Applied Ergonomics 56 (2016) 160–169

Contents lists available at ScienceDirect

Applied Ergonomics

journal homepage: www.elsevier.com/locate/apergo

Improving the detectability of oxygen saturation level targets for preterm neonates: A laboratory test of tremolo and beacon sonifications

Marie-Lys Deschamps [a], Penelope Sanderson [a, b, c, *], Kelly Hinckfuss [a], Caitlin Browning [a], Robert G. Loeb [d], Helen Liley [c, e], David Liu [b, c]

[a] School of Psychology, The University of Queensland, Australia
[b] School of ITEE, The University of Queensland, Australia
[c] School of Medicine, The University of Queensland, Australia
[d] Department of Anesthesiology, University of Arizona, USA
[e] Mater Mothers Hospital, Australia

A R T I C L E I N F O

Article history:
Received 17 November 2015
Received in revised form
22 March 2016
Accepted 23 March 2016
Available online 14 April 2016

Keywords:
Patient monitoring
Pulse oximetry
Sonification
Auditory display

A B S T R A C T

Recent guidelines recommend oxygen saturation (SpO_2) levels of 90%–95% for preterm neonates on supplemental oxygen but it is difficult to discern such levels with current pulse oximetry sonifications. We tested (1) whether adding levels of tremolo to a conventional log-linear pulse oximetry sonification would improve identification of SpO_2 ranges, and (2) whether adding a beacon reference tone to conventional pulse oximetry confuses listeners about the direction of change. Participants using the Tremolo (94%) or Beacon (81%) sonifications identified SpO_2 range significantly more accurately than participants using the LogLinear sonification (52%). The Beacon sonification did not confuse participants about direction of change. The Tremolo sonification may have advantages over the Beacon sonification for monitoring SpO_2 of preterm neonates, but both must be further tested with clinicians in clinically representative scenarios, and with different levels of ambient noise and distractions.

1. Introduction

When the oxygen saturation of the arterial blood of a preterm neonate receiving oxygen support is too high or too low, it can lead to serious injury or even death. To avoid adverse outcomes, recent clinical guidelines recommend that neonatal arterial oxygen saturation levels determined using pulse oximetry (SpO_2) should stay within a narrow 90%–95% target range (Saugstad and Aune, 2014). However, as will be discussed, the current audible pulse oximetry tones are limited in how effectively they alert clinicians to SpO_2 changes away from that range.

In a recent study, Hinckfuss et al. (2016) added a reference tone, or "beacon", at regular intervals to the current pulse oximetry tones whenever SpO_2 levels were outside the neonatal target range. Non-clinician participants listening to simplified scenarios could identify general SpO_2 ranges and discriminate the target and non-target SpO_2 states more accurately with the beacon than without. However, potential disadvantages of the beacon enhancement were noted.

The study reported herein provides a further test of the beacon enhancement to the pulse oximetry tones. In addition, the present study tests another way of signalling whenever SpO_2 levels are outside the target range—by changing the quality of the pulse tones with 'tremolo', a vibrating characteristic added to the sound.

1.1. Pulse oximetry and the needs of preterm neonates

Pulse oximetry is a clinical monitoring technique that non-invasively estimates a patient's arterial oxygen saturation level by analysing light transmission though a vascular bed, such as in a finger or, for an infant, in the wrist, palm, or foot. Along with a visual display of HR and SpO_2, most pulse oximeters use 'sonification'—a continuous auditory display. A tone is played with every pulse, allowing estimation of pulse rate in beats per minute. The pitch of the tones represents SpO_2. A series of tones increasing

* Corresponding author. School of Psychology, The University of Queensland, Australia. Tel.: +61 407 288 695, +61 7 3346 9529.
E-mail address: uqpsande@uq.edu.au (P. Sanderson).

http://dx.doi.org/10.1016/j.apergo.2016.03.013

or decreasing in pitch indicates rising or falling SpO_2 levels, respectively.

The current pulse oximetry sonification works well when the highest SpO_2 levels are also the target levels, but not if the target levels are in a lower range. In healthy adults, oxygen saturation levels typically remain between 100% and 97%. Therefore, for adults it is usually a good sign if the pitch of the pulse oximeter tones is increasing, and bad if the pitch is decreasing.

However, in preterm neonates, too much supplemental oxygen can lead to oxygen toxicity, development of bronchopulmonary dysplasia (Saugstad, 1997; Saugstad and Aune, 2014), and retinopathy of prematurity which can lead to blindness (Chow et al., 2003; Saugstad and Aune, 2014; Tin et al., 2001). Conversely, hypoxia (oxygen insufficiency) can damage organs, including the brain, contributing to cerebral palsy (Askie et al., 2011; Collins et al., 2001) and infant mortality (Dawson et al., 2010; Finer and Leone, 2009; Ford et al., 2006; Lim et al., 2014; Saugstad and Aune, 2014).

Therefore for preterm neonates monitored with the current pulse oximeter sonification, an increasing tone pitch could indicate either that the oxygen saturation level is becoming too high and the infant is at risk of oxygen toxicity, or that oxygen saturation is safely moving upwards towards the target zone. Likewise, a decreasing tone pitch could indicate either that oxygen saturation levels are moving safely downwards into the target range, or that hypoxia is imminent.

A further problem relates to knowing when the preterm neonate's SpO_2 has entered or exited the target range. Recent clinical guidelines recommend that neonatal SpO_2 should remain between 90% and 95% (Saugstad and Aune, 2014), but debate about the exact range continues (Manja et al., 2015). The current pulse oximetry technology does not provide SpO_2 range information in a manner that is suited to the clinical monitoring environment (Goos et al., 2013; Janata and Edwards, 2013; Lim et al., 2014; Tin and Lal, 2015). A recent study in two NICUs revealed that neonatal SpO_2 levels remained within a target range only 31% of the time, and were frequently outside it for prolonged periods (Lim et al., 2014). The longer and further a preterm infant's SpO_2 is outside target range, the higher the neonate's risk of adverse outcomes (Askie et al., 2011; Carlo et al., 2010; Lim et al., 2014).

1.2. Further challenges

Current pulse oximetry sonification relies on the clinician's ability to discern relative and absolute pitch (Brown et al., 2015). Multisensory training has been suggested as a way to improve clinicians' perception of the pitch of pulse oximetry tones (Schlesinger et al., 2014). However, this approach is unlikely to work across different pulse oximeter models, which use different mappings of tones to SpO_2 (Loeb et al., 2016; Santamore and Cleaver, 2004). Creating an auditory display design that makes deviations from target SpO_2 range obvious without the need for training could be a more robust approach to addressing the problem.

