

Auditory Icon Alarms Are More Accurately and Quickly Identified than Current Standard Melodic Alarms in a Simulated Clinical Setting

Richard R. McNeer, M.D., Ph.D., Danielle Bodzin Horn, M.D., Christopher L. Bennett, Ph.D., Judy Reed Edworthy, Ph.D., Roman Dudaryk, M.D.

ABSTRACT

Background: 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 s, difference 3 s [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 current standard alarms than icon alarms. (**ANESTHESIOLOGY 2018; 129:58-66**)

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 (Geneva, Switzerland) published a standard in 2003 (most recently revised in 2012) known as International Electrotechnical Commission 60601, which specifies basic safety of medical electrical equipment and governs almost all medical equipment across the globe.¹ Parts 1 to 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 current standard alarms). However, the current standard alarms have been shown to function poorly by researchers in the fields of human factors and psychology.²⁻⁵ 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 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.¹

What We Already Know about This Topic

- Current standard audible medical alarms are difficult to learn and distinguish from each other
- Auditory icons are a new type of alarm that mimic the underlying meanings they are meant to represent

What This Article Tells Us That Is New

- In a simulated intensive care unit using primarily anesthesiology residents as test subjects, the ability to learn and identify standard and icon alarms was tested
- In this setting, icon alarms were easier to learn and identify than standard alarms, while standard alarms were more likely to be perceived as having higher fatigue and task load

Several studies have demonstrated that the current standard alarms are therefore difficult to learn (especially in the musically uninitiated), and alarms within the set are easily confused with one another²⁻⁵—factors potentially contributing to alarm fatigue, and certainly the cause of unnecessary confusion.⁶ Device manufacturers are not required to adopt the International Electrotechnical Commission standard and can implement proprietary alarm sounds that at least

This article is featured in “This Month in Anesthesiology,” page 1A. Supplemental Digital Content is available for this article. Direct URL citations appear in the printed text and are available in both the HTML and PDF versions of this article. Links to the digital files are provided in the HTML text of this article on the Journal’s Web site (www.anesthesiology.org). This article has a video abstract.

Submitted for publication May 19, 2017. Accepted for publication March 23, 2018. From the Department of Anesthesiology (R.R.M., R.D.), and the Department of Anesthesiology, Jackson Memorial Hospital (D.B.H.), Miller School of Medicine, and Music Engineering Technology Program, Frost School of Music (C.L.B.), University of Miami, Miami, Florida; and the Cognition Institute, Plymouth University, United Kingdom (J.R.E.).

Copyright © 2018, the American Society of Anesthesiologists, Inc. Wolters Kluwer Health, Inc. All Rights Reserved. Anesthesiology 2018; 129:58-66

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 Fifth Symphony
Oxygenation	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”
Equipment failure	Attempted “pull start” of a lawn mower

demonstrate equivalence.¹ However, no clear precedent has been established on how to best test the effectiveness of novel alarm sounds.

Development of an updated version of International Electrotechnical Commission 60601-1-8 is currently underway, and “auditory icons”⁶ (referred to here as icon alarms) are considered for replacement of current standard 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 current standard alarms, icon alarms were found to be easier to learn and discriminate when studied in nonclinical, computer-based settings using lay, nonclinical participants.^{6,7} Additionally, icon alarms were easier to localize in an experimental setting.⁶ On the basis of these results, the International Electrotechnical Commission 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 (written personal communication from Dave Osborn, B.S.E.E., M.E.E., chair of International Electrotechnical Commission Alarms Joint Working Group, Philips, Salem, Massachusetts, April 2016).

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 current standard alarms in a simulated intensive care unit using anesthesia providers as subjects.

Materials and Methods

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

Study Design Overview and Outcome Measures

To evaluate the relative effectiveness of current standard alarms and icon alarms, we used a simulated two-bed intensive care unit. The study had a between-subjects factor represented by “group” (current standard or icon alarm set exposure) and a within-subjects factor represented by replicated measure—there were two sessions (fig. 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 a methodology described previously.⁸ Experiments were conducted from October 13, 2016, to December 16, 2016, in the early afternoon period.

