A PSYCHOACOUSTIC SOUND DESIGN FOR PULSE OXIMETRY

Sebastian Schwarz

University of Hamburg
Institute of Systematic Musicology
Neue Rabenstr. 13, 20354 Hamburg, Germany
bav7222@studium.uni-hamburg.de

Tim Ziemer

University of Hamburg
Institute of Systematic Musicology
Neue Rabenstr. 13, 20354 Hamburg, Germany
tim.ziemer@uni-hamburg.de

ABSTRACT

Oxygen saturation monitoring of neonates is a demanding task, as oxygen saturation (SpO_2) has to be maintained in a particular range. However, auditory displays of conventional pulse oximeters are not suitable for informing a clinician about deviations from a target range. A psychoacoustic sonification for neonatal oxygen saturation monitoring is presented. It consists of a continuous Shepard tone at its core. In a laboratory study it was tested if participants (N = 6) could differentiate between seven ranges of oxygen saturation using the proposed sonification. On average participants could identify in 84% of all cases the correct SpO_2 range. Moreover, detection rates differed significantly between the seven ranges and as a function of the magnitude of SpO_2 change between two consecutive values. Possible explanations for these findings are discussed and implications for further improvements of the presented sonification are proposed.

1. INTRODUCTION

In a clinical environment auditory displays can be very beneficial for patient monitoring, especially when visual attention is committed with another task [1]. The translation of input data to sound is called sonification, which is considered as the central element of an auditory display [2]. As sound is a temporal medium, process monitoring seems to be a very promising candidate for sonifications [3]. In a monitoring situation temporally-related data has to be observed and it is important to recognize changes in the current state of the process to be able to intervene appropriately in time [3]. In a clinical context auditory displays are already very common. For example there exists a huge variety of different alarms for patient monitoring. However, there seem to be drawbacks using them [4]. Apart from auditory alarms, auditory displays have the potential to inform the listener continuously about the current state of a patient, rather than putting him in a sudden state of alert [5]. This way the issue about when information is presented can be avoided and moreover the sonification also informs about normal states of the process, while attention is not attracted in an inappropriate way [6]. For example in the case of pulse oximetry, auditory displays seem to be of great use for patient monitoring, as they can shorten reaction times [5] and improve performances in time-shared tasks [5], [7].

Pulse oximeters are used to monitor oxygen saturation (SpO_2) and to prevent unwanted deviations [8]. The realization of a high level of SpO_2 was often supported by the aim to avoid negative consequences of hypoxemia and tissue hypoxia [9]. However, optimal oxygen saturation differs significantly across ages [1], [10]. Mainly patients at the extremes of age are at high risk of potential detriments of hyperoxia [10], [11]. In a meta-analysis the effect of functional oxygen saturation targets in premature infants was examined, which revealed an increased relative risk for mortality and necrotizing enterocolitis and a reduced relative risk of severe retinopathy of prematurity for a low compared to a high oxygen saturation target [12]. According to these results, the functional SpO_2 should lie between 90- and 95% in case of a gestational age under 28 weeks until 36 weeks postmenstrual age [12]. It is therefore of high importance to keep the oxygen saturation level in newborns in a particular range [1]. However, the maintenance of SpO_2 in a particular range using a pulse oximeter seems to be difficult, as could be shown in the case of preterm infants [13], [14]. In a conventional pulse oximeter a tone can be heard on each heartbeat and the pitch of the tone is varying with the oxygen saturation [15]. With the oxygen saturation rising or falling, the pitch is accordingly going up or down. Although most manufacturers include a variable pitch tone in their pulse oximeters, the acoustic properties of this tone are not standardized [16], which can lead to confusion interpreting the sonification [17]. For example the mapping between SpO_2 and frequency can be linear or logarithmic, whereby pitch perception is logarithmic rather than linear in nature [18]. Accordant to that, anaesthetists could estimate absolute oxygenation values as well as the size of oxygenation level differences significantly more accurate with a logarithmic pitch scale than with a linear scale [18]. Nonetheless, considering the specific demands on oxygen supply for neonates, a clinician would need more direct information, if and to what extent the SpO_2 level is moving out of a target range, unless he regularly checks the SpO_2 level on a visual monitor [1].

