doi:10.1111/j.1365-2044.2005.04502.x

Learnability and discriminability of melodic medical equipment alarms*

P. M. Sanderson,[1] A. Wee[2] and P. Lacherez[3]

1 Professor of Cognitive Engineering and Human Factors and 3 ARC Research Associate, ARC Key Centre for Human Factors; 2 PhD candidate, School of Information Technology and Electrical Engineering, The University of Queensland, St Lucia, Queensland, Australia

Summary

Melodic alarms proposed in the IEC 60601-1-8 standard for medical electrical equipment were tested for learnability and discriminability. Thirty-three non-anaesthetist participants learned the alarms over two sessions of practice, with or without mnemonics suggested in the standard. Fewer than 30% of participants could identify the alarms with 100% accuracy at the end of practice. Confusions persisted between pairs of alarms, especially if mnemonics were used during learning ($p = 0.011$). Participants responded faster ($p < 0.00001$) and more accurately ($p = 0.002$) to medium priority alarms than to high priority alarms, even though they rated the high priority alarms as sounding more urgent ($p < 0.00001$). Participants with at least 1 year of formal musical training identified the alarms more accurately ($p = 0.0002$) than musically untrained participants, and found the task easier overall ($p < 0.00001$). More intensive studies of the IEC 60601-1-8 alarms are needed for their effectiveness to be determined.

Correspondence to: Penelope Sanderson
**Presented in part at the 10th IFIP TC13 International Conference on Human–Computer Interaction (Interact 2005), Rome, September 2005.*
Accepted: 19 October 2005

Novel displays and alarms are being explored for use on medical equipment in critical care environments, but relatively little is known about their effectiveness [1, 2]. The IEC 60601-1-8 standard for medical equipment alarms [3], initially released in 2003 for international use and revised in 2005, offers equipment manufacturers an option to create melodic alarms that distinguish the physical or physiological system that each alarm represents.

In the 1980s, Patterson *et al.* proposed melodic alarms for oxygenation, ventilation, cardiovascular functioning, temperature, perfusion and infusion [4]. The melodic alarms were controversial [5] and untested in a clinical setting. Later, Block proposed melodies that had names semantically associated with each alarm source (e.g. oxygenation and 'Love is blue') and found that anaesthetists could rapidly achieve good learning when melodies' names were supplied [6]. Further melodic alarms discussed by Block *et al.* [7] and incorporated in IEC 60601-1-8 distinguish alarms in eight organ systems – Patterson *et al.*'s original six plus a power failure alarm and a general alarm. There is a three-note melody played once for medium priority alarms and a five-note version of the melody played twice for high priority alarms. Mnemonics are provided to help learners associate the alarms' labels with the melodies used. Evaluations of the melodic alarms were not performed to guide the development of the standard, despite a stated need [7]. However, evaluations of the sounds have emerged since the standard was published [8, 9].

Patient safety relies upon alarms being easily distinguished. The IEC 60601-1-8 standard includes 17 melodic alarms but prior research indicates that listeners can only learn to recognise four to six audible alarms, increasing to 10 or 12 alarms after a week of practice [10]. Moreover, previous research shows that alarms with similar rhythms and repetition [11] or similar duration and amplitude [12] such as the IEC 60601-1-8 melodic alarms, are more likely to be confused. Finally, most previous research on discriminating between alarms has investigated alarm labels rather than mnemonics for mapping labels to alarm sounds. However, when alarms are learned with meaningful labels rather than neutral labels, there is a change in the pattern of alarms that are confused with each other [12].

Given the above, it may be difficult for users to learn the IEC 60601-1-8 melodic alarms and the mnemonics supplied may not aid learning. We studied the ability of participants to learn the IEC 60601-1-8 melodic alarms either with or without the help of the mnemonics, looking at the effectiveness of the alarms (whether they caused confusion), the usefulness of the mnemonics to aid learning, and the role of prior musical training.

Methods

Thirty-three undergraduates from the School of Psychology at The University of Queensland participated in the investigation for course credit. Ethical clearance was provided through the School of Psychology's Ethics Committee and written informed consent was obtained from all participants. Six participants in each learning condition satisfied our criterion for being considered musically trained (at least 1 year of formal musical training). All participants reported normal hearing.

