

2018-07

# Auditory icon alarms are more accurately and quickly identified than standard melodic alarms in a simulated clinical setting

Mcneer, R

<http://hdl.handle.net/10026.1/11275>

---

10.1097/aln.0000000000002234

Anesthesiology

Lippincott, Williams & Wilkins

---

*All content in PEARL is protected by copyright law. Author manuscripts are made available in accordance with publisher policies. Please cite only the published version using the details provided on the item record or document. In the absence of an open licence (e.g. Creative Commons), permissions for further reuse of content should be sought from the publisher or author.*

Auditory icon alarms are more accurately and quickly identified than current standard melodic alarms in a simulated clinical setting.

**1. Richard R. McNeer, M.D., Ph.D.**

**Title:** Associate Professor of Clinical Anesthesiology

**Affiliation:** Department of Anesthesiology, University of Miami Miller School of Medicine

**Role:** Richard McNeer helped design the study, conduct the study, analyse the data, perform the statistical analysis, and write the manuscript

**Attestation:** Richard McNeer has seen the original study data, reviewed the analysis of the data, approved the final manuscript, and is the author responsible for archiving the study files.

**2. Danielle Bodzin Horn, M.D.**

**Title:** Anesthesiology Resident

**Affiliation:** Department of Anesthesiology, Jackson Memorial Hospital, University of Miami Miller School of Medicine

**Role:** Danielle Bodzin Horn helped conduct the study, analyse the data, and write the manuscript

**Attestation:** Danielle Bodzin Horn has seen the original study data, reviewed the analysis of the data, and approved the final manuscript.

**3. Christopher L. Bennett, Ph.D.**

**Title:** Assistant Professor

**Affiliation:** Music Engineering Technology, Frost School of Music, University of Miami

**Role:** Christopher Bennett helped design the study, conduct the study, analyze the data, and write the manuscript

**Attestation:** Christopher Bennett has seen the original study data, reviewed the analysis of the data, and approved the final manuscript.

#### **4. Judy Reed Edworthy, PhD**

**Title:** Professor of Applied Psychology

**Affiliation:** Cognition Institute, Plymouth University, UK

**Role:** Judy Edworthy helped design the study, conduct the study, analyse the data, perform the statistical analysis, and write the manuscript

**Attestation:** Judy Edworthy has seen the original study data, reviewed the analysis of the data, approved the final manuscript, and is the author responsible for archiving the study files.

#### **5. Roman Dudaryk, M.D.**

**Title:** Associate Professor of Clinical Anesthesiology

**Affiliation:** Department of Anesthesiology, University of Miami Miller School of Medicine

**Role:** Roman Dudaryk helped design the study, conduct the study, analyse the data, and write the manuscript

**Attestation:** Roman Dudaryk has seen the original study data, reviewed the analysis of the data, and approved the final manuscript.

**Institution:** Department of Anesthesiology, University of Miami Miller School of Medicine

**Corresponding Author:**

Richard R. McNeer, M.D., Ph.D.

Department of Anesthesiology, University of Miami Miller School of Medicine

Division of Trauma and Critical Care, Department of Anesthesiology

PO Box 016370 (M820), Miami, FL 33101

Phone: 305 585 1191

FAX: 305 545 6195

Email: mcneer@miami.edu

**Funding:** Association for the Advancement of Medical Instrumentation

**Conflicts of Interest:** none

**Number of words in Abstract:** 249

**Number of words in Introduction:** 500

**Number of words in Discussion:** 1399

Complete word count: 5625

Number of tables: 5

Number of figures: 3

Number of supplemental digital content files: 2

**Abbreviated Title:** Auditory Icon Alarm Effectiveness in Simulated Clinical Settings

## ABSTRACT

**Introduction.** Current standard audible medical alarms are difficult to learn and distinguish from one another. Auditory icons represent a new type of alarm that has been shown to be easier to learn and identify in laboratory settings by lay-subjects. In this study, we test the hypothesis that icon alarms are easier to learn and identify than standard alarms by anesthesia providers in a simulated clinical setting.

**Methods.** Twenty anesthesia providers were assigned to standard or icon groups. Experiments were conducted in a simulated intensive care unit. After a brief group-specific alarm orientation, subjects identified patient-associated alarm sounds during the simulation and logged responses via a tablet computer. Each subject participated in the simulation twice and was exposed to 32 alarm annunciations. Primary outcome measures were response accuracy and response times. Secondary outcomes included assessments of perceived fatigue and task load.

**Results.** Overall accuracy rate in the standard alarm group was 43% (mean) and in the icon group was 88% (mean). Subjects in the icon group were 26.1 (odds ratio [98.75% CI, 8.4 to 81.5;  $P < 0.001$ ] times more likely to correctly identify an alarm. Response times in the icon group were shorter than in the standard alarm group (12 vs 15 seconds, difference 3 seconds [98.75 %CI 1 to 5;  $P < 0.001$ ]).

**Conclusions.** Under our simulated conditions, anesthesia providers more correctly and quickly identified icon alarms than standard alarms. Subjects were more likely to perceive higher fatigue and task load when using IEC alarms than icon alarms.

## INTRODUCTION

Audible alarms are essential sounds within the clinical soundscape important for patient monitoring. They play a vital role in patient safety by alerting caregivers of patient or medical equipment state changes. The International Electrotechnical Commission (IEC) published a standard in 2003 (most recently revised in 2012) known as IEC 60601, which specifies basic safety of medical electrical equipment and governs almost all medical equipment across the globe.<sup>1</sup> Parts 1-8 of the standard specify performance requirements for alarm sounds and systems and contain an example set of auditory alarms that complies with the normative portions of the standard (referred to here as IEC alarms). However, these IEC alarms have been shown to function poorly by researchers in the fields of human factors and psychology.<sup>2-5</sup> Each alarm sound is a distinct melody meant to facilitate appreciation of the alarm meaning or etiology. Although the melodic contour varies across the different alarms in the IEC alarm set, other aspects of composition and instrumentation are fixed, including timbre/pitch, key, duration, rhythm and tempo leading to very little acoustical variation, or heterogeneity, within the set.<sup>1</sup> Several studies have demonstrated that the IEC alarms are therefore difficult to learn (especially in the musically uninitiated) and alarms within the set are easily confused with one another<sup>2-5</sup>—factors potentially contributing to alarm fatigue, and certainly the cause of unnecessary confusion.<sup>6</sup> Device manufacturers are not required to adopt the IEC standard and can implement proprietary alarm sounds that at least demonstrate equivalence.<sup>1</sup> However, no clear precedent has been established on how to best test the effectiveness of novel alarm sounds.

