Work 41 (2012) 3580-3585
DOI: 10.3233/WOR-2012-0668-3580
IOS Press

Training time estimation to improve alarm reactions

James P. Bliss, Ph.D.[a,*], Rachel Liebman[a] and Eric Chancey[a]
[a]*Psychology Department, Old Dominion University, 5115 Hampton Boulevard, Norfolk, VA, USA*

Abstract. Prior researchers have demonstrated that training may be an effective strategy for improving operator reactions to alarm systems with less than perfect reliability. Of the training strategies tested, recognition of temporal patterns in prior sensor activations seems to offer the greatest promise for improving the speed and appropriateness of subsequent alarm reactions. The current research was completed to clarify which of three temporal interval training methods leads to the most appropriate alarm reactions. Fifty-six undergraduates evaluated whether alarms occurring after sensor activations were true or false, based on elapsed time between the sensor signals and the alarm signals. Participants completed five training sessions to learn to estimate time intervals using simple repetition training, performance feedback, or performance feedback plus subdivision cues. Contrary to expectations, results indicated that participants did not benefit differentially from temporal interval training. Differences between pre- and posttest interval estimation performance was similar among groups, and training groups performed comparably when reacting to signals. Participants generally focused on advertised alarm system reliability, responding more appropriately and more quickly to lower reliability alarms. Future researchers and designers should replicate these findings with realistic tasks and real-world complex task operators to determine their generalization.

Keywords: Alarms, Training, Reliability, Time, Reactions

1. Introduction

Unreliable signaling systems pose a challenge for respondents because they must determine the likelihood of individual signal validity given limited evidence. As a result, respondents often adopt inefficient strategies for reactions. Some opt to match the perceived signaling system reliability, some opt to respond to all signals, some opt to respond to no signals, and some adopt a mixture of these strategies [1]. Complicating the matter is the necessity to divide attention between a primary task and the secondary signal reaction task.

For many years, researchers have worked to understand the implications of low signal reliability. Following several years of work, Breznitz published his text, "The Psychology of False Alarms," documenting the effects of low signal reliability on physiological measures such as heart rate, as well as on behavioral measures like threat avoidance [2]. A decade after his work, Bliss [1] and Getty, Swets, Pickett, and Gonthier [3] provided the first evidence that low signal reliability causes predictable effects on signal reaction time, frequency, accuracy, and signal reaction appropriateness. In the years that followed, many researchers demonstrated the impact of low signal reliability in applied environments such as aviation [4], mining [5], medicine [6], and security monitoring [7].

A common finding by researchers has been that task operators typically rely on task related cues to help them determine the validity of individual signals and, in turn, the reliability of an overall signaling system. For example, anecdotal reports have suggested that research participants have relied on the actions of others, the urgency of the signals themselves, the presence of additional system data, and prior signaling system behavior (reliability) as cues to indicate the probable likelihood that a given signal is true. This has prompted researchers to advocate "Likelihood Alarm Displays" that could embed information from such cues into the signal

*Corresponding author. E-mail: jbliss@odu.edu.

itself to aid operators as they judge signal validity. [8] In the absence of Likelihood Alarm Displays, people often look for secondary cues to indicate whether a signal is valid. For example, upon hearing a fire alarm, people may look for smoke or may watch the actions of others to determine its validity.

As the complexity of an operational system rises, the number of additional cues present also rises. For example, in a chemical processing plant or building security application, signaling systems often consist of multiple linked sensors that contribute to the overall indication of a problem. Frequently, operators must interpret the underlying pattern of sensor activations, or the relationship of sensor activations to alarm activation, to determine whether an alarm signal truly indicates a dangerous event. [9]

2. Use of Training as a Strategy to Improve Signal Responses

Over the years, researchers have investigated strategies to improve alarm reactions. Some of these have included manipulations of signal content, urgency, or timeliness; others have focused on task parameters such as priority or workload; some have concentrated on operator states like attention or teaming. In the applied domain, it is typical for agencies and trainers to advocate generalized reaction strategies as a way to ensure swift, reliable signal reactions (even when global training strategies may not be optimal in all situations). For example, some researchers have advocated responding to all alarms in medical environments, even with the understanding that some may not represent authentic danger. [10]

In addition to training operators to respond to all signals, regardless of suspected validity, other training strategies are possible. These could include training operators to look for additional information, examine past signaling system behavior, consult other operators, or investigate related systems. Bliss and Gilson have stressed the need for research to investigate the efficacy of cue training as a way to improve operator reactions [11].

