URGENCY ANALYSIS OF AUDIBLE ALARMS IN THE OPERATING ROOM

Christopher Bennett
Anesthesiology
Miller School of Medicine
University of Miami
`cbennett4@med.miami.edu`

Richard McNeer
Anesthesiology
Miller School of Medicine
University of Miami
`mcneer@miami.edu`

Colby Leider
Music Engineering
Frost School of Music
University of Miami
`cleider@miami.edu`

ABSTRACT

Recent studies by researchers, governmental agencies, and safety organizations have recognized a deficiency in the performance of medically related audible alarms [1–4]. In the clinical setting, care providers can suffer from *alarm fatigue*, a condition in which audible alarms in an operating room are perceived as a nuisance. In this study, we explore the auditory features associated with current audible alarms using tools from the music information retrieval community, and then we examine how those auditory features correlate to listeners' perception of urgency. The results show that aperiodic changes in the auditory spectrum over time are the most salient contributor to the perception of urgency in sound. These results could inform the development of a novel standard regarding the composition of medical audible alarms.

1. INTRODUCTION

The need to reevaluate audible medical alarms has been identified by several organizations, including The Joint Commission (www.jointcommission.org, the accreditation and certification organization of U.S. hospitals), the U.S. Food and Drug Administration (www.fda.gov), the Anesthesia Patient Safety Foundation (www.apsf.org), and the American Society of Anesthesiologists (www.asahq.org). Serious errors associated with audible medical alarm perception have also received much recent press [1–4]. Previous attempts have been made by the International Electrotechnical Commission (IEC) to standardize these alarms by providing normative and informative guidelines [5]. These recommendations stipulated that alarms should consist of a series of pulsed tones, each forming a three-note melody.

Under IEC guidelines, melodies are reserved for one of several sentinel events for two levels of priority: *cautionary* and *emergency*. Emergency status, which is meant to engender a greater sense of urgency in the listener, is differentiated from cautionary status by simply lengthening and repeating the melodies and increasing the tempo. However, several previous studies have suggested that these IEC alarms are ineffective at conveying the appropriate level of urgency and are difficult to learn [6–8]. Worse yet, these alarms are often perceived as a nuisance in the intraoperative environment, often leading to physicians' manual silencing of audible alarms, which presents a potentially serious patient safety issue [9].

In this study, we sought to observe the acoustic features of audible alarms currently used in intraoperative environments. Because alarm sounds vary by equipment manufacturer, the current IEC-recommended alarms were used [5]. As mentioned, IEC alarms are differentiated by various short melodies. For example, a cardiovascular event is represented by a major triad in first inversion and ventilation by a major triad in second inversion. (For a complete list of examples, see Table 1.) In addition to the IEC alarms, an experimental alarm set that incorporates additional auditory features besides melody and tempo was synthesized and used in this study for comparison.

Event	Melody	Description
General	$\hat{1} - \hat{1} - \hat{1}$	Ostinato
Oxygenation	$\hat{8} - \hat{7} - \hat{6}$	Falling pitch
Ventilation	$\hat{1} - \hat{6} - \hat{4}$	NBC chime
Cardiovascular	$\hat{1} - \hat{3} - \hat{5}$	*Kumbaya*
Temperature	$\hat{1} - \hat{2} - \hat{3}$	Major scale
Drug Infusion	$\hat{8} - \hat{2} - \hat{5}$	Quartal
Perfusion	$\hat{1} - \hat{4}_{\#} - \hat{1}$	Tritone
Power	$\hat{8} - \hat{1} - \hat{1}$	Octave

Table 1. IEC alarms by source with melodic and descriptive annotations.

The first objective of this study was to observe the primary salient auditory features of the IEC alarm set and to compare it to those of the experimental set. The researchers hypothesized that features would differ significantly between the two sets because the IEC group uses only two auditory dimensions (melody and tempo), and the experimental group employs additional dimensions. The second objective was to investigate the acoustical correlates of urgency perception, thereby laying a foundation for a more robust and comprehensive audible medical alarm protocol.

In this paper, we first present our experimental methods, followed by results and a discussion of the results. We conclude with a description of ongoing and future work.