Development of an updated version of IEC 60601-1-8 is currently underway, and ‘Auditory icons’<sup>6</sup> (referred to here as icon alarms) are considered for replacement of current IEC alarms. Icon alarms are commonplace and acoustically complex sounds that mimic the underlying meanings they are meant to represent. For example, the auditory icon alarm for ‘File deletion’ on a personal computer is typically designed to sound like the crumpling-up of a waste paper. Conceptually similar icon designs are easily relatable to medical alarms (Table 1.). Relative to the abstract and tonally similar IEC alarms, icon alarms were found to be easier to learn and discriminate when studied in nonclinical, computer-based settings using lay-, nonclinical-participants.<sup>6, 7</sup> Additionally, icon alarms were easier to localize in an experimental setting.<sup>6</sup> On the basis of these results, the IEC Alarms Joint Working Group, which is in a position to recommend the specific details of any update to the standard, has called for the development and testing of a set of icon alarms to be considered for adoption into the standard (personal communication from Dave Osborn, chair of IEC Alarms Joint Working Group).

In this report, we describe methodology for testing clinician responsiveness to alarms within a simulated clinical setting as a measure of alarm effectiveness. We specifically test the hypothesis that icon alarms are easier to learn and identify than the IEC standard alarms in a simulated intensive care unit (ICU) using anesthesia providers as subjects.

## **METHODS**

This study was approved by institutional review boards at the University of Miami-Miller School of Medicine and the Jackson Health System. Written informed consent was obtained from all subjects prior to participating in the study.

### **Study Design Overview and Outcome Measures**

To evaluate the relative effectiveness of IEC and icon alarms, we used a simulated, 2-bed, intensive care unit (ICU). The study had a between-subjects factor represented by 'Group' (IEC or icon alarm set exposure) and a within-subjects factor represented by replicated measure—there were two sessions (refer to Figure 1). This mixed-design allowed us to assess the effects of group and repeated exposure on subject performance. Two primary outcomes were chosen to assess alarm effectiveness: Alarm identification accuracy (binary response) and time to respond to an alarm annunciation (response time). In addition to the primary outcome measures, we studied a secondary set of outcome measures that included subject perception of task load and fatigue using methodology described previously.<sup>8</sup> Experiments were conducted from October 13, 2016 to December 16, 2016 in the early afternoon period.

### **Icon Alarm Set Design**

Icons alarms are real-world sounds that are somehow associated with the process that they represent. The advantage of icons is that they are immediately intuitive, even upon first audition, and therefore should be easy to learn. This derives from the design principle of directly conveying a concept instead of an encoded message, the latter



being the case with current IEC alarms. With medical alarms, the conveyed “concepts” derive from the category of alarm, and the IEC specifies eight such categories: cardiovascular, ventilation, perfusion, drug administration, oxygenation, temperature, equipment failure, and a general “catch all”. For this study we focused on an example set of icon alarms described in Table 1 (also refer to slide show presentation, Supplemental Digital Content 1 and 2, with embedded audio of icon and IEC alarms used in this study) all of which were studied previously in a laboratory setting except for the ‘general alarm’. In order to standardize the perceptual loudness within and between the alarm sets, each individual alarm was processed to maximize audibility through level dynamic range compression and normalization. In addition, each alarm was embedded (within the first second) with an auditory pointer comprising three notes followed by two notes after a gap, with the entire sequence repeated after a longer gap. This pattern is a rhythmic element that is specified in current standard alarms and serve to draw the attention of the user to the presentation of the alarm.<sup>9</sup>

### **Intensive Care Unit Simulator Setup**

The simulated intensive care unit (ICU) consisted of 2 beds, each with a simulated patient (mannequin). A custom multimedia graphical user interface (GUI) described in detail elsewhere<sup>8, 10</sup> was associated with each patient and placed adjacent to left of bedside from patient perspective. For this study, the GUI was modified to add touchscreen functionality and was installed on two tablet computers (Microsoft Surface 3), which mimicked touch functionality found on most modern patient monitor displays. The GUI functioned to visually display simulated patient vital sign and ventilator parameters, to sonify a variable pitch pulse oximeter auditory display, and to annunciate

audible alarms when alarm thresholds were reached based on static simulation scripts (Table 2). The GUI also allowed subjects to respond via the touchscreen to alarm annunciations, and, therefore, functioned to log timestamp and alarm response type which was needed to determine the primary outcome measures of response time and binary response, respectively. The simulated environment contained items typically found in ICUs including infusion pumps, IV poles, associated tubing, stretchers, a crash cart, etc. Devices not available to us such as the dialysis and extra-corporeal membrane oxygenation (ECMO) machines were indicated using written placards. At the foot of each bed was a mobile desk with a paper chart for the patient. The simulated ICU is similar acoustically to the actual clinical settings at our institution and is capable of playing a clinical ‘background’ soundscape along with script-specific alarm annunciation and pulse oximetry display during experiments.<sup>8, 11</sup> For the current study, no background soundscape was utilized and all simulation sounds (i.e., pulse oximeter display and alarm sounds) were generated by the GUI.

