91	0.004
Latency to detect SpO_2 critical transitions†	28.5 [12.0, 44.2]	1.5 [1.3, 2.2]	–	−3.80	<0.001
Latency to detect all SpO_2 range transitions†	30.1 [19.2, 44.6]	2.7 [1.7, 4.0]	–	−4.25	<0.001
Latency to detect out of target range transitions for other vital signs‡	7.5 [6.0, 12.1]	6.3 [4.0, 12.0]	89.0	−.597	0.550
Accuracy of SpO_2 range identifications‡	57.1 [40.0, 70.0]	100.0 [90.0, 100.0]	210.0	3.92	<0.001
Accuracy of range identifications for other vital signs†	100.0 [100.0, 100.0]	100.0 [100.0, 100.0]	–	0.00	1.0

condition than when in the standard plus alarm condition. However, a Wilcoxon signed rank test showed there was no statistically significant difference across display conditions in the accuracy of participants' entries (P=0.573).

Subjective judgements

Participants found it easier and were more confident of their judgments of SpO_2 parameters and their interpretation of the SpO_2 auditory display when using the enhanced display than when using the standard display plus alarm (Table 4). They judged that the enhanced display would lead to less harmful patient outcomes, was more dependable, reliable, trustworthy, and helpful for patient monitoring than the standard display plus alarm. Finally, they reported that it was easier to perform the patient categorisation task with the enhanced display than with the standard display plus alarm.

In open-ended responses, participants reported that it was easier to identify saturation from the added sound dimensions in the enhanced display rather than with the pitch-only standard display plus alarm, and that they found it easier to perform the categorisation task when in the enhanced condition than when in the standard condition. For example, one participant wrote that 'I felt I could tell purely on the auditory sounds what range the sats were, leaving me more head space to concentrate on the computer task'.

The LIWC analysis showed that participants used more positive language about the enhanced display (78.0) than about the standard display plus alarm (1.0). A score of 100 reflects solely positive language to describe emotions felt

Table 4 Subjective ratings of SpO_2 identification tasks, auditory displays and patient categorisation task using 9-point Likert items. Values are median [95% confidence interval]. Q1, Q2, Q11: 1 – extremely hard, 9 – extremely easy; Q3, Q4, Q5: 1 – not at all confident, 9 – very confident; Q6–Q10: 1 – not at all, 9 – definitely (result in harm, dependable, trustworthy, helpful). †Wilcoxon signed rank test. ‡Sign test.

Question	Standard display	Enhanced display	W value	Z value	P value
Q1. Ease of identification of SpO_2 range at any point in the scenario†	4.0 [3.0, 5.0]	8.0 [7.0, 9.0]	210.0	3.95	<0.001
Q2. Ease of identification of when SpO_2 transitioned between ranges†	3.0 [2.0, 4.0]	8.0 [8.0, 9.0]	210.0	3.94	<0.001
Q3. Confidence of identification of SpO_2 range†	4.0 [3.0, 5.0]	8.0 [7.0, 8.0]	–	4.25	<0.001
Q4. Confidence of identification SpO_2 range transitions‡	3.5 [3.0, 4.0]	8.0 [7.0, 9.0]	210.0	3.94	<0.001
Q5. Confidence of interpretation of SpO_2 auditory display†	4.0 [3.0, 6.0]	7.5 [7.0, 8.0]	197.5	3.48	<0.001
Q6. Perception that SpO_2 auditory display output may result in harmful outcomes‡	5.0 [2.0, 7.0]	2.0 [1.0, 3.0]	–	−3.62	<0.001
Q7. Dependability of SpO_2 auditory display†	6.0 [5.0, 7.0]	8.0 [7.0, 8.0]	149.5	3.47	0.001
Q8. Reliability of SpO_2 auditory display‡	6.0 [4.0, 7.0]	7.5 [7.0, 8.0]	–	4.13	<0.001
Q9. Trustworthiness of SpO_2 auditory display to provide accurate SpO_2 information†	5.0 [4.0, 6.0]	7.0 [7.0, 8.0]	168.5	3.65	<0.001
Q10. Helpfulness of SpO_2 auditory display for monitoring patient‡	7.0 [6.0, 7.0]	8.0 [8.0, 8.0]	–	3.62	<0.001
Q11. Ease of performing patient categorisation task†	3.0 [2.0, 4.0]	5.0 [4.0, 6.0]	142.5	3.14	0.002

about the display; a score of 0 reflects solely negative language.

