The recognizabiity and localizability of auditory alarms: setting global medical device standards

Judy Edworthy[1], Scott Reid[1], Siné McDougall[2], Jonathan Edworthy[1], Stephanie Hall[1], Danielle Bennett[1], James Khan[1] & Ellen Pye[1]

[1] Cognition Institute, Plymouth University, Drake Circus, Plymouth, Devon PL4 8AA, UK

[2]Department of Psychology, Faculty of Science and Technology Bournemouth University, Poole, Dorset, BH12 5BB

Running head: Medical audible alarms

Manuscript type: Research article (multiple studies) Exact

Word count: 7946 words

Acknowledgements: This research was supported financially by the Association for the Advancement of Medical Instrumentation, Arlington, VA 22203-1633 with a grant to the first author

Objective: Four sets of eight audible alarms matching the functions specified in IEC 60601-1-8 (2012) were designed using known principles from auditory cognition, with the intention that they would be more recognizable and localizable than those currently specified in the standard.

Background: The audible alarms associated with IEC 60601-1-(2012), a global medical device standard, are known to be difficult to learn and retain, and there have been many calls to update them. There are known principles of design and cognition which might form the basis of more readily recognizable alarms. There is also scope for improvement in the localizability of the existing alarms.

Method: Four alternative sets of alarms matched to the functions specified in IEC 60601-1-8 (2012) were tested for recognizability and localizability, and compared with the alarms currently specified in the standard

Results: With a single exception, all prototype sets of alarms outperformed the current IEC set on both recognizability and localizability. Within the prototype sets, 'auditory icons' were the most easily recognized, but the other sets, using word rhythms and simple acoustic metaphors, were also more easily recognized than the current alarms. With the exception of one set, all prototype sets were also easier to localize.

Conclusion: Known auditory cognition and perception principles were successfully applied to a known audible alarm problem

Application: This work constitutes the first (benchmarking) phase of replacing the alarms currently specified in the standard. The design principles used for each set demonstrates the relative ease with which different alarm types can be recognized and localized.

Keywords: audition; auditory displays; learning; medical device technologies

Precis: Four sets of audible alarms matched to the functions specified in IEC 60601- 1-8, a global medical device standard, were designed using known principles of successful audible alarm design. When tested for recognizability and localizability, all sets (with one exception for localizability) outperformed the current alarms specified in the standard. This work represents the first phase of updating the alarms specified in that standard

Accepted for publication: 15 April 2017

INTRODUCTION

The global medical device standard IEC 60601-1-8 (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) has considerable reach in the sphere of medical instrumentation. According to the American National Standards Institute (ANSI) the standard 'Specifies basic safety and essential performance requirements and tests for alarm systems in medical electrical equipment and medical electrical systems and provides guidance for their application. This is accomplished by defining alarm categories (priorities) by degree of urgency, consistent alarm signals and consistent control states and their marking for all alarm systems' (webstore.ansi.org). The standard was first published in 2006 and was republished in 2012 with some refinements and an amendment to the fine detail of the audible alarms, namely an increase in the length allowable for the onset rise time of the alarm pulses.

The standard specifies seven alarm risk categories and one general category (Kerr, 1983; Kerr & Hayes, 1985, Figure 1). In addition, the standard contains a 'reserved' set of eight audible alarm sounds for these categories. These alarm sounds are mapped to the eight categories, and in line with the prioritizing suggested in the standard, there are both high- and medium-priority versions of the eight alarm sounds. The differentiation between these categories is achieved by presenting the high priority alarm as a 5-pulse rhythmic unit (da-da-da---da-da) twice in quick succession, whereas the medium priority sound is references by a single, 3-pulse unit presented at a slower pace than the high priority alarm (da—da—da). A low priority or information sound is also specified, which can be in the form either of a hostess ding-dong call or a single tone. Reminder signals are also specified separately.

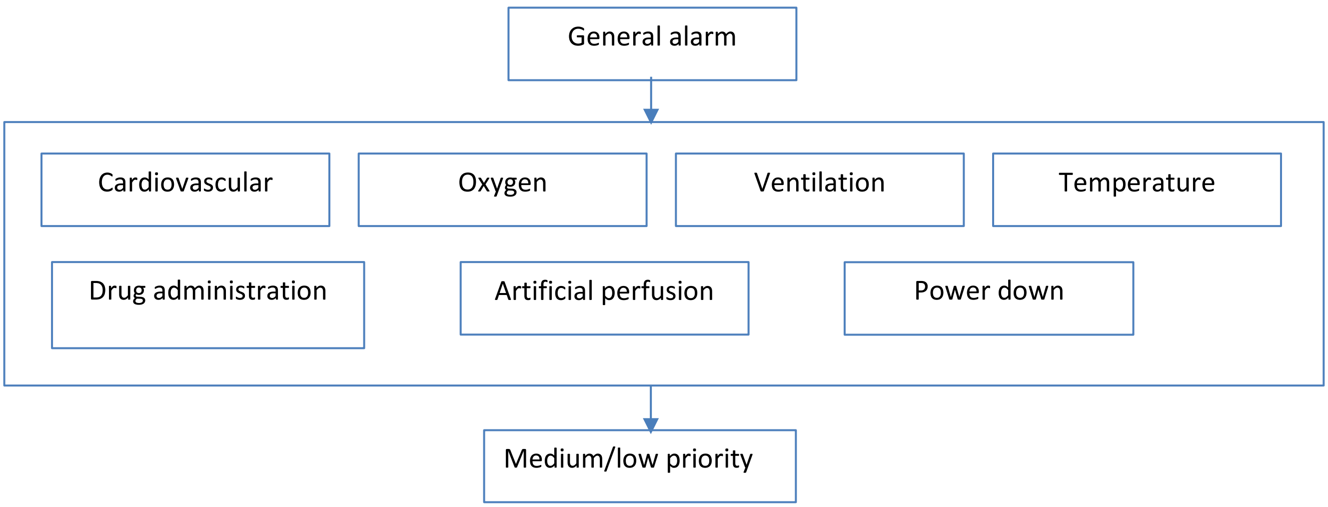


Figure 1: General alarm structure and categories of IEC 60601-1-8

The alarms currently supporting the standard were put forward by Block et al (2000) and are as indicated in Appendix 1a. The alarms are tonal and are sometimes referred to as 'melodies'. They are constructed from pulses of sound with at least four harmonics (pure tones) in order to provide some resistance to masking, and to aid localizability. The alarms follow some, but not all, of the known (at the time of design) principles of imbuing an appropriate sense of urgency into the alarms **(**Arrabito et al, 2004; Edworthy et al 1991; Finley & Cohen, 1991; Haas & Casali, 1995; Haas & Edworthy, 1996; Guillaume et al, 2003; Hellier et al 1993; Momtahan, 1991).

Studies around 2006-2008 demonstrated that clinicians found it difficult to learn and distinguish between the alarms even after repeated exposure on more than one testing occasion (Lacherez, Seah & Sanderson, 2007; Sanderson, Wee & Lacherez, 2006; Wee & Sanderson, 2008). Furthermore, clinicians with musical training performed better at the learning task than those without (Sanderson et al, 2006; Wee & Sanderson, 2008). The key reasons for this learning difficulty is the high level of similarity between the sounds (which only differ according to their pitch patterns, see Appendix 1a) and the lack of link between sound and function. Underscoring the acknowledgement of the suboptimality of the current alarm set is the designer's own stated opinion (now published a few years ago) that the current alarms need to be updated and replaced (Block, 2008).

There is thus a strong imperative to improve and update the alarm sounds currently supporting IEC 60601-1-8. Huge advances in both technology and understanding have been made since the inception of the existing alarm sounds, which were (even at the time of adoption) known (or could be predicted) to be sub-optimal. This makes the use of better, richer, more ergonomically-designed alarms a possibility and a safety imperative (particularly given the length of time needed to incorporate new elements in a standard and the subsequent longevity of those changes.

The current alarms bear little implicit relationship to their meanings so any association therefore hasto be learned (Petocz et al, 2008). Some attempt was made to provide a link between sound and meaning in the current IEC 60601-1-8 alarm sounds by suggesting ways in which the sound mapped to its meaning via its pitch pattern. For example, the pitch pattern of the ventilation sound (Appendix 1a) can be thought as being akin to the rise and fall of breathing (as the alarm rises and falls in pitch pattern) and the power down alarm falls in pitch from start to end, which can be likened to the failure of a power source. Strong association between sound and meaning help people to learn alarms more quickly. This is achieved by the alarm sound being a metaphor for the function it is describing. ‘Good’ metaphors (an interesting question of itself) can lead to very quick recognition needing only one or two exposures to the sound (Belz et al, 1999; Edworthy, Page et al, 2014; Graham, 1999; Leung et al 1997; Perry, Stevens et al 2007; Stephan et al, 2006; Ulfvengren, 2003). One of our goals in designing prototype sets of alarm sounds is to move on from small-scale tweaking of tonal alarm sounds (and consequently small-scale tweaking of the standard) which generally achieve only small improvements, to sets of alarm sounds which can be learned after very brief exposure. For example, ‘auditory icons’, which are usually everyday sounds with clear metaphors and meaning, can sometimes be recognized after only one or two exposures to the sound. The possible use of auditory icons as clinical alarms therefore has potential benefit in an age where sound reproduction and quality can be both excellent and inexpensive.