2. METHODS

2.1 Stimuli

Two sets of alarm sounds were used: IEC alarms and experimental alarms. The experimental alarm set was synthesized by applying various audio effects to a pure-tone carrier; we devised this second set of sounds in order to introduce additional auditory dimensions. These additional auditory effects included amplitude modulation, waveshaping, frequency modulation, phase randomization, and other basic sound synthesis and processing techniques. We constructed a listening test in which alarm sounds were presented to users in a random order; while listening to the alarms, users completed a simple questionnaire. Stimuli were presented once to the subject as 16-bit, 44.1-kHz WAV files and presented over two loudspeakers in a music-rehearsal space.

Section 2.1 summary: IEC/ISO alarms were used as the control set and newly developed alarms as the experimental set. These were played to subjects who completed a questionnaire about the alarm sounds.

2.2 Subjective Experimental Protocol

A total of 21 undergraduate and graduate students in our music technology program were enrolled. Subjects were presented with eight IEC alarms and seven experimental alarms total. Each IEC alarm was presented at both the cautionary and emergency levels, and each experimental alarm was presented at nine levels of auditory effect strength (20% to 100%, in steps of 10%). As an example of effect strength, consider amplitude modulation, in which the effect strength is determined by the modulation depth. These sounds can be found online at http://mue.music.miami.edu/soundSurvey. Subjects were asked to perceptually rate the alarms on a nine-point Likert scale [10] according to their perceived sense of urgency, anxiety, attention, and severity. These terms were predefined for subjects according to the definitions given in Table 2.

Descriptor	**Definition**
Urgency	*A sense of requiring immediate action*
Anxiety	*A sense of apprehension due to uncertainty or doubt*
Attention	*A sense of drawing your focus or observation*
Severity	*A sense of harshness or intensity*

Table 2. Definitions of emotional descriptors used in the Likert questionnaire.

Section 2.2 summary: Music students were asked to rate the IEC and experimental alarm sets on a Likert scale based on their perceived sense of urgency, anxiety, attention, and severity.

2.3 Data Analysis

Subject responses were first tested for normality by applying a Lilliefors test. If the responses were tagged as being significantly different from normal distribution, then those responses were normalized using a power transformation to reduce intra-subject variability [11]. A power transform is rank-preserving but stabilizes the variance to make the distribution more Gaussian. This was useful for comparing responses across subjects. Next, inter-subject statistics were calculated to tag those responses that fell outside of ±3 standard deviations from the inter-subject mean. These outliers were removed from further analysis. Finally, to deal with missing responses, the intra-subject mean was used in place of an empty element for statistical computations.

Section 2.3 summary: Subjects' emotional response ratings were normalized, then outliers were removed, and finally blank responses were filled in.

2.4 Auditory Feature Selection

Auditory feature selection consisted of three primary phases: automatic feature extraction, automatic feature selection, and informed feature parsing. Automatic feature extraction was conducted using MIRToolbox [12]. Each alarm was segmented using one-second frames, and we computed 25 common audio features for each frame. These features describe the dynamics, rhythm, spectral, timbral, and tonal characteristics of each frame. In order to make direct comparisons between features, the inter-frame statistics for each feature were analyzed instead of the feature vectors themselves. These statistics included mean, standard deviation, slope, and periodicity.

After obtaining the feature statistics, automatic feature selection was performed. Features were parsed and included for consideration only if the following criteria were met: features were significant at $p<0.05$, features and urgency ratings had a correlation coefficient $\rho>0.25$, and linear regression of the features against the urgency ratings produced a square correlation coefficient of $R^2>0.25$. These minimum criteria ensured that the selected auditory features would be at least moderately and reproducibly correlated to the four perceptual dimensions of urgency, anxiety, attention, and severity. Features were then sorted using a statistical mapping between the subjects' emotional ratings and the two sets of alarm sounds using methods developed by Lartillot, et al. [13].

Section 2.4 summary: Several auditory features were computed for each alarm sound. Next, these features for each alarm were compared to how that alarm was rated by the subjects on average. Features were removed if they were not significantly correlated to one of the emotional scales. The remaining features were ranked in order of correlation.

3. RESULTS

3.1 Subject Responses

In our initial analysis of subject responses, we performed a squared multiple correlation to test for collinearity among the descriptor dimensions. The results showed that urgency was collinear ($R^2 = 0.95$) with all of the other descriptors. As a result, the other descriptions (anxiety, attention, and severity) were removed from further analysis.

