of the pulse tones was increased in the critical range. In contrast, a continuous change in the acoustic properties of the tones was achieved by gradually increasing tremolo in the low range as SpO_2 decreased. The brightness of the tones gradually increased in the critical range with a decreasing level of SpO_2, whereas tremolo remained constant. Results show that the accuracy of identifying absolute SpO_2 values was significantly better for a sonification with a stepwise change of acoustic properties (M = 53.7%, SD = 12.2%) than for a sonification with a continuous change of acoustic properties (M = 47.9%, SD = 11.7%). The participants also performed significantly better at identifying target transitions and SpO_2 ranges, if the acoustic properties of the tones were changed in a stepwise way. According to Collett et al. (2019), the results can be explained by the fact that stepwise changes in the tones provide clearer and more pronounced auditory cues for the current SpO_2 range and range transitions. This is an important finding for future sonifications. Nevertheless both approaches could be used complementary. A stepwise change of the pulse tones could be complemented by a continuous change of an additional psychoacoustic dimension.

1.3 The sonifications of the present study

The present work proposes a novel psychoacoustic sonification for the pulse oximeter. Compared to previous approaches the psychoacoustic sonification is designed to provide the listener with information about the current oxygen saturation at an even higher level of accuracy, while still clearly communicating important SpO_2 thresholds. This objective is addressed by combining a stepwise with a continuous change of psychoacoustic parameters. A stepwise approach is used to make important thresholds apparent to the listener, while a continuous change is supposed to provide a fine grained resolution of the oxygen saturation. In addition, a sonification based on the model of a conventional pulse oximeter is developed and used for the control condition in the listening test of the present study. Finally, an alternative sonification is being developed with the aim of taking the perceived pleasantness in the sound design into account. In the present study, the oxygen saturation is subdivided into seven ranges (see Figure 8). Ranges 2 to 6 lie within the optimal range of 90-95% SpO_2, while range 1 lies above and range 7 lies below the optimal range. In contrast to other studies (see chapter 1.2.6), the target area is divided into five further ranges. The psychoacoustic sonification design is intended to allow the differentiation of a larger number of SpO_2 ranges and thus to achieve a more accurate resolution of the target area. Here, it is assumed that it may be important for the clinician to be able to hear whether the oxygen saturation is in the center or more in the upper or lower part of the target range.

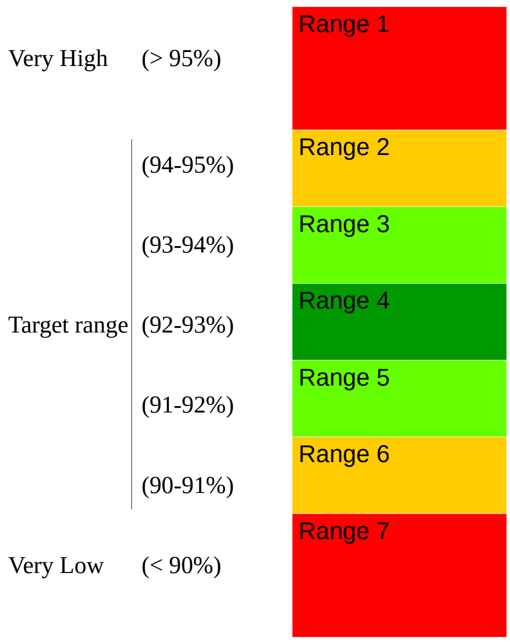

Figure 8. Division of the oxygen saturation into seven ranges. The optimum range lies between 90 - 95%. The corresponding saturation range is provided on the left of each colour-coded area.

