International Journal on Robotics, Automation and Sciences

Testing of Auditory Clinical Alarms in ICU/CCU

Sasikala Thangavelu, Eko Supriyanto*, and Jasmy Yunus

***Abstract*— Almost all medical devices in ICU/CCU have a built-in clinical alarm system to alert when there are changes in a patient's condition. The objective of this study is to investigate the effectiveness of the existing alarm system in ICU/CCU. Two summative usability tests were conducted to test the effectiveness of existing and new alarm signals based on IEC 60601-1-8:2006 standard. Further formative test is conducted to study the perception of urgency associated with a number of tones in the alarm signals. The findings indicate that the existing auditory alarm signal in ICU/CCU does not indicate the urgency of the alarm conditions. The simulation test indicates that the respondents preferred 282Hz, 500Hz and 800Hz for low, medium and high-risk alarm respectively. The one-sample proportion z test on urgency mapping indicates that the proportion of responses for the highest risk is more than 50% for a single tone test signal. These results show that a single tone test signal being perceived as the highest risk is regardless of frequency. It can be concluded the auditory alarm designed based on this IEC 60608-1-8:2006 standard is not effective. As such it is proposed that the incorporation of the new alarm frequencies and tones will improve the effectiveness of the alarm signal.**

Keywords—Clinical Alarm System, Auditory.

I. INTRODUCTION

With the advancement of healthcare technology, there is a proliferation of medical devices in healthcare institutions. These devices have a built-in clinical alarm system to monitor and alert when there are changes in a patient's condition or malfunction of the devices. Each of these devices has its own audio and visual alarm and there is no standardization among the manufacturers [1]. Efforts to harmonize alarm systems in medical equipment have led to the issuance of a new standard IEC60601-1-8 in 2003.

The IEC60601-1-8 which was published in the year of 2003, is a collateral standard on general requirements, tests and guidance for alarm systems in medical electrical equipment and systems [2]. This standard replaces the ISO 9703-1:1992 standard on anesthesia and respiratory care alarm signals which include Part 1 on visual alarm signals, Part 2 on the auditory alarm signals, Part 3 on the application of alarms and EN 475:1995. It is a voluntary standard and specifies the requirements of alarm systems and alarm signals in medical electrical equipment and medical electrical systems. Under the 3rd edition of IEC60601-1, the alarm system standard is a mandatory requirement for all medical devices [3].

The second edition of IEC 60601-1-8 was published in 2006 which details the general requirements, tests and guidance for alarm systems in medical electrical equipment and medical electrical systems and covers alarm systems incorporated into medical equipment. The amendment in this edition details the requirements for alarm signal inactivation states and reminder signals [4]. In the context of alarm signals, this version refers to indicator lights for visual alarms signal and melody or optional verbal sounds for auditory alarm signals. It is the intention of this study to look into the effectiveness of auditory alarm signals based on this standard.

The IEC 60601-1-8 gives guidance on alarm conditions assigned as high, medium, or low priority. The corresponding high priority alarm signals are composed of two repetitions of five-note melodies, medium priority alarm signals are composed of three-note melodies while low priority alarms are composed of two note melodies. In addition, mnemonics which are provided to support the learning of alarm signals, ensure difference between the alarms due to the different devices and help clinicians to remember the nature of the devices [5].

It is highlighted in this IEC standard that any new audible alarms need to be validated before implementation [4]. However, the melodies and methods for urgency-encoding are not validated in the clinical real world or in the simulated clinical settings [6]. Furthermore, the standard does not offer a

*Corresponding author. Email : eko@utm.my
Sasikala Thangavelu and Jasmy Yunus are with Medical Device Authority Malaysia (e-mail: sasikala@mdb.gov.my).
Eko Supriyanto is with the Advanced Diagnostics and Progressive Human Care Research Group, IJN-UTM Cardiovascular Engineering Centre and School of Biomedical Engineering and Health Sciences, Faculty of Engineering, Universiti Teknologi Malaysia, 81310 Johor, Malaysia (e-mail: eko@utm.my).

International Journal on Robotics, Automation and Sciences (2020) 2:6-14
https://doi.org/10.33093/ijoras.2020.2.2
Manuscript received: 1 Nov 2018 | Accepted: 23 April 2019 | Published: 30 Oct 2020
Published by MMU PRESS. URL: http://journals.mmupress.com/ijoras

validation method [4]. Moreover, no formal tests or empirical evaluation of the melodic alarms were conducted to verify their effectiveness when the standard was published [7]. This study also highlighted that no human factors studies on alarms were conducted before the initial publication of the standard in the year of 2003. Sanderson et al. presented an empirical study to evaluate the learnability and discriminability of the melodic alarms that were proposed by IEC 60601-1-8 [6]. They compared the performance of participants with and without access to the mnemonics and found very little benefit for mnemonics. They speculated that the use of mnemonics increased confusion. Sanderson et al., William et al. and Lacherez et al. indicated that the melodic alarms were difficult to be identified and distinguished [6,8,9]. However, Sanderson et al. did not use trained medical volunteers in their study [6]. Edworthy et al. further raised concerns in auditory design theory [10,11]. Their studies also indicated no significant effect of mnemonics on speed and accuracy of identifying the alarms or on retention. The only observable difference between conditions was that the participants in the mnemonic condition were confused with several pairs of alarms whereas the participants in the non-mnemonic conditions had a confusion pattern that was more idiosyncratic [6].