However, a redesign of the pulse oximetry sound for neonates must take into account the problem of adding extra sounds to an already demanding auditory environment. A major area of concern in all neonatal monitoring contexts is excess noise, which can lead to attentional narrowing, resulting in impaired monitoring performance (Hockey, 1997) as well as poorer health outcomes for patients (Long et al., 1980). Pulse oximetry devices use conventional auditory alarms alongside the variable-pitch sonification, so that their users also suffer from the well-documented disadvantages of auditory alarms (Edworthy, 2013; Edworthy et al., 2014; Lim et al., 2014). Neonates have higher heart rates than do adults, so variable pitch tones are emitted more frequently. In many NICUs, pulse oximeter variable-pitch tones are silenced to decrease the noise level, and intermittently-emitted earcons have been proposed as an alternative (Janata and Edwards, 2013). Within NICUs, clinicians rely on conventional threshold alarms to alert them to oxygenation deviations, given the impracticality of monitoring multiple pulse oximetry sonifications. During neonatal transport and resuscitation, clinicians are more likely to use the pulse tones as well as conventional threshold alarms.

Sonification offers several advantages. It may help to reduce reliance on alarms and may convey more information to the clinician. Sonification also supports eyes-free monitoring, where the patient's vital signs can be monitored in the clinician's peripheral awareness while other important tasks are performed (Watson and Sanderson, 2004; Woods, 1995). When a sonification is designed appropriately, changes in the sound should draw the clinician's attention to deteriorations in a patient's vital signs (Loeb and Fitch, 2002; Sanderson, 2006; Sanderson et al., 2009; Sanderson et al., 2005). A well-designed sonification should let clinicians maintain continuous awareness of the neonate's oxygen saturation levels during visually demanding tasks.

1.3. Recent enhancements of pulse oximetry sonification for neonates

Recently, Hinckfuss et al. (2016) examined whether a beacon, or reference tone, added to a simulated "log-linear" pulse oximetry sonification could improve detection of oxygen saturation deviations compared with the log-linear sonification alone. A log-linear sonification maps fixed increments of SpO_2 to a fixed percentage increase in frequency [Hz], thereby mapping a linear scale to a logarithmic one. Several conventional pulse oximetry sonifications use a log-linear mapping (Loeb et al., 2016). Previous research indicates that participants can identify SpO_2 values more accurately with a log-linear mapping than with a linear mapping (Brown et al., 2015), making the log-linear mapping a better control condition.

In the Hinckfuss et al. (2016) study, the beacon functioned as both an alert and a reference tone. It was played immediately before every fourth tone when oxygen saturation fell outside the 90%–95% target zone, alerting the listener to a non-target state. The pitch of the beacon was the same as the 93% oxygen saturation level, which is the middle of the target range, offering the listener a reference tone. When oxygen saturation levels were above the target range, the beacon was lower in pitch than the higher pitched heart rate tones, indicating the SpO_2 should decrease. When oxygen saturation levels were below the target range, the beacon was higher in pitch than the lower pitched heart rate tones, indicating that SpO_2 should increase.

When using the log-linear sonification with the beacon, participants could identify in which of five ranges SpO_2 fell with a 25-percentage point advantage compared with the log-linear sonification alone. However, the beacon sonification has potential drawbacks. First, it adds extra sounds to an already sound rich environment. Second, because the beacon is played only on every fourth heart rate tone, there may be a short delay between oxygen saturation going out of range, and the auditory signal of that change. Third, the listener must infer how far SpO_2 has deviated from the target range from the difference in pitch between the beacon and the tone.

In addition, there were shortcomings in the Hinckfuss et al. (2016) experiment and its findings. First, the effect of the beacon was tested within subjects: a clear between-subjects test would provide additional support for its effectiveness. Second, participants were not given a visual anchor of the SpO_2 range at the start of each trial, potentially making range identification unrealistically difficult. Third, a minor concern was that participants detected the direction of change with the beacon sonification slightly less accurately than with the log-linear sonification, although the 2-percentage point difference for the latter was well below the clinical effect size of 10 percentage points or more that had been chosen by the authors. A resolution of the above shortcomings would put the beacon design on a stronger basis for a future clinical test.

There are many further ways of indicating acoustically whether SpO_2 is in or out of the target range, and by how much. For example, tremolo can be produced by amplitude modulations that create a vibrating or corrugated quality to a tone. As tremolo frequency increases, in cycles per tone or per second (Hz), the vibrating becomes more rapid. In a subsequent pilot study with a between-subjects design, Hinckfuss et al. (2015; Experiment 3) added three cycles of tremolo to tones in the low and high SpO_2 ranges, and five cycles of tremolo to tones in the very low and very high SpO_2 ranges (see Table 1 for ranges). A third dependent variable was also introduced: participants' accuracy at detecting when SpO_2 first exited or entered its target range, as this moment is clinically important. Participants could identify range 12 percentage points better with the tremolo sonification than with the log-linear sonification, but even better with the beacon sonification. Moreover, participants could detect transition into or out of the target SpO_2 range significantly better with the beacon than with the log-linear sonification, but the tremolo sonification did not yield a significant improvement. The beacon sonification confused a minority of participants regarding the direction of SpO_2 change, with the result that the average accuracy for identifying direction of change with the beacon sonification was as much as 17 percentage points worse than for the log-linear sonification, but accuracy was good for participants using the tremolo sonification.

The current study extends this prior research in a between-subjects study. First, a stronger implementation of tremolo is used, to make it more perceivable. Accordingly, we seek an improvement of at least 15 percentage points for the tremolo sonification over the log-linear sonification, and will define 15 percentage points as the minimum improvement expected in all comparisons. Second, the effect of the beacon on direction identification is tested again to determine whether or not the confusions experienced by some participants in Hinckfuss et al. (2016; Experiment 2) and Hinckfuss et al. (2015; Experiment 3) are repeatable, and a cause for concern.

1.4. The present study

The three hypotheses to be tested for both the tremolo and beacon sonifications—referred to collectively as enhanced sonifications—are as follows, making a total of six hypotheses:

- Participants using an enhanced sonification will show a 15-percentage point or more advantage in how accurately they identify the SpO_2 range (very low, low, target, high, very high) over participants using a log-linear sonification alone.
- Participants using an enhanced sonification will identify the direction of change (decreasing, steady, increasing) significantly less accurately than participants using a log-linear sonification alone, if the enhancement is confusing.
- Participants using an enhanced sonification will show a 15-percentage point advantage or more in how accurately they detect transitions into or away from the target range, compared with participants using a log-linear sonification alone.

2. Method

2.1. Power analysis

The three experimental conditions were termed the LogLinear, Beacon, and Tremolo conditions. The primary outcome measures were range identification, direction identification, and target transition identification for LogLinear vs. Beacon and for LogLinear vs. Tremolo. A power analysis was conducted using G*Power seeking an effect size of at least 15-percentage points improvement in accuracies on all three measures. The experiment was powered to test six two-tailed t-tests of independent groups, with an alpha level of .00833 each (.05/6), a statistical power of 0.8, and an equal participant ratio across conditions. Using means and SD from Experiment 3 in Hinckfuss et al. (2015) it was determined that a total sample size of 93 participants was required, 31 per sonification condition. Secondary outcome measures included tests of how quickly participants detected transitions into and out of the target range, and participants' subjective judgments of the difficulty for each task and their level of confidence in their judgments.

2.2. Participants

The Ethics Committee of the School of Psychology at The University of Queensland granted approval for the experimental protocol. Participation was voluntary: students gave informed consent and received either course credit, or an AUD$10 gift voucher. Participants were excluded if they were not undergraduate students at The University of Queensland, or if they did not report normal hearing.