In a recent study a novel pulse oximetry sonification for neonatal oxygen saturation monitoring was proposed [1]. In two experiments it was tested, if nonclinician's ability to identify a target range of SpO_2 (90-95%) would improve with a modified version of a conventional pulse oximeter with a logarithmic mapping between SpO_2 and pitch. Two different redesigns of the conventional sonification were compared to the control condition. For the first sonification the pitch differences became very small in the target zone and increasingly large outside the target zone. This design didn't improve range identification accuracy compared to the control condition. In a second redesign [1] a fixed-pitch reference tone was included, when SpO_2 was outside of the target range. The pitch of this reference tone corresponded to the pitch at a SpO_2 level of 93% and it preceded every fourth pulse. This sonification significantly improved the accuracy of SpO_2 range identification in comparison to the control condition (85% vs. 60%). Consequently a modified sonfication seems to be beneficial for the listeners ability to detect a specific range of SpO_2. In a subse-

https://doi.org/10.21785/icad2019.024

quent study different levels of tremolo were added to a conventional pulse oximeter to test, if this would help listeners to identify SpO_2 ranges, direction of change and target transitions [19]. SpO_2 ranges were subdivided into five ranges, a target range and two ranges below and above the target range. In the target range no tremolo was used, whereby three cycles of tremolo were added each time a SpO_2 range was reached, that deviated further form the target range. SpO_2 ranges and transitions into and out of the target range were identified more accurately with the advanced sonification, than with the conventional sonification of a pulse oximeter. According to this, adding tremolo to a conventional pulse oximeter seems to be beneficial for identifying SpO_2 ranges and might even be more effective than the use of a reference tone [19]. Similarly in another study, tremolo and brightness were used to differentiate three SpO_2 ranges [20]. Participants of this study could successfully identify SpO_2 ranges (*Mdn* = 100 %), as well as transitions into and out of the target range (*Mdn* = 100 %).

This work proposes a novel sonification for pulse oximetry to convey information about current SpO_2 of neonates receiving supplemental oxygen. Unlike the examples discussed above, this design deviates further from the auditory display of a conventional pulse oximeter, as a Shepard tone [21] forms the basis of the sonification. Among other things, this approach is motivated by the aim to differentiate a larger number of SpO_2 ranges. In a listening test the effectiveness of the proposed sonification for identifying seven different SpO_2 ranges was tested. On the basis of the results of the listening test, further adjustments of the sonification are discussed.

2. THE SONIFICATION

The sonification is derived from the psychoacoustic sonification for navigation that has been introduced in [22] and discussed in a clinical context in [23]. The technical implementation is explained in [24]. The central element of the sonification is a continuous Shepard tone. In a preliminary study the Shepard tone has proven to be helpful in finding a target region [22]. As it might be important for a clinician to be able to estimate the distance of current SpO_2 from a predefined target range, a Shepard tone was used instead of the conventional mapping of SpO_2 to pitch. The Shepard tone contains the carrier frequencies

$$f_n = f_0 2^n \text{Hz}, \qquad (1)$$

whereby $n = 0, \ldots, N-1$. In total the Shepard tone contains six carrier frequencies with $f_0 = 100$ Hz. If SpO_2 values are above or below the center of the target range, the Shepard tone is rising or falling in frequency respectively. This way the information about SpO_2 being below or above the center of the target range is conveyed by a simple binary coding. All carrier frequencies are rising or falling with the function

$$f(\phi) = f_0 2^{\frac{\phi N}{2\pi}}. \qquad (2)$$

This way neighboring carrier frequencies are always one octave apart. In Eq. (2) ϕ is the phase of one cycle, such that the frequency rises from f_0 to f_N. The phase ϕ is defined as

$$\phi(\theta, t) = \arg[\sin(2\pi\theta t)], \qquad (3)$$

whereby θ is a function of the distance to the center of the SpO_2 target range. This way the speed of the Shepard tone (rising or falling) is dependent on the distance to the center of the SpO_2 target range (90-95%), such that the speed increases the further SpO_2 deviates from the center. The amplitude of one frequency is weighted by a simple bell shaped curve. Consequently the amplitude of partials close to f_0 and at f_N are gradually reaching 0. A temporal envelope curve is used to create a pulse like sound, as the Shepard tone is supposed to get integrated in the sound design of conventional pulse oximeters. The frequency interval every pulse goes through, is increasing or decreasing with the Shepard tone gaining or losing speed respectively. This way a continuous mapping for the distance of current SpO_2 from the center of the target range is provided. A logarithmic mapping from distance to speed is used, such that a 1% change of SpO_2 would result in an approximately equal change of the perceived frequency interval. As the partials of the Shepard tone are continuously rising or falling, it is likely to happen, that the phase is varying between different pulses. Therefore, it is important that the Shepard tone is reseted to the starting point of its period *T* with every pulse of the oximeter. This means that the point of origin is held constant for every pulse, avoiding possible confusion, as the period of the Shepard tone contains no additional information.

