



Article

To Beep or Not to Beep? Evaluating Modalities for Multimodal ICU Alarms

Vanessa Cobus *  and Wilko Heuten

OFFIS—Institute for IT, Escherweg 2, 26122 Oldenburg, Germany; wilko.heuten@offis.de

* Correspondence: vanessa.cobus@offis.de; Tel.: +49-441-9722219

Received: 25 January 2019; Accepted: 6 March 2019; Published: 9 March 2019



Abstract: Technology plays a prominent role in intensive care units (ICU), with a variety of sensors monitoring both patients and devices. A serious problem exists, however, that can reduce the sensors' effectiveness. When important values exceed or fall below a certain threshold or sensors lose their signal, up to 350 alarms per patient a day are issued. These frequent alarms are audible in several locations on the ICU, resulting in a massive cognitive load for ICU nurses, as they must evaluate and acknowledge each alarm. "Alarm fatigue" sets in, a desensitization and delayed response time for alarms that can have severe consequences for patients and nurses. To counteract the acoustic load on ICUs, we designed and evaluated personal multimodal alarms for a wearable alarm system (WAS). The result was a lower response time and higher ratings on suitability and feasibility, as well as a lower annoyance level, compared to acoustic alarms. We find that multimodal alarms are a promising new approach to alert ICU nurses, reduce cognitive load, and avoid alarm fatigue.

Keywords: head-mounted display; wearable computing; ICU alarms; multimodal alarms

1. Introduction

Intensive care units (ICU) are equipped with highly sophisticated technical systems and devices for patient monitoring, respiratory and cardiac support, pain management, emergency resuscitation, and other life support measures. To ensure uninterrupted monitoring of patients with life-threatening illnesses and injuries or after major surgical procedures, these medical devices issue visual and acoustic alarms for various reasons. Each alarm has an individual sound, with the pitch and frequency of the sounds increasing with the priority of the alarm.

Research has shown that the number of alarms can rise to up to 350 per patient a day [1]. Since most ICUs foster a ubiquitously audible alarm distribution, each of these alarms sounds from a central working and monitoring station and, depending on the local alarm policy within the hospital, also from the patient's room—audible for every person in the ICU but difficult for the source to be identified [2]. Each alarm, therefore, needs to be evaluated and acknowledged by the caring nurse. The majority of the issued alarms, however, require no intervention from the other nurses and distract them from their current task. In addition to the unnecessarily increased cognitive workload, the high number of alarms results in desensitization and lower response time of healthcare professionals. This condition is called alarm fatigue [1].

Alarm fatigue may have severe consequences not only for the patients but also for the healthcare professionals. Staff involved in a negative patient event (e.g., due to a missed alarm) can suffer from

trauma called “second victim syndrome,” when the person who feels responsible for the failure suffers from guilt and depression, perhaps even occupational disability [3].

Wilken et al. [4] mentioned possible reasons for the high number of alarms and the resulting alarm fatigue, e.g., the use of default values for alarm thresholds. Default values can cause many unnecessary alarms by signaling that a vital data point exceeds the threshold but does not pose a threat for the patient. Other reasons may include inadequate use of electrodes or sensors, which can cause false alarms by e.g., falling off or disconnecting.

Using smart alarm-delay algorithms, daily electrode changes, or alarm management education already shows positive effects in reducing the number of unnecessary alarms [5]. The remaining unnecessary alarms, however, are still audible for every person in the ICU, distracting both patients and nurses.

Previous work in several domains has shown success using visual or tactile alerts [6–8]. A body-worn device, for example, can allow just the responsible nurse to receive the alarm. We assume that alarms can also be conveyed reliably via other modalities. In our work, we investigate the suitability of light, vibration, and sound via bone conduction speakers for critical care alarms using a wearable alarm system (WAS) in the form of a head-mounted display (HMD).

In prior design studies, we found suitable light and vibration patterns to represent three different urgency levels for patient alarms. We conducted a user study under task conditions that mimic concrete loads of nursing tasks with 12 nurses in an ICU lab. We compared different modalities for representative monitoring alarms on our HMD to the state-of-the-art: ubiquitous sound. Our results show that visual and tactile alarms presented on an HMD performed better than speakers regarding the factors of reaction time, error rate, perceived suitability, perceivability, and level of annoyance. Moreover, our prototype shows good usability and comfort. Based on this research, we propose a multimodal alarm design to present three different types of urgency.

This paper is organized as follows: The next section (Section 2) gives an overview of our general approach to reduce the alarm load on ICUs. Section 3 shows related work that helped us shape our own work. In Section 4, we describe our requirements for a multimodal wearable alarm system (WAS), based on a requirements analysis published in 2018 [9]. Sections 5 and 6 are based on previous publications [10,11], which motivated and documented the design of vibrotactile and peripheral light alarm patterns and their general applicability in this context. The results from these earlier studies suggested the patterns we used in the final evaluation. The evaluation and the results are presented in Section 7 of this work. Finally, in Section 8, we conclude and give examples for future work.

2. Approach

Our overall goal is to reduce the alarm load on ICUs by developing a wearable alarm system. This includes the design of:

1. a personal alarm distribution, which means that only the responsible nurses will be alerted by their patient alarms, and
2. a multimodal alarm signaling to reduce the cognitive workload by delivering different urgency levels with different modalities [12].

For our multimodal alarm design, we focus on the alarm classification of common patient monitoring systems (PMS): critical, uncritical, and technical alarms.

The alarms shall be conveyed with tactile and visual cues via an HMD. HMDs fulfill several safety and hygienic requirements common for systems in local hospitals, e.g., keeping the forearms free and enable hands-free interaction. Moreover, they can be easily integrated in the nursing workflow, which includes frequent moving between patient rooms, as they display information directly in the user’s (peripheral) vision.

Additionally, necessary information for the nurse to evaluate the alarm will be displayed. After that, the nurse will be able to interact touch-free with the system to respond adequately to the alarm.

We want to describe this in a future scenario: Jane is a nurse on a surgical ICU. In today's shift, she is responsible for three patients. While she is changing the medication for her first patient, an alarm for a second patient (John Doe) occurs on her WAS (see Figure 1). The system shows her that alarm is caused by a displaced sensor. Since she knows that this patient is in stable condition, she can silence this alarm for a specific time to finish her task. Afterwards, she walks to her second patient to adjust the sensor and acknowledge the alarm.

We imagine our concept to be integrated into data glasses such as Google Glass.

Especially for safety-critical environments, it is important to include the users in the design process of a new system. For that reason, we planned several user studies to design and evaluate modalities for critical care alarms. Figure 2 shows an overview of the studies described in this paper.

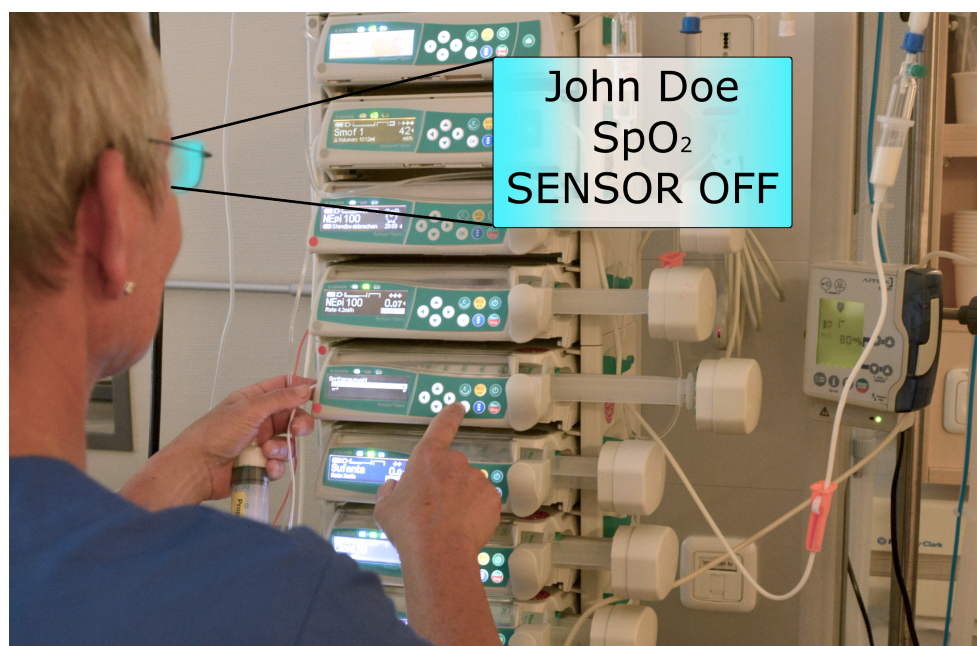


Figure 1. Future scenario.

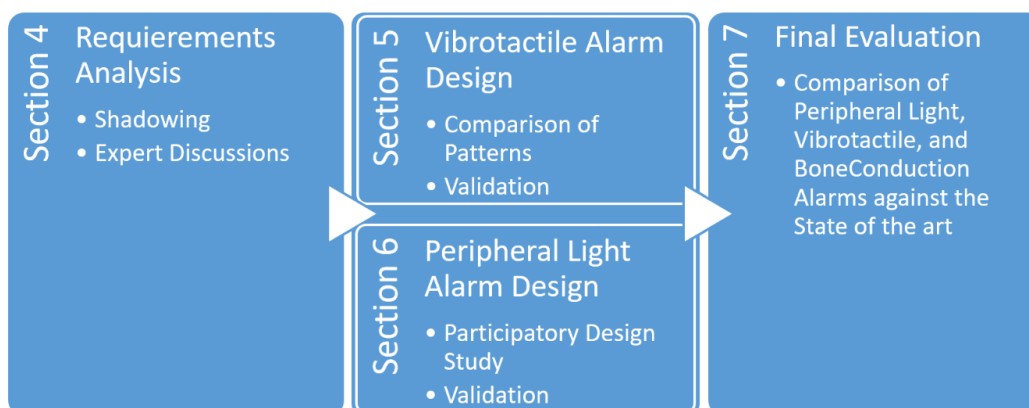


Figure 2. Planned user studies.

3. Related Work

In this section, we present related work that helped us shape our design solution. First, we present research that focused on a personal alert for nurses, which is the base of our research. The next part summarizes work that indicates the feasibility as well as suitability of HMDs for the medical domain. Finally, we show relevant research for each modality we explored for our WAS.

3.1. Alarm Distribution

In hospitals, especially ICUs, the high number of alarms is a well-known problem. Due to the common spatial distribution, alarms are loud, distracting, and hard to localize [2]. As the nursing workflow includes moving frequently between patient rooms and other locations, an alternative promising approach is to forward alarms directly to responsible healthcare providers. One example system for this approach is a pager. The portable device notifies nurses and, especially, physicians about relevant changes in the health status of their patients with vibrotactile and audible cues. Cvach et al. [13] developed a novel alarm escalation algorithm that distinguishes between a crisis and non-crisis condition of high priority alarms. If the first nurse does not react to a crisis alarm in a certain period of time, a second nurse will receive the alarm. Different from our alarm model, in the Cvach model, the charge nurse will be notified if there is no reaction within 60 s. For a non-crisis alarm, the algorithm causes a delay before the first escalation step. That algorithm was implemented as a secondary alarm notification system on pagers and tested for six months on two surgical progressive care units. The approach significantly decreased the mean alarm frequency and duration on the participating ICUs and shows the importance of distributed alerts. In 2014, Brander et al. developed a conceptual design of a mobile healthcare device to improve the information flow in hospitals by forwarding information to the responsible nurses [14]. The device takes the form of a nurse watch with three buttons. One red button is supposed to trigger an emergency alarm, a yellow one to call for assistance, and a white one to mute or forward alarms. Similar to pagers, the system uses vibrotactile and audible cues to notify the user. Nurses from three hospitals were involved during the whole development process. This work shows requirements for a mobile healthcare device based on a user-centered approach. Although portable devices such as pagers or a nurse watch can improve the distribution of alarms in hospitals, they have the disadvantage that they have to be put inside pockets. As nursing tasks are often stressful and physically demanding, the vibrotactile signal may go undetected [15]. With additional visual cues, a multimodal alert may reduce the cognitive workload and decrease the perceived alarm load for nurses [12].