Icon Alarm Set Design

Icon 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 the current standard alarms. With medical alarms, the conveyed “concepts” derive from the category of alarm, and the International Electrotechnical Commission standard specifies eight such categories: cardiovascular, ventilation, artificial 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, <http://links.lww.com/ALN/B708>, and 2, <http://links.lww.com/ALN/B709>, with embedded audio of icon and current standard alarms used in this study, respectively), 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 serves to draw the attention of the user to the presentation of the alarm.⁹

Intensive Care Unit Simulator Setup

The simulated intensive care unit consisted of two beds, each with a simulated patient (mannequin). A custom multimedia graphical user interface described in detail elsewhere^{8,10} was associated with each patient and placed adjacent to the

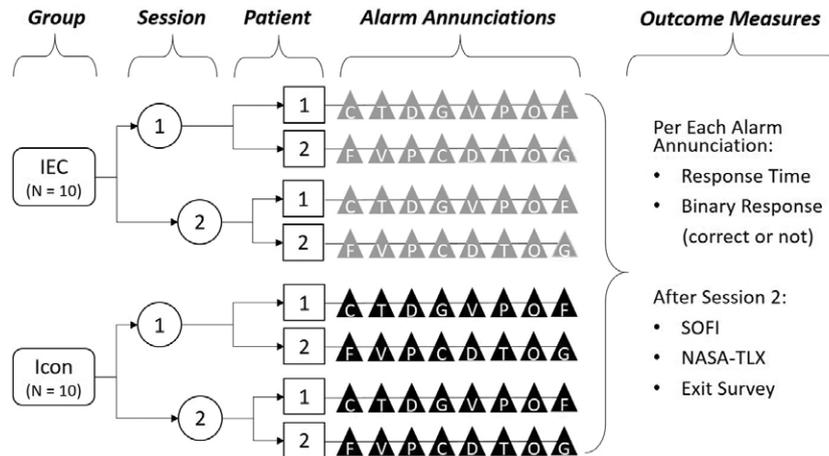


Fig. 1. Schematic showing experimental design and nested data structure. During experiments, each subject was exposed to a total of 32 alarm annunciations, and a maximum of 640 cases (20 subjects \times 32 alarms) was obtained. *Gray and black triangles* represent International Electrotechnical Commission (IEC) and icon alarm annunciations, respectively. C = cardiovascular; D = drug administration; F = equipment failure; G = general alarm; NASA-TLX = National Aeronautics and Space Administration Task Load Assessment Questionnaire; O = oxygenation; P = artificial perfusion; SOFI = Swedish Occupation Fatigue Inventory; T = temperature; V = ventilation. Refer to table 2 for exact times of alarm annunciation.

left of the bedside from a patient perspective. For this study, the graphical user interface was modified to add touchscreen functionality and was installed on two tablet computers (Microsoft Surface 3, Microsoft, USA), which mimicked touch functionality found on most modern patient monitor displays. The graphical user interface 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 graphical user interface 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 intensive care units including infusion pumps, support poles, associated tubing, stretchers, a crash cart, and so forth. Devices not available to us such as the dialysis and extracorporeal membrane oxygenation 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 intensive care unit 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.^{8,11} 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 graphical user interface.

Experimental Procedure

Subjects consisted of clinical anesthesia residents and anesthesia attending physicians who were recruited the day of

scheduled experiments and randomized to current standard or icon groups based on order of arrival to the simulation laboratory (odd, current standard; 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 (Supplemental Digital Content 1, <http://links.lww.com/ALN/B708>, and 2, <http://links.lww.com/ALN/B709>, with embedded audio of icon and current standard alarms used in this study, respectively). The presentations were subject-paced and took 5 to 10 min to complete. Then subjects were escorted to the simulated intensive care unit 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 (one 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 (fig. 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 re-presented, 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