3. Goal of the Current Research

The research reported here is a follow-on project to examine the utility of training temporal estimation skills to alarm respondents. In prior research, we tested participants to react to alarms by referring to previous sensor activations [12]. One group focused on the activity of a single movement sensor. Another group was trained to focus on the spatial pattern of sensor activations among simulated building rooms. The third group was trained to focus on the temporal pattern represented among sensor activations. The fourth group received no training. Results showed that participants learned and utilized information best in the temporal training group. This led us to believe that training temporal interval recognition might be particularly beneficial for participants who must discriminate true from false alarms. In essence, we trained participants to use elapsed time between a sensor signal and an alarm signal as a cue indicating the validity of the alarm signal.

4. Training Temporal Interval Estimation

An important aspect within multiple applied domains is the operator's ability to estimate time. Automated warning and alert systems that are relied upon too heavily may go unchecked. In cases where these systems malfunction and fail to alert critical events, adverse consequences ensue. Situations such as these have been explored in instances of automation misuse [13]. In some cases, however, the system malfunction may have been detected if the operator had noticed lapses in the occurrences of alerts and warnings over the passage of time.

For example, in 1995 the automated navigation system of the Panamanian cruise ship *Royal Majesty* malfunctioned and consequently the vessel ran aground after it went off course for 24 hours [14]. Subsequent investigation revealed that the fathometer alarm (usually set at 3 meters) was improperly set at 0 meters, effectively rendering the alarm useless. Furthermore, the crew relied solely on the position-fix alarm, an autopilot feature designed to alert deviations in the set course, instead of monitoring position instrumentation. An important aspect to consider is the crews' inability to notice the absence of alerts that "should" have occurred, had these systems been functioning in a reliable capacity. A potential solution to counteract operator inaction may be attained by training focused on time estimation. The notion of training operators to recognize trends in monitoring data to respond to subsequent signals more accurately has been suggested [11] and empirically evaluated [12] before.

Using a simulated security monitoring task, Bliss et al. found marginal support for a performance benefit for participants who were provided with alarm reaction training strategies when compared to a group that received no training. One of the strategies employed was a form of time estimation. Participants were asked to monitor a building schematic with sensors located in each room. Based on the pattern of activations of these sensors, participants were required to respond to a subsequent alarm as being true or false. The alarm system was associated with either 20% or 40% reliability. Participants were told that if the time separations between successive sensor activations became progressively shorter, this was indicative of a true alarm.

Time estimation training has been investigated by other researchers as well. Using a dual-task paradigm, Brown was able to successfully reduce interference between a distracter task and a time estimation task by training participants on the secondary time estimation task [15]. Participants were presented with a computerized display that remained on screen for a set amount of time (between 6 and 14 seconds inclusive). Participants were then asked to reproduce this time interval via the spacebar for both the beginning and ending of the estimated amount of elapsed time. Participants were given either no feedback on the accuracy of their response or an exact indication of the variance from the actual amount of time (i.e. how much over or under the participant had estimated the time in seconds). The results of subsequent testing sessions following the feedback training revealed a performance benefit when compared to the group that received no training. The author suggests that the training used promoted automaticity of the time estimation task and therefore reduce the resources expended on making time interval judgments.

The most promising method for achieving temporal interval recognition seems to be practice sessions with intervening signals that evenly subdivide the target time interval. For example, if the interval to be learned is ten seconds, intervening signals might be presented after each second. Such subdivision is commonly practiced in music to teach musicians to properly count time. One popular method for this is the Takadimi technique. [16]

The overall goal for this research was to determine whether participants could learn to estimate elapsed time and use that information to correctly evaluate alarm signals as true or false based on the elapsed time between a prior sensor signal and the subsequent alarm signal.

5. Hypotheses

From the prior research discussed above, we expected to observe three things. First, participants who experienced alarm reaction training would react more appropriately to true and false alarms than those who did not receive training. Second, participants with subdivision temporal training would respond more appropriately to true and false alarms than those in other training groups. Third, we expected that participants would respond more appropriately to the higher reliability signaling system.

6. Method

6.1. Design

We structured our experiment according to a 3 X 2 mixed design. The first variable was the type of alarm reaction training, manipulated between groups. There were three temporal training conditions: training with feedback (viewing how closely a temporal interval was approximated), training with feedback and a subdivision aid (an auditory tone), and a control group that repeatedly tried to match a temporal interval without aid or feedback.

