ENCODING URGENCY IN LEGACY AUDIO ALERTING SYSTEMS

Derek Brock and James A. Ballas

Naval Research Laboratory
Washington, DC
USA
brock|ballas@itd.nrl.navy.mil

Daniel C. McFarlane

Lockheed Martin Advanced Technology Laboratories
Cherry Hill, NJ
USA
mcfarlane@acm.org

ABSTRACT

Despite ongoing modernization efforts, the U.S. Navy expects that it will continue to make highly effective use of legacy systems for many years to come. This and a mandate to maintain fully mission capable platforms have made the service slow to place new audio alerting technologies in command and control environments, despite their demonstrated effectiveness in the laboratory and elsewhere. However, recent upgrade programs for decision support systems have brought with them opportunities to revise and improve standing audio alert techniques. In this paper, the authors describe how the legacy audio component of a Navy decision support workdesk was revised to encode appropriate levels of urgency for incoming action and information alerts. The preliminary design process and issues germane to it are discussed, and the results of an empirical design study are presented. In addition, the implemented solution and the results of a subsequent empirical evaluation are briefly described and discussed.

1. INTRODUCTION

The incorporation and use of auditory alerts in operational settings, where failure to take note of specific information or events can have critical consequences, has generally been more *ad hoc* than principled in practice. However, until recently there have been comparatively few empirically based guidelines for the design of auditory warnings. This historical lack of systematic guidance has often led to the use of either too many sounds or inappropriate sounds in a variety of important circumstances [1].

As an example of this point, in the operational setting that served as the basis for the effort described in this paper, one of the audio alerting strategies used by the information display system sounds a relatively harsh buzzer continuously until the operator takes a specific action. Not surprisingly, there are anecdotal reports of operators in the environment taking matters into their own hands and clipping the buzzer's wires.

This particular system involves an informational alert queue that is associated with an advanced situation status display. The authors were participants in a larger team whose task was to redesign the functioning of the alert queue based on the doctoral work of McFarlane [2] [3] [4] to better manage an increasingly overwhelming number of routine and critically urgent notfications. Part of the effort included a redesign and implementation of the existing audio alerting scheme that would take several levels of informational urgency into account. The redesign, which involved a literature review and an empirical evaluation of a group of candidate alerts, was tightly constrained by the operational limitations of the existing hardware. In particular, there was no capacity for the use of a modern sound card. In spite of this audio limitation, a validated range of new alerts was successfully implemented, and the improved information alert queue was completed and will be placed in service in the near future. The plan of this paper then is to describe the authors' auditory design process and to report and discuss the relevant findings of their evaluations.

2. RELEVANT RESEARCH AND DESIGN CONSTRAINTS

The design process began with an assessment of the existing auditory warning paradigm and a review of recent research on manipulating the perception of urgency in nonspeech sounds.

As one of several core responsibilities in the targeted environment, Navy watchstanders are expected to manage a text-based, alert message queue that delivers a broad range of situationally-relevant information about activities in the current theatre of operations and the status of shipboard systems. Internally, these messages are numerically prioritized, but in present practice, a distinction is only made for the highest priority alerts. As each alert is enqueued, a piezo-electric buzzer under the control of the information display software is sounded. The alerts that require immediate attention trigger the buzzer to stay on until the operator takes action to acknowledge, or "surface," the message. Alerts with lesser priority are lumped together, and for each of these, the buzzer is sounded just once for an eighth of a second. Hence, in the existing paradigm, only two auditory levels of urgency are employed. There are plans to expand the auditory capabilities in future systems of this type, but for now, the current display hardware's ability to generate sound is constrained to on-off control of this buzzer.

One of the intents of the alert queue redesign is to make greater use of the internal alert prioritization scheme. In addition to playing a role in a newly developed automated assessment mechanism, numerical priority assignments will now be shown in a fully redesigned user interface, and the incoming information alerts with priorities other than those requiring immediate attention will be divided into three discrete categories of conceptually decreasing urgency for purposes of auditory alerting.

The limitations of the audio hardware dictated a time-based approach to urgency encoding. In particular, the design of any new sounds to be associated with the watchstander's alert queue would be necessarily constrained to what could be accomplished through software control of the buzzer. This suggested adopting an approach similar to that of Patterson's 1982 proposed guidelines for aircraft warning systems [5]. Warning sounds in Patterson's more general scheme [6] should be composed of one or more "bursts" that are repeated a number of times with specifically varied intensities and starting points over the course of as much as 40 or more seconds or until the warning is acknowledged. Each burst should itself be composed of repetitions of one or more shorter "pulses" whose intensities and starting points are also deliberately and systematically varied. These pulses, which function as the atomic units of a warning sound, should be short acoustic waveforms with a specific amplitude envelope and length (a few hundred milliseconds at most), and intentionally distinct characteristics with regard to pitch and timbre. Consequently, when a burst is composed of differing pulses, it should sound much like a "brief atonal melody with syncopated rhythm" [6].