Table 2. Simulation Scripts

Minute	Alarm Category		Alarm Etiology
	Patient 1	Patient 2	
1 to 3	—	—	—
4	Cardiovascular	—	Heart rate change from 62 to 58
5	—	Equipment failure	No identifiable etiology
6	Temperature	—	Temperature change from 37.4 to 37.6°C
7	—	Ventilation	ETCO ₂ change from 30 to 24 mmHg
8	Drug administration	—	Antibiotic infusion pump inferred
9	—	Artificial perfusion	ECMO device inferred
10	General alarm	—	No identifiable etiology
11	Ventilation	—	ETCO ₂ change from 29 to 15 mmHg
12	—	Cardiovascular	Blood pressure change from 91/50 to 88/48 mmHg
13	—	Drug administration	Epinephrine infusion pump inferred
14	Artificial 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 to 89%
18	Equipment failure	—	No identifiable etiology
19	—	General alarm	No identifiable etiology
20	—	—	—

ECMO = extracorporeal membrane oxygenation; ETCO₂ = end-tidal carbon dioxide.

sound (either current standard or icon) was annunciated a total of four times per subject over the course of two sessions. Each subject was, therefore, exposed to a total of 32 alarm annunciations during the experiment (fig. 1).

Subjective Instruments

At the end of session 2, subjects completed two validated psychometric instruments: the Swedish Occupation Fatigue Inventory^{8,12} and the National Aeronautics and Space Administration Task Load Assessment Questionnaire.^{8,13} Subjects also completed an exit survey consisting of six questions that 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”; G Power, University of Dusseldorf, Germany). Previously, results of a repeated-measures, laboratory-based study comparing identifiability of five alarm sets (including International Electrotechnical Commission and icon alarm sets) reported effect sizes in terms of partial eta squared (η_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 ($\eta_p^2 = 0.193$).⁶ 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 (current standard or icon). We also used this expected effect size in consideration of the effect of group (current standard *vs.* icon) on alarm reaction time averaged for each alarm set.

Using the Bonferroni approach, the 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 SPSS Statistics software (version 24; IBM, USA). To analyze primary outcome results, a generalized linear mixed model was selected for the following reasons.¹⁴ This approach (1) is able to account for nested and hierarchical data (fig. 1); (2) can consider dependent variables that are parametric (*e.g.*, response time) and binary; (3) can account for both fixed and random effects; and (4) when 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 generalized linear mixed model 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 follows published suggested guidelines.¹⁴ To reduce the risk of type I error, significance was adjusted as in the power analysis to alpha = 0.0125. Additionally, a generalized linear

mixed model can reduce type I error by its accounting of random effects.¹⁴ An important disadvantage of the generalized linear mixed model is that common measures of effect size (e.g., Cohen's d and η_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% CI 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% CI.

Secondary outcome measures were analyzed using descriptive statistics. Individual items from the Swedish Occupation Fatigue Inventory, National Aeronautics and Space Administration Task Load Assessment Questionnaire, and exit survey are reported as mean values and 95% CI. P values are also reported for pairwise comparisons. Cronbach's alpha was calculated to assess internal consistency for the Swedish Occupation Fatigue Inventory and National Aeronautics and Space Administration Task Load Assessment Questionnaire (i.e., that all items in each instrument measured the same construct).

Results

Twenty subjects consisting of 17 clinical anesthesia residents (7 year 1 residents, 2 year 2 residents, and 8 year 3 residents) and 3 attending physicians participated in the study. Over the course of the entire experiment, 640 alarms (cases) were announced—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 current standard 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 generalized linear mixed model 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 current standard alarms, “general alarm” (61%), “oxygenation” (77%), and “cardiovascular” (70%) were associated with the highest accuracy rates, while “artificial perfusion” (17%) and “equipment failure” (19%) were associated with the lowest. Six of the eight icon alarms were identified correctly 80% of the time or more, while “equipment failure” was associated with the lowest accuracy rate (69%) of the group (fig. 2). Overall, subjects identified icon alarms more accurately and quickly than the current standard alarms (table 3), and an effect of training level on subject performance was not observed (fig. 3). In particular, subjects in the icon group were 26.1 (98.75% CI, 8.5 to 81.5) times more likely to respond correctly to alarm annunciations ($P < 0.001$) and responded sooner by 3 (98.75% CI, 1 to 5) s than subjects in the current standard group ($P < 0.001$; table 4). Most subjects (7 of 10 for each group) performed better in session 2 than in session 1 irrespective of alarm grouping (fig. 3). Overall, subjects were 2.2 (98.75%