Discussion

This simulator study demonstrates many measurable benefits of an enhanced pulse oximeter auditory display whose sound properties change when saturation goes into low and critical ranges. Anaesthetists more accurately detected SpO_2 range transitions when they heard the varying pitch display enhanced with tremolo and brightness than when they used a standard pulse oximeter display where only pitch varied, even though the standard display had an alarm. They also detected range transitions faster and maintained a better sense of saturation range when using the enhanced display. Furthermore, participants were more confident of their ability to detect changes in saturation ranges and to identify those ranges with the enhanced display, and they were highly positive in their ratings and comments about it.

Previous laboratory studies that tested the enhanced and standard displays with non-clinician and clinician participants found similar results.[10,12,13] However, those studies were conducted in non-representative environments with only the participant and experimenter present, and in an artificial situation where only saturation was monitored for short durations. This investigation extends our series of studies, and demonstrates the benefits of the enhanced auditory display in the more realistic setting of a fully staffed simulated operating theatre during situations with multiple ongoing tasks and interruptions for an extended period of time. The simulator study even had a larger effect size (r=0.87 for SpO_2 range transitions and 0.95 for range identifications) than did the two laboratory studies (r=0.66 and 0.63 for range transitions, and 0.41 and 0.55 for range identifications).[23]

One explanation for participants' superior performance with the enhanced display is that a discrete acoustic change in the pulse tone captures participants' attention and recalibrates their sense of current SpO_2 value when it transitions into new range. Redundant acoustic features can improve identification of auditory stimuli,[24,25] and previous research shows that discrete changes of those features at clinically important SpO_2 boundaries support better performance than do incremental changes at each SpO_2 value.[10,12,13,26]

More accurate identification (100% *vs* 73.2%) and faster detection (3.3 *vs* 27.4 s) of the transition between an SpO_2 of 97% and 96%, based only on the sound of the PO, may be of clinical relevance. Although a patient may not be in immediate danger when SpO_2 has transitioned from the target to the low range, the enhanced display provides a signal that SpO_2 is trending towards a critical range and that intervention may be required before the critical threshold is reached. Importantly, if remedial action is successful, changes in the display will signal when target levels are regained.

Faster detection (1.5 *vs* 28.5 s) of the transition between an SpO_2 of 90% and 89% may also be of clinical relevance, and is especially impressive because an alarm sounded at that transition with the standard display but not the enhanced display. The enhanced display may help anaesthetists maintain appropriate SpO_2 status, and reduce noise levels by avoiding auditory alarms for critical SpO_2 breaches.

Twenty-five years ago, cognitive psychologist David Woods[27] wrote about the 'alarm problem' in dynamic fault management. Auditory alarms call attention to the situation that a problem exists, but they provide little or no further information about the source of the problem or the state of the process. The pulse oximeter tone is a continuously informing auditory display that identifies the source of the signal and state of oxy-haemoglobin saturation. The standard tone conveys well that saturation is changing, but not that it has crossed a threshold into a different clinical range. Current pulse oximeters call attention to these range transitions with alarm sounds, but this study demonstrates a different philosophy of enhancing a continuously informing display to call attention to range transitions. The results show that anaesthetists respond faster to the enhanced display than to a standard display with an alarm.

There is another potential advantage to the enhanced display, in that SpO_2 range may be more easily recognised if the auditory display has been obscured or is unavailable for a time. For example the pulse oximetry signal may be masked by the intermittent loud noise of an orthopaedic drill. Between episodes of drilling, SpO_2 range may be identified by sound alone more accurately with the enhanced display than with the standard display.

The enhanced display also appears to have provided more capacity for the patient categorisation tasks. Although there was no difference in accuracy of data entry, participants made 23% more entries in the enhanced condition than in the standard condition. This suggests that when using the enhanced display, participants were willing to monitor patient vital signs in peripheral awareness while performing other tasks.[7,27] Supporting this interpretation are participants' judgments that they were more confident of their judgements of SpO_2 with the enhanced display and believed it was more dependable, reliable and helpful for patient monitoring than the standard display. Qualitative data revealed that they also felt more emotionally positive about the enhanced display. It is important to survey end users' attitudes about a new auditory display because if users judge a display unfavourably, they may be reluctant to use it despite any possible benefits it may provide.[28]

One limitation of this study is that a full sensitivity and specificity analysis of range detections could not be done. The primary outcome of SpO_2 transition detection accuracy can be taken as a measure of sensitivity, where sensitivity=true positives/(true positives+false negatives). Specificity cannot be calculated because the extended periods between range transitions in these scenarios offered no basis for measuring true negatives. However, a false discovery rate could be calculated, where false discovery rate=false positives/(false positives+true positives). A *post hoc* exploratory analysis of the false discovery rate for SpO_2 target transitions revealed that participants' median probability of making a false discovery was significantly lower with the enhanced display (Mdn = 0.00, 95% CI [0.00, 0.13]) than with the standard display (Mdn = 0.29, 950.17,0.47]), Z = 3.06, P = 0.001. This finding suggests that the enhanced display supports a more effective allocation of attention, and it is consistent with the other advantages demonstrated for the enhanced display.