The standard is also concerned with psychoacoustic issues. The standard specifies that the alarm sounds should have a fundamental frequency of between 150 and 1000Hz, and should contain at least four harmonics within the range 300-4000Hz. This is to help aid localization and resistance to masking (Patterson, 1982). The physics of sound localization and masking is complex (Blauert, 1997; Zwicker & Fastl, 2013) but by and large (with some caveats) both are helped by increasing the harmonic richness or harmonic density of the sound, which is most simply interpreted as the number of harmonics contained within the sound. Very generally speaking, the more harmonically dense a sound is, the easier it will be both to localize and to provide resistance to masking. The small amount of research specifically concerning the localizability of alarm sounds which does exist (Alali, 2011; Vaillancourt, Nélisse et al, 2013; Catchpole, McKeown & Withington, 2004 demonstrates that more harmonically rich and dense sounds produce better localization, or fewer errors in localization. Broadband noise performs particularly well.

Another benefit of the specification of harmonic structure in the standard is that the range it specifies for the fundamental frequency of the alarm (the pitch that is heard) is lower than would normally be used in alarms, which also helps in improving localizability and in reducing the aversiveness of the alarm (Edworthy, 2017).

Thus while the alarms currently supporting IEC 60601-1-8 are an improvement on the most mediocre types of shrill, acoustically poor and hard-to-learn alarms, they are still in need of updating. In this paper we present the initial benchmarking of four alternative sets of alarms that might form the basis of this update. As there is no standardized or accepted method for evaluating an auditory alarm, the obvious place to start in the development and benchmarking of proposed new alarms is to test them in a way that allows direct comparison to be made with the current IEC 60601-1-8 alarms. The only experimental data available for these is learning data, so our first study compares the learnability (more specifically, recognizability) of other potential designs with the current sounds. Here, we focus on recognizabiity after a single exposure to each alarm sound within each alarm set, rather than conducting a full-blown learning trial, as this level of exposure is more typical of the way alarm sounds are introduced into the workplace (indeed, the first time a clinician hears an alarm sound in a new piece of equipment could well be in a real clinical situation, and we know that some types of sound are very easy to learn). Earlier studies with sounds which provide good metaphors (particularly auditory icons) also suggest that reliable recognizability can be achieved after a single, or at least a very small number of, exposures to the sound. In our study, participants are exposed to the alarms on multiple occasions but are only told its meaning on the first presentation of the sound. Our second study investigates the localizability of the alarms, as predictions can be made about the relative localizability of the various sets tested. Localizabiity is important for clinical alarms in contexts where carers may not always be present at a bedside, and/or where there is more than one bed in a unit, as is typical of many Intensive Care Units. If an alarm has enhanced localizability when, say, a nurse is in an ICU ward and needs to identify which patient requires attention, the most efficient way of achieving this is that their ear is drawn to the correct patient rather than via some other indirect, more circuitous rout

STUDY 1: ALARM RECOGNIZABILITY

Alarm sets

This study is concerned with the comparison of four sets of potential alarm sounds designed to represent different ways of improving both variability and the sound-meaning relationships within an alarm set. It is not a factorial experiment aimed at determining the extent to which acoustic parameters and other elements affect learnability, but rather uses the available literature to design possible sets which would be expected to be an improvement on the current set.

Two key factors known to be important in influencing the degree to which alarms (or indeed any sounds intended to convey meaning) can be learned and/or recognized are the degree to which the sound and its referent are related, and the degree of variability in the sounds, which can be thought of as the number of dimensions along which the sounds vary. In the study presented here, this variability is achieved through evaluation which is to some extent the subjective view of the experimenters but is currently a topic of more formal evaluation in our laboratory.

As all of the prototype alarm sets were developed as potential replacements for the current alarms, the degree of variability in each set was intended to be substantial. The sound-referent relationships are determined in different ways for each set, but in each set there is a principle or set of principles on which the relationship has been developed. Finally, the design remits adopted for the four sets of experimental sound meant that three of them (auditory icon, auditory icon plus ident, and word rhythm) were harmonically dense while the current IEC alarms and the ‘resilient’ alarms are more harmonically sparse by comparison. The IEC standard specifies that the pulses of the alarm bursts must contain at least four harmonics, and alarms with this configuration are often used in practice.

The design of the alarm sets reported here started from the agreed position that each of the sets of alarms should cover the following ten functions as specified in the current standard: General; Perfusion; Cardiovascular; Drug administration; Oxygen; Power down; Temperature; Ventilation; medium priority; and low priority. Each set consisted of ten alarm sounds, one for each of those ten functions.

The principles driving the four prototype alarm sets are shown are described below, and the detail of the acoustic and temporal features are shown in Appendix 1(a) to 1(d). The eight function alarms are all intended as the high-priority version of that alarm. In addition to the eight high-priority function alarms, a single medium and a single low priority alarm were added. In the standard, there are high- and medium-priority versions of each of the eight alarms, and a single low-priority alarm. However, the medium-priority alarms are variants of the high-priority alarms and during the planning stages of this work it was established that the status of the high-priority alarms was of much greater significance than the medium priority alarms, and that in future versions of the standard consideration might be given to a single medium-priority alarm as well as a single low-priority alarm, which already exists within the standard.

'Word rhythms'

These alarms attempt to copy the rhythms of the words of the functions that they represent. They were intended to function as mnemonics for the alarm meanings expressed in a word or words. For example, the cardiovascular alarm is represented by a 6-pulse unit mimicking the word 'cardiovascular', power down is represented by a three-pulse unit in the rhythm 'Po-wer----down' and so on (see Appendix 1b). This linking should improve learnability relative to the IEC set, and also improve the differentiation between the alarms (and conversely also be the possible source of confusion between alarms with the same number of syllables). The sounds were constructed so as to have different timbres from one another and to contain a relatively large number of harmonics. All of the sounds (aside from the low and medium priority sounds) contained at least a dozen discernible harmonics, often in excess of 20. The low-and medium-priority sounds contained fewer harmonics (less than a dozen). Of course, word rhythms are highly language dependent so would be different for different languages depending on how the categories translate. When a standard is published it is initially published in English and in French, and once adopted is routinely translated into other languages such as Chinese, Japanese, German and so on before being adopted in those countries. It is important to note that we are testing only an English version of the words corresponding to the categories in this study.

‘Auditory icons’ and ‘auditory icons plus ident’

‘Auditory icon’ is a term used to describe any sort of sound that has an obvious link to the hazard it is representing. It is a rather loose term but in general is used to refer to everyday sounds which can act as metaphors for the hazards that they represent. Depending on how well the link is made, alarms which are auditory icons can be very easy to learn – in many cases next to no learning is required. We selected seven sounds to act as metaphors for the seven specific functions for which alarms were required. For example, we use a drumming sound for cardiovascular, a pillbox shaking for drug administration and so on (see Appendix 1c). We adapted the current (IEC) general sound into a much quicker and repeated version of the 5-pulse alarm as the auditory icon for ‘general’. This sound is well-known in its slower form to clinicians and so can serve as an auditory icon as its meaning is well-learned (Graham, 1999; Petocz et al, 2008).

We constructed a second set of auditory icons by adding an ‘ident’ to the sounds described above. ‘Ident’ is short for ‘identifier’, which is typically used in a visual format. Visual idents are short and usually concise visual images, typically used by TV stations, to allow quick identification of the channel. We are using the auditory equivalent of this by providing a short and concise signal embedded in the auditory icon that should allow identification (and potential classification) of the auditory icon, in this case as an alarm.

The ident used was the ‘auditory icon’ version of the general (Table 1(c), ‘General’). We included the ‘auditory icon + ident’ as an additional set as the addition of the ident reinforces the status of the sound as an alarm – a possible implementation and aesthetic issue in future testing and evaluation.

Because the auditory icons used are real-world sounds, they possess a complex and usually rich harmonic structure. Identifying the precise number of harmonics in each of the sounds is thus difficult to achieve, but all of the sounds used possessed more than 16 harmonics, often many more than this. The only exception to this was the general and the low-and medium-priority sounds, which were more abstract in nature and contained between 6 and 12 harmonics.

‘Resilient’

Advances in technology and sound processing and storage mean that it is now considerably easier to store and play complex sounds like those embodied in the word rhythm and the auditory icon sounds. The design of the ‘resilient’ set is based on the assumption that sometimes a low-fidelity sound device may be used, which may compromise the quality of the alarm sounds. Here, the meanings of the alarms are achieved through the use of simple metaphors that are unlikely to be degraded no matter how bad the device used to reproduce the sound. The ‘resilient’ set of sounds are a more acoustically simple set of sounds produced recorded at a sampling rate of 8kHz rather than 44.1 kHz (the rate for the other sets of sounds), and possess fewer harmonics than the other alarm sets (other than the IEC alarms; see Appendix 1d). The meanings are achieved either through a simple metaphor such as a falling pitch (used for power down), or different numbers of pulses, in a similar way to the ‘word rhythm’ alarms. Harmonically, these sounds are less complex and dense than the word rhythm and auditory icon alarms. Most of the sounds possessed six harmonics. Because these alarms represent greater variation across the set than the IEC alarms, the precise values of these harmonics (in terms of Hz) could vary as the sound progressed (for example they could increase or decrease in value, if the sound went up or down in pitch).

IEC alarms

The IEC alarms tested were constructed exactly as specified in the current standard and were as shown in Appendix 1(a). Each of the sounds possessed 5 harmonics in total. The values of the frequencies were fixed throughout the duration of the sound.

Method

Materials

Each participant was asked to learn only one of the five sets of alarms: the IEC alarms, the ‘word rhythm’ alarms, the ‘auditory icons’ alarms, the ‘auditory icons plus ident’ alarms or the ‘resilient’ alarms. The alarms were as indicated in Appendices 1(a) to 1(d). They were normalized to sound at a loudness level of approximately 75-80dB(A) at the ear, measured with a Koolertron Sound Level Meter.