In addition to making specific recommendations concerning the number of spectral components and appropriate sound levels for warning sounds, Patterson also suggested that within warning sounds and within bursts pitch, intensity, and the temporal parameter of speed could be used to vary perceived urgency, but provided no explicit, quantified advice in this regard. Subsequent studies by Edworthy *et al*. [7] and Hellier *el al*. [8] have addressed this aspect of auditory alert design directly.

In the first of these studies, Edworthy and her colleagues systematically varied a number of acoustic parameters that could be applied to the composition of pulses and to the structure of bursts. Of these, three temporal parameters, namely, speed, rhythm, and number of repetitions, are relevant for auditory alerts subject to the hardware constraint described above. In this research, which involved a series of separate experiments, listeners heard only individual bursts, as opposed to full, Patterson-style auditory warnings made up of several bursts. They were asked in each experiment first to rank from highest to lowest the perceived urgency of a set of bursts in which either two or three auditory parameters had been manipulated and then to give each burst a numerical urgency rating on a scale of 0 to 100. For the three temporal parameters of interest, it was found that greater pulse speed within a burst (i.e., number of pulses per unit of time) translates into additional perceived urgency, that syncopating pulses rather than spacing their onsets in a regular manner translates into lower perceived urgency, and that increasing the number of repetitions of pulse groupings or patterns also increases perceived urgency. An additional, speed-related finding was that perceived urgency can also be increased and decreased by respectively accelerating and decelerating pulse onsets within a burst. The final experiment in this study examined whether auditory parameters related to urgency can be predictably combined and confirmed that, in fact, they can be when the ordinal relationships among parameter values are known.

Hellier *et al.'s* study ([8]) addressed the manipulation of acoustic parameters from a psychophysical perspective and also involved an experiment to quantify the effects of combining parameters. In this effort, the researchers sought to scale the subjective effect of individual parameter manipulations on perceived urgency using the relation expressed by Stevens' power law [9],

$$\psi = kS^n, \tag{1}$$

wherein the exponent n, which is understood to be positive, quantifies the extent to which an objectively measurable change in a stimulus value S (e.g., the value of an acoustic parameter) produces a change in the subjective judgment of an associated property ψ (e.g., perceived urgency). (Note, k is a stimulus-specific constant.) Assuming the scale of ψ is theoretically equivalent for manipulations of different parameters, the value of n for a particular S_i can be construed as an index of that parameter's efficiency with regard to changes in ψ for values of S_i greater than the point at which the slope of Equation 1 reaches unity, which can be expressed as

$$S = \left(\frac{1}{kn}\right)^{\frac{1}{n-1}}. \tag{2}$$

That is, the larger n_i is, the smaller ΔS_i must be above this point to produce a desired change in ψ. Conversely, knowing the value of n for different Ss makes it theoretically possible to equate changes in ψ between one S and another, which would have useful ramifications for designers who wish to distinguish among or combine these parameters.

Only two of the acoustic parameters examined in Hellier *et al.'s* study, speed and repetition, are relevant to the effort presented here. Urgency judgments for these and the other parameters that were investigated proved to be systematic and quantifiable over the ranges that were used. The values of n found for speed and repetition were 1.35 and 0.502, respectively, and were the largest in the study. In theory, smaller changes in these parameters should be needed in comparison to the others to produce equivalent changes in perceived urgency. In the study's parameter combination experiment, the strength of this proposition was tested with a set of 27 bursts, across which three, theoretically equal and equally increasing urgency levels (low, medium, and high) of speed, repetition, and a third parameter, pitch, were covaried. These urgency levels were expected to combine additively, and since the medium and high parameter values each represented a 30% increase in urgency over the level below it, it was also expected that the set of 27 bursts would be partitioned into equivalence classes of increasing urgency. For instance, a burst with high speed and repetition values and a low pitch would be equivalently urgent to a burst with a high speed and medium repetition and pitch values. Although this prediction did not fully prove to be the case (a much wider range of urgency was associated with the contribution of pitch than with speed or repetition whose associated ranges were essentially the same), meaning that the combination of theoretically equivalent parameters is not fully orthogonal for purposes of encoding urgency, the correlation between the empirical and expected rankings of these bursts was nevertheless highly significant, which, in turn, suggests that the parameters of interest in the present work are, to some extent, additive as encoders of urgency.

3. DESIGN PROCESS

In addition to consulting [5], [6], [7], and [8], the authors also listened to several bursts, based on the parameters described in [7] that were designed for a recent series of auditory warning experiments carried out by Guillaume and her colleagues [10]. Using a multidimensional technique for obtaining urgency judgments and a similar set of stimuli in which the same parameters were again combined, Guillaume *et al*. validated the significant correlation between predicted and observed urgency rankings reported in [7].

