2018-06-01

Acoustic features of auditory medical alarms - an experimental study of alarm volume

Schlesinger,, JJ

http://hdl.handle.net/10026.1/11630

10.1121/1.5043396
Journal of the Acoustical Society of America
Acoustical Society of America

Acoustic features of auditory medical alarms – An experimental study of alarm volume

Joseph J. Schlesinger, Department of Anesthesiology, Vanderbilt University School of Medicine, 1211 21st Ave South, Medical Arts Building, Suite 422, Nashville, TN, 37212

Sarah H. Baum, Department of Psychology, University of Washington, Seattle, WA, Main Office Mailbox 357270, T-471 Health Sciences Center, Seattle, Washington, 98195

Katherine Nash, Department of Neuroscience, Vanderbilt University, 2201 West End Ave, Nashville, TN, 37212

Marissa Bruce, Department of Molecular Cell and Biology, Vanderbilt University, 2201 West End Ave, Nashville, TN, 37212

Daniel Ashmead, Department of Hearing and Speech Sciences, Vanderbilt University, 1211 21st Avenue South, Medical Center East, South Tower, Nashville, TN, 37212

Matthew S. Shotwell, Department of Biostatistics, Vanderbilt University, 2525 West End Ave, Nashville, TN, 37212

Judy R. Edworthy, Department of Psychology, Plymouth University, Drake Circus, Plymouth PL4 8AA, UK

Mark T. Wallace, Department of Neuroscience, Vanderbilt University, 465 21st Avenue South, Nashville, TN, 37232

Matthew B. Weinger, Department of Anesthesiology, Vanderbilt University, 1211 21st Avenue South, Medical Arts Building, Suite 706, Nashville, TN, 37212

ABSTRACT

Audible alarms are a ubiquitous feature of all high-paced, high-risk domains such as aviation and nuclear power where operators control complex systems. In such settings, a missed alarm can have disastrous consequences. It is conventional wisdom that for alarms to be heard, “louder is better,” so that alarm levels in operational environments routinely exceeds ambient noise levels. Through a robust experimental paradigm in an anechoic environment to study human response to audible alerting stimuli in a cognitively demanding setting, akin to high-tempo and high-risk domains, clinician participants responded to patient crises while concurrently completing an auditory speech intelligibility and visual vigilance distracting task as the level of alarms were varied as a signal-to-noise ratio above and below hospital background noise. There was little difference in performance on the primary task when the alarm sound was -11 dB *below* background noise as compared with +4 dB *above* background noise – a typical real-world situation. Concurrent presentation of the secondary auditory speech intelligibility task significantly degraded performance. Operator performance can be maintained with alarms that are *softer* than background noise. These findings have widespread implications for the design and implementation of alarms across all high-consequence settings.

I. INTRODUCTION

Alarms that draw attention to dangerous situations are prominent in all high consequence industries: aviation, ground transportation, nuclear power, healthcare, etc.[1,2] In high-tempo, high-risk, safety-critical situations where a few operators are responsible for controlling complex systems, a missed signal or alarm can cost human lives.[3] In healthcare, for example, a four-month 2010 review of the US Food and Drug Administration (FDA) Manufacturer and User Facility Device Experience (MAUDE) database found 73 alarm-related deaths and a ten-year (2004-2014) review revealed 844 injuries related to alarm mismanagement, catapulting these issues to high-profile media attention.[4]

Given the importance of alarms, it is not surprising that they are ubiquitous and used liberally, and the ‘better-safe-than-sorry’ approach can lead to other problems. For example, conventional beliefs and, often, guidelines on alarm signal implementation hold that alarms must be louder than background (ambient) noise levels in order to be adequately perceived. This of course is an overly simplified view. We can hear a soprano signing over a large orchestra even though she is objectively not as loud as that orchestra because of the relationship between the (relatively high) frequency components of her voice and the (generally lower) frequency components of the orchestral sound, mediated by the operation of the auditory filter.[5,6] Available guidance on the design and evaluation of auditory alarms in fact takes validated models of the auditory filter and demonstrates how the levels of the individual components of auditory alarm signals should be adjusted or designed to be within an appropriate audibility band given the background noise over which it is intended to be heard.[6,7] Thus, the audibility of an alarm sound doesn’t depend just on the overall background noise level, but the spectrum of the background noise and its relationship to the spectrum of the alarm signal. However, practice has not typically followed in that a) many auditory signals still in use do not possess many frequency components, and may possess only one or two which are much louder than the others, on which its entire audibility relies and b) the take-home message of the earlier, detailed work (that alarms should overall be louder than their background noise, by a considerable margin) leads to alarms which are too loud, by virtue of point a.

That alarms might be audible when they are overall lower in loudness than their background noise is further suggested by the idea of stochastic resonance,[7] whereby the presence of noise enhances the perception of a sensory signal. For sounds heard in noise, the effects are usually most pronounced when the noise level is lower than the signal to be detected, so may not be relevant when we are considering weak signals in the presence of noise which is louder than the signal. However, the possibility of stochastic resonance playing a role might also be considered in this context.[8]

This ubiquitous but untested assumption regarding alarm volume relative to background noise has created a vicious cycle of increasing sound intensity, particularly in the less well controlled sound environments, resulting in increased alarm-related incidents.[9] Alarm fatigue, another aspect of alarms and alarm signals which is often talked about but is not well understood in terms of its components, is generally conceived of as desensitization to alarms resulting from the number of audible alarms and associated noise load.[10-13] Alarm fatigue is also believed to be a factor in many missed or delayed responses.[11] Further, the increased noise from numerous alarms can also increase operator stress, hamper decision-making,[10,14] predispose to miscommunication,[15] and may have negative health effects including hearing damage and even cardiovascular morbidity from chronic increases in sympathetic tone.[16]

More basically, we know that unnecessary noise increases stress and reduces performance,[17] so any measures which can reduce noise levels will be beneficial.

Based on human research in auditory and multisensory (i.e., multiple sensory systems) perception, we hypothesized that the operational dictum that ‘louder is better’ is wrong.[18] More precisely, we address the issue as to how loud an auditory alarm signal actually needs to be in order to be detected relative to background noise, in order that the relationship between the alarm and the background noise can be reconceptualised. For example, what (misinterpreted) guidelines might suggest is that alarms thought to be inaudible might actually be audible, and if set at the level suggested in the (misinterpreted) guidance, they might be so loud as to start interfering with performance. Thus the issue is one of calibration. To test this hypothesis, we created an experimental paradigm in an anechoic chamber to evaluate audible alarms in a simulation of a relevant operational environment – the management of acute changes in cardiovascular physiology during patient care.[19] We presented domain knowledgeable participants with a primary task of making appropriate treatment decisions in response to alarms triggered by changes in ‘vital signs’ presented on a video display. The participants concurrently performed two secondary tasks designed to tax their attentional and decisional resources; the coordinated response measure (CRM), a well-validated speech recognition task[20] and a randomly presented visual vigilance task. Changes in auditory alarm signal intensity ranged from negative to positive signal-to-noise ratios ([SNRs] – the ratio of the strength of a signal carrying information to that of interference), relative to the ambient noise level fixed at 60 dB. The primary outcomes were alarm response time, treatment selection accuracy, and a composite performance measure, the inverse efficiency score (IES), which is a calculated ratio of response time and accuracy.[21]

A. The alarm tested

In this study we use a single auditory alarm signal, the high acuity (red) alarm from a Philips MP-70 patient monitor. The spectrum of the alarm is shown in **Figure 1**. Figure 1 shows the alarm against a typical background noise level of 60dB(A), as used in this study, with three different Signal-to-Noise ratios when measuring the overall loudness of the noise and the alarm (rms). It can be seen that most of the frequency components of the alarm are well below the noise level, but that there are two components at about 980 Hz and 2881 Hz which dominate the sound (and will be the only audible components of the sound in any reasonable amount of noise). Thus the audibility of the alarm depends entirely on these two components. At an SNR (alarm-to-noise) of +4dB, the alarm should be highly audible, possibly too loud. At -11dB(A), the spectral comparisons suggest that the alarm should still be audible. At -27db(A), the alarm should be inaudible.