1.3.1 A psychoacoustic sonification

1.3.1.1 Sonification design

A Shepard tone is used as the basis for the sonification. In a study by Ziemer and Schultheis (2018a), a Shepard tone was used successfully for a navigation task, where the speed of the rising or falling tone indicated the distance to a target in one of two dimensions (see also Ziemer et al., 2017; Ziemer & Schultheis, 2018b). Therefore it is considered that the Shepard tone might also be useful in estimating the distance of the instantaneous SpO_2 from a target region. The Shepard tone used in the present study contains six partials, whose frequency is defined as:

$$f_n = f_0 2^n \mathrm{Hz}, \tag{3}$$

where n = 0,...,N-1 and N = 6. The lowest partial f_0 is equal to 100 Hz. For SpO_2 values above the center range the frequency of the partials is rising and for SpO_2 values below the center range their frequency is falling. This way a simple binary coding is used to convey the information about SpO_2 being below or above the center of the target range. The rising or falling motion of the partials follows the following function:

$$f(\phi) = f_0 \sqrt[12]{2}^{\frac{\phi 12 N}{2\pi}}. \tag{4}$$

Two adjacent partials thus always have a distance of one octave. Here $\sqrt[12]{2}$ was chosen as the basis of the exponential function, since the sonification is based on a semitone scale of a twelve-tone equal temperament tuning system (see the following chapter). The phase ϕ increases for a rising motion of the partials continuously from 0 to 2 π such that within one period the frequency of the lowest partial increases from f_0 to f_N. The phase ϕ is defined as:

$$\phi(\theta, t) = arg[sin(2\pi\theta t)]. \qquad (5)$$

θ is a function of the distance between the current SpO_2 value and the center of the target range. The greater the distance, the faster the frequency of the partials changes. As in the study by Shepard (1964), the partials move through a bell-shaped curve, which weights the amplitude of the partials. The amplitude gradually approaches zero as the partials approach f_0 or f_N. To give the Shepard tone a pulse-like sound, a temporal envelope curve is used. Within one pulse the Shepard tone moves through a frequency interval, which is determined by its speed. With increasing speed of the Shepard tone, the frequency interval becomes larger. This way a continuous mapping between the SpO_2 distance to the center of the target range and the frequency interval within a pulse tone is provided. Since the phase of the Shepard tone provides no additional information about the current SpO_2 value, the Shepard tone is reset to the starting point of its period for each pulse.

To indicate that the current SpO_2 value is within the target range (90-95%) a background noise is presented, which sounds continuously and thus not only within the time window of each pulse. This way the most important information about the current oxygen saturation can be conveyed to the listener, while the cognitive workload is kept to a minimum. Here pink noise is used, which is further filtered by a comb filter (see chapter 2.4 for more details). Further information about the current SpO_2 value is provided by the movement of the Shepard tone. A rising or falling motion of the Shepard tone indicates SpO_2 values above and below the center range respectively. Within the center range the speed of the Shepard tone is set to 0, which results in a pulse of constant frequency. However, in the target range (90-95%), two further ranges above (range 2 and 3) and below (range 5 and 6) the center range are to be distinguished (see Figure 8). In order to identify these ranges the listener must rely on the size of the frequency interval that the Shepard tone passes through within one pulse. The speed of the Shepard tone and thus also the size of the frequency interval reach their maximum at 90- and 95% of SpO_2. Leaving the target range is indicated by the discontinuation of the background noise. In addition, roughness of the Shepard tone is increased, both above and below the target range.

1.3.1.2 Psychoacoustic background

In the following, the psychoacoustic background of the continuous change of the sonification as a function of the oxygen saturation is explained, whereby the Shepard tone is the central element of this translation. Furthermore, the psychoacoustic background of the stepwise changes of the sonification, more precisely the use of the background noise and the change in the roughness of the Shepard tone is discussed.