It is well documented that the current IEC standard of auditory clinical alarms is in need of updating [12]. Poor quality of the standard causing the alarms difficult to learn, easily confusable and ineffective at conveying the appropriate level of urgency [13,14,15]. This will lead to alarm fatigue, which associated with the human factor and ergonomic problems, and has a potential safety risk to the patient [16,17,18].

Studies have been conducted to discuss the improvement of current auditory clinical alarms standard as well as developing potential alternative alarm signals [19,20,21]. Some of the previous studies used clinical personnel as subjects and none of the studies were conducted in a realistic clinical environment. In contrast, this paper looks into the test of effectiveness of the auditory alarm signals design test based on IEC60601-1-8 with the clinical users in an ICU environment.

Three simulation tests were conducted in this study to test the effectiveness of auditory alarm signals.

- Simulation A: Testing of existing alarm signals in ICU/CCU with the objectives to study the ability of clinicians to identify the source and the risk associated with the existing auditory alarms.
- Simulation B: Testing the new auditory alarm signals based on IEC60601-1-8. This simulation is conducted to study the ability of clinicians to recognize the risk in the new alarm signals and the urgency mapping associated with those.
- Simulation C: Testing to establish the relationship between perceived urgency and number of tones. This is to study the perception of urgency associated with the number of tones in the alarm signals.

II. METHODOLOGY

The Continuous Medical Education (CME) room in the ICU was simulated in this study. TENMA 72-6635 was used to measure the sound pressure level of the alarm signals with ambient sound turned on throughout the sessions.

Z-tests and charts were created using Microsoft Excel 2007.

A. Simulation A

Audio recordings of the existing auditory alarm sounds were carried out in the ICU with a clinician for one week. Recordings with clear background speech or conversations were deleted. Five devices commonly available in ICU/CCU were selected, the alarm sounds were recorded and played back to check on the learnability among the clinicians in ICU. The list includes a feeding pump, a patient monitor, a ventilator, a perfusion and hemodialysis machine. The respondents were required to identify the medical devices that generated the alarm sound and the associated risk. A total of 27 clinicians, nurses and post-basic nursing students in ICU participated in this study. This is a summative usability testing of the existing auditory alarm of the medical devices commonly available in ICU/CCU.

B. Simulation B

The test based on the principles of usability engineering and international standard IEC62366:2007 was conducted in a simulated ICU/CCU room with new auditory alarm and 100 clinicians. This simulation test methodology can be divided into 5 phases;

- Phase 1: Generation of alarm signal,
- Phase 2: Selection of alarm sound based on frequency
- Phase 3: Learning phase and testing of learnability
- Phase 4: Learnability testing
- Phase 5: Urgency testing of alarm signal

Nine sets of auditory alarm signals with specifications based on IEC60601-1-8 were composed using Audacity software version 2.0.5. Frequency ranges from 200Hz to 1000Hz were recommended for alarm signal in medical devices. In this study, frequency ranges from 150Hz to 350Hz was designated to low-risk alarm signals, frequency ranges from 351Hz to 399Hz was designated to medium-risk alarm signals and frequency ranges from 501Hz to 1000Hz was designated to high-risk alarm signals. The selected frequencies for the alarm signals were based on a logarithmic scale: 252Hz, 282Hz, 317Hz with 2 tones for low-risk alarm; 400Hz, 447Hz and 500Hz with 3 tones for medium-risk alarm and 635Hz, 708Hz and 800Hz with 5 tones for high-risk alarm.

The interburst interval (t_b), rise time (t_r), fall time (t_f), pulse spacing (t_s) and pulse duration (t_d) of the alarm signal were generated based on IEC60601-1-8:2006 recommendation as follows:

- low-risk alarm:
  - t_b = 16s,16s, 5s and 17s
  - t_d = 200ms; t_s = 200ms; t_r = 20ms; t_f < 180ms
- medium-risk alarm:
  - t_b = 10s,12.5s and 15s
  - t_d = 200ms; t_s = 200ms; t_r = 20ms; t_f < 180ms
- high-risk alarm:
  - t_b = 2.5ms, 5.0s and 7.5s
  - t_d = 200ms; t_s = 100ms; t_r = 20ms t_f < 180ms
  - t_s 1, t_s 2, t_s 4, t_s 6, t_s 7, t_s 9 = 00ms; t_s 3, t_s 8 = 400ms

The temporal characteristics of generated auditory alarm signals waveform are shown in Figure 1.

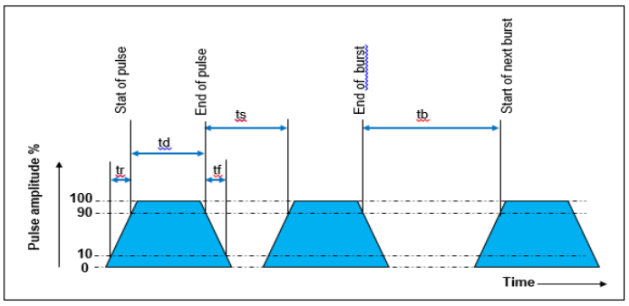

FIGURE 1. Proposed temporal characteristics of auditory alarm signals (extract from IEC60601-1-8:2006).

Part I: Learnability Phase. The participants were required to learn 9 different alarm sounds in the CME room. The researcher played back the alarm signals in the simulation room using a Sony VAIO laptop and speakers. In the learning phase, the respondents were introduced to the alarm signals and the associated risk assignments: low, medium and high-risk alarm conditions.