The temporal parameters in the examples auditioned by the authors included regular, slowing, and syncopated rhythms and the use of repetition. Each was informally evaluated for the present design task in terms of how it might be implemented under the constraints imposed by software control of the targeted hardware's buzzer. To do this, a sound patch tool was used to compose several prototype bursts from a 200 ms snippet of a buzzer sound found online whose character was similar to the actual buzzer. After listening to a few of these, it was decided that the strict use of a single

200 ms pulse was an unnecessary constraint, and that, just as music involves notes of various durations, short and long pulses could be used, particularly in bursts involving syncopation.

This parallel with music led to some additional insights concerning rhythmic affect, phrasing, and syncopation. Although the limitation of simple on/off control of the buzzer meant that it would not be possible to incorporate amplitude-based rhythmic accents in the buzzer burst designs (the use of accents within a burst is an acoustic parameter that was not considered in any of the studies discussed above), this same control allowed for the incorporation of small timing differences in pulse onsets and offsets that succeeded in giving the impression of accents and pulse groupings, even when an effectively regular rhythm was used. It is worth noting that similarly subtle timing differences among note groupings are typically seen in musical performances captured by musical instrument digital interface (MIDI) tools. In addition, there is a small body of empirical evidence and modeling work that suggests that rhythmic meter is perceived even when accents are explicitly absent [11]. The use of subtle timing differences to imply the presence of accents also led to the notion of bursts as rhythmic phrases that have a gestalt. In particular, by subtly lengthening, or in some cases shortening, a burst's final pulse, it proved possible to give the prototypes a more emphatic sense of completion. Finally, the use of syncopation in one of the example bursts from [10], together with the notions of phrasing and rhythmic accents, raised a question concerning the perceptual affect of note duration order in syncopation.

Although syncopated phrases can be composed in any number of ways, the basic ingredients of syncopation are notes of short and long duration ordered in such a way as to contrast with a plain or regular rhythm. In music, the juxtaposition of a short and a long note, or vice versa, is often characterized as a dotted figure. When dotted figures are repeated, they produce a syncopated walking effect: a kind of galloping when the short note comes first (*Dit-dah, Dit-dah, Dit-dah,* to use a jazz rhythmic terminology [12] in which *Dits* are abruptly short leading notes and *dahs* are long) and a skipping effect when the short note comes second (*Doo-dit, Doo-dit, Doo-dit*, with *Doos* used here for leading long notes). Which of these syncopations is inherently more urgent? While intuition suggests galloping over skipping, it was decided to answer the question empirically with a formative evaluation of a small group of prototype buzzer bursts in which some of the urgency encodings would include instances of these two patterns.

An additional purpose for evaluating a set of prototype bursts was to ensure that the conclusions of [7], [8], and [10] were valid when only temporal parameter manipulations were used in combination to encode urgency. The parameters chosen for the formative evaluation were rhythm, speed, and repetition. After careful consideration, it was decided that speed manipulations involving accelerating and decelerating pulse onsets were inappropriate for the present work. In particular, the obtained rankings of bursts involving acceleration in combination with other parameters were inconsistent between [7] and [10]. In the latter effort, bursts involving acceleration were judged to be more urgent than they were predicted to be, which suggests that the use of bursts employing this manipulation alongside other bursts whose urgency encoding is reliably high could be unacceptably confusing to users, given that the working specification required an unambiguous ordering of urgencies. On the other hand, all of the bursts involving deceleration in these studies were predicted and consistently judged to be least urgent. Indeed, after listening to a decelerating burst that was used in [10], it was decided that the presence of this manipulation could even suggest a lack of urgency, which made its use unacceptable.

Ultimately, five prototype buzzer bursts were chosen for the formative evaluation. Three involved repetitions of the two syncopation patterns described above, and the other two involved more straight forward manipulations of speed and repetition. A more thorough description of the bursts can be found in Table 1, where, on the basis of the findings in [7] and [8] and a neutral position on the contribution of different types of syncopations, they are numerically ranked by their theoretical degree of encoded urgency. A description of the method used for calculating and ranking the theoretical urgencies is given in Section 4.3.

burst 1	[5.5]
rhythm, speed:	regular, fast
description:	nine pulses: eight short (165ms) followed by a final short (205ms)
grouping:	four pulses: all short
spacing:	5 ms between each group
repetitions:	2.25
total length:	1535ms
burst 2	[7.5]
rhythm, speed:	syncopated, moderate to fast
description:	six pulses: three short (165ms) alternating with two long (290ms) followed by a final long pulse of 325ms
grouping:	two pulses: one short, one long
spacing:	5 ms between each group
repetitions:	3
total length:	1400ms
burst 3	[8.5]
rhythm, speed:	syncopated, moderate
description:	six pulses: two long (340ms) alternating with two short (190ms) followed by one long (390ms) and one short (190ms)
grouping:	two pulses: one long, one short
spacing:	30 ms between each group
repetitions:	3
total length:	1700ms
burst 4	[11.5]
rhythm, speed:	regular, slow
description:	two pulses: one long (425ms) followed by another long (415ms)
grouping:	two pulses
spacing:	125 ms between each pulse
repetitions:	1 (pattern is not repeated)
total length:	965ms
burst 5	[12.0]
rhythm, speed:	syncopated, moderate (slower than burst 3)
description:	four pulses: one long (340ms), one short (190ms), one long (370ms), ending with one short (190ms)
grouping:	two pulses: one long, one short
spacing:	90 ms between each group
repetitions:	2
total length:	1180ms