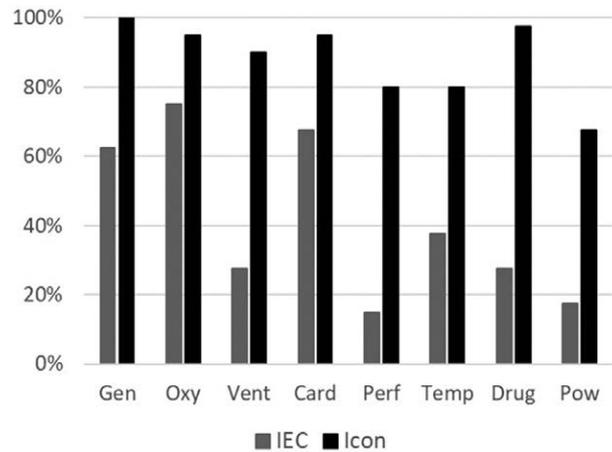


Fig. 2. Comparison of individual alarm sounds in International Electrotechnical Commission (IEC) standard and icon sets. The overall accuracy rates shown do not account for the nested data structure (fig. 1) and are, therefore, averaged across subjects and sessions. Card = cardiovascular; Drug = drug administration; Gen = general; Oxy = oxygenation; Perf = artificial perfusion; Pow = equipment failure; Temp = temperature; Vent = ventilation.

CI, 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 (98.75% CI, 1 to 3) s quicker ($P < 0.001$; table 4).

Secondary Outcomes

Reliability of test results as measured by Cronbach's alpha suggests that the Swedish Occupation Fatigue Inventory and National Aeronautics and Space Administration Task Load Assessment Questionnaire 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 current standard group reported a

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 in seconds (98.75% CI)	17 (14–19)*	13 (12–14)
Session 2		
Correct responses out of 160 cases	75	149
Overall percentage correct	47	93
Mean response time in seconds (98.75% CI)	14 (13–16)	11 (11–12)

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.

*Fifteen cases under the International Electrotechnical Commission (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.

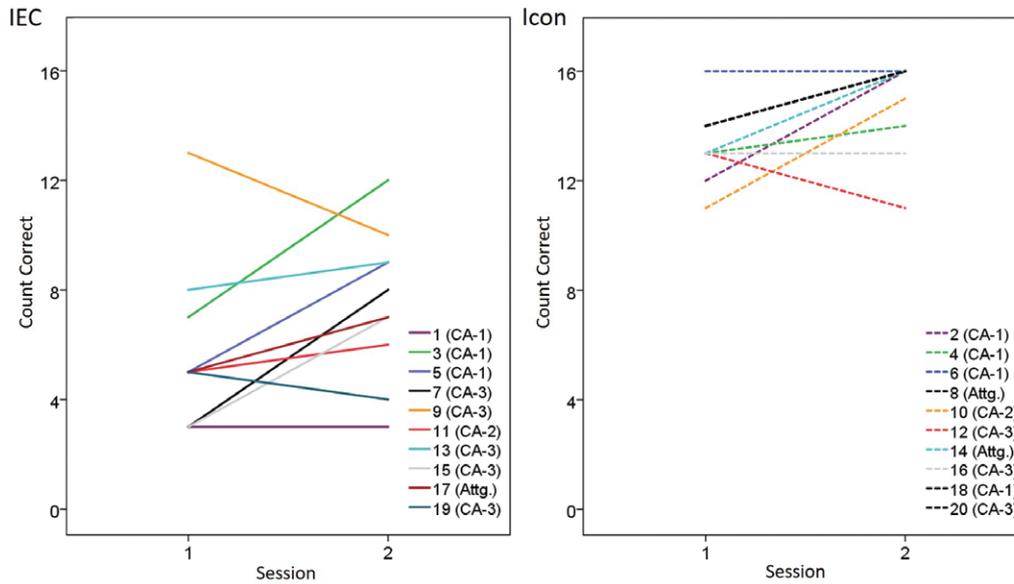


Fig. 3. Individual subject performance during the course of experiments, International Electrotechnical Commission (IEC) standard group and icon group. Here accuracy is depicted in total number of counts (maximum of 16 counts per subject per session). For example, subject 6 in the icon group identified every alarm annunciation correctly (32/32), while subject 1 from the IEC standard 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*. Attg. = attending physician; CA-x = clinical anesthesia year of residency.