2.3. Design

The independent variable was sonification condition (LogLinear vs. Beacon vs. Tremolo), which was manipulated on a between-subjects basis. To control for potential effects of music training, participants were classified as either musically trained (more than 12 months training) or not musically trained (less than or equal to 12 months training) using a questionnaire, and were then assigned at random to one of the three sonification conditions.

Participants completed four blocks of 12 trials each, where each trial lasted for 30 s. For experimental blocks 1 and 2, participants made range identification and direction identification decisions. For experimental blocks 3 and 4, participants made range identification and target transition identification decisions. Decision types were not counterbalanced across blocks because they simply

Table 1
Description of the five SpO_2 ranges showing additions to the LogLinear sonifications with the Beacon and Tremolo sonifications.

Range	SpO_2 (%) from	SpO_2 (%) to	Beacon (Reference tone every 4th tone)	Tremolo (Tremolo cycles per tone)
Very High	99	100	Present	6
High	96	98	Present	3
Target	90	95	Absent	0
Low	84	89	Present	3
Very Low	80	83	Present	6

represented different types of measures with which to evaluate the effectiveness of the sonifications. However, the order of test trials within blocks was randomised differently for each participant.

2.4. Apparatus

The training and experimental tasks were run on an 11-inch MacBook Air laptop computer with a CDSONIC (AE-120E) external speaker set attached. Both the training and experimental scenarios were run from a standalone application programmed in LiveCode™ Community version 6.6 (RunRev, Edinburgh).

2.5. Tasks

For all three tasks outlined below, participants indicated their response by clicking on the relevant on-screen button using a computer mouse.

2.5.1. Range identification

Participants indicated whether the level of oxygen at the completion of the scenario was in the *Very High*, *High*, *Target*, *Low*, or *Very Low* range (see Fig. 1). For each trial, the starting oxygen range level was displayed in the centre of the screen, but it disappeared as soon as the sounds began.

2.5.2. Direction identification

Participants indicated whether the oxygen saturation level was *Increasing*, *Decreasing*, or *Steady* during the last half of each 30-s trial. To indicate the halfway point (15 s) of each trial, a yellow square appeared next to the direction answer box, and was visible on the screen during the remainder of the trial.

2.5.3. Target transition identification

Participants clicked the on-screen button to note the moment at which the oxygen level either entered or left the target range. If a trial included a target transition, there was only one target transition in the trial. For other trials, the oxygen level did not enter or leave the target range, so participants were expected to make no response.

2.6. Stimuli

2.6.1. Test trials

Throughout each 30-s trial, the SpO_2 values varied smoothly, changing in no more than 1% increments, or the values stayed constant in parts. In all conditions SpO_2 was indicated by tone pitch, plus beacon or tremolo additions if relevant. All scenarios simulated a constant HR of 120 bpm, representing the normal pulse rate of a neonate.

2.6.2. Practice trials

Before the range and direction identification trials started in blocks 1 and 2, practice trial 1 presented the full sequence of 21 oxygen tones from *Very High to Very Low* and practice trial 2 presented all tones from *Very Low* to *Very High*. Before the range and target transition identification trials started in blocks 3 and 4, practice trial 1 began in the *Very High* range and presented one transition from *High* into Target, and practice trial 2 presented one transition beginning in the Target range and moving into *Low*, finally finishing in *Very Low*. In both cases, the next five practice trials were similar to the experimental trials but the current oxygen range was displayed on screen throughout the trial, allowing the participants to become familiar with the mapping of tones to each oxygen range.

2.6.3. Tones

All tones were created with a custom-built MatLab application, and the tremolo was added in Audacity™ (The Audacity Team, FossHub).

1. **LogLinear.** To represent SpO_2 levels ranging from 80% to 100%, pure sine wave functions were used to generate twenty-one frequencies, spanning 525 Hz–881 Hz. The proportion change in frequency for each of the 21 SpO_2 levels from the level below

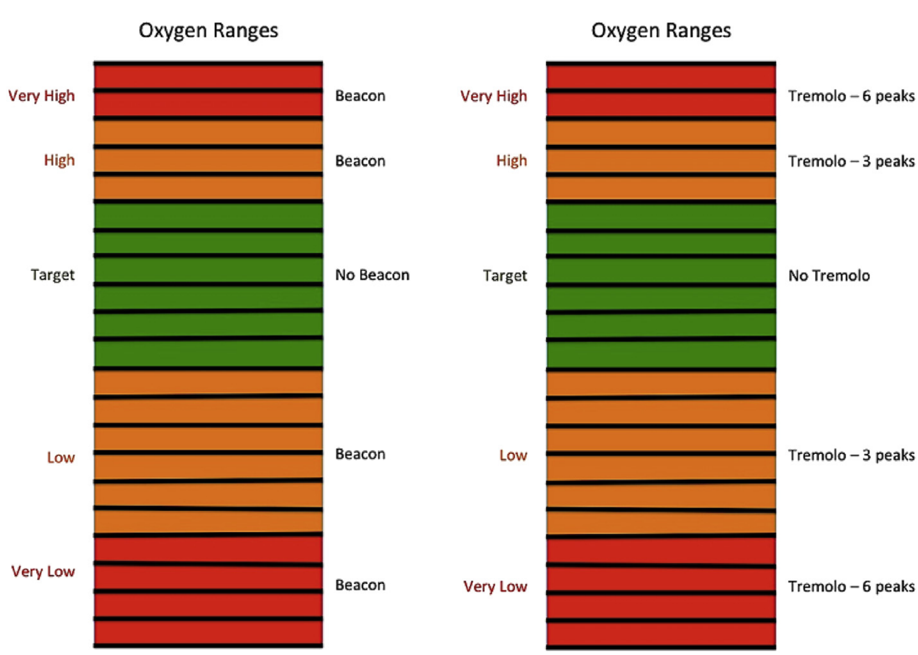

Fig. 1. Visual aid of the five oxygen ranges shown to participants in the Beacon and Tremolo conditions. The LogLinear visual aid was similar, but with no information on the right side of the graph, as it was not applicable.

it was 2.75%, and therefore equal in ratio across the whole range. Each tone was 200 ms in length, including a 15 ms fade in and 15 ms fade out.

2. **Beacon.** The same pulse tones were used as in LogLinear, but with the addition of a beacon, or reference tone, sounding 120 ms prior to every fourth tone for SpO_2 values outside the target range of 90%–95% (see Table 1 and Fig. 2). The beacon was a sine wave function at a frequency of 735 Hz, associated with a 93% oxygen saturation level. The beacon was 100 ms in length, including a 10 ms fade in and 10 ms fade out.
3. **Tremolo.** The same pulse tone frequencies were used as in LogLinear, but for SpO_2 levels falling outside the target range of 90%–95% a tremolo effect was added to the tone. For *Low* and *High* levels, 3 cycles of tremolo amplitude modulation, with 90% wet, were added to each 200 ms tone. (The *wet* level describes how much sound power is lost between the maximum and minimum amplitude of the tremolo.) For *Very Low* and *Very High* levels, 6 cycles of tremolo with 90% wet were added to each 200 ms tone (see Table 1). Sound pressure levels were adjusted so that total tone power was the same across ranges, regardless of tremolo.