Participants

One hundred and ninety-four participants took part in the study. The participants were recruited either from Plymouth University's paid public pool, which has a large age range, or were current Psychology undergraduate students at the University of Plymouth, UK, which typically has a smaller age range. Approximately half of the participants were aged under 21, a quarter under 25 and the rest above 25, spread evenly over the age range to 72 years of age. This pattern was retained for the individual conditions tested, so that the age profile for each of the alarm sets was approximately the same. Forty-four participants took part in the IEC condition, 24 in the word rhythm, 45 in the auditory icon, 44 in the auditory icon plus beacon, and 38 in the resilient condition. Table 1 summarizes the number of participants in each age group for each condition.

	21 or under	22-25	26-35	36-45	46-55	56-65	Over 65
Study 1							
IEC	21	10	3	4	2	1	1
WR	13	6	2	2	0	1	0
AI	24	10	4	2	3	1	1
AI + I	20	12	6	2	2	2	0
Res	17	11	5	3	1	1	0
Study 2							
IEC	10	7	2	2	1	1	0
WR	12	6	3	2	1	1	1
AI	16	8	2	2	1	1	0
AI + I	10	6	1	2	1	1	0
Res	13	5	2	2	0	1	0

Table 1: Distribution of participants' ages in each of the conditions in each of the two studies

This research complied with the British Psychological Society's Code of Ethics and was approved by the Faculty of Health and Human Sciences Research Ethics Committee at Plymouth University, UK. Informed consent was obtained from each participant.

Apparatus

Each participant was tested by being seated at a Desktop Viglen DQ67SW computer with a Realtek High Definition Audio 24-bit, 48000 Hz (Studio Quality) sound card. Responses were made on a Philips 221PLPY monitor, and participants listened to the sounds through Behringer HPM 1000 headphones. The sounds were played as .wav files and were presented at a fixed loudness level (75-80db(A) at the ear) for all participants. The computer selected the condition for each of the participants at random, so that the experimenter did not know which condition each participant had been given to perform. As the selection of condition was random, fewer sets of responses were obtained for some conditions than for others.

Procedure

Each of the participants was required to sit on a chair facing a computer, in their own cubicle area facing a computer screen. They were given a set of headphones to wear before being asked to follow the on-screen instructions when they were comfortable and prepared to start the study. Once they had agreed to participate and had ticked the relevant on-screen responses, they were randomly assigned one of five alarm sets via the computer program and began the study.

Once the participant was ready to start the study, he/she was presented with each of the ten alarms within their assigned set one by one through the headphones, with the name/ function of each one presented simultaneously on the screen. They were asked to try to remember the sound and its function. Once this was complete, the experiment proper began. In each trial one of the ten alarm sounds was presented to the participant via the headphones. They were asked to click on the appropriate name of the sound (of ten) presented on-screen that they believed to be the correct alarm. If they were correct they moved on to the next trial, where a different alarm was presented and the participant was again asked to select the name of the sound. If they were incorrect, they were informed and presented with the sound and the screen for a second time and asked to respond again. If correct at this second turn, they then moved on to the next trial but if incorrect again they were presented with the sound for a third and final time. If incorrect for a third time they moved on to the next trial and the final response was recorded as incorrect.

There were ten blocks of trials in which each alarm was heard once, resulting in a total of 100 alarms in each test. The order of the ten alarms in each of the blocks was randomized, as also was the layout of the alarm names on the screen. This was done so that the participant would neither be able to predict which alarm would sound next, or learn to associate specific positions on the screen with the specific alarms.

Once the test was over, participants were given a debrief to read and were free to leave the study. The procedure took from 30 to 60 minutes depending on condition, as participants responded more quickly and accurately in some conditions than in others.

Results

Each participant heard each of the ten alarms alarm ten times, and had up to three attempts at each sound before moving on to the next sound. The data presented here are the correct/incorrect first time responses only. Initial analyses included responses to all ten sounds (including the Medium and Low priority alarms (MP and LP)). The results showed that performance was very good for these two sounds (almost at 100% throughout, for all five sets of alarms) and so these were removed from the following analyses for the purposes of clarity. The MP and LP sounds were very similar (and in some cases identical) across the five sets, in any case. These alarms also act as a calibration of the responses across the alarm sets, which were heard by different participants.

An alarm set condition (1-5) x sound (1-8) x block (1-10) mixed analysis of variance was conducted on participants' accuracy of responding to sound sets with condition as a between-subjects factor and sound and learning block as within- subjects factors. The results of this analysis of variance are shown in Table 2. This shows that many of the effects are highly significant. The effect of sound set is highly significant, meaning that some sets of alarms were easier to recognize than others. The effect of block was also significant, meaning that participants got better at recognizing the sounds the more they heard them, as one would expect. The significant effect for 'sound' means that some sounds were easier to recognize than others. There were also a number of interactions.

	Effect	Df	F	Probability	ηp^2	Effect size
1	Sound set condition	4,189	77.60	p<.001	.622	Large
2	Block	9, 1701	45.07	p<.001	.193	Med
3	Sound	7,1323	8.96	p<.001	.045	Small
4	Sound set condition x	36,189	1.72	p=.005	.035	Small
5	Sound set condition x	28,189	12.83	p<.001	.214	Med
6	Block x Sound	63,11907	1.57	p=.003	.008	Small
7	Sound set condition x	252,11907	1.29	p=.001	.027	Small

Table 2: Summary of effects found in the mixed 5 (sound set condition) x 8 (sound) x 10 (block) ANOVA conducted to examine participants' accuracy in the learning task in Study 1

When many effects are highly significant, as they are in this study, and as is demonstrated by the probability values in Table 2, it is best to examine how large these effect sizes are relative to one another. This is estimated using partial eta squared (ηp^2) which is essentially an estimate of the amount of the variation in accuracy scores accounted for by each effect. Effect sizes using partial eta squared in SPSS are deemed large if they are .41 or larger; medium if .18 or larger; and small if between .08 and less than .18 (c.f. Cohen, 1992, and Fritz, Morris & Richler, 2012, for a discussion of these values; we have used the more conservative estimates adopted by Fritz et al here).

Once the size of the effects is considered three key messages emerge from this analysis. The first is that the sound set individuals heard has a very large effect on the accuracy with which participants are able to recognize the sounds (effect 1; see Figure 1 and Table 2). Secondly, and perhaps unsurprisingly, individuals steadily improve across blocks of trials (effect 2; see Figure 1 and Table 2). Thirdly, there are variations in the efficacy of the sounds within sound sets (effect 5; see Figure 2 and Table 2). These are discussed in greater detail below.

Figure 1 shows the percentage correct in each sound condition across the 10 blocks of trials (condition x block interaction; Effect 4, Table 2). It shows that accuracy and overall recognition of the alarm set is poorest for the existing IEC alarms, and best for the two auditory icon conditions, with the word rhythm and the resilient sounds in between.

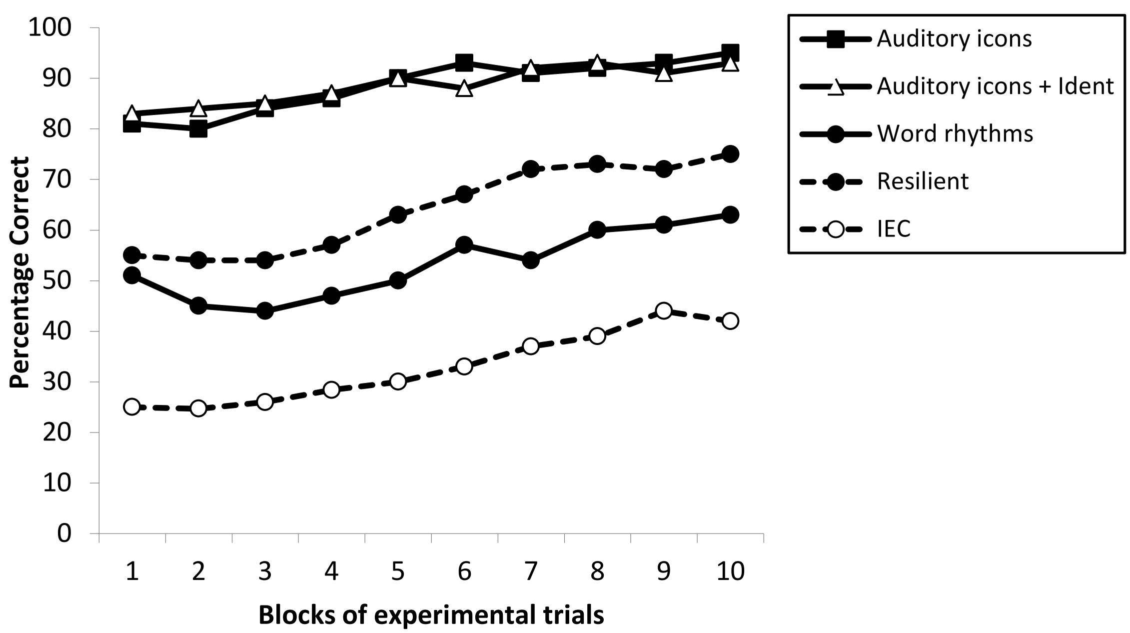


Figure 1: Percentage correct in in each alarm sound condition across 10 blocks of experimental trials in the recognition task in Study 1

Student Newman-Keuls comparisons were carried out to examine the extent to which differences in accuracy between each individual alarm set reached significance (see Howell, 2009 for a discussion of post-hoc comparisons). The comparisons revealed that, while the auditory icon and auditory icon + ident conditions were recognized equally well ($p>.05$), all other experimental conditions differed significantly from one another ($ps<.05$). Performance is highest in the auditory icon conditions and there is very little difference as to whether or not idents are included in the sound. In these conditions, participants can attribute alarm meaning using the metaphors conveyed by the real world sounds used and have high levels of performance from the outset (approximately 80% in block 1) and reach asymptote after approximately 5 trials. Performance in the ‘resilient’ condition is significantly worse than the auditory icon conditions, but significantly better than the ‘word rhythm’ condition. Performance in the ‘word rhythm’ condition is in turn significantly better than in the IEC condition, which is significantly worse than all other conditions.