Part 2: Test Phase. In the test phase, the participants were requested to listen to the nine same alarm signals again after one week. The alarm sounds were played, and the participants were asked to indicate whether it is low, medium and high-risk alarm conditions accordingly.

Part 3: Urgency mapping. Three sets of alarm signals for low, medium and high-risk alarm conditions were composed with three different interbursts (t_b). In this study, the t_b assigned for low-risk alarm were 16ms,16.5ms and 17ms, medium risk by 10ms,12.5ms and 15ms and high risk by 2.5ms, 5.0ms, and 7.5ms based on repetition time based on IEC60601-1-8. The participants were required to identify the alarm signals with suitable t_b for low, medium and high-risk alarm conditions based on the perceived urgency.

C. Simulation C

Test signals with three sets of medium frequencies (400Hz, 447Hz, 500Hz) and three sets of high frequencies (635Hz, 708Hz, and 800Hz) were generated with tone duration t_d and constant amplitude. Each test signal has 4 sets of tones: single tone (T1), 2-tones (T2), 3-tones (T3) and 5-tones (T5). Pulse duration (t_d) was fixed at 400ms and 200s.

A total of 60 clinicians, nurses, medical assistant and post-basic nursing students in ICU participated in this study. They were briefed on the meaning and representation of low-risk, medium-risk, and high-risk alarm sounds. The alarm sound effect was created for each test signal and the respondent needed to rank the perceived urgency of the sound and representation of the alarm sound from the highest risk (R1) to the lowest risk (R4).

III. RESULTS

A. Results of Simulation A

16 different alarm sounds were recorded from different medical devices with different models and brands in ICU. The devices include feeding pump, 5 types of ventilator, 3 types of perfusor, 5 types of patient monitor and 2 types of dialysis unit.

The z-test for one-proportion was used to test if the proportion of correct identification exceeds 50%.

H0: p=0.50

H1: p≠0.50

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$$

where $\hat{p}$ =sample proportion, n is sample size and p_o =0.50

At 5% significance level, we reject the null hypothesis if z≥1.96 or z≤ -1.96 or p-value <0.05.

27 clinicians could recognise the alarm sounds correctly from feeding pump, Ventilator 2, Patient Monitor 3, 4, 6, 7 and Dialysis 1, 2, 3 as shown in Table 1. Majority of clinicians could recognise the alarm sounds from Patient Monitor 1, 2, 5, Perfusor 3 and Ventilator 3. However, only 30% of clinicians could recognize the alarm sounds correctly from Perfusor 4.

TABLE 1. One sample proportion test results for device recognition.

Device	X	n	p	z	p-value
Feeding Pump	27	27	1.00	5.196	0.0000
Patient Monitor1	24	27	0.89	4.041	0.0001
Ventilator 2	27	27	1.00	5.196	0.0000
Ventilator 3	24	27	0.89	4.041	0.0000
Perfuser 3	25	27	0.93	4.426	0.0000
Perfuser 4	8	27	0.30	-2.117	0.0343
Ventilator 1	22	27	0.81	3.272	0.0011
Patient Monitor2	25	27	0.93	4.426	0.0000
Patient Monitor3	27	27	1.00	5.196	0.0000
Patient Monitor4	27	27	1.00	5.196	0.0000

Patient Monitor5	24	27	0.89	4.041	0.0001
Dialysis 1	27	27	1.00	5.196	0.0000
Dialysis 2	27	27	1.00	5.196	0.0000
Dialysis 3	27	27	1.00	5.196	0.0000
Patient Monitor 6	27	27	1.00	5.196	0.0000
Patient Monitor 7	27	27	1.00	5.196	0.0000

Risk recognition test results were shown in Table 2. The findings indicated that a high proportion of clinicians could recognise the alarm risks from feeding pump (81%) and Dialysis Machines 1, 2, 3 (96%, 100%, 100%). However; lower proportion of clinicians could recognise risk from Perfusor 3 (15%, 26%), Ventilator 2, 3 (44%, 59%), and Patient Monitor 1, 2, 3, 4, 5, 6 (59%, 63%, 56%, 48%, 59%, 63%).

TABLE 2. Ne-sample proportion test results for risk recognition.

Device	X	n	p	z	p-value
Feeding Pump	22	27	0.81	3.272	0.0011
Patient Monitor1	16	27	0.59	0.962	0.3359
Ventilator 2	12	27	0.44	-0.577	0.5637
Ventilator 3	16	27	0.59	0.962	0.1680
Perfusor 3	4	27	0.15	-3.657	0.0003
Perfusor 4	7	27	0.26	-2.502	0.0124
Ventilator 1	16	27	0.59	0.962	0.3359
Patient Monitor2	17	27	0.63	1.347	0.1779
Patient Monitor3	15	27	0.56	0.577	0.5637
Patient Monitor4	13	27	0.48	-0.192	0.8474
Patient Monitor5	14	27	0.52	0.192	0.8474
Dialysis 1	26	27	0.96	4.811	0.0000
Dialysis 2	27	27	1.00	5.196	0.0000
Dialysis 3	27	27	1.00	5.196	0.0000
Patient Monitor 6	26	27	0.96	4.811	0.0000
Patient Monitor 7	27	27	1.00	5.196	0.0000

Most of the respondents were able to recognize the alarm sounds of most of the devices from these tests. Almost all the respondents were able to recognise the alarm sounds of the physiological patient monitor, ventilator and dialysis devices.