The second independent variable was the stated reliability of the alarm system. Reliability was manipulated within subjects with two levels: 20% reliable (an alarm signal had a 20% chance of indicating a true problem) and 40% reliable (an alarm signal had a 40% chance of indicating a true problem). These levels were specified to reflect conventional alarm reliability differences used in past research [17] and to resemble the low reliability rates commonly associated with security alarm systems [18].

Dependent measures included participants' score on the primary word search task (number of words correctly circled in 30 minutes), the percentage of alarms reacted to appropriately (responded to if true and canceled if false), the average time in seconds taken for participants to acknowledge the alarms, and the percentage of alarms eliciting a response during the experimental session.

6.2. Participants

Based on a power analysis, we determined that testing 60 participants would ensure experimental power of .80. [19] Fifty-six undergraduate students (21 male, 35 female) from Old Dominion University participated for class credit. The mean age of participants was 22.56 (s= 4.7, min=17, max=45). No participant indicated hearing impairment or color vision deficiency. Participants indicated a mean of 25.2 hours of weekly computer use (s=14.6) and a mean of 3.6 hours of weekly video game use (s=8.6). Participants received course credit for participating.

6.3. Materials

The alarm task was constructed and performed on a desktop Macintosh computer. Participants indicated their responses to the alarm system by using a mouse. Participants acknowledged sensor activations by clicking an "ACK." icon. Following each of 10 sensor activations, an "INTRUDER" alarm signal would follow. Participants then to "RESPOND" or "IGNORE" the alarm based on whether they believed it occurred less than 60 seconds (true alarm) or more than 60 seconds (false alarm) after the preceding sensor activation (see Figure 1). Sensor activations were accompanied by a two-second, 1000-Hz. tone; visual alarms were accompanied by the fire bell sound from a Boeing 757. Two word searches that were used as the primary task were downloaded from the internet site, http://www.puzzle-club.com. Participants received no word bank.

Prior to the tasks, participants completed an Informed Consent Form and a Background Questionnaire to indicate their level of computer and video game knowledge. They also received extensive training to gain skill at estimating time intervals. Participants in all conditions completed a pretest, five sessions, and a posttest. During each training session, the pretest and the posttest they estimated 30, 45, 60, 75, and 90 second intervals (randomly presented) by pressing the mouse key when they believed the interval was complete. Participants in the feedback+aid condition were informed of their guess accuracy (elapsed time in seconds) and were provided a 1500 hz. tone every five seconds as an aid. Participants in the feedback only condition received the feedback but not the subdivision tone. Control participants completed all training sessions but received neither the feedback nor the tone.

6.4. Procedure

After completing the informed consent form and demographic questionnaire, participants were randomly assigned to one of the three training groups. Participants then read instructions for the word search and alarm reaction tasks and a description of the task scenario. They then received a pretest, five sessions of training about time estimation based on their experimental groups, and a posttest. Participants then practiced the alarm reaction task.

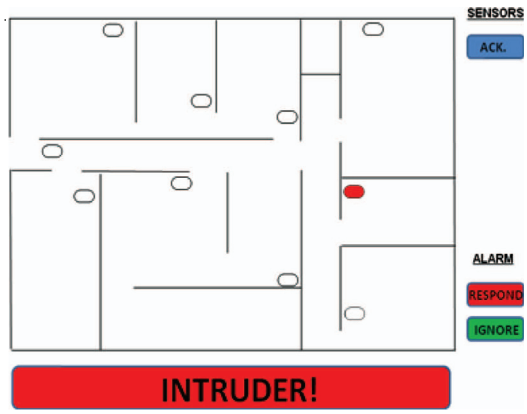

Figure 1. The Sensor and Alarm Reaction Task.

The experiment required participants to complete a word search task while monitoring a building schematic and responding to alternating sensors and alarms. Based on the perceived time interval between sensors and alarms participants either responded to them (indicating they believed them to be true) or ignoring them (indicating they believed them to be false). Sensor activations were indicated by a 1000-Hz. tone and a color change from white to red. An alarm was indicated by the B-757 fire bell and a red block with the word "intruder" at the bottom of the screen (see Figure 1). In each session, there were ten sensor activations and ten subsequent alarm activations. Actual validity of the alarms reflected the particular reliability condition.

