AUDITORY ALARMS, MEDICAL STANDARDS, AND URGENCY

Penelope Sanderson, Alexandra Wee, Eunice Seah, and Philippe Lacherez

Key Centre for Human Factors and School of ITEE
The University of Queensland, St Lucia, Australia

psanderson@itee.uq.edu.au

ABSTRACT

A newly-released international standard for medical equipment alarms, IEC 60601-1-8, incorporates a long-standing suggestion that alarms should indicate their source through distinctive melodies. In this paper we examine this suggestion. We describe the proposed alarms, outline the history of the idea, and review recent research on the effectiveness of the alarms, some of it performed in our laboratory. Finally we discuss the concept of "urgency mapping" for alarms, noting where it may and may not be effective.

1. INTRODUCTION

The problems with auditory alarms have been noted in scores of articles, including some published at earlier ICAD conferences [1]. Our focus is the effectiveness of auditory alarms for medical environments, particular the operating room (OR) and intensive care unit (ICU).

In September 2005 an international standard was published for auditory alarms on medical equipment—IEC 60601-1-8. There are many good recommendations in the standard that will be useful to manufacturers for many years. A novel recommendation is that manufacturers implement melodic alarms whose source can be recognized from the melody played. Such alarms are an attempt to convey the meaning of situations to clinicians as soon as possible so they can direct their interventions effectively from the start rather than having to waste time seeking the source of an alarm.

In this paper we outline the history of the idea of melodic alarms, outline some recent research performed to evaluate the alarms, including some research from our laboratory, and we conclude with some observations on urgency mapping.

2. MELODIC ALARMS

Patterson [2] (in [3]) initially proposed melodic alarms for medical environments. The alarms would consist of a sequence of notes of different pitches in a distinctive rhythm and urgency would be indicated by playing the notes more rapidly. Kerr [3] built on Patterson's ideas and outlined different possible approaches to alarm design: single all-purpose alarms, priority-based alarms, equipment-based alarms, risk-based alarms, and risk-and-response based alarms. Selecting the latter, Kerr proposed alarms for hypoxia, ventilator problems, cardiovascular problems, interruption to perfusion, drug administration problems, and thermal risk, each in a low and high level of alarm state and distinguished by melodic changes.

The possibility of melodic alarms was subsequently discussed by ISO and CEN committees for possible inclusion in medical equipment alarms standards, focusing on the six categories of oxygenation, ventilation, cardiovascular status, temperature, artificial perfusion, and infusion (drug administration). However, the idea of melodic alarms was heavily criticized as potentially making a bad situation worse [4]. Despite this, Block [5] somewhat lightheartedly proposed alterative melodies for the six alarms, this time based on popular tunes whose titles could be readily with the six alarm categories (eg. "love is blue" for oxygenation). In an informal study with 79 anesthetists, Block found that after one exposure to the alarm sounds with their labels, anesthetists performed better than chance on their first attempt to identify the alarms in a surprise test. There was a significant improvement on the second attempt, with 53% of anesthetists getting all six correct. Block pointed out the general benefit of associating words with melodies for easier recognition and recall—a suggestion later adopted for the IEC 60601-1-8 alarms. Block noted the need for further research into the idea, involving clinicians, HF engineers, psychoacousticians, and musicians. Unfortunately such an effort did not occur in the way Block envisaged.

The ISO 9703-2 standard for auditory alarms signals on anesthesia and respiratory care equipment was released in 1994. The standard did not recommend melodic alarms, but it recommended rhythms for alarms at different levels of urgency. Notably, a medium-priority alarm used three notes ("3"). High-priority alarms used three notes followed by two notes after a gap and then repeated after a longer gap ("3-2, 3-2"). The high-priority alarm was played at a faster rate than the medium priority alarm.

Some years later in the context of their work on the IEC 60601-1-8 standards committee for medical equipment alarms, Block, et al. [6] proposed a new set of melodic alarm sounds that satisfied the rhythmic "3-2, 3-2" and "3" constraints imposed by the earlier ISO 9703-2 standard. Rather than creating associations between melodies and medical equipment alarms through associations with titles of songs, instead they created associations through the functioning of the medical equipment or physiological process itself, which would be well-known to healthcare workers (see Table 1). For example, an oxygen alarm would sound when oxygen saturation in the blood falls from 100% (its maximum). Accordingly, the melodic theme is a decreasing series of notes [**OXYGEN-HI-P.WAV**]. All alarms were in the key of C major and had the same timbre.

