Getting alarm sounds into a global standard: a case study with reflections

Judy Edworthy, Richard McNeer, Christoper Bennett, Roman Dudaryk, Siné McDougall, Joseph Schlesinger, Matthew Bolton, Jonathan Edworthy, Roman Dudaryk, Elif Ožcan Vieira, Andrew Boyd, Scott Reid, Michael Rayo, Melanie Wright, David Osborn

1. Introduction

Audible alarms are very important across high-workload industries and their use in those environments is not always driven by settled science, but by other factors such as customer reaction, budget, lack of expertise in design and application of knowledge, inflexible approaches to known problems, and so on. Therefore there are many examples of high-workload, safety-critical environments where the audible alarms leave much to be desired in terms of both their implementation and their design, though increasingly there are also many examples of thoughtful and designed implementations. In clinical environments the problem of bad alarm implementation has reached colossal proportions, where patient deaths have been attributed to 'alarm fatigue' (Drew et al, 2014; Sendelbach & Funk, 2013). Until a national summit in the US in 2011 (http://www.aami.org/events/eventdetail.aspx?ItemNumber=1153&loggedOut=True), little was being done about the general problem of alarm over-use but now there are well-documented and successful attempts to reduce the problem of over-alarming in general (Cvach, 2012; Welch et al, 2011; Whalen et al, 2014). The audible alarms themselves have traditionally left a lot to be desired from the point of view of design, but now that those bigger problems are slowly being resolved, the time is right to improve upon the audible alarms as well. In this paper we describe a project intended to upgrade and update the audible alarms in a global medical device standard.

The challenge of carrying out what is essentially an applied, customer-based problem while maintaining the best scientific approach one can muster is a challenge. This issue is

highlighted by Morrow and Durso (2011) in their editorial on a special issue of JEP:Applied on Cognitive Issues in Healthcare. They introduce their papers thus:

'…we focus on the need for research that is sufficiently comprehensive to identify threats to patient safety, yet specific enough to explain how provider and patient factors interact with task and health context to engender these threats. Such research should be theory-based, yet also problem-driven; exert experimental control over theoretically relevant variables, yet also involve participants, tasks, and contexts that represent the problems of interest. A tension exists between theory-based, experimentally controlled research on the one hand, and problem-driven research with representative situations on the other' (p.191)

The challenge in terms of audible alarm design is to bring the scientific evidence to bear on the problem, but also to commit at some point during the process to a specific set or set of sounds so that a research database can be built around them.

The evidence base for auditory alarm design is considerably more advanced than the typical sorts of alarms which are used in practice might suggest. Bridging the 'valley of death' between theory and application is always a problem, made more acute in auditory work given the difficulty of talking to non-experts (often the client) about sound in any abstract way, and given the predisposition that clients have to like or dislike a sound designed for a specific application – which can sometimes have nothing to do with the sound itself, but something to do with the implementation. For example, the typical response to any new clinical alarm, particularly from the nursing community, is 'We already have more than enough, we don't want any more!' Understandably, the fact that in time new and better alarms will replace those currently in use is not of great relevance to those who might have to work through the transition.

2. The standard: IEC 60601-1-8

IEC 60601 is a set of standards concerned with the safety of medical electrical equipment (so covers almost all medical equipment). Parts 1 to 8 specify the basic safety and essential performance requirements and tests for the alarm systems contained within that equipment. Thus this standard governs almost all medical equipment across the globe. It was published first in 2006, then updated in 2012, had something of an update in 2015 and is due for another, major update by the end of 2019. The key feature of the standard in terms of audible alarms is that it specifies the acoustic and structural elements of the audible alarms that should accompany specific medical hazards or categories (IEC, 2012).

The reserved set of alarms was designed with the best of intentions (Block et al, 2000), based on some aspects of what was known about alarm design at the time (but not all). The sounds embodied important acoustic features that would increase their resistance to masking (compared at least with single harmonics), and improve their general acceptability over the earlier beeps, buzzers and bells. The structure of the alarms and their categories is shown in Table 1. There are eight categories of risk specified. Each has a high- and a medium-priority form. In our studies only the high-priority version was tested, though generic medium- and low-priority alarms were also tested for this update.