II. METHODS

A. Overview

Using a model of monitoring performance, we developed a paradigm that tasks participants with treating clinical scenarios while also completing domain-relevant auditory and visual tasks to address speech intelligibility and vigilance, respectively. Each perilous situation is associated with an alarm of varying loudness relative to normalized (application of constant amount of gain to bring the peak amplitude to a consistent target level) discipline-relevant hospital background noise.

B. Participants

The study paradigm was refined using 14 attending anesthesiologist participants (10 men and 4 women, 31 to 51 years old) who gave written informed consent as approved by our

Institutional Review Board. Study participants were a different cohort of 31 consenting anesthesiologists, 1 faculty physician (56 years of age) and 30 residents in training 26 to 30 years old, comprising 20 men and 11 women. All study participants had a near-threshold of hearing the alarm from 30 to 38 dB (-30 to -22 dB SNR from ambient background noise).

C. Apparatus

Testing took place in an anechoic chamber (4·65 x 6·55 x 7·47 m tall, wire mesh floor 1·7 m from bottom). Each participant sat at the center of a circular array of 64 equally spaced loudspeakers (Meyer Sound MM-4), at ear level and 1·95 m from the center. The participant faced a central yellow light emitting diode (LED), positioned directly under one of the loudspeakers (**Figure 2**). Two additional loudspeakers (JBL 8110) were positioned just above the full loudspeaker array, 15° left and 60° right of the LED. An 18-in color monitor was located 30° to the right of the LED just below the loudspeakers. Sessions were controlled by custom software written in MATLAB (MathWorks, R2015a) on a Dell PC running Windows 7. Sound was generated with two Tucker-Davis Technologies (TDT) RP2 processors linked with 4 TDT PM2 16-channel multiplexers. Additional sound output was through TDT audio components and Crown amplifiers (D-75 and XLS1000). Conveniently mounted on the arm of the chair at which each participant sat was a customized computer keyboard that supported the specific study tasks as described in the next section.

D. Tasks

Throughout the session, pre-recorded discipline-relevant background noise from an intensive care unit was played continuously through the two JBL auxiliary loudspeakers and two ring speakers (located 15° and 105° left, and 60° and 150° right of the participant's facing direction). A sound level of 60 dB was chosen based on the literature,[10,14,22] preliminary acoustic measurements in our actual intensive care units, and pilot testing in the anechoic chamber. An average spectrum of 29 seconds of the noise is shown in **Figure 1**, sampled at 24,414 Hz. There was little variation in that spectrum over time, and it averaged around 60dB(A) for the duration of the study.

The participant was instructed to engage in three interleaved or concurrent tasks (**Figure 2**). The primary task was monitoring for emergency events. The two secondary tasks were responding to spoken phrases from the validated Coordinate Response Measure (CRM)[20] *and* responding to the illumination of a yellow LED, which was construed as an unspecified but critical signal. The alarm monitoring task simulated a typical domain-relevant patient monitoring system, with the visual display showing physiological values (diastolic and systolic blood pressure, heart rate, respiration rate, and blood oxygen level, refreshed at 1 Hz) representing the different 'patient' conditions. The vital signs varied randomly within the 'normal' range except during the four simulated emergency events – isolated sinus bradycardia (low heart rate), isolated tachycardia (high heart rate), tachycardia with hypotension (low blood pressure), and bradycardia with hypotension (**Table 1**). Each event was accompanied by the same audible alarm signal, projected from a loudspeaker directly above the visual display, presented at different variable sound levels, calculated as different signal-to-noise ratios (SNRs) compared to the ongoing hospital background noise. The participant was expected to 'treat' each emergency event using one of four labelled keys on the keyboard, corresponding to different appropriate drug choices (atropine, esmolol, phenylephrine, and ephedrine, respectively).

The CRM task measures speech intelligibility ability, which is the foundation for communicating in any complex multi-member team-based endeavour. The CRM corpus has gained broad acceptance as a research tool for investigating speech intelligibility in background competition and has been widely used in studies of informational masking.[23] The CRM has been used in studies in which speech is masked by speech because the format of the speech materials allows the listener to lock onto a target phrase signified by its call sign even when competing sentences from the same corpus are presented simultaneously.[20,23] Choosing two or more talkers results in speech-shaped noise versus a single-talker interferer yielding a non-monotonic psychometric function.[23] We constructed the salient features of our CRM for this study based on previous work by Eddins et al[23] and Bolia et al[20] as well as parallels to real-life clinical scenarios. The clinical correlation, besides the clinical task, was that speech intelligibility is paramount in clinical situations in emergency and non-emergency situations in practice locations such as the operating room and intensive care unit.[15] Given that, we chose to follow the two-talker CRM paradigm in previous work[23] instead of four-talker babble or cafeteria noise (remembering we utilized background hospital noise to simulate our intended environment). The CRM task consists of pre-recorded spoken sentences with the carrier phrase, “Ready [call sign], go to [color] [number] now” (e.g. “Ready Baron, go to blue eight now”). In this study, three phrases were presented concurrently from a single loudspeaker located behind the participant, each with different call signs, colors, and numbers spoken by three different males. The full CRM phrase set consists of 256 combinations of eight call signs, four colors, and the number 1 through 8. The sound level of each CRM phrase was 0 dB relative to the background hospital noise level. One of the three enunciated phrases always used the call sign “Baron.” The participant was instructed to report the color and number of the phrase that had that designated call sign using clearly marked buttons on the keyboard. The participant received visual feedback via brief flashes of centrally located LED’s – green for correct selection of both color and number or amber if either were incorrect.

The visual distraction task was structured as a classical vigilance response task in which the participant was to press a designated key press whenever the LED light went *off*. Once lit, the yellow LED remained on for a variable time (M = 5s, SD=2s, minimum 1s). After turning off, it remained off until the participant pressed the key. To discourage participants from tapping the vigilance key randomly in an effort to keep it on, the LED was switched off if the key was pressed while the LED was on. One purpose of this task was to prevent participants from directing their visual attention continuously to the vital signs monitor, better emulating real-world clinical conditions.

E. Study conditions

For each participant, all twenty combinations of four types of emergency events (specified above) and five alarm levels were presented. We presented auditory alarms at each participant's individual threshold (between -30dB and -21dB), -20 dB, -11 dB, -2 dB, and +4 dB) relative to hospital background noise at 60 dB.

There were ten trials per condition. The 200 trials were sequenced in 10 blocks of 20 trials each; with each block containing a random ordering of the 20 combinations of four emergency types and five alarms levels. Although the session was structured by "trials," these were connected seamlessly so that participants experienced a single running event sequence lasting approximately 70 minutes, with breaks offered every 15 minutes. Trial duration averaged 20s (SD=5s, range 12s to 28s), during which there were two CRM presentations and one emergency event. All emergency events lasted 6s, and were constrained to begin at least 2s after trial onset and end at least 2s before trial offset. CRM presentations began at least 1s after trial onset and ended 1s before trial offset. At least 2·5s elapsed from the end of the first CRM presentation to the start of the next. Aside from these constraints, the timing of the emergency events and CRM presentations was random. The LED vigilance light was on at the very beginning of the study, and then went on or off as described above, without regard for trial boundaries. Therefore, CRM presentations and LED vigilance events overlapped with some but not all emergency events.