Next, scale validity was confirmed by testing for intersubject correlation (ρ) and Cronbach's α. In general, ρ indicates the degree of linear dependence across subjects and varies from –1 (anti-correlated) to +1 (perfect positive correlation), while α measures the internal consistency or agreement of a psychometric test score across a population, and it varies from 0 (inconsistent responses) to 1 (perfect consistency). A high degree of internal consistency between subject responses was found, both in ρ (0.39) as well as α (0.92).

Section 3.1 summary: The test methodology was validated by exhibiting consistent subject responses. 'Urgency' was highly collinear with the remaining emotional scales, so the others were removed from further analysis.

3.2 Independent Analysis of Alarm Sets

After the combined subjective data were shown to exhibit internal consistency, we proceeded to analyze the features computed within the IEC and experimental alarm sets. The `mirmap` command from MIRToolbox was used to discover which of approximately 300 standard audio features (including statistics on feature vectors) showed strongly positive or negative correlation to the mean urgency rating. Additionally, only features exhibiting statistical significance and sufficient independence ($r_{xy}>0.6$) were selected. On the IEC alarm set, the only feature that correlated strongly to urgency was the magnitude of the spectral centroid periodicity ($\rho=0.9$). For the IEC alarm set, the cardiovascular alarm at the emergency level exhibited the strongest perceptual urgency and the highest rhythmic attack slope.

Analysis of the experimental data set yielded a more robust and descriptive set of correlated features. Five features exhibited strong correlation to mean urgency rating: standard deviation and mean of the rhythmic attack slope (i.e., variation in the "transientness" over time), entropy of the "majorness"/"minorness," variation of the tonal centroid, and the mean spectral roughness. These correlations are illustrated in Figure 1.

Section 3.2 summary: IEC alarms exhibited one auditory feature correlated with urgency, and the experimental alarms exhibited five (as shown in Figure 1).

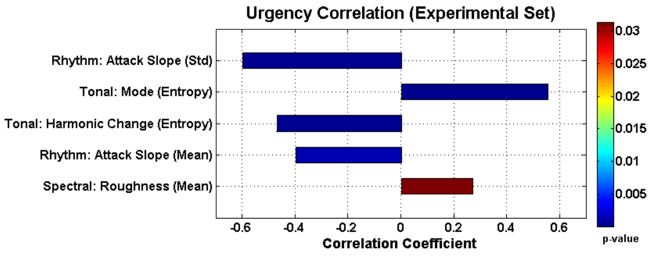

Figure 1. The five most significantly correlated features relating to perceived urgency in the experimental alarm set.

3.3 Post-Hoc Analysis of the Combined Alarm Set

Data from the listening tests of each alarm set were then combined and analyzed together instead of separately to seek additional insights. We performed automatic feature selection by statistically mapping the combined set of features to the combined set of perceived urgency ratings. We found eleven features that met the criteria of $\rho>0.25$ and $p<0.05$. Next, R^2 values were computed by determining how linearly the feature correlated to urgency. However, after removing those features with poor R^2 values, only three features remained, as shown in Figure 2: periodic entropy of spectral flatness (ρ=–0.56*, R^2=0.45), standard deviation of tonal chromagram centroid (ρ=–0.49*, R^2=0.32), and standard deviation of spectral irregularity (ρ=–0.36*, R^2=0.31). Curiously, the three most correlated features in the combined data set were different than those reported by the analysis of each alarm set individually. However, of these, the standard deviation of spectral irregularity ranked sixth in the IEC alarm set and fourth in the experimental alarm set.

Section 3.3 summary: When the IEC and experimental alarm sets and responses were combined, it was found that three features correlated linearly with perceived urgency. These three features were all related to the change of spectral characteristics over time.

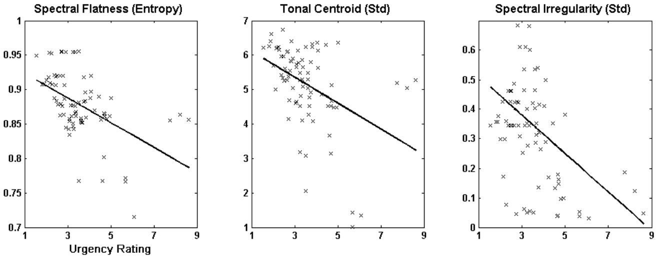

Figure 2. Linear regression of auditory features against perceived urgency.