Figure 1 also demonstrates one other medium-sized effect, which is the effect of block. The figure clearly shows that performance improves as participants hear the alarms repeated; for all conditions, they improve from block 1 to block 10 in a fairly systematic way. The results table (Table 1) shows a small sound set x block interaction, which means that the rate at which participants recognized the sounds varied somewhat across the sound conditions (as can be seen by the relative slopes of the lines in Figure 1, where the word rhythm and resilient sounds in particula diverge with increasing block number). However, Figure 1 also makes it clear that after a single hearing of the sounds, participants were able to name more or fewer at the starting point (block 1) dependent on the sound set condition in which they participated.

Figure 2 illustrates the efficacy of each sound in each of the sound sets (i.e. the sound x sound set condition interaction; effect 5, Table 2, a medium-sized effect). It shows that while the general trends between sound conditions remain, there are clear differences in the efficacy of individual sounds. Some of the most obvious effects are that the ‘general’ alarm sound is relatively well understood across conditions, including the IEC condition. Second, the ‘power down’ sound is particularly effective in the resilient condition and is particularly poor in the word rhythm condition (perhaps because many of the sound functions have 3 syllables). In the word rhythm condition the ‘drug administration’ sound is particularly effective. And finally, both the auditory icon and auditory icon + ident conditions show very similar variations in efficacy across different sounds, with ‘oxygen’ being the least effective.

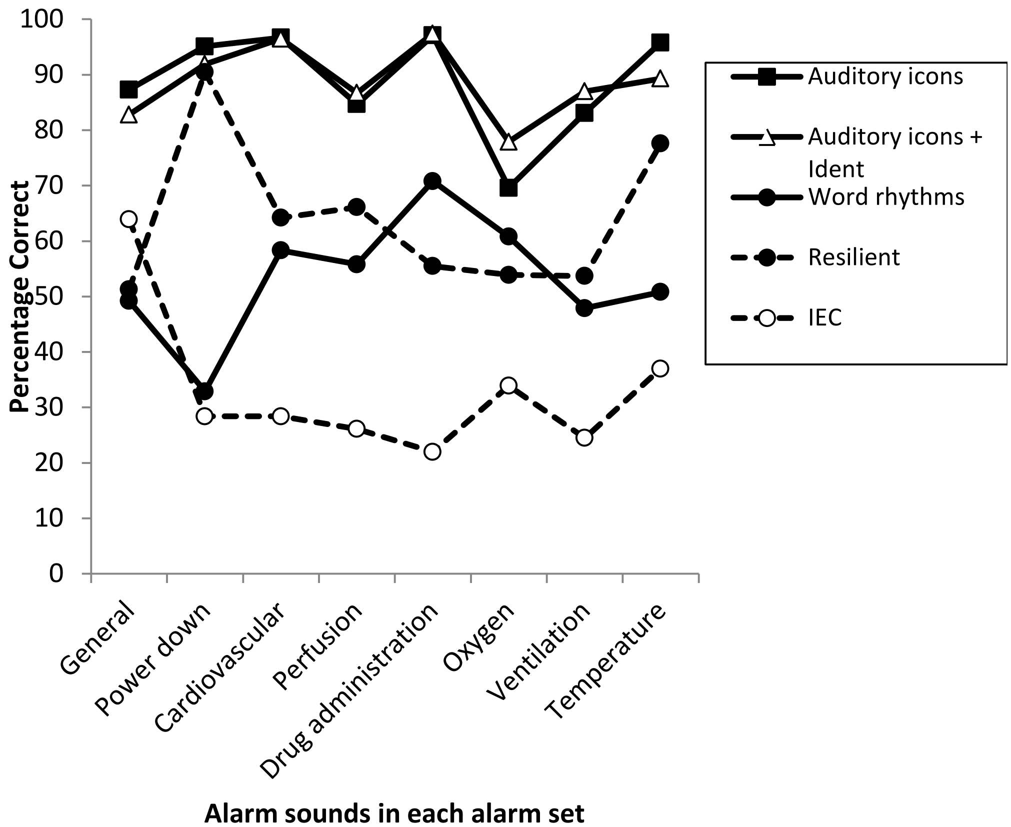


Figure 2: Percentage correct for each sound in each alarm sound condition in the learning task in Study 1

The results for the word rhythm and the resilient alarms show more variation than the other sets, whereby specific sounds in each of those two groups were recognized better than others. In the case of the word rhythm alarms, the relatively poor level of performance for ‘power down’ might be attributed to the fact that it is a three-pulse unit, and may be confused with other three-pulse alarms in this set, which are ‘perfusion’ and ‘oxygen’. For the resilient alarms, both the ‘power down’ and ‘temperature’ sounds appeared to be easier to learn than the others, suggesting that a simple pitch sweep may be more effective than differentiating between sounds on the basis of numbers of pulses.

Perhaps the most striking feature of the sound x sound set interaction is however that none of the sounds in the other three sets outperforms any of the auditory icon or auditory icon + ident alarms. At the other extreme, performance for each of the IEC sounds (other than the general alarm sound) are below performance for all of the sounds in all of the other sets.

Medium and low priority sounds

The ease with which medium and low priority sounds were learned was considered in separate analyses because their characteristics differed relatively little between conditions. Responses to the medium priority alarm were at ceiling throughout, with an overall accuracy of 95%, varying between 93-97% correct across trial blocks 1-10 and between 91-97% correct between conditions. As a result, no further analyses were conducted on medium priority recognizability data.

There was more variation in accuracy with the low priority alarm (overall mean=82.00% SD=24%). This data was therefore subjected to a mixed ANOVA with sound condition (1-5) as a between-subjects factor and blocks of trials (1-10) as a within-subjects factor. Given the lack of variation in the nature of this alarm between conditions (see Appendix 1a-d), it is not surprising that the effect of condition on accuracy was not significant, $F(4,190)=1.28$, $p=.280$, $\eta p^2 =.026$.Participants did, however, improved in their ability to recognize this alarm across blocks of trials, $F(9,190)=13.56$, $p <.001$, $\eta p =.067$. Table 3 shows how accuracy increased across trials and planned repeated contrasts revealed that mean accuracy differed significantly only between blocks 3 and 4, $F(1,190)=11.69$, $p=.001$, ηp $=.058$.

Block	Mean	SD
1	.70	.46
2	.69	.46
3	.73	.45
4	.84	.37
5	.85	.36
6	.85	.36
7	.88	.44
8	.86	.35
9	.91	.28
10	.91	.28
Total	.82	.24

Table 3: Mean and standard deviations of participants' proportion correct when responding to the low priority alarm in the learning task in Study 1

STUDY 2: LOCALIZATION

There is value in understanding the degree to which alarms and alarm sets intended for clinical use can be localized. Often, for example in a multibed ICU, it would be useful to be able to use localization cues to quickly and directly identify the relevant bed rather than having to go through some other more indirect route. For example, we detect the direction from which the transporter in the airport with the broadband noise alarm is coming simply by using the auditory cues coming from the vehicle itself, rather than locating it through some secondary mechanism such as an announcement giving a direction, or a visual display. Thus the relative localizability of the alarm sounds is presented as a feature of the benchmarking of the alarms tested in this paper. The localizability issue is also linked with masking, another key psychoacoustic issue.

Many audible alarms consist of tones with only a single or very few harmonics. These are typically poor auditory alarms, being both hard to localize and offering only weak resistance to masking (Hasanain et al, 2017). If a noise or other sound with the same, or close to, the same frequencies of the alarm signals at the same time and is louder than all or some of the harmonics in the alarm, the alarm will be masked and go unheard. There is thus a ‘safety in numbers’ principle when considering the harmonic density of an alarm sound. This is recognized in IEC 60601- 1-8 by specifying that the alarms should possess at least four components within the range 300-4000Hz and a lower fundamental, giving the sound at least four, and potentially more, harmonics.

A small number of studies have been carried out on the relative localizability of alarm sounds for use in vehicles. Catchpole et al (2004) carried out a set of studies where tonal and broadband noise alarms were arranged in various combinations. They demonstrated that the localizability of white noise performed best, a pure tone performed worst, and noise added to a tone could improve its localizability. Alali (2011) made direct comparisons between tonal and broadband alarms and demonstrated that a tonal alarm had a greater distance range than the broadband alarm (which may or may not be an advantage). Vaillancourt et al (2013) took a number of both objective and subjective measurements from three types of backup (reversing) alarm which varied in their degree of harmonic density – from a tonal alarm with few harmonics, to a multitone alarm with a greater number of harmonics, to a broadband noise alarm with many harmonics. They measured the responses in different listening conditions: without Hearing Protection Devices (HPD); with headphones; and with earplugs. Aside from one or two anomalies the localization data revealed more confusions (front/back and right/left) with the tonal alarms than the multitone alarms, and fewer with the broadband alarms in comparison with the other two types. Thus the available literature on localizability of alarms demonstrates broadly, as theory would suggest, that the greater the harmonic content or denseness of the sounds, the easier they are to localize.