Majority of the respondents could recognise the risk associated with the Dialysis. However, they were unable to recognise the risks associated with the alarm sounds especially from the patient monitor and perfusor. From this outcome, it can be concluded that the existing auditory alarm system does not provide enough information on the risk associated with the alarm conditions for the users to identify accurately. As such there is a gap in user recognition to identify the associated risk of alarm signal in current alarm sounds.

The lack of this information could lead towards delayed response time or alarms being ignored by the users. Further investigation needs to be carried out on how to improve this risk recognition in the existing alarm system.

B. Results of Simulation B

Part I: Suitable Alarm Tones for Low, Medium and High-Risk Conditions. The data were analysed using IBM SPSS Statistics 22. Results from descriptive analyses were reported in frequencies and percentages. Table 3 indicated the agreement on suitable alarm frequencies assigned for low, medium and high risks were provided. For low risk, 58(58.0%) of the participants agreed that 282Hz was the most suitable, followed by 34(34.0%) who preferred 317 Hz and 8(8.0%) who preferred 252 Hz alarm signal. For medium risk, 54(54.0%) agreed that 500Hz was the most suitable, followed 42(42.0%) who preferred 447 Hz and 4(4.0%) who preferred 252 Hz. For high risk, 76(76.0%) agreed that 800Hz was the best, followed 23(23.0%) who preferred 708 Hz and one (1%) participant suggested 635 Hz.

TABLE 3. Suitable Alarm Tones for Low, Medium and High Risk.

Suitable alarm tones for	Agreement*
2 tones Low Risk	
252 Hz	8(8.0%)
282 Hz	58(58.0%)
317 Hz	34(34.0%)
3 tones Medium Risk	
400 Hz	4(4.0%)
447 Hz	42(42.0%)
500 Hz	54(54.0%)
10 tones High Risk	
635 Hz	1(1.0%)
708 Hz	23(23.0%)
800 Hz	76(76.0%)

*Frequency(%)

Part 2: Learnability of the auditory alarm signals based on risk. The results for learnability were provided in Table 4. For 2 tone low risk, 99.0% of the participants were able to recognise 282Hz, 97% recognised 317Hz and 90% recognised low frequencies correctly. For 3 tone medium risk, 500Hz tones were most easily recognisable with 99.0% able to recognise it correctly. For 5 tones high-risk signal, 800Hz tones were the most recognised signal.

TABLE 4. Learnability.

Alarm tones	Correct response *
2 tones Low Risk	
252 Hz	90(90.0%)
282 Hz	99(99.0%)
317 Hz	97(97.0%)
3 tones Medium Risk	
400 Hz	62(62.0%)
447 Hz	94(94.0%)
500 Hz	96(96.0%)
5 tones High Risk	
635 Hz	62(62.0%)
708 Hz	95(95.0%)
800 Hz	97(97.0%)

*Frequency(%)

Interburst interval (t_b). The findings on the suitable interburst interval for alarm systems to represent low, medium and high urgency were provided in Table 5. For low risk 252 Hz, 94(94.0%) of the participants chose 16 seconds, while 3 (3.0%) chose 16.5 seconds. For 282Hz, 98(98.0%) of the participants chose 16 seconds, while 1 (1.0%) chose 16.5 seconds. For 317Hz, all the participants chose 16 seconds. Based on the given comments, almost all the participants preferred the interburst interval to be less than 16 seconds.

For medium risk 400 Hz, 96(97.0%) of the participants chose 10 seconds, while 2 (2.0%) chose 12.5 seconds.

For 447Hz, 98(99%) of the participants chose 10 seconds, while 1 (1.0%) chose 12.5 seconds. Similarly, for 500Hz, 98(98.0%) of the participants chose 10 seconds, while 1 (1.0%) chose 12.5 seconds. Based on the given comments, 98% of the participants preferred the time to be less than 10 seconds. For high-risk 635 Hz, 97(97.0%) of the participants chose 2.5 seconds, while 2 (2%) chose 5 seconds. For 708Hz, 98(99.0%) of the participant chose 2.5 seconds, while 1 (1.0%) chose 5 seconds. Similarly, for 800Hz, 98(98.0%) of the participants chose 2.5 seconds, while 1 (1.0%) chose 5 seconds. Based on the given comments, almost all participants preferred the time to be less than 2.5 seconds.

C. Results of Simulation C

Medium risk classification. The perceived risk associated with various numbers of tones were conducted for the following medium frequencies; 400Hz, 447Hz, and 500Hz.

TABLE 5. Suitable interburst interval (t_b) Time For Low, Medium and High.