The melodies outlined in Table 1 are a compromise between the rhythmic constraints of the prior ISO 9703-2 standard and the need to find sufficiently discriminable melodies with contours that could still be clearly associated with the functioning of the medical equipment or physiological process being signaled. Surprisingly, no systematic testing with human participants informed the final design of the IEC 60601-1-8 melodic alarms. It is only in the few months immediately prior to the release of the standard, or afterwards, that evaluations have appeared, all too late to influence the standard. In the next section we outline the findings of these evaluations.

Alarm	Melody*† and mnemonic		Other information in support of mnemonic
	Medium priority	High priority	
General (GE)	C4-C4-C4	C4-C4-C4—C4-C4 (repeated)	Fixed pitch, traditional (usual) 9703 sound
Oxygen (OX)	C5-B4-A4 'OX-Y-GEN'	C5-B4-A4—G4-F4 (repeated) 'OX-Y-GEN A-LARM'	Slowly falling pitches; top of a major scale; falling pitch of an oximeter
Ventilation (VN)	C4-A4-F4 'VEN-TI-LATE' 'RISE-AND-FALL'	C4-A4-F4—A4-F4 (repeated) 'VEN-TI-LATE A-LARM' 'RISE-AND-FALL AND-FALL'	Old 'NBC chime'; inverted major chord; rise and fall of the lungs
Cardio-vascular (CV)	C4-E4-G4 'CAR-DI-AC'	C4-E4-G4—G4-C5 (repeated) 'CAR-DI-AC A-LARM'	Trumpet call; call to arms; major chord
Temperature (TE)	C4-D4-E4 'TEM-P'RA-TURE'	C4-D4-E4—F4-G4 (repeated) 'TEM-P'RA-TURE A-LARM'	Slowly rising pitches; bottom of a major scale; related to slow increase in energy or (usually) temperature
Infusion (IN)	C5-D4-G4 'IN-FU-SION'	C5-D4-G4—C5-D4 (repeated) 'IN-FU-SION A-LARM'	Jazz chord (inverted 9th); drops of an infusion falling and 'splashing' back up
Perfusion (PE)	C4-F#4-C4 'PER-FU-SION'	C4-F#4-C4—C4-F#4 (repeated) 'PER-FU-SION A-LARM'	Artificial sound; tri-tone; similar to 'yo-ee-oh' of the Munchkins in 'The Wizard of Oz'
Power failure (PF)	C5-C4-C4 'POW-ER FAIL' 'GO-ING DOWN'	C5-C4-C4—C5-C4 (repeated) 'POW-ER GO-ING DOWN'	Falling pitch as when the power has run down on an old Victrola

Table 1: Description of the melodic alarms introduced in Block et al.(2000) and recommended in IEC 60601-1-8. Approximate total duration for the medium priority alarms is 920 msec and for each repetition of the high-priority alarms is 1250 msec.

3. EVALUATION OF EFFECTIVENESS

There have been several evaluations of the IEC 60601-1-8 melodic alarm recommendations. All have been performed by research groups unassociated with the work of the committee. The studies suggest that the melodic alarms may be difficult to use in practice, but studies examining the alarms in a clinical context have not yet been reported.

3.1. Williams and Beatty study

Williams and Beatty [7] trained 21 non-clinicians to identify the Block et al. [6] alarms using just the mnemonics (eg 'CAR-DI-AC A-LARM') and without the further supporting information that explains the mapping of the melody to the alarm source (eg 'trumpet call; call to arms; major chord'). Participants learned the alarms in a first session and then returned a week later for further learning and test. During testing, identification accuracy ranged from 10% for the medium-priority ventilator (VN) alarm to 61% for the medium-priority oxygen (OX) alarm (full names and codes for alarms are in Table 1). Accuracy for the simple GE alarms was 85% and 90% for medium- and high-priority alarms respectively.

There were systematic confusions between alarms, such as between CV and TE, IN and VN, PE and VN, PE and IN, CV and OX. However, participants rated all the high-priority alarms as sounding significantly more urgent than the medium-priority alarms, and they rated the medium-level alarms as sounding more urgent than the low-priority alarm.

3.2. Sanderson, Wee and Lacherez study

In a study that also ran over two sessions about a week apart, Sanderson et al. [8] trained 33 non-clinicians on the high-priority and medium-priority alarms, using the alarms in [6, 8] as specified within the IEC 60601-1-8 standard. One group of participants learned the alarms with full support of the mnemonics and the further supporting information shown at right in Table 1. A second group of participants learned without any mnemonic and without any further supporting information.