2.7. Visual/auditory training aids

For all conditions, a graphic representation of oxygen level ranges (see Fig. 1), and a screen shot of each task were provided during training. For the Beacon condition, a visual aid was used to help the participant understand the beacon feature (see Fig. 2). For the Tremolo condition, a visual and auditory aid (see Fig. 3) was used to help the participant understand the tremolo feature.

2.8. Questionnaires

2.8.1. Formal musical training questionnaire

Participants' formal musical training was measured with a 6-item questionnaire before the start of training. Participants were categorised as musically trained if their total formal music training exceeded 12 months.

2.8.2. Task 1 (4-item) questionnaire

At the completion of blocks 1 and 2, participants indicated the level of difficulty of the range and direction identification tasks on 9-point Likert scales (1 = *extremely difficult* and 9 = *extremely easy*), and their level of confidence in their ability to perform each task (1 = *not at all confident* and 9 = *very confident*).

2.8.3. Task 2 (2-item) questionnaire

At the completion of blocks 3 and 4, participants indicated the level of difficulty of the transition identification task and their level of confidence in their ability to perform the task on 9-point Likert scales, with end labels as above.

2.9. Procedure

The training and testing of each participant proceeded as follows.

1. **Introduction.** Participants read the information sheet and signed a consent form.
2. **Questionnaire 1 (music) and assignment to sonification.** Participants completed the formal musical training questionnaire and were randomly assigned to sonification condition.
3. **Training phase 1.** Training began with a verbal description of the experiment. Visual aids were used to explain oxygen ranges, tones, and the experimental task. Participants completed 7 practice trials in a fixed order, and received verbal feedback at the end of each trial.
4. **Experimental testing phase 1.** Participants completed range identification and direction identification tasks in two blocks (1 and 2) of 12 trials each, presented in an order randomised across participants.
5. **Questionnaire 2.** Participants completed the Phase 1 questionnaire.
6. **Training phase 2.** Training began with a verbal description of the range identification and target transition tasks, supported by a visual aid showing a screen shot of the experimental task. Participants completed 7 practice trials in a fixed order, and received verbal feedback at the end of each trial.
7. **Experimental testing phase 2.** Participants completed range identification and target transition identification tasks in two blocks (3 and 4) of 12 trials each, presented in an order randomised across participants.
8. **Questionnaire 3.** Participants completed the Phase 2 questionnaire.

3. Results

Of the 104 participants recruited, data from four participants were dropped from the analysis due to equipment failure during testing. Analyses were performed on the data for the remaining 100 participants. For the LogLinear, Beacon, and Tremolo conditions, respectively, the proportion of male participants was 28%, 28%, and 25%; mean (SD) years of age was 21.3 (3.2), 19.7 (2.7), and 19.9 (2.0); and mean (SD) for years of musical training was 2.8 (3.2), 2.7 (3.4), and 2.8 (3.5).

3.1. Primary outcomes

Residuals analyses were conducted as part of initial parametric tests to determine the suitability of each primary outcome measure

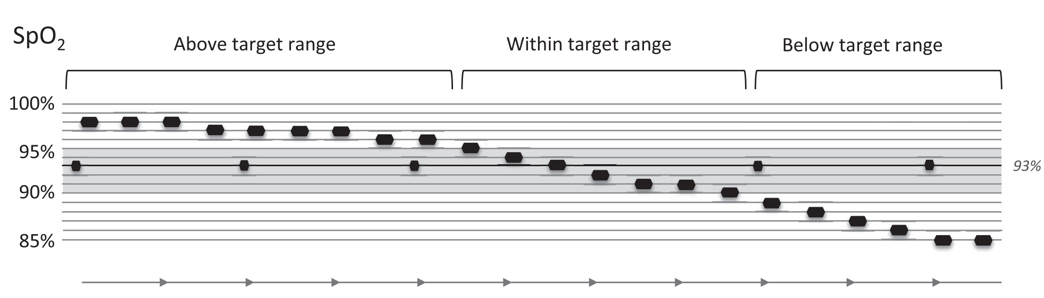

Fig. 2. Visual aid used to train Beacon participants. The beacon, or reference tone, is shown as a short tone preceding every fourth pulse tone when SpO_2 is above or below the target range, but not in the target range.

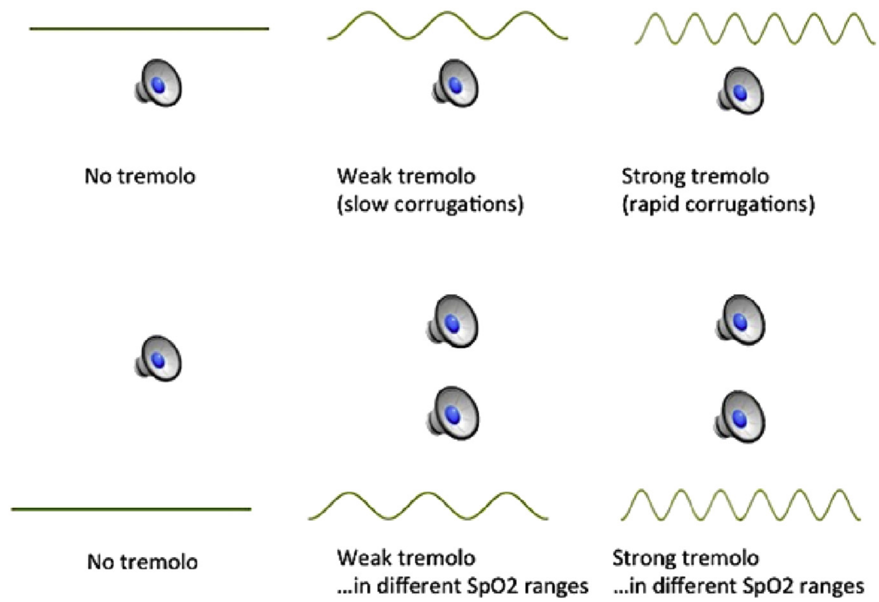

Fig. 3. Visual and auditory aid used to train Tremolo participants. The top row indicates the difference between no tremolo, weak tremolo (3 cycles) and strong tremolo (6 cycles). The bottom row indicates examples of tremolo tones in the 4 out-of-target ranges.

for parametric analysis. Examinations of plots for normality, homoscedasticity of variance, and independence of residuals revealed that the dependent measures did not meet one or more of the assumptions, so all primary outcome analyses were conducted using non-parametric tests. Medians, lower quartiles, and upper quartiles are shown in Table 2 and are graphed in the figures.

3.1.1. Range identification

The range identification data from all four experimental blocks were combined for the analysis. Mann–Whitney U tests indicated that range identification accuracy was significantly higher for participants using Tremolo than for participants using LogLinear, $U = 52.00$, $z = -6.256$, $p < .001$, $r = .63$ (see Fig. 4). Range identification accuracy was significantly higher for participants using Beacon than for participants using LogLinear, $U = 78.50$, $z = -6.051$, $p < .001$, $r = .61$.