We created 16 melodic alarms that conform to the specifications in IEC 60601-1-8, which are based on those of Block *et al.* (Table 1) [7]. The alarms were created using CSOUND [13] and were processed on a Pentium® 4 1.9 GHz personal computer with a Soundblaster® live 5.1 digital soundcard. The eight high priority alarms were a series of five pulses (notes) consisting of three initial pulses separated by a slight pause before a further two pulses. After a longer pause the pattern was repeated to make a '3–2, 3–2' pattern overall. Pulse rise time (time from zero to maximum amplitude) and pulse fall time were each 10 ms. Pulse sustain time (time at maximum amplitude) was 100 ms and pulse width (pulse sustain time + 10% pulse rise time + 10% fall time) was 102 ms. The silence between each adjacent pulse in the group of three pulses and the group of two pulses was 30 ms. The interval between the beginning of the fall time for one pulse and the end of the rise time for the next pulse was 50 ms. Spacing between the third and fourth pulses was 498 ms. For the eight medium priority alarms, all values were twice those for the high priority alarms. (Spacing between pulses of the medium priority alarms was inadvertently 25 ms shorter than recommended by the standard; however, the intended perceptual difference between the medium and high priority alarms was preserved, as the results attest). The melodies in Table 1 were superimposed on these patterns of pulses. The lowest note was middle C (C4: 278.4 Hz) and the highest note was the octave above middle C (C5: 556.8 Hz). The sounds were presented via Harmon/Kardon HK695 speakers at an average amplitude of 60 dB SPL(A). The speakers were located about 90 cm from the participant. Alarm labels and the mnemonics were displayed on a 19-inch touch screen monitor.

Participants were allocated at random to the mnemonic (n = 18) or non-mnemonic (n = 15) learning condition.

Table 1 IEC 60601-1-8 alarms with associated melody, mnemonic labels, and other information in support of the mnemonic

Alarm	Melody*,† and mnemonic		Other information in support of mnemonic
	Medium priority	High priority	
General	C4-C4-C4	C4-C4-C4—C4-C4 (repeated)	Fixed pitch, traditional (usual) 9703 sound
Oxygen	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	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
Cardiovascular	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 (or delivery of energy)	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 (drug delivery)	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 (artificial perfusion)	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 failures	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

*C4 is middle C; C5 is the C above middle C.
†Sound files are available on request to Penelope Sanderson.

For each participant the experiment ran over two sessions that were conducted 4–7 days apart. At the start of Day 1, the experimenter introduced each of the 16 alarms and its meaning to the participant. In the mnemonic condition, the experimenter also sang each alarm melody and explained the mnemonic. Learning-test cycles proceeded until two successive tests were completed with 100% accuracy or the maximum learning time (35 min) was reached. In each learning phase, the computer monitor displayed all 16 alarm labels, with the high priority alarms arranged vertically on the right and the medium priority alarms arranged vertically on the left. Participants touched the label of the alarm they wished to hear. They could listen to each alarm only once per learning phase, but in any order, and could skip hearing specific alarms. If participants were in the mnemonic condition, then during the learning phase they also saw the words to the alarm tune (e.g. 'CAR-DI-AC A-LARM') and the description provided in the rightmost column of Table 1 (e.g. 'Trumpet call; call to arms; major chord'). In each test phase the alarms were played to the participant in random order without any identifying information. After each alarm sounded the participant touched the label on the computer screen that matched the alarm. At the end of each test phase the computer displayed the names of alarms the participant had not identified correctly. The participant then proceeded to the next learning phase. Learning-test cycles continued until the learning criterion was met (see above). At the end of Day 1 participants answered a questionnaire that asked them how urgent the high *vs* medium priority alarms seemed, how easy it was to distinguish the alarm sounds, whether they thought the mnemonic helped them to remember the sounds, and how pleasant the sounds were.

On Day 2, participants completed two initial test phase trials to determine how much they remembered from the previous session. Then they performed learning-test cycles until two successive tests were completed with 100% correct, until 35 min had elapsed, or until seven learning-test cycles were completed. Participants then completed a further questionnaire. Overall, the time provided to learn the IEC 60601-1-8 alarms was about the maximum that would be expected in standard in-service training for hospital staff on new medical equipment.