## **Experimental Procedure**

Subjects consisted of clinical anesthesia (CA) residents and anesthesia attending physicians who were recruited the day of scheduled experiments and randomized to IEC or icon groups based on order of arrival to the simulation laboratory (odd-IEC and even-icon). Order of arrival depended on ad hoc provision of relief of subjects from clinical duty in the operating room. This relief task was implemented by an individual(s) not affiliated with the study. Subjects were asked to review an instructional multimedia presentation on a computer in a simulation staging area that detailed the session instructions, presented brief medical histories of the two simulated patients, and

provided group specific exposure/training to alarm sounds (see Supplemental Digital Content 1 and 2, with embedded audio of icon and IEC alarms used in this study). The presentations were subject-paced and took 5-10 minutes to complete. Then subjects were escorted to the simulated ICU and asked to watch over two patients while a clinician actor went to find supplies to place an arterial catheter. Subjects had access to each patient's chart at the foot of the bed. The simulation session lasted 20 min during which two static scripts (1 per patient) were synchronized and run simultaneously (Table 2). A total of 16 alarms were annunciated—each alarm category was represented twice per session, once per patient (Figure 1). At the conclusion of Session 1, arrangements were made for subjects to return about 1 week later to participate in a second session. As with Session 1, subjects were asked to review the same group-specific multimedia presentation before starting Session 2. For Session 2, the same simulated patients were represented, but the progress notes were updated to reflect that about one week had passed. Sessions 1 and 2 both followed the same simulation scripts. At the completion of Session 2, subjects completed two validated psychometric instruments and an exit survey (See Subjective Instruments). Each alarm sound (either IEC or icon) was annunciated a total of 4 times per subject over the course of 2 sessions. Each subject was, therefore, exposed to a total of 32 alarm annunciations during the experiment (Figure 1).

### **Subjective Instruments**

At the end of Session 2, subjects completed 2 validated psychometric instruments: the Swedish Occupation Fatigue Inventory (SOFI)<sup>8, 12</sup> and the NASA Task Load Assessment

Questionnaire (NASA-TLX).<sup>8, 13</sup> Subjects also completed an exit survey consisting of 6 questions which assess usability of the alarms.

## **Power Analysis**

In preparation for this study, a power analysis was performed with G\*Power 3.1.9.2 (test family: 'F-tests', statistical test: 'ANOVA Repeated measures, within-between interactions'). Previously, results of a repeated-measures, laboratory-based study comparing identifiability of 5 alarm sets (including IEC and icon alarm sets) reported effect sizes in terms of partial eta squared ( $\eta_p^2$ ) which represents the fraction of variation in observed outcome that is attributable to the independent variable(s) and ranges from 0 (no effect) to 1. That study showed a large main effect size for group ( $\eta_p^2 = 0.622$ ) and a medium interaction effect size ( $\eta_p^2 = 0.193$ ).<sup>6</sup> We conservatively chose an expected medium effect ( $\eta_p^2 = 0.2$ ) of 'group' or 'session' on the alarm response accuracy within the entire set (IEC or icon). We also used this expected effect size in consideration of the effect of group (IEC vs icon) on alarm reaction time averaged for each alarm set. Using the Bonferroni approach, alpha level was adjusted considering four measured outcomes (the measured effects of group and session on response time and binary response) to 0.0125(0.05/4). Power was set at 0.90 and correlation among repeated measures was conservatively set at 0.5. Based on this, a sample size of 20 was calculated to be sufficient.

**Statistical analysis.** All statistical analyses were conducted using IBM SPSS Statistics software (version 24). To analyze primary outcome results, a generalized linear mixed model (GLMM) was selected for the following reasons:<sup>14</sup> i. GLMM is able to account for

nested and hierarchical data (see Figure 1); ii. GLMM can consider dependent variables that are parametric (e.g., response time) and binary; iii. GLMM can account for both fixed and random effects; iv. Compared to other statistical tests of repeated measures, incomplete (missing) data pertaining to a subject are not excluded from analysis. Since both Group and Session were factors of interest, and because Session was a replicated repeated measure (each subject remained in same Group for both sessions), both factors were considered to be fixed effects. Subjects were set as a random effect. A fixed intercept and a random intercept were specified. A Diagonal repeated covariance type was selected for analysis and is the default used in GLMM with repeated measures by SPSS. This model specification was used to conduct two separate statistical analyses: one measured the effects of Group and Session on binary response, and the other measured the effects of Group and Session on response time. Reporting of results follow published suggested guidelines.<sup>14</sup> To reduce the risk of Type 1 error, significance was adjusted as in the power analysis to  $\alpha = 0.0125$ . Additionally, GLMM can reduce Type 1 error by its accounting of random effects.<sup>14</sup> An important disadvantage of GLMM is that common measures of effect size (e.g., Cohen's  $d$  and  $\eta_p^2$ ) are not obtainable. Therefore, effect sizes are reported as follows. For binary responses, effect size is reported as odds ratio accompanied by 98.75% confidence intervals as is customary when reporting logistic results. For response time, an unstandardized effect size is reported as the difference between the means accompanied by 98.75% confidence intervals.

Secondary outcome measures were analyzed using descriptive statistics. Individual items from the SOFI, NASA-TLX and exit survey are reported as mean values

and 95% confidence intervals. P-values are also reported for pairwise comparisons. Cronbach's alpha was calculated to assess internal consistency for SOFI and NASA-TLX (i.e., that all items in each instrument measured the same construct).

## RESULTS

Twenty subjects consisting of 17 clinical anesthesia (CA) residents (7 CA-1's, 2 CA-2's, and 8 CA-3's) and 3 attending physicians participated in the study. Over the course of the entire experiment, 640 alarms (cases) were annunciated—320 per alarm group. Data for 15 (2.3%) cases were missing, and these were attributed to subjects failing to log responses. These cases occurred during Session 1 in the IEC group. There were no missing data for the icon group. Failed responses were counted as 'incorrect' when assessing response accuracy and were not used to calculate response times. Therefore, in the GLMM analyses, 640 and 625 cases were processed to assess response accuracy and response time, respectively.