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

	Log Odds of a Correct Response	Mean (98.75% CI) in Seconds	SE	OR (98.75% CI)	P Value
Mixed model for binary response*					
Group (icon vs. IEC)	3.26	—	0.45	26.1 (8.4–81.5)	< 0.001
Session (2 vs. 1)	0.78	—	0.21	2.2 (1.3–3.7)	< 0.001
Mixed model for response time†					
Group (IEC minus icon)	—	3 (1–5)	1	—	< 0.001
Session (1 minus 2)	—	2 (1–3)	0	—	< 0.001

*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 (98.75% CI) = 1.6 (0.1 to 1.8), $P = 0.057$. †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, $P = 0.033$.

Icon = group ($N_{\text{subjects}} = 10$) exposed to example icon alarms; IEC = ($N_{\text{subjects}} = 10$) exposed to International Electrotechnical Commission standard alarms; OR = odds ratio; SE = standard error.

higher score in the Swedish Occupation Fatigue Inventory questionnaire for “lack of energy” (3 [95% CI, 1 to 4] vs. 1 [95% CI, 0 to 2]; $P = 0.028$). Subjects in the current standard group reported experiencing higher levels of task load on the National Aeronautics and Space Administration Task Load Assessment Questionnaire questionnaire along all items, especially for “performance” (12 [95% CI, 9 to 16] vs. 6 [95% CI, 4 to 8]; lower is better; $P = 0.003$) and “frustration” (14 [95% CI, 10 to 19] vs. 7 [95% CI, 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 current standard group (5 [95% CI, 4 to 6] vs. 2 [95% CI, 1 to 3]; $P < 0.001$), and the same group found the alarm sounds more helpful (5 [95% CI, 5 to 6] vs. 4 [95% CI, 2 to 5]; $P = 0.016$; table 5).

Discussion

After a brief exposure to alarm sounds, anesthesia providers identified icon alarms more accurately than the International Electrotechnical Commission standard alarms during clinical simulation. This indicates that icon alarms were easier to learn, which corroborates results obtained previously from laboratory-based experiments that used non-clinical subjects.^{6,7} Our subjects also identified icon alarms more quickly than current standard alarms. Although it is unclear if the effect observed here is clinically relevant, we believe, 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 current standard alarms.

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

	Mean (95% CI)		P Value§
	IEC (N = 10)	Icons (N = 10)	
Swedish Occupation Fatigue Inventory*			
Lack of energy	3 (1–4)	1 (0–2)	0.028
Physical exertion	0 (0–0)	0 (0–1)	0.34
Physical discomfort	0 (0–1)	0 (0–1)	0.71
Lack of motivation	1 (0–2)	1 (0–2)	0.77
Sleepiness	2 (0–3)	2 (1–3)	0.95
National Aeronautics and Space Administration Task Load Assessment†			
Mental demand: How mentally demanding was the task?	13 (10–16)	9 (5–12)	0.08
Physical demand: How physically demanding was the task?	5 (2–7)	3 (1–5)	0.24
Temporal demand: How hurried or rushed was the pace of the task?	6 (3–10)	4 (1–7)	0.20
Performance: How successful were you in accomplishing the task?	12 (9–16)	6 (4–8)	0.003
Effort: How hard did you have to work to accomplish the task?	13 (10–16)	9 (6–12)	0.06
Frustration: How insecure, discouraged, irritated, stressed, and annoyed are you?	14 (10–19)	7 (2–12)	0.025
Exit survey‡			
1. How many audio alarms do you think you heard in total?	12 (9–14)	13 (11–16)	0.21
2. To what extent were you aware of the audio alarms?	7 (6–7)	7 (7–7)	0.12
3. How easy was it to work out what the alarms meant?	2 (1–3)	5 (4–6)	< 0.001
4. How easy was it to hear the alarms?	6 (5–7)	6 (5–7)	0.41
5. How helpful did you find the audio alarms?	4 (2–5)	5 (5–6)	0.016

*Responses to items are on a 7-point Likert scale and lower values are better. †Responses are on a 20-point visual scale and lower values are better. ‡Response to question 1 is open-ended, responses to questions 2 through 5 are on a 7-point Likert scale, and larger values are better. §Values less than 0.05 are in bold. IEC = International Electrotechnical Commission.