In our research, we focus on exploring vibrotactile and peripheral visual alarms to convey different levels of urgency. Moreover, we aim to transmit the commonly used alarm tones via bone conduction speakers to avoid the distraction of patients and other nurses who are not responsible for that specific alarm. Finally, the multimodal alarms will be integrated into an HMD.

3.2. Head-Mounted Displays in the Medical Domain

The technology for smart glasses developed rapidly. Natalia Wrzesińska gave an overview of the use of smart glasses in healthcare in 2015 [16]. She points out that the majority of the current studies used *Google Glass*. One example is the work of Wolfgang Vorraber et al. [17]. Via a *Google Glass* application, they monitored patient vital data of a participant during radiological interventions. The results showed that using smart glasses improved the efficiency and awareness of the task at hand by reducing head and neck movements toward the patient monitor. Wrzesińska assumes that wearables, especially smart glasses, have the potential to improve effectiveness of healthcare and education, although there is still a need for more investigations.

Mentler et al. [18] sum up actual use cases for HMDs in healthcare and the usability challenges that arise for researchers and developers: interaction design, information visualization, and context of use. Moreover, they suggest including all stakeholders during the development.

A recent work by Pascale et al. [19] has shown that HMDs can support healthcare professionals maintaining awareness of their patients' health status without affecting their nursing task. Their HMD studies revealed that the participants reacted faster to alarms and answered situation awareness questions more accurately than those in the control condition.

These works showed that an HMD is a suitable form factor for a WAS to combine the delivery of visual and tactile stimuli in the medical domain.

3.3. Vibrotactile Information Representation in Medical Domains

The following works build the base of our vibrotactile alarm design, presented in Section 5. In 2005, Ng et al. [20] developed a vibrotactile prototype placed on the forearm to inform the user about the patient's heart rate during surgery. Their study results have shown that the prototype provides a much better result compared to the acoustic alarm scheme. In subsequent questioning of the subjects, it was noted that the prototype has more effectively caught the attention in a noisy environment. The comfort of the prototype, however, has been criticized for restricting freedom of movement due to the fixed elastic mounting strips for the vibration motors and the loose wiring.

Other work that focuses on tactile feedback on the wrists for medical applications was presented by Rossa et al. [21]. They present a wrist-worn prototype and multiple vibration patterns to guide a surgeon's hand for a cancer treatment procedure (brachytherapy). The study results showed that the vibration patterns could be successfully identified. Moreover, with a success rate of about 80%, the device could work in tandem with a needle steering algorithm to help surgeons in precision-demanding tasks.

The prototype developed by McLanders et al. [22] aims to keep the wearer informed about a patient's heart rate and oxygen saturation with vibrotactile cues on the upper arm. The results of their evaluation showed that the participants recognized over 90% of the changes in heart rate and oxygen saturation and that the comfort factor was rated as appropriately positive.

A vibrotactile belt for medical applications was introduced by Dosani et al. [23]. The device aims to support anesthesiologists in monitoring the vital data of patients with four vibration motors (front right, front left, back right, back left). Each motor represents a vital sign. The prototype was evaluated in the field during anesthesia application. Participants correctly detected 89.5% of the alarms. In a subsequent usability survey, participants gave positive feedback and reported that detecting the alarms became easier with increased wearing time. The presented work indicates that vibrotactile information representation is a promising approach for the medical context. The developed vibration patterns could convey information to healthcare professionals in a noiseless way.

The presented work indicates that vibrotactile information representation is a promising approach for the medical context. The developed vibration patterns could convey information to healthcare professionals in a noiseless way. Based on these works, we designed multiple vibration patterns to evaluate, if they are also suitable to represent different levels of urgency in a lab study with simulated nursing tasks (see Section 5).

3.4. Peripheral Light

Past research showed that peripheral light is a suitable modality to represent information within ambient systems. Chang et al. investigated the use of a noise-sensor light alarm in a newborn ICU [7]. The device in the form of a flower was installed on a central wall of the ICU. It lights up when the noise

level exceeds 65 dBA. The study results indicated that this peripheral light alarm has positive effects in reducing the environmental sound in the newborn ICU.

In 2014, Fortmann et al. [24] developed a wearable device to remind users to drink water. The device in the form of a bracelet indicates the elapsed time since the user's last water consumption using peripheral light and finally an additional vibrotactile notification as a reminder event. Study results strongly confirmed that drinking inputs were made by more than 90% of participants before the vibrotactile reminder event. This work demonstrates that light can be used to grab attention if positioned in the user's peripheral vision, e.g., on an HMD.

In 2015, Matvienko et al. [25] developed guidelines to map information to light patterns. They structured similarities in existing light encodings and, based on that, defined four information classes (*Progress*, *Status*, *Spatial*, and *Notification*) based on that. Matching those information classes, they defined everyday life scenarios (e.g., elapsing time as progress information, temperature as a status information or urgent/low-priority notifications) and did a two-part participatory design study. The focus was placed on color and brightness of light as well as LED position. The derived light patterns have been evaluated with a second group of participants.

This work deduces options on how to encode these information classes and derives nine design guidelines for ambient light systems; it is, however, limited to portable devices lying on a table.

3.5. Bone Conduction

Headphones are a well-known way to listen privately to audio. Headphones, however, are not integrable into the nursing workflow due to frequent necessary communications [9]. Different from the common mode of audio conveyance (air conduction), bone conduction is a tactile stimulus that leads signals to the inner ear through the bones of the skull [26]. This mode of audio signaling has the advantage that sounds can be conveyed to a single user and conveyed without blocking the user's ear canals.

McBride et al. [27] developed a map for interfaces using bone conduction. Their studies showed that bone conductive speakers should be placed in the area around the ears. Since our future work aims to use *Google Glass*, we chose the mastoid as the most suitable position for the speakers. In that way, the speakers can be easily integrated into the temples of glasses without being too cumbersome to wear.

4. Requirements

Our requirements analysis consisted mainly of two parts: a shadowing session in a 13-bed surgical ICU and two group discussions, one with four and one with three healthcare professionals from different hospitals. The concrete procedure was described in a prior publication [9]. Key questions for the sessions were: (1) of the nurses in the ICU, who should get which alarms and (2) how can the alarms be acknowledged or forwarded. As a further question, we asked how the nurses would like to be alerted. From the results, we could derive an alarm distribution and escalation model similar to the algorithm of Cvach et al. [13]:

- (1) Low-priority (uncritical) and technical alarms will be forwarded to the responsible nurse with a 60-s delay.
- (2) If there is no reaction within 60 s, the alarm will be forwarded to a second nurse.
- (3) If the second nurse does not react within 60 s, the alarm will be forwarded to the remaining nurses (see Figure 3 left).
- (4) High-priority (critical) alarms will be forwarded to the responsible nurse and the responsible physician immediately.
- (5) If there is no reaction within 60 s, the alarm will be forwarded to the remaining nurses (see Figure 3 right).

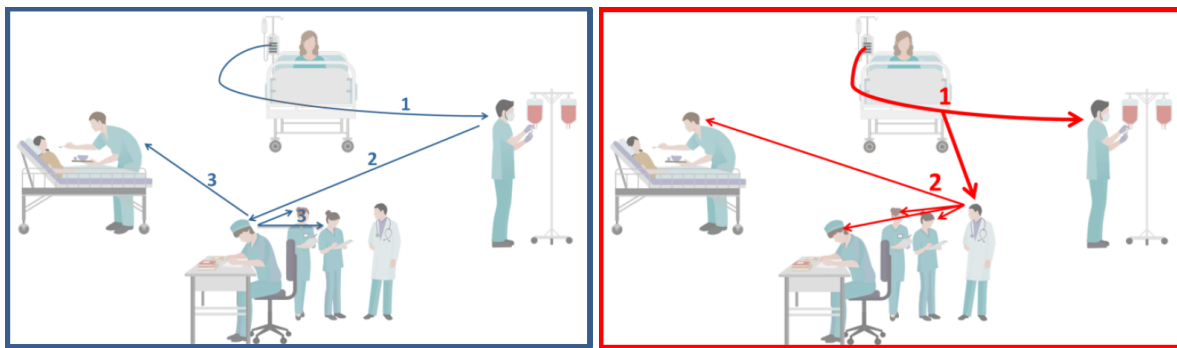


Figure 3. Alarm distribution model ((left) uncritical alarms; (right) critical alarms) [9].

The responsible nurse should have the option to acknowledge, to silent or to forward the alarm; to call for assistance and for an emergency call. An emergency call behaves like the high priority alarm and will be forwarded to the remaining nurses and the responsible physician.

A wearable alarm system (WAS) device that implements this algorithm should fulfill the following requirements to be integrable into the ICU workflow:

To comply with applicable hygiene and clothing standards, the WAS must not be applied to the hands or forearms. It should be shock and water resistant to withstand various circumstances in intensive care units. The nurse should be able to clean and wipe–disinfect the surface of the device to prevent germs or viruses from being transferred from one patient to another. It should be made of allergy-free and breathable material to avoid adverse reactions and sweating. For cost-saving reasons, the hardware components should be easy to detach so they can be used by multiple intensive care nurses. The system should be easily applicable and, moreover, resizable to fit different intensive care nurses. In addition, it should sit tight to the body so that it does not slip or get lost during work. The size of the device should be as small as possible. The WAS should reliably alert with three levels of urgency to distinguish between high priority (critical), low priority (uncritical), and technical alarms. The critical alarms should still be delivered acoustically. The alarms must be easily and quickly identifiable. Finally, the device must be easily integrable into the nursing workflow without having negative influence on the quality of nursing [9].

In the following sections, we present the implementation of our requirements as well as design studies and evaluations of unobtrusive wearable alarms.

5. Design of Vibrotactile Alarms

This section sums up the approach and the results from our prior work [10].

Vibrotactile feedback has been investigated for several years. The related work we presented in Section 3.3 provided vibration patterns evaluated for multiple use cases. In a two-part study, we investigated which patterns best suit representation of three alarm categories: technical, uncritical and critical alarms.

5.1. Apparatus

As mentioned in the work of Myles et al. [28], the head is a sensitive region for vibrotactile cues. Since we wanted to compare multiple vibration patterns, we built a first prototype for the upper arm to avoid a negative bias due to discomfort. It consists of an armlet with three relocatable vibration motors. The prototype can be seen in Figure 4. The detailed technical setup can be found in [10].

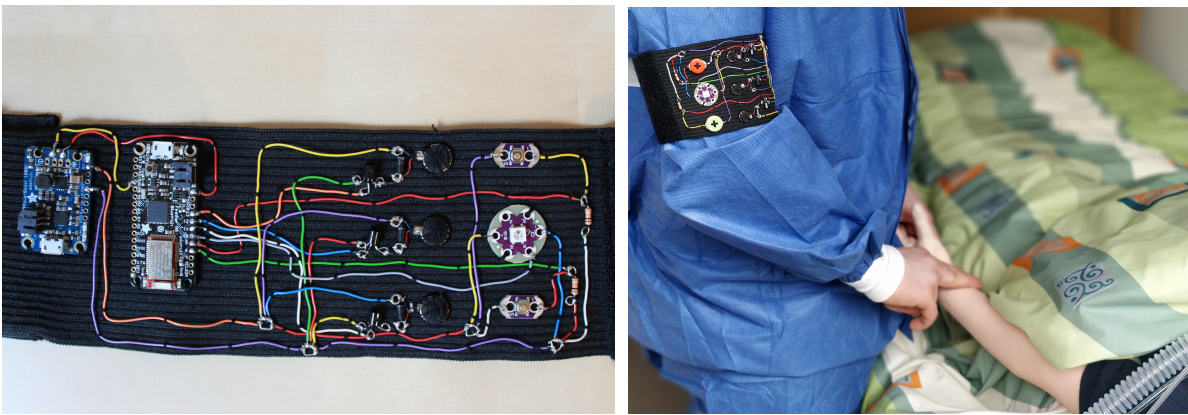


Figure 4. Vibrotactile alarm display as an armlet [10].