Data were tested for significance using two-way tests with $\alpha = 0.05$, using Statistica™ v6.1. Data for response accuracy and speed were analysed with a split-plot factorial ANOVA. Between-subjects factors were learning condition (mnemonic *vs* non-mnemonic) and level of musical training (musical *vs* non-musical). Within-subjects factors were alarm priority (high *vs* medium) and learning-test phase. On Day 1, the first two tests (Day 1 start) and last two tests (Day 1 end) were used. On Day 2 the initial two recall tests (Day 2 start) and the last two relearning tests (Day 2 end) were used. Differences between means were tested using the Tukey HSD test, which preserves $\alpha = 0.05$ across a set of comparisons. To test whether the frequency of confusions between alarms depends on condition, the Chi-squared test was used. Responses to questionnaires were tested either in a factorial ANOVA with the between-subjects factors of learning condition and musicality, or in a split-plot factorial ANOVA that added the within-subjects factor of Day. An *a priori* power analysis based on previously published data for a between-subjects comparison of alarm label salience [12] showed that the present study could detect a 2% improvement in accuracy with a power of > 90%.

Results

Response accuracy

Not all participants successfully learned the alarms to the criterion of 16/16 correct in two successive learning-test cycles. On Day 1, 8/18 (44%) mnemonic participants and 4/15 (27%) non-mnemonic participants met the criterion and on Day 2, only 5/18 (28%) mnemonic participants, and 4/15 (27%) non-mnemonic participants met it.

Figure 1 shows the speed and accuracy of discriminating alarms at the start and end of Days 1 and 2 for participants in the mnemonic and non-mnemonic conditions. There was no significant effect of learning condition (mnemonic 5.7 and non-mnemonic 5.2 correct responses out of 8 questions; $p = 0.39$). However, the response was more accurate for medium priority alarms than for high priority alarms (5.6 and 5.3 correct responses, respectively, out of 8 questions; $p = 0.002$). Participants with musical training responded much more accurately overall than those without musical training (6.5 and 4.4 correct responses, respectively, out of 8 questions; $p = 0.0002$). Accuracy increased significantly within each day but dropped between days, leading to a significant effect of learning-test phase (Day 1 start 4.2, Day 1 end 6.0, Day 2 start 5.2 and Day 2 end 6.5 correct responses out of 8 questions; $p < 0.0001$). There was a significant improvement in accuracy within Day 1 and within Day 2, and from the start of Day 1 to the start of Day 2. However, accuracy fell between the end of Day 1 and the start of Day 2, and at the end of Day 2 accuracy was no better than at the end of Day 1 (all comparisons significant at $p < 0.001$). None of the two-, three- or four-way interactions in the ANOVA was significant.

Response latency

Neither learning condition (mnemonic 5494 ms and non-mnemonic 5947 ms; $p = 0.45$) nor musical training