III. DATA ANALYSIS

A. Primary task – Alarm monitoring and treatment selection

For each participant, and for each combination of alarm condition and co-occurrence of a CRM and vigilance task, inverse efficiency scores (IES) were calculated as the ratio of the average response time relative to the fraction of correctly addressed primary tasks. Lower IES scores represent better performance. The inverse efficiency score was initially introduced as a measure of approximate number system (ANS) acuity, and is calculated by diving the mean response time (RT) of correct responses by the proportion of correct responses.[24] The IES is used primarily to account for a speed-accuracy trade-off (e.g., accuracy can often be improved at the expense of response time and vice versa). In addition, the IES has an intuitive interpretation as the average amount of time required to achieve a correct response in a sequence of consecutive trials (i.e. the smaller the efficiency score, the higher the ANS acuity).[25,26]

Linear and logistic mixed effects regression analysis were used to quantify the effect of alarm signal-to-noise ratio (SNR) on the amount of time required to respond to a clinical task (alarm event), the odds of selecting a correct response to the clinical event presented, and the corresponding IES, adjusting for the possibility of concurrent CRM or vigilance task distractions. Due to skew in the distribution of response time and IES, these variables were log-transformed prior to regression analysis. A three-knot natural spline was used to model the effect of alarm SNR. The "no association" null hypothesis regarding alarm SNR was assessed using a Wald-type chunk test. Pairwise interactions between alarm SNR and the co-occurrence of either distracting task (CRM, vigilance) were also considered. Interactions were retained when there was strong evidence, as determined by a likelihood ratio test. A random intercept indexed by study subject was used to account for heterogeneity among participants (e.g., some participants were consistently quicker to respond than others, regardless of alarm level). The effects of alarm SNR were summarized using pointwise bootstrap (normal approximation) 95% confidence bands, and stratified by CRM co-occurrence. The effects of CRM or vigilance task co-occurrence were summarized using Wald-type 95% confidence intervals and tests for the mean difference and odds ratio (OR). P-values less than 0·05 were considered statistically significant. Thus, the type-I error rate was preserved at 5%. We did not attempt to control the familywise type-I error probability using a multiple comparisons procedure.

B. Secondary task – CRM

Linear and logistic mixed effects regression analysis were used to quantify the effect of alarm SNR on the average amount of time required to respond to a CRM task and the odds of selecting the correct response, adjusting for the concurrence of an unaddressed clinical task (i.e. with alarm), alarm SNR (if alarm was concurrent), and concurrence of a vigilance task. No interactions were considered. These analyses were otherwise treated similarly to those for the clinical task.

IV. RESULTS

A. Primary task – Alarm monitoring and treatment selection

Among the 31 study participants, 25 completed all 200 trials, 3 completed 160, and 1 each completed 180, 140, and 120 trials. Across all trials, 51% of alarm tasks were addressed without interruption by a CRM or vigilance task, 24% were interrupted by CRM task, 15% by a vigilance task, and 10% by both a CRM and vigilance task. **Table 2** summarizes participant performance on the alarm monitoring task, stratified by alarm SNR. The associations between alarm SNR and primary task performance were statistically significant – both the accuracy *and* speed of the treatment choices, and the corresponding inverse efficiency score, were significantly improved at sound levels greater than the near-threshold of hearing (**Table 3**). **Figure 3** illustrates the associations between alarm SNR and the primary task IES. This shows that there was little difference in performance on the primary task when the alarm sound was -11 dB *below* background noise as compared with +4 dB *above* background noise. Specifically, the estimated probability of correctly addressing the primary task when there was no concurrent distracting task was only 0.7% smaller at -11 dB relative to +4 dB (risk ratio 95% CI: 0·98, 1·02). Likewise, the estimated mean response time was just 0·04 s longer (95% CI: -0·03, 0·12), and the estimated mean IES was 0·04 s longer at -11 dB versus +4 dB (95% CI: -0·04, 0·12), when there was no concurrent distracting task. Thus, provided the alarm signal was audible (as **Figure 1** suggests it would be at -11dB), performance was no further enhanced by increasing its loudness.

Concurrent presentation of the secondary auditory CRM task significantly degraded performance. The odds of correctly addressing the primary task were decreased by 29% (95% CI: 16, 39), mean response time was slower (0·79 s, 95% CI: 0·73 s, 0·84 s), and mean IES was longer (0·30 s, 95% CI: 0·24 s, 0·35 s). In contrast, concurrent presentation of the visual secondary task (a visual vigilance task) did not significantly affect primary task performance, nor were there any significant interactions.

B. Secondary Task - CRM

The likelihood of correctly addressing the CRM task was not significantly decreased when there was a concurrent secondary vigilance task (OR: 0·93; 95% CI: 0·85,1·03). However, alarm loudness (when there was a concurrent primary task) significantly affected the likelihood of correctly addressing the CRM task (p = 0·002; **Table 3**). **Figure 4** illustrates the association between CRM task accuracy with concurrent clinical task alarm level. The positive alarm SNR (i.e., conventional levels) condition was associated with the *poorest* CRM performance under these conditions. In addition, the estimated probability of correctly addressing the CRM task was 12% greater at -11 dB versus +4 dB (95% CI: 2, 24). Thus, the higher, positive alarm SNR was associated with poorer performance relative to lower SNRs, suggesting that the higher alarm loudness level impeded, rather than helped, performance on the CRM task.

Neither the co-occurrence of a primary task nor the associated alarm SNR were significantly associated with CRM task response time (**Figure 5**). The co-occurrence of a vigilance task did not significantly affect CRM response time.

V. DISCUSSION

A. Acoustics and alarm design

We describe a new experimental paradigm modeled on the types of tasks an anesthesiologist might be expected to perform while monitoring auditory alarm signals to study the effects of alarms on human performance. Primarily, we question the typical approach and understanding of the signal-to-noise ratios of auditory alarm signals and background noise and how the levels of auditory alarms should be set. Using medical alarms and the performance of anesthesiologists as a model, the results of this study demonstrate that auditory alarms do not need to be louder overall than background sound levels to elicit accurate and reliable responses. Specifically, both response time and accuracy of the treatment selection to an abnormal clinical condition was preserved from 4 dB *louder* to 11 dB *softer* than the 60 dB of background noise. The presence of the secondary auditory (CRM) task adversely affected both response time and response accuracy, but appeared to do so in a comparable manner across SNR levels. Further, CRM task accuracy degraded when alarm sounds were +4 dB above background levels suggesting an interfering effect on the speech perception task at alarm sound levels typical of real-world conditions.

Giving context of our CRM results to other work, Eddins demonstrated that performance on this paradigm at 0 dB SNR yielded about a 55% correct response rate. The addition of feedback in our paradigm was to ensure attentional allocation and drive to perform in our competitive cohort of clinicians.[27] Our alarm stimulus, not interfering with human speech, would not appreciably mask the target talker.[28] Thus, our slightly higher performance data and approach is informed by previous work utilizing the CRM, the nature of our paradigm, and pilot work showing that approximately 65% performance with feedback strikes a balance of attentional allocation with the desire to improve without hitting a performance a ceiling or conversely eliciting frustration and burnout from the task.[29]

As **Figure 1** suggests, the auditory alarm signal was audible at -11dB SNR, as one of the frequency components was still well above threshold at that frequency. Below this SNR, both components became inaudible. At an SNR of +4dB, the most prominent component of the alarm was about 30dB above threshold at that frequency, which according to Patterson (1982) would be so loud as to interfere with task performance. This does seem to be the case here, where the secondary CRM task was impeded at this higher level (though the effect may also be partially due to masking by the alarm signal). The results suggest that the solution to setting this particular alarm at an appropriate level would be re-calibrate the relationship between the alarm signal and the background noise so that audibility is deemed to have started at an SNR of -11dB, and that a positive SNR is simply too loud.

The findings for this study are to some extent specific to the alarm tested because it has a particular spectrum and will thus represent a specific relationship between the noise background and the components of that alarm sound. However, the alarm used is fairly typical of alarms often used in medical equipment in that it relies on one or two relatively high frequency, loud components for its audibility, rather than a balanced spectrum with more, but more appropriate, components. Other alarms with different spectra may produce slightly different results depending on how the energy of the sound is distributed across its spectrum. This is a topic that could be modeled or tested in future studies. Nevertheless, what this alarm demonstrates is the gaping mismatch between evidence-based recommendations about how the spectrum of an alarm sound should be designed and set in relation to possible background noise scenarios. In practice, our findings suggest that alarms can be set at the minimum or near-minimum audible level, which can be determined through a simple listening test.

Our data also demonstrate that ‘louder is not better’. For the primary task, provided the alarm is audible, there is no benefit to increasing audibility to performance on the primary task. Thus, auditory alarms can be set at minimum levels of audibility with no detriment to performance. Indeed, the results of the secondary task performance suggest that louder is worse, as performance on the secondary task declines as alarm audibility is increased – though whether this is a direct masking effect or is some function of the participants being overloaded by the tasks, as it is well known that increased noise reduces performance in high-workload scenarios.