Based on literature [10], we implemented eight sets, each consisting of three vibration patterns to represent three different urgency levels. A visualization of each set is shown in Figure 5. The patterns are distinguished by vibration frequency, number of repeated vibrations, transition of the vibration intensity, and position of the vibrating motor. The urgency level increases from Pattern 1 to Pattern 3.

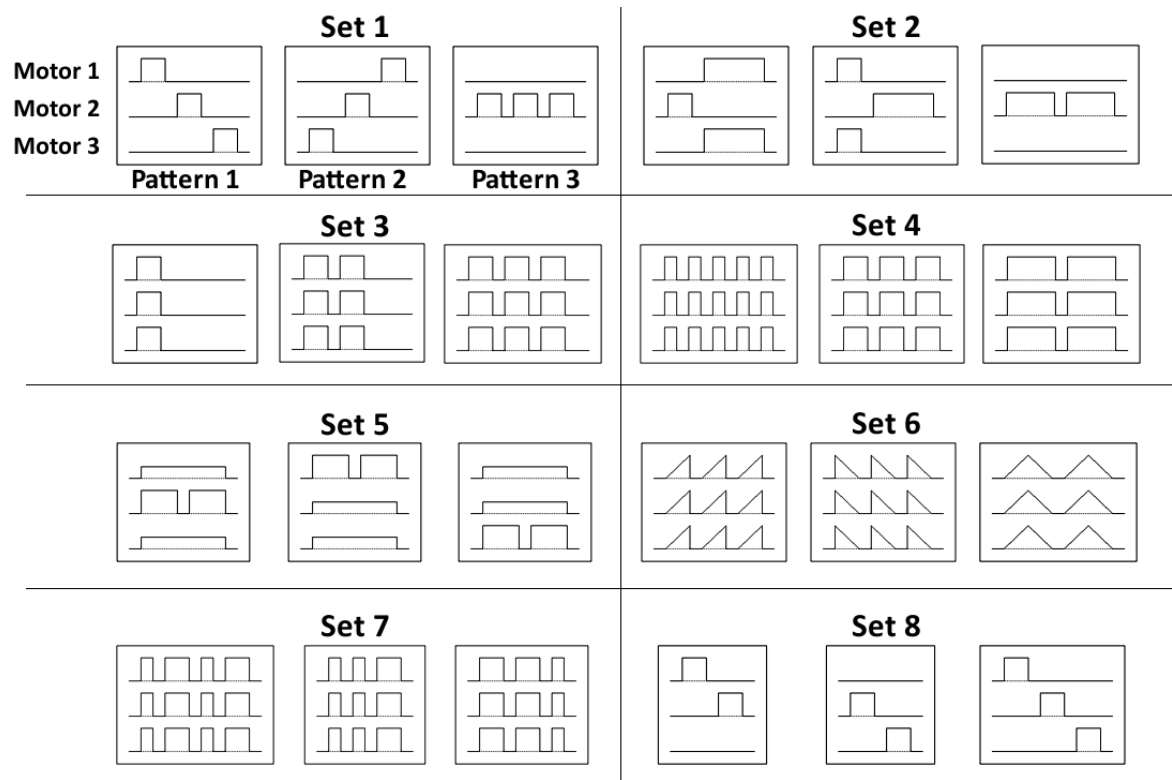


Figure 5. Overview of the implemented vibration pattern sets [10].

5.2. Evaluation

We divided our evaluation into two parts with a between-subject design. To find a suitable set of vibration patterns, we first evaluated the perception of the patterns regarding error rates, response time, learnability, distinctness, and diversity of urgency presentation. After that, we evaluated the overall prototype with actual nurses during cognitively and physically demanding tasks.

5.2.1. Evaluation of Vibration Patterns

We conducted a lab study to find out which set of vibration patterns best suits our wearable system. For this study, we invited 12 participants (five female), between 27 and 53 years old (average 35.9 years). Since there is no need for profession-specific knowledge to evaluate vibration patterns, we did not focus on the target group to recruit participants. The study itself consisted of two parts, an initial learning phase and the actual evaluation of vibration patterns. In the learning phase, each set was first explained to the participant on a print-out with a visual representation and then presented on the body-worn prototype. The order of the sets was randomized. After this phase, the vibration patterns were evaluated within the respective sets. Therefore, we sent each pattern of a set three times in randomized order via smartphone to the device. By sending the trigger, a timer was started. As soon as the participant was able to identify a pattern, s/he had to tap on the matching visual representation. Subsequently, we stopped the timer and thus the pattern remotely. We recorded the response time and the error rate for the identification of each pattern within a set. Additionally, the participant assessed the perceived perceptibility, distinctness, learnability and the diversity of urgencies for the vibration patterns in a five-point Likert scale.

Results

Figure 6 shows the average reaction times and error rates of the participants for each pattern and set. The Shapiro–Wilk test revealed a normal distribution of the data ($p < 0.01$). The fastest response times were measured for Set 4, which has a median of 1.19 s with a low dispersion. The response time of Set 3 was slightly but significantly worse (Wilcoxon signed-rank: $p < 0.05$) with a median of 1.59 s but with a similar dispersion. Generally, the reaction times of Set 3 and Set 4 showed significant differences in comparison with all other sets ($p < 0.05$).

Regarding the error rates, the data are also normally distributed ($p < 0.05$). Figure 6 shows that the vibration patterns in Set 3 were detected correctly in all cases. A similarly good recognition rate shows Set 4 with an error rate of 1.85%. However, there are no significant differences.

The subjective rating of perceptibility showed that all sets were perceived at least as good (median 4.0) or really good (median 5.0), whereas Sets 3, 4 and 7 had the lowest standard deviation. Regarding the distinctness of the vibration patterns of each set, Sets 3 and 4 were ranked best with a median of 5.

		Response Time				ERROR %		Perceptibility		Distinctness		Learnability		Urgency	
		MEDIAN		SD											
		Pattern	Set	Pattern	Set	Pattern	Set	MEDIAN	SD	MEDIAN	SD	MEDIAN	SD	MEDIAN	SD
Set 1	P1	2.87		1.17		11.11									
	P2	2.38	2.24	1.01	1.10	8.33	6.48	4.50	0.78	3.00	1.15	4.00	0.87	3.50	1.24
	P3	1.70		0.93		0.00									
Set 2	P1	2.46		1.61		8.33									
	P2	2.28	2.14	1.11	1.23	27.78	13.89	5.00	0.51	4.00	1.29	4.00	1.00	4.00	1.06
	P3	1.89		0.45		5.56									
Set 3	P1	1.42		0.34		0.00									
	P2	1.64	1.59	0.33	0.42	0.00	0.00	5.00	0.29	5.00	0.29	5.00	0.00	5.00	0.39
	P3	1.75		0.48		0.00									
Set 4	P1	0.88		0.24		0.00									
	P2	1.32	1.19	0.36	0.46	2.78	1.85	5.00	0.00	5.00	0.45	5.00	0.45	5.00	0.49
	P3	1.31		0.52		2.78									
Set 5	P1	3.61		2.42		38.89									
	P2	3.80	3.93	2.78	2.44	47.22	38.89	4.00	1.00	1.00	0.89	3.00	0.90	2.00	1.14
	P3	4.59		2.26		30.56									
Set 6	P1	2.29		1.51		11.11									
	P2	1.92	2.29	1.07	1.38	19.44	18.52	4.50	0.67	3.50	0.87	4.00	1.09	3.00	1.00
	P3	2.55		1.42		25.00									
Set 7	P1	2.19		1.77		2.78									
	P2	1.95	2.12	1.07	1.29	2.78	4.63	5.00	0.29	4.00	0.83	4.50	0.52	4.00	0.79
	P3	1.94		0.60		8.33									
Set 8	P1	3.09		1.25		8.33									
	P2	2.61	2.70	1.53	1.30	8.33	7.41	5.00	0.79	4.00	0.98	4.00	1.04	3.00	1.06
	P3	2.40		1.08		5.56									

Figure 6. Response time and error rates [10].

The participants rated Sets 3 and 4 also as easiest to learn (median 5) followed by Set 7 (median 4.5).

The perceptibility of diversity of urgencies was rated best for Sets 3 and 4 (median 5.0). The remaining sets got an average rating of 3.0 and 4.0, except Set 5 with an average rating of 3.0.

Overall, Sets 3 and 4 stand out with the best response time and the lowest error rate. Set 4 shows a slightly better response time than Set 3. Figure 7 shows that the reaction time of the Set 3 patterns increases from Patterns 1 to 3 because participants had to wait for the number of vibrations to identify the pattern. This may also be the reason for the better response time of Set 4. With Set 4, the pattern can be identified directly with the frequency of the vibrations, which results in a constantly low response time. In contrast, Set 3 shows a lower error rating with 0%. However, Set 4 shows an acceptable error rate as well (1.85%). Set 7 shows similar results in response time and error rates, but participants mentioned that Pattern 3 seems confusing and should be adjusted. The remaining vibration patterns can be excluded by a too high error rate for our safety-critical application. The order of the errors did not indicate a learning effect. Thus, Set 3 and Set 4 are implemented for the evaluation of the WAS.

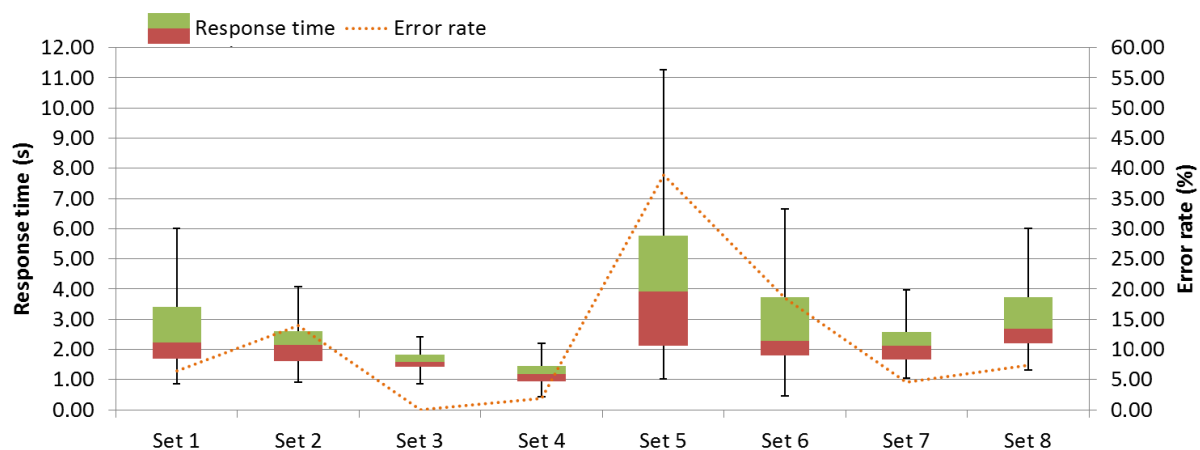


Figure 7. Response time and error rates [10].