Sanderson et al. assessed how many participants reached the learning criterion of correctly identifying all 16 alarms correctly on two successive occasions, when alarms were presented in random order. On Day 1, only 44% of the mnemonic participants and 27% of the non-mnemonic participants reached the learning criterion. On Day 2 only 28% of the mnemonic and 27% of the non-mnemonic participants reached the criterion.

Curiously, Sanderson et al. found that responses to the medium-priority alarms were faster and more accurate than responses to the high-priority alarms. It would be safer for responses to be faster and more accurate to high-priority alarms than to medium-priority alarms, but this was not so. No latency data were reported by Williams and Beatty [7] and a statistical test of any difference in accuracy between the high- and medium-priority alarms was not performed, so a direct comparison cannot be made between the studies. A runs test on the descriptive statistics in [7], however, reveals no accuracy difference between the high- and medium-priority alarms.

Although the methods for classifying confusions as "strong" differed across studies, Sanderson et al. found strong confusions between certain pairs of alarms that partly matched those found by Williams and Beatty. The strongest confusions for both the mnemonic and non-mnemonic participants in the Sanderson et al. study were CV and TE, and IN and VN, as Williams and Beatty had found. Confusions between the other pairs that Williams and Beatty had found were weaker or nonexistent. Interestingly, the mnemonic participants' confusions were focused around a few key alarm pairs, whereas the non-mnemonic participants' confusions were spread over a much wider variety of alarm pairs. Some alarms sound similar, so they are confused on an acoustic basis regardless of whether mnemonics are used to support learning. However, mnemonic support may channel participants towards some confusions, making them more entrenched, and away from other confusions because the mnemonics "standardise" the confusions.

A further factor examined by Sanderson et al. was the effect of prior musical training on participants' ability to learn the alarms. Participants with just one year of formal musical training were much more accurate at identifying the alarms than those with no musical training. It would be a concern if musical training were needed for effective use of the melodic alarms.

Finally, Sanderson et al. asked participants to rate the relative urgency of the medium and high-priority alarms on a 7-point scale. Ratings for the high-priority alarms was 4.8 and for medium-priority alarms was 3.1, which was statistically significant. Interestingly the musically trained participants judged the high-priority alarms as relatively more urgent than the non-musically trained participants did.

3.3. Lacherez, Seah and Sanderson study

It is important that the IEC 60601-1-8 melodic alarms are tested with healthcare personnel rather than solely with naïve volunteers as was the case for the preceding two studies. Moreover, the alarms should be tested under conditions slightly more representative of healthcare environments. We have engaged in a series of studies that are introducing such factors in a systematic way. The study by Lacherez, Seah, and Sanderson [9] is the first in a series addressing such issues.

Lacherez et al. examine the effectiveness of the IEC 60601-1-8 alarms under learning conditions often encountered in busy healthcare systems. Fourteen nurses based in a major metropolitan tertiary-care hospital were participants. The IEC 60601-1-8 alarms were learned without the benefit of mnemonics. Although this is contrary to the intention behind the IEC 60601-1-8 standard, healthcare practitioners advise that many healthcare workers will learn the alarms serendipitously on the job rather through structured training. In this initial study we therefore examined how effectively the nurses could learn to identify the alarms without using mnemonics.

In addition, once the learning phase was over, Lacherez et al. examined how effectively nurses could identify the alarms when two alarms overlapped and while the nurses performed a distractor task (a mental arithmetic task presented on a screen with continuous performance feedback). Although IEC 60601-1-8 recommends a centralized alarm control that stages the sounding of alarms in time, it is unlikely that centralized control will be used as the alarms appear on all the disparate forms of equipment in critical care areas. It is important to know whether the benefits of having informative alarms may be nullified if the alarms overlap in time. In this study we therefore used just the high-priority alarms so that we could examine all the combinations of overlaps in a reasonable amount of time. [**OXYGEN-CARDIOVASCULAR-OVERLAP.WAV**]

Results indicate that nurses learning without mnemonic or further support did not perform better than the non-nurses learning under the same conditions in the Sanderson et al.[8] study. Only 14% of the nurses reached the learning criterion on Day 2 (compared with 27% of the equivalent group in the Sanderson et al. study). However the performance of musically-trained nurses was again significantly better than that of non musically-trained nurses. A much greater proportion of nurses confused certain pairs of alarms than in the equivalent group in the Sanderson et al. study. Strong mutual confusions were seen between CV and TE, and between PE and PF. In addition, in a series of one-way confusions, TE was mistaken for OX, IN for PE and VN for TE and IN.