Table 1: IEC 60601-1-8 High Priority Alarm Characteristics (from Edworthy et al, 2017a)

Function of Alarm	Alarm Characteristics
General	A burst of three regularly spaced pulses (each pulse ranging between 100ms – 300ms), followed by a burst of two regularly spaced pulses in the following pattern: c c c – c c
Power down	A burst of three regularly spaced pulses (each pulse ranging between 100ms – 300ms), followed by a burst of two regularly spaced pulses in the following pattern: C c c – C c
Cardiovascular	A burst of three regularly spaced pulses (each pulse ranging between 100ms – 300ms, followed by a burst of two regularly spaced pulses in the following pattern: c e g – g C
Perfusion	A burst of three regularly spaced pulses (each pulse ranging between 100ms – 300ms), followed by a burst of two regularly spaced pulses in the following pattern: c f# c – c f#
Drug Administration	A burst of three regularly spaced pulses (each pule rangingbetween 100ms – 300ms), followed by a burst of two regularly spaced pulses in the following pattern: C d g – C d
Oxygen	A burst of three regularly spaced pulses (each pulse ranging between 100ms – 300ms), followed by a burst of two regularly spaced pulses in the following pattern: C b a – g f
Ventilation	A burst of three regularly spaced pulses (each pulse ranging between 100ms – 300ms), followed by a burst of two regularly spaced pulses in the following pattern: c a f – a f
Temperature	A burst of three regularly spaced pulses (each pulse ranging between 100ms – 300ms), followed by a burst of two regularly spaced pulses in the following pattern: C d e – f g

A key problem with the design was that the alarms, which sound like short, tonal melodies, all possess the same number of pulses and the same rhythm, making them very hard to distinguish between (Lacherez et al, 2007; Sanderson et al, 2006; Wee & Sanderson, 2008). The lack of diversity between the sounds is a major contributor to the known problems with learning and recognizing these alarms, and the finding is no surprise given that our ability to distinguish between stimuli depends on the number of dimensions along which they vary (Miller, 1956). A shared rhythm is also a key component of a listener's confusion between sounds (Patterson 1982). Calls to update and improve the sounds have been numerous, with the designer of the sounds himself issuing an apologia for the current sounds (Block, 2008).

It has become clear that almost anything would be better than the current alarms, which presents its own problem. Atyeo and Sanderson (2015) demonstrated that a similar set of alarms designed prior to the 2006 version of IEC 60601-1-8 (designed for an earlier version of the standard, Patterson et al 1986) outperform the current alarms, and other evidence shows that a random set of alarms, with no association to the meanings or functions of the alarms, were easier to learn than the current alarms (Edworthy et al, 2014). This earlier set of sounds was rejected on a non-empirical basis which allowed interested parties to call into a telephone line and listen to the sounds, and then to voice an opinion. However, that was the 1980s and patently, replacing the current alarms with alarms that simply perform better than the current alarms – even those designed in the 80s which turned out to be better than the alarms in the standard - is not enough.