5.2.2. Validation of the Vibrotactile Alarms

The results of the former study indicated that for Set 3 a lower error rate but longer duration to identify a pattern comes about than for Set 4. Since this study was conducted without a simulated load, we want to investigate if there are similar results with actual nurses during simulated nursing tasks. This led us to the following hypotheses:

Hypothes 1 (H1). *The response time of the participants to the vibration patterns of Set 3 is equal to that of Set 4.*

Hypothes 2 (H2). *The response time of the participants to the vibration patterns of Set 4 is lower than that of Set 3.*

Hypothes 3 (H3). *The error rate of the participants for Set 3 is lower than that for Set 4.*

To confirm those hypotheses, we conducted a within-subject lab study with 12 participants (10 female) with an average age of 37.6. All participants were fully trained nurses and three of them trained ICU nurses. The study was divided into a short training phase in which the prototype, its functions and the vibration patterns were introduced to the participant, followed by the actual study. The study design consists of two main conditions—one with a cognitive load and another with a physical load for the participant. This approach was supposed to represent two typical kinds of loads that humans in general and nurses and physicians in particular often bear. In the cognitive task, the participant sat at a table and had to filter out a given four-letter word in a series of similar four-letter words and mark it with a pen. In the physical task, the participant had to remove the duvet cover of a blanket and then cover it again. Both conditions were divided into two subconditions by presenting Set 3 and Set 4. The order of the conditions was counter-balanced. During the task, the participant wore the WAS on the upper arm and got alarms of the respective set with different urgencies delivered. The alarms were triggered remotely via a smartphone. As soon as an alarm was triggered, a timer started automatically. While focusing on the task, the participant had to name its urgency level. Furthermore, s/he had to press the red button for a critical alarm (Pattern 3) and the green button for an uncritical alarm (Pattern 1 or 2). We recorded the error rate of the pressed buttons as well as the response time from sending the trigger to the identification of a pattern. Each kind of alarm was triggered three times in a randomized order.

Results

In Figure 8, the response times and error rates for both conditions are visualized. The response times and error rates are both normally distributed ($p < 0.05$).

			Cognitive Load						Physical Load					
			Set 3			Set 4			Set 3			Set 4		
			Low	Middle	High	Low	Middle	High	Low	Middle	High	Low	Middle	High
Response Time	MEDIAN	Pattern	2.08	2.32	2.35	2.46	2.68	1.88	2.63	2.42	2.70	2.56	2.63	2.17
		Set	2.29			2.34			2.58			2.51		
	SD	Pattern	0.82	0.69	0.55	0.82	0.62	0.45	0.64	0.47	0.62	0.91	0.79	0.61
		Set	0.69			0.73			0.58			0.80		
	ERROR	Pattern	0.00	0.00	0.00	11.11	11.11	0.00	2.78	0.00	0.00	16.67	13.89	5.56
		Set	0.00			7.41			0.93			12.04		

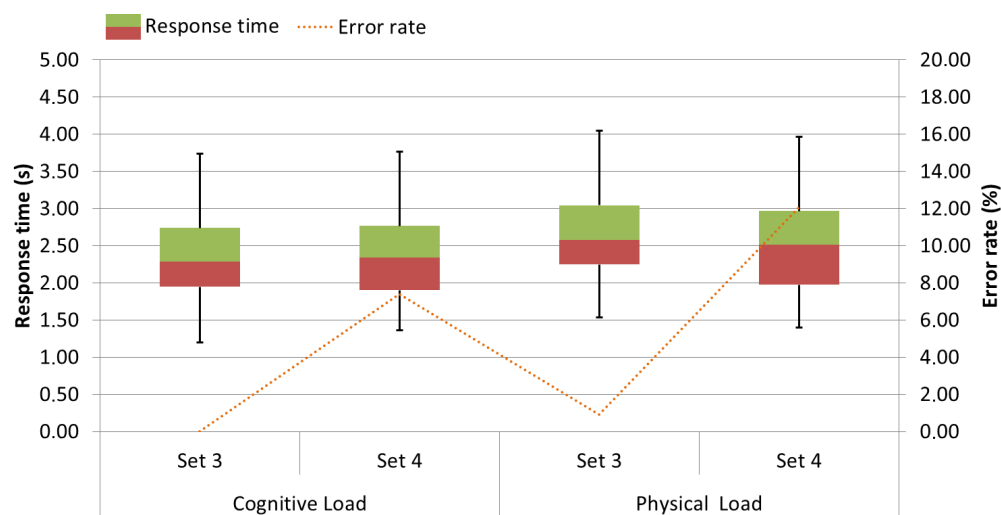


Figure 8. Results of the experiment [10].

Compared to the vibrating pattern study, in which the participants were able to concentrate fully on the vibration patterns, the response times to the two sets during the execution of both tasks are noticeably higher. During the cognitive load, the response time of both sets is almost equal with a median of 2.29 s for Set 3 and 2.34 s for Set 4.

During the physical load, the response times were generally slightly higher. Both sets showed similar response times with a mean of 2.58 s for Set 3 and 2.51 s for Set 4. However, Set 4 shows a higher dispersion in the response time. There were no significant differences.

Regarding the error rate, the results of Set 3 were similar to the first study. During the cognitive load, no mistakes were made. In the physical task, the error rate for Set 3 is 0.93%. In comparison to the first study, the error rate for Set 4 increased, with 7.4% during the cognitive load and 12.04% during the physical load. A Wilcoxon signed-rank test revealed that the error rate for Set 4 is significantly higher than for Set 3 ($p < 0.05$).

5.3. Discussion

The evaluation of the prototype with nurses showed positive results regarding the error rate and response time in identifying a vibrotactile pattern. With at least two sets of vibration patterns, we could

represent three different levels of urgency. From the study results, we learned that there is a higher error rate as well as a higher response time for both sets during physically demanding tasks.

Since there are no significant differences in the participants' response time for both sets, we have to reject = H2 and accept H1, respectively. Regarding the error rate, Set 3 shows significantly better results than Set 4, thus, we can accept H3 with $p < 0.05$. Since we aim to develop a WAS for an ICU, which means a safety-critical environment, the error rate for Set 4 is not acceptable. Therefore, the patterns of Set 3 will serve to represent alarms for our WAS.

6. Design of Peripheral Light Alarms

Many PMS highlight the source of the alarm (e.g., the relevant vital parameter) on the information display with the colors red (critical), yellow (uncritical) and blue (technical). However, the design space for light patterns contains more parameters than just the hue. Prior research has shown that other parameters, such as, e.g., the blinking frequency or brightness, affects the perceived urgency of the delivered information [25]. For that reason, we conducted a participatory design study to design light patterns that represent three levels of urgency, based on the given color mapping.

6.1. Apparatus

To design and evaluate peripheral light cues, we developed an HMD, based on safety glasses with a diffused peripheral LED display next to each eye. The prototype is shown in Figure 9. The detailed technical setup can be found in [11].

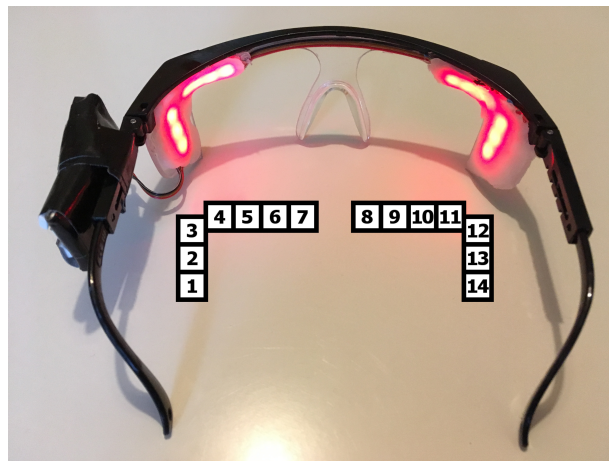


Figure 9. Head-mounted display and the numeration of LEDs [11].

6.2. Evaluation

We conducted a study to design alarm cues for the alarm categories “critical alarm”, “uncritical alarm” and “technical alarm”. We divided our study into two conditions, a design study and a validation of the patterns.

Due to their very demanding shift work, conducting studies with nurses as participants may present several issues for them, e.g., the time needed to participate in the study or the competing commitments for clinical practice.

Since this study took place in a preliminary stage of exploring light patterns to represent different urgent alarms, we conducted the first user studies outside the target group to finally take the findings from our work to nurses.

6.2.1. Participatory Design Study

For the design study, we invited ten participants (six female), between 18 and 33 years old, without a specific background. Each participant had a normal or rectified vision (through contact lenses). None of them were color blind. In the first condition, the participants were asked to design two urgent light patterns for the color red and two less urgent light patterns for the colors yellow and blue. We made these colors mandatory because they are already established in several ICUs for the different alarm categories.

We provided a laptop on which the light pattern was programmed via the *Arduino IDE*. To simplify the design of the light patterns, we predefined functions that let the participants adapt: (1) the brightness levels (from 0 to 255), (2) the brightness transition (stepwise/smooth), (3) the duration of the lighting/smooth transition, and (4) individual LEDs that should light up.

These functions, a description, and a scheme with the numeration of the LEDs on the prototype were placed on a table next to the study laptop, always visible for the participants. Every design was directly uploaded and shown on the prototype and, if necessary, corrected. The participants were asked to think aloud during the design process and to justify their design solutions afterward. In the end, they had to choose, for each color, which pattern they prefer.

Results

Since every participant developed individual light patterns, we derived the following similarities:

Stepwise Transition The light pattern includes a stepwise brightness transition.

Smooth Transition The light pattern includes smooth brightness gradient.

Different LED Positions The light pattern includes the use of different LED positions.

The frequency of the general use of each parameter is shown in Table 1. The number in parentheses states the frequency of the used parameter within the preferred patterns.

Table 1. Used parameters within all designed light patterns [11].

	Red	Yellow	Blue
Stepwise Transition	12 (4)	2 (1)	2 (0)
Smooth Transition	1 (1)	3 (1)	2 (1)
Stepwise + Smooth Transition	3 (3)	5 (2)	6 (5)
Different LED Positions	4 (2)	10 (6)	10 (4)

Regarding the preferred light patterns, the majority of the red patterns included a stepwise brightness transition or at least a combination with a smooth transition (e.g., a fading out). The majority of the yellow patterns included the use of different LED positions such as a chasing light or only the outer LEDs blinking. The blue patterns were mostly designed with a combination of stepwise and smooth transitions as well as with different LED positions.

During the thinking aloud process, the following statements have frequently been made: three participants stated that the lateral LEDs appear to be brighter than the top ones. Three other participants mentioned that the urgency was intrinsically encoded by the color. It was especially remarked that yellow is more urgent than blue (5), yellow and red are more urgent than blue (4) and that red is more urgent than yellow and blue (2). Nearly all participants (9) perceived higher blinking frequencies as more urgent, and five participants considered dimmed (“soft”) brightness modification as less urgent. Four participants referred to more LEDs switched on and to an increasing brightness as being more urgent, respectively. Finally, three participants stated that they had based their pattern designs on commonly known alarms.

The results indicate that the color blue appears less urgent than yellow or red and yellow appears generally less urgent than red. This complies with the general perception of colors. Furthermore, a lower urgency was represented with smooth brightness transitions. This has to be considered for the final design.

Deriving from the results, we implemented five light patterns for each color, which are shown in Figure 10. The patterns are distinguished by blinking frequency, brightness, brightness transition, and the position of the blinking LED.