In the second study, we investigate the localizability of the five sets of alarms using a simple paradigm. Broadly speaking, the auditory icons, auditory icons + ident and word rhythm sounds can be thought of as harmonically complex or dense, and the IEC and the resilient alarms can be thought of as simple and harmonically sparse, possessing typically no more than six harmonics (though even these are not as sparse as many alarms currently in use). We would expect this to affect the localizability of the alarm sets whereby the former three sets should be easier to localize than the latter two.

Method

Materials

The same five sets of alarms used for Study 1 were used in this study (Appendices 1(a) to 1(d)). We removed the two non-high priority alarms (medium and low priority) from each set as the logic of the speaker set-up was better suited to using eight rather than ten alarms, and also the localizability of medium and low priority alarms is of less interest than the eight high-priority alarms (quick and accurate localizability is logically not important for alarms other than those which are designated high priority).

Participants

A total of 124 participants took part in this study. As the participants were recruited in the same way as Study 1, the age spread of the participants was approximately the same as in this study, with approximately half of the participants being under 21 years of age, approximately a quarter in the age range 21-25 and the rest older, up to a maximum age of 70 years. Twenty-seven participants took part in the 'Word rhythm' condition, 23 took part in the 'IEC' condition, 30 took part in the 'Auditory icon' condition, 21 took part in the 'Auditory icon + ident' condition and 23 took part in the 'resilient' condition (Table 1). Al l participants were asked if they possessed any known hearing loss and none reported any.

Apparatus

Eight EasyACC model LX-839 speakers measuring 14.5 x 7.4 x 8.6 cm were each mounted on Amazon Basics 60 inch lightweight camera tripods. These were set around an empty room at 45 degree intervals as shown in Figure 3, with the distances and heights as shown in the figure. A chair for the participant was placed in the middle of the speakers. A customized program was written to run on a Windows tablet. On the tablet, participants saw a reproduction of the layout of the speakers as eight circles equally spaced in a larger circle on the screen. During the study, alarm sounds were heard from each of the speakers on multiple occasions and on each presentation the participant was required to indicate which speaker had sounded by selecting the relevant circle. The speakers were labelled 1 to 8 with speaker 1 being directly in front of the participant and speaker 5 directly behind the participant. A block consisted of 64 trials, with each alarm being played from each speaker once in a block.

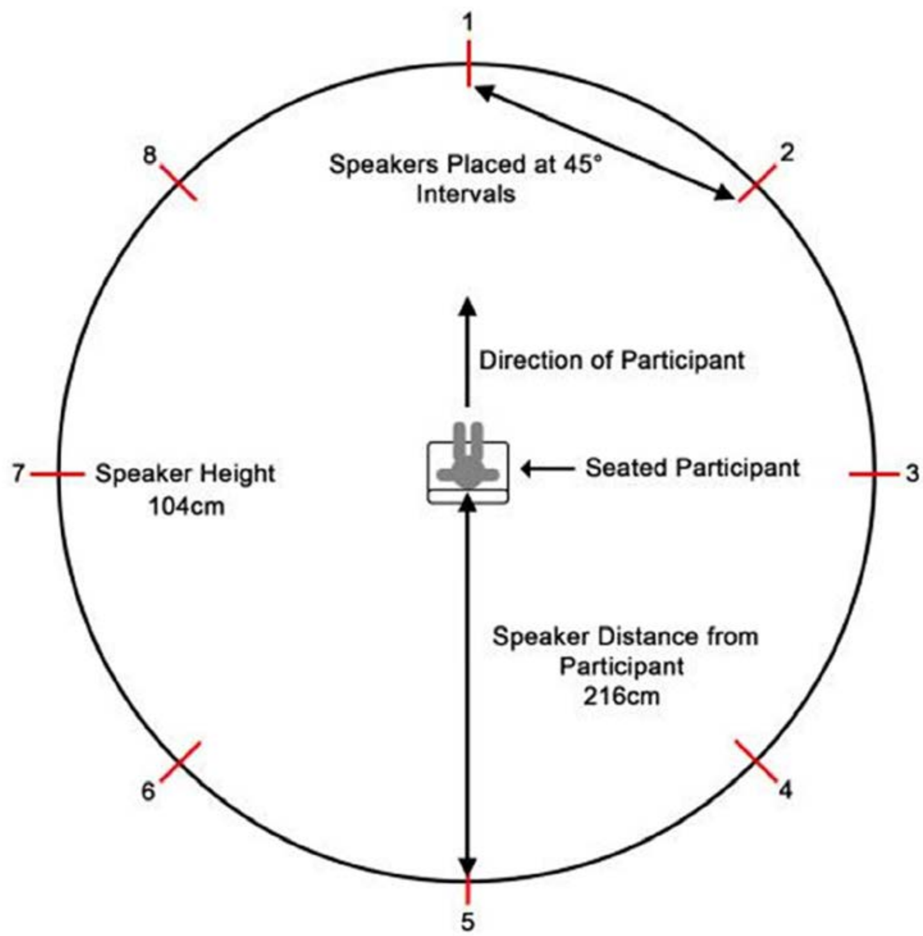


Figure 3: Laboratory configuration for the localisation task in Study 2

Procedure

Participants were seated in the middle of the room, surrounded by the speaker set- up.

They were briefed that they would hear a series of sounds coming from the set of speakers, and their task was to indicate which speaker location the sound had come from by touching and responding on the relevant circle on the screen. They were told nothing about the sounds other than that they were possible prototypes for future medical alarms. They were allowed to move and turn their head but were not allowed to stand up or move about the room. This procedure was adopted because we wanted the paradigm in some way to reflect naturally-occurring behavior (one would turn one's head to locate a sound) but not to accentuate the usefulness of loudness variation in localizing a sound by being able to move about in space (i.e. walking). Turning one's head would have some effect on the relative loudness, but little by comparison to being able to move about. After eight practice trials, where each of the eight sounds was heard once, from a different speaker (so all eight speakers were tested in the practice trial) the experiment proper began. Participants heard 64 sounds in a single block, consisting of each sound from each speaker played once. Each participant took part in a total of three blocks of 64 trials, resulting in 192 trials in total. Participants were permitted only one response, and responses were timed out eight seconds after the start of the sound. There was an interval of two seconds between each trial, and a pause of one minute at the end of each block. After the experiment was completed, participants were thanked and debriefed. The whole procedure took from 30 to 45 minutes depending on how quickly participants responded to the alarms, which varied from condition to condition.

Results

We present both reaction time and accuracy data in this section, as both measures are of ecological relevance to the issue of localizing and responding to alarms. The variables of interest were the sound set, the block (blocks 1, 2 or 3), and the individual sounds within the sets.

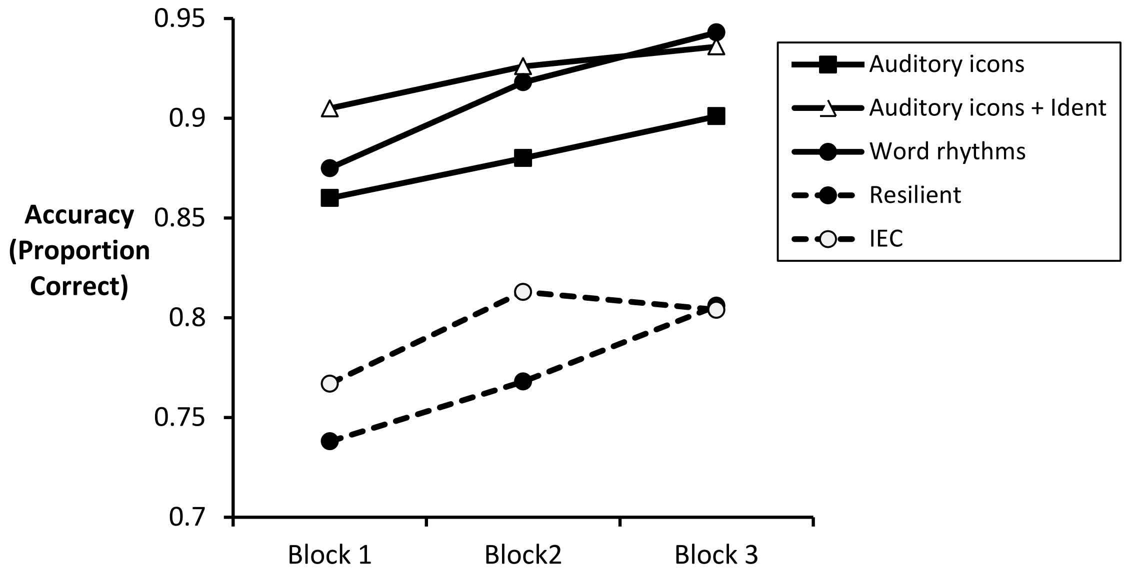
Localisation Accuracy

Participants were scored 1 for every correct response and 0 for every incorrect response, which was then converted to a percentage correct score. A sound set condition (1-5) x sound (1-8) x speaker (1-8) x block (1-3) mixed analysis of variance was conducted on the percentage correct data. Table 4 summarizes the main effects, interactions, and effect sizes for this data. Three- and 4-way interactions are not included in the table because they were either not significant or only marginally so. Effect sizes for these interactions were also very small, as ceiling effects were observed particularly with respect to speaker positions 1, 2, 7 and 8 (see Figures 4 and 5). Table 3 shows that the main effects of sound set, speaker, and block were notable in terms of effect size, though all of these effects were small (Cohen, 1992; Fritz et al, 2012).