In a listening experiment Shepard (1964) could show a circularity in the judgement of relative pitch, which provides evidence that pitch can be separated in the two components height and tonality or chroma. The complex tones generated by Shepard preserved tonality but could not be ordered with respect to height. This finding suggests that the perceived pitch is not adequately represented by a rectilinear scale. In fact the tones could be arranged equidistantly around a circle in such a way, that the next clockwise tone is perceived higher in pitch while the next counterclockwise tone is perceived lower in pitch (Shepard, 1964). This way the Shepard tone can create the illusion of an infinitely rising or falling pitch, while the fundamental octave of the perceived pitch remains ambiguous (Ziemer et al., 2017). Although this illusion cannot be heard with the psychoacoustic sonification, the Shepard tone provides a number of advantages. Auditory brightness and loudness of the Shepard tone remain rather constant, so that these variables can be used for other input parameters (Ziemer et al., 2017). For the present sonification both brightness and loudness of the pulse tones do not provide the listener with any relevant information. Therefore, a change of brightness or loudness could be falsely interpreted as a change in the data underlying the sonification. As the Shepard tone allows to keep these parameters nearly constant, possible confusion is avoided.

Based on evidence on auditory scene analysis (see Bregman, 1990) it is very likely that the partials of the Shepard tone are integrated into one single stream. As explained by Ziemer et al. (2017), the carrier frequencies fuse because they are separated by octaves (see also chapter 1.1.9.3). The octave interval between the carrier frequencies is maintained even when their frequency is rising or falling. Here the principle of common fate is important. The frequency of the carriers changes in synchrony, facilitating their grouping by the auditory system (Ziemer et al., 2017). This principle applies also in case of the FM modulated Shepard tone – roughness of the Shepard tone was increased by using FM synthesis (see chapter 2.4) –, as the frequency of the carrier frequencies and of the generated sidebands vary in synchrony. Moreover, these sidebands do not form a harmonic spectrum, since each carrier frequency is modulated with the same modulation frequency (Ziemer et al., 2017). In case the sidebands would form a harmonic spectrum, they could get integrated into a second auditory stream. This would make it difficult for the listener to capture the information conveyed by the sonification, due to the difficulty of following several streams simultaneously (Ziemer et al., 2017). Last but not least the sonification of a pulse oximeter is presented in mono via a single loudspeaker. This facilitates the grouping into one stream, as the partials of the Shepard tone all originate from the same location in space (see chapter 1.1.9, Ziemer et al., 2017).

It is very likely that the individual pulse tones get integrated into one single stream by the auditory system. Again the common spatial location of the pulse tones facilitates the formation of a stream. Also the temporal proximity of the pulse tones helps the auditory system to group the individual pulses. Another important factor that influences the grouping into different streams is the similarity between the pulse tones. All pulse tones share the same carrier frequencies and duration and for every pulse tone the phase of the Shepard tone is set to 0. Moreover the loudness and the brightness of the Shepard tone remains fairly constant for every tone. The frequency interval the Shepard tone passes through varies continuously as a function of the SpO_2 level. As SpO_2 level is expected to change in a continuous manner without large and sudden jumps, the interval of the pulse tones should only change slowly and gradually, which facilitates integration according to the principle of continuity. A sudden change of the sound quality accurs, when the SpO_2 level is leaving the target range, as the roughness of the Shepard tone is increased. According to Bregman (1990) the factors that influence grouping processes of auditory scene analysis can reinforce and compete with each other. For the auditory system, the sudden change in roughness should be an indication of new event in the environment. In this case, the new event is that the current SpO_2 deviates from the optimum range. However, there are still many factors that support an integration of the pulse tones, such that they are perceived as a continuation of the previous pulse sounds within the optimum range. Last but not least, all of the factors mentioned here that facilitate the integration of the pulse tones into a single stream also facilitate the segregation between the pulse tones and other sounds in the environment.

