UPDATING AN INTERNATIONAL MEDICAL DEVICE STANDARD:
A PROCESS FOR AUDIBLE ALARMS

Judy Reed Edworthy, Cognition Institute, Plymouth University, UK PL4 8AA
Richard McNeer, Miller School of Medicine, University of Miami, Miami, FL 33101
Christopher Bennett, Frost School of Music, University of Miami, Miami, FL 33146
Roman Dudaryk, Miller School of Medicine, University of Miami, Miami FL 33101

Many human factors and ergonomics problems are associated with clinical alarms, usually referred to as 'alarm fatigue'. Among these problems is the nature of auditory signals used to attract attention, as these signals are often difficult to learn, easily confusable, and sometimes prone to masking. Symptomatic of this problem is the poor quality of the audible alarms associated with a global medical device safety standard, IEC 60601-1-8. A project aimed at improving and updating these sounds according to best practice is being carried out. This paper charts the progress of this venture and summarizes the results and the published papers which present those results.

INTRODUCTION

The global medical device standard IEC 60601-1-8 (2006, 2012) is a horizontal standard governing all medical devices with electrical components. It is thus relevant to the overwhelming majority of such devices. Parts 1-8 of the standard cover many aspects of safety, including visual and auditory alarms. The standard was updated in 2012 and is due for another update in 2019. One focus of the current update is to improve the auditory alarms specified in the standard. This paper deals with the process of updating those auditory alarms.

IEC 60601-1-8 AUDITORY ALARMS

IEC 60601-1-8 was first published in 2006, and republished with some modifications in 2012 (IEC, 2006, 2012). It is due to be published and updated again in 2019. The standard covers many issues relevant to medical device safety and includes a specific focus on clinical alarms and alerts. The standard also specifies a set of auditory alarms for the specific clinical hazards denoted in the standard. These eight alarms cover the clinical situations/hazards denoted in Figure 1.
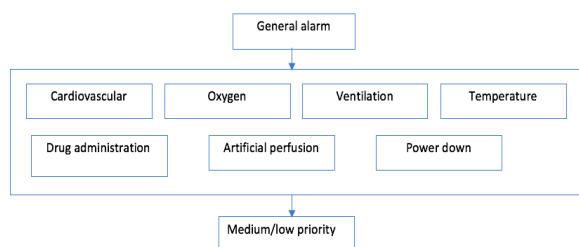

Figure 1: Clinical hazards covered by IEC 60601-1-8

The standard specifies the tonal, temporal, and amplitude structure (with some latitude for some of the values) for the eight auditory alarms to a high level of specificity (Block, Rouse, Hakala & Thompson, 2000). There is one alarm for each of the following hazards/categories: General,

Cardiovascular, Oxygenation, Ventilation, Temperature, Drug Administration, Artificial Perfusion, and Power Down). The standard also differentiates between high, medium, and low urgency versions of each of these alarm where urgency is represented by variants of the alarms, rather than by obviously different alarms.

Within an urgency grade (for example, high priority) there is very little variation across the set of eight alarms, making them difficult to learn and retain (Sanderson, Wee & Lacherez, 2006; Wee & Sanderson, 2008). Notably, they are all tonal patterns with the same rhythm and the same pulse structure, meaning that the small variation in tonal pattern is the only way of telling the alarms apart. It is not surprising therefore that research findings (collected only after the standard was published) indicated that the alarms were difficult to learn and retain, that musicians did better than nonmusicians in remembering the alarms, and that the "General" alarm, which has a fixed pitch pattern, is the only alarm that is easy to learn and recognize.

Even at the time of design, research evidence would suggest that this was not a good design remit to follow but the alarms have remained in the standard, together with the acknowledgement of the designer of the shortcomings of those signals (Block, 2008). In due course, sufficient traction over the whole problem (or set of problems) related to clinical alarms has been generated to tackle the specific problem of the auditory alarms.

ALARM FATIGUE

'Alarm fatigue' is a term which is often used as an expression of all that is bad about clinical alarms: it encapsulates the various issues of high false alarm rates, alarms that are excessively shrill and loud, alarms which are masked by other alarms and sounds, alarms which cannot be understood and interpreted, and any other problems caused by alarms not functioning as they might. 'Alarm fatigue' can be thought of as a set of related issues around clinical alarms, with a range of potential solutions from different domains. The narrative of alarm fatigue also seems to hint that the problem of alarm fatigue will disappear if the number of

alarms (particularly false alarms) is reduced. Very little is said about the auditory (and visual) signals themselves (Kristensen, Edworthy and Özcan Vieira, 2017) and one of the few studies which has attempted to measure the effect of likely causes of alarm fatigue on reasonable measures of nurses' alarm fatigue failed to find any relationship between alarm fatigue and the number of alarms presented (Deb & Claudio, 2015). This suggests that there is more to alarm fatigue than simply the number of alarms heard. Thus reducing the number of alarms in any clinical environment might reduce the problem but will not eliminate it. For example, it seems logical that if the meaning of an alarm isn't clear, then it still won't be clear whether the clinician is exposed to 4 alarms or 100 alarms per hour (though there will only be 4 unidentified alarms!). The problem of alarm meaning will remain, even when other measures which serve to reduce the problem have been successfully implemented.