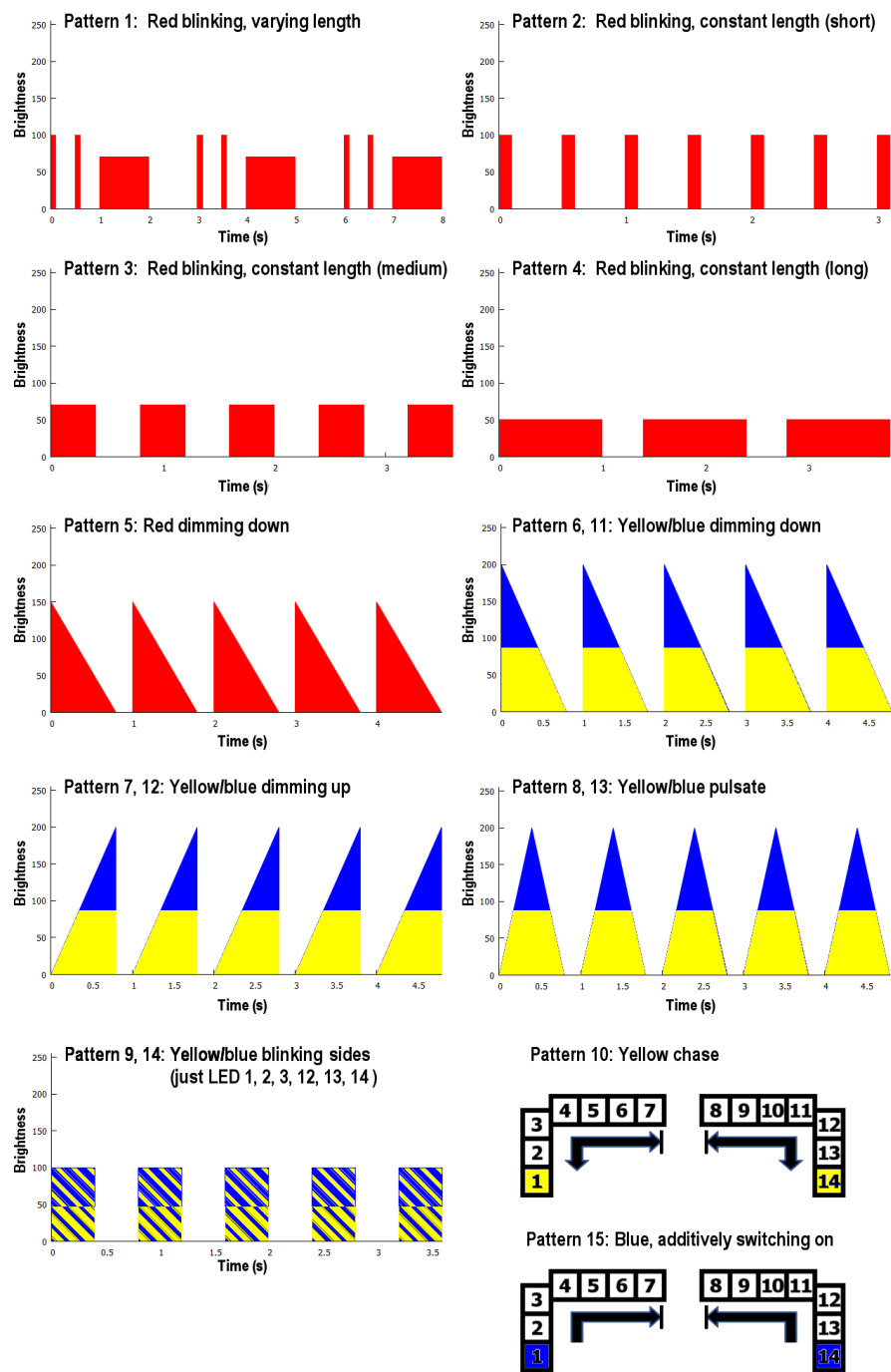


Figure 10. Overview of the implemented light patterns [11].

The frequent use of stepwise transitions for high priority alarms confirms the guidelines for urgent notifications of Matviienco et al. [25]. Therefore, we designed four light patterns using stepwise brightness transitions. They differ in the frequency of blinking and the brightness. One pattern is based on the preferences in the designed red patterns including a combination of stepwise and smooth transitions. Since yellow and blue patterns shall appear similarly urgent, we implemented similar patterns. Three of them include a combination of stepwise and smooth transition, one is a pulse, and one is a chasing light. Durations and brightness values are based on former pretests.

6.2.2. Validation of the Peripheral Light Alarms

In a further study, we wanted to evaluate which of the shown patterns is best suited for representing three different types of alarms. Moreover, we wanted to find out whether blue light patterns appear, generally, independent from the design of the pattern, less urgent than red or yellow. This study served for evaluating the implemented light patterns (see Figure 10) with regard to subjectively perceived urgency, comfort and distraction. Moreover, we wanted to derive at least one feasible light pattern for each alarm.

For this evaluation, we invited 20 participants (11 female), between 18 and 41 years old. Each participant had a normal or rectified vision (through contact lenses). None of them were color blind. During the user study, the 15 light patterns (see Figure 10) were shown to the participants on the HMD in intervals of one minute. Each pattern was repeated three times; the order of the patterns was randomized. To prevent the participants from getting tired or irritated, we split the study into three conditions with different precision-demanding tasks [29,30]. The tasks should also represent a load similar to that on ICUs (e.g., giving injections). The order of the conditions was counterbalanced.

The first task was a wire loop game in which the participant has to remove plastic items from cavities inside the patient with a pair of tweezers without touching the edges of that cavity [29]. If s/he touches an edge, the game board gives a visual and audible signal. In the second task, the participants had to play another wire loop game in which they must try to guide a wand along a wire without touching it. As soon as the wand touches the wire, a sound occurs (see Figure 11). In the third task, the participants had to refill syringes with exact predefined amounts of water. Between each condition the participant was given the ability to pause for a while.



Figure 11. Participant doing a precision task wearing the prototype [11].

Each time the participant noticed a light pattern, s/he was asked to rate it regarding its perceived level of urgency, pleasantness, and distraction from the performed task. This was done based on a five-point Likert scale. These factors were chosen to find a suitable light pattern that appears urgent for the user but not distracting from the actual task. Since the HMD should be worn during a whole shift, we also paid attention to the comfort factor of a light pattern. The participant should also state one first association s/he had regarding the pattern. The results were logged by the researcher to minimize the distraction caused by the rating process. To help the participant remember, the rating criteria and their Likert scales a print-out of them was placed in viewing distance. At the end of the study, the participant was asked to note their age and gender on the protocol. There was also space given for further annotations.

Results

Quantitative Results

In the following, we use *LP1–LP15* for Pattern 1–Pattern 15. We regard patterns rated with an average score higher than 3.5 as relevant. For the color red, all patterns except the constantly long blinking pattern (LP4) were perceived as urgent, with a range between 4.25 (SD = 0.67) for LP2 and 3.71 (SD = 0.98) for medium long blinking pattern (LP3). Of the yellow patterns LP6, LP7, LP8 were perceived as urgent but with no significant differences between the patterns. LP15, the additively on-switching LED, was the only blue light pattern perceived as urgent with a rating of 3.88 (SD = 0.95). There was a significance between LP 15 and all other blue patterns except LP11 ($p < 0.01$).

Regarding the distraction, LP1 (mean = 3.87, SD = 1.02), LP2 (mean = 3.5, SD = 0.9) and LP7 (mean = 3.8, SD = 1.05) were perceived as distracting. LP4 was significantly less distracting than LP1 and LP2 (both $p < 0.01$). Within the blue patterns, LP15 was significantly more distracting than LP12, LP13 and LP14 (all $p < 0.01$).

None of the red nor one for yellow light patterns were perceived as comfortable. However, the constantly long blinking pattern (LP4) was perceived as significantly more comfortable than LP1 ($p = 0.0031$) and LP2 ($p = 0.0052$) and LP9 is more comfortable than LP7 ($p = 0.0025$) and LP8 ($p = 0.0049$). All blue light patterns except LP15 were perceived as comfortable, with a range between 3.87 (SD = 1.11) for LP11 and 3.65 (SD = 1.04) for LP13. There were no significant differences. An overview of the results can be seen in Figure 12.

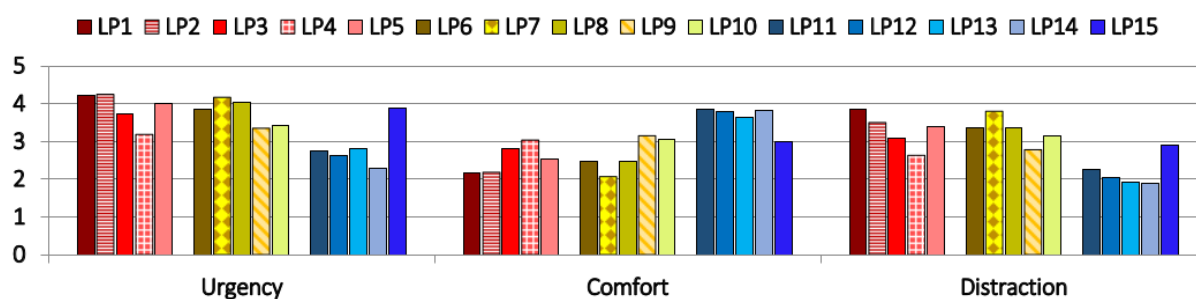


Figure 12. Overview of the results for each pattern [11].

Regarding the color groups, a Wilcoxon signed-rank test showed significant differences of the perceived urgency, distraction and comfort between blue and red or yellow patterns, overall $p < 0.01$, which means that blue light patterns are generally less urgent and distracting but more comfortable. Red and yellow in general showed no significant differences in any of the factors. Nevertheless, there

are combinations of red (LP1, LP3 or LP5) and yellow (LP9) patterns that show significant differences ($p < 0.01$). The results are visualized in Figure 13).

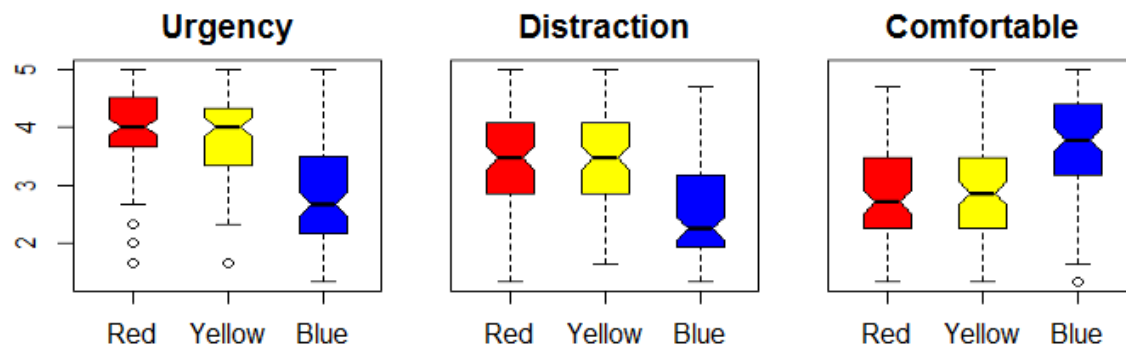


Figure 13. Summary of the perception between colors [11].

Pearson correlation test revealed a correlation between the perceived urgency and distraction ($r(298) = 0.70$, $p < 0.01$). Moreover, there is a negative correlation between urgency and comfort ($r(298) = -0.54$, $p < 0.01$).

Qualitative Results

After the presentation of each pattern, the participants were asked to mention an association with it. For red patterns, participants mentioned overall 74 times an association with “alarms”, “danger” or “emergencies” and 23 times an association with “warnings” or “errors”. All 20 participants associated LP2, and 18 participants LP1 and 15 LP5 with alarms. LP3, the constantly medium long blinking pattern, had 13 associations with “alarms” and 10 with warnings like “stop!”. LP5 was called “bright” or “dazzling” 10 times. Another association with the red patterns was “annoying”/“too long”, which was mentioned overall 30 times, evenly distributed.

The same association was made with the yellow patterns 21 times. The most prominent association with the yellow patterns was “bright”/“dazzling” with 76 mentions, which affected mainly LP6 and LP8 with 18 mentions and LP7 with 24 mentions. LP10 was associated with “party” or “fair” 19 times and with “confusing” nine times. The association with “alarms” was made 33 times, evenly distributed between LP6, LP7, LP8 and LP9.

The blue patterns were mostly associated with “alarms”, like a police blue light (74 mentions). Another association, mentioned 62 times, was “pleasant”. This was related to all blue patterns except LP15. This pattern was called “too long” or “hectic” 17 times.

6.3. Discussion

The results showed that the implemented blue light patterns appear overall less urgent than red or yellow ($p < 0.01$). Since blue patterns shall represent technical alarms and may indicate that a sensor does not measure the data reliably, ignoring a blue alarm due to an erroneously underrated urgency could lead to missing a critical incident. Thus, this result may indicate that light is not sufficient to represent an urgent blue alarm reliably. One option could be to extend this alarm by another sensory stimulus.

On the other hand, it is conceivable that nurses may perceive the blue pattern as urgent after a period of familiarization. This must be evaluated in a long-term study.

A correlation test revealed that the perceived urgency correlates with the perceived comfort of a light pattern ($r(298) = 0.70, p < 0.01$). This means for us that we have to compromise between those factors while choosing a light pattern for each alarm type.