Commentary 1

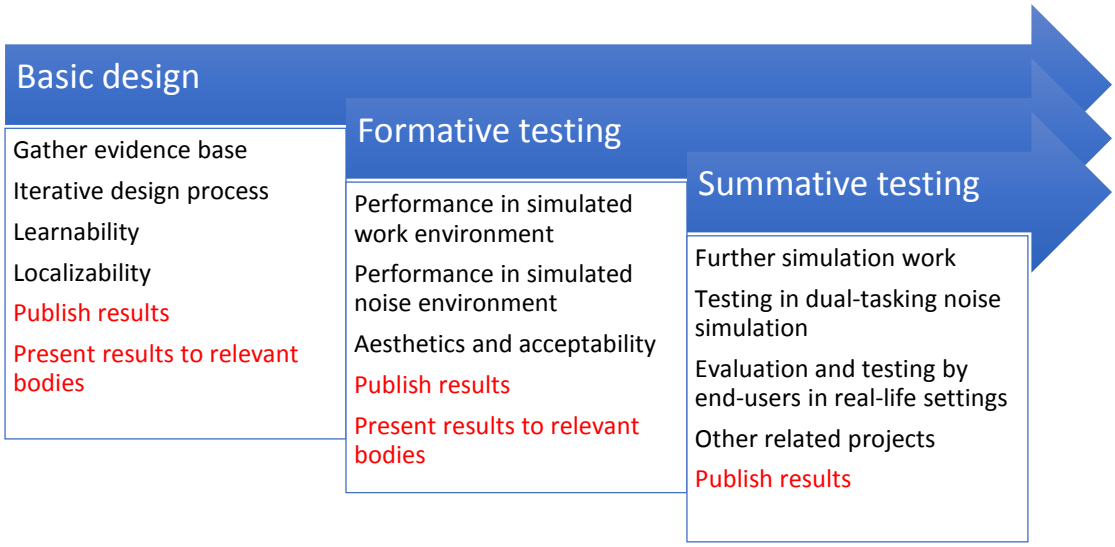
Despite knowing of the existence of the 'IEC 60601-1-8 alarm problem' for years prior to the start of the project, it was important to conduct the project with the endorsement of the body charged with updating the standard rather than conducting the work in isolation and then presenting it to that body, and to wait for a head of steam to build up over any potential replacement. The body in this case is the IEC 60601-1-8 and AAMI 60601-1-8 standards committees, through an IEC alarms joint working group. Access to this group was made possible by the Association for the Advancement of Medical Instrumentation (AAMI) having an open policy on membership of its own parallel IEC 60601-1-8 committee, AAMI 60601-1-8. The first author joined and began attending meetings. AAMI sometime later made a grant to the first author to carry out the initial development work.

Changing and updating standards is akin to the proverbial changing of the course of a ship using a teaspoon. The process of bringing about change in standards is very slow, requires a lot of attention, and any changes made now will be in place for many years in the future so there is considerable pressure to do a good job.

Also, the fate of earlier work heightens our awareness of non-empircally-based criticisms and potential scuppering, which is best met with empirically-based answers. Thus a key element of our strategy is to create a published and accessible database at every point in the process.

3. The process

Figure 1 shows the process we have adopted in developing the alarm sounds



Whereas medical equipment alarms have traditionally been produced by cheap sounding devices, many medical devices are now equipped with good quality speakers. Sound storage is also much cheaper, all of which means that, potentially, any sound can be used as an alarm, and the sound reproduction can be of high quality. This doesn't make the work of the designer any easier, indeed it focuses the effort required to demonstrate that any new alarms are not only 'better', but 'the best', or among the best, possible. A key question is what constitutes 'best'. Here, we have to start with learnability (whether or not it is important, though it probably is) as learnability is the only data we have on the current alarm sounds and comparisons are a good starting point, indeed essential in making the preliminary arguments for adoption of any new sounds.

3.1. Basic design

Developing memorable alarm sounds is fairly straightforward. There is ample evidence to show that the concrete-abstract continuum plays a big part in the learnabiity of

sounds, and there are many examples of ‘auditory icon’ alarm design which outperform abstract alarms (Belz et al, 1999; Edworthy et al, 2014; Graham, 1999; Keller & Stevens, 2004; Leung et al, 1997; Perry et al, 2007; Petocz et al, 2008; Stephan et al, 2006; Ulfvengren, 2003). Our starting point was therefore to develop several sets of sounds which used different types of metaphors for the alarm functions, and to compare them to the current alarms, which have no, or very minimal, metaphors. It is clear also that the acoustic variability of sound sets also affects people’s ability to learn them, so care was taken to have good variation within each set – in the case of the auditory icons, for example, it is quite tempting to end up with a set of ‘slushy’ sounds which might run the risk of confusion with one another, so this was consciously avoided by having some percussive sounds.

We tested four sets of alarms on both learnabiity and localizability.

We developed four sets of candidate sounds, one based on the word rhythms of the functions (an idea embodied but not well exploited in the current sounds), one based on simple metaphors (and low reproduction quality), and two based on auditory icons such as a rattling pill bottle for ‘drug administration’. One of the sets of auditory icons was simply the icons themselves, and the other was the icons plus a ‘pointer’, an abstract sound indicating the urgency, whereas the icon identifies the nature of the problem (cardiac, ventilation etc). The learnability data for the sound sets can be seen in Figure 2. All of our designs were more memorable than the existing set (all lines on the graph were significantly different from one another except the two at the top), but there was variation across our experimental sets also, with the auditory icon sets being the most memorable. The performance data suggests that we have covered the range of responses here, in that the performance for the auditory icons was almost at ceiling level from the start, and the current IEC alarms were very difficult to learn.