### Primary Outcomes

Alarm identification accuracy varied with alarm category for each group. For the IEC alarms, 'general alarm' (61%), 'oxygenation' (77%) and 'cardiovascular' (70%) were associated with the highest accuracy rates while 'perfusion' (17%) and 'power failure' (19%) were associated with the lowest. Six of the 8 icon alarms were identified correctly 80 or more percent of the time, while 'power failure' was associated with the lowest accuracy rate (69%) of the group (Figure 2). Overall, subjects identified icon alarms more accurately and quickly than IEC alarms (Table 3), and an effect of training level on subject performance was not observed (Figure 3). In particular, subjects in the icon group were 26.1 (8.5 to 81.5) times more likely to respond correctly to alarm annunciations ( $P < 0.001$ ) and responded sooner by 3 (1 to 5) seconds than subjects in the IEC group ( $P < 0.001$ ) (Table 4). Most subjects (7 out of 10 for each group) performed better in Session 2 than in Session 1 irrespective of alarm grouping (Figure

3). Overall, subjects were 2.2 (1.3 to 3.7) times more likely to respond correctly in Session 2 than in Session 1 ( $P < 0.001$ ) and response times were 2 (1 to 3) seconds quicker ( $P < 0.001$ ) (Table 4).

## **Secondary Outcomes**

Reliability of test results as measured by Cronbach's alpha suggest that SOFI and NASA-TLX instruments each measured a single construct, i.e., fatigue ( $\alpha = 0.723$ ) and task load ( $\alpha = 0.798$ ), respectively. Relative to subjects in the icon group, subjects in the IEC group reported a higher score in the SOFI questionnaire for 'Lack of Energy' (3 [1 to 4] versus 1 [0 to 2];  $P = 0.028$ ). Subjects in the IEC group reported experiencing higher levels of task load on the NASA-TLX questionnaire along all items, especially for 'Performance' (12 [9 to 16] versus 6 [4 to 8]; lower is better;  $P = 0.003$ ) and 'Frustration' (14 [10 to 19] versus 7 [2 to 12];  $P = 0.016$ ). Results of the exit survey suggest subjects in the icon group found it easier to work out an alarm's meaning than subjects in the IEC group (5 [4 to 6] versus 2 [1 to 3];  $P < 0.001$ ) and the same group found the alarm sounds more helpful (5 [5 to 6] versus 4 [2 to 5];  $P = 0.016$ ) (Table 5).

## **DISCUSSION**

After a brief exposure to alarm sounds, anesthesia providers identified icon alarms more accurately than IEC alarms during clinical simulation. This indicates that icon alarms were easier to learn which corroborates results obtained previously from laboratory-based experiments that used nonclinical subjects.<sup>6, 7</sup> Our subjects also identified icon alarms more quickly than IEC alarms. Though it is unclear if the effect observed here is clinically relevant, we feel, on principle, that any decrease in time required to correctly



detect reversible adverse events is desirable in terms of patient safety. Secondly, we observed that subjects perceived less task load and fatigue when using icon alarms and found them more useful than IEC alarms. Collectively, these results suggest that the set of icon alarms tested here as an example would not only be more effective, but could be less likely to contribute to alarm fatigue than the current IEC alarm set in real-world clinical settings.

In our practice, and most likely in general, clinicians are not given formal introduction to and training in the use of medical alarms. In a previous study, formal training of subjects to learn IEC alarms resulted in accuracy rates between 10 and 61%.<sup>15</sup> Ideally, alarm sounds should require minimal, if any training, before effective implementation in clinical settings. We expect this goal to be more attainable with icon alarms because they more intuitively encode alarm meaning. After subjecting our subjects to a brief 5-10 minute orientation period, we observed overall accuracy rates for the icon alarms of between 68% (power failure) and 100% (general alarm). In comparison, our overall accuracy for the IEC alarms ranged from 15% (perfusion) to 75% (oxygenation). These results demonstrate that a brief informal orientation may be sufficient to prepare clinicians for use of icon alarms. Although we observed a modest improvement in subject performance after a second orientation period for both IEC and icon alarms, it seems that additional and more regimented training sessions would be required for the current IEC set if accuracy rates are to approach those of the icon set.

Our findings are based on comparison between alarm sets (i.e., IEC vs icon). However, some icon alarms tested here were easier to identify than others (refer to Figure 2) indicating that there is scope for improvement in individual alarm function.

Some of the alarm categories may lend themselves to more obvious metaphors than others. Additionally, the effectiveness of an alarm depends on the other alarms with which it is heard.<sup>16</sup> For example, a ‘watery’ sound will be easier to identify if it is the only one in the set, and harder to identify if there are two or more ‘watery’ sounds also within the set. This study was not designed to detect and characterize these types of intra-set interactions.

Manufacturers are able to use proprietary alarms as long as they conform to the normative portions of the standard which specify sequences of tones and demonstrate that the alarms are as effective as the IEC alarms.<sup>1</sup> We were unable to find literature surveying audible alarms on medical devices, but we have anecdotally observed that common patient monitor systems and ventilator/workstations are equipped with proprietary alarm sounds that are tonal in nature like the current IEC alarms. As a result, the effectiveness of proprietary alarms is likely to be closer to that of the IEC alarms than to the icon alarms. Hence, we chose the IEC alarm set as a control for this study, though it is possible that some proprietary alarms are more effective than IEC alarms in clinical practice. At our institution, few devices use the IEC alarms, and this also informed our selection for control because subjects could be considered to be relatively naïve to the IEC alarms thus putting IEC and icons sets on more equal footing with regard to previous alarm exposure history. Additionally, the design problems surrounding IEC alarms are well described and have helped inform the rationale and design of icon alarms. An expectation is that if the next standard alarms are demonstrably more effective than the current IEC alarms, manufacturers will be more likely to implement them upon adoption into the standard. Alternatively, manufacturers

may continue development and implementation of novel proprietary alarms, however, considering the higher mark set here with icon alarms to meet equivalency, this scenario seems less likely.