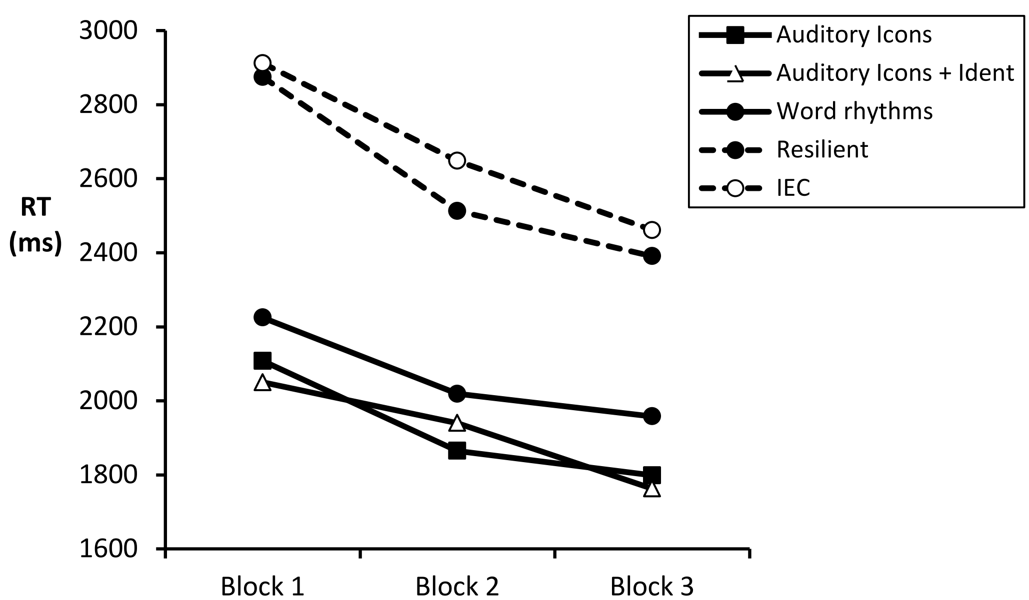
	Effect	Df	F		2	
1	Sound set condition	4,120	6.32	p<.001	.174	Small
2	Block	2,240	15.27	p<.001	.113	Small
3	Sound	7, 840	6.48	p<.001	.051	-
4	Speaker	7, 840	20.81	p<.001	.148	Small
5	Block x Sound	14, 1680	1.79	p<.035	.015	-
6	Block x Sound set	8, 240	0.82	p=.589	.026	-
7	Block x Speaker	14, 1680	4.34	p<.001	.035	-
8	Sound x Speaker	49, 5831	2.02	p<.001	.017	-
9	Sound x Sound set	28, 840	3.63	p<.001	.108	-
1	Speaker x Sound set	28,840	1.48	p=.053	.047	-

Table 4: Summary of primary effects found in the mixed 5(sound set condition) x 8(sound) x 8(speaker) x 3(block) ANOVA conducted to examine participants' percentage accuracy in the localization task in Study 2

A significant effect was found for the alarm set presented (Figure 4(a); effect 1 in Table 3). Here, responses were more accurate for auditory icons, auditory icons + ident and word rhythms than they were for the current IEC alarms and the resilient set (Figure 4(a))



(a) Localization accuracy

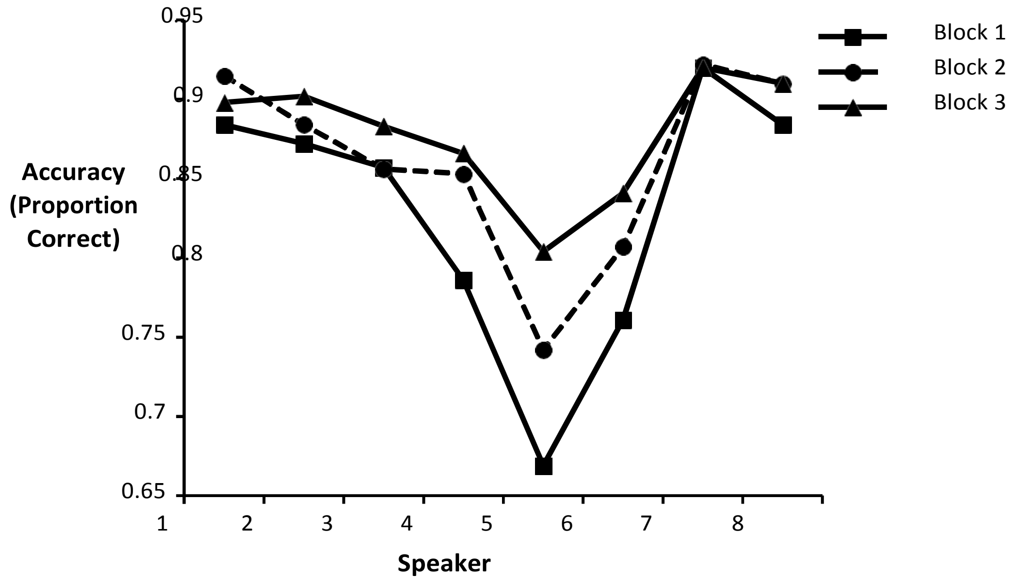


(b) Localization speed

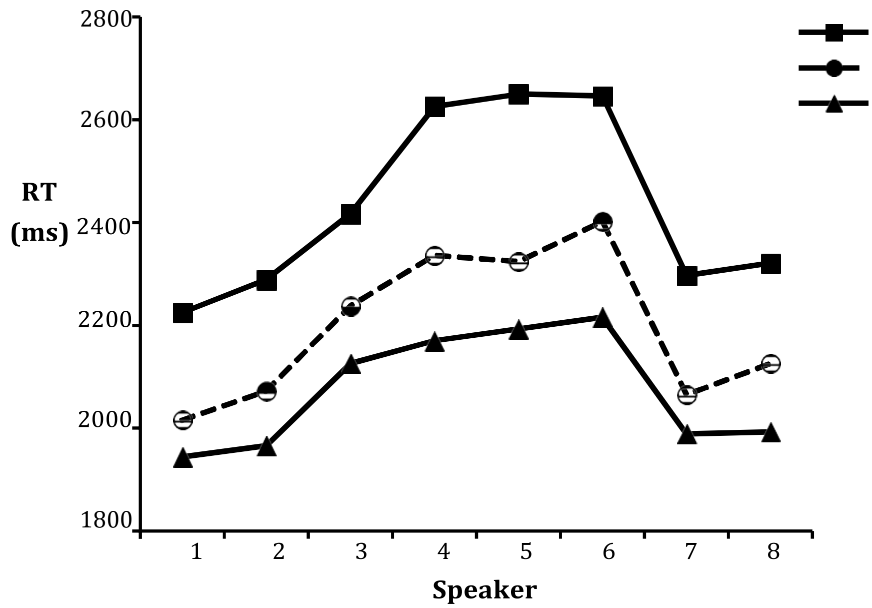
Figure 4: Accuracy (proportion correct) and speed (ms) for each sound set in blocks of experimental trials in the localisation task in Study 2

Student-Newman-Keuls post-hoc comparisons showed that localization accuracy was at a similarly high level for responses to auditory icons, auditory icons + ident and word rhythms, and that the accuracy of responses to the IEC sounds and resilient sounds were similar to one another ($p>.05$) but significantly lower than for the other sound types ($p<.05$). The ANOVA also revealed a significant man effect for the position of the speaker (Figure 5(a); effect 4 in Table 4). Here, accuracy was highest for sounds from speaker positions 1, 2, 7 and 8 and poorest for sounds from positions 4, 5, and 6. Thus performance was better when the sounds were in front of or to the side of the participant, rather than behind them (see Figure 3). Planned repeated contrasts were carried out to examine the differences in accuracy observed between speakers in more detail and these are summarized on the right-hand side of Table 6. As might be expected from Figure 5(a), these reveal that localization accuracy differs between speakers 4 and 5, 5 and 6, and 6 and 7.

Finally, a significant effect was found for block (Table 4, effect 2, Figures 4(a) and 5(a)). Student Newman-Keuls post-hoc comparisons showed that performance was significantly less accurate in block 1 than block 2, and significantly more accurate in block 3 than in either block 1 or block 2 ($p < 0.05$).



(a) Localization accuracy



(b) Localization speed

Figure 5: Accuracy (proportion correct) and speed (ms) for each of 8 speakers in blocks of experimental trials in the localization task in Study 2

Localization Speed

Where responses were correct, the speed of response was also noted and analyzed. A sound set condition (1-5) x sound (1-8) x speaker (1-8) x block (1-3) mixed analysis of variance of speed of was carried out and the results summarized in Table 4. Note again that 3- and 4-way interactions are not included in Table 5. Generally, these were not significant, or were only marginally so, and effect sizes were very small. Once again, the effect sizes suggest that the three main effects of sound set, block, and speaker dominate the findings, with some effect for the interaction between sound and sound set.

Reflecting response accuracy, participants responded more quickly to the word rhythm, auditory icon and auditory icon + ident conditions than to the resilient and the IEC alarms. Figure 4(b) shows that participants respond more quickly to the auditory icons, auditory icons + ident, and to the word rhythms but more slowly to the IEC sounds and resilient sounds. Student-Newman-Keuls post-hoc comparisons revealed that localization speed was at a similarly high level for those presented with auditory icons, auditory icons + ident and word rhythm sounds and that the performance of those presented with the IEC sounds and resilient sounds was significantly slower ($p<.05$). Comparison with Figure 4(a) suggests that slower responses were associated with less accurate responses, so reflect a general performance decrement with the latter two sets rather than a speed-accuracy trade-off.