This effect of alarms at typical volumes on other auditory tasks may be due to divided auditory attention and/or auditory masking.[28] In high-consequence industries, where team communication can be paramount, worsening of speech perception and errors of interpretation may lead to deleterious consequences.

B. Clinical correlates of auditory medical alarms

In 1859, Florence Nightingale wrote that, “Unnecessary noise, then, is the most cruel absence of care which can be inflicted either on sick or well.”[30] Indeed, she knew that loud and uninformative sounds can be maladaptive for patients and clinicians. Medical intervention is necessary to improve patient health, but so is “therapeutic neglect” – letting patients rest is part of the recovery process. Perceived sound loudness and measured sound loudness are not equivalent, White et al describe that nurses perceive noise to be 14.1 dB higher than the actual noise level at the nursing station, and 9.3 dB greater than the noise between patient rooms.[31] As there is a difference in perceived and measured sound, initial work completed by Buxton and colleagues parsed the sound sources contributing to the overall sound level exposure, measuring sleep via polysomnography in a sleep laboratory. Buxton found that alarms at 70 dBA caused arousal in 100% of subjects in non-rapid eye movement (NREM) stage N2 sleep, and conversational speech produced a 50% arousal rate at just 50 dBA in both N2 and REM sleep.[32,33] Besides the measurable aspects of patient care, such as sleep, patient perception of care is also crucial. In a survey of ICU patients, 40% recalled ICU noise and 85% reported being disturbed by it.[34] Attenuating the disturbing aspects of the acoustic environment, interventions to create a softer acoustic environment may create a restorative period – defined as a minimum of five minutes with the maximum noise level over a period of time limited to 55 dB and the raw noise (the minute-to-minute peak values reached by sound pressure levels) limited to 75 dB.[35] Quiet time creates a restorative period, a period when sound is at a level less likely to cause arousal. But does trying to increase currently modifiable sources of noise truly help?

Recommendations for quiet time have existed for over 20 years.[36] Since patient monitor alarms cannot yet be turned down (only silenced), these interventions typically include restricting or limiting visitors, staff movement, treatments, closing doors or curtains, and decreasing noise and light. Gardner et al found that quiet time led to a 10.3 dB difference between units and improved sleep;[37] however, sound levels quickly returned to baseline within 30 minutes of the conclusion or quiet time. Although this is encouraging, the quiet time interventions still did not achieve the WHO noise recommendations.[38] A missing piece of the puzzle is the newfound knowledge presented in this study, it *is* safe to turn down the alarms and achieve the recommended WHO noise recommendations.

The ICU nurse typically manages two critically ill patients, with attention allocation split between two patients and two patient rooms. The higher acuity patient may have more monitoring devices and more alarms. Lawson et al found that turning up the alarms on the infusion pumps simply increased the sound level exposure in that patient's room, but not in the adjacent room where the nurse may be attending to his other patient.[39] The use of earplugs to diminish the effect of alarms on the patient may improve sleep, but the potential detrimental effects of decreasing all environmental auditory stimuli on neuropsychological outcomes such as ICU delirium has not been elucidated.[40]

Improving the alarmscape in the ICU will likely improve patient sleep, but sleep is not the only marker of improvement for patients and clinicians – as utilizing sleep as a primary outcome (as observed in clinical medicine, outside of a sleep lab) is fraught with using different evaluation methods, and over/under estimation of the quality of sleep.[32,41] However, Sveinsson et al found that sleep deprivation is a potential precipitating factor for delirium in cardiac surgical patients,[42] and Helton et al found that patients with sleep deprivation were significantly more likely to develop delirium than patients without sleep deprivation.[43] There is a feedforward mechanism between sleep deprivation and delirium admixed with ICU environment factors (noise, light, circadian disruption, patient care activities, stress and sensory deprivation), stress response (critical illness, mechanical ventilation, pain, sepsis), and direct effects on the brain (medications, dementia, sepsis, head trauma, advanced age, alcoholism).[44] It is no longer good enough to discharge patients from the ICU alive, we must be mindful of neuropsychological outcomes such as ICU delirium and PTSD anchored to critical illness, and what we can do to modify and ameliorate those negative outcomes.[45] A meta-analysis from 1997-2012 shows the prevalence of acute psychological risk factors for PTSD range from 8-27%.[46] Clinical risk factors include use of benzodiazepines, duration of sedation, and mechanical ventilation.[47-50] Psychological risk factors include stress and fear experienced acutely in the ICU, and frightening memories of the admission.[46] As described earlier, Hofhuis et al exhibited that patients remember and are disturbed by the ICU alarmscape.[34]

The work presented herein shows that alarm volume should be dynamic, it is safe to *turn down the volume* to improve patient safety. Sound is a complex signal and the acoustic features of sound must be dissected and studied before coalescing into an auditory unisensory stream and then with multisensory information.[51,52] Through a rigorous approach based in neuroscience and human factors applications, this work serves as a foundation to improve alarm design and patient care.

C. Study limitations

Using the anechoic chamber, we studied auditory signals in a controlled acoustic setting. However, as in any controlled study, the experimental conditions do not capture the full complexity and challenges of the real-world environment.[53] Nonetheless, this approach allowed us to efficiently conduct a prospective randomized controlled human experiment that would not have been possible otherwise. This experimental paradigm may provide generalizable data about human responses to auditory signals; the paradigm could be easily modified to study skilled operators in other high-consequence industries. All our participants had normal hearing acuity. Individuals working in healthcare and industrial settings may have mild hearing impairment.

The participants knew that the study was about auditory medical alarms but they were unaware that the hypotheses centered on alarm volume. Our participants were anesthesiologist physicians in training who all had at least one full year of residency training, which included appreciable prior exposure to the expected tasks. Although they had variable levels of clinical experience, there was no effect of years of training on observed performance. Participants offered feedback that the clinical task was as realistic as clinical tasks in high-fidelity medical simulation – an environment used for medical education, human factors engineering research, and maintenance of medical board certification. The apparent ceiling effect of clinical task performance secondary to physicians possessing a reflexive and appropriate response to treating physiologic aberration requires the perception of a salient alarm. However, super-threshold auditory presentation is not needed for performance, and may contribute to fatigue and decrease the ability to benefit from multisensory input.

D. Implications for design of next generation alarms

Understanding the optimal signal-to-noise ratios for auditory signal detection is critical to future alarm design and implementation. While the poor positive predictive value of alarms remains a problem in healthcare,[54] regardless of the domain, auditory (and other sensory) signals need to be detected and informative while ameliorating operator fatigue and maintaining vigilance.[10] Further directions for auditory information delivery will include personalized auditory devices and ambisonics, a full sphere surround sound technique. Personalized audio devices (e.g. museum audio tours) create independent sound zones, which could be useful in many jobs (e.g., anesthesiology, military command-control, etc.) as well as in everyday situations.[55] Controlled experiments using paradigms like ours will be essential to understanding which auditory signal processing approaches will optimize human performance under different conditions.

VI. CONCLUSION

The key finding of this study was that primary task performance was maintained even when alarm volume was noticeably lower than background sound levels. Our results suggest that it may be safe to decrease alarm volumes in operational settings. Sound is a complex signal and past problems with auditory alarms are partially attributable to the inability to appreciate this complexity.[22,56] This study provides new experimental evidence to inform alarm management strategies to optimize the design of auditory alarms, particularly in high-tempo, high-consequence situations. This approach serves as a foundation for parsing the sound signal, starting with the signal-to-noise ratio, to improve the use of auditory alarms across many applications to enhance human performance and health.[57]

Table 1. Vital sign ranges displayed based on the clinical event condition*

Clinical Event Condition	ECG (beats per minute)	$Sp0_2$ (%)	Respiration (breaths per minute)	Systolic blood pressure (Torr)	Diastolic blood pressure (Torr)
Normal	60-100	90-100	7-17	90-160	50-98
Isolated sinus bradycardia	*30-39*	90-100	7-17	90-160	50-98
Isolated tachycardia	*101-160*	90-100	7-17	100-160	50-98
Tachycardia with hypotension	*101-160*	90-100	7-17	*50-89*	*30-49*
Bradycardia with hypotension	*50-59*	90-100	7-17	*50-89*	*30-49*

Table Legend: Blood pressure – Systolic is the upper/larger blood pressure value and Diastolic is the smaller/lower value, measured in Torr or mm of mercury. Bradycardia – lower than normal heart rate; ECG – electrocardiogram; Hypotension – higher than normal blood pressure; SpO_2 –Oxygen saturation measured from the pulse in the finger; Tachycardia – higher than normal heart rate.