Table 1: *Characteristics of the five prototype bursts evaluated in the listening study, ordered by their predicted urgency values (shown in square brackets), with burst 1 being the most urgent.*

4. LISTENING STUDY

Evaluation of the five prototype auditory alerts was carried out as an empirical listening study using judgment procedures that were largely based on the approach used in [7]. Here, participants were asked first to rank the bursts from most to least urgent and then, as a separate task, to estimate the magnitude of each burst's urgency. Two of the authors carried out a pilot run of the procedures on themselves. The test itself was carried out in a sound attenuated booth using two different sound editing tools on a notebook computer to play the stimuli for each of the judgment procedures. The bursts were presented with headphones.

4.1. Method and experimental design

Ten volunteers from the staff at the Naval Research Laboratory (NRL) participated in the study, most of whom were in their late twenties or early thirties. All were asked if they had any hearing problems and none were reported. Since all of the stimuli were to be presented at the same level throughout the study, it was judged that no further evaluation of the participants' hearing was necessary. None of the participants had any prior awareness of the study or its purpose.

The listening process was divided into two parts. In the first part, referred to as the rank ordering task, participants ranked the sounds in terms of their ability to convey urgency. Participants were asked first to listen to all five sounds and then to decide which of these he or she felt sounded the most urgent of the group. The sounds were presented to each listener in a different order. Next, the selected sound was removed and participants were asked to listen to the four remaining sounds and decide which of these now sounded the most urgent[1]. This iterative judgment procedure continued until only one sound was left, which was then taken to be judged the least urgent. The second part of the listening process was a magnitude estimation task. For this procedure, participants were asked to listen to each sound again and to make a numerical estimate of its urgency on a scale of 0 to 100, with 100 being characterized as "the most urgent sound possible" and 0 as "the most nonurgent sound that you can imagine." The order of presentation was different for each listener in this procedure too.

4.2. Results

Results of the listening study are presented in Table 2. Only one of the two rank orderings determined by the judgment tasks was consistent with the theoretically predicted rank ordering, but further examination of the data proved to be useful for making design decisions. The Spearman rank-order correlation coefficient between the ordering obtained in the rank ordering task and the predicted ordering was significant at $r_s = 0.9$ ($p = 0.05$), as were the values of Kendall's coefficient of concordance W (0.274, $p \leq 0.05$) and Kendall's T_C (0.46, $p \leq 0.0003$) for these judgments[2]. Respectively, the latter measures indicate the degree of association among the participants' rank orderings and the correlation between these orderings and the predicted ordering. As for the rank ordering that was obtained in the magnitude estimation task by reducing each estimation to an integer rank, the Spearman correlation with the predicted order was significant at $r_s = 1$, and so, too, were Kendall's W and T_C at 0.4735 ($p \leq 0.01$) and 0.59 ($p \leq 0.00003$), respectively. Additionally, the Spearman correlation between the two methods of ordering the bursts was significant (0.9, $p = 0.05$). Collectively, these measures suggest that the judgments made in the two listening tasks were quite stable across participants, and that the perceived urgency of each burst corresponded exceptionally well, if not perfectly, with its theoretical rank.

[1] Because of the design intent to quickly communicate prioritized urgencies, participants in this rank ordering portion of the study were asked to choose the most urgent sound immediately after hearing a single presentation of each successive group of sounds; concerns about ordering effects and auditory memory were addressed by varying the initial, within-group order of the sounds for each listener.

[2] Unlike [7], Kendall's W is used here instead of W' because the rank ordering task was not carried out as a series of paired comparisons.

Burst urgency judgments					
	rank ordering task		magnitude estimation task		
pred.	mean rank	s.d.	mean est.	s.d.	geo. mean
1	1.8(1)	1.03	82.5(1)	9.79	81.92(1)
2	2.9(3)	0.99	70.0(2)	14.48	68.51(2)
3	2.8(2)	1.32	57.2(3)	23.35	51.96(3)
4	3.7(4)	1.84	55.8(4)	27.68	49.14(5)
5	4.0(5)	0.94	51.2(5)	14.98	49.34(4)

Table 2: *Results of the two urgency judgment procedures in the NRL listening study, showing the observed mean rank, the rank determined by this measure (in parenthesis), and the standard deviation of this mean for each burst in the rank ordering task, and for each burst in the magnitude estimation task, the mean estimation, the rank determined by this measure (in parentheses), the standard deviation of this mean, and finally, the geometric mean estimation and the rank determined by this measure (again, in parentheses). The theoretically predicted rank ordering is shown in the first column.*

4.3. Discussion

Despite these correlations, on its surface the study did not fully resolve all of the issues it hoped to examine. In particular, the rank orderings of bursts 2 and 3, which were expected to address the perceptual affects of contrasting syncopations, were inconsistent between the two judgment tasks. In addition, despite the correlation of burst 4's obtained and theoretical orderings, the large variabilities associated with its respective means were unexpected. Consideration of both of these matters was warranted before the study's results could be used as a basis for the prioritized auditory alert design.