In the transfer trials, if two alarms overlapped 75% or more, nurses' ability to identify the alarms, and to identify them in the right order, was at chance, whereas the ability to identify sequential alarms or single alarms was significantly better. On the transfer trials where the distractor arithmetic task was used, nurses took longer to respond but their accuracy was unaffected.

3.4. Further work under way and planned

A further study currently being conducted by Alexandra Wee examines issues unanswered by the preceding work. One group of nurses learns with mnemonics and further supporting information whereas the other group has no mnemonics or supporting information. After one session of learning and a refresher session a week later, the nurses identify the alarms while performing a distractor task. This study will provide a definitive answer to the question of whether mnemonics help nurses learn the alarms (very preliminary results suggest not), what level of performance can be expected in about 40-60 minutes of training, and how much alarm identification performance suffers when other tasks must be performed.

A further study is planned for 2006 with funding from the Australian and New Zealand College of Anaesthetists, in collaboration with Dr C Thompson who served on the IEC 60601-1-8 standards committee. The first goal will be to test the effectiveness of the IEC 60601-1-8 alarms with anesthetist participants working under different levels of auditory and cognitive distraction. A second goal will be to test some relatively simple design changes to the IEC 60601-1-8 alarms that have been proposed in hopes of improving their learnability, decreasing their confusability, and sharpening the distinction between the high- and medium-priority alarms.

3.5. Summary and discussion

It is clear from the above studies that the IEC 60601-1-8 melodic alarms may pose some problems in practice. First, it is unclear how much training, and of what kind, will most efficiently lead to acceptable levels of accuracy for clinical use. Second, it is unclear whether healthcare workers will benefit from the mnemonics supplied more than naïve participants do, although the answer is near at hand. Third, it is unclear whether learning will be robust when other tasks must be performed.

At present, the IEC 60601-1-8 alarms have many similarities, making it difficult to discriminate among them. (1) they are all in the key of C major. (2) They all have the same rhythmic structure ("3" for medium priority or "3-2, 3-2" for high priority). (3) They all have the same timbral structure. A long tradition of research in earcon design indicates that auditory patterns that are similar along the above three dimensions are hard to discriminate when heard singly and especially when heard concurrently [10]. The alarms' consistent rhythmic structure has apparently been inherited from the ISO 9703 standard as a Level 1 earcon property [10], making it even more important for the other dimensions to vary. In summary, the kind of human factors studies outlined above should have been initiated during the work of the IEC 60601-1-8 standards committee rather than afterwards when it is too late.

4. URGENCY MAPPING

As noted at the start of this paper, melodic alarms are an attempt to provide a more informative environment in critical care areas. Even before the release of the IEC 60601-1-8 standard and the research supported here, however, other concerns were being raised about how effective melodic alarms could possibly be in attracting attention when needed, in competition with all the other demands on a clinician's attention. It has been argued [11, 12] that attention will always be required to identify the source of melodic alarms such as those proposed by Block et al.[6] and found in IEC 60601-1-8. When the anesthetist's attention is elsewhere, perception of unattended stimuli will be less likely and identification will be slower and less accurate. Instead, an alarm with urgency mapped into the physical properties of the alarm (via pitch, speed or timbre) is more likely to sound urgent and therefore attract attention. Those physical properties should be matched to the urgency of the alarm itself—a design principle known as urgency mapping [13].

The above argument deserves considerably more attention than it is possible to give here. Several issues can, however, be raised. First, the exact contrast being made is unclear. Melodic alarms of course come with three levels of urgency—low, medium and high. The strongest form of the urgency mapping proposal is to label alarms solely by their level of urgency rather than by their identity. Second, in some domains urgency depends on the clinical context [14] and cannot always devolve to the physical attributes of alarms as assessed out of context. A single mapping of situation urgency with physical alarm urgency will not be adequate. Third, urgency mapping does not solve the original problem of clinicians needing to know which alarm has sounded. Fourth, alarms can have strong subjective importance for a healthcare worker because of the potential for doing harm to the patient. Research shows that unattended auditory cues with strong subjective significance may be enough to reorient attention [15]. Fifth, if alarms are undiagnostic or unselective, as they tend to be, urgency mapping may not attract attention and appropriate responses. Research shows that participants match their response rate to the hit rate of an alarm, and the latter can be very low [16]. Sixth, various groups have used anesthetists to rank order the urgency of different subsystem failures and thereby of the alarms that should sound [17, 18]. However, critical studies have not yet been done in true clinical contexts to determine whether melodic alarms are more effective in attracting attention at the right time than urgency mapped alarms. Seventh, some studies indicate that the predictions of urgency mapping are not always borne out empirically [19].