* Items in italicized font in the table were the clinically abnormal values to which the participants responded.

Table 2. Clinical task summary statistics across study participants.

Alarm Level (db)	N	Accuracy	Avg. Resp. Time (s)	IES (s)
-30	2	0.57 (0.56, 0.59) [0.55, 0.60]	3.5 (3.3, 3.6) [3.2, 3.8]	6.1 (5.7, 6.5) [5.3, 6.9]
-29	2	0.66 (0.54, 0.78) [0.41, 0.90]	2.6 (2.4, 2.9) [2.2, 3.1]	4.9 (3.7, 6.2) [2.4, 7.4]
-28	1	0.70 (0.70, 0.70) [0.70, 0.70]	4.1 (4.1, 4.1) [4.1, 4.1]	5.8 (5.8, 5.8) [5.8, 5.8]
-27	3	0.93 (0.68, 0.96) [0.44, 1.00]	2.5 (2.3, 3.2) [2.1, 3.8]	2.7 (2.4, 5.7) [2.1, 8.7]
-26	6	0.71 (0.70, 0.82) [0.62, 0.86]	2.8 (2.5, 3.3) [2.1, 3.5]	3.8 (3.1, 4.7) [3.0, 5.0]
-25	6	0.40 (0.28, 0.61) [0.20, 0.96]	3.7 (3.3, 3.9) [2.1, 4.1]	10.0 (5.8, 15.0) [2.2, 17.9]
-24	2	0.96 (0.94, 0.98) [0.93, 1.00]	2.3 (2.2, 2.4) [2.1, 2.5]	2.4 (2.2, 2.5) [2.1, 2.7]
-23	1	0.97 (0.97, 0.97) [0.97, 0.97]	2.0 (2.0, 2.0) [2.0, 2.0]	2.1 (2.1, 2.1) [2.1, 2.1]
-22	4	0.76 (0.72, 0.80) [0.64, 0.87]	2.9 (2.9, 3.2) [2.9, 4.0]	3.8 (3.6, 4.5) [3.3, 6.2]
-21	1	0.28 (0.28, 0.28) [0.28, 0.28]	3.4 (3.4, 3.4) [3.4, 3.4]	12.0 (12.0, 12.0) [12.0, 12.0]
-20	30	0.88 (0.79, 0.93) [0.30, 0.97]	2.7 (2.3, 3.2) [1.8, 3.9]	3.2 (2.5, 3.9) [1.9, 8.3]
-11	31	0.92 (0.87, 0.95) [0.57, 1.00]	2.6 (2.3, 2.9) [1.6, 3.7]	2.9 (2.4, 3.3) [1.8, 5.2]
-2	31	0.95 (0.85, 0.97) [0.55, 1.00]	2.6 (2.3, 2.8) [1.7, 3.6]	2.8 (2.3, 3.2) [1.8, 4.9]
4	31	0.92 (0.87, 0.93) [0.53, 1.00]	2.6 (2.2, 2.9) [1.4, 3.6]	2.9 (2.5, 3.3) [1.5, 5.3]

Data shown are the **median (IQR) [range]**. “N” is the number of participants observed at the corresponding “Alarm Level (db)”. For each participant, “Accuracy” was computed as the fraction of clinical task presentations that were addressed correctly.

Table 3. Regression results for clinical and CRM tasks.

	Clinical Task			CRM Task	
Description	Accuracy P-value	Response Time P-value	IES P-value	Accuracy P-value	Response Time P-value
Alarm SNR	<0.001	<0.001	<0.001	0.002	0.125
CRM Task	<0.001	<0.001	<0.001	--	--
Vigilance Task	0.505	0.404	0.101	0.166	0.079
Alarm -by- CRM Interaction	0.205	0.745	0.134	--	--
Alarm -by- Vigilance Interaction	0.554	0.808	0.232	--	--

Tests were implemented using regression methods described in the statistical analysis section.

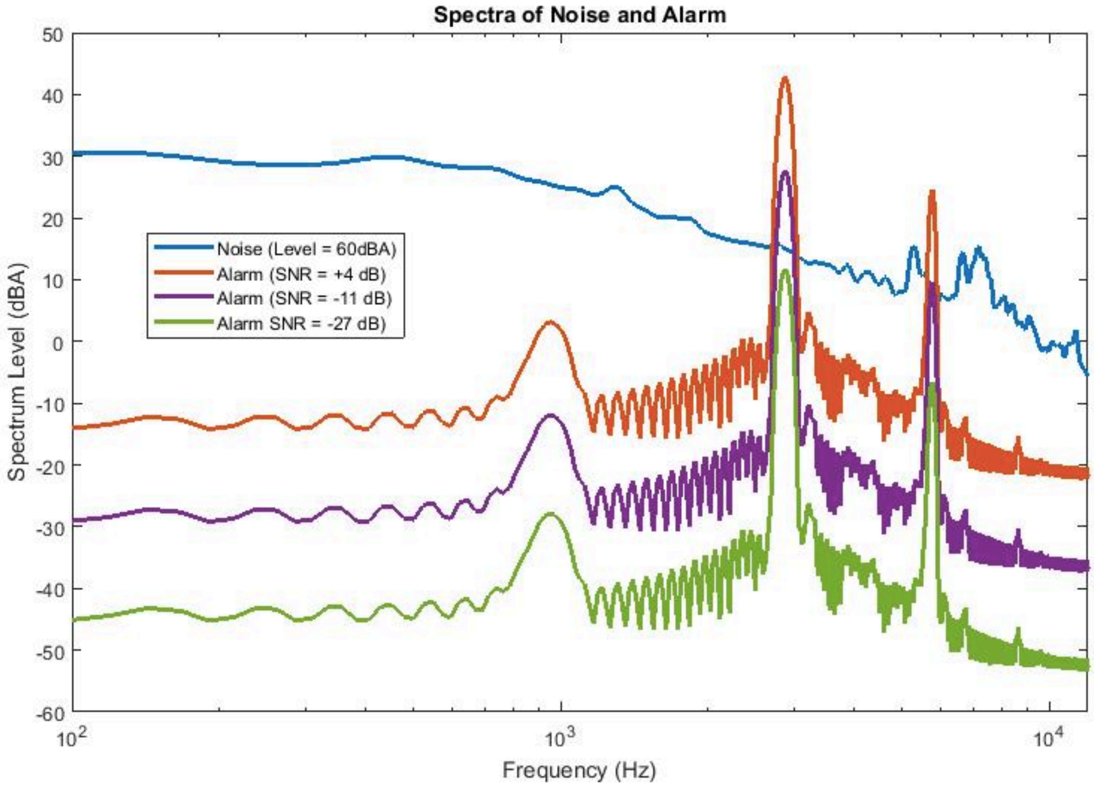

Figure 1: Spectrum of Philips MH-70 high acuity alarm at three SNR levels relative to the noise background (+4dB, -11dB and -27 dB)