The aim of this sonification was to enable the listener to differentiate between seven different ranges of SpO_2 illustrated in Figure 1. This is achieved by subdividing the target range (90-95%) into five ranges, consisting of a center range (92-93%) and two ranges below (90-91% and 91-92%) and above (93-94% and 94-95%) the center range. The remaining two SpO_2 ranges are defined as below ($<$ 90%) or above ($>$ 95%) the target range. SpO_2 ranges are numerated starting with range 1 at the top (see Figure 1). Pink noise is used to indicate that SpO_2 is within the target range. It provides a continuous background sound, such that it does not only occur within the time window of every pulse. Pink noise is used, as it is considered to be more pleasing to hear than white noise. This way the most critical information about the current SpO_2 is provided by placing only a minimum of cognitive workload on the clinician. Further information about the position of SpO_2 can be inferred by the direction and the speed of the Shepard tone. Within the center range (92-93%) the speed of the Shepard tone is set to 0, resulting in a pulse with a constant pitch. Deviations below or above the center range result in an increasingly falling or rising speed respectively. Thus, by identifying a rising or falling motion of the Shepard tone, a clinician should be able to locate current SpO_2 below or above the center range. To further differentiate between the remaining two ranges below (90-91% and 91-92%) and above (93-94% and 94-95%) the center range, the listener has to rely on the size of the interval the particular pulse goes through. The speed of the Shepard tone reaches its maximum at 90% and 95% of SpO_2 respectively, such that further deviations of SpO_2 do not result in an additional increase of speed. SpO_2 values outside the target range (90-95%) are made audible by the vanishing of the pink noise, whereby the direction of the Shepard tone still indicates, if current SpO_2 is below or above the center range. Nonetheless, a redundant coding is chosen, to make the ranges below ($<$ 90%) and above ($>$ 95%) the target range more distinguishable. A redundant coding by a second parameter can increase the robustness of the auditory display, as it may reinforce the representation parameter [25]. For SpO_2 values below the target range, frequency modulation is used to increase the perceived roughness of the Shepard tone, whereas for SpO_2 values above the target range the sound is not further manipulated. By using FM-synthesis to create roughness the perceived inharmonicity, roughness and noisiness increases with an increasing modu-

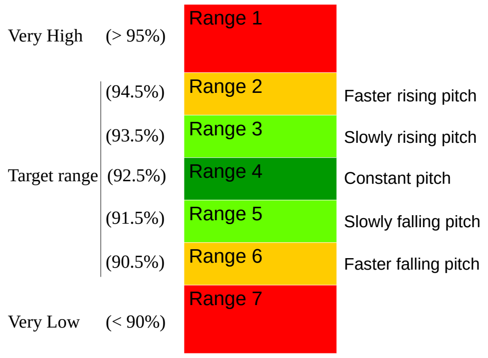

Figure 1: Subdivision of SpO_2 ranges. The target range (90-95%) is further subdivided into five SpO_2 ranges.

lation depth, such that the sound is perceived as more urgent [2]. Consequently the proposed sonification suggests a higher need for action in the case of hypoxia than in the case of hyperoxia. A demo video can be found on the second authors Youtube channel (`https://youtu.be/5kwzCunbLrA`).

3. METHOD

A convenience sample was recruited, consisting of students (N = 5) and staff (N = 1) of the Institute of Systematic Musicology at the University of Hamburg. In total 6 participants (1 female and 5 male) with an average age of 27.6 years (age range: 22-32 years) took part in the listening test. With only 6 participants the sample was rather small and not very representative, which should be kept in mind, while interpreting the results. All participants were non clinicians and except for one participant had no or little experience with sonifications. Participants were seated around two broadband loudspeakers, approximately 2-3 meters away. Due to economic reasons, all participants were tested simultaneously and were therefore instructed not to communicate with each other during the listening test, to prevent potential bias in the individual performances. The primary outcome variable was the detection rate calculated as the percentage of correct identified SpO_2 ranges. As described earlier, the principle of the proposed sonification is continuous between 90 and 95% of oxygen saturation. More precisely the frequency interval the Shepard tone went through got continuously bigger between 93- and 95% and 92- and 90% of SpO_2. As the participants had to discriminate between two different SpO_2 ranges in each of these cases, they could solely rely on the magnitude of the corresponding interval to do so. In the proximity of the transition from one range to the other it would be almost impossible to identify the correct range by hearing alone. Therefore, all values in the range of 93- to 95% and 90- to 92% were replaced by the mean of the corresponding range. The value of 90.2% was for example replaced by the corresponding value of 90.5%, which is the mean of 90 and 91%.

At first the sonification was explained, in particular the theoretical background and the applied mapping of data and sound, which was supported by auditory examples. After that the participants took part in a training session, which lasted about 5 minutes. In this session, participants had to listen to the modified pulse oximeter, which produced a pulse-like sound with a heart frequency of 60 Hz. Since it was assumed that the identification of the correct SpO_2 range each second would be too demanding for an untrained person, the value of the oxygen saturation was changed every two pulses. This way participants had two seconds for every SpO_2 value to identify the correct range. SpO_2 values were chosen arbitrarily, to cover all relevant ranges in a relatively short amount of time. Altogether, the training session consisted of four blocks, whereby in each block participants had to identify the correct SpO_2 range of five consecutive SpO_2 values. For each SpO_2 value the participants had to tick the correct box in a 7x5 table, whereby each row corresponded to one of the seven oxygen ranges and each column to one of the five SpO_2 values. After each part of the training session a feedback in terms of the correct answers was provided and a short break of approximately 30 seconds was taken. To indicate the start of a sequence, two pulses with the corresponding sound of 92.5% of oxygen saturation were always played at the beginning.