3.1.2. Direction identification

Mann–Whitney U tests indicated that direction identification accuracy was not significantly different for participants using Tremolo compared to those using LogLinear, $U = 446.50$, $z = -1.089$, $p = .276$, $r = .11$ (see Fig. 5). Direction identification accuracy was not significantly different for participants using Beacon compared to those using LogLinear $U = 461$, $z = -1.271$, $p = .204$, $r = .12$.

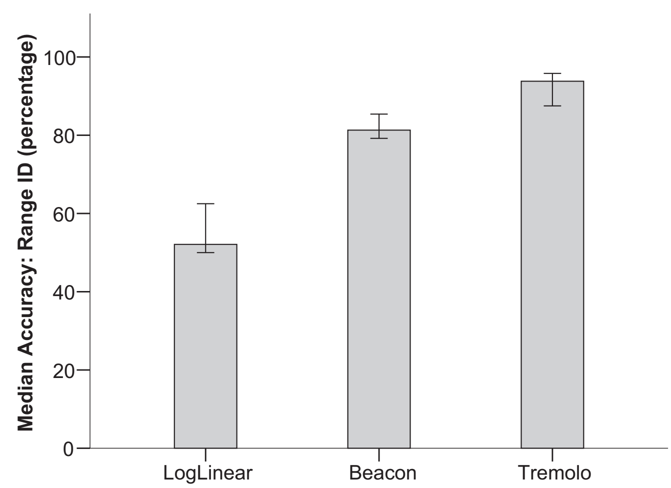

Fig. 4. Range identification median accuracy for LogLinear, Beacon, and Tremolo sonifications. Error bars are non-parametric 95% confidence intervals.

3.1.3. Transition identification

Mann–Whitney U tests indicated that transition identification accuracy scores were significantly higher for participants using Tremolo compared to those using LogLinear, $U = 35.00$, $z = -6.526$,

Table 2
Medians (Mdn) of the primary outcomes measures plus transition identification RT.

	LogLinear			Beacon			Tremolo		
	Mdn	*LQ*	*UQ*	*Mdn*	*LQ*	*UQ*	*Mdn*	*LQ*	*UQ*
Range ID	52%	50%	58%	81%[a]	79%	84%	94%[a]	90%	96%
Direction ID	88%	79%	96%	96%	88%	100%	96%	92%	96%
Transition ID	71%	63%	79%	100%[a]	96%	100%	96%[a]	96%	100%
Transition RT (s)	7.6	5.9	9.0	2.8[a]	2.5	3.0	1.2[a,b]	0.9	1.8

Note.
LQ = Lower Quartile; *UQ* = Upper Quartile.
[a] Indicates significantly different from LogLinear at $p < .001$ level.
[b] Indicates significantly different from Beacon at $p < .001$.

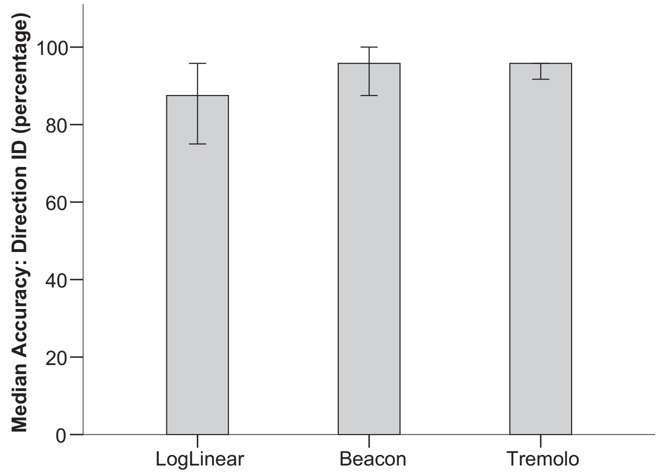

Fig. 5. Direction identification median accuracy for LogLinear, Beacon, and Tremolo sonifications. Error bars are non-parametric 95% confidence intervals.

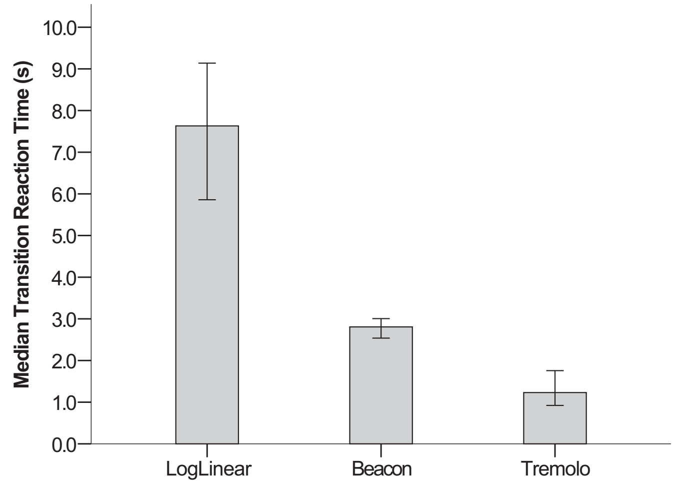

Fig. 7. Median transition reaction times (s) for LogLinear, Beacon, and Tremolo sonifications. Error bars are non-parametric 95% confidence intervals.

$p < .001$, $r = .65$ (see Fig. 6). Transition identification accuracy scores for participants using Beacon were significantly higher than for those using LogLinear, $U = 13.00$, $z = -6.962$, $p < .001$, $r = .69$.

3.2. Secondary outcome measures

3.2.1. Transition reaction time

Mann–Whitney U tests were performed to determine if there were differences in median transition reaction times between all three pairs of conditions. Transition reaction times were faster for participants using Tremolo than for participants using LogLinear, $U = 47.00$, $z = -6.311$, $p < .001$, $r = .63$ (see Fig. 7). Transition reaction times were faster for participants using Beacon than for participants using LogLinear, $U = 150.00$, $z = -5.146$, $p < .001$, $r = .51$. Finally, transition reaction times were faster for participants using Tremolo than for participants using Beacon, $U = 34.00$, $z = -6.67$, $p = < .001$, $r = .67$.

3.2.2. Questionnaires

Mann–Whitney U tests were performed to compare participants' perceptions of task difficulty and their confidence in completing tasks across conditions. Results are in Table 3. For range identification, there were no differences between participants using LogLinear versus Beacon for perceived task difficulty or confidence level. However, participants using Tremolo found range identification easier compared with participants using LogLinear or Beacon, and they were more confident than participants using LogLinear. For the direction identification task, there were no differences between conditions in perceived difficulty or level of confidence to perform the task. For the transition identification task, participants using the Beacon or Tremolo sonification found the task easier and were more confident compared with participants using LogLinear.