Overall, we may find that urgency mapping is more appropriate in some work domains (eg military helicopters) than others. The role of domain differences in determining the best kind of alarm design was noted by Kerr [3] but has not received a systematic analysis since then.

5. CONCLUSIONS

In summary, the melodic alarms proposed in IEC 60601-1-8 appear to be difficult to learn and easily confusable. The results of the studies reported here underscore the need for effective, user-centred, formative testing of novel auditory display concepts before they are promulgated in international standards.

Although the IEC 60601-1-8 alarms need further development, it may be premature to conclude that alarms that convey information about their source are unworkable and that alarms should be designed solely around predicted urgency. Further research on critical care clinicians using melodic alarms in rich clinical contexts are needed before we really understand whether melodic alarms are appropriately designed to attract the clinician's attention when needed.

6. REFERENCES

[1] A. Guillaume, E. Jacob, B. Leonore, C. et al, "Analysis of auditory warning signals in the operating room," in *Proceedings of ICAD2005* Limerick, Ireland, 2005.

[2] R. D. Patterson, "Guidelines for auditory warning systems on civil aircraft," 1982.

[3] J. H. Kerr, "Warning devices," *British Journal of Anaesthesia*, vol. 57, pp. 696-708, 1985.

[4] M. Weinger, "Proposed new alarms standards may make a bad situation worse (letter)," *Anesthesiology*, vol. 74, pp. 791-792, 1991.

[5] F. E. Block, "Evaluation of Users' Abilities to Recognize Musical Alarm Tones," *Journal of Clinical Monitoring*, vol. 8, pp. 285-290, 1992.

[6] F. E. Block, J. D. Rouse, M. Hakala, and C. L. Thompson, "A proposed new set of alarm sounds which satisfy standards and rationale to encode source information," *Journal of Clinical Monitoring and Computing*, vol. 16, pp. 541-546, 2000.

[7] S. Williams and P. C. W. Beatty, "Measuring the performance of audible alarms for anaesthesia," *Physiological Measurement*, vol. 26, pp. 571-581, 2005.

[8] P. Sanderson, A. Wee, and P. Lacherez, "Learnability and discriminability of melodic medical equipment alarms," *Anaesthesia*, vol. 61, pp. 142-147, 2006.

[9] P. Lacherez, L. Seah, and P. Sanderson, "Overlapping melodic alarms are almost indiscriminable," *Human Factors,* accepted pending minor revision.

[10] D. McGookin and S. Brewster, "Understanding concurrent earcons: Applying auditory scene analysis principles to concurrent earcon recognition," *ACM Transactions on Applied Perception*, vol. 1, pp. 120-155, 2004.

[11] T. A. Mondor and G. A. Finley, "The perceived urgency of auditory warning alarms used in the hospital operating room is inappropriate," *Canadian Journal of Anaesthesia*, vol. 50, pp. 221-228, 2003.

[12] G. R. Arrabito, T. A. Mondor, and K. J. Kent, "Judging the urgency of non-verbal auditory alarms: A case study," *Ergonomics*, vol. 47, pp. 821-840, 2004.

[13] J. Edworthy, S. Loxley, and I. Dennis, "Improving auditory warning design: Relationship between warning sound parameters and perceived urgency," *Human Factors*, vol. 33, pp. 205-231, 1991.

[14] F. J. Seagull and P. M. Sanderson, "Anesthesia alarms in context: An observational study," *Human Factors*, vol. 43, pp. 66-78, 2001.

[15] R. J. Dolan, "Emotion, cognition, and behaviour," *Science*, vol. 298, pp. 1191-1194, 2003.

[16] J. P. Bliss and R. D. Gilson, "Human probability matching behaviour in response to alarms of varying reliability," *Ergonomics*, vol. 38, pp. 2300-2312, 1995.

[17] T. C. Walsh and P. C. W. Beatty, "Establishing scales of perceived severity for clinical situations during anaesthesia," *British Journal of Anaesthesia*, vol. 95, pp. 339-343, 2005.

[18] G. A. Finley and A. J. Cohen, "Perceived urgency and the anaesthetist: responses to common operating room monitor alarms," *Canadian Journal of Anaesthesia*, vol. 38, pp. 958-964, 1991.

[19] E. E. Wiese and J. D. Lee, "Auditory alerts for in-vehicle information systems: The effects of temporal conflict and sound parameters on driver attitudes and performance," *Ergonomics*, vol. 47, pp. 965-986, 2004.