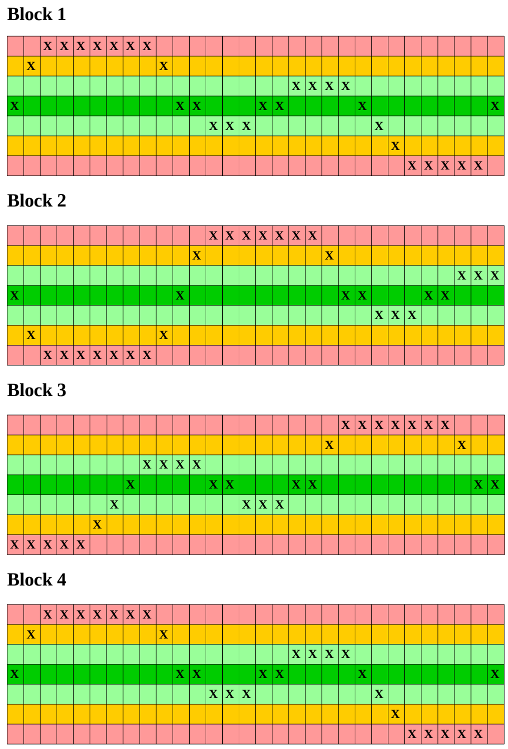

Figure 2: In each of the four blocks participants had to identify 30 SpO_2 values by ticking the correct box in a 7x30 table. The 7 rows correspond to the 7 SpO_2 ranges and each column to one SpO_2 value, which changed for every second pulse with a frequency of 30 Hz. The sample solution for each block is depicted above.

After the training was completed, the actual experimental task was performed, which lasted for approximately 10 minutes. In

Table 1
Median (Mdn), upper (HQ) and lower (LQ) quartiles for the detection rate of each SpO_2 range

	Mdn	*LQ*	*HQ*
Range 1	89%	60%	96%
Range 2	18%	12%	25%
Range 3	100%	78%	100%
Range 4	100%	100%	100%
Range 5	100%	87%	100%
Range 6	87%	78%	87%
Range 7	100%	100%	100%

Note. Detection rates were calculated as percentage of correct SpO_2 range identifications.

Table 2
Effect sizes (r) for multiple post hoc comparisons

	Range 1	Range 2	Range 3	Range 4	Range 5	Range 6	Range 7
Range 1		-.62	-.30	-.55*	-.47	-.06	-.55*
Range 2			-.61***	-.61***	-.62***	-.61*	-.61***
Range 3				-.26	-.08	.00	-.26
Range 4					-.26	-.56	.00
Range 5						-.31	-.26
Range 6							-.56

Note. P-values were calculated by a post hoc test after Conover (1999). Bonferroni adjustment method was used; *p<.05, **p<.01, ***p<.001.

contrast to the training session, participants had to identify 30 SpO_2 ranges in each of the four blocks and no feedback was given after each sequence. The four blocks are illustrated in Figure 2. In addition, the SpO_2 values were generated by a sine function. A smooth function was used, because it was considered to be in line with the fluctuations of oxygen saturation in an actual clinical setting. To account for possible training affects trial 1 and 4 were identical. Moreover, trial 3 was the reversal of trial 1 to examine possible effects of the direction of SpO_2 movement. In trial 2 the sine function was shifted about 2/3 π to the right. For the evaluation of the experimental task each tick, which was not placed in the correct box, that is the row and the column had to be correct, was considered as a wrong answer.

All significance tests were conducted at a significance level of α = .05. Detection rates were calculated as the percentage of correct SpO_2 range identifications for each participant over all 4 trials. To examine possible differences between different SpO_2 ranges, detection rates were also calculated for each SpO_2 range respectively. As an inspection of the corresponding qq-plots revealed deviations from normality a Friedman rank sum test was applied and subsequent multiple comparisons were conducted by a post hoc test after Conover (1999) [26]. The Bonferroni correction was applied, in which the p values were multiplied by the number of comparisons. In addition, it was tested, if different SpO_2 increment sizes did have an effect on the detection rates. Again a Friedman rank sum test was applied, as the corresponding qq-plots did not form a straight line. A post hoc test after Conover (1999) and the Bonferroni correction were used for multiple comparisons as well. To examine possible training effects between trial 1 and 4 the Wilcoxon signed rank test was applied, as the sampling distribution of the differences between scores did not look normal on a qq-plot. Moreover, detection rates between trial 1 and 3 were compared to account for any effect of direction of SpO_2 movement. The Wilcoxon signed rank test was used as well, as the corresponding qq-plot showed deviations from normality. Furthermore, it was of particular interest, if participants could identify an SpO_2 value being either within or outside the target range. Therefore, all given answers were additionally evaluated on a binary basis, whereas only the confusion between SpO_2 values within and outside the target range was treated as an incorrect answer (*inside/outside error*).