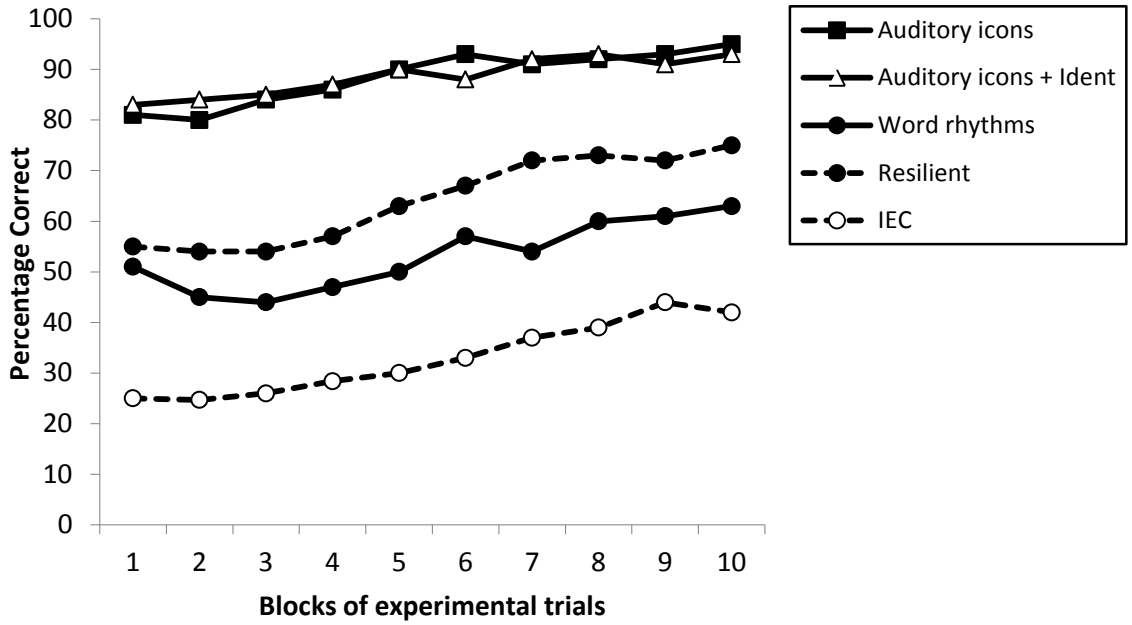


Figure 2: Percentage correct responses for each set of alarms, across ten trials (from Edworthy et al 2017a)

The candidate alarms were also varied in their harmonic complexity and denseness, as by and large more harmonically dense sounds are easier to localize. Very few tests of alarm localizability have been conducted (Alali, 2011; Catchpole et al, 2004; Vaillancourt et al, 2014), though localizability is often a pertinent issue in medical care (for example in a multibed ICU). Our results confirmed that the more harmonically dense alarms were easier to localize, and that the least complex, the current alarms, were poorest at localizability (Edworthy et al, 2017a).

Reflection 2

The findings from the basic study (Edworthy et al 2017a) were presented to the standards alarms joint working group in April 2016. They were also presented to the AAMI 60601-1-8 committee in June 2016, and to a meeting of the AAMI alarms coalition in July

2016. The empirical evidence was presented along with the sounds. As a consequence of this, the alarms joint working group decided they wishes to go ahead with the auditory icons plus pointer design, and supplied a list of activities, some formative and some summative, they would like to see undertaken prior to the committee recommending the adoption of the alarms in the standard. A further grant from AAMI to the first author was negotiated on this basis.