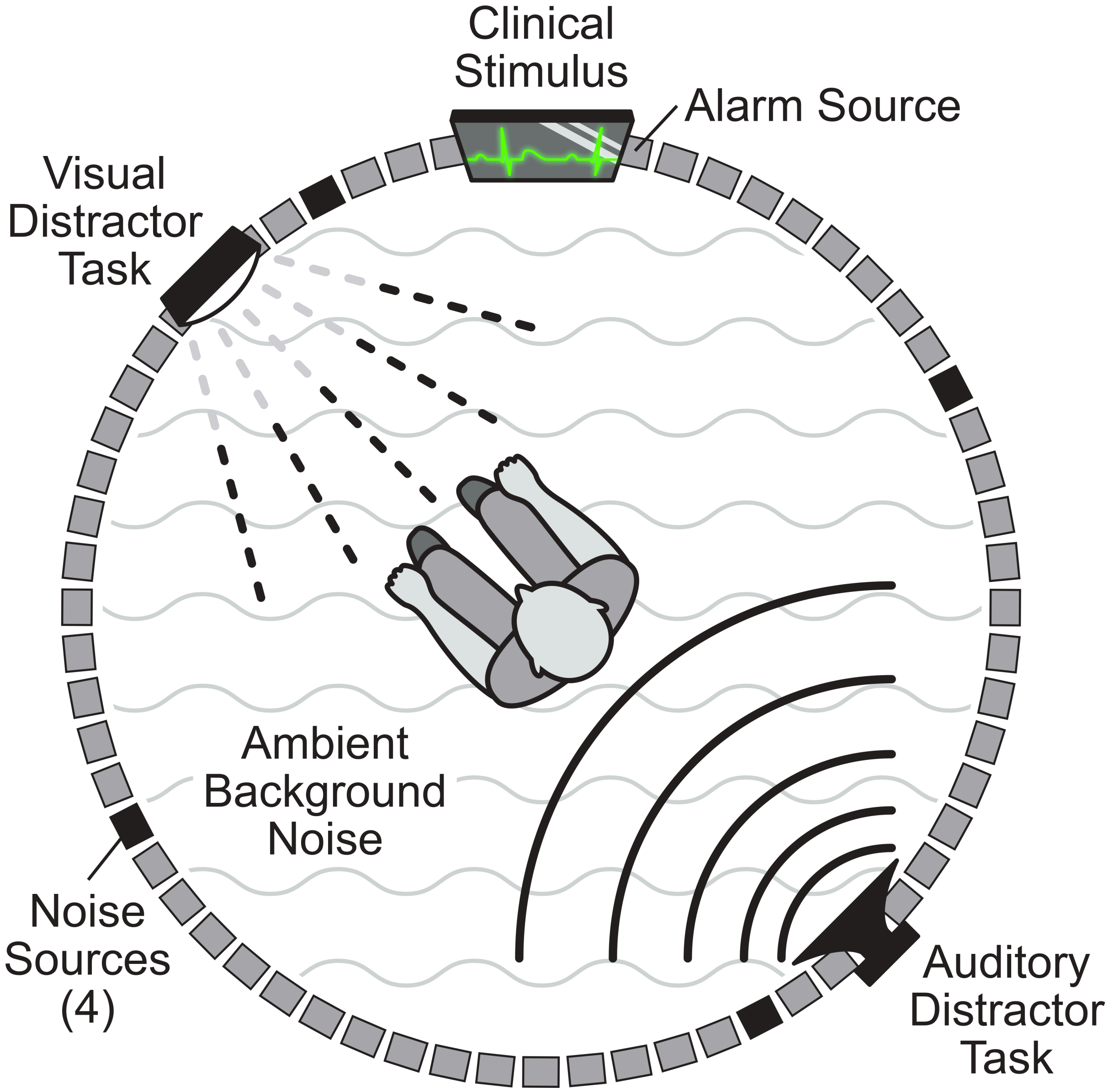

Figure 2: **Study Configuration inside the Anechoic Chamber**. The study's experimental paradigm included three interleaved tasks – the primary task was the correct treatment response based on the physiological vital signs presented on a visual display. The participants also had to respond to a visual distractor task and an auditory distractor task, the Coordinate Response Measure (CRM).[20] Pre-recorded discipline-relevant background noise was played through speakers at 60dB located 15° and 105° left, and 60° and 150° right of the participant's facing direction. There were five alarm SNRs and four types of emergency events. The participant was instructed to respond with equal urgency to all three tasks.

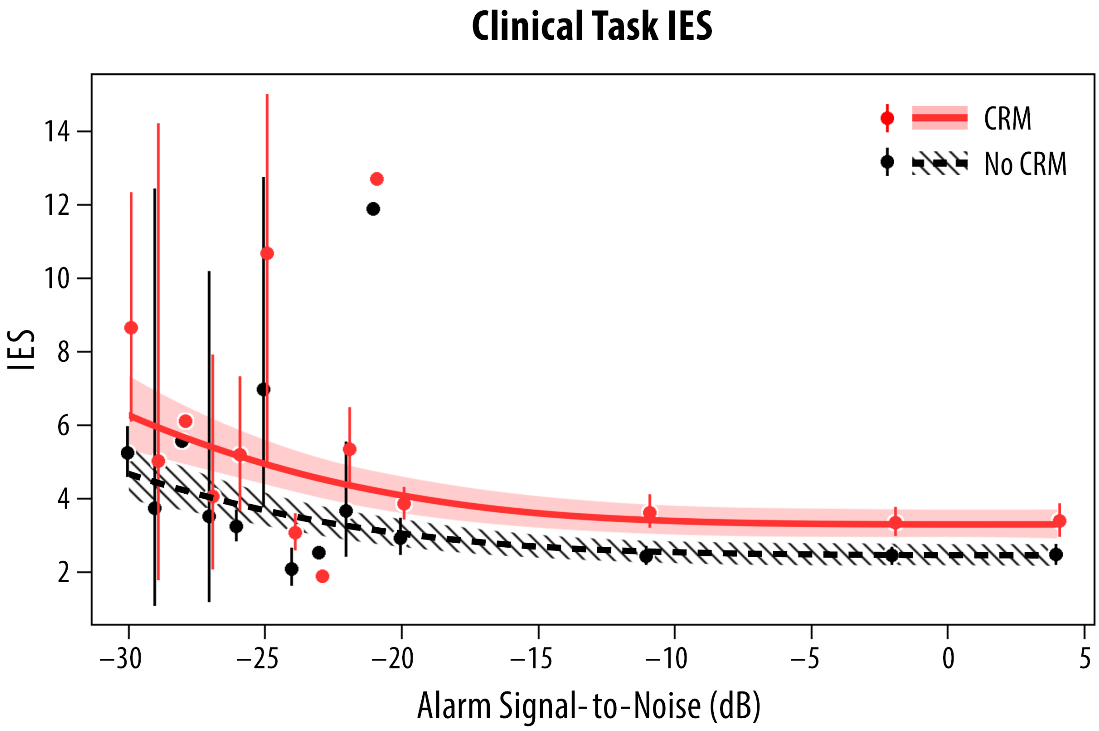

Figure 3. **Primary Task Efficiency is Preserved Down to -11 dB Below Background Noise Levels**. The model-estimated effect of alarm SNR in dB on the likelihood of correctly and rapidly treating the presented clinical event depicted as the average inverse efficiency score (IES=average response time/accuracy [lower is better]). Shaded regions represent model-based pointwise 95% confidence bands under the conditions when there was or was not a concurrent distracting auditory (CRM) task. The plotted points are raw averages of individual IES values with 95% confidence interval. The CRM task significantly degraded performance accuracy.