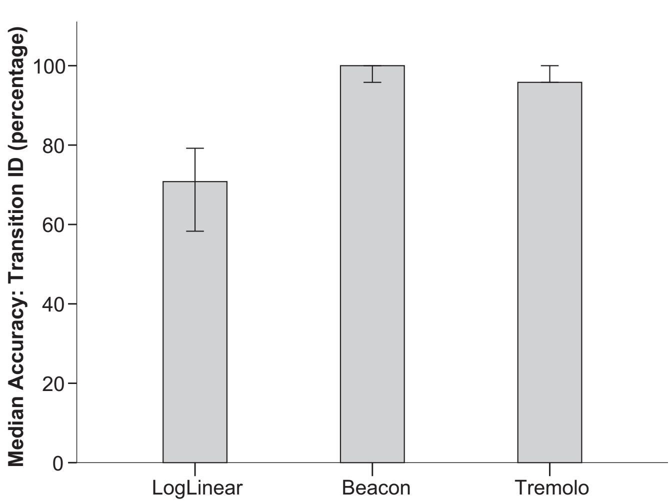

Fig. 6. Transition identification median accuracy for LogLinear, Beacon, and Tremolo sonifications. Error bars are non-parametric 95% confidence intervals.

4. Discussion

The results for Tremolo sonification are discussed, followed by the findings for the Beacon sonification. Then implications, limitations, and future research are discussed.

4.1. Tremolo vs. LogLinear

The present study operationalized and tested a Tremolo sonification intended to help clinicians monitor SpO_2 in pre-term neonates receiving supplemental oxygen. As predicted, non-clinical participants using the Tremolo sonification to monitor sequences of tones identified SpO_2 range significantly more accurately (94%) than did participants using the LogLinear sonification (52%). Participants using Tremolo showed an advantage of 42 percentage points, which was distinctly larger than the 15-percentage point difference sought.

Participants using the Tremolo sonification identified SpO_2 direction with 96% median accuracy, compared to 88% for participants using the LogLinear sonification. These findings contrast with Experiment 3 in Hinckfuss et al. (2015), where direction identification with Tremolo was slightly less accurate than for LogLinear. There is clearly no indication that the introduction of a Tremolo pulse oximeter sonification display would create potential confusion about direction of change and increase reliance on visual monitoring.

The findings for transition identification were also consistent with predictions. Participants using Tremolo sonification showed 96% accuracy at detecting when SpO_2 went into or out of the target range, compared to 71% for participants using the LogLinear sonification. The 25-percentage point advantage for participants using

Table 3
Medians (Mdn) of the 6 questionnaire items for each sonification level. For questions 1, 2 and 5, 9 is "Extremely easy". For questions 3, 4 and 6, 9 is "Extremely confident".

	LogLinear			Beacon			Tremolo		
	Mdn	*LQ*	*UQ*	*Mdn*	*LQ*	*UQ*	*Mdn*	*LQ*	*UQ*
Q1: Range (H/E)	4.75	4.25	5.50	4.00	3.50	5.75	6.50[a,b]	6.50	7.00
Q3: Range (C)	5.25	4.50	6.00	5.50	5.00	6.00	6.50[a]	6.50	7.25
Q2: Direction (H/E)	7.25	6.50	7.50	7.50	6.50	7.50	7.00	6.00	7.50
Q4: Direction (C)	7.00	6.50	7.50	7.50	6.50	8.00	6.50	6.00	7.00
Q5: Transition (H/E)	3.50	3.00	4.50	7.50[a]	7.00	8.00	7.00[a]	6.50	7.50
Q6: Transition (C)	4.00	3.50	4.50	7.50[a]	7.00	8.00	7.00[a]	6.50	7.50

Note.
LQ = Lower Quartile; *UQ* = Upper Quartile; (H/E) = Hard versus Easy; (C) = Confidence.
[a] Indicates significantly different from LogLinear at $p < .002$ level.
[b] Indicates significantly different from Beacon at $p < .001$.

the Tremolo sonification over the LogLinear sonification was again larger than the 15-percentage point advantage sought. These findings suggest that a Tremolo sonification offers clearer acoustic signalling of the transitions into and out of the target SpO_2 zone, an important condition to support preattentive monitoring (Sanderson, 2006).

An exploratory analysis indicated that participants in the Tremolo condition detected transitions with a median latency of 1.2 s, more than 6 s faster than participants in the LogLinear condition with 7.6 s. Questionnaire results reflected all the above findings. Participants using the Tremolo sonification rated it easier to identify range and to detect transitions than did participants using LogLinear sonification, and their levels of confidence were higher.

These findings suggest that a Tremolo pulse oximeter sonification shows promise for neonatal monitoring. The findings build on those of Brown et al. (2015) who found that participants showed a 30% improvement in their ability to detect oxygen range using a LogLinear sonification compared with a Linear sonification. The present study used a LogLinear sonification as a control condition yet was still able to show a 42-percentage point advantage of Tremolo over LogLinear for range identification, and a 25-percentage point advantage for transition identification.

These findings support and extend the research of Hinckfuss et al. (2015; Experiment 3) in which a more modest advantage of 12 percentage points was found for Tremolo over LogLinear. The greater benefit found in the present experiment for range identification with the Tremolo sonification may be explained by the changes in how the tremolo was implemented, with 3 versus 6 cycles at 90% wet distinguishing the non-target ranges rather than the 3 versus 5 cycles at 40% wet used previously. These changes improved the auditory distinction between the low and very low ranges, and between the high and very high ranges.

Nonetheless, a comment by the final Tremolo participant run in the study warrants discussion. The participant commented that towards the end of the experiment he noticed that the pilot light on the speakers flickered slightly with changes in sound power, including changes in sound power associated with the tremolo. If other participants in the Tremolo condition had noticed the changes in the pilot light and been able to interpret it, then the superiority of the Tremolo condition could conceivably be attributed to participants monitoring the pilot light rather than the tremolo sound. However, this possibility is unlikely, for the following reasons. First, an inspection revealed that the variation in brightness of the pilot light was very subtle, and unlikely to have been noticed or used to any effect by participants. Second, benefits from Tremolo were already evident in Experiment 3 of Hinckfuss et al. (2015), which was performed with internal laptop speakers, and the operationalization of tremolo in the present experiment made the ranges even more formally distinct. Third, a recent small study in our laboratory using internal laptop speakers (no pilot light) rather than external speakers has reproduced the present pattern of findings, with median range identification accuracy at 95% for the Tremolo condition, 79% for the Beacon condition, and 56% for the LogLinear condition. Therefore it is exceptionally unlikely that changes in the pilot light could explain the current Tremolo results.

4.2. Beacon versus LogLinear

As also seen in previous studies, participants using the Beacon sonification identified range significantly more accurately (81%) than participants using the LogLinear sonification (52%)—a 29-percentage point advantage that is well above the 15-percentage point advantage sought. The findings are in agreement with Hinckfuss et al. (2016; Experiment 2), who showed that range identification accuracy improved by 25 percentage points with a Beacon pulse oximeter sonification compared to LogLinear sonification. The current findings support Schulte and Block's (1992) suggestion that a beacon, or "heralding" tone could improve accuracy for detecting clinically important oxygen saturation ranges, compared to variable pitch sonification alone. Nonetheless, the questionnaire results suggest that the Beacon and LogLinear participants found range identification equally difficult, and had similar levels of confidence in their performance.