Before beginning the first 30-minute experimental session, all participants were told the historical reliability of the alarm system (20% or 40% true alarms, counterbalanced). Following the first experimental session participants were given a ten minute break and then were given a new word search (counterbalanced) and reliability level. Following the second session, participants were debriefed and dismissed.

7. Results

Our first step was to screen data for outliers (primary or secondary task data falling beyond 3 standard deviations from the group mean that reflected system malfunction or errors). We determined that there were no outliers and that all data were normally distributed for observed performance variables (primary task score, alarm reaction appropriateness, alarm reaction time, and alarm response frequency).

Next, we computed an Analysis of Variance (ANOVA) to determine whether alarm system reliability and training condition influenced alarm reaction appropriateness. Using p=.05 as the significance criterion, there was no significance found for the interaction or main effects, though the main effect of alarm system reliability on reaction appropriateness approached significance, $F(1,53)=3.404$, $p=.071$, partial $\eta^2=.06$, observed power=.441 (see Figure 2). This reflects that reaction appropriateness was somewhat higher for lower reliability alarms.

The Analyses of Variance for average alarm response frequency revealed no significant interaction between training group and alarm system reliability and no main effect of training group ($p>.05$); however, there was a significant main effect of reliability, $F(1,53)=32.469$, $p<.001$, partial η^2 = .38, observed power=1.00. This suggests that participants responded more often to alarms generated by the system understood to be more reliable.

Analyses for alarm reaction time and for primary task score showed no significant interactions or main effects, $p>.05$. Likewise, a one-way ANOVA to investigate training improvements as a function of training group was not significant, $p>.05$.

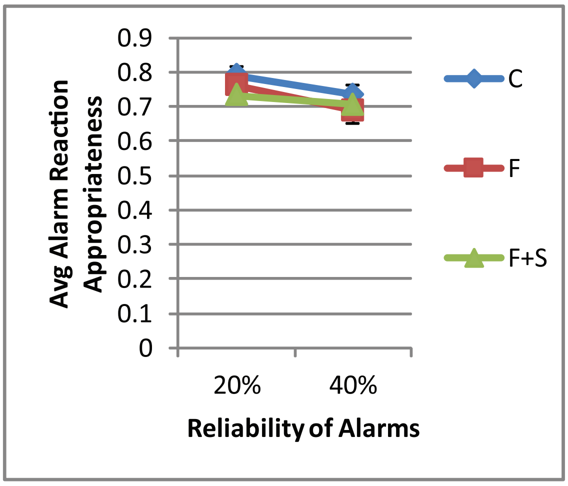

Figure 2. Average Alarm Reaction Appropriateness as a Function of Alarm Reliability and Training Group (C=control; F=feedback; F+S=feedback and subdivision).

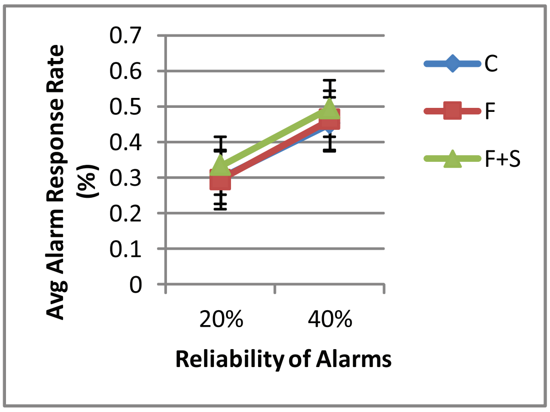

Figure 3. Average Alarm Response Rate as a Function of Alarm Reliability and Training Group (C=control; F=feedback; F+S=feedback and subdivision).

8. Discussion

From prior research, we had expected temporal interval training to positively influence alarm reaction decisions. Quite to the contrary, however, training seemed to make little difference, as control group participants reacted equally appropriately as those trained using recommended methods [15]. There are two potential explanations for this. The first is that the 60-sec time interval threshold used, though realistic in terms of sensor-alarm relationships, may have been too lengthy an interval to benefit from training, as others have suggested [15]. Another possibility is that the complexity or workload associated with our dual-task paradigm may have left few cognitive resources available for participants to devote to temporal interval estimation. Further research should be devoted to determining which of these explanations is most credible.

The fact that participants in all groups failed to benefit equally from the five intense training sessions may reflect poor motivation on the part of the sample tested, or may reflect a fundamental failure of training to influence alarm reaction appropriateness in this case. Again, additional research using trained task operators and a more realistic task may help to clarify this.