Alarm tones	Preference [Frequency (%)]
Low Risk 252 Hz	
Set 1 (16 seconds)	94(94.0%)
Set 2 (16.5 seconds)	3(3.0%)
Set 3 (17 seconds)	0
Low Risk 282 Hz	
Set 1 (16 seconds)	98(98.0%)
Set 2 (16.5 seconds)	1(1.0%)
Set 3 (17 seconds)	0
Low Risk 317 Hz	
Set 1 (16 seconds)	99(99.0%)
Set 2 (16.5 seconds)	0
Set 3 (17 seconds)	0
Medium Risk 400 Hz	
Set 1 (10 sec)	97(97.0%)
Set 2 (12.5 sec)	2(2.0%)
Set 3 (15 sec)	0
Medium Risk 447Hz	
Set 1 (10 sec)	98(98.0%)
Set 2 (12.5 sec)	1(1.0%)
Set 3 (15 sec)	0
Medium Risk 500Hz	
Set 1 (10 sec)	98(98.0%)
Set 2 (12.5 sec)	1(1.0%)
Set 3 (15 sec)	0
High Risk 635 Hz	
Set 1 (2.5 sec)	93(93.0%)
Set 2 (5.0 sec)	5(5.0%)
Set 3 (7.5 sec)	0
High Risk 708 Hz	
Set 1 (2.5 sec)	98(98.0%)
Set 2 (5.0 sec)	1(1.0%)
Set 3 (7.5sec)	0
High Risk 800 Hz	
Set 1 (2.5 sec)	98(98.0%)
Set 2 (5.0 sec)	1(1.0%)
Set 3 (7.5 sec)	0

Table 6 represented the urgency ranking based on the number of tones for medium frequencies of 400Hz, 447Hz, and 500Hz test signals. Results showed that a high majority (90% and above) ranked single tone test signal (T1) as the highest risk (R1) for all three frequencies. The respondents also ranked two tone test signal (T2) as second highest risk (R2), three tones test signal (T3) test signal as third highest risk (R3) and five tones (T5) test signal as lowest risk (R4) for all three frequencies.

One-sample proportion z test was carried out to test the following hypothesis for single tone (T1) test signal:

Ho: The proportion of responses of R1 for single tone was 50%

Ha: The proportion of responses of R1 for single tone was more than 50%

The one-sample proportion z test was also carried out to test the perceived risk for T2, T3, and T5 test signal. For two tones (T2) test signals:

Ho: The proportion of responses of R2 for two-tones

was 50%

Ha: The proportion of responses of R2 for two tones was more than 50%

For three tones (T3) test signal:

Ho: The proportion of responses of R3 for three tones was 50%

Ha: The proportion of responses of R3 for three tones was more than 50%

For five tones (T5) test signal:

Ho: The proportion of responses for R4 was 50%

Ha: The proportion of responses for R4 was more than 50%.

TABLE 6. Medium Risk: Perceived risk for various tones.

FREQ	Tone	Perceived Risk: R1 highest to R4 Lowest Risk				Most Preferred Tone for Medium Risk	z-value
		R1(%)	R2(%)	R3(%)	R4(%)		
400Hz	T1	45(90%)	5(10%)	0	0	2(4%)	5.657**
	T2	5(10%)	45(90%)	0	0	33(66%)	5.657**
	T3	0	0	50(100%)	0	15(30%)	7.071**
	T5	0	0	0	50(100%)	0%	7.071*
447Hz	T1	46(92%)	4(8%)	0	0	2(4%)	5.940**
	T2	4(8%)	46(90%)	0	0	34(68%)	5.657**
	T3	0	0	50(100%)	0	14(38%)	7.071**
	T5	0	0	0	50(100%)	0	7.071**
500Hz	T1	45(90%)	5(10%)	0	0	2(4%)	5.657**
	T2	5(10%)	45(90%)	0	0	33(66%)	5.657**
	T3	0	0	50(100%)	0	15(30%)	7.071**
	T5	0	0	0	50(100%)	0	7.071**

Notes: T1: Single tone; T2:2Tones; T3:3 tones; T5:5 tones test signals;

**p<0.01

Based on the ranking of the tone test signals at 447Hz, 92% of the respondents indicated that a single tone test signal had the highest risk, 90% ranked two tones test signal as the second highest risk, 100% indicated three tones test signal as the third highest risk and 100% ranked the four tones test signal as the lowest risk. It was also noted that for medium frequency test signal, 68% of respondents selected 2 tones test signal to be assigned to medium risk conditions.

The findings showed that the perceived urgency of the stimulus signal increased as the number of tones decreased. The findings of this simulation test indicated that the perceived risk associated with a stimulus was dependent on the number of tones, the single tone auditory signal had the highest risk, followed by two tones, three tones, and five tones had the lowest risk. Results also indicated that 66% of the respondents had identified two tones test signal to represent medium risk frequency.

The one-sample proportion z test found out that the proportion of responses for R1 was more than 50% for single tone test signal. The z-test was repeated for the proportion of perceived risk of R2, R3, and R4 for two, three and five tones respectively. All statistical tests were highly significant ($p<0.01$). These results provided empirical evidence that the perceived risk was associated with the number of tones, with single tone test signal being perceived as the highest risk regardless of frequency. The z-test provided statistical evidence that the perceived risk was associated with the number of tones, with single tone being perceived as the highest risk regardless of frequency.

High-risk classification. Table VII indicated the risk classification for the 635Hz, 708Hz and 800Hz test signals based on the number of tones assessed by clinicians. Table 7 represented the high-risk classification for 800Hz, 708Hz, and 635Hz tones. For 800Hz test signal, 94% of the respondents identified that single tone test signal had the highest risk, 94% ranked two tones test signal as the second highest risk, 100% identified three tones test signal as the third highest risk and 100% ranked the five tones test signal as the lowest risk. In Table 7, 88% of the respondents identified 800Hz tone with single tone test signal to be assigned for the high risk.

The outcomes of these formative usability simulation tests showed that the single tone test signal was perceived as the highest risk, followed by two tones, three tones, and five tones test signals. For both medium and high-risk frequencies assignment, the respondents perceived single tone test signal as high risk.