4. RESULTS

On average participants could identify in 84% (about 102 of 120 answers) of all 120 SpO_2 values the correct range. The chances to randomly guess the correct box were 1/7 $\approx$ 14%. In 98% (about 118 of 120 answers) of all cases participants could identify either the correct range or its neighbor range. Chances of choosing the correct field or its neighbor with a random guess are 19/49 $\approx$ 38%. To find out which part of the sonification was most ambiguous for the participants, detection rates were calculated for each SpO_2 range respectively (see Table 1). Detection rates of the participants changed significantly over SpO_2 ranges ($\chi^2(6)$ = 24.96, $p < .001$). The results of multiple comparisons are summarized in Table 2. In addition, detection rates were varying significantly as a function of the SpO_2 increment size ($\chi^2(4)$ = 19.66, $p < .001$). An overview of the detection rates for different SpO_2 increment sizes and the post hoc test of multiple comparisons is given in Table 3 and 4 respectively. To further examine, if participants found it particu-

Table 3

Median (Mdn), upper (HQ) and lower (LQ) quartiles for the detection rate of different SpO_2 increment sizes

	Mdn	*LQ*	*HQ*
Two ranges up	20%	5%	35%
One range up	75%	75%	84%
No change	97%	95%	98%
One range down	75%	72%	77%
Two ranges down	20%	20%	35%

Note. Detection rates were calculated as percentage of correct SpO_2 range identifications.

Table 4

Multiple post hoc comparisons of detection rates of different SpO_2 increment sizes

Value 1	Value 2	*p*	*r*
No change	One range up	.003**	-.62
No change	Two ranges up	<.001***	-.62
No change	One range down	.018*	-.62
No change	Two ranges down	<.001***	-.61
One range up	One range down	1	-.06
Two ranges up	Two ranges down	1	-.16
One range up	Two ranges up	.003**	-.61
One range down	Two ranges down	.018*	-.55

Note. Increment sizes: -2 (two ranges down), -1 (one range down), 0 (no change),+1 (one range up), +2 (two ranges up). P-values were calculated by a post hoc test after Conover (1999). Bonferroni adjustment method was used; *p<.05, **p<.01, ***p< .001.

larly difficult to identify SpO_2 ranges above the center, detection rates were compared between SpO_2 ranges above and below the center range. After examination of the corresponding qq-plots, a nonparametric test was chosen, as the data points did not form a straight line. The Wilcoxon signed rank test indicated, that participants detection rates were lower above the center (*Mdn* = 78%) than below the center (*Mdn* = 94%) of SpO_2 saturation ranges (*p* = .031, *r* = -.62).

In addition to that, it was of particular interest, if the Shepard tone was a useful choice to convey information about current SpO_2 being below or above the center and the current direction of movement of SpO_2. Of all 720 answers given there was only one case, where a participant mixed up the corresponding SpO_2 ranges below and above the center range. In three cases there was a false evaluation of the direction of SpO_2 movement and in seven cases a change of the SpO_2 range was not recognized. Interestingly all these mistakes were made by one participant. Only participant 3 had a detection rate below 80% (96 of 120 answers). This participant accounted for approximately 37% (40 of 109 incorrect answers) of all falsely identified SpO_2 ranges. Already in the training session participant 3 had together with participant 6 the highest occurring error rate. Overall, participant 3 performed distinctly worse than all other participants. About 6% (about 7 of 120 answers) of the answers of all participants were false, due to an inside/outside error. They accounted for around 39% (43 of 109 incorrect answers) of all incorrect answers. Approximately 84% (36 of 43 inside/outside errors) of all inside/outside errors occurred due to a confusion between range 1 and 2 and around 5% (2 of 43 inside/outside errors) due to a confusion between range 6 and 7. Participant 3 accounted for about 51% of all inside/outside errors. There was no observable training effect, as trial 1 (*Mdn* = 88%), and trial 4 (*Mdn* = 91%) did not differ significantly in their detection rates (*p* = .371, *r* = -.26). Moreover, there was no difference between the detection rates of trial 1 (*Mdn* = 88%) and 3 (*Mdn* = 93%), which indicated that there was no effect of the direction of SpO_2 movement (p = .418, *r* = -.23).