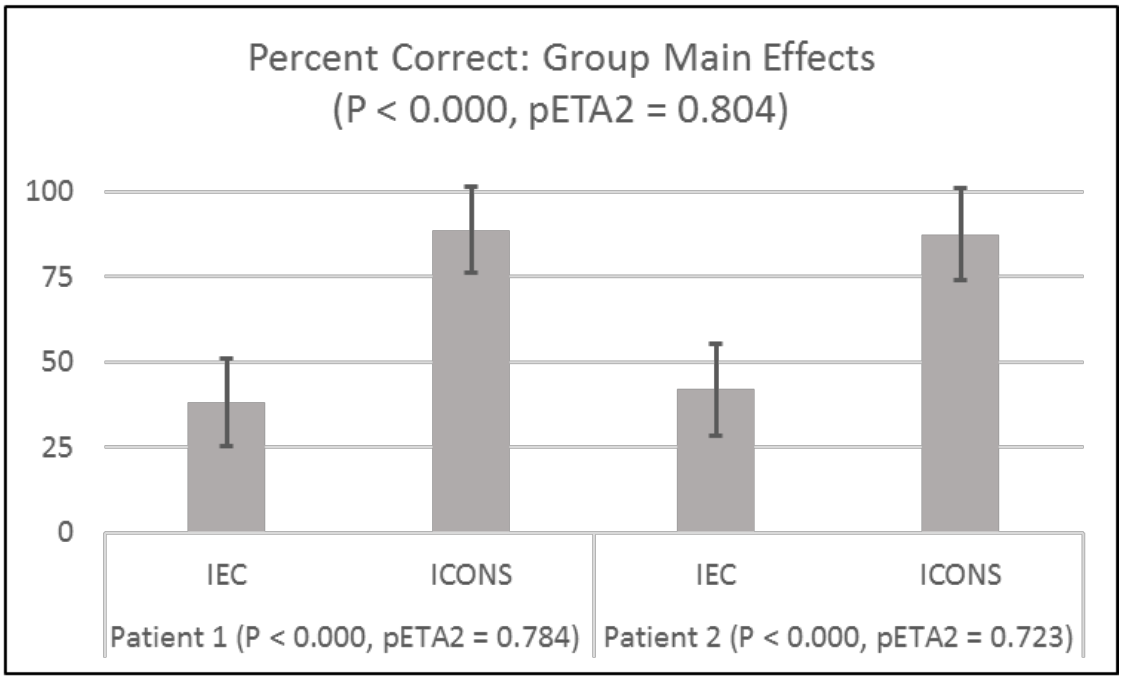
Another unexpected consequence is that there appears to be a substantial amount of dissent over the categories of risk themselves. We have approached this by writing a paper to open out discussion of the categories themselves (Edworthy et al, 2017b), and again AAMI has made a grant available, this time to Dr Wright, to address this issue.

3. 2 Formative testing

Mindful of Morrow & Durso's call for the use of contexts, tasks and participants of relevance (2011), the formative testing involves more realistic tasks, using clinically-trained participants. Using a range of already-developed and published techniques (Bennett & McNeer, 2012; Bennett et al 2015; McNeer et al 2016) a paradigm was developed whereby trained anesthesiologists carried out a short simulation task, where they were required to monitor two patients and respond to alarms by indicating the nature of the alarm (its category) and with their reaction times also measured. Prior to this they were given a brief exposure to either the auditory icon plus pointer alarms, or the current IEC alarms. Results indicated very early on that the auditory icons produced faster and more accurate responses than the current IEC alarms (McNeer et al, 2017a, b). Results of the early trials can be seen in Table 2. Secondary workload and fatigue measures were also taken in these studies and there is some evidence that the auditory icons are also less frustrating and impede performance less. Here, we may be tapping into 'alarm fatigue'. This is important,

because though the concept of alarm fatigue is generally accepted, and there certainly is a clinical alarm problem, the details of its manifestation and dimensions are somewhat sketchy (Deb & Claudio, 2015; Kristensen et al, 2016).