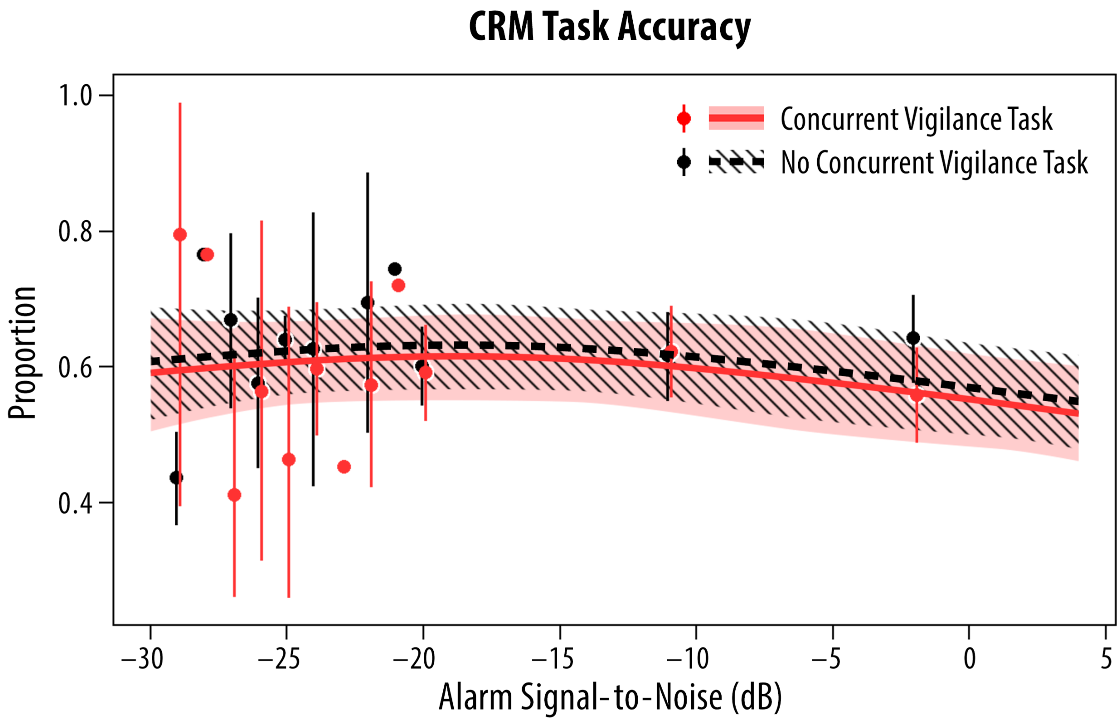

Figure 4. **Secondary Auditory Task Accuracy Deteriorated at Typical Alarm SNRs**. The model-estimated effect of alarm signal-to-noise ratio (SNR) on the likelihood of correctly responding to the Coordinate Response Measure (CRM) secondary auditory task. Shaded regions represent model-based pointwise 95% confidence bands under the conditions when there was and was not a concurrent distracting visual vigilance task. The plotted points are the raw averages of individual accuracies (i.e., correctly addressed CRM tasks) with 95% confidence interval. Secondary task accuracy deteriorated at the highest alarm SNR (i.e., 4 dB above background, which is typical of real-world situations).

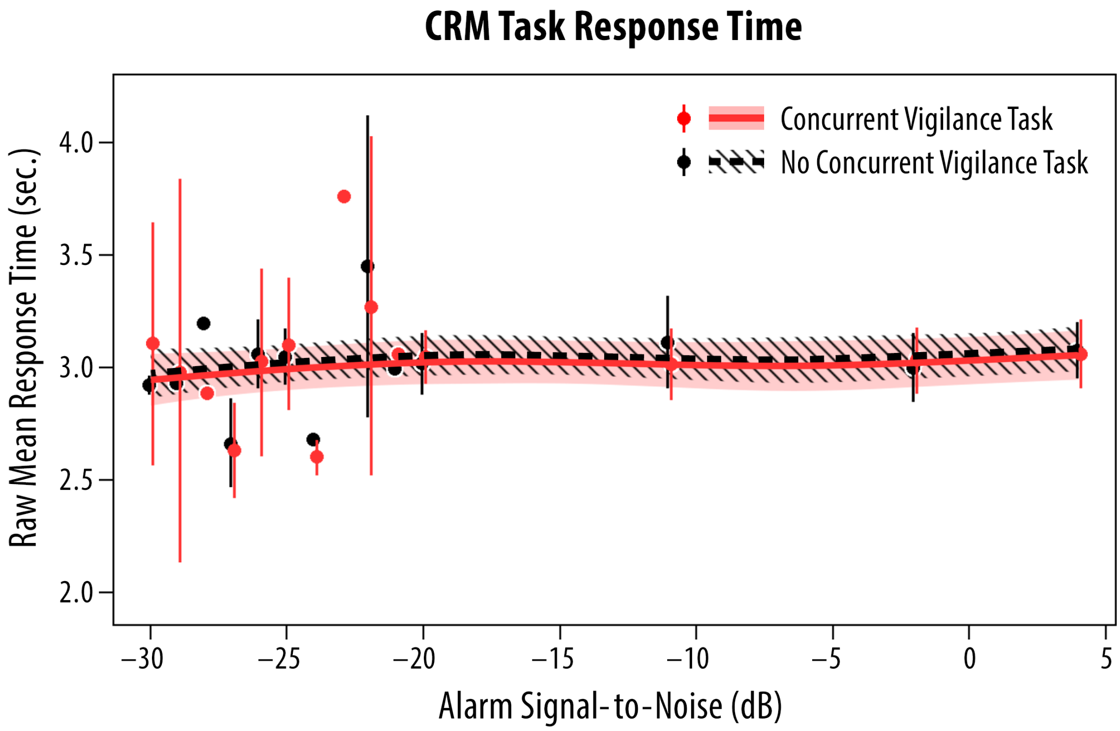

Figure 5. **Secondary Auditory Task Response Time Was Not Affected by Alarm SNR**. The alarm signal-to-noise ratio (SNR) in decibels (dB) did not significantly affect the response time to the Coordinate Response Measure (CRM) secondary auditory task. Shaded regions represent model-based pointwise 95% confidence bands for mean response time under the conditions when there was and was not a concurrent distracting visual vigilance task. The plotted points are the raw averages of individual mean response times with 95% confidence interval. The occurrence of a visual vigilance task did not significantly affect response time

Online Supplementary Picture

Anechoic Chamber Laboratory

References

1. Bliss JP, Freeland MJ, Millard JC. Alarm related incidents in aviation: a survey of the aviation safety reporting system database. Paper presented at: Proceedings of the Human Factors and Ergonomics Annual Meeting (1999)1999.
2. Dadashi N, Wilson J, Golightly D, Sharples S, Clarke T. Practical use of work analysis to support rail electrical control rooms: A case of alarm handling. *J Rail Rapid Trans.* 2012;0:1-13.
3. Lawton R, Ward NJ. A systems analysis of the Ladbroke Grove rail crash. *Accid Anal Prev.* 2005;37(2):235-244.
4. Weinger MB, Berry JM. *Anesthesia Equipment: Principles and Applications.* Philadelphia: Saunders (Elsevier); 2013.
5. Glasberg BR, Moore BC. Derivation of auditory filter shapes from notched-noise data. *Hearing research.* 1990;47(1-2):103-138.
6. Patterson RD. Auditory filter shapes derived with noise stimuli. *The Journal of the Acoustical Society of America.* 1976;59(3):640-654.
7. Lugo E, Doti R, Faubert J. Ubiquitous crossmodal Stochastic Resonance in humans: auditory noise facilitates tactile, visual and proprioceptive sensations. *PloS one.* 2008;3(8):e2860.
8. Zeng FG, Fu QJ, Morse R. Human hearing enhanced by noise. *Brain research.* 2000;869(1-2):251-255.
9. Wiczorek R, Manzey D. Supporting attention allocation in multitask environments: effects of likelihood alarm systems on trust, behavior, and performance. *Human factors.* 2014;56(7):1209-1221.

10. Cvach M. Monitor alarm fatigue: an integrative review. *Biomedical instrumentation & technology / Association for the Advancement of Medical Instrumentation.* 2012;46(4):268-277.

11. Sendelbach S, Funk M. Alarm fatigue: a patient safety concern. *AACN Adv Crit Care.* 2013;24(4):378-386; quiz 387-378.

12. Kristensen MS, Edworthy J, Ozcan E, Denham S. Alarm fatigue in the perception of medical soundscapes. Paper presented at: Proceedings of EuroNoise (2015)2015.

13. Deb S, Claudio D. Alarm fatigue and its influence on staff performance. *IIE Trans Healthc Syst Eng.* 2015;5:183-196.

14. Choiniere DB. The effects of hospital noise. *Nursing administration quarterly.* 2010;34(4):327-333.

15. Weinger MB, Englund CE. Ergonomic and human factors affecting anesthetic vigilance and monitoring performance in the operating room environment. *Anesthesiology.* 1990;73(5):995-1021.

16. van Kempen EE, Kruize H, Boshuizen HC, Ameling CB, Staatsen BA, de Hollander AE. The association between noise exposure and blood pressure and ischemic heart disease: a meta-analysis. *Environmental health perspectives.* 2002;110(3):307-317.

17. Kryter KD. The effects of noise on man. *J Speech Hear Disord Monogr Suppl.* 1950;1:1-95.

18. Ross LA, Saint-Amour D, Leavitt VM, Javitt DC, Foxe JJ. Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cerebral cortex.* 2007;17(5):1147-1153.