5. DISCUSSION

Overall the results of the listening test are very promising, as the six participants could differentiate seven ranges of SpO_2 saturation well above chance. Although participants received only a short training in advance, they were able to continuously track SpO_2 saturation in each of the four trials. Interestingly the detection rates of all SpO_2 ranges differed significantly from one another. Multiple post hoc comparisons revealed that participants performed better in identifying range 7 than range 1. A reason for this finding might be the design of the sonification. As described above, perceived roughness of the Shepard tone was increased, as soon as SpO_2 values were below the target range (90-95%). On the contrary the acoustic properties of the Shepard tone remained the same, after reaching the upper threshold of the target range. Thus, participants had to recognize the discontinuation of the background noise to detect deviations of SpO_2 above the target range. The fact that values below the target range have been identified more accurately than values above the target range is evidence, that a redundant coding improves detectability. It is possible that participants simply missed the onset or offset of the continuous background noise.

Although this did likely happen on both sides of the target range, transitions below 90% of SpO_2 could still be identified by recognizing the change of roughness of the Shepard tone alone. As the results indicate, participants had greater difficulties to identify SpO_2 ranges in the upper part of the sonification, meaning all SpO_2 ranges above the center range. It is therefore plausible, that participants perceived the sonification of SpO_2 above the center as more ambiguous than below the center. These results underline the importance of redundant coding, to make important thresholds more obvious to the user.

In addition, participants performed distinctly worse in identifying SpO_2 values in range 2 than in all other ranges except range 1. As stated above, the asymmetric design of the sonification probably accounted for participants greater difficulties to detect range 1 in comparison to range 7. This might have also affected the recognition of SpO_2 values in range 2. As participants had to continuously track SpO_2 values the correct identification of a SpO_2 range depended highly on the correct recognition of the previous SpO_2 range. Thus, an increased insecurity concerning range 1 most probably also affected the performance in range 2. Moreover, the detection rate varied as a function of the SpO_2 increment size. More precisely participants had greater difficulties in recognizing the correct change of SpO_2 ranges, if the SpO_2 value jumped two ranges up or down, than if it simply moved one range upwards or downwards respectively. If the preceding SpO_2 value happened to be in the same range, participants performed better than with a preceding SpO_2 value one or two ranges away. This finding might provide an additional explanation for the distinctly worse performance concerning SpO_2 range 2. SpO_2 values in range 2 and 6 were more often preceded by an SpO_2 value two ranges away, than any other SpO_2 range. In fact 50% of all preceding SpO_2 values of range 2 and 6 happened to be two ranges away, thus making it more difficult to identify the correct SpO_2 range. Nonetheless, only detection rates for range 2 were considerably lower than for all other ranges except for range 1. Therefore, it is likely that because of the specific design of the sonification as stated above, participants perceived a greater degree of ambiguity concerning range 1 and 2. As already mentioned, around 6% of all given answers were false, due to an inside/outside error, whereas about 84% of all inside/outside errors occurred due to a confusion between range 1 and 2. This result underlines the already mentioned difficulty to discriminate range 1 and 2. Only in two cases there was a confusion between range 6 and 7, whereas these mistakes likely occurred as an aftereffect. The design of the sonification consequently proved to be useful to inform the listener about SpO_2 being inside or below the target range. On the downside, it appeared to be more difficult for the participants to differentiate between SpO_2 values being inside or above the target range, mainly due to a confusion between range 1 and 2.

The Shepard tone proved to be a useful choice to inform the listener about being below or above the center range, the overall direction of current SpO_2 movement and about deviations outside a critical target range. As already mentioned in the results, only participant 3 made mistakes that disagree with this conclusion. Interestingly participant 3 accounted for around 51% of all inside/outside errors and for about 37% of all falsely identified SpO_2 ranges. Apart from possible differences in individual abilities, the specific design of the listening experiment might contribute to such a distinctly worse performance. As described in the method section, participants had to continuously track SpO_2 values, which were changing every second pulse for 30 times in each block. Therefore, the listening test was highly susceptible to aftereffects. For example, if a single SpO_2 value was missed during the listening test, all subsequent ticks made in the corresponding table were shifted one column to the left. Especially if a SpO_2 value was missed or falsely added at the beginning of a trial this could lead to considerably lower detection rates. This is most probably the reason for such huge performance differences between participant 3 and all the other participants.