In spite of the intuition that burst 2's short-long syncopation would be perceived as more urgent than the converse order of pulse durations in burst 3, the affects of both syncopations were treated as being equal in the calculation of the predicted rank ordering. The calculation itself was done in the following way: Using the speed and pulse group repetition exponents (respectively, 1.35 and 0.502) and the associated data reported in [8], an average value for the constant k in Equation 1 was determined for each of these parameters. From this, ψ_{speed} and $\psi_{repetition}$ were calculated for each of the five bursts. Using these values, the bursts were then ranked by speed and by repetition. Additionally, the bursts were ranked by rhythm according to the dominance of regularity over syncopation reported in [7]. Ties were assigned the average of the ranks that would have been assigned had the ties not occurred. These three rankings were then summed for each burst, and the order of the resulting totals was taken as the theoretical rank ordering, with the lowest value being the most urgent. (The computed totals are shown in square brackets in Table 1.)

Aside from their contrasting syncopations, the only difference between bursts 2 and 3 was the value of their respective speed parameters. Consequently, if one of the syncopations was, in fact, more perceptually urgent than the other, the obtained distance between their ranks would be expected to be larger or smaller than was implied by their theoretically calculated distance. Clearly, something like this happened in the rank ordering task data, where with less variance, burst 2 was judged to be essentially a rank below its predicted rank, and burst 3's rank remained effectively unchanged, albeit with more variability. However, even though a similar pattern of variability was seen in the magnitude estimation task, the mean urgency judgments of these bursts were entirely consistent with their predicted rank ordering.

While there may not be a full explanation for the disparity in these data, some useful conclusions can still be drawn. First, it could be that both types of syncopation are perceived as being equally urgent. Certainly, if either of these patterns are steadily repeated at the same speed and without accents, they can be easily confused with one another. Since the test did not allow for a head-to-head comparison of the syncopation parameter alone, it is possible that the source of the confound is rooted in one or more parameter interactions. Nevertheless, since speed is shown in both [7] and [8] to be a robust and economical encoder of urgency, if perceived urgency is in fact immune to the difference between the syncopations examined here, then the ordering of bursts 2 and 3 in the magnitude estimation data is essentially the result that would be expected. Second, and somewhat encouraging, is the variability in the data, which shows that listeners in both judgment tasks were collectively more consistent in their judgments about burst 2 than about burst 3. Thus, even though the mean of burst 2 in the rank ordering task came in just below that of burst 3, the tighter spread associated with this statistic suggests that, for whatever reason, more listeners in this particular paradigm were clear about where burst 2 fit into the scheme of things. Taken together, these interpretations suggest that in an ordered, three-alert subset of the five prototypes evaluated in the study, burst 2 is a better choice than burst 3 to follow burst 1.

A possible explanation for the large variabilities associated with the mean rank ordering and magnitude estimations for burst 4 may be an ecological factor examined in [10] that was overlooked in the design of the study's five prototype bursts, namely, the perceived urgency of sounds that are already familiar to listeners as auditory alerts in one or another environment. As it turned out, burst 4 happened to sound very much as if a common office or intercom buzzer had been pressed twice to command someone's attention. In [10], it was found that sound designs that are evocative of commonly used auditory alerts in real-world settings are frequently rated more urgent by listeners than their theoretical urgency values would predict. An examination of the participants' individual judgments showed that listeners in both tasks rated burst 4 the most urgent of the bursts more often than any other except for burst 1. Part of the explanation for the substantial variabilities associated with burst 4, then, may be that the similarity of this burst with a familiar alert sound influenced some listeners' judgments of perceived urgency and, so, took precedence over the urgency encoding of its acoustic parameters relative to the other bursts. Since a major aim of the auditory design was to achieve an unambiguously ordered set of alert sounds, this possibility alone made burst 4's inclusion in the prioritized design unlikely. In contrast to the variabilities associated with burst 4's means, which are the largest in the study, the variability of listeners' judgments about burst 5 proved to be comparable in size to that of burst 2, suggesting that listeners were collectively more certain of burst 5's relative urgency than they were about burst 4, much like the distinction between bursts 2 and 3. As a consequence, given the extent of burst 4's variability and its possible perceptual ambiguity, burst 5 was chosen to complete the three-alert audio design proposal.