3.4 Multivariable Linear Regression

Perceptual urgency responses were used to sort the alarms from lowest perceived urgency to highest perceived urgency. The same sorting order was used for each of the three selected auditory features that were highly correlated to perceived urgency. Next, linear regression was calculated from these two data sets to produce a line of best fit (Figure 2) for each feature.

We performed multivariable linear regression to compute the vector of weighting coefficients, $\boldsymbol{W}$, that best fits the equation

$$\boldsymbol{u} = \boldsymbol{W} \cdot \boldsymbol{f} \tag{1}$$

Here, $\boldsymbol{u}$ represents the mean inter-subject urgency rating vector, where $\boldsymbol{u}[0]$ is the collective mean rating of the first alarm, and so on for every alarm sound. The vector $\boldsymbol{f}$ contains the three auditory features described in §3.3, whereby $\boldsymbol{f}^{\mathrm{T}} = \{SFE, SIS, TCS\}$, and *SFE* is the spectral flatness entropy, *SIS* is the spectral irregularity standard deviation, and *TCS* is the tonal chromagram centroid standard deviation.

Performing a least-squares regression results in the following weighting coefficients yields

$$\boldsymbol{W} = \{7.5, -2.8, -0.4\}. \tag{2}$$

Together, SFE and SIS account for 80% of the predicted urgency. These two features were plotted against perceived urgency in a 3D scatter plot, along with a mesh of predicted urgency, in Figure 3.

Section 3.4 summary: An equation for predicting perceived urgency was formulated based on the three most correlated features selected from a set of hundreds of features. The weighting coefficients for those features, **W**, *was determined.*

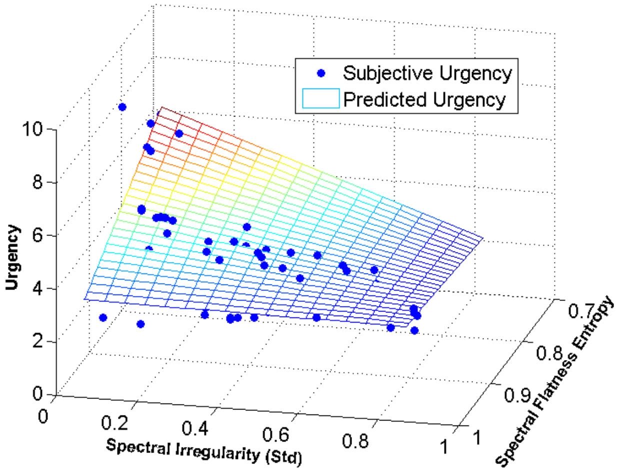

Figure 3. Perceived urgency (•) and predicted urgency (□) are shown. Urgency can be predicted based on the auditory features of inter-frame spectral irregularity standard deviation and inter-frame spectral flatness entropy.

4. DISCUSSION AND FUTURE WORK

Our hypothesis that an experimental alarm set could be constructed that utilized more auditory dimensions was validated by independent objective and subjective analysis of the IEC alarm set and a new, experimental alarm set. The IEC set exhibited only one statistically significant feature that correlated to user perception of urgency, whereas the experimental data set exhibited five. This indicates that a broader set of features can be considered when constructing audible alarms, thereby providing a larger number of "handles" to manipulate when constructing new alarms. This could be a useful tool in addressing the alarm problem, as it has been suggested that more heterogeneity among alarms could improve identification of alarms [14].

Combined analyses of both data sets indicate that changes in an alarm's spectral features over time are the largest contributor to perceived urgency. In each case (periodic entropy of spectral flatness, standard deviation of tonal chromagram centroid, and standard deviation of spectral irregularity), the feature represents an inter-frame statistic. Furthermore, in each case, the feature is anti-correlated to urgency, indicating that fluctuating spectral content is a key to determining perceived urgency.

There are a large number of audible alarms present in the clinical arena, and many of these are false alarms. This large number of false alarms leads to clinicians routinely ignoring them by suffering from alarm fatigue (a concept similar to listener fatigue) [15]. It has been suggested that improving the encoding of alarm information and reducing the number of false alarms could help reduce alarm fatigue [16]. This study provides the framework for establishing new audible alarms that do properly convey urgency, thus augmenting the transfer of information to the clinician.