Clinical environments are notoriously noisy.<sup>8, 17-19</sup> Therefore, in addition to learnability, the ability of an alarm sound to be heard in the presence of background noise (audibility) is an important criterion for assessing its adoption into a new standard. We intentionally conducted our simulations in the absence of background noise, though we acknowledge that in clinical practice there are likely to be interactions between learnability and audibility. Because the work presented here is an early step toward updating the global alarms standard, it is important to document the systematic testing of candidate alarms. Audibility of icon alarms and the relationship between audibility and identifiability in the presence of background noise remain to be characterized experimentally.

## **Limitations**

In addition to those already mentioned, there were several additional limitations inherent in this study. Although it was designed to be more ‘clinically’ realistic than the previously reported laboratory studies,<sup>6, 7, 16</sup> this study, nonetheless, only approximated an ICU setting. The simulated patients were chosen to be representative of typical critically-ill patients, however, vital signs and machine state changes followed static scripts, and subjects were not required to intervene or interact with patients or simulator props and resources. Subjects were told that their clinical performance would not be evaluated, and it is probable that some adopted a mindset consistent with completing the narrow task of identifying the alarm sounds as quickly and accurately as possible. A more

realistic experience could require subjects to complete clinically relevant distractor tasks in addition to alarm identification. An additional limitation is that physicians but not nurses were used as subjects, though it is generally recognized that nurses endure the most exposure to alarm sounds and have the highest risk of alarm fatigue.<sup>20, 21</sup> We acknowledge this as a significant limitation of the current study. Since we focused on alarm perception rather than on clinical response to underlying etiology (e.g., interpretation, diagnosis and intervention), disparity in subject performance based on training level is less likely to have been a factor and was not observed in our data (refer to Figure 3). Additionally, to date, lay-subjects and anesthesia providers appear to perform similarly when comparing IEC and icon alarms. Nonetheless, we cannot be certain if our results are extrapolatable to real-world clinical scenarios.

Another potential limitation is our decision to use alarm categories as classified in IEC 60601-1-8 which were based on work by Kerr.<sup>22</sup> There is increasing discussion of a need to modify the alarm categories, and audible alarm function may depend partly on the classification system.<sup>23</sup> In the current study, we used the standard categories out of necessity as the focus was to compare icon alarms to the IEC alarms. We believe we have definitively demonstrated that icon alarms which were the front-runners in the previous laboratory-based studies function better than the IEC alarms in a simulated ICU. Future investigation of icon alarms, we propose do not need to include an IEC alarm set arm and can concentrate on improving icon alarm design by comparing additional versions of icon alarms. This approach need not be constrained to a certain alarm classification system leaving open the possibility of parallel effort to update and improve both alarm sounds and the classification system. Additional refinements to

alarm sets must also incorporate input from many stakeholders beyond the designers and end-users of alarms, including manufacturers, regulatory organizations (e.g., Joint Commission and Occupational Safety and Health Administration), industry groups (e.g., Association for the Advancement of Medical Instrumentation), and standards organizations (e.g., IEC/ISO, American National Standards Institute). Lastly, future alarm sets should attempt to comply with international guidelines that govern the sound level within clinical environments, such as those set by the World Health Organization<sup>24</sup> and US Environmental Protection Agency.

## **CONCLUSION**

Relative to the IEC melodic alarms, auditory icon alarms were easier to learn and more quickly identified in a simulated clinical environment. Subjects were more likely to perceive higher fatigue and task load when using IEC alarms than icon alarms.

## References

1. Anonymous **IEC 60601-1-8:2006**

**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.** ; 2015

2. Wee AN, Sanderson PM: Are melodic medical equipment alarms easily learned? *Anesth Analg* 2008; 106: 501,8, table of contents

3. Lacherez P, Seah EL, Sanderson P: Overlapping melodic alarms are almost indiscriminable. *Hum Factors* 2007; 49: 637-45

4. Sanderson PM, Wee A, Lacherez P: Learnability and discriminability of melodic medical equipment alarms. *Anaesthesia* 2006; 61: 142-7

5. Atyeo J, Sanderson PM: Comparison of the identification and ease of use of two alarm sound sets by critical and acute care nurses with little or no music training: a laboratory study. *Anaesthesia* 2015; 70: 818-27

6. Edworthy J, Reid S, McDougall S, Edworthy J, Hall S, Bennett D, Khan J, Pye E: The Recognizability and Localizability of Auditory Alarms: Setting Global Medical Device Standards. *Hum Factors* 2017; 59: 1108-27

7. Edworthy J, Page R, Hibbard A, Kyle S, Ratnage P, Claydon S: Learning three sets of alarms for the same medical functions: a perspective on the difficulty of learning alarms specified in an international standard. *Appl Ergon* 2014; 45: 1291-6
8. McNeer RR, Bennett CL, Dudaryk R: Intraoperative Noise Increases Perceived Task Load and Fatigue in Anesthesiology Residents: A Simulation-Based Study. *Anesth Analg* 2016; 122: 512-25
9. Lorenz RH, Schuster H: Auditory Pointers, Works in Audio and Music Technology, 1st edition. Edited by Berndt A. , TUDpress, 2015, pp 1-30
10. McNeer RR, Dudaryk R, Nedeff NB, Bennett CL: Development and Testing of Screen-Based and Psychometric Instruments for Assessing Resident Performance in an Operating Room Simulator. *Anesthesiol Res Pract* 2016; 2016: 9348478
11. Bennett CL, Dudaryk R, Ayers AL, McNeer RR: Simulating environmental and psychological acoustic factors of the operating room. *J Acoust Soc Am* 2015; 138: 3855-63
12. Ahsberg E, Gamberale F, Gustafsson K: Perceived fatigue after mental work: an experimental evaluation of a fatigue inventory. *Ergonomics* 2000; 43: 252-68
13. Carswell CM, Lio CH, Grant R, Klein MI, Clarke D, Seales WB, Strup S: Hands-free administration of subjective workload scales: acceptability in a surgical training environment. *Appl Ergon* 2010; 42: 138-45