As there have been many success stories demonstrating how the number of alarms can be successfully reduced in clinical environments (for example at Johns Hopkins Hospital, AAMI, 2012), the need to improve the auditory alarm signals themselves has become more evident. Thus the impetus to improve and update the audible alarms associated with IEC 60601-1-8 can be seen as one more aspect of the multifaceted response to the 'alarm fatigue' problem.

THE PROCESS

Designing and benchmarking the sounds

During the first phase, we engaged in an evidence-based iterative design process to formulate several different design alternatives which we posited would perform better than the current IEC alarms. After a design phase, each resultant set of alarms was tested for learnability and localizability. As with all stages of the process, the other key enterprises were the publication of the findings in the public domain, and presentation of the findings to the relevant committees in order to focus and drive the work.

The design, development and preliminary testing of candidate sounds and sound sets is presented in two articles (Edworthy, Page, Hibbard, Kyle, Ratnage, & Claydon (2014); Edworthy, Reid, McDougall, Edworthy, Hall, Bennett, Khan, & Pye (2017)). These papers demonstrate that different types of sounds vary in their degree of learnable largely as a function of the degree to which the sound is a metaphor for its function, a finding that was already well-known in the literature. Thus, we established (not for the first time) that auditory icons were easier to learn than any other type of sounds tested, and therefore appeared to be the best candidate for further development and scrutiny (for previous work which has demonstrated the superiority of auditory icons, see Belz, Robinson & Casali, 1999; Graham, 1999; Keller & Stevens, 2004; Leung, Smith, Parker, & Martin, 1997; Perry, Stevens, Wiggins, and Howell 2007; Petocz, Keller, & Stevens, 2008; Stephan, Smith, Martin, Parker, & McAnally, 2006; and Ulfvengren, 2003). The key reason for auditory icons being easier to learn than other alarm types is that there is a clear link between sound and function – if the sounds are wisely

chosen – but also partly because auditory icons tend to be acoustically varied when a set of them is used. We have recently demonstrated that both these factors contribute to the learnability of alarm sounds (McDougall, Sinimeri, Edworthy, Goodliffe, & Bradley, 2017).

We do not know for sure that the learnability of an alarm is of practical importance (though *a priori* we would expect it to be). However, the only data that exists on the current set is their learnability, so for the purposes of comparison it seemed logical to start here. Our studies demonstrated that the current IEC alarms were significantly harder to learn than any of the candidate sets tested. The other sets also varied considerably in their learnability, with auditory icons being the most easily learned. This finding was not surprising, but nevertheless needed to be demonstrated.

In addition to testing learnability, the first phase of the benchmarking also included the testing of localizability. This is the propensity of the sound to be localized/located within a set of possible locations. There is no data on localizability for the current IEC alarms, but again we reasoned that localizability might be important as in the hospital environment it is often necessary to locate a particular bed in a multibed ward (for example, an ICU or a recovery room) and thus improved or enhanced localizability might also be important.

Localizability of sound is largely a function of the harmonic density of a sound – white noise is the most localizable sound, and it consists of all frequencies – so it is possible to predict the relative localizability of sounds designed according to different remits. Research studies on alarm localizability have shown that more harmonically dense alarms produce more accurate localization (Catchpole, McKeown, & Withington, 2004; Vaillancourt, Nélisse, Laroche, Giguére, Boutin, & Laferrière, 2013). Three of our sets of sounds were more harmonically dense than the other two sets (including the current IEC set) and indeed our results indicated that the more harmonically dense alarms resulted in more accurate localization than those lower in harmonic density. Included in the harmonically dense sets were, again, auditory icons.

The initial benchmarking therefore established that all of the alarm sets that we put forward as alternatives performed significantly better than the existing IEC auditory alarms (except in one case, where a set of alarms we had designed to be suitable for low-end equipment and therefore were acoustically simple, did not outperform the existing IEC alarms on localizability, as we could have predicted).