Interestingly, participants were sensitive to the advertised historical reliability of the alarm system. In fact, they apparently paid greater attention to the one-time advertisement of reliability than they did to repeated temporal discrimination training to improve reactions. This seems to again provide evidence that alarm respondents may adopt general, heuristic based reaction strategies prior to a session rather than change strategies for individual signals in real time.

The results presented here speak poignantly to realistic signal reaction situations, and may support the idea that any training should focus on general heuristic strategies rather than case-by-case signal evaluation.

References

[1] Bliss, J.P., Gilson, R.D., & Deaton, J.J. (1995). Alarm response behavior under conditions of varying alarm reliability. *Ergonomics, 38*(11), 2300-2312.

[2] Breznitz, S. (1984). *Cry wolf: The psychology of false alarms*. Hillsdale, N.J.: Lawrence Erlbaum Associates.

[3] Getty, D.J., Swets, J.S., Pickett, R.M. & Gonthier, D. (1995). System operator response to warnings of danger: A laboratory investigation of the effects of the predictive value of a warning on human response time. *Journal of Experimental Psychology: Applied, 1*(1), 19-33.

[4] Pritchett, A. (2001). Reviewing the role of cockpit alerting systems. *Human Factors & Aerospace Safety, 1*, 5-38.

[5] Mallett, L., Vaught, C., & Brnich, M.J. (1993). Sociotechnical communication in an underground mine fire: A study of warning messages during an emergency evacuation. *Safety Science, 16*, 709-728.

[6] Xiao, Y. & Seagull, F. J. (1999). An analysis of problems with auditory alarms: Defining the roles of alarms in process monitoring tasks, In E. Haas and J. Edworthy (eds.), *The Ergonomics of Sound: Selections from Human Factors and Ergonomics Society Annual Meetings 1985-2000*, Santa Monica, CA, 256-260.

[7] Marra, A., & Playford, A. (2009). Home alarms: A false sense of security? The Palm Beach Post News [ONLINE]. Downloaded from http://www.palmbeachpost.com/news/crime/home-alarmsa-false-sense-of-security-59710.html November 30, 2009.

[8] Sorkin, R. D., Kantowitz, B. H., & Kantowitz, S. C. (1988). Likelihood alarm displays. *Human Factors, 30*(4), 445-459.

[9] Papadopoulos, Y., & McDermid, J. (2001). Automated safety monitoring: A review and classification of methods. *International Journal of Condition Monitoring and Diagnostic Engineering Management, 4*(4), 1-32.

[10] Xiao, Y., Seagull, F.J., Nieves-Khouw, F., Barczak, N., & Perkins, S. (2004). Organizational-historical analysis of the “failure to respond to alarm” problems. *IEEE Transactions on Systems, Man, & Cybernetics – Part A: Systems and Humans, 34*(6), 772-778.

[11] Bliss, J.P., & Gilson, R.D. (1998). Emergency signal failure: Implications and recommendations. *Ergonomics, 41*(1), 57-72.

[12] Bliss, J. P., & Chancey, E. (2010). The effects of alarm system reliability and reaction training strategy on alarm responses. *Proceedings of the Human Factors and Ergonomics Society 54th Annual Meeting*, (pp. 2248-2252). San Francisco, CA.

[13] Parasuraman, R. & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors, 39,* 230-253.

[14] National Transportation Safety Board. (1997*). Marine accident report – Grounding of the Panamanian passenger ship Royal Majesty on Rose and Crown Shoal near Nantucket, Massachusetts, June 10, 1995* (NTSB/MAR97/01). Washington, DC: Author.

[15] Brown, W. S. (2008). The attenuation effect in timing: Counteracting dual-task interference with time-judgment skill training. *Perception , 37*, 712-724.

[16] Houlahan, M., & Tacka, P. (2008). *From sound to symbol: Music fundamentals*. London: Oxford University Press.

[17] Dixon, S. & Wickens, C.D. (2003). *Imperfect automation in unmanned aerial vehicle flight control.*(AHFD-03-17/MAAD-03-1). Savoy, IL: University of Illinois, Aviation Research Lab.

[18] Sampson, R. (2002). *False burglar alarms*. (Problem-Oriented Guides for Police Series, Report No. 5). Washington, D.C.: U.S. Department of Justice Office of Community Oriented Policing Services.

[19] Keppel, G. and Wickens, T.D. (2004). *Design and analysis (4th ed.)*. Englewood Cliffs, NJ: Prentice-Hall.