A



B

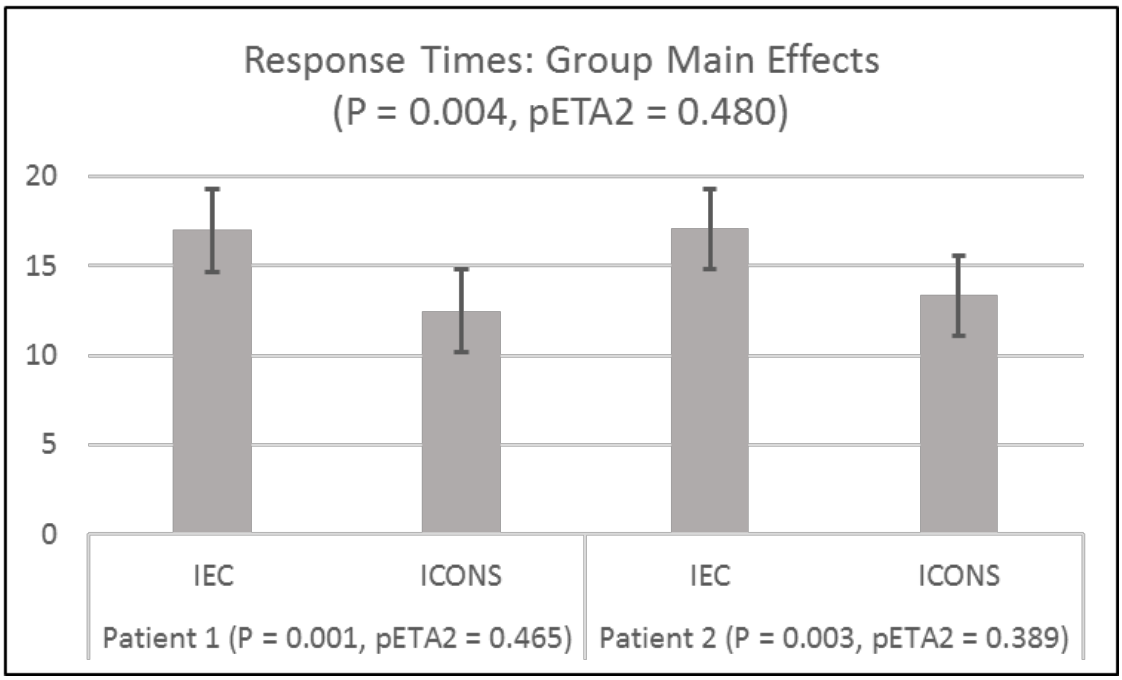


Figure 2: Mean percentage correct identification and mean reaction times to the new sounds ('Icons') or the current IEC set ('IEC') (from McNeer et al, 2017a)

Other studies currently being carried out as part of the formative (and now summative) testing are looking at the audibility of the alarms in realistic listening conditions; findings thus far indicate that the sounds work well in relatively low signal-to-noise ratios (a finding being demonstrated for alarms more generally in other studies, Stevenson et al, 2013) and that the presence of the pointer enhances audibility.

Reflection 3

Because the alarms are intended for the update of the standard and therefore access to them will be of commercial advantage, the final sounds will be released to medical instrumentation companies via a website, through AAMI (the final details of this process are yet to be decided). Several companies are keen to do their own testing on the sounds once they are released.

3.3. Summative testing and other work

The summative testing will follow the broad protocols of the formative testing, with additional researchers testing the sounds in a range of clinical environments using protocols yet to be developed, as well as accepted and published protocols (Stevenson et al, 2013). There is other, related work also. Dr Bolton is currently leading an AHRQ-funded project grant looking at the issue of masking of auditory alarms with specific reference to IEC 60601-1-8 (Hasanain et al, 2017). The project uses a model-checking approach hitherto unused in auditory masking studies. Naturally this project is aware of both the current alarms and the projected new alarms, which helps its validity as a practical instrument and

also pushes the functionality of the software to more complex tasks. We are also carrying out more theoretical studies on the contributions of strength of metaphorical link and auditory diversity in sound sets, as these two dimensions are thought to be large contributors to the effectiveness of any set of alarms.

Reflection 4

The work is on-track to be completed to the satisfaction of the IEC alarms joint working group well before the updated standard is published. By this time, there will be several published papers documenting the performance of the alarms from basic testing to their performance in simulated environments, their performance in noise and in other, increasingly realistic, tasks. Of course, the project won't have reached a satisfactory conclusion until the alarms and the relevant advice is embodied within the standard and there is still a way to go and other possible unknown threats along the way.

We anticipate that our work will improve patient safety and work performance, as well as contributing to the science of alarm design and implementation.

References

Alali KA. (2011) Azimuthal Localization and Detection of Vehicular Backup Alarms Under Electronic and Non-electronic Hearing Protection Devices in Noisy and Quiet Environments. Ph.D. thesis. Blacksburg, Virginia, USA: Virginia Polytechnic Institute and State University; 2011.