Regarding all factors, the constantly short blinking pattern (LP2) is the most feasible pattern for red alarms. Even though it is the second most distracting light pattern, it appears most urgent. Red alarms require immediate reaction; accordingly, this alarm needs to grab attention. Adapting brightness and frequency, this alarm could become more comfortable. For the low priority alarm, we consider LP9, the constantly blinking lateral LEDs. As uncritical alarms are the most frequent alarms in ICU, we chose the most comfortable and less distracting light pattern for this alarm, which was still associated with alarms. As a technical alarm, we consider the constantly pulsating pattern, LP13, which is the second most urgent and also the second most comfortable pattern.

7. Experiment

We conducted an experiment to evaluate (1) the performance, suitability and feasibility of multimodal alarms compared to the state of the art (speakers) and (2) the usability and comfort rating of a multimodal head-mounted alarm display. The study was conducted under task conditions that mimic the load of care tasks.

7.1. Participants

For the study, we acquired 12 ICU nurses (seven female) from two different hospitals, between 22 and 51 ($M = 33.8, SD = 10.3$) years. Their years of experience in ICUs ranged from 3 months to 19 years ($M = 8.92$ years, $SD = 6.92$). Five of them had a corrected rectified vision through glasses, one of whom was blind in one eye. None of the participants were color blind.

7.2. Apparatus

Since we are in an early state of development, there are several safety factors that kept us from testing in the field. For that reason, we conducted the study in an unused ICU treatment room of a cooperating hospital, equipped with a patient bed, a desk, and a chair (see Figure 14). To create a realistic environment, we prepared a sound file as background noise containing talking and laughing people, clinking glass containers, ringing telephones and the sounds of a respirator. We explicitly excluded further alarm sounds of clinical devices.

Moreover, we prepared additional speakers in the corner of the room with a distance of ~ 1 m to the working place and ~ 1.8 m to the patient bed to simulate the state of the art using a notebook with the acoustic alarms as mp3-files. To present the multimodal alarms, we augmented our prototype from the light patterns study (see Figure 15). It consists of an *Adafruit Feather m0 WiFi* as micro controller combined with a *Music Maker FeatherWing* as amplifier, powered by a 3.7 V LiPo battery. In addition, 12 RGB LEDs are attached in the peripheral field of view (six on each side, three above, three next to the outsides of the eyes), diffused with *Gorilla Plastic*. We attached two ERM vibration motors ($\varnothing 10$ mm) on the temples behind the ear for vibrotactile cues. Two bone conduction speakers ($14 \text{ mm} \times 21.5 \text{ mm}$) were attached on the temples above the ear (one on each side) which convey acoustic alarms without blocking the ear channels. To ensure a firm hold for different head sizes, we attached an elastic band on the prototype.



Figure 14. Study setup in an intensive care treatment room.

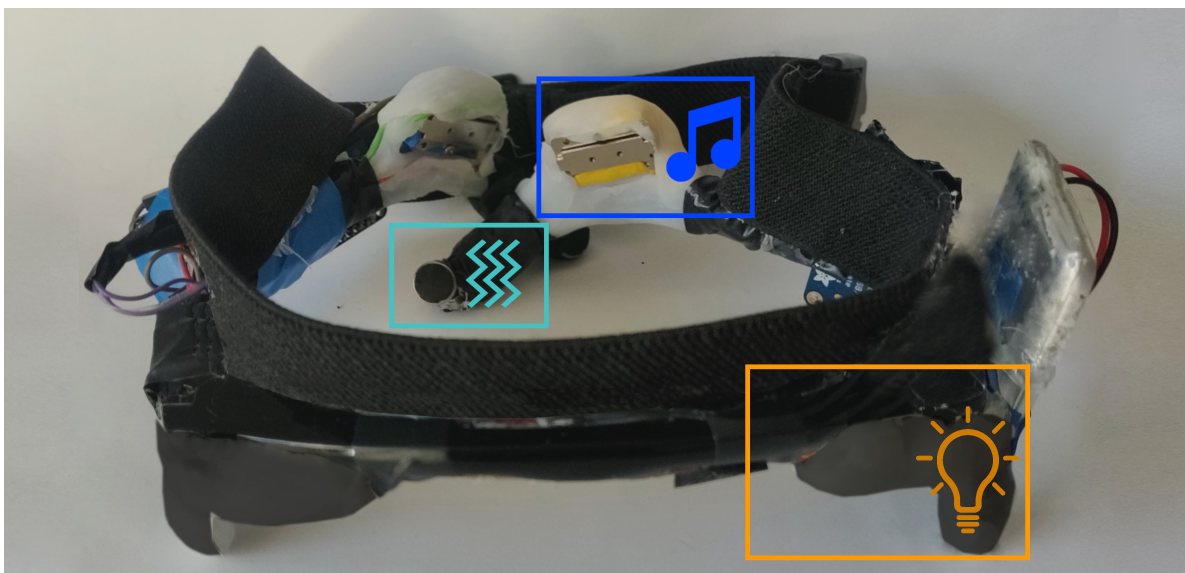


Figure 15. Multimodal head-mounted display for visual, vibrotactile and personal audible alarms.

We used the following alarm designs for our experimental conditions based on our previous studies described in Sections 5 and 6. The acoustic alarms are based on the sounds of a commercial patient monitoring system. *Light patterns:* To represent critical alarms, we used a red blinking. All LEDs light up for 0.1 s with a 0.4 s break in between and a brightness value of 100. For the uncritical alarm, we used a yellow blinking. Just the side LEDs light up for 0.4 s with a 0.4 s break in between and a brightness value of 100. The technical alarm is a blue pulsating of all LEDs from brightness value 0 to 200 and back to 0 over 0.8 s with a 0.2 s break.

Vibration patterns: With increasing priority of the alarm, the number of vibrations increases. Respectively, the technical alarm consists of one, the uncritical alarm of two, and the critical alarm

of three recurring vibrations with a length of 400 ms and a pause of 100 ms between the vibrations. The pattern itself repeats after an 800 ms pause.

Sound patterns: The sound patterns are based on the original sound files of a commercial PMS. With increasing priority of an alarm, the loudness and frequency and pitch of the beep increase. The patterns are visualized in Figure 16.

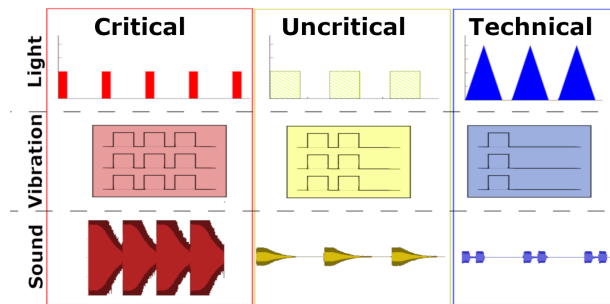


Figure 16. Visualization of the used patterns for each modality.

The vibrotactile and light alarm patterns were hard-coded on the Feather. The acoustic alarms were stored as mp3-files on a micro-SD-card, read by the FeatherWing. We tested loudness, brightness and intensity of each condition and the background noises in prior pilot tests. These alarms were triggered via an Android app using a WiFi-connection to a web server that was started up by the Feather.

7.3. Study Design

For our study, we used a within subject design with the alerting medium as independent variables. It consisted of four main conditions, based on the used modalities: visual via peripheral light, tactile via vibration, and auditory via bone conduction speakers as experimental conditions, and finally auditory via speakers as control condition.

The essential key tasks of a nurse include i.e., the shift handover, a routine control, and the admission of new patients, which are mainly cognitively demanding. Furthermore, there are physically demanding tasks like mobilization of the patients or accompaniment of their transport, and, finally, there are precision demanding tasks like the application of medication or dressing changes. For that reason, we divided our study into three tasks that are cognitively, physically and precision demanding to represent realistic workloads according to the workflow.

For the *cognitive task*, we prepared several cross-multiplication problems that should be solved using the Rule of Three. This method is commonly used in nursing education and for calculating the dose of medication for patients. To avoid discrepancies between the math knowledge of the participants, we provided an explanation, including an example for the Rule of Three as well as a calculator.

For the *physical task*, the participants were asked to perform a mobilization on a training manikin (Resusci Anne), which weighs 12 kg and is about 1.5 m tall (see Figure 17, left). We chose this manikin to avoid fatigue effects caused by repeating the task, since awake patients usually support the nurses and heavier patients will be mobilized with at least two healthcare professionals.

The *precision task* consisted of a wire loop game in which the participant had to remove plastic items with a pair of tweezers from cavities inside the patient without touching the edges of that cavity. If s/he touched an edge, the game board gave a visual and audible signal and the participant had to choose a different item (see Figure 17, right). This hand-eye-coordination task should represent a load similar to giving injections or taking blood and was performed in former studies, e.g., by Englert et al. [29].

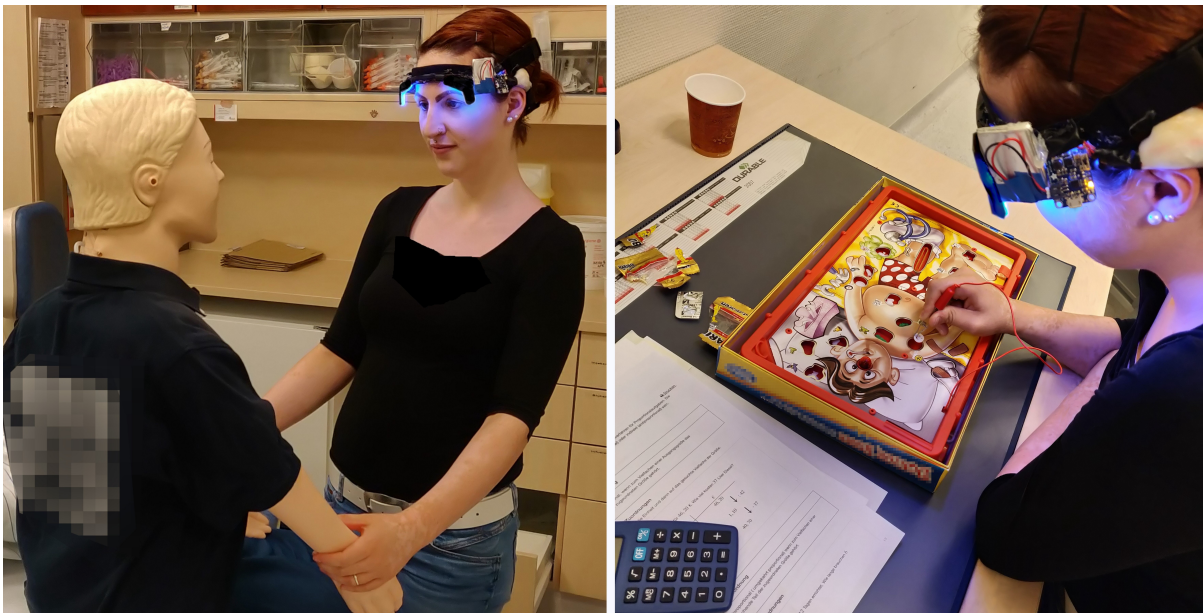


Figure 17. Left side: physical task, right side: precision task.

During each task, all conditions were evaluated. Therefore, every participant performed each task and experienced three alarm levels per condition: critical, uncritical and technical alarms. The order of the conditions, alarm levels, and tasks was counter-balanced randomized. A possible setup for one participant is visualized in Figure 18.

Participant 1	Physical Task	Light	Urgency 1	Urgency 2	Urgency 3
		Vibration	Urgency 2	Urgency 3	Urgency 1
		Bone Conduction	Urgency 3	Urgency 1	Urgency 2
		Speakers	Urgency 1	Urgency 2	Urgency 3
	Cognitive Task	Light	Urgency 2	Urgency 3	Urgency 1
		Bone Conduction	Urgency 3	Urgency 1	Urgency 2
		Speakers	Urgency 1	Urgency 2	Urgency 3
		Vibration	Urgency 2	Urgency 3	Urgency 1
	Precision Task	Vibration	Urgency 3	Urgency 1	Urgency 2
		Light	Urgency 1	Urgency 2	Urgency 3
		Speakers	Urgency 2	Urgency 3	Urgency 1
		Bone Conduction	Urgency 3	Urgency 1	Urgency 2

Figure 18. Exemplary study setup for one participant.