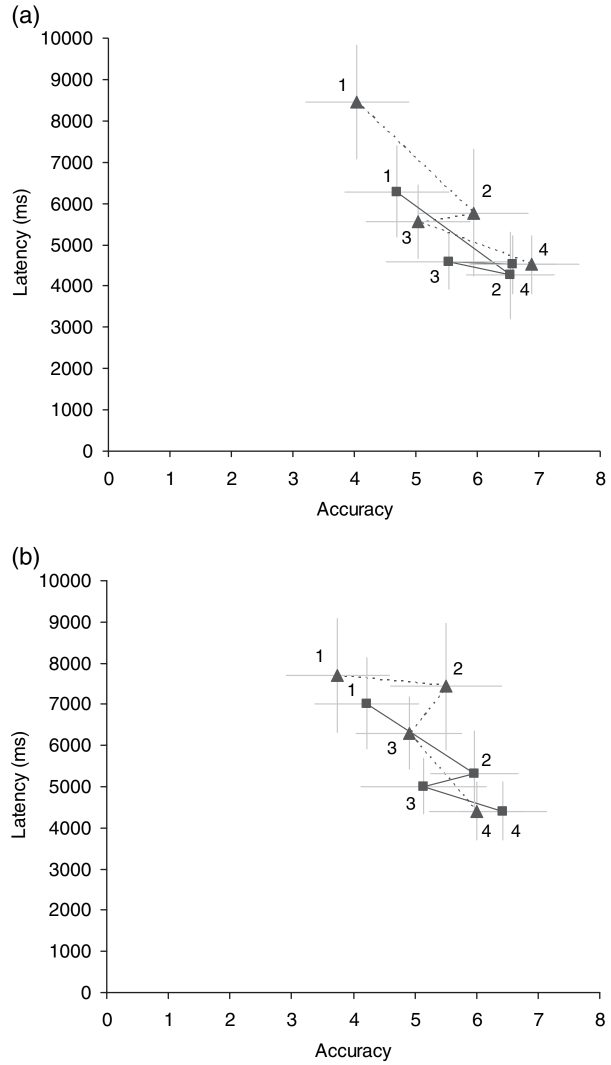

Figure 1 Latency and accuracy for (a) mnemonic and (b) non-mnemonic conditions, with accuracy out of a maximum possible eight correct for high (–▲–) and medium (–■–) priority alarms. 1 = Day-1 start, 2 = Day-1 end, 3 = Day-2 start, and 4 = Day-2 end. Error bars (in grey) are 95% CI. The bottom right of each figure represents fast, accurate performance.

(musical training 5300 ms and no musical training 6140 ms; p = 0.16) had an effect on the response latency. Participants responded faster to medium priority alarms than to high priority alarms (5174 ms and 6267 ms, respectively; p < 0.00001). There was a significant effect of learning-test phase (p < 0.00001) and a significant speed-up across all phases except from the end of Day 1 to the start of Day 2 (Fig. 1).

There was an interaction of alarm priority with learning-test phase, indicating that participants reached a top speed much sooner when responding to medium priority alarms than to high priority alarms (p = 0.00002). An interaction of alarm priority with musicality indicated that high priority alarms slowed the non-musical participants' response more than that of musical participants (p = 0.022). An interaction of alarm priority, phase and condition indicated that non-musical participants in the non-mnemonic condition showed much less speeding up of response from the start to the end of Day 1, compared with participants in the mnemonic condition and compared with all musical participants (p = 0.017).

Confusions

The number of confusions between alarms was collated for the learning-test phases on Day 1 and Day 2. Fewer than 3% of the confusions were between high and medium priority alarms, so the results were combined. There were many confusions between alarms on Day 1. Fig. 2 presents results for the learning-test phases on Day 2 by which time overall speed and accuracy were approaching an asymptote and confusions should have been reduced. The cardiovascular and temperature alarms in particular were frequently confused with each other, as were the infusion and ventilation alarms.

Participants learning under mnemonic conditions showed a narrower range of confusions than those learning under non-mnemonic conditions (Fig. 2). In the mnemonic condition, 80% of participants' confusions were shared by at least one other participant, whereas in the non-mnemonic condition only 29% of participants' confusions were shared by at least one other participant (p = 0.011).

Questionnaire

Participants rated the high priority alarms as sounding more urgent than the medium priority alarms (4.8 on a 7-point scale *vs* 3.1, respectively; p < 0.00001). More musically trained participants rated the high priority alarms as relatively more urgent (p = 0.031), and reported that it was easier to distinguish the alarms (4.8 *vs* 2.9, respectively; p < 0.00001) than did the non-musically trained participants.

Discussion

At the end of two learning sessions, spaced a week apart, fewer than 30% of our participants could identify the 16 IEC 60601-1-8 melodic alarms with 100% accuracy in two successive tests. In addition, there were persistent confusions between certain alarms, corroborating previous findings [8, 9].

Unexpectedly, and despite the experiment's high statistical power, the mnemonics provided in IEC 60601-1-8 had no effect on average speed and accuracy of response. Previous research on the effectiveness of