19. Bennett CL, Dudaryk R, Ayers AL, McNeer RR. Simulating environmental and psychological acoustic factors of the operating room. *The Journal of the Acoustical Society of America.* 2015;138(6):3855-3863.

20. Bolia RS, Nelson WT, Ericson MA, Simpson BD. A speech corpus for multitalker communications research. *The Journal of the Acoustical Society of America.* 2000;107(2):1065-1066.

21. Bruyer R, Brysbaert M. Combining Speed and Accuracy in Cognitive Psychology: Is the Inverse Efficiency Score (Ies) a Better Dependent Variable Than the Mean Reaction Time (Rt) and the Percentage of Errors (Pe)? *Psychol Belg.* 2011;51(1):5-13.

22. Edworthy J. Medical audible alarms: a review. *Journal of the American Medical Informatics Association : JAMIA.* 2013;20(3):584-589.

23. Eddins DA, Liu C. Psychometric properties of the coordinate response measure corpus with various types of background interference. *The Journal of the Acoustical Society of America.* 2012;131(2):EL177-183.

24. Dietrich JF, Huber S, Klein E, Willmes K, Pixner S, Moeller K. A Systematic Investigation of Accuracy and Response Time Based Measures Used to Index ANS Acuity. *PloS one.* 2016;11(9):e0163076.

25. Bartelet D, Vaessen A, Blomert L, Ansari D. What basic number processing measures in kindergarten explain unique variability in first-grade arithmetic proficiency? *J Exp Child Psychol.* 2014;117:12-28.

26. Sasanguie D, De Smedt B, Defever E, Reynvoet B. Association between basic numerical abilities and mathematics achievement. *Br J Dev Psychol.* 2012;30(Pt 2):344-357.

27. Dunn LB, Iglewicz A, Moutier C. A conceptual model of medical student well-being: promoting resilience and preventing burnout. *Acad Psychiatry.* 2008;32(1):44-53.

28. Hasanain B, Boyd AD, Edworthy J, Bolton ML. A formal approach to discovering simultaneous additive masking between auditory medical alarms. *Appl Ergon.* 2017;58:500-514.

29. Hyman SA, Michaels DR, Berry JM, Schildcrout JS, Mercaldo ND, Weinger MB. Risk of burnout in perioperative clinicians: a survey study and literature review. *Anesthesiology.* 2011;114(1):194-204.

30. Nightingale F. [Notes on nursing: what it is and what it is not. 1]. *Sogo Kango.* 1974;9(1):92-98.

31. White BL, Zomorodi M. Perceived and actual noise levels in critical care units. *Intensive & critical care nursing : the official journal of the British Association of Critical Care Nurses.* 2017;38:18-23.

32. Buxton OM, Ellenbogen JM, Wang W, et al. Sleep disruption due to hospital noises: a prospective evaluation. *Annals of internal medicine.* 2012;157(3):170-179.

33. Stafford A, Haverland A, Bridges E. Noise in the ICU. *Am J Nurs.* 2014;114(5):57-63.

34. Hofhuis JG, Spronk PE, van Stel HF, Schrijvers AJ, Rommes JH, Bakker J. Experiences of critically ill patients in the ICU. *Intensive & critical care nursing : the official journal of the British Association of Critical Care Nurses.* 2008;24(5):300-313.

35. Tegnestedt C, Gunther A, Reichard A, et al. Levels and sources of sound in the intensive care unit - an observational study of three room types. *Acta anaesthesiologica Scandinavica.* 2013;57(8):1041-1050.

36. Edwards GB, Schuring LM. Pilot study: validating staff nurses' observations of sleep and wake states among critically ill patients, using polysomnography. *American journal of critical care : an official publication, American Association of Critical-Care Nurses.* 1993;2(2):125-131.

37. Gardner G, Collins C, Osborne S, Henderson A, Eastwood M. Creating a therapeutic environment: a non-randomised controlled trial of a quiet time intervention for patients in acute care. *Int J Nurs Stud.* 2009;46(6):778-786.
38. Dennis CM, Lee R, Woodard EK, Szalaj JJ, Walker CA. Benefits of quiet time for neuro-intensive care patients. *J Neurosci Nurs.* 2010;42(4):217-224.
39. Lawson N, Thompson K, Saunders G, et al. Sound intensity and noise evaluation in a critical care unit. *American journal of critical care : an official publication, American Association of Critical-Care Nurses.* 2010;19(6):e88-98; quiz e99.
40. Hu RF, Jiang XY, Hegadoren KM, Zhang YH. Effects of earplugs and eye masks combined with relaxing music on sleep, melatonin and cortisol levels in ICU patients: a randomized controlled trial. *Critical care.* 2015;19:115.
41. Maidl CA, Leske JS, Garcia AE. The influence of "quiet time" for patients in critical care. *Clin Nurs Res.* 2014;23(5):544-559.
42. Sveinsson IS. Postoperative psychosis after heart surgery. *The Journal of thoracic and cardiovascular surgery.* 1975;70(4):717-726.
43. Helton MC, Gordon SH, Nunnery SL. The correlation between sleep deprivation and the intensive care unit syndrome. *Heart & lung : the journal of critical care.* 1980;9(3):464-468.
44. Weinhouse GL, Schwab RJ, Watson PL, et al. Bench-to-bedside review: delirium in ICU patients - importance of sleep deprivation. *Critical care.* 2009;13(6):234.
45. Patel MB, Jackson JC, Morandi A, et al. Incidence and Risk Factors for ICU-related Posttraumatic Stress Disorder In Veterans and Civilians. *American journal of respiratory and critical care medicine.* 2016.

46. Wade D, Hardy R, Howell D, Mythen M. Identifying clinical and acute psychological risk factors for PTSD after critical care: a systematic review. *Minerva anestesiologica.* 2013;79(8):944-963.

47. Jackson JC, Pandharipande PP, Girard TD, et al. Depression, post-traumatic stress disorder, and functional disability in survivors of critical illness in the BRAIN-ICU study: a longitudinal cohort study. *The Lancet Respiratory medicine.* 2014;2(5):369-379.

48. Pandharipande P, Cotton BA, Shintani A, et al. Prevalence and risk factors for development of delirium in surgical and trauma intensive care unit patients. *The Journal of trauma.* 2008;65(1):34-41.

49. Pandharipande P, Shintani A, Peterson J, et al. Lorazepam is an independent risk factor for transitioning to delirium in intensive care unit patients. *Anesthesiology.* 2006;104(1):21-26.

50. Pandharipande PP, Girard TD, Ely EW. Long-term cognitive impairment after critical illness. *The New England journal of medicine.* 2014;370(2):185-186.

51. Gillard J, Schutz M. Composing alarms: considering the musical aspects of auditory alarm design. *Neurocase.* 2016;22(6):566-576.

52. Stevenson RA, Wilson MM, Powers AR, Wallace MT. The effects of visual training on multisensory temporal processing. *Experimental brain research.* 2013;225(4):479-489.

53. Stevenson RA, Schlesinger JJ, Wallace MT. Effects of divided attention and operating room noise on perception of pulse oximeter pitch changes: a laboratory study. *Anesthesiology.* 2013;118(2):376-381.

54. Paine CW, Goel VV, Ely E, et al. Systematic Review of Physiologic Monitor Alarm Characteristics and Pragmatic Interventions to Reduce Alarm Frequency. *J Hosp Med.* 2016;11(2):136-144.

55. Pasco Y, Gauthier PA, Berry A, Moreau S. Interior sound field control using generalized singular value decomposition in the frequency domain. *The Journal of the Acoustical Society of America.* 2017;141(1):334.

56. Edworthy JR, Schlesinger JJ, McNeer RR, Kristensen MS, Bennett CL. Classifying Alarms: Seeking Durability, Credibility, Consistency, and Simplicity. *Biomedical instrumentation & technology / Association for the Advancement of Medical Instrumentation.* 2017;51(s2):50-57.

57. Carayon P, Bass E, Bellandi T, Gurses A, Hallbeck S, Mollo V. Socio-Technical Systems Analysis in Health Care: A Research Agenda. *IIE Trans Healthc Syst Eng.* 2011;1(1):145-160.