Participants using the Beacon sonification did not identify direction significantly more accurately (96%) than participants using the LogLinear sonification (88%), but clearly there was no trend to be less accurate, as had been observed in a previous study. The present results therefore do not replicate the difficulties evident in Hinckfuss et al. (2015; Experiment 3). Direction identification accuracy with the Beacon sonification was much higher in the present study than in the latter study where participants using the Beacon sonification showed the lowest directional accuracy scores at 73%. The present results for the Beacon sonification were closer to the absolute levels of performance in Hinckfuss et al. (2016; Experiment 2) where direction identification accuracy was 96% for the LogLinear and 94% for the Beacon sonification.

Participants using the Beacon sonification showed higher accuracy (100%) in detecting the moment of oxygen level transition into or out of target range than participants using the LogLinear sonification (71%). The 29-percentage point difference in performance is, again, larger than the 15 percentage points sought. In addition, Beacon participants rated the transition identification task easier and their confidence higher than LogLinear participants. Overall, the results suggest that the extra beacon tone provides a strong, unambiguous signal for listeners—through its absence or presence—that oxygen has transitioned into or out of the target range.

An exploratory analysis of latencies showed that participants using the Beacon sonification detected transitions almost 5 s more quickly (2.8 s) than those using the LogLinear sonification (7.6 s). The questionnaire data revealed that participants using the Beacon sonification found the transition identification task easier and they had more confidence performing the task compared to participants using the LogLinear sonification. However, the ratings of ease and confidence for range identification by participants using the Beacon sonification were not higher than the ratings by participants using the LogLinear sonification.

4.3. Implications, limitations, and future research

Together with the results of Hinckfuss et al. (2016) the current findings provide strong evidence that current pulse oximetry sonification could be improved to support the monitoring of preterm neonates receiving supplemental oxygen. Log-linear mappings of SpO_2 to pitch alone do not provide an unambiguous signal that alerts clinicians to oxygen levels that are deviating from a prespecified target zone. Each new sonification tested herein—Tremolo and Beacon—has advantages and disadvantages, and further testing is needed.

A first limitation is that non-clinical participants were used and the scenarios did not reflect clinical situations. It is important that future research with the Beacon and Tremolo sonifications should test clinicians in more clinically representative situations. Related to this is the question of whether the Beacon and Tremolo ranges should be fixed, or whether clinicians should be able to adjust the ranges for the clinical circumstances and a patient's needs. The principles tested in the present experiments are clearly flexible in how they can be used.

Second, a sonification must be well-tolerated by those within the acoustic environment (Walker and Nees, 2011; Watson and Sanderson, 2004). When the beacon is mapped onto every fourth pulse tone for a neonatal HR of 120 bpm, an extra 30 sounds per minute are added to the acoustic environment. Although the Beacon sonification produces a strong indication that deterioration is starting, the most detectable sound may not necessarily the best solution (Watson and Sanderson, 2007). In addition, it is unclear how effective the rhythmic Beacon sonification will remain when HR is irregular or rapidly changing, potentially breaking the rhythm that is heard with a steady HR. In ongoing research we are using different periodicities of the beacon to distinguish low vs. very low and high vs. very high non-target ranges, to lessen its overall frequency and to improve its informativeness.

Third, the Tremolo sonification distinguishes the different non-target ranges without adding extra tones, which might offer an advantage over the Beacon sonification. However, during resuscitation and transportation, with increased noise and cognitive load, the tremolo may be harder to detect. Stevenson et al., (2013) found that increases in cognitive load and noise reduced anaesthesiologists' ability to detect changes by 17%. Our participants were tested without distraction or secondary task load, whereas during resuscitation and transportation, other cognitively demanding critical tasks are performed. Further research is required to examine the effectiveness of the Beacon and Tremolo sonifications under conditions that more accurately reflect noise and workload in the neonatal monitoring context and environment.

Fourth, participants only had a very short training period (although longer than the formal training that clinicians receive) but accuracy scores for range, direction, and transition identification with the Tremolo display were all very high, ranging between 94% and 96%. Therefore implementing a Tremolo sonification display in a clinical setting might not require extensive training time and costs. The time and expense of multisensory training to improve pitch perception on existing pulse oximeters in an effort to improve patient monitoring (Schlesinger et al., 2014) may be unnecessary if a sonification design can provide even better improvements that work well in the clinical environment.

Fifth, the present experiments tested participants' ability to identify SpO_2 ranges, and not their ability to identify exact SpO_2 levels, as was done in previous research (Brown et al., 2015; Morris and Mohacsi, 2005). Further experiments should test whether the Beacon display, in particular, also improves participants' ability to identify exact levels of SpO_2.

5. Conclusion

In summary, adding tremolo to the tones of a conventional pulse oximeter sonification leads to faster and more accurate detection of deviations of simulated SpO_2 levels from a target range. Tremolo might even offer advantages over a Beacon sonification for SpO_2 range identification. Further research with clinical participants in a simulated clinical setting is needed to determine if enhanced sonifications can lead to easier and more accurate identification of oxygen saturation levels in a dynamic environment.

Acknowledgements

The authors gratefully acknowledge funding from the Australian Research Council through Discovery Project grant DP140101822.

References

Askie, L.M., Brocklehurst, P., Darlow, B.A., Finer, N., Schmidt, B., Tarnow-Mordi, W., 2011. NeOProM: neonatal oxygenation prospective meta-analysis collaboration study protocol. BMC Pediatr. 11, 1–9. http://dx.doi.org/10.1186/1471-2431-11-6.

Brown, Z., Edworthy, J., Sneyd, J.R., Schlesinger, J., 2015. A comparison of linear and logarithmic auditory tones in pulse oximeters. Appl. Ergon. 51, 350–357. http://dx.doi.org/10.1016/j.apergo.2015.06.006.

Carlo, W.A., Finer, N.N., Walsh, M.C., Rich, W., Gantz, M.G., Laptook, A.R., et al., 2010. Target ranges of oxygen saturation in extremely preterm infants. N. Engl. J. Med. 362, 1959–1969. http://dx.doi.org/10.1056/NEJMoa0911781.

Chow, L.C., Wright, K.W., Sola, A., Group, C.O.A.S., 2003. Can changes in clinical practice decrease the incidence of severe retinopathy of prematurity in very low birth weight infants? Pediatrics 111, 339–345. http://dx.doi.org/10.1542/peds.111.2.339.

Collins, M.P., Lorenz, J.M., Jetton, J.R., Paneth, N., 2001. Hypocapnia and other ventilation-related risk factors for cerebral palsy in low birth weight infants. Pediatr. Res. 50, 712–719. http://dx.doi.org/10.1203/00006450-200112000-00014.

Dawson, J.A., Kamlin, C.O., Vento, M., Wong, C., Cole, T.J., Donath, S.M., et al., 2010. Defining the reference range for oxygen saturation for infants after birth. Pediatrics 125, e1340–e1347. http://dx.doi.org/10.1542/peds.2009-1510.

Edworthy, J., 2013. Medical audible alarms: a review. J. Am. Med. Inf. Assoc. 20, 584–589. http://dx.doi.org/10.1136/amiajnl-2012-001061.