Further limitations were that participants did not provide direct patient care and knew that they were being studied in an artificial situation where they verbalised vital sign range changes that they observed. We considered only paediatric cases, and scenarios were relatively short with an unusually high rate of saturation range changes. Demands on anaesthetists' cognitive processes may be affected by case type, length of case, and elements such as uncertainty, dynamism, and risk associated with the operating theatre environment.

Furthermore, for the purposes of the experiment we did not provide a visual readout of the numerical SpO_2 value, but provided only the auditory display, whereas anaesthetists normally monitor patients' vital signs using both auditory and visual displays. However, the pattern of results is strong enough that we suspect that the enhanced display would be beneficial to anaesthetists in real operating theatres during actual, longer, and less dynamic cases.

This simulator study demonstrates that an auditory display enhanced with tremolo and brightness to distinguish SpO_2 ranges supports more accurate and timely identification of SpO_2 levels than does a standard display based on varying pitch plus alarms. Such a display may allow anaesthetists to monitor patients more effectively so that remedial actions can be made proactively, potentially reducing the incidence of hypoxaemia. The principles underlying the enhanced display have been tested in a series of studies in which the benefits have proven to be robust. Incorporating the enhanced display in future and existing pulse oximeters would require only software changes. In this study, the ranges were fixed, but like alarms, the thresholds in clinical devices could be adjustable to accommodate different situations and patients.

Authors' contributions

Conception and design of study: PMS, EP, NABP, RGL
Operationalised the study: EP, NABP, PMS, IM, RGL
Participant recruitment: EP, NABP
Data analysis: EP, PMS, ISS, IM, FPB
First draft of manuscript: EP
Clinical review and manuscript revision: EP, NABP, PMS, ISS, IM, FPB, RGL

Declarations of interest

PMS: co-inventor of a respiratory sonification (Sanderson and Watson, US Patent 7070570).

RGL has received $1000 per year to be on the Masimo, Inc. Scientific Advisory Board.

The other authors have no conflicts of interest.

Funding

Australian and New Zealand College of Anaesthetists [Grant number S18/003] (CI-A NABP).

Acknowledgements

The authors thank Fiona MacFarlane, Jason Shoutrop, Glen Bannister, and administration staff, Anaesthesia and Pain Management Services, Queensland Children's Hospital (QCH), South Brisbane, Queensland for facilitating participant recruitment. We are grateful to Ben Lawton, Tricia Pilotti, and staff from the Simulation Training on Resuscitation for Kids (SToRK) at QCH for assistance with simulator operation. Our thanks also to Wendy Fennah, Denise Wilson and Peta Tippett from the Learning and Workforce at QCH for arranging simulation facilities. James Bishop and Peter Mills, Clinical Skills Development Service, Royal Brisbane and Women's Hospital, Herston, also assisted with equipment and simulator set up. Rose Voyzey and Caitlin Hunter and staff from the QCH Operating Suite were invaluable in providing surgical and anaesthetic equipment for each session. We also appreciate assistance with statistical analysis from Anne Bernard from QFab BioInfomatics, The University of Queensland, St Lucia. Finally, we thank the following students from The University of Queensland: Jelena Zestic, Garry Mann and Tom Davidson who acted operating team roles, and Polly Fong who assisted with analysis of qualitative data.

Supplementary material

Supplementary material to this article can be found online at https://doi.org/10.1016/j.bja.2020.05.038