Before turning to the outcome of the implementation and evaluation phases of this effort, the geometric means that resulted from the magnitude estimation task data are worth considering for their apparent corroboration of a psychophysical justification for manipulating acoustic parameters to encode information such as urgency or importance. These means, which are the Nth root of the product of the N estimations for each burst, indicate what the mean magnitude estimation for each burst is relative to the scalable parameter manipulations in the series. As can be seen in Table 2, the rank ordering of the five bursts obtained from this statistic is not fully consistent with the theoretically predicted ordering. What is interesting about this result is what it reflects about perceived urgency relative to the three acoustic parameter types employed in the study. Since the parameter values involved in each burst were deliberately combined to compose a predictable ordering of urgencies based on the results of [7] and [8], and at least two, speed and repetition, operate on ratio scales, it is arguable from the ordering of the geometric means that it is largely the combination of these two parameters that informed the perceptual affect of each burst. In other words, the slight disagreement of this statistic with the predicted ordering raises the possibility that the importance of the rhythm parameter's contribution may be small at best. Indeed, if the rank orderings for the rhythm parameter are left out of the sums used to determine the five bursts' theoretical rank ordering, the resulting order matches that of the geometric means. Certainly, it would be difficult to assert that rhythm, *per se*, is quantifiably scalable, but more importantly, it may be that the perceptual affect of rhythm has more in common with notions of sound identity than with notions of degree. If at first this appears to challenge one of the findings in [7], it may be that the reason regular rhythms are judged to be more urgent than syncopated rhythms when both are presented at the same speed is that the presence of syncopation naturally draws attention to the dimension of rhythm by virtue of its lack of complete regularity and, so, mildly distracts from the immediate cognition of the central information content encoded in the sound.

5. IMPLEMENTATION DECISIONS AND ADDITIONAL USES OF SOUND

Based on the preceding analysis of the listening study data, bursts 1, 2, and 5 were proposed as the new, prioritized audio alerting scheme for the preliminary design of the revised action and information alert queue described at the beginning of the paper.

As various changes in the system's display of information were being worked out, the design team decided that an additional auditory alert whose purpose was orthogonal to the role of the new bursts was needed to solve an attentional issue in the redesign. While the purpose of the urgency encoded buzzer patterns is to draw prioritized attention to incoming situational advisories as they are enqueued, this additional use of the buzzer would alert operators to a separate display of critical-system, status-change buttons that blink for four seconds at a rate of 1 Hz when an associated update occurs. Since this blink rate was substantially slower than the pulse speed of any of the newly proposed prioritized bursts, arguably any operational confusion about the identity and purpose of an additional burst with a correspondingly slow rate was unlikely. Accordingly, the authors proposed a straightforward, auditory reinforcement of the blink rate with the buzzer, which was adopted immediately.

At this point, a functioning prototype of the redesigned alert system and user interface was implemented in simulation software and evaluated with a group of 8 participants with relevant operational experience to assess the effectiveness of the design and gauge the response of the user community. Although the experimental design at this stage of the effort was unable to accommodate a sound manipulation, the participants were asked to comment on the new uses of sound in exit interviews, resulting in a valuable but mixed assessment of the preliminary auditory design.

5.1. Redesign of the prioritized bursts

Chief among the participant's concerns about the auditory alerts were the lengths of the prioritized bursts, which were thought to be too long, and a perception that the burst with the lowest priority could be interpreted as two separate alerts. These and numerous other findings led to a range of modifications to the preliminary design specification, including a redesign of the prioritized auditory alerts. The new specification retained the original mapping of urgency encoded buzzer patterns to action and information alert priorities but called for a set of much shorter bursts, with durations of no more than 500 ms, and the inclusion of an additional "reminder" alert (a simple, 125 ms pulse proposed by the authors) that would sound periodically when unacknowledged action or information alerts were pending in the queue. In addition, due to a varying range of operational and command concerns raised in the participant interviews, which included additional operator duties involving radio communications, other uses of sound in the operational environment, command protocols, and the noise and pace of business, it was further decided that the implemented system would provide a configuration option that allowed operators to toggle between the existing auditory display (described at the beginning of Section 2) and the new uses of sound.