The central parameter of the sonification is pitch. As described in chapter 1.3.1.1 the pitch changes depending on the current oxygen saturation. The listener must be able to tell from the direction of the pitch change whether the oxygen saturation is above or below the center range. In addition, the magnitude of the change provides information about the distance of the current oxygen saturation to the center (92.5%) of the target range. By intuition, it should be possible to understand that a rising motion of the Shepard tone indicates a deviation of SpO_2 above the center range and a falling motion of the Shepard tone indicates a deviation of SpO_2 below the center range. This principle is also similar to the sonification of a conventional pulse oximeter, where an increase in oxygen saturation is accompanied by an increase in the frequency of the pulse tones. A positive polarity is chosen between the size of the frequency interval that the Shepard tone passes through and the distance of the current SpO_2 value from the center range. Again, this relationship should be intuitively understandable, as a musical interval can be understood as the relative distance between two tones in terms of their frequencies. The selection of pitch as the central parameter of the sonification represents an optimal choice, as the human auditory system is very sensitive to changes in pitch (see chapter 1.1). The just noticeable variation in frequency (JNVF) is almost constant at a value of about 3.6 Hz for a modulation frequency of 4 Hz and carrier frequencies below 500 Hz (for a level of 60 phon). Above 500 Hz the JNFV is about 0.7% of the carrier frequency. This allows a sensation of very small changes in frequency, especially in the medium to high frequency range. At 50 Hz the JNVF is about 120 Cent, which is about a semitone in music. Cent is a logarithmic unit, which can be used to quantify musical intervals (see Ellis & Hipkins, 1884). At 400 Hz the JNVF is at about 16 Cent, much smaller than the interval of a semitone. The human auditory system is therefore able to sense very small changes in pitch. For the sensation of pitch other parameters like sound pressure level are also important (see chapter 1.1.3), but these are not varied in this sonification. Besides, it is generally unfavourable to use parameters that are not orthogonal to each other in the same sonification.

In the psychoacoustic sonification, the change in pitch depends on the speed at which the partials of the Shepard tone move through their bell shaped envelope curve. In the center range (range 4) the speed is set to 0, such that the pitch is constant in this range. Since the auditory system is very sensitive to changes in pitch, a constant pitch among other rising or falling tones should be easy to identify. In the center range the phase ϕ of the Shepard tones is set to 0°, so that $f_0 = 100$ Hz, $f_1 = 200$ Hz, $f_2 = 400$ Hz, etc.. For instance, a change in frequency of $f_2 = 400$ Hz by 3.6 Hz (see above) would correspond to a change in oxygen saturation of slightly above 0.01%, if one tracks back the processing chain described in chapter 2.4. In general, the change in frequency is 10 Cent for a change in oxygen saturation of 0.01%. Changes in oxygen saturation on this scale are probably not relevant from a practical point of view. In the present study, the participants are supposed to detect a change in oxygen saturation of 0.1% (see chapter 2). A change in oxygen saturation of 0.1% gives a change in frequency of 100 Cent, which is a semitone with a frequency ratio of $\sqrt[12]{2}$. For the lowest partial $f_0 = 100$ Hz a change in oxygen saturation of 0.1% results in a change of its frequency by about 6 Hz, which is above the JNVF as reported above. In general, it may be beneficial to assign a relevant SpO_2 change (here 0.1%) to a specific musical interval (here a semitone) (see Brown et al., 2015). This also takes into account the fact that the perception of pitch is logarithmic rather than linear in nature, as reflected in the mel scale (Fastl & Zwicker, p. 113). Also, Brown et al. (2015) showed that anaesthetists were able to estimate absolute SpO_2 values and differences in SpO_2 levels significantly more accurately with a logarithmic pitch scale than with a linear pitch scale. At a SpO_2 value of 90%, i.e. a deviation of 2.5% from the center of the target range, the interval comprises 2500 Cent, which is equivalent to two octaves and one semitone.

Scaling decisions are based on the aim to provide a fine grained resolution of the instantaneous SpO_2 value and to provide a meaningful change in frequency for a certain change in SpO_2. This is achieved in that even a slight change of 0.1% in oxygen saturation results in a perceptible change of the interval the Shepard tone passes through. A change in oxygen saturation by 0.1% always leads to an increase or decrease of the interval by a semitone.