Atyeo, J., & Sanderson, P. M. (2015). 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*, *70*(7), 818-827.

Belz, S. M., Robinson, G. S., & Casali, J. G. (1999). A new class of auditory warning signals for complex systems: Auditory icons. Human Factors: The Journal of the Human Factors and Ergonomics Society, 41(4), 608-618.

Bennett, C. L., & McNeer, R. R. (2012). PT-SAFE: A software tool for development and annunciation of medical audible alarms. Anesthesia & Analgesia, 114(3), 576-583.

Block Jr, F. E. (2008). "For if the trumpet give an uncertain sound, who shall prepare himself to the battle?"(I Corinthians 14: 8, KJV). Anesthesia & Analgesia, 106(2), 357-359.

Block FE, Rouse JD, Hakala M, Thompson CL. (2000) A proposed new set of alarm sounds which satisfy standards and rationale to encode source information. J Clin Monit Comp, 16:541– 6

Catchpole, K. R., McKeown, J. D., & Withington, D. J. (2004). Localizable auditory warning pulses. Ergonomics, 47(7), 748-771.

Cvach, M. (2012). Monitor alarm fatigue: an integrative review. *Biomedical instrumentation & technology*, *46*(4), 268-277.

Deb, S., & Claudio, D. (2015). Alarm fatigue and its influence on staff performance. *IIE Transactions on Healthcare Systems Engineering*, *5*(3), 183-196.

Drew, B. J., Harris, P., Zègre-Hemsey, J. K., Mammone, T., Schindler, D., Salas-Boni, R., ... & Hu, X. (2014). Insights into the problem of alarm fatigue with physiologic monitor devices: a comprehensive observational study of consecutive intensive care unit patients. *PloS one*, *9*(10), e110274.

Edworthy, J., Reid, S., McDougall, S., Edworthy, J., Hall, S., Bennett, D., ... & Pye, E. (2017a). The Recognizability and Localizability of Auditory Alarms: Setting Global Medical Device Standards. *Human Factors*, 0018720817712004.

Edworthy, J. R., Schlesinger, J. J., McNeer, R. R., Kristensen, M. S., & Bennett, C. L. (2017b). Classifying alarms: Seeking durability, credibility, consistency, and simplicity. *Biomedical Instrumentation & Technology*, *51*(s2), 50-57.

Edworthy, J., Hellier, E., Titchener, K., Naweed, A., & Roels, R. (2011). Heterogeneity in auditory alarm sets makes them easier to learn. International Journal of Industrial Ergonomics, 41(2), 136-146.

Edworthy, J., Page, R., Hibbard, A., Kyle, S., Ratnage, P., & Claydon, S. (2014). Learning three sets of alarms for the same medical functions: A perspective on the difficulty of learning alarms specified in an international standard. *Applied ergonomics*, *45*(5), 1291-1296.

Graham, R. (1999). Use of auditory icons as emergency warnings: evaluation within a vehicle collision avoidance application. Ergonomics, 42(9), 1233-1248.

Hasanain, B., Boyd, A. D., Edworthy, J., & Bolton, M. L. (2017). A formal approach to discovering simultaneous additive masking between auditory medical alarms. Applied Ergonomics, 58, 500-514.

International Electrotechnical Commission. (2006, 2012). IEC 60601-1-8: Medical electrical equipment–General requirements, tests and guidance for alarm systems in medical electrical equipment and medical electrical systems.

Keller, P., & Stevens, C. (2004). Meaning from environmental sounds: types of signal-referent relations and their effect on recognizing auditory icons. Journal of experimental psychology: Applied, 10(1), 3.

Kristensen, M , Edworthy, J & Ozcan Vieira, E (2017) In search of acoustical signs of alarm fatigue. *Sound Effects*, special issue on *Sound and Listening in Healthcare and Therapy,* 6(1), 88-104

Lacherez, P., Seah, E. L., & Sanderson, P. (2007). Overlapping melodic alarms are almost indiscriminable. Human Factors: The Journal of the Human Factors and Ergonomics Society, 49(4), 637-645.