The implementation schedule did not permit the conduct of further listening studies, so the redesign of the prioritized bursts endeavored to preserve the underlying characteristics of the three, longer bursts the authors had initially proposed on the basis of listeners' judgments. At this point in the process, though, the strong correlations found in the five-burst listening study and the consistent findings of [7], [8], and [10] arguably justified a design approach that relied chiefly on calculations of the new bursts' theoretical urgencies. Because of the requirement for shorter pattern durations, generally shorter pulses were employed, and the use of repetition was abandoned. An additional short pulse was added to the beginning of the short-long, syncopated, second burst (creating a short-short-long pattern) to strengthen its identity and functionally increase its average pulse speed. And finally, all spacing between pulses was eliminated, leaving the ramps of pulse onset and offset envelopes to convey pulse segmentation. Theoretical urgencies were then calculated for each of the new bursts in the same way as before, using the average values of k_{speed} and $k_{repetition}$ derived from [8] to determine the corresponding values of ψ for these parameters (note that the contribution of repetition for each of the new bursts is simply the value of k), which were then used to determine parameter-specific ranks for each burst. Rhythm-specific ranks, per the results of [7], were also assigned as before, and the sum of these three ranks determined each burst's theoretical rank ordering, with the lowest value being the most urgent. Table 3 describes the new bursts in the same manner as Table 1.

burst 1 (final)		[4.0]
rhythm, speed:	regular, fast	
description:	five pulses: four short (85ms) followed by a final short (125ms)	
grouping:	all five pulses	
spacing:	none	
repetitions:	1	
total length:	465ms	
burst 2 (final)		[6.5]
rhythm, speed:	syncopated, moderate to fast	
description:	three pulses: two short (85ms) followed by a long (260ms)	
grouping:	all three pulses	
spacing:	none	
repetitions:	1	
total length:	430ms	
burst 3 (final)		[7.5]
rhythm, speed:	syncopated, moderate	
description:	two pulses: one long (260ms) followed by a short (85ms)	
grouping:	both pulses	
spacing:	none	
repetitions:	1	
total length:	345ms	

Table 3: *Characteristics of the final (implemented), prioritized burst design, ordered by their predicted urgency values (shown in square brackets), with burst 1 being the most urgent.*

5.2. Subjective evaluation of the final implementation

The implemented system was evaluated in two sessions that cumulatively involved 20 participants, all having operational experience. The evaluation involved the analysis of performance data collected during and after the participants' execution of the action and information alerting task in both the old and new system configurations. Each condition was driven by a separate but functionally equivalent scenario involving realistic Navy fleet operations in a fictional encounter. Subjective data was gathered through subject mater expert judgments, standard workload index tools, system-specific questionnaires, and participant exit interviews. Limited access to the participants' time precluded an experimental design that also manipulated the use sound. Instead, data collected in regard to the auditory component was captured in the form of questionnaire and interview responses. Again, the new uses of sound received a mixed subjective evaluation from the participants. Those who commented on sound-related matters in their exit interview, expressed both indifference about the auditory priority scheme during their exercise with the new system and a mild appreciation of its potential had they been given more practice and time to work with it. Another perspective acknowledged the likely benefit but questioned whether the new sound scheme and other enhancements truly simplified the operator's task.

In the first of two system-specific questionnaires, participants were asked to rate the utility of 29 features of the new system's redesigned user interface on a scale of 1 to 7. The mean ratings ranged from 6.21. to 3.85. Four concerned features of the auditory display. Of these, the 125 ms reminder sound feature, signaling the presence of pending, unacknowledged advisories in the action and information alert queue, was rated the most useful at 5.31, followed by the auditory reinforcement of the system status change buttons' blink rate at 5.25. Trailing these were the ability to toggle the system between the old and new auditory displays at 4.25 and the association of different urgency encoded buzzer patterns with the respective priorities of incoming action and information alerts, which received the lowest utility score overall at 3.85. In the second system-specific questionnaire, participants were asked on a scale of -3 to 3, with 0 being neutral, whether application support for 25 dimensions of the operational task was respectively better in the old system or in the new system. Here, the mean ratings ranged from 1.95 to -0.03, with two rating preferences for systematic uses of sound. Both of these were close to 0. Participants showed a slight preference (0.02) for the auditory reinforcement of the system status change buttons' blink rate and a similarly slight aversion (-0.03) to the new prioritized buzzer and reminder sound scheme. Table 4 summarizes these evaluation results.

Questionnaire evaluations of implemented auditory display			
Utility of auditory display features			
	[scale: 1 (low) to 7 (high)]		
	mean rating	s.d.	s.e.
pending alert reminder sound	5.31	1.66	0.083
status change sound	5.25	1.74	0.085
toggle for old and new sounds	4.25	1.88	0.940
prioritized sounds	3.85	1.69	0.085
Preferred auditory support for operational task			
[scale: -3 (old system) to 0 (neutral) to 3 (new system)]			
	mean rating	s.d.	s.e.
status change sound	0.02	1.32	0.066
alert queue sounds	-0.03	1.66	0.083

Table 4: *Participant ratings of revised auditory display features and preferred auditory support schemes, showing the mean rating, the standard deviation, and the standard error of the mean for each element. The ranges and meanings of the rating scales employed are given in square brackets.*

6. CONCLUSIONS

Although the subjective opinions of participants in the final evaluation regarding the new uses of sound were mixed at best, this result was not unexpected. Indeed, the operational performance demands that are placed upon watchstanders in modern tactical decision environments are extraordinary, and this fact was confirmed repeatedly in other comments made in the final evaluation's exit interviews. These settings are often noisy, stressful, and high-paced, and most watchstanders have numerous additional responsibilities, some of which include the continuous monitoring of radio communications. In addition, watchstanders can hear, and be confused by, each other's buzzers, and there are numerous other critical uses of sound in shipboard environments. Given the scope of these factors, it would be unusual for any substantial change in the function of the watchstander's tools to be met with incautious or enthusiastic approval. Such changes bring with them the added burden of learning a new paradigm. However, like any job that involves acquired skill in the use of a complex system, the true measure of the auditory, interactional, and transparent enhancements of the alert queue task resulting from this effort will only come with time.