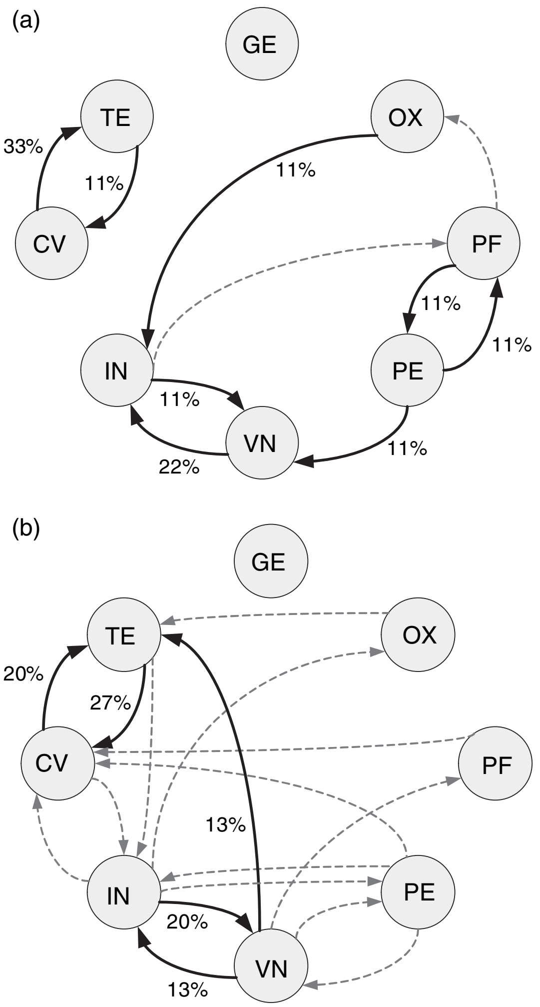

Figure 2 Confusions in (a) mnemonic and (b) non-mnemonic conditions. Numbers indicate the percentage of participants who on more than 25% of the Day-2 test trials confused the alarm at the start of the arrow with the alarm at the end of the arrow. Dotted lines represent confusions that were unique to just one participant. GE, general; OX, oxygenation; PF, power failure; PE, perfusion; VN, ventilation; IN, infusion; CV, cardiovascular; TE, temperature.

labels in helping people learn medical alarms suggests that performance is better if people can generate their own labels [14], whereas in IEC 60601-1-8 the alarm labels are fixed and the mnemonics explain the mapping of the melodies to the labels. In addition, some of the IEC 60601-1-8 mnemonics are culturally and demographically specific. For example, the NBC chime, the Munchkins in the Wizard of Oz, and the sound of a Victrola are best known by an older North American population.

A novel finding is that participants using the mnemonics showed a narrower range of confusions between alarms than those not using mnemonics, most confusions being shared by at least two participants with very few confusions being specific to just one participant. Participants not using the mnemonics showed a broader range of confusions, most of which were specific to just one participant. In both conditions, certain confusions persisted into Day 2, such as between the cardiovascular and temperature alarms, and between the infusion and ventilation alarms. The alarms in each pair of alarms have similar musical contours that are easy to confuse. The confusions could have clinical consequences. Infusion and ventilation are often used together in critical care contexts, so both alarms would frequently have to be discriminated. Cardiovascular monitoring and temperature monitoring are used together less often, but the latter alarms could be mistaken for the more prevalent cardiovascular alarms. Listeners may develop unwarranted confidence in their identification of melodic alarms and so may be less likely to turn to check the source. This possibility should be tested in simulator studies.

A further novel finding is that participants with just 1 year of formal musical training discriminated the melodic alarms almost 50% more accurately and reported the alarms to be easier to learn than did participants without musical training. This is probably because the alarms require the ability to distinguish simple major-scale melodies and to remember them, which is part of musical training. Effective discrimination of alarms in critical care contexts should not depend on musical training.

Participants accurately distinguished between high and medium priority alarms, confusing the two priority levels on only 3% of trials. This is encouraging, as previous research has revealed that distinctions between alarms of different priorities are sometimes not apparent to listeners [15]. Many researchers argue that alarms should convey levels of urgency through tonal features such as loudness or timbre [15, 16] but the IEC 60601-1-8 alarm priorities were discriminated with no such features. Unexpectedly, our participants were faster and more accurate at recognising medium priority alarms than high priority alarms. This is a concern because high priority alarms should be more quickly and accurately discriminated than medium priority ones. The faster pace of the high priority alarms may have made it difficult for participants to encode the melody and retrieve the label. Moreover, the final two pulses of the high priority alarms, which do not map to the first three pulses in a regular way, may have confused participants. The IEC 60601-1-8 standard's adoption of the ISO 9703 '3–2, 3–2' alarm pattern, with no differentiation between alarms in rhythm or pattern of repetition, may therefore make it difficult to learn to discriminate between the alarms.

Future research should address limitations of the present study by examining medically trained personnel, realistic clinical scenarios, longer periods of practice, and the effect of having to do timeshared tasks. Medically trained participants may better appreciate how the mnemonics map the IEC 60601-1-8 melodies onto physiological and physical alarm sources, and so may be able to use the alarms more effectively. However, for simple perceptual discriminations such as those explored here, medical training may produce relatively little difference [17].