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 standard 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 the current standard alarms resulted in accuracy rates between 10 and 61%.¹⁵ 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- to 10-min orientation period, we observed overall accuracy rates for the icon alarms of between 68% (equipment failure) and 100% (general alarm). In comparison, our overall accuracy for the current standard alarms ranged from 15% (artificial 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 current standard and icon alarms, it seems that additional and more regimented training sessions would be required for the current standard set if accuracy rates are to approach those of the icon set.

Our findings are based on comparison between alarm sets (*i.e.*, current standard *vs.* icon). However, some icon alarms tested here were easier to identify than others (fig. 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.¹⁶ 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 intraset interactions.

Manufacturers are able to use proprietary alarms as long as they conform to the normative portions of the standard that specify sequences of tones and demonstrate that the alarms are as effective as the current International Electrotechnical Commission alarms.¹ 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 standard alarms. As a result, the effectiveness of proprietary alarms is likely to be closer to that of the current standard alarms than to the icon alarms. Hence, we chose the International Electrotechnical Commission alarm set as a control for this study, although it is possible that some proprietary alarms are more effective than the current standard alarms in clinical practice. At our institution, few devices use the current standard alarms, and this also informed our selection for control because subjects could be considered to be relatively naive, thus putting current standard and icon sets on more equal footing with regard to previous alarm exposure history.

Additionally, the design problems surrounding the current standard alarms are well described and have helped inform the rationale and design of icon alarms. An expectation is that if the next set of standard alarms are demonstrably more effective than the current ones, 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.^{8,17-19} 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, although 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,^{6,7,16} this study nonetheless only approximated an intensive care unit 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, although it is generally recognized that nurses endure the most exposure to alarm sounds and have the highest risk of alarm fatigue.^{20,21} 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 (fig. 3). Additionally, to date, lay subjects and anesthesia providers appear to perform similarly when comparing current standard 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 International Electrotechnical Commission 60601-1-8, which were based on work by Kerr.²² There is increasing discussion of a need to modify the alarm categories, and audible alarm function may depend partly on the classification system.²³ In the current study, we used the standard categories out of necessity as the focus was to compare icon alarms to the current standard 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 current standard alarms in a simulated intensive care unit. We propose that future investigations of icon alarms do not need to include a current standard 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 a 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, Washington, D.C.), industry groups (*e.g.*, Association for the Advancement of Medical Instrumentation, Arlington, Virginia), and standards organizations (*e.g.*, International Electrotechnical Commission [Geneva, Switzerland], International Organization for Standardization [Geneva, Switzerland], and, American National Standards Institute [Washington, D.C.]). Last, 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 (Geneva, Switzerland)²⁴ and the U.S. Environmental Protection Agency (Washington, D.C.).

Conclusions

Relative to the International Electrotechnical Commission 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 International Electrotechnical Commission alarms than icon alarms.

Research Support

Supported by the Association for the Advancement of Medical Instrumentation, Arlington, Virginia.

Competing Interests

The authors declare no competing interests.

Correspondence

Address correspondence to Dr. McNeer: Division of Trauma and Critical Care, Department of Anesthesiology,

University of Miami Miller School of Medicine, P.O. Box 016370 (M820), Miami, Florida 33101. mcneer@miami.edu. This article may be accessed for personal use at no charge through the Journal Web site, www.anesthesiology.org.

References

1. International Organization for Standardization: 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. #iso:std:iec:60601:-1-8:ed-2:v1:en. Available at: <https://www.iso.org/obp/ui/>. Accessed April 19, 2018
2. Wee AN, Sanderson PM: Are melodic medical equipment alarms easily learned? *Anesth Analg* 2008; 106:501-8
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., Dresden, Germany, 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-Farrés 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. *Int J Ind 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(5 pt 1):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(suppl 2):50-7
24. Falk SA, Woods NF: Hospital noise—Levels and potential health hazards. *N Engl J Med* 1973; 289:774-81