References

1. Pretto JJ, Roebuck T, Beckert L, Hamilton G. Clinical use of pulse oximetry: official guidelines from the Thoracic Society of Australia and New Zealand. *Respirology* 2014; **19**: 38–46
2. Strachan L, Noble D. Hypoxia and surgical patients—prevention and treatment of an unnecessary cause of morbidity and mortality. *J R Coll Surg Edinb* 2001; **46**: 297–302
3. Ehrenfeld J, Funk L, Schalkwyk J, Merry A, Sandberg W, Gawande A. The incidence of hypoxemia during surgery: evidence from two institutions. *Can J Anaesth* 2010; **57**: 888–97
4. Baillard C, Boubaya M, Statescu E, et al. Incidence and risk factors of hypoxaemia after preoxygenation at induction of anaesthesia. *Br J Anaesth* 2019; **122**: 388–94
5. Ford S, Birmingham E, King A, Lim J, Ansermino JM. At-a-glance monitoring: covert observations of anesthesiologists in the operating room. *Anesth Analg* 2010; **111**: 653–8
6. Schulz CM, Schneider E, Fritz L, et al. Visual attention of anaesthetists during simulated critical incidents. *Br J Anaesth* 2011; **106**: 807–13
7. Watson MO, Sanderson P. Sonification supports eyes-free respiratory monitoring and task time-sharing. *Hum Factor*. 2004; **46**: 497–517
8. Loeb RG, Brecknell B, Sanderson P. The sounds of desaturation: a survey of commercial pulse oximeter sonifications. *Anesth Analg* 2016; **122**: 1395–403
9. Morris RW, Mohacsi PJ. How well can anaesthetists discriminate pulse oximeter tones? *Anaesth Intensive Care* 2005; **33**: 497–500
10. Paterson E, Sanderson PM, Brecknell B, Paterson NAB, Loeb RG. Comparison of standard and enhanced pulse oximeter auditory displays of oxygen saturation: a laboratory study with clinician and nonclinician participants. *Anesth Analg* 2019; **129**: 997–1004
11. Stevenson RA, Schlesinger JJ, Wallace MT. Effects of divided attention and operating room noise on perception of pulse oximeter pitch changes: a laboratory study. *Anesthesiology* 2013; **118**: 376–81
12. Paterson E, Sanderson P, Paterson NAB, Liu D, Loeb RG. The effectiveness of pulse oximetry sonification enhanced with tremolo and brightness for distinguishing clinically important oxygen saturation ranges: a laboratory study. *Anaesthesia* 2016; **71**: 565–72
13. Paterson E, Sanderson PM, Paterson NAB, Loeb R. The effectiveness of enhanced pulse oximetry sonifications for conveying oxygen saturation ranges: a laboratory comparison of five auditory displays. *Br J Anaesth* 2017; **119**: 1224–30

14. Deschamps M-L, Sanderson P, Hinckfuss K, et al. Improving the detectability of oxygen saturation level targets for preterm neonates: a laboratory test of tremolo and beacon sonifications. *Appl Ergon* 2016; **56**: 160–9
15. Hinckfuss K, Sanderson P, Loeb RG, Liley H, Liu D. Novel pulse oximetry sonifications for neonatal oxygen saturation monitoring: a laboratory study. *Hum Factor*. 2016; **58**: 344–59
16. Zestic J, Brecknell B, Liley H, Sanderson P. A novel auditory display for neonatal resuscitation: laboratory studies simulating pulse oximetry in the first 10 minutes after birth. *Hum Factor*. 2019; **61**: 119–38
17. Sanderson PM, Watson MO, Russell WJ, et al. Advanced auditory displays and head-mounted displays: advantages and disadvantages for monitoring by the distracted anesthesiologist. *Anesth Analg* 2008; **106**: 1787–97
18. *Medical electrical equipment, Part 1–8: general requirements, tests and guidance for alarm systems in medical electrical equipment and medical electrical systems*. Geneva, Switzerland: International Electrotechnical Commission; 2006. IEC 60601–1–8:2005
19. Hasfeldt D, Laerkner E, Birkelund R. Noise in the operating room—what do we know? A review of the literature. *J Perianesth Nurs* 2010; **25**: 380–6
20. Faul F, Erdfelder E, Lang A-G, Buchner AG. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav Res Method* 2007; **39**: 175–91
21. Cohen J. A coefficient of agreement for nominal scales. *Education Psychol Measure* 1960; **20**: 37–46
22. Pennebaker JW, Boyd RL, Jordan K, Blackburn K. *The development and psychometric properties of LIWC 2015*. Austin, TX: University of Texas at Austin; 2015
23. Cohen J. A power primer. *Am Psychol Assoc* 1992; **112**: 155–9
24. Edworthy J, Meredith C, Hellier E, Rose D. Learning medical alarms whilst performing other tasks. *Ergonomics* 2013; **56**: 1400–18
25. Rayo MF, Patterson ES, Abdel-Rasoul M, Moffatt-Bruce SD. Using timbre to improve performance of larger auditory alarm sets. *Ergonomics* 2019: 1–33
26. Collett R, Salisbury I, Loeb RG, Sanderson PM. Smooth or stepped? Laboratory comparison of enhanced sonifications for monitoring patient oxygen saturation. *Hum Factor*. 2019; **62**: 124–7
27. Woods DD. The alarm problem and directed attention in dynamic fault management. *Ergonomics* 1995; **38**: 2371–93
28. Edworthy J. Medical audible alarms: a review. *J Am Med Inform Assoc* 2013; **20**: 584–9

Handling editor: Christa Boer