Limitations and Prospects

In total six participants took part in the listening test, whereby the sample consisted of students and staff of the Institute of Systematic Musicology at the University of Hamburg. In a subsequent study it would be desirable to have a larger sample, including participants without a musical background. There was no control group and any findings need further corroboration. Moreover, clinicians might interpret the sonification differently, because of a broader medical background knowledge. Also the setting of the listening test differed from a clinical environment, especially as there was little background noise and participants could concentrate solely on listening to the SpO_2 sonification. As for example an anesthesiologist has to divide his attention across different tasks, Paterson, Sanderson, Paterson, Liu and Loeb (2016) tested effects of a secondary task on identification of SpO_2 ranges using an enhanced sonification of the pulse oximeter [27]. Performances for SpO_2 range identification deteriorated more for a LogLinear sonification than for the enhanced sonification of the pulse oximeter, although the difference did not reach significance [27]. This way the applicability of an enhanced sonification of the pulse oximeter can be evaluated under more realistic conditions.

As described above, SpO_2 values for the listening test were generated by using a sine function. It was assumed that a smooth function would provide a more realistic change of SpO_2 over time, but this needs the evaluation of a clinically trained person. By using a sine function, conditions were not identical for each SpO_2 range, as for example the average distance to the previous value differed as a function of the SpO_2 range. This might lead to misleading conclusions, when the sonification is evaluated in terms of each single SpO_2 range. The design of the listening test was very susceptible to aftereffects. As already discussed above, these kind of mistakes probably accounted for a considerable percentage of all mistakes made by participant 3. Therefore, a different design for the listening test might be helpful to prevent bias caused by aftereffects.

The results indicate, that participants found it particularly difficult to identify SpO_2 range 1 and 2. As discussed above, SpO_2 values above 95% could only be identified by the discontinuation of a continuous background noise, in contrast to range 7, where the perceived roughness of the Shepard tone was increased. Therefore, it might be beneficial to increase the perceived roughness for SpO_2 values above 95% as well. This might also contribute to a better detection rate of SpO_2 values in range 2. Alternatively beating could be applied as suggested in [22].

The proposed sonification of the pulse oximeter could be extended to nine different SpO_2 ranges by implementing two levels of roughness above and below the target range. This way clinicians could differentiate between urgent and less urgent deviations of SpO_2 from the target range. This would be similar to the enhanced sonification of the pulse oximeter by Deschamps et al. (2016), where four different SpO_2 ranges outside the target range were sonified by adding two levels of tremolo to a LogLinear pulse oximeter [19]. However, the need of such a fine grained subdivision (nine different ranges) in the case of oxygen saturation monitoring of neonates needs to be evaluated by a clinically trained person. Furthermore, the sonification principle is designed to be continuous. It would be interesting to see how well the SpO_2 value could be interpreted on a continuous scale.

6. REFERENCES

[1] Hinckfuss, K., Sanderson, P., Loeb, R. G., Liley, H. G., & Liu, D. (2016). Novel Pulse Oximetry Sonifications for Neonatal Oxygen Saturation Monitoring: A Laboratory Study. *Human Factors, 58(2)*, 344-359. https://doi.org/10.1177/0018720815617406

[2] Ziemer T., Black D., & Schultheis, H. (2017). Psychoacoustic sonification design for navigation in surgical interventions. *173rd Meeting of Acoustical Society of America and 8th Forum Acusticum*, Boston, Massachusetts. https://doi.org/10.1121/2.0000557

[3] Vickers, P. (2011). Sonification for Process Monitoring. In T. Hermann, A. Hunt, & J. G. Neuhoff (Ed.), *The Sonification Handbook* (pp. 455-491). Berlin: Logos Publishing House.

[4] Edworthy, J. (2013). Medical audible alarms: a review. *Journal of the American Medical Informatics Association, 20(3)*, 584-589. https://doi.org/10.1136/amiajnl-2012-001061

[5] Sanderson, P. M., Watson, M. O., & Russell, W. J. (2005). Advanced Patient Monitoring Displays: Tools for Continuous Informing. *Anesthesia & Analgesia, 101(1)*, 161-168. https://doi.org/10.1213/01.ANE0000154080.67496.AE

[6] Sanderson, P. M., Liu, D., & Jenkins, S. A. (2009). Auditory displays in anesthesiology. *Current Opinion in Anaesthesiology, 22(6)*, 788-795. https://doi.org/10.1097/ACO.0b013e3283326a2f

[7] Watson, M., & Sanderson, P. (2004). Sonification Supports Eyes-Free Respiratory Monitoring and Task Time-Sharing. *Human Factors, 46(3)*, 497-517. https://doi.org/10.1518/hfes.46.3.497.50401

[8] Ruiz, T. L., Trzaski, J. M., Sink, D. W., & Hagadorn, J. I. (2014). Transcribed oxygen saturation vs oximeter recordings in very low birth weight infants. *Journal of Perinatology, 34(2)*, 130-135. https://doi.org/10.1038/jp.2013.157

[9] Sjberg, F., & Singer, M. (2013). The medical use of oxygen: a time for critical reappraisal. *Journal of Internal Medicine, 274(6)*, 505-528. https://doi.org/10.1111/joim.12139