It is worth noting the degree to which our new designs improved performance for both learnability and localizability. Whereas the existing IEC alarms were still not well learned (below 50% performance) after ten exposures to them, performance with the auditory icons was about 80% correct after a single exposure, increasing to 90-95% after a couple of trials. The performance for localizability is also noteworthy. Whereas participants mislocalized one in ten of the auditory icons, they mislocalized one in four of the current IEC alarms. In a subsequent study (Edworthy, Reid, Peel, Lock, Williams, Newbury, Foster, & Farrington, 2018, in press) where workload was manipulated whilst participants were required to

localize only the auditory icon alarms, participants mislocated one in four alarms only when they had to attend to the alarms both in noise and while simultaneously having to perform either a reading or a mathematics task. Thus localization performance was approximately the same for the auditory icons when under medium to high workload as it was for the existing IEC alarms when the participant was required to do nothing other than to locate the position of the alarm. In practical terms, this is a large premium.

Obtaining appropriate sanction and approval

A key part of the work was to conduct the research with the full knowledge (and ultimately the sanction) of the bodies who are able to effect an update to the standard. Ultimately this is managed through the IEC committee concerned with this particular standard, and the network which surrounds the standard. In this case, that is the corollary AAMI (Association for the Advancement of Medical Instrumentation) 60601-1-8 committee, and an overarching alarms committee known as the IEC Joint Working Group on alarms which has input to both of these (and other) committees. To this end, funding to carry out the work was sought by the first author and obtained from AAMI, and the program of work approved by the Chairs of the relevant committees. Once the benchmarking work was complete (the learnability and localizability work), the findings were presented to these committees, who gave feedback and direction for the work to follow. After considerable consultation and discussion, the committee decided that it would like to move forward with the best performing sounds, which were the auditory icons. Specifically, they preferred the 'auditory icons plus pointer' set of sounds, which consist of an auditory icon with a short abstract sound embedded within it (called the 'pointer'). The purpose of this pointer is first to indicate that the sound is an alarm (as auditory icons are everyday sounds, their status as alarms might not be clear in some contexts) and also to allow urgency differentiation through the use of pointers of different priorities.

A further program of work was developed on the basis of the committees' expectations, which was then funded by AAMI and approved again by the Chairs of the committees. Thus the research team and the bodies able to recommend acceptance of the updated alarm sounds worked towards a common aim from early on in the project.

Simulation

The next phase of the work was to carry out simulation work on the sounds as a further step in formative testing, then moving on to summative testing.

Performance in a simulated work environment. Our simulations involved testing the alarms both in a clinical simulation and in a lab-based noise simulation. We did not initially want to test the sounds in both a clinical simulation and a noise environment, in order that we might understand the effects of those two big steps towards full clinical testing clearly before placing them together in a more complex and realistic simulation.

The auditory icon plus pointer set of sounds was initially tested in a controlled simulation which compared the current IEC alarms with those proposed new alarms. Whilst we were keen to drop testing of the current alarms, it was important to keep the current alarms in the testing in the early phases of the work in order to provide some kind of comparison for the new alarms.

In the first simulation, participants, who were mostly resident anesthesiologists at the Jackson Memorial Hospital, Miami, were given a Powerpoint presentation either of the current IEC alarms or the proposed new set. This presentation told them the meanings of the eight alarms as in Figure 1, and played them the alarm sounds. They then participated in a 20-minute simulation where each of the alarms sounded at least once, and where they were required to indicate the meaning of the alarm. Their responses (correct/incorrect) and their response time were recorded. At this point we were simply interested in whether participants could recognize the alarms that they had been presented with prior to the simulation.

Our results were quite compelling (McNeer, Bennett, Horn, Dudaryk, & Edworthy, 2017a; McNeer, Bennett, Horn, Dudaryk, & Edworthy, 2017b; McNeer. Bodzin Horn, Bennett, Edworthy, & Dudaryk, in press). Correct recognition of the proposed new alarms was approximately double that of the IEC alarms (around 80% as opposed to below 40%) and response times were about a third lower. This also needs to be seen in the context that some of the current IEC alarms might already have been familiar, at least in the sense that some of the participants are likely to have been exposed to them in their clinical work.

Interesting also was that participants were asked about their subjective fatigue and workload, and on a couple of measures there were significant differences between the groups dependent on which set of alarms they had been exposed to. Here, the proposed new alarms scored lower on workload and fatigue-related issues, suggesting that as well as enhancing performance, the use of auditory icons as alarms might tap into the more general issue of 'alarm fatigue'. For example, if an alarm is easy to learn and/or is meaningful, then it may be able to reduce alarm fatigue beyond that which might be achieved by a similar number of less meaningful alarms.

This first of the simulation studies served as a proof-of-concept for the idea of using auditory icons as alarms, and indeed suggested that their use as alarms in real clinical practice might be possible, and that furthermore there might be considerable advantage in doing so.