The one-sample proportion z test was carried out to test that the proportion of responses for R1 is more than 50% for a single tone test signal. The z-test was repeated for the proportion of perceived risk of R2, R3, and R4 for two, three and five tones respectively. All statistical tests were highly significant ($p<0.01$). These results provided empirical evidence that the perceived risk was associated with the number of tones, with single tone test signal being perceived as the highest risk regardless of frequency.

The findings showed that the perceived urgency increased as the number of stimulus tones decreased. The findings of this simulation test indicated that the perceived risk associated with a stimulus was dependent on the number of tones. Single tone had the highest risk, followed by two tones, three tones, and five tones had the lowest risk applies for both medium and high frequencies samples. These findings differ from standard IEC 60601-1-8:2006 recommendation which specified five tone signals for high, three tone signals for medium and two tones for low priority alarm signals.

TABLE 7. High Risk: Perceived risk for various tones.

FREQ	Tone	Perceived Risk:R1 highest to R4 Lowest Risk				Most Preferred Tone for High Risk	
		R1	R2	R3	R4		z-value
635Hz	T1	46(92%)	4(8%)	0	0	42(84%)	5.940**
	T2	4(8%)	46(90%)	0	0	7(14%)	5.940**
	T3	0	0	50(100%)	0	1(2%)	7.071**
	T5	0	0	0	50(100%)	0	7.071**
708Hz	T1	47(94%)	3(6%)	0	0	44(88%)	6.223**
	T2	3(6%)	47(94%)	0	0	6(12%)	6.223**
	T3	0	0	50(100%)	0	0	7.071**
	T5	0	0	0	50(100%)	0	7.071**
800Hz	T1	47(94%)	3(6%)	0	0	44(88%)	6.223**
	T2	3(6%)	47(94%)	0	0	6(12%)	6.223**
	T3	0	0	50 (100%)	0	0	7.071**
	T5	0	0	0	50(100%)	0	7.071**

Notes: T1:Single tone;T2:2Tones;T3:3 tones; T5:5 tones

**p<0.01

IV. Findings and Discussion

A. Simulation A

From the tests, it could be concluded that most of the respondents were able to recognize the alarm sounds of most of the devices. Almost all the respondents were able to recognise the alarm sounds of the physiological patient monitor. However, the high percentage of respondents were unable to recognise the risk associated with the alarm sounds and were unable to differentiate the low, medium and high risk. The patient monitor indicated the highest percentage of respondents were unable to recognise the associated risk of the alarm sounds. Based on the outcomes, it could be concluded that the existing auditory alarm system did not provide enough information on the risk associated with the alarm conditions for the users to identify accurately. As such the existing alarm sounds were not easily be identified and differentiated according to the patient conditions. The lack of this information could lead towards delayed response time or alarms being ignored by the users.

B. Simulation B

From the simulation tests, it could be concluded that the respondents preferred 282Hz for low risk, 500Hz for medium risk and 800Hz for high risk. This was in line with the research studies which indicated the higher the frequency of an auditory signal, higher will be the perceived risk [4].

From the conducted learnability tests, it could be concluded that more than 90.0% of the clinicians could identify all three low alarm tone conditions clearly compared to medium risk and high-risk alarm tones. These findings concurred with studies by Sanderson et al. that low-risk alarms were easily remembered compared to high-risk alarms design based on IEC, [6]. Research also indicated that these signals were difficult to learn and distinguish [5,11]. Standards committee have identified the effectiveness of the IEC 60601-1-8 melodic alarms for alerting healthcare professionals to be potentially dangerous events [6,7,8,9]. These studies by Sanderson et al., Lacherez et al., Williams et al. used non-clinicians who had no experience in the studies, whereas Alexandra et al. used trained nurses in the studies. All these studies proposed that the melodies were not easy to learn and that some alarms were often confused with others. Sanderson et al. speculated that the use of mnemonics increased confusion. All these four relevant studies described in this paper strongly suggested that learnability and discriminability issues in the IEC 60601-1-8 melodic alarms should be addressed before being introduced into healthcare settings. Further, Alexandra et al. stated that no human factors studies or usability testing regarding these alarms were conducted before the initial publication of the standard in 2003 [7].

The respondent of the test proposed interburst interval less than the time recommended by IEC standard for low, medium and high risk. These results concurred with findings from other studies that the alarm signals based on IEC recommendations did not signify any urgency nor indicated the physiological condition of the patient [14,22,23]. As such this design had resulted in less than adequate alarm response [6,8,9,15,24,25]. In the simulation study, the findings indicated that the clinician proposed that the response time should be less than the time recommended by IEC standard. It also indicated that the interburst interval recommended did not signify the urgency of the alarm signals.

From the outcome of the simulation tests, it could be concluded that the auditory alarm signals that based on IEC 60601-1-8:2006 recommendations were not effective and were not acceptable to the users IEC60601-1-8:2006. Numerous studies have indicated that this standard did not address the design of the alarm sounds appropriately and has resulted in less effective alarm sounds [6,11,22,24,25]. The short tone melodies recommended by this standard were rather confusing [1,26]. Numerous reviews and concerns raised by Block et al. also indicated that the alarm signals based on recommendations by IEC60601-1-8:2006 were not acceptable and could be improved [4,27,28]. The failure of this design protocol based on this standard made it crucial and timely to propose a new user alarm interface design based on human factor engineering principles.