The JNVF serves only as an orientation. The frequency modulation used by Fastl and Zwicker (2007) is obviously different from a rising or falling Shepard tone. For example, the JNVFs mentioned here refer to a modulation frequency of 4 Hz. At higher modulation frequencies, the JNVFs are significantly larger (see Fastl & Zwicker, 2007, p. 185). Similarly, the JNVF could depend on the speed of the Shepard tone, which is varied in the psychoacoustic sonification. Furthermore, the partials of the Shepard tone form a harmonic spectrum, i.e. they differ from individual sine tones. A change in frequency can be noticed via frequency changes of all partials. In any case the human auditory system has a very high frequency resolution. Up to 16 kHz the human auditory system can distinguish 640 adjacent pitches, which was calculated on the basis of just-noticeable frequency variations (Zwicker & Fastl, 2017, p. 184). Therefore, changes in the frequency of the partials of the Shepard tone should be easily detected by the listener. Perceiving chroma or tonality is limited to the audible frequency range below about 5000 Hz (Gelfand, 2009, p. 223). It is important that the listener is able to hear a change in chroma, as a change of SpO_2 of 0.1% corresponds to a fixed musical interval. The partials of the Shepard tone should therefore always sufficiently cover the frequency range below 5000 Hz. This is the case with the present sonification, since the highest partial has a frequency of 3200 Hz.

The sonification principle is designed to be continuous. A continuous increase in oxygen saturation results in a continuous increase in the frequency interval every pulse goes through. In contrast to the sonification of conventional pulse oximeters, this sonification provides a reference value. The Shepard tone is reset to the starting point of its period for each pulse tone. This starting point provides a reference for the listener to estimate the size of the interval. As described in chapter 1.3.1.1 a continuous background noise can be heard within the target range. Here pink noise is used, because it is more pleasant to hear than white noise. Pink noise has been used, for example, to improve sleep quality (Papalambros et al., 2017). An increased pleasantness of the background noise is particularly important, if patients or staff are exposed to the sounds of the sonification over a long period of time. The pink noise was further filtered by a comb filter to create a more artificial sound. This is to better distinguish the noise from other noise sources, such as respiratory machines or air conditioning systems. The use of the comb filter increases the probability of producing a distinguishable timbre compared to other background noises. Besides the spatial information – the noise originates from the loudspeaker of the pulse oximeter and not from any other machines – a different timbre provides an additional cue that allows a stream segregation between the pink noise of this sonification and other possible background noise. Furthermore, masking of the sonification by other noise sources becomes more difficult.

It is important that for critical SpO_2 values the sonification conveys the urgency to act. The roughness of the Shepard tone is increased by FM synthesis below 90 and above 95% of SpO_2. The number and amplitudes of sidebands generated by FM synthesis depend on the modulation depth (see chapter 2.4). Increasing the modulation depth increases "the number and amplitudes of inharmonic frequency components" and thus the perceived roughness and the perceived inharmonicity and noisiness (Ziemer et al., 2017, p. 6). An increased inharmonicity is generally perceived as more urgent (Edworthy, 2003). The sonification thus indicates a higher need for action for SpO_2 values outside the target range. As described in chapter 2.4 a modulation frequency of 120 Hz and a modulation depth of 100 Hz is used. For a pure tone with a center frequency of 1500 Hz, a sound pressure level of 70 dB_{SPL} and a frequency deviation of $\pm$ 700 Hz a maximum roughness can be achieved for a modulation frequency of about 70 Hz (Fastl & Zwicker, 2007, p. 260). Roughness also increases approximately linearly with the logarithm of the frequency deviation (Δf). Thus, larger values of Δf and a lower modulation frequency could achieve much higher levels of roughness. On the one hand, by further increasing the roughness for SpO_2 values outside the target range, it would be even easier to distinguish the two levels of roughness used in this sonification. On the other hand, an increasing inharmonicity and noisiness leads to a lower pitch strength (see chapter 1.1.3). This could negatively affect the perception of pitch change of the Shepard tone. However the direction of pitch change is essential to decide whether the current SpO_2 value is above or below the target range.