Leung, Y. K., Smith, S., Parker, S., & Martin, R. (1997). Learning and retention of auditory warnings. Proceedings of the International Community for Auditory Display, 1997. http://www.icad.org/Proceedings/1997/LeungSmith1997.pdf

McNeer, R. R., Bennett, C. L., & Dudaryk, R. (2016). Intraoperative noise increases perceived task load and fatigue in anesthesiology residents: a simulation-based study. Anesthesia & Analgesia, 122(6), 2068-2081.

McNeer, R, Bennett, C.L., Horn, D.B., Dudaryk, R & Edworthy, J (2017a) The Aggregate Performance of Novel Auditory 'Icon' Alarms in a Simulated Intensive Care Unit Environment.Poster presented at International Anesthesia Research Society ,Washington DC, May 2017

McNeer, R, Bennett, C.L., Horn, D.B., Dudaryk, R & Edworthy, J (2017b) Use of Ordinal Regression Analysis for Assessing and Comparing Individual Alarm-type Performance in a Simulated Intensive Care Unit Environment. Poster presented at International Anesthesia Research Society ,Washington DC, May 2017

Miller, G. A. (1956). The magical number seven, plus or minus two: some limits on our capacity for processing information. *Psychological review*, *63*(2), 81.

Morrow, D. G., & Durso, F. T. (2011). Patient safety research that works: Introduction to the special issue on human performance and health care. *Journal of Experimental Psychology: Applied*, *17*, 191-194.

Patterson, RD (1982) Guidelines for auditory warnings systems on civil aircraft. Civil Aviation Authority paper 82017: Civil Aviation Authority.

Patterson RD, Edworthy J, Shailer MJ, Lower MC, Wheeler PD (1986) Alarm sounds for medical equipment in intensive care areas and operating theatres. Report No. AC598: Institute of Sound and Vibration Research, Southampton, UK

Perry, N. C., Stevens, C. J., Wiggins, M. W., & Howell, C. E. (2007). Cough once for danger: Icons versus abstract warnings as informative alerts in civil aviation. Human Factors: The Journal of the Human Factors and Ergonomics Society, 49(6), 1061- 1071.

Petocz, A., Keller, P. E., & Stevens, C. J. (2008). Auditory warnings, signal-referent relations, and natural indicators: re-thinking theory and application. Journal of experimental psychology: Applied, 14(2), 165.

Sanderson, P. M., Wee, A., & Lacherez, P. (2006). Learnability and discriminability of melodic medical equipment alarms. Anaesthesia, 61(2), 142-147.

Sendelbach, S., & Funk, M. (2013). Alarm Fatigue A Patient Safety Concern. *AACN advanced critical care*, *24*(4), 378-386.

Stephan, K. L., Smith, S. E., Martin, R. L., Parker, S. P., & McAnally, K. I. (2006). Learning and retention of associations between auditory icons and denotative referents: Implications for the design of auditory warnings. Human Factors, 48(2), 288-299.

Stevenson, R. A., Schlesinger, J. J., & Wallace, M. T. (2013). Effects of divided attention and operating room noise on perception of pulse oximeter pitch changes; A laboratory study. The Journal of the American Society of Anesthesiologists, 118(2), 376-381.

Ulfvengren, P. (2003). Associability: A comparison of sounds in a cognitive approach to auditory alert design. Human factors and aerospace safety, 3(4), 313-331.

Vaillancourt, V., Nélisse, H., Laroche, C., Giguère, C., Boutin, J., & Laferrière, P. (2013). Comparison of sound propagation and perception of three types of backup alarms with regards to worker safety. Noise and Health, 15(67), 420.

Wee, A. N., & Sanderson, P. M. (2008). Are melodic medical equipment alarms easily learned?. Anesthesia & Analgesia, 106(2), 501-508.

Welch, J. (2011). An evidence-based approach to reduce nuisance alarms and alarm fatigue. *Biomedical Instrumentation & Technology*, *45*(s1), 46-52.

Whalen, D. A., Covelle, P. M., Piepenbrink, J. C., Villanova, K. L., Cuneo, C. L., & Awtry, E. H. (2014). Novel approach to cardiac alarm management on telemetry units. *Journal of Cardiovascular Nursing*, *29*(5), E13-E22.