Our results suggest that there is cause for concern about the IEC 60601-1-8 melodic alarms. Persistent confusions remain at the end of learning that are not lessened by mnemonics. Moreover, preliminary data on participants' ability to recognise alarms while performing timeshared tasks suggest that accuracy drops when workload increases. In general, thorough user studies should be performed before novel interface designs are promulgated in international medical equipment standards.

Acknowledgements

For information about the work of the IEC 60601-1-8 standards committee, we gratefully acknowledge Dr Chris Thompson, Royal Prince Alfred Hospital, Sydney. For help with preparation of the experimental protocol, the second author in particular thanks Dr Marcus Watson at the ARC Key Centre for Human Factors.

References

1 Sanderson P, Watson M, Russell WJ. Advanced patient monitoring displays: Tools for continuous informing. *Anesthesia and Analgesia* 2005; **101**: 161–8.
2 Edworthy J, Hellier E. Fewer but better auditory alarms will improve patient safety. *Quality and Safety in Health Care* 2005; **14**: 212–5.
3 IEC. *International Standard IEC 60601 (2005): Medical Electrical Equipment – Part 1–8: General Requirements for Safety – Collateral Standard. General Requirements, Tests and Guidance for Alarm Systems in Medical Electrical Equipment and Medical Electrical Systems*. Geneva: International Electrotechnical Commission, 2005.
4 Patterson RD, Edworthy J, Shailer MJ, Lower MC, Wheeler PD. Alarm sounds for medical equipment in intensive care areas and operating theatres. *Institute of Sound and Vibration Research*, Report no. AC598, 1986.
5 Weinger M. Proposed new alarms standards may make a bad situation worse. *Anesthesiology* 1991; **74**: 791–2.
6 Block FE. Evaluation of users' abilities to recognize musical alarm tones. *Journal of Clinical Monitoring* 1992; **8**: 285–90.
7 Block FE, Rouse JD, Hakala M, Thompson CL. A proposed new set of alarm sounds which satisfy standards and rationale to encode source information. *Journal of Clinical Monitoring and Computing* 2000; **16**: 541–6.
8 Wee A, Sanderson P. Testing new alarms for medical electrical equipment. Lecture Notes in Computer Science. *Proceedings of the Tenth IFIP TC13 International Conference on Human–Computer Interaction, Interact2005*. 2005; **3585**: 1145–8.
9 Williams S, Beatty PCW. Measuring the performance of audible alarms for anaesthesia. *Physiological Measurement* 2005; **26**: 571–81.
10 Patterson RD. *Auditory Warnings on Civil Aircraft. The Learning and Retention of Warnings*. Civil Aviation Authority Contract Report 7D/S/0142, R3. London, CAA, 1980.
11 Patterson RD. Auditory warning sounds in the work-environment. *Philosophical Transactions of the Royal Society of London Series B – Biological Sciences*. 1990; **327**: 485–92.
12 Edworthy J, Meredith C. Influence of verbal labeling and acoustic quality on the learning and retention of medical alarms. *International Journal of Cognitive Ergonomics* 1997; **1**: 229–43.
13 Boulanger R, ed. *The Csound Book. Perspectives in Software Synthesis, Sound Design, Signal Processing, and Programming*. Cambridge, MA: MIT Press, 2000.
14 Edworthy J, Hards R. Learning auditory warnings: The effects of sound type, verbal labelling and imagery on the identification of alarm sounds. *International Journal of Industrial Ergonomics* 1999; **24**: 603–18.
15 Mondor TA, Finley GA. The perceived urgency of auditory warning alarms used in the hospital operating room is inappropriate. *Canadian Journal of Anaesthesia* 2003; **50**: 221–8.
16 Edworthy J, Loxworthy S, Dennis L. Improving auditory warning design: Relationship between warning sound parameters and perceived urgency. *Human Factors* 1991; **33**: 205–31.
17 Sanderson P, Tosh N, Philp S, Rudie J, Watson M, Russell WJ. Effects of ambient music on simulated anaesthesia monitoring with visual and auditory displays. *Anaesthesia* 2005; **60**: 1073–80.