The experiment took approximately 60–90 min including setup, initial learning phases and final interview.

The nurses received their actual hourly wage for participation.

7.4. Measures

We measured the following dependent variables to compare the experimental conditions to the control conditions:

General suitability of the modality, suitability of the modality for the performed task, perceivability, distinguishability, comfort, and level of annoyance: (5-point Likert scale, 5—the most suitable—annoying): for each task after each condition, every participant rated the respective factor for the modality.

Perceived workload: for each condition during each task, every participant filled out a Raw-TLX Scale [31].

Error rate: each time a participant named the wrong or no alarm level, we counted that as an error and logged it to the specific alarm.

Reaction time: we measured the time between presentation of the certain alarm and the identification by the participant.

Usability: at the end of the study, every participant filled out a System Usability Scale [32].

Comfort of the prototype: at the end of the study, every participant filled out a Comfort Rating Scale [33].

7.5. Procedure

After obtaining the informed consent, we collected the participants' demographic data (age, gender, years of ICU experience) and explained the overall procedure.

For each condition, we introduced the respective alarm design to the participants and gave them an initial learning phase. The study started when the participant felt confident with the alarm design. The background noise was started and the three alarms of the respective condition were sent to the prototype or speaker. By sending the alarm, a timer was started automatically. As soon as the participant recognized an alarm, s/he had to name the type (red for critical, yellow for uncritical or blue for technical alarm) and we stopped the timer. After each alarm, we asked the participant to rate the suitability of the design for the respective alarm level, the suitability for the performed task and the perceivability on a 5-point Likert scale. At the end of the certain condition, we asked the participant to fill out a Raw-TLX scale and to rate the general suitability, task suitability, perceivability, distinguishability of the alarms, comfort and level of annoyance for the respective condition, also on a 5-point Likert scale.

This procedure was repeated for each condition during each task. At the end of the study, the participants were asked to fill out a system usability scale as well as a Comfort Rating Scale. Finally, we asked the participant which modalities they would prefer for their personal alarm system.

7.6. Results

In the following, we present our results in the respective evaluated factor. A Shapiro–Wilk test revealed that all data are normally distributed, so we used a Wilcoxon signed-rank test to test for significant differences with a significance level of 1%. A summary of the results can be seen in Figure 19.

		Light	Vibration	BC	Speakers
Task Suitability	Suitability	4	3	4	3
	Cognitive	4	4	4	3
	Physical	4.5	4	4	3
	Precision	5	3	4	3
	Perceivability	5	5	4	3
	Distinguishability	5	3	4	3
	Comfort	4	3	4	3
	Annoyance	2	3	3	3
	Error rate	0	0.028	0.093	0.093
	ReactionTime(s)/SD	2.02/0.6	3.15/1.06	3.37/1.7	3.71/1.74

Figure 19. Summary of the results.

Suitability of the modality: For the general suitability, the participants rated light and bone conduction sound (in the following named as BC) significantly best, with a median of 4 ($p < 0.01$). Regarding light, the comments included that this modality is “really secure” (P1), “safe” (P1, P3, P7), “clear, unique” (P2), and “not distracting” (P1, P3–P7, P9). A representative quote of a participant for BC was: “Somehow distracting. However, it SHOULD distract me in a specific way. It is so weird because it is in my head. Maybe I could get used to it.” (P2)—Vibration and speakers were both rated with 3, which means a medium suitability.

Task suitability: For the task suitability, we regarded the ratings in the respective task. For each task, light and BC performed significantly better than the control condition, which had an overall rating of 3/medium ($p < 0.01$). Except for the cognitive task, light performed with a rating of 4.5 for the physical and 5 for the precision task even better than all conditions ($p < 0.01$). However, one participant (P9) was concerned during the physical task about how the patient would perceive that. Vibration performed better than the control condition for the cognitive ($p < 0.01$) and physical task (not significant, $p = 0.03$). During the precision task, two participants (P4, P12) were concerned that they would be afraid to get out of place with the needle during taking blood. For this task, vibration was rated as medium suitable.

Perceivability: Regarding the perceivability, all experimental conditions performed significantly better than the control condition. Comparing light and vibration (both 5, best perceivable), there is no significant difference but eight participants mentioned a rather negative or annoying perceivability for vibration. In contrast, light was described as “Fast perceivable, especially with background noises” (P1).

Distinguishability: The distinguishability of the alarms was rated significantly best (median = 5) compared to all conditions for light ($p < 0.01$). Worst distinguishable were vibration and speakers, with a medium rating (3). For vibration, one participant (P8) mentioned: “Sure, they are distinguishable. However, I have to concentrate on that and count the vibrations, so they are actually not!”

Comfort: The participants rated light ($p < 0.01$) and BC with as “comfortable”. Two participants (P2, P11) were concerned that light could be exhausting for the eyes, and one of them mentioned that s/he was “tired and had headache during the shift. Bright light doesn’t make it any better. Maybe there should be a mode to switch to something different.” (P11). Vibration and the control condition were rated as medium comfortable. *Level of annoyance:* Regarding this factor, just light showed a better score, with 2 (not annoying). One participant (P12) described light as “distracting but positively distracting. Opposite to the speakers, I see that immediately, I don’t have to interrupt my task.” All other conditions were rated medium.

Error rate: The most errors were made with acoustic alarms. Both speakers and BC showed an error rate of 9.3%. Each mistake was a confusion between the uncritical and technical alarm. Five participants mentioned that they got used to hearing an alarm and look to the PMS display to realize which one it is. With one error, vibrotactile alarms showed an error rate of 2.8%. No error was made for light alarms.

Generally, the time stamps for the mistaken patterns did not indicate a learning effect for the patterns. *Reaction time:* With 2.02 s (SD = 0.6), the participants had the fastest reaction time in identifying the light alarms. Generally, all experimental conditions performed significantly better than the control condition (3.72 s, SD = 1.74).

Perceived workload: The analysis of the Raw-TLX showed that there are overall high scores, which means a high workload. However, there is also a high standard derivation for each factor (see Figure 20). The participants perceived the lowest work load in all categories with light alarms. Compared to the control condition, vibration showed worse results in all categories except in “Overall Performance”. BC performed just slightly better than the control condition except for the factor “Frustration”. However, these results are not significant.

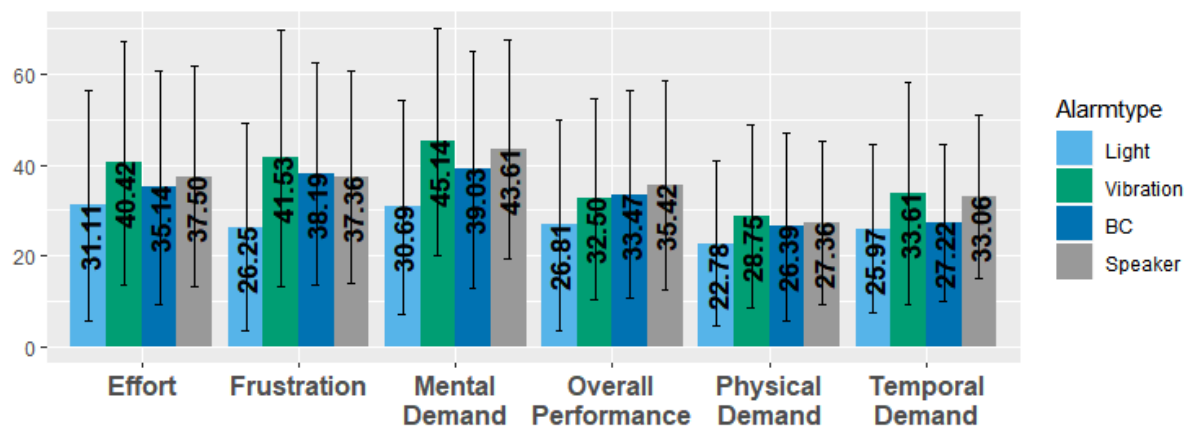


Figure 20. Rating of the perceived work load.

Usability and Comfort: With a SUS-score of 80 (SD = 8.53), the participants rated the prototype with good usability. The median of the Comfort Rating is 4 (0 means best comfort), which means good comfort, considering the prototype's early state development. However, Figure 21 shows that the lowest rating applies to the attachment of the prototype (median = 16.5).

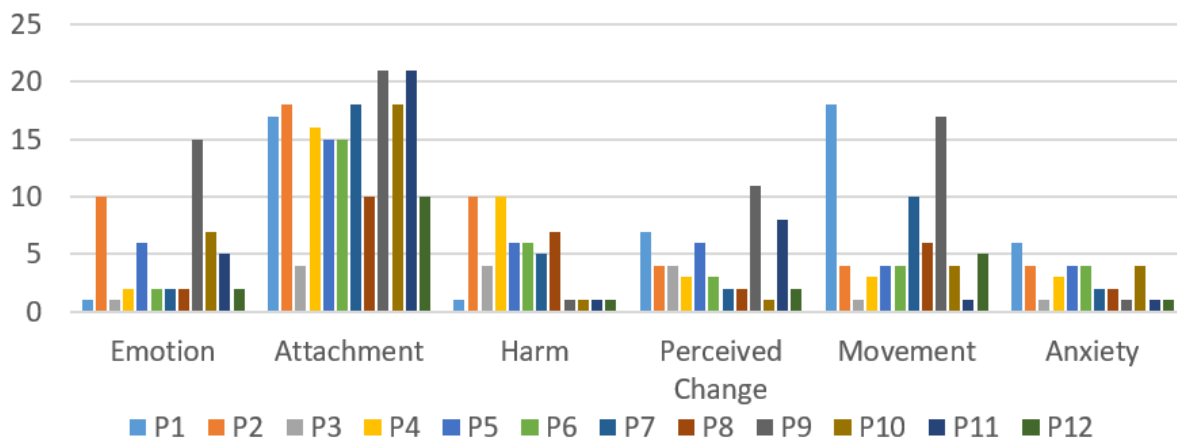


Figure 21. Results of the Comfort Rating Scale.

Qualitative feedback: Regarding the final feedback, none of the participants wanted to adhere to the state of the art. The majority of the participants (11 of 12) preferred the light alarms. Two participants mentioned that light would be sufficient for each alarm level. However, five of them remarked that they could imagine that other nurses could prefer vibrotactile alarms and we should consider making the alerting method customizable. One participant preferred BC as single method to be alerted, since s/he reacted sensitively to light, especially, when s/he is sick. The other participants preferred a combination of modalities for specific alarms. For example, 9 of 12 participants wanted a combination of light and BC for critical alarms. Ten of 12 participants preferred just a visual solution for uncritical alarms. Regarding the technical alarm, there was no clear result. Five participants preferred the light alarms for this, two the acoustic, and two the vibrotactile solution. Two participants suggested a combination of vibration and light and one participant considered a combination of all modalities because the technical alarm is usually underestimated and ignored, since a missing sensor could also cause a critical alarm.

7.7. Discussion

Regarding all factors, we suggest using a combination of light and BC for critical alarms, and just light to represent uncritical and technical alarms. However, even if light performed best in all categories, there was a high variability of the preferred signal types. Therefore, we should consider enabling an opt-out for light alarms and, at least for uncritical and technical alarms, and opt-in for a vibrotactile alerting, which needs to be configured individually. This option could also be useful to switch the modality for specific tasks.

Even if the time stamps of the measured errors did not indicate that there was a learning effect, it should be considered that this may be caused by the small sample size and could be clarified in a long-term study. Moreover, the high error rate for sounds may be explained by the fact that nurses are generally used to getting additional information on the monitoring system for the specific alarm to identify its urgency.