These simulation studies were then followed up by other simulation studies which tested variants of the auditory icons to ensure that the most effective versions of the auditory icons for each of the clinical hazards (Figure 1) had been generated. A series of simulations were carried out where clinically-trained participants (different in each condition) were presented with and asked to respond to one of three different versions of the alarm for each of the eight alarm categories, and the data summed to show the best and worst performing. When completed, the best and worst performing sounds for each function were compared to show the benefits of specific auditory icons over one another for the clinical categories. We

did indeed find differences between the possible auditory icons for each of the functions. Also, by and large the original icon idea (as in Edworthy et al 2017) turned out to be the most appropriate icon when compared with other possibilities. This confirmed the usefulness of the initial iterative design procedure.

As with the benchmarking work, we aim to put all our data into the public domain eventually. The first study is now in press (McNeer et al, 2018) and the later work will be submitted in due course.

Performance in noise. Our initial testing in noise involved carrying out a standard psychophysical test of the audibility of the auditory icons plus pointer in noise generated from recordings of real ICU noise. Again, three versions of the sounds (three different auditory icons) were tested and the best performing alarms identified. The results closely followed those for the simulation study in terms of best and worst performing sounds. However, the most important thing to note in these studies was that even the weakest signal-to-mask ratio gave a value of -10dB SPL, meaning that the sound which performed worst in noise (needed to be relatively louder) was audible when that sound was 10dB SPL lower than that of the noise in which it was required to be detected. The high-priority pointer itself was detectable when presented in noise that was four times louder than the sound itself.

Again, we are in the process of writing up these studies for publication.

NEXT PHASE

Further testing and development

Now that the benchmarking and simulation work is complete, the sounds have been published for reference in a pre-final draft of the standard and have been commented upon from those able to vote on the progress of the standard. The sounds are also available for any company or laboratory (or any other enterprise) wishing to carry out their own tests on the sounds.

The next phase of the experimental work is to carry out further summative work in clinical settings and to collect usability and acceptability data from clinicians who would encounter the sounds once the standard is adopted. It is expected that the program of work will continue both within the project and outside of the project together with other interested parties.

A related project of note is one being carried out at the State University of New York at Buffalo, which is developing a formal methods approach to the auditory masking of alarms, another key issue in the alarm fatigue narrative (Hasanain, Boyd, Edworthy, & Bolton, 2017). This project has harnessed the potential of formal methods (a computer technique) in the field of auditory masking for the first time. Here, the technique allows the modelling of all possible situations where two, three or more alarms might begin at approximately the same time, and show those specific conditions where one or more of the alarms might be masked by one or more of the other alarms. The project focuses on the current IEC 60601-1-

8 alarms, and has shown that masking of one alarm by another is possible with the current alarm sounds. In due course, it is anticipated that the method will be developed in a way which will allow contemplation of the proposed new alarms in addition.

The standard

As well as having well-tested and benchmarked auditory alarms which will be available for use (either as a download or in tabular form), the standard will have a raft of other improvements in terms of its auditory alarm components.

First, those references connected to the development of the alarms which are either published or are in press will be listed in the standard so that the reader can refer to those articles. Second, the benchmarking data will be provided in tabular form, so that if manufacturers wish to develop their own versions of the alarms, they will be able to see what kind of performance can be achieved with the alarm set specified in the standard. This benchmarking data will summarize the data on (at least) learnability, localizability, audibility and performance in simulation as described in this paper. Third, annexes will be provided on how to generate and test alarm sounds if manufacturers wish to develop their own.

The standard will be published towards the end of 2019.

ACKNOWLEDGEMENTS

REFERENCES

AAMI (Association for the Advancement of Medical Instrumentation) (2012) Safety Innovations: Using Data to Drive Alarm System Improvement Efforts. AAMI White paper, http://www.premiersafetyinstitute.org/wp-content/uploads/Johns-Hopkins-White-Paper.pdf

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.

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.

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

Edworthy, J., Reid, S., McDougall, S., Edworthy, J., Hall, S., Bennett, D., Khan, J, & Pye, E. (2017). The Recognizability and Localizability of Auditory Alarms: Setting Global Medical Device Standards. *Human Factors*, 59(7), 1108-1127

Edworthy, J, Reid, S, Peel, K, Lock, S, Williams, J, Newbury, C, Foster, J & Farrington, M (in press) The impact of workload on the ability to localize audible alarms. *Applied Ergonomics*

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 & Özcan 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

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

McDougall, S, Sinimeri, D, Edworthy, J, Goodliffe, J & Bradley, D (2017) Searching for meaning in sound: How alarm sounds are learned and interpreted in visual environments. Presented at the HFES European chapter meeting 2017, Rome, September 2017.

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

McNeer, R, Bodzin Horn, D, Bennett, C L, Reed Edworthy, J, & Dudaryk, R (in press) Auditory icon alarms are more accurately and quickly identified than IEC 60601-1-8 melodic alarms in a simulated clinical setting *Anethesiology*

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.

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.

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.