C. Simulation C

This test focused on the relationship between the number of tones and ranking of perception with an ordinal scale according to the risk associated with the psychoacoustic effect of sound.

The perceived risk associated with various numbers of tones were conducted for the following medium frequencies; 400Hz, 447Hz, and 500Hz.

Based on the ranking of the tones, for all the three frequencies, the respondents indicated that a single tone has the highest risk, two tones as the second highest risk, three tones as the third highest risk and the four tones as the lowest risk. It was also noted that majority of the respondents selected 2 tones to be assigned to medium risk conditions.

The outcomes of the simulation test and analysis of the high-frequency bands 663Hz, 708Hz, and 800Hz indicated that the single tone auditory signal is perceived as the highest risk, followed by two tones, three tones, and five tones. For the high-risk assignment, the respondents preferred a single tone.

The findings indicated that the number of tones is linked with urgency. The findings showed that the perceived urgency increases as the stimulus number of tones decreases. The findings of this simulation test indicated that the perceived risk associated with a stimulus is dependent on the number of tones and single tone auditory signal has the highest risk, followed by two tones, three tones, and five tones.

V. Conclusion

In the first simulation test, the users did not perceive urgency in the existing auditory alarm. The results of this study concur with the previous studies by Sanderson P. et al., Williams S.et al., Lacherez P. et al. and Alexandra N. et al. that auditory alarm based on IEC60601-1-8 are not effective [6,9,13,15].

In the second simulation test, new alarms were generated based on IEC60601-1-8 and the users identified suitable frequencies for low, medium and high alarm risk conditions. They were able to recognise the low, medium and high-risk conditions of the new auditory alarm but commented that the interburst t_b allocations are not suitable. These auditory signals did not indicate urgency.

As such there is a need to establish and quantify the relationship between changes in objective parameters of the alarm signal to the subjective perception of urgency. Further research on psychophysics needs to be undertaken to derive the techniques to establish urgency mapping in alarm designs. Studies by Edworthy et al. and Hellier et al. drew several important distinctions on urgency mapping such as the relationship between perceived urgency and the acoustic parameters of the sound that could be used to address urgency in auditory alarm design [29,30].

The third simulation test indicates that the hierarchy of perceived urgencies with different numbers of tones contradicts IEC60601-1-8 recommendations. The single tone was perceived to indicate the highest risk followed by two and three, four and five tones. This new finding needs to be further investigated with the users before introducing it in the design. This is to ensure the alarm signals do not have a startling effect and prolonged alarm signal do not cause alarm fatigue or habituation effect [31,32].

It can be concluded that the existing alarm and alarm based on IEC60601-1-8 standard are found not to be effective. This is because respondents of this study were unable to recognize the risk associated with the sound of the existing alarm. The respondents also preferred shorter interburst interval compared to the IEC60601-1-8 standard. On the other hand, this study found that the number of tones is linked with urgency, with single tone was perceived to indicate the highest risk. Further investigation of this finding can be the foundation of the development of new and more effective auditory alarm in ICU/CCU. One limitation in this study was the respondents for the three simulations were not the same individuals, which may cause inconsistencies in the result.

Acknowledgment

This work is not supported by any funding agencies.

References

[1] C. Meredith, and J. Edworthy, “Are there too many alarms in the intensive care unit? An overview of the problems,” *Journal of Advanced Nursing*, vol. 21, no. 1, Jan 1995, pp. 15-20.

[2] International Electrotechnical Commission, “Medical electrical equipment -- Part 1-8: General requirements for safety -- Collateral standard: General requirements, tests and guidance for alarm systems in medical electrical equipment and medical electrical systems,” *IEC std 60601-1-8:2003*, 2003.

[3] International Electrotechnical Commission, “Medical electrical equipment -- Part 1-8: General requirements for basic safety and essential performance -- Collateral standard: General requirements, tests and guidance for alarm systems in medical electrical equipment and medical electrical systems,” *IEC std 60601-1-8:2006*, 2006.

[4] International Electrotechnical Commission, “Medical electrical equipment - Part 1: General requirements for basic safety and essential performance,” *IEC std 60601-1:2005*, 2005.

[5] F. E. Block, J. D. Rouse, M. Hakala, and C. L. Thompson, “A proposed new set of alarm sounds which satisfy standards and rationale to encode source information,” *Journal of Clinical Monitoring and Computing*, vol. 16, no. 7, Sep 2000, pp. 541-546.

[6] P. M. Sanderson, A. Wee, and P. Lacherez, “Learnability and discriminability of melodic medical equipment alarms,” *Anaesthesia*, vol. 61, no. 2, Feb 2006, pp. 142-147.

[7] A. N. Wee, and P. M. Sanderson, “Do mnemonics help nurses learn melodic medical equipment alarms?,” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 50, no. 10, Oct 2006, pp. 1039-1043.

[8] P. Lacherez, E. Limin Seah, and P. Sanderson, “Overlapping melodic alarms are almost indiscriminable” *Human Factors*, vol. 49, no. 4, Aug 2007, pp. 637-645.

[9] S. Williams, and P. C. Beatty, “Measuring the performance of audible alarms for anaesthesia,” *Physiological Measurement*, vol. 26, no. 4, May 24, 2005, pp. 571.

[10] J. Edworthy, and E. Hellier, “Fewer but better auditory alarms will improve patient safety,” *BMJ Quality & Safety*, vol. 14, no. 3, Jun 1, 2005, pp. 212-215.