Effect		Df	F	Probability	ηp^2	Effect size
1	Sound set condition	4, 119	7.47	p<.001	.201	Med
2	Block	2, 238	65.43	p<.001	.355	Med
3	Sound	7, 833	3.45	p=.001	.029	-
4	Speaker	7, 833	40.37	p<.001	.253	Med
5	Block x Sound	14, 1666	2.92	p<.001	.024	-
6	Block x Sound set condition	8, 238	1.47	p=.169	.047	-
7	Block x Speaker	14, 1666	2.62	p=.001	.022	-
8	Sound x Speaker	49, 5831	1.91	p<.001	.016	-
9	Sound x Sound set	28,833	6.69	p<.001	.184	Small
1	Speaker x Sound set	28,833	0.98	p=.502	.032	-

Table 5: Summary of primary effects found in the mixed 5(sound set condition) x 8(sound) x 8(speaker) x 3(block) ANOVA conducted to examine participants' speed of responding (in ms) in the localisation task in Study 2

Table 5 also indicates a medium-sized effect for position of speaker (see also Figure 5 and Table 5). Figure 5(b) shows that participants responded fastest to speakers in positions 1, 2, 7 and 8 and relatively poorly to those in positions 4, 5 and 6. Table 6 summarizes the results of the planned repeated contrasts carried out to examine the differences between speakers in more detail. As might be expected from Figure 5, these reveal that localization speed differs between speakers 2 and 3, 3 and 4, and 6 and 7.

	Localization speed			Localization accuracy		
Speaker comparison	**F**	**p**	η_p^2	**F**	**p**	η_p^2
1 vs 2	2.37	.126	.020	.571	.452	.005
2 vs 3	**27.73**	**.000**	**.189**	2.22	.139	.018
3 vs 4	**17.39**	**.000**	**.127**	3.73	.056	.030
4 vs 5	.191	.663	.002	**27.80**	**.000**	**.188**
5 vs 6	1.09	.299	.009	**9.95**	**.002**	**.077**
6 vs 7	**123.30**	**.000**	**.509**	**43.60**	**.000**	**.266**
7 vs 8	1.19	.278	.010	2.65	.106	.022

Table 6: Summary of planned comparisons examining differences in localization speed and accuracy between speakers in Study 2

For the accuracy data, performance significantly declined as speaker position progressed from front to back, and then improved as the position progressed from back to front. For reaction time, performance declined significantly earlier on in the speaker sequence, from 2 to 3 and again from 3 to 4. For both accuracy and speed, performance again improves as the speakers progress back around the front from positions 6 to 7. As for the alarm set main effect, both speed and accuracy degrade together.

Finally, there is a middle-sized significant effect for block (Figures 4(b) and 5(b)). As for the accuracy data, performance was faster in block 2 than in block 1, and faster in block 3 than in either block 2 or block 1 ($p < 0.05$).

General Discussion

The data presented in this paper provide initial benchmarking data for four sets of alarms that serve as prototypes for replacements for those currently supporting the international standard IEC 60601-1-8. The prototype alarm sets have been compared with the current IEC alarms and for all comparisons except one – that between the localizability of the 'resilient' set and the current IEC set – have been shown to outperform the current alarms.

In terms of recognizability, all of the four prototype sets are more readily recognized than the current IEC set. However, the data show that some of the sets are more readily recognized than others. The auditory icon sets are more readily recognized than the resilient set, and these are all more readily recognized than the word rhythm sets. Although the level of metaphor or sound-referent link is difficult to conceptualise across these sets in any systematic way (for example, how does one compare a word metaphor – a mnemonic – with a sound that might evoke a visual image?), taken as a whole, what this finding suggests is that the more direct the nature of the metaphor used to represent the function, the easier it is to intuit that association - OR - that some metaphors are better than others (whatever 'better' might mean). This finding is in line with previous research (Keller & Stevens, 2007; Petocz et al, 2008). Alternatively, or in addition, it might be that the ease with which an alarm set can be learned is also governed by the variability within the set (Edworthy et al, 2011).

Although variability was not systematically controlled in these studies, it is likely that the word rhythm and resilient sets were less varied than the auditory icon sets, and the IEC set is certainly the least varied. The effects of these factors, as well as the nature of how the relationship between sound and meaning change with exposure, are in the process of being investigated further in this laboratory.

The practical outcome of the work presented here is that some types of alarm sounds require next to no exposure in order to be recognized. Of course, the alarm sets represent only one example of each design remit, and there could be other manifestations of the same remits which lead to more, or less, recognizable alarms in comparison. For example, the 'word rhythm' alarms would probably be more difficult for a non-native user of English, and certainly a lot more difficult for people with no facility for the English language. Also, there may be other auditory icons that work less well, or may work better. This topic would benefit from further investigation, though there is some research to guide this topic (Keller and Stevens, 2004).

Our first study dealt with recognizability rather than with learning explicitly, as participants were presented with the name of each of the sounds only once, at the beginning of the study. Our study suggests that though they were not presented with the name again, performance did improve with each successive block and so some learning did appear to take place for all sets of alarms. Currently in our lab we are looking at how repeated presentation of sounds improves when participants are made aware of the correct answer when their response is incorrect.

The data also reinforce the known effects of harmonic content on localizability. The three sets with the larger number of harmonics – the two auditory icon sets and the word rhythm set – produced greater accuracy and speed of localization than those with fewer harmonics. The IEC alarms as specified in the standard are required to possess a minimum of five harmonics (and in practice usually do), and the resilient alarms were deliberately designed with fewer harmonics than the other prototype sets. This lower number of harmonics resulted in the reduction of the localizability of those two sets in comparison to the others. These findings are also in line with previous studies, which demonstrate that, all other things being equal, sounds with more harmonics are easier to localize than those with fewer. Another consequence of alarms with more harmonics is that they will be more resistant to masking, making it more difficult to miss alarms when they sound at the same time. In the clinical environment it is often the case that several alarms (either from the same piece of equipment, or from different pieces of equipment) tend to signal at the same time because they are indicating a problem which can often have more than one dimension (for example, an increase of heart rate along with an increase in temperature) and so there is often a risk that one alarm will mask, or will be masked by, another. Testing of the prototype sounds in scenarios where they may be masked is one line of enquiry that should be pursued in the next set of tests.

Related to the issue of masking is the issue as to how far the alarm sounds might travel in a typical clinical environment. This feature of the alarm sounds has yet to be tested, and it is important to know how any new alarm sounds will work in this respect as carers are often working at some physical distance from a patient needing care (they may, for example, be in a different room). However, as a general rule, lower frequencies travel further so any sound with many harmonics, particularly of lower frequency, should travel at least as well as harmonics of existing, typically high-frequency, alarm sounds.

One important caveat that it is important to note is that our participants were generally younger in age than a typical sample of clinicians (see Table 1). Because the young have minimal hearing loss in comparison with older adults, this is likely to have inflated the responses obtained here in comparison to that which a professional clinical population might produce. We might expect there to be some differences between our participant population and a more age-diverse clinical population as hearing in general declines with age, particularly with higher frequency sounds. A further attraction of using auditory icons is that their spectra is usually fairly wide, often with plenty of low-frequency energy (depending on the sound used), rendering them potentially more resistant to potential inaudibility as a function of age. The participants tested here were also not clinically trained, and it is clearly important to test these sounds with clinical populations during the next phases.

Another aspect of our studies which requires some discussion is that we used a randomized method of selecting participants in each of the conditions, rather than a counterbalanced method. We selected this method because we did not want experimenters to be aware of which of the conditions each participant was undertaking in case they gave clues to participants as to how easy or difficult the task would be. The consequence of this method is that different numbers of participants were tested in each of the five conditions. While this is not a problem (as we had sufficient participants in each of the conditions) a future refinement which combines a counterbalanced method with blinded experimenters would be a methodological improvement.

The benefit of having auditory alarms which are easier to recognize and localize seem fairly obvious in an environment which is typically high-workload, safety-critical, and both mentally and emotionally demanding. This paper presents the first step in benchmarking prototype replacement alarm sounds for IEC 60601-1- 8 and has already gone beyond the data currently available for the existing alarm sounds, which provide data only on learnability. Here we have provided data also on localizability.

There is no standardized or agreed method of evaluating auditory alarms, although parts of the process (for example, that of eliciting from respondents what sorts of sounds might make the 'best' alarm sounds; Edworthy & Stanton, 1995) have been suggested in the past. The next steps in developing and evaluating some or all of these alarm sets will be part of a program of work which will include testing in simulated environments where the audible alarms can be readily inserted into simulated medical scenarios (Bennett et al 2015; Bennett & McNeer, 2012; McNeer et al, 2016), and testing in an appropriate dual-task paradigm (Stevenson et al, 2013). The designs can also be incorporated into a model-checking approach to masking. which should ultimately be able to establish potential masking between the IEC alarms, other non-alarm sounds, and other medical device alarm sounds in a clinical environment (Hasanain et al, 2017).