[10] Habre, W., & Petk, F. (2014). Perioperative use of oxygen: variabilities across age. *British Journal of Anaesthesia, 113*, ii26-ii36. https://doi.org/10.1093/bja/aeu380

[11] Saugstad, O. D. (2005). Oxidative Stress in the Newborn - A 30-Year Perspective. *Neonatology, 88(3)*, 228-236. https://doi.org/10.1159/000087586

[12] 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(1)*, 55-63. https://doi.org/10.1159/000356561

[13] Lim, K., Wheeler, K. I., Gale, T. J., Jackson, H. D., Kihlstrand, J. F., Sand, C., & Dargaville, P. A. (2014). Oxygen Saturation Targeting in Preterm Infants Receiving Continuous Positive Airway Pressure. *The Journal of Pediatrics, 164(4)*, 730-736. https://doi.org/10.1016/j.jpeds.2013.11.072

[14] Sink, D. W., Hope, S. A. E., & Hagadorn, J. I. (2011). Nurse:patient ratio and achievement of oxygen saturation goals in premature infants. *Archives of Disease in Childhood - Fetal and Neonatal Edition, 96(2)*, F93-F98. https://doi.org/10.1136/adc.2009.178616

[15] Schulte, G. T., & Block, F. E. (1992). Can people hear the pitch change on a variable-pitch pulse oximeter? *Journal of Clinical Monitoring, 8(3)*, 198-200. https://doi.org/10.1007/BF01616776

[16] Loeb, R. G., Brecknell, B., & Sanderson, P. M. (2016). The Sounds of Desaturation: A Survey of Commercial Pulse Oximeter Sonifications. *Anesthesia & Analgesia, 122(5)*, 1395-1403. https://doi.org/10.1213/ANE.0000000000001240

[17] Santamore, D. C., & Cleaver, T. G. (2003). The Sounds of Saturation. *Journal of Clinical Monitoring and Computing, 18(2)*, 89-92. https://doi.org/10.1023/B:JOCM.0000032698.47717.06

[18] Brown, Z., Edworthy, J., Sneyd, J. R., & Schlesinger, J. (2015). A comparison of linear and logarithmic auditory tones in pulse oximeters. *Applied Ergonomics, 51*, 350-357. https://doi.org/10.1016/j.apergo.2015.06.006

[19] Deschamps, M.-L., Sanderson, P. M., Hinckfuss, K., Browning, C., Loeb, R. G., Liley, H. G., & Liu, D. M. K. I. (2016). Improving the detectability of oxygen saturation level targets for preterm neonates: A laboratory test of tremolo and beacon sonifications. *Applied Ergonomics, 56*, 160-169. https://doi.org/10.1016/j.apergo.2016.03.013

[20] Paterson, E., Sanderson, P. M., Paterson, N. A. B., Liu, D., & Loeb, R. G. (2016). The effectiveness of pulse oximetry sonification enhanced with tremolo and brightness for distinguishing clinically important oxygen saturation ranges: a laboratory study. *Anaesthesia, 71(5)*, 565-572. https://doi.org/10.1111/anae.13424

[21] Shepard, R. N. (1964). Circularity in Judgments of Relative Pitch. *The Journal of the Acoustical Society of America, 36(12)*, 2346-2353. https://doi.org/10.1121/1.1919362

[22] Ziemer, T., & Schultheis, H. (2018). Psychoacoustic auditory display for navigation: an auditory assistance system for spatial orientation tasks. *Journal on Multimodal User Interfaces*. http://doi.org/10.1007/s12193-018-0282-2

[23] Ziemer, T. & Schultheis, H. (2018). A Psychoacoustic Auditory Display for Navigation. *24th International Conference on Auditory Display (ICAD 2018)*, 136-144. http://doi.org/10.21785/icad2018.007

[24] Ziemer, T., Schultheis, H., Black, D., & Kikinis, R. (2018). Psychoacoustical Interactive Sonification for Short-Range Navigation. *Acta Acustica United With Acustica, 104(6)*, 1075-1093. http://doi.org/10.3813/AAA.919273

[25] Ferguson, S., Cabrera, D., Beilharz, K., & Song, H. J. (2006). Using psychoacoustical models for information sonification. *Proceedings of the 12th International Conference on Auditory Display*, London, UK. http://hdl.handle.net/1853/50694

[26] Conover, W.J. (1999) *Practical Nonparametric Statistical*. 3rd Edition, John Wiley & Sons Inc., New York, 428-433.

[27] Paterson, E., Sanderson, P., Paterson, N., Liu, D., & Loeb, R. (2016). The effect of a secondary task on identification accuracy of oxygen saturation ranges using an enhanced pulse oximetry sonification: A laboratory study. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 60(1)*, 628-632. https://doi.org/10.1177/1541931213601143