In order to ensure that the listener still reliably detects SpO_2 values outside the target range, a redundant coding is chosen. The listener can hear SpO_2 deviations from the target range by either detecting an increased level of roughness or the discontinuation of the background noise. A redundant coding can increase the robustness of an auditory display, as it can make important thresholds more obvious to the user (Ferguson et al., 2006). By changing the acoustic properties of the pulse tones continuously within the target range the listeners attention is not attracted in an unfavourable manner, allowing him/her to engage in additional tasks. When the SpO_2 value leaves the target range the sound quality of the sonification shows a sudden change, as both the roughness of the Shepard tone is increased and the background noise is not present anymore. This way the attention of the listener is directed to the sonification in the case of critical SpO_2 values. Last but not least the envelope of the Shepard tone is symmetrical around its maximum. The maximum in this sonification is at 800 Hz, thus avoiding excess of high frequency energy. This also avoids a strong sensation of sharpness, which would reduce the sensory pleasantness of the sonification (see Fastl & Zwicker, 2007, p. 243).

1.3.2 A conventional sonification

For the control condition of the present study a sonification is used, which is based on a conventional pulse oximeter. Here the model MX800 from the manufacturer Philips is chosen. The model MX800 is chosen because it uses a logarithmic mapping between SpO_2 and frequency, which is advantageous compared to a linear mapping for monitoring oxygen saturation (see e.g. Brown et al., 2015). The fundamental frequency of the MX800 is at 662 Hz for an oxygen saturation of 100%, whereas the frequency remains constant within one pulse beat. With each 1% decrease in oxygen saturation, the frequency decreases by 2.9% (Loeb et al., 2016). In the present study, the fundamental frequency is varied as a function of oxygen saturation as follows:

$$y[\mathrm{Hz}] = 662\left(0.971^{x}\right), \tag{6}$$

where y is the fundamental frequency and x = 100 - s, with s being equal to the current SpO_2 value. For an oxygen saturation of 80% this results in a fundamental frequency of about 367 Hz. According to a spectrogram created by Loeb et al. (2016), a pulse tone of the MX800 model consists of the fundamental frequency and four higher harmonics. Therefore, four higher harmonics are used in the present study, with their amplitude decreasing with increasing frequency. Furthermore, an alarm is added for SpO_2 values outside the optimal range, although Loeb et al. (2016) report no alarm for the MX800 model. The sound design of the alarm is based on a study by Paterson et al. (2019), where a similar alarm was used for the control condition. Using an alarm for critical SpO_2 values is typical for the sonification of many conventional pulse oximeters. In intensive care units, the variable pitch is often deactivated to reduce the noise level so that only the alarm can be heard (Paterson et al., 2019). Paterson et al. (2019) also showed that subjects could identify SpO_2 ranges and absolute SpO_2 values more accurately when using a sonification with a variable pitch and an alarm, than when using a sonification with a variable pitch but no alarm. Therefore, by using a logarithmic mapping between frequency and SpO_2 and an alarm for critical SpO_2 values, the most promising variant of a conventional sonification is used. The alarm consists of three short tones, whereas each of the tones lasts about 400 ms and has a fundamental frequency of 530 Hz. The alarm sounds once, as soon as the oxygen saturation leaves the optimum range, i.e. if the oxygen saturation exceeds a value of 95% or falls below a value of 90%. At a fundamental frequency of 530 Hz the alarm can be clearly distinguished from the pulse tones of the sonification. At 95% SpO_2 the fundamental frequency of the pulse tones lies at about 571 Hz and at 90% SpO_2 at about 493 Hz. With a JND