[11] J. Edworthy, and E. Hellier, “Alarms and human behaviour: implications for medical alarms,” *BJA: British Journal of Anaesthesia,* vol. 97, no. 1, May 12, 2006, pp. 12-17.

[12] J. R. Edworthy, R. McNeer, C. Bennett, and R. Dudaryk, “Updating an International Medical Device Standard: A Process for Audible Alarms,” *Proceedings of the International Symposium on Human Factors and Ergonomics in Health Care*, vol. 7, no. 1, Jun 2018, pp. 11-16.

[13] R. R. McNeer, J. Bohórquez, Ö. Özdamar, A. J. Varon, and P. Barach, “A new paradigm for the design of audible alarms that convey urgency information,” *Journal of Clinical Monitoring and Computing*, vol. 21, no. 6, Dec 2007, pp. 353-363.
T. A. Mondor, and G. A. Finley, “The perceived urgency of auditory warning alarms used in the hospital operating room is inappropriate,” *Canadian Journal of Anaesthesia*, vol. 50, no. 3, Mar 1, 2003, pp. 221-228.

[14] A. N. Wee, and P.M. Sanderson, “Are melodic medical equipment alarms easily learned?,” *Anesthesia & Analgesia*, vol. 106, no. 2, Feb 1, 2008, pp. 501-508.

[15] P. C. Beatty, and S. F. Beatty, “Anaesthetists' intentions to violate safety guidelines” *Anaesthesia*, vol. 59, no. 6, Jun 2004, pp. 528-540.

[16] K. J. Ruskin, and D. Hueske-Kraus, “Alarm fatigue: impacts on patient safety,” *Current Opinion in Anesthesiology*, vol. 28, no. 6, Dec 1, 2015, pp. 685-690.

[17] K. R. Johnson, J. I. Hagadorn, and D. W. Sink, “Alarm safety and alarm fatigue,” *Clinics in Perinatology*, vol. 44, no. 3, Sep 1, 2017, pp. 713-728.

[18] J. Edworthy J, and C. Baldwin, “Medical Audible Alarms and IEC 60601-1-8,” *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, vol. 60, no. 1, Sep 2016, pp. 634-635.

[19] S. Thangavelu, J. Yunus, E. Ifeachor, and J. Edworthy, “Improving Alarm Response in ICU/CCU,” *International Journal of Simulation -- Systems, Science & Technology*, vol. 16, no. 4, Aug 2015, pp. 1.1-1.9.

[20] J. R. Edworthy, R. R. McNeer, C. L. Bennett, R. Dudaryk, S. J. McDougall, J. J. Schlesinger, M. L. Bolton, J. D. Edworthy, E. Özcan, A. D. Boyd, and S. K. Reid, “Getting Better Hospital Alarm Sounds Into a Global Standard,” *Ergonomics in Design*, vol. 26, no. 4, Oct 2018, pp. 4-13.

[21] J. Edworthy, “Designing effective alarm sounds,” *Biomedical Instrumentation & Technology*, vol. 45, no. 4, Jul 2011, pp. 290-294.

[22] G. A. Finley, and A. J. Cohen, “Perceived urgency and the anaesthetist: responses to common operating room monitor alarms,” *Canadian Journal of Anaesthesia*, vol. 38, no. 8, Nov 1, 1991, pp. 958-964.

[23] F. E. Block, ““For if the trumpet give an uncertain sound, who shall prepare himself to the battle?”(I Corinthians 14: 8, KJV),” *Anesthesia & Analgesia*, vol. 106, no. 2, Feb 2008, pp. 357-359.

[24] J. Edworthy, “Medical audible alarms: a review,” *Journal of the American Medical Informatics Association*, vol. 20, no. 3, Oct 25, 2012, pp. 584-589.

[25] S. Siebig, S. Kuhls, M. Imhoff, J. Langgartner, M. Reng, J. Schölmerich, U. Gather, and C. E. Wrede, “Collection of annotated data in a clinical validation study for alarm algorithms in intensive care—a methodologic framework,” *Journal of Critical Care*, vol. 25, no. 1, Mar 2010, pp. 128-135.

[26] J. Bliss, M. Dunn, and B. S. Fuller, “Reversal of the cry-wolf effect: An investigation of two methods to increase alarm response rates,” *Perceptual and Motor Skills*, vol. 80, no. 3_suppl, Jun 1995, pp. 1231-1242.

[27] F. E. Block, L. Nuutinen, and B. Ballast, “Optimization of alarms: a study on alarm limits, alarm sounds, and false alarms, intended to reduce annoyance,” *Journal of Clinical Monitoring and Computing*, vol. 15, no. 2, Feb 1999, pp. 75-83.

[28] J. Edworthy, S. Loxley, and I. Dennis, “Improving auditory warning design: Relationship between warning sound parameters and perceived urgency,” Human Factors, vol. 33, no. 2, Apr 1991, pp. 205-231.

[29] E. J. Hellier, J. Edworthy, and I. A. Dennis, “Improving auditory warning design: Quantifying and predicting the effects of different warning parameters on perceived urgency,” *Human Factors*, vol. 35, no. 4, Dec 1993, pp. 693-706.

[30] W. M. Hartmann, “Signals, sound, and sensation,” *Springer Science & Business Media*, Sep 14, 2004.

[31] Y. Yuki, “Alarm system optimization for increasing operations productivity,” *ISA Transactions*, vol. 41, no. 3, Jul 2002, pp. 383-387.