14. Casals M, Girabent-Farres M, Carrasco JL: Methodological quality and reporting of generalized linear mixed models in clinical medicine (2000-2012): a systematic review. *PLoS One* 2014; 9: e112653
15. Williams S, Beatty PC: Measuring the performance of audible alarms for anaesthesia. *Physiol Meas* 2005; 26: 571-81
16. Edworthy J, Hellier E, Titchener K, Naweed A, Roels R: Heterogeneity in auditory alarm sets makes them easier to learn. *International Journal of Industrial Ergonomics* 2011; 41: 136-46
17. Busch-Vishniac IJ, West JE, Barnhill C, Hunter T, Orellana D, Chivukula R: Noise levels in Johns Hopkins Hospital. *J Acoust Soc Am* 2005; 118: 3629-45
18. Kracht JM, Busch-Vishniac IJ, West JE: Noise in the operating rooms of Johns Hopkins Hospital. *J Acoust Soc Am* 2007; 121: 2673-80
19. Ginsberg SH, Pantin E, Kraidin J, Solina A, Panjwani S, Yang G: Noise levels in modern operating rooms during surgery. *J Cardiothorac Vasc Anesth* 2013; 27: 528-30
20. Korniewicz DM, Clark T, David Y: A national online survey on the effectiveness of clinical alarms. *Am J Crit Care* 2008; 17: 36-41
21. Honan L, Funk M, Maynard M, Fahs D, Clark JT, David Y: Nurses' Perspectives on Clinical Alarms. *Am J Crit Care* 2015; 24: 387-95



22. Kerr JH: Symposium on anaesthetic equipment. Warning devices. Br J Anaesth 1985; 57: 696-708

23. Edworthy JR, Schlesinger JJ, McNeer RR, Kristensen MS, Bennett CL: Classifying Alarms: Seeking Durability, Credibility, Consistency, and Simplicity. Biomed Instrum Technol 2017; 51: 50-7

24. Falk SA, Woods NF: Hospital noise--levels and potential health hazards. N Engl J Med 1973; 289: 774-81

## Figures

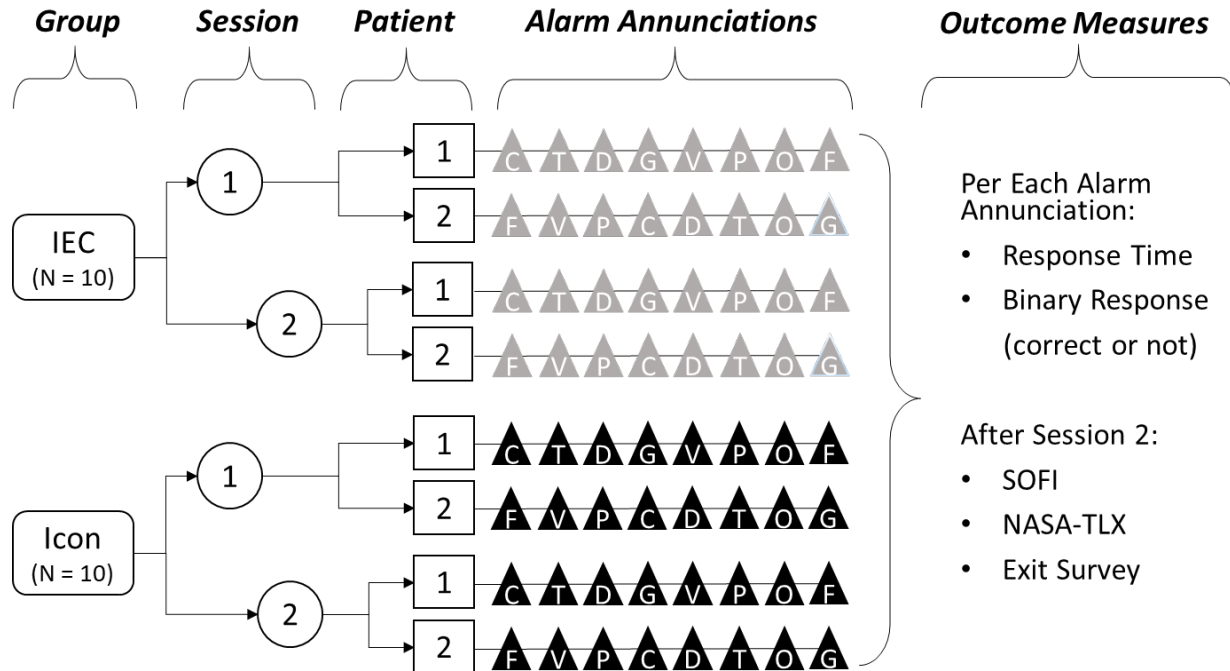


Figure 1. Schematic showing experimental design and nested data structure. During experiments, each subject was exposed to a total of 32 alarm annunciations, and a maximum of 640 cases (20 subjects X 32 alarms) were obtained. Gray and black triangles represent IEC and Icons alarm annunciations, respectively. 'O' = Oxygenation, 'V' = Ventilation, 'C' = Cardiovascular, 'P' = Artificial Perfusion, 'T' = Temperature, 'D' = Drug Administration, 'F' = Equipment Failure, and 'G' = General Alarm. Refer to Table 2 for exact times of alarm annunciation.