This study ties together concepts from music information retrieval, such as auditory feature selection, with auditory displays commonly found in a clinical setting. Leveraging these findings, we are currently working on a comprehensive syntax for operating room alarms that is able to convey multiple usable dimensions of data in addition to urgency. This might be done, for example, by mapping one auditory feature to urgency and another feature to indicate the alarm recipient (e.g., anesthesiologist, surgeon, nurse). Ultimately, we hope this work will lead to a robust alarm protocol that will minimize alarm fatigue in the operating room, thereby increasing patient safety.

5. REFERENCES

[1] L. Kowalczyk: "Groups Target Alarm Fatigue at Hospitals—Accrediting Panel Finds Problem Worsening," *The Boston Globe* 18 Apr, 2011.

[2] Emergency Care Research Institute: "Top 10 Health Technology Hazards for 2011," *Health Devices* Vol. 39, No. 11, pp. 389–390, 2010.

[3] M.A. Olympio: "APSF Workshop Recommends New Standards," *APSF Newsletter Winter*, Vol. 19, No. 4, pp. 41, 52–53, 2005.

[4] T. Swartz (ed.): "Technology Overload: An Alarming Danger – Preventing Alarm Fatigue To Improve Patient Safety," *Patient Safety Monitor Journal,* Vol. 12, No. 3, pp. 9–10, 2011.

[5] IEC 60601-1-8: "Audible Alarms in Medical Equipment," *International Electrotechnical Commission*, Geneva, Switzerland, 13 Nov, 2006.

[6] R.R. McNeer, J. Bohórquez, Ö. Özdamar, A.J. Varon, and P. Barach: "A New Paradigm for the Design Of Audible Alarms That Convey Urgency Information," *J Clin Monitor Comp*, Vol. 2, No. 6, pp. 353–63, 2007.

[7] T.A. Mondor and G.A. Finley: "The Perceived Urgency Of Auditory Warning Alarms Used in the Hospital Operating Room Is Inappropriate," *Gen Anesth*, Vol. 50, No, 3, pp. 221–228, 2003.

[8] A.N. Wee and P.M. Sanderson: "Are Melodic Medical Equipment Alarms Easily Learned?" *Anesth Analg*, Vol. 106, No. 2, pp. 501–508, 2008.

[9] P.C. Beatty and S.F. Beatty: "Anaesthetists' Intentions to Violate Safety Guidelines," *Anaesthesia,* Vol. 59, No. 6, pp. 528–540, 2004.

[10] R. Likert: "A Technique for the Measurement of Attitudes," *Psychology*, Vol. 22, pp. 55, 1932.

[11] G.E.P. Box, D.R. Cox: "An Analysis of Transformations," *J Roy Statist Soc*, Vol. 26, pp. 211–252, 1964.

[12] O. Lartillot, P. Toiviainen, and T. Eerola: "A Matlab Toolbox for Music Information Retrieval," in C. Preisach, *et al.* (eds.), *Data Analysis, Machine Learning and Applications*, Studies in Classification, Data Analysis, and Knowledge Organization, Springer-Verlag, New York, 2008.

[13] O. Lartillot, T. Eerola, P. Toiviainen, and J. Fornari: "Multi-Feature Modeling of Pulse Clarity: Design, Validation, and Optimization," *Proceedings of the International Symposium on Music Information Retrieval*, pp. 521–526, 2008.

[14] J. Edworthy, E. Hellier, K. Titchener, A. Naweed, and R. Roels: "Heterogeneity in Auditory Alarm Sets Makes Them Easier to Learn," *Int J Ind Ergonomics*, Vol. 41, pp. 136–146, 2011.

[15] F. Schmid, M.S. Goepfert, D. Kuhnt, V. Eichorn, S. Diedrichs, H. Reichenspurner, A.E. Goetz, and D.A. Reuter: "The Wolf is Crying in the Operating Room: Patient Monitor and Anesthesia Workstation Alarming Patterns During Cardiac Surgery," *Anesth Analg*, Vol. 112, pp. 78–83, 2011.

[16] J. Edworthy, E. Hellier: "Fewer But Better Auditory Alarms Will Improve Patient Safety," *Qual Saf Health Care*, Vol. 14, No. 3, pp. 212–215, 2005.