A final issue which may need further refinement is the use of the eight categories specified in Kerr's (1985) original exposition of the categories used as the basis for the alarms, which is used as the basis of not only the current alarms but the prototypes put forward here. In the approach developed by Kerr, alarm sounds are associated with possible causes of tissue damage. This means that different alarm sounds are associated with the underlying physiological functions, rather than equipment. The implication of this therefore is that a 'cardiovascular' sound, for example, could be produced by more than one piece of equipment. It also means that some pieces of equipment could produce alarms from more than one category. How these multiple alarms for the same function might be presented and subsequently detected and recognized by the clinician is a topic on which there is very little research, and indeed the categories themselves have come under scrutiny (Edworthy et al, 2017). The work presented in this paper simply addresses the design of the alarm sounds themselves, and the principles that surround that design process. Having established that auditory icons are easier to recognize than other classes of alarm sound (and that there are differences between other types of potential alarm sound), the more practical issues as to how the alarms might function in a clinical environment, and new ways of thinking about how alarm situations might be classified, can be developed. For example, one of the key features of Kerr's thinking was that alarm sounds should be restricted to 8-10 maximum, because of the known learning problems with abstract alarms. If the ease of recognizing auditory icons extends to larger numbers of sounds (and more recent (and this) evidence suggests that it does), then it might not be necessary to restrict the number of alarms used so dramatically. However, proliferation should probably be avoided as the potential for masking is likely to be increased if there is the potential for a larger number of categories. At the point of writing we simply do not know whether there is an optimal number of alarms based on the desire for alarm information on the one hand, and the desire to avoid masking problems, on the other.

Key points:

- This work represents the first phase of benchmarking medical auditory alarms intended to replace those currently specified in the standard as the alarms currently supporting the standard are known to be difficult to learn and retain
- Four sets of audible alarms were designed to match the eight functions (plus medium and low priority) specified in a global medical device standard, IEC 60601-1-8
- The four prototype sets were designed to have acoustic variability across the set, and to use different ways of representing a link between alarms and their functions
- All prototype sets were easier to recognize than the current set, with auditory icons performing best
- All (aside from a simple set) were also easier to localize than the current set

References

Arrabito, G. R., Mondor, T. A., & Kent, K. J. (2004). Judging the urgency of non-verbal auditory alarms: A case study. Ergonomics, 47(8), 821-840.

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.

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.

Bennett, C. L., Dudaryk, R., Ayers, A. L., & McNeer, R. R. (2015). Simulating environmental and psychological acoustic factors of the operating room. The Journal of the Acoustical Society of America, 138(6), 3855-3863.

Blauert, J. (1997) Spatial hearing. Boston: MIT Press

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.

Cohen, J. (1992). A power primer. Psychological bulletin, 112(1), 155.

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 (2017) Ch. 23: Designing Auditory Alarms. In Information Design: Research and Practice (Ed Black, A, Luna, P, Lund, O & Walker, S), 377-390. London: Routledge.

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., Loxley, S., & Dennis, I. (1991). Improving auditory warning design: Relationship between warning sound parameters and perceived urgency. Human Factors: The Journal of the Human Factors and Ergonomics Society, 33(2), 205-231

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.

Edworthy, J., & Stanton, N. (1995). A user-centred approach to the design and evaluation of auditory warning signals: 1. Methodology. Ergonomics,38(11), 2262-2280.

Finley, G. A., & Cohen, A. J. (1991). Perceived urgency and the anaesthetist: responses to common operating room monitor alarms. Canadian Journal of Anaesthesia, 38(8), 958-964.

Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: current use, calculations, and interpretation. Journal of Experimental Psychology: General, 141(1), 2.

Guillaume, A., Pellieux, L., Chastres, V., & Drake, C. (2003). Judging the urgency of nonvocal auditory warning signals: perceptual and cognitive processes. Journal of experimental psychology: Applied, 9(3), 196.

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

Haas, E. C., & Casali, J. G. (1995). Perceived urgency of and response time to multi- tone and frequency-modulated warning signals in broadbandnoise. Ergonomics, 38(11), 2313-2326.

Haas, E. C., & Edworthy, J. (1996). Designing urgency into auditory warnings using pitch, speed and loudness. Computing & Control Engineering Journal,7(4), 193-198.

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.

Hellier, E. J., Edworthy, J., & Dennis, I. (1993). Improving auditory warning design: Quantifying and predicting the effects of different warning parameters on perceived urgency. Human Factors: The Journal of the Human Factors and Ergonomics Society, 35(4), 693-706.

Howell, D. C. (2012). Statistical methods for psychology. Cengage Learning.

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.

Kerr JH. Warning devices. Br J Anaesth 1985;57:696 –708

Kerr, J. H., & Hayes, B. (1983). An “alarming” situation in the intensive therapy unit. Intensive care medicine, 9(3), 103-104.

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.

Momtahan, K. L. (1991). Mapping of psychoacoustic parameters to the perceived urgency of auditory warning signals. National Library of Canada= Bibliothèque nationale du Canada.

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.

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.

Zwicker, E., & Fastl, H. (2013). Psychoacoustics: Facts and models (Vol. 22). Springer Science & Business Media.

Appendix 1: Characteristics of alarm sounds in each sound set

Appendix 1a: IEC 60601-1-8 High Priority Alarm Characteristics

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 pulse ranging between 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
Medium priority	The first three pulses of the general alarm played at a lower pitch and a slower speed (c – c – c)
Low priority	Hostess ‘ding-dong’: a two pulse unit: e--c

Appendix1b: Word Rhythm Alarm Characteristics

Function of Alarm	Alarm Characteristics
General	Updated version of the current IEC60601-1-8 general alarm, including the same temporal pattern. The basic pulse structure is a three-note major triad chord (G3-B3-D4) with the lowest F_O 196Hz (G3).
Power down	A three-pulse burst, made more harmonically complex by overlapping a musical fourth on to the original. C5-G4-Bb4 in an irregular rhythm
Cardiovascular	A burst containing six pulses, made more harmonically complex by overlapping a musical fifth on to the original. D4-E4-C4-G4-C4-C4 pattern in an irregular rhythm
Perfusion	A three-pulse burst, made harmonically more complex by overlapping a musical fourth on to the original. G4-C5-F4 pattern in an irregular rhythm
Drug Administration	A burst containing six pulses, made more harmonically complex by overlapping a musical fourth on to the original. D5-C5-C5-Bb4-D5-Bb4 pattern in an irregular rhythm
Oxygen	A burst containing three pulses. C5-G4-G4 in an irregular rhythm
Ventilation	A four-pulse burst with a fixed pitch of G3, with an irregular rhythm
Temperature	A four-pulse burst, made more harmonically complex by overlapping a musical fifth on to the original. G4-B4-D5-A4 with an irregular rhythm
Medium priority	A four-pulse burst constructed as a two-tone chord G3-B3). Regularly spaced with xylophone quality
Low priority	Single-tone version of Medium priority sound

Appendix 1c: Auditory Icon Alarm Characteristics

Function of Alarm	**Alarm Characteristics**
General	A faster version of the current 5-pulse IEC60601-1-8 general alarm. Repeated once, with a total of 10 pulses each lasting 70ms within the bursts. Pitch is A4
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 ‘heartbeat’ sound with no discernible frequency. Six pulses formed from 3 2-pulse units indicating 3 heartbeats
Perfusion	A ‘water bubbling’ sound, 2 pulses each approximating 2 seconds in length
Drug Administration	The sound of a continuously rattling ‘pillbox’
Oxygen	The sound of an aerosol, 4 pulses each spaced 600ms apart
Ventilation	The sound of a single deep breath out
Temperature	The sound of ‘frying on a stove top’
Medium priority	Simple 3-pulse tone at A3, 100ms pulse length with 50ms between pulses
Low priority	Single pulse version of Medium priority, longer

Appendix 1d: 'Resilient' Alarm Characteristics

Function of Alarm	**Alarm Characteristics**
General	A faster version of the current 5-pulse IEC60601-1-8 general alarm. Repeated once, with a total of 10 pulses each lasting 70ms within the bursts. Pitch is A4
Power down	A 2000ms tone starting at F4, falling to Bb3
Cardiovascular	A 6-pulse burst ('Car-di-o-vas-cu-lar) with a pitch change between pulses 3 and 4. Tone based on E4, pulses 200ms in length with 100ms gaps
Perfusion	6 x 2-pulse oscillating sound based on C5, each oscillation lasting approximately 350ms
Drug Administration	A 3-pulse burst. Each pulse has a pitch shift of 7 semitones starting on G4. 500ms pulse length with 350ms gaps
Oxygen	2 x 3-pulse burst (Ox-y-gen), tone based on C#5 with 0.3ms/0.2ms0.2 first/second/third syllables
Ventilation	2x a 4-pulse burst ('Ven-ti-la-tion'), tone based on F#5, with 0.2/0.2/0.3/0.2 first/second/third/fourth syllables
Temperature	A 2000ms tone starting at F4, rising to A4
Medium priority	Simple 3-pulse tone at A3, 100ms pulse length with 50ms between pulses
Low priority	Single pulse version of Medium priority, longer

Biographies

Judy Edworthy is professor of applied psychology and Director of the Cognition Institute at Plymouth University, UK. She holds a PhD from Warwick University, UK (1984). She has published many papers on the design, evaluation and behaviour surrounding audible alarms in many different contexts. She also researches visual warnings and the psychology of music.

Scott Reid is an undergraduate student at Plymouth University working on alarm design, learning, and localizability

Sine McDougall is professor of psychology at Bournemouth University, UK. Her PhD is from Goldsmith's College, London, UK (1993). Her research has focused on how we understand, learn, and use icons and signs used on computer interfaces and on traffic and public information signs. This includes how we attend to, and search for, traffic signs and the use of icons as cues for decision-making in complex environments.

Jonathan Edworthy has a BA in Music Technology form Plymouth University (2015). He works as a sound designer and recording technician

Stephanie Hall is an undergraduate student at Plymouth University working on alarm design, learning, and localizability

Danielle Bennett has a BSc in Psychology from Plymouth University (2016)

James Khan has a BSc in Psychology form Plymouth University (2016)

Ellen Pye has a BSc in Psychology from Plymouth University (2016)