Edworthy, J., Page, R., Hibbard, A., Kyle, S., Ratnage, P., Claydon, S., 2014. Learning three sets of alarms for the same medical functions: a perspective on the difficulty of learning alarms specified in an international standard. Appl. Ergon. 45, 1291–1296. http://dx.doi.org/10.1016/j.apergo.2013.10.003.

Finer, N., Leone, T., 2009. Oxygen saturation monitoring for the preterm infant: the evidence basis for current practice. Pediatr. Res. 65, 375–380 doi: 0031-3998/09/6504-0375.

Ford, S.P., Leick-Rude, M.K., Meinert, K.A., Anderson, B., Sheehan, M.B., Haney, B.M., et al., 2006. Overcoming barriers to oxygen saturation targeting. Pediatrics 118, S177–S186. http://dx.doi.org/10.1542/peds.2006-0913P.

Goos, T.G., Rook, D., van der Eijk, A.C., Kroon, A.A., Pichler, G., Urlesberger, B., et al., 2013. Observing the resuscitation of very preterm infants: are we able to follow the oxygen saturation targets? Resuscitation 84, 1108–1113. http://dx.doi.org/10.1016/j.resuscitation.2013.01.025.

Hinckfuss, K., Sanderson, P., Liu, D., Browning, C., Loeb, R.G., Liley, H., 2015. Novel pulse oximetry sonifications to improve eyes free monitoring: three laboratory studies. In: Proceedings of the 59th Annual Meeting of the Human Factors and Ergonomics Society (HFES2015). Los Angeles, CA: 26-30 October, 536-536.

Hinckfuss, K., Sanderson, P., Loeb, R.G., Liley, H., Liu, D., 2016. Novel pulse oximetry sonifications for neonatal oxygen saturation monitoring: a laboratory study. Hum. Factors 58, 344–359. http://dx.doi.org/10.1177/0018720815617406.

Hockey, R., 1997. Compensatory control in the regulation of human performance under stress and high workload: a cognitive-energetical framework. Biol.

Psychol. 45, 73–93.

Janata, P., Edwards, W.H., 2013. A novel sonification strategy for auditory display of heart rate and oxygen saturation changes in clinical settings. Hum. Factors J. Hum. Factors Ergon Soc. 55, 356–372. http://dx.doi.org/10.1177/0018720812455433.

Lim, K., Wheeler, K.I., Gale, T.J., Jackson, H.D., Kihlstrand, J.F., Sand, C., et al., 2014. Oxygen saturation targeting in preterm infants receiving continuous positive airway pressure. J. Pediatr. 164, 730–736 e731. http://dx.doi.org/10.1016/j.jpeds.2013.11.072.

Loeb, R.G., Brecknell, B., Sanderson, P., 2016. The sounds of desaturation: a survey of commercial pulse oximeter sonifications. Anesth. Analg. http://dx.doi.org/10.1213/ANE.0000000000001240.

Loeb, R.G., Fitch, W.T., 2002. A laboratory evaluation of an auditory display designed to enhance intraoperative monitoring. Anesth. Analg. 94, 362–368. http://dx.doi.org/10.1213/00000539-20020000-00025.

Long, J.G., Luey, J.F., Philip, A.G.S., 1980. Noise and hypoxemia in the intensive care nursery. Pediatrics 65, 143–145.

Manja, V., Lakshminrusimha, S., Cook, D.J., 2015. Oxygen saturation target range for extremely preterm infants: a systematic review and meta-analysis. JAMA Pediatr. 169, 332–340. http://dx.doi.org/10.1001/jamapediatrics.2014.3307.

Morris, R.W., Mohacsi, P.J., 2005. How well can anaesthetists discriminate pulse oximeter tones? Anaesth. Intensive Care 33, 497–500.

Sanderson, P.M., 2006. The multimodal world of medical monitoring displays. Appl. Ergon. 37, 501–512. http://dx.doi.org/10.1016/j.apergo.2006.04.022.

Sanderson, P.M., Liu, D., Jenkins, S.A., 2009. Auditory displays in anesthesiology. Curr. Opin. Anaesthesiol. 22, 788–795. http://dx.doi.org/10.1097/ACO.0b013e3283326a2f.

Sanderson, P.M., Watson, M.O., Russell, W.J., 2005. Advanced patient monitoring displays: tools for continuous informing. Anesth. Analg. 101, 161–168. http://dx.doi.org/10.1213/01.ANE.0000154080.67496.AE.

Santamore, D.C., Cleaver, T.G., 2004. The sounds of saturation. J. Clin. Monit. Comput. 18, 89–92 doi: 10.1023/B:JOCM.0000032698.47717.06.

Saugstad, O.D., 1997. Bronchopulmonary dysplasia and oxidative stress: are we closer to an understanding of the pathogenesis of BPD? Acta Paediatr. 86, 1277–1282.

Saugstad, O.D., Aune, D., 2014. Optimal oxygenation of extremely low birth weight infants: a meta-analysis and systematic review of the oxygen saturation target studies. Neonatology 105, 55–63. http://dx.doi.org/10.1159/000356561.

Schlesinger, J.J., Stevenson, R.A., Shotwell, M.S., Wallace, M.T., 2014. Improving pulse oximetry pitch perception with multisensory perceptual training. Anesth. Analg. 118, 1249–1253. http://dx.doi.org/10.1213/ANE.0000000000000222.

Schulte, G.T., Block Jr., F.E., 1992. Can people hear the pitch change on a variable-pitch pulse oximeter? J. Clin. Monit. 8, 198–200. http://dx.doi.org/10.1007/BF01616776.

Stevenson, R.A., Schlesinger, J.J., Wallace, M.T., 2013. Effects of divided attention and operating room noise on perception of pulse oximeter pitch changes: a laboratory study. Anesthesiology 118, 376–381. http://dx.doi.org/10.1097/ALN.0b013e31827d417b.

Tin, W., Lal, M., 2015. Principles of pulse oximetry and its clinical application in neonatal medicine. Semin. Fetal Neonat. Med. 20, 192–197. http://dx.doi.org/10.1016/j.siny.2015.01.006.

Tin, W., Milligan, D.W.A., Pennefather, P., Hey, E., 2001. Pulse oximetry, severe retinopathy, and outcome at one year in babies of less than 28 weeks gestation. Arch. Dis. Child. Fetal Neonat. Ed. 84, F106–F110.

Walker, B.N., Nees, M.A., 2011. Theory of sonification. In: Hermann, T., Hunt, A., Neuhoff, J.G. (Eds.), The Sonification Handbook. Logo Verlag Berlin GmbH, Berlin, Germany, pp. 9–40.

Watson, M., Sanderson, P., 2004. Sonification supports eyes-free respiratory monitoring and task time-sharing. Hum. Factors 46, 497–517. http://dx.doi.org/10.1518/hfes.46.3.497.50401.

Watson, M., Sanderson, P., 2007. Designing for attention with sound: challenges and extensions to ecological interface design. Hum. Factors 49, 331–346. http://dx.doi.org/10.1518/001872007X312531.

Woods, D.D., 1995. The alarm problem and directed attention in dynamic fault management. Ergonomics 38, 2371–2393. http://dx.doi.org/10.1080/00140139508925274.