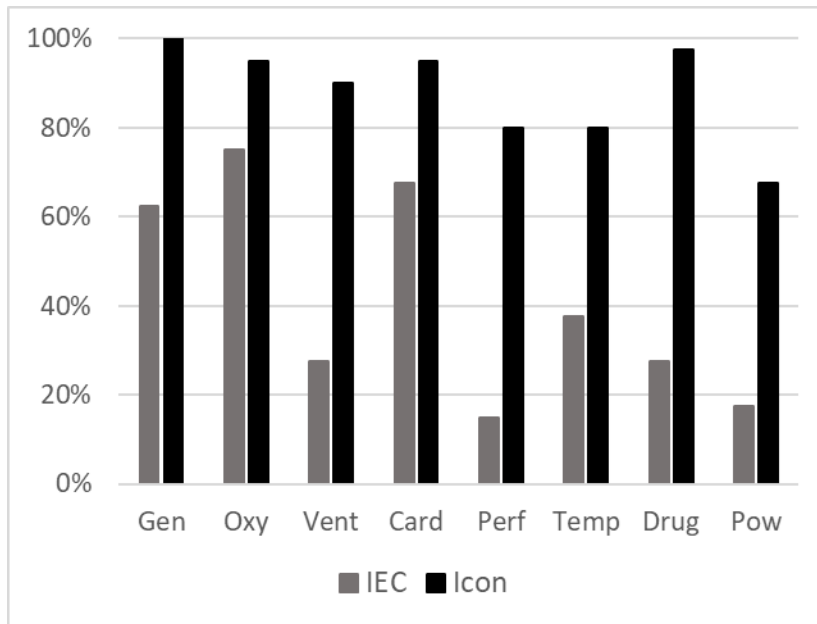


Figure 2. Comparison of individual alarm sounds in IEC and Icon sets. The overall accuracy rates shown do not account for the nested data structure (refer to Figure 1) and are, therefore, averaged across subjects and sessions.

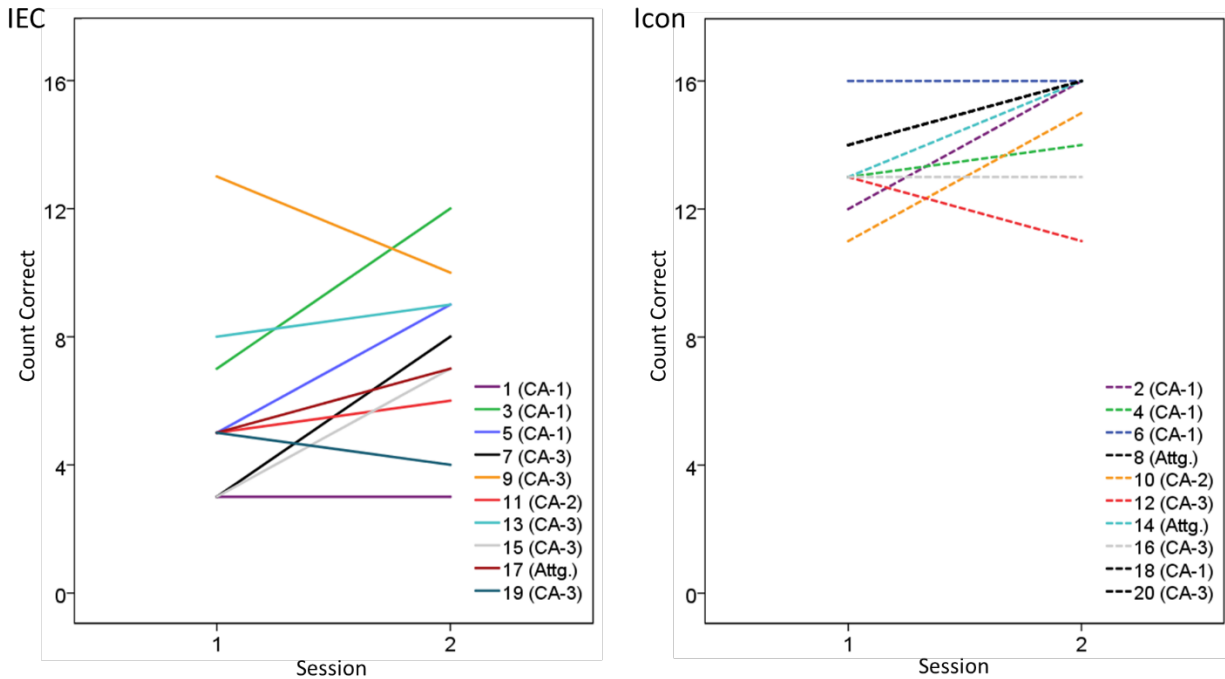


Figure 3. Individual subject performance during the course of experiments; IEC group and Icon group. Here accuracy is depicted in total number of counts (maximum of 16 counts per subject per session). For example, subject 6 in the Icon group identified every alarm annunciation correctly (32/32), while subject 1 from the IEC group was wrong 26/32 times. Subjects 8, 18, and 20 from the Icon group shared the same number of correct counts for sessions 1 and 2; data for these subjects are represented by the same dashed black line. 'CA-x' refers to clinical anesthesia year of residency. 'Attg.' refers to attending physician.

Table 1. List of novel auditory icons with description for the eight alarm categories.

Category	Auditory Icon Characteristics#
General	Doorbell chime version of fate motif in Beethoven's 5th symphony
Oxygen	Intermittent jet ventilation-two pulses
Ventilation	Heavy breathing for one respiratory cycle
Cardiovascular	Heart beating with no discernable frequency
Artificial Perfusion	Flowing liquid
Temperature	Tea kettle whistling
Drug Administration	Rattling 'pillbox'
Power Down	Attempted 'pull start' of a lawn mower

Table 2. Simulation Scripts

<u>Minute</u>	<u>Alarm Category</u>		<u>Alarm Etiology</u>
	Patient 1	Patient 2	
1-3	-	-	-
4	Cardiovascular	-	Heart rate change from 62 to 58
5	-	Power Failure	No identifiable etiology
6	Temperature	-	Temperature change from 37.4 to 37.6 °C
7	-	Ventilation	EtCO2 change from 30 to 24 mmHg
8	Drug Administration	-	Antibiotic Infusion pump inferred
9	-	Perfusion	ECMO device inferred
10	General Alarm	-	No identifiable etiology
11	Ventilation	-	EtCO2 change from 29 to 15 mm Hg
12	-	Cardiovascular	Blood pressure change from 91/50 to 88/48 mm Hg
13	-	Drug Administration	Epinephrine Infusion pump inferred
14	Perfusion	-	Dialysis machine inferred
15	-	Temperature	Temperature change from 35.6 to 35.4 °C
16	Oxygenation	-	Saturation change from 90 to 89%
17	-	Oxygenation	Saturation change from 91-89%
18	Power Failure	-	No identifiable etiology
19	-	General Alarm	No identifiable etiology
20	-	-	-

EtCO2 = end tidal carbon dioxide

ECMO = extracorporeal membrane oxygenation

Table 3. Identification Accuracy and Response Times.