The high means of the Raw-TLX have to be regarded in combination with the high standard derivation. This may be caused by the fact that some participants attended our study right after their shift, which could mean that they might have been exhausted. Another possible reason could be the low number of participants. If we take a closer look at the single factors of the Comfort Rating Scale, the overall median represents the rating for the factors "Emotion", "Harm", "Perceived Change", "Movement", and "Anxiety". However, the factor "Attachment" was rated with a median of 16.5, which means a physical feel of the device on the body. This is caused by the early state of development, in which we focused on the functionality of the prototype.

Although we could show significant results that support the use of personal multimodal alarms, our results are still limited. At this point, we did not evaluate the combination of the modalities, and, moreover, the task performance. Even if the qualitative feedback didn't show any indications in case of light and BC, we still need to measure whether the multimodal alarms have a negative influence in performing nursing tasks.

Moreover, it needs to be evaluated if multimodal alarms perform as well in a long-term study. Thus, we can exclude fatigue due to the new alarm signaling.

Another point that needs to be evaluated is the long-term use of HMDs in the medical context. There is a risk that "yet another technical device" will be neglected as soon as the novelty effect ends. To avoid this, a WAS would have to be made mandatory for the specific ICU as part of the general work clothes.

However, for a long-term evaluation, the WAS needs an approval for medical devices to be tested in the field. Until then, we aim to create a more realistic test lab for preliminary explorations.

8. Conclusions and Future Work

In this paper, we presented the design of vibrotactile and peripheral light patterns to represent ICU alarms. Moreover, we showed the evaluation of different modalities for a personal multimodal alerting on ICUs via an HMD.

Under task conditions that mimic the specific load of nursing tasks, we could show that visual and tactile alarms performed better and were faster to identify by nurses than the state of the art. In particular, light and bone conduction speakers performed well regarding the factors suitability, task suitability, perceivability, distinguishability and comfort.

Based on our results, we could derive a final multimodal alarm design, which will be evaluated with a realistic alarm frequency as a next step. The qualitative feedback we got from nurses indicated that there is a general need to change from a ubiquitous, obtrusive and distracting alert system to a personal, noiseless alarm distribution. We could observe that our participants were positively surprised about the new method of alerting.

Even though the prototype was in an early state of development, there were just concerns regarding the attachment of the HMD. This will be solved in future work.

However, there were concerns about the acceptance of peripheral light alarms of patients and relatives we have to consider. Overall, we have to take care about evaluating the social acceptability of such a device. Moreover, our alarm design needs to be evaluated in a long-term study with a realistic number of alarms.

In the future, we will integrate the multimodal alarms into data glasses to test their feasibility in combination with displayed vital data. Finally, we aim to develop touch-free interaction methods to acknowledge, forward, or silence alarms.

Author Contributions: Writing—Original Draft Preparation, V.C.; Writing—Review and Editing, W.H.

Funding: This research was funded by Bundesministerium für Bildung und Forschung (BMBF) grant number (16SV7501).

Acknowledgments: At this point, we want to thank the AlarmRedux Research Team, which gave constructive feedback to our study designs and prototypes. In particular, we want to thank the healthcare professionals who took their limited time to participate in our studies.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Ruskin, K.J.; Hueske-Kraus, D. Alarm fatigue: Impacts on patient safety. *Curr. Opin. Anaesthesiol.* **2015**, *28*, 685–690. [[CrossRef](#)] [[PubMed](#)]
2. Block, F.E. “For if the trumpet give an uncertain sound, who shall prepare himself to the battle?” (I Corinthians 14:8, KJV). *Anesth. Analgesia* **2008**, *106*, 357. [[CrossRef](#)] [[PubMed](#)]
3. Grissinger, M. Too many abandon the “second victims” of medical errors. *P&T* **2014**, *39*, 591–592.
4. Wilken, M.; Hüske-Kraus, D.; Klausen, A.; Koch, C.; Schlauch, W.; Röhrig, R. Alarm fatigue: causes and effects. *Stud. Health Technol. Inform.* **2017**, *243*, 107–111. [[PubMed](#)]
5. Winters, B.D.; Cvach, M.M.; Bonafide, C.P.; Hu, X.; Konkani, A.; O’Connor, M.F.; Rothschild, J.M.; Selby, N.M.; Pelter, M.M.; McLean, B.; et al. Technological Distractions (Part 2): A Summary of Approaches to Manage Clinical Alarms with Intent to Reduce Alarm Fatigue. *Crit. Care Med.* **2018**, *46*, 130–137. [[CrossRef](#)] [[PubMed](#)]
6. Altosaar, M.; Vertegaal, R.; Sohn, C.; Cheng, D. AuraOrb: Using social awareness cues in the design of progressive notification appliances. In Proceedings of the 18th Australia Conference on Computer-Human Interaction: Design: Activities, Artefacts and Environments, Sydney, Australia, 20–24 November 2006; pp. 159–166.
7. Chang, Y.J.; Pan, Y.J.; Lin, Y.J.; Chang, Y.Z.; Lin, C.H. A noise-sensor light alarm reduces noise in the newborn intensive care unit. *Am. J. Perinatol.* **2006**, *23*, 265–272. [[CrossRef](#)] [[PubMed](#)]
8. Lee, S.C.; Starner, T. BuzzWear: Alert perception in wearable tactile displays on the wrist. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Atlanta, GA, USA, 10–15 April 2010; pp. 433–442.
9. Cobus, V.; Boll, S.; Heuten, W. Requirements for a Wearable Alarm Distribution System in Intensive Care Units. In *Zukunft der Pflege, Tagungsband der 1. Clusterkonferenz 2018—Innovative Technologien für die Pflege. oops*; BIS-Verlag: Oldenburg, Germany, 2018; pp. 185–189.
10. Cobus, V.; Ehrhardt, B.; Boll, S.; Heuten, W. Vibrotactile Alarm Display for Critical Care. In Proceedings of the 7th ACM International Symposium on Pervasive Displays, Munich, Germany, 6–8 June 2018; ACM: New York, NY, USA, 2018; pp. 11:1–11:7. doi:10.1145/3205873.3205886. [[CrossRef](#)]
11. Cobus, V.; Meyer, H.; Boll, S.; Heuten, W. Towards Reducing Alarm Fatigue: Peripheral Light Pattern Design for Critical Care Alarms. In Proceedings of the 10th Nordic Conference on Human-Computer Interaction, Oslo, Norway, 29 September–3 October 2018. doi:10.1145/3240167.3240218.
12. Wickens, C.D. Multiple resources and mental workload. *Hum. Fact.* **2008**, *50*, 449–455. [[CrossRef](#)]
13. Cvach, M.M.; Frank, R.J.; Doyle, P.; Stevens, Z.K. Use of pagers with an alarm escalation system to reduce cardiac monitor alarm signals. *J. Nurs. Care Qual.* **2014**, *29*, 9–18. [[CrossRef](#)]

14. Brander, S.; von Schewen Sterndal, K. Development of Wearable Healthcare Device: A User-Centred Project Focused on Meeting Discovered Needs of Nurses. Master's Thesis, Machine Design (Dept.), KTH, Stockholm, Sweden, 2014.
15. Chapman, C.E.; Bushnell, M.C.; Miron, D.; Duncan, G.H.; Lund, J.P. Sensory perception during movement in man. *Exp. Brain Res.* **1987**, *68*, 516–524. [[CrossRef](#)]
16. Wrzesinska, N. The use of smart glasses in healthcare—Review. *MEDtube Sci.* **2015**, *3*, 31–34.
17. Vorraber, W.; Voessner, S.; Stark, G.; Neubacher, D.; DeMello, S.; Bair, A. Medical applications of near-eye display devices: An exploratory study. *Int. J. Surg.* **2014**, *12*, 1266–1272. [[CrossRef](#)] [[PubMed](#)]
18. Mentler, T.; Wolters, C.; Herczeg, M. Use cases and usability challenges for head-mounted displays in healthcare. *Curr. Direct. Biomed. Eng.* **2015**, *1*, 534–537. [[CrossRef](#)]
19. Pascale, M.T.; Sanderson, P.; Liu, D.; Mohamed, I.; Brecknell, B.; Loeb, R.G. The Impact of Head-Worn Displays on Strategic Alarm Management and Situation Awareness. *Hum. Fact.* **2019**. doi:10.1177/0018720818814969. [[CrossRef](#)] [[PubMed](#)]
20. Ng, J.Y.C.; Man, J.C.F.; Fels, S.; Dumont, G.; Ansermino, J.M. An evaluation of a vibro-tactile display prototype for physiological monitoring. *Anesth. Analg.* **2005**, *101*, 1719–1724. [[CrossRef](#)] [[PubMed](#)]
21. Rossa, C.; Fong, J.; Usmani, N.; Sloboda, R.; Tavakoli, M. Multiactuator haptic feedback on the wrist for needle steering guidance in brachytherapy. *IEEE Robot. Autom. Lett.* **2016**, *1*, 852–859. [[CrossRef](#)]
22. McLanders, M.; Santomauro, C.; Tran, J.; Sanderson, P. Tactile displays of pulse oximetry in integrated and separated configurations. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, Chicago, IL, USA, 27–31 October 2014; SAGE Publications: Sauzen Oaks, CA, USA, 2014; Volume 58, pp. 674–678.
23. Dosani, M.; Hunc, K.; Dumont, G.A.; Dunsmuir, D.; Barralon, P.; Schwarz, S.K.; Lim, J.; Ansermino, J.M. A vibro-tactile display for clinical monitoring: Real-time evaluation. *Anesth. Analg.* **2012**, *115*, 588–594. [[CrossRef](#)]
24. Fortmann, J.; Cobus, V.; Heuten, W.; Boll, S. WaterJewel: Design and Evaluation of a Bracelet to Promote a Better Drinking Behaviour. In Proceedings of the 13th International Confererence on Mobile and Ubiquitous Multimedia, Melbourne, Australia, 25–28 November 2014; pp. 58–67.
25. Matviienko, A.; Cobus, V.; Müller, H.; Fortmann, J.; Löcken, A.; Boll, S.; Rauschenberger, M.; Timmermann, J.; Trappe, C.; Heuten, W. Deriving Design Guidelines for Ambient Light Systems. In Proceedings of the 14th International Confererence on Mobile and Ubiquitous Multimedia, Linz, Austria, 30 November–2 December 2015; pp. 267–277.
26. Stenfelt, S.; Goode, R.L. Bone-conducted sound: Physiological and clinical aspects. *Otol. Neurotol.* **2005**, *26*, 1245–1261. [[CrossRef](#)]
27. McBride, M.; Letowski, T.; Tran, P. Bone conduction reception: Head sensitivity mapping. *Ergonomics* **2008**, *51*, 702–718. [[CrossRef](#)]
28. Myles, K.; Kalb, J.T. *Guidelines for Head Tactile Communication*; Technical report; Army Research Lab Aberdeen: Adelphi, MA, USA, 2010.
29. Englert, C.; Bertrams, A. Too exhausted for operation? Anxiety, depleted self-control strength, and perceptual—Motor performance. *Self Identity* **2013**, *12*, 650–662. [[CrossRef](#)]
30. Yadunath, R.V.; Jeyapaul, R. Trip-Hit Accidents and Safety: Human Error Psychology and Influence of the Subconscious Mind in Preventing and Causing Trip Hit Accidents. Available online: <http://inet.vidyasagar.ac.in:8080/jspui/bitstream/123456789/250/1/42.Trip-Hit> (accessed on 25 January 2019).
31. Hoonakker, P.; Carayon, P.; Gurses, A.P.; Brown, R.; Khunlertkit, A.; McGuire, K.; Walker, J.M. Measuring workload of ICU nurses with a questionnaire survey: The NASA Task Load Index (TLX). *IIE Trans. Healthc. Syst. Eng.* **2011**, *1*, 131–143. [[CrossRef](#)]

32. Brooke, J. SUS-A quick and dirty usability scale. *Usabil. Eval. Ind.* **1996**, *189*, 4–7.
33. Knight, J.F.; Baber, C. A tool to assess the comfort of wearable computers. *Hum. Fact.* **2005**, *47*, 77–91. [[CrossRef](#)] [[PubMed](#)]



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).