It is important to note that the application of empirical principles derived from the stimulus-response approach employed here, as well as in the work of others on which it is based, does not guarantee that performance resulting from the final auditory design will be optimal. In stimulus-rich, multitask environments, numerous contingent and meta-cognitive factors, including cognitive strategy, cultural biases, experiential history, and situational context, often undermine or limit the performance targets that can be practicably achieved. In spite of their demonstrated benefits, auditory alerts are typically viewed as annoying and unnecessary, and when they are meant not only to notify but also to convey related but separate information, such as degree of urgency and where to look in the present design, their ultimate effectiveness depends on complex perceptual, attentional, cognitive, and attitudinal processes that are difficult to reliably assess outside of longitudinal studies in context. The nature of the effort reported here and the crucial mandates of the milieu in which it was carried out are, in fact, illustrative of this point and of the challenges of getting new, empiricaly informed auditory designs into actual use in critical operational settings. As a result, it is a significant advance to place an alternative into operation, and for corresponding reasons, small incremental changes are often the best that can be achieved in this type of applied research and development.

In summary, the purpose of this paper has been to review the considerable process involved in the design and implementation of a multipurpose, urgency-encoded, audio alerting scheme in the restricted context of a legacy Navy decision support workdesk. Working within the limits imposed by the existing system's hardware, and drawing on the previous research of [5], [7], [8], and [10], this effort has shown that it is possible to design and predictably encode recognizably different levels of urgency in otherwise parametrically static alert sounds using only combined manipulations of time-based acoustic parameters. In addition, the effort addressed as yet unresolved aspects of the perceptual affect and contribution of rhythm parameter manipulations and underscored the importance of considering the effect of ecological factors on urgency encoding in auditory alerts.

This work was supported by the Office of Naval Research under work request N0001404WX20835. The authors would like to thank Vera Zaychik, Brett Breslow, Greg Trafton, Chris Corvo, Mike Thomas, Adria Hughes, Kath Stibler, Brian McClimens, and Farilee Mintz for their assistance.

7. REFERENCES

[1] J. A. Ballas, "Delivery of information through sound," in G. Kramer (Ed.) *Auditory Display: Sonification, Audification, and Auditory Interfaces*, Addison Wesley, Reading, MA, pp. 79-94, 1994.

[2] D. C. McFarlane, *Interruption of People in Human-Computer Interaction*, Doctoral Dissertation, George Washington University, Washington, DC, 1998.

[3] D. C. McFarlane, "Comparison of four primary methods for coordinating the interruption of people in human-computer interaction," Human-Computer Interaction, 17(1), pp. 63-139, 2002.

[4] D. C. McFarlane and K. A. Latorella, "The scope and importance of human interruption in human-computer interaction design," Human-Computer Interaction, 17(1), pp. 1-61, 2002.

[5] R. D. Patterson, "Guidelines for auditory warning systems on civil aircraft," CAA Paper 82017, London Civil Aviation Authority, 1982.

[6] R. D. Patterson, "Guidelines for the design of auditory warning sounds," in Proceedings of the Institute of Acoustics, 11(5), pp. 17-24, 1989

[7] J. Edworthy, S. Loxley, and I. Dennis, "Improving auditory warning design: Relationship between warning sound parameters and perceived urgency," Human Factors, 33(2), pp. 205-231, 1991.

[8] E. J. Hellier, J. Edworthy, and I. Dennis, "Improving auditory warning design: Quantifying and predicting the effects of different warning parameters on perceived urgency," Human Factors, 35(4), pp. 693-706, 1993.

[9] S. S. Stevens, "On the psychophysical law," Psychological Review, 64, pp. 153-181, 1957.

[10] A. Guillaume, L. Pellieux, V. Chastres, and C. Drake, "Judging the urgency of nonvocal auditory warning signals: Perceptual and cognitive processes," Journal of Experimental Psychology, 9(3), pp. 196-212, 2003.

[11] E. F. Clarke, "Rhythm and timing in music," in D. Deutsch (Ed.) *The Psychology of Music, 2nd Ed.*, Academic Press, San Diego, CA, pp. 473-500, 1999.

[12] T. D. Mason, *The Art of Hearing: Aural Skills for Improvisers*, Hal Leonard, Milwaukee, WI, 1997.