	IEC (N = 10)	Icon (N = 10)
Session 1		
Correct Responses out of 160 Cases	57	133
Overall Percentage Correct	36%	83%
Mean Response Time (98.75% CI)	17 (14 to 19) <sup>#</sup>	13 (12 to 14)
Session 2		
Correct Responses out of 160 Cases	75	149
Overall Percentage Correct	47%	93%
Mean Response Time in seconds (98.75% CI)	14 (13 to 16)	11 (11 to 12)

<sup>#</sup>15 cases under the IEC condition were left blank during Session 1. These binary responses were counted as 'incorrect' when calculating percentage and were ignored when calculating response times.

Note: There were 10 subjects per group and each subject was exposed to 16 alarm annunciations per session with each of the 8 alarm categories annunciated twice. Therefore, unless otherwise noted, each data element corresponds to 160 cases.

Table 4. Mixed Model Effects of Group and Session on Alarm Detection Accuracy and Response Times

	<b>Log Odds of a Correct Response</b>	<b>Mean (98.75% CI) in Seconds</b>	<b>SE</b>	<b>Odds Ratio (98.75% CI)</b>	<b>P- Value</b>
<b>Mixed Model for Binary Response<sup>&amp;</sup></b>					
Group (Icon vs IEC)	3.26	-	0.45	26.1 (8.4 to 81.5)	<0.001
Session (2 <sup>nd</sup> vs 1 <sup>st</sup> )	0.78	-	0.21	2.2 (1.3 to 3.7)	<0.001
<b>Mixed Model for Response Time<sup>#</sup></b>					
Group (IEC minus Icon)	-	3 (1 to 5)	1	-	<0.001
Session (1 <sup>st</sup> minus 2 <sup>nd</sup> )	-	2 (1 to 3)	0	-	<0.001

<sup>&</sup>Additional mixed effects logistic results for binary response. Fixed effects: Intercept estimate (SE) = -0.170 (1.510), OR (98.75% CI) = 0.8 (0.0 to 34.5), P = 0.909. Random covariance: Intercept estimate = 2.119, Subject estimate (SE) = 0.491 (0.258), OR = 1.6 (0.1 to 1.8), P = 0.057.

<sup>#</sup>Additional mixed effects results for response time. Fixed effects: Intercept estimate (SE) = 16 (5), 98.75% CI = 4 to 29, P = 0.001. Random covariance: Intercept estimate = 24, Subject estimate (SE) = 2 (1), 98.75% CI = 1 to 6, P = 0.033.

IEC refers to group (N<sub>subjects</sub> = 10) exposed to International Electrotechnical Commission standard alarms.

Icon refers to group (N<sub>subjects</sub> = 10) exposed to example icon alarms.

SE = Standard Error, CI = Confidence Interval



Table 5. Assessments of Fatigue and Task Load and Results of Exit Survey

	Mean (95% CI)		
	IEC [N = 10]	Icons [N = 10]	P-Value
Swedish Occupation Fatigue Inventory <sup>&amp;</sup>			
Lack of Energy	3 (1 to 4)	1 (0 to 2)	<b>0.028</b>
Physical Exertion	0 (0 to 0)	0 (0 to 1)	0.34
Physical Discomfort	0 (0 to 1)	0 (0 to 1)	0.71
Lack of Motivation	1 (0 to 2)	1 (0 to 2)	0.77
Sleepiness	2 (0 to 3)	2 (1 to 3)	0.95
NASA Task Load Assessment <sup>@</sup>			
Metal Demand: How mentally demanding was the task?	13 (10 to 16)	9 (5 to 12)	0.08
Physical Demand: How physically demanding was the task?	5 (2 to 7)	3 (1 to 5)	0.24
Temporal Demand: How hurried or rushed was the pace of the task?	6 (3 to 10)	4 (1 to 7)	0.20
Performance: How successful were you in accomplishing the task?	12 (9 to 16)	6 (4 to 8)	<b>0.003</b>
Effort: How hard did you have to work to accomplish the task?	13 (10 to 16)	9 (6 to 12)	0.06
Frustration: How insecure, discouraged, irritated, stressed, and annoyed are you?	14 (10 to 19)	7 (2 to 12)	<b>0.025</b>
Exit Survey <sup>#</sup>			
1. How many audio alarms do you think you heard in total?	12 (9 to 14))	13 (11 to 16)	0.21
2. To what extent were you aware of the audio alarms?	7 (6 to 7))	7 (7 to 7)	0.12
3. How easy was it to work out what the alarm's meant?	2 (1 to 3)	5 (4 to 6)	<b>&lt; 0.001</b>
4. How easy was it to hear the alarms?	6 (5 to 7)	6 (5 to 7)	0.41
5. How helpful did you find the audio alarms?	4 (2 to 5)	5 (5 to 6)	<b>0.016</b>

<sup>&</sup>Responses to items are on a 7-point Likert scale and lower values are better.

<sup>@</sup>Responses are on a 20-point visual scale and lower values are better.

<sup>#</sup>Response to question 1 is open ended, responses to questions 2 thru 5 are on a 7-point Likert scale and larger values are better.

<sup>%</sup>Values less than 0.05 are in bold.

CI = Confidence Interval