EMBRY-RIDDLE Aeronautical University
SCHOLARLY COMMONS

International Journal of Aviation, Aeronautics, and Aerospace

Volume 10 | Issue 1 Article 6

2023

Utilizing the Alarm Taxonomy and Classification System (ATACS) to Redesign Landing Gear Warnings

Stephen Rice
ERAU, scrice@outlook.com
Ryan Lange
Embry-Riddle Aeronautical University, langer2@my.erau.edu
Sean R. Crouse
Embry-Riddle Aeronautical University, sean.r.crouse@gmail.com
Scott R. Winter
Embry-Riddle Aeronautical University, scott.winter@mac.com
Ryan J. Wallace
Embry-Riddle Aeronautical University, ryan.wallace@erau.edu

Scholarly Commons Citation
Rice, S., Lange, R., Crouse, S. R., Winter, S. R., & Wallace, R. J. (2023). Utilizing the Alarm Taxonomy and Classification System (ATACS) to Redesign Landing Gear Warnings. *International Journal of Aviation, Aeronautics, and Aerospace, 10*(1). DOI: https://doi.org/10.58940/2374-6793.1794

Alarms are typically used to provide information or warnings to operators. For example, modern aircraft are equipped with various alarms and alerts intended to reduce operator workload, increase safety, and improve efficiency. A poorly designed alarm may impair its ability to attract the operator's attention (International Organization for Standardization [ISO], 2003), and an unreliable alarm that produces false alarms or misses events may negatively influence operator trust (Breznitz, 1982; Dixon et al,, 2007; Rice, 2009; Wickens & Dixon, 2007). Thus, it is critical to design alarms so operators can interact optimally with them.

Including users in the alarm design process may improve overall system performance (Nielsen & Levy, 1994); however, when an alarm fails to live up to expectations, users traditionally describe the alarm in qualitative terms (e.g., it does not work, it is terrible, it is worthless, etc.). If the user wants the alarm redesigned, they have to express their concerns about the alarm in a similar qualitative manner. This can create confusion between the user and vendor, resulting in cost overruns and time delays. This paper will present a *use case* of a new taxonomy, methodology, scoring system, and data storage process that any researcher can use to design effective alarms in any domain.

Literature Review

Alarms are a form of automation that fall under Stage 2 or Stage 3 automation, as defined by Parasuraman et al. (2000). Tonal alarms typically fall under Stage 2 automation, which alerts a user to the state of the world (e.g., low fuel warning), while speech alarms may cross over into Stage 3 automation, which provides verbal guidance to the user (e.g., ground proximity warning telling the pilot to pull up).

Alarms in aviation are so common that many operators fail to appreciate what it takes to design an effective alarm. In fact, government agencies often spend hundreds of thousands of dollars creating design templates and handbooks explicitly meant to design new alarms or redesign old ones. Multiple handbooks available provide information on how to design alarms; however, these handbooks typically provide qualitative feedback in the form of Do's and Don'ts. The problem with this approach is that it doesn't allow the user, as a subject matter expert (SME), to provide quantitative feedback on what specifically is wrong with an alarm. In many cases, the first time a user is informed about a new alarm is when it has already been installed. Including users in the design process is critical and providing them with a quantitative method of expressing their concerns and critique of the alarm. Otherwise, users will continue to be stuck with suboptimal alarms that often cause more problems than they solve.

In aviation, alarm-related incidents have been widely reported in air traffic control (Wickens et al., 2009), with many of these incidents involving the Minimum Safe Altitude Warning (MSAW). There have also been incidents associated with the Ground Proximity Warning System (GPWS) and the Traffic Collision

Avoidance System (TCAS) (Bliss et al., 1999). Some of these issues are due to alarm fatigue (Edworthy, 2013), trust (Keller & Rice, 2010; Meyer, 2001; 2004; Rice, 2009), perceived urgency (Arrabito et al., 2004; Burt et al., 1995) and loss of situational awareness (Endsley, 2018). Accurately measuring an alarm's efficacy in the early stages of development can be a crucial factor in preventing these types of incidents.

Outside of aviation, alarms are widely used in technology-driven domains. For example, healthcare professionals must observe and maintain awareness of a rapid data flow reflecting their patients' conditions. Modern medical equipment is often designed with various alarms that notify doctors, nurses, and other staff of relevant information about patient status. These users are frequently exposed to hundreds of alarms daily, which can cause alarm fatigue and negatively impact patient safety. A 2016 survey of healthcare professionals reported that at least 30% of respondents experienced adverse patient outcomes related to clinical alarms (Clark, 2016). With an average of 150-400 alarms generated per patient per shift, it is unsurprising that nurses can spend up to 35% of their working time recognizing, identifying, and interpreting the meaning of an alarm (Lewandowska et al., 2020; Li et al., 2018). Poorly designed alarms can further increase the workload of these professionals.

Some prior research has been conducted to evaluate the application of alarms within various systems. Jian et al. (2000) developed a scale measuring the level of user trust in automated systems. Trust is a critical component of human-automation interaction (Bliss, 2003; Hughes et al., 2009; Rice & Keller, 2009) and is included in the taxonomy presented here. Singh et al. (1993) produced a scale indicating the potential for automation complacency, while Arrabito et al. (2004) and Burt et al. (1995) have discussed ways of measuring perceived alarm urgency. Urgency is a sub-factor of Priority, which is also included in the current taxonomy. Other studies have focused on alarm's acceptability, using a single-item rating (Taylor & Wogalter, 2012), and this characteristic is also included in the current taxonomy. Other measures focus on alarm system performance (Dorgo et al., 2021) but not necessarily on the measurement of alarm efficacy.

While these scales are incredibly useful for their intended purposes, they do not provide a methodology to rate the overall efficacy of an alarm or break down where the problems are occurring. Lange et al. (2022) produced an overall alarm efficacy scale, which measures efficacy on a single-factor scale using seven items that users respond to on a Likert-scale. This single score can be used to quantitatively assess the overall efficacy of both tonal and speech alarms; however, it does not provide detailed information about the different characteristics of the alarm nor a way for users to indicate what is wrong with the alarm precisely. The current ATACS process fills this gap.

Most relevant to the current study, a previous taxonomy describes a process by which one can design/redesign alarms in anesthesia (Rice et al., under review); however, that model has four major limitations. First, it is solely limited to use by clinicians in anesthesia and does not encompass any other field. Second, the taxonomy lacks the characteristics necessary for use with alarms in other fields. Third, the process does not necessarily apply to users and SMEs in other arenas. Lastly, the authors did not describe a *use case*, or provide a dataset, to show how the process works on an actual alarm. The current ATACS process fills these gaps by upgrading the taxonomy, process, and scoring system to fit any field. In addition, we provide a *use case* to show how the process works on a real-world alarm in aviation.

Current Study

Previous research has produced a plethora of guidance on how to design alarms and how to rate them on various single-factor scales. However, to date, no taxonomy or classification system exists that allows users to communicate their opinions about alarms in specific and quantitative terms in a language understandable to all parties involved (users, engineers, designers, and human factors professionals) across all domains. In the coming section, we explain how this taxonomy can be used to help design new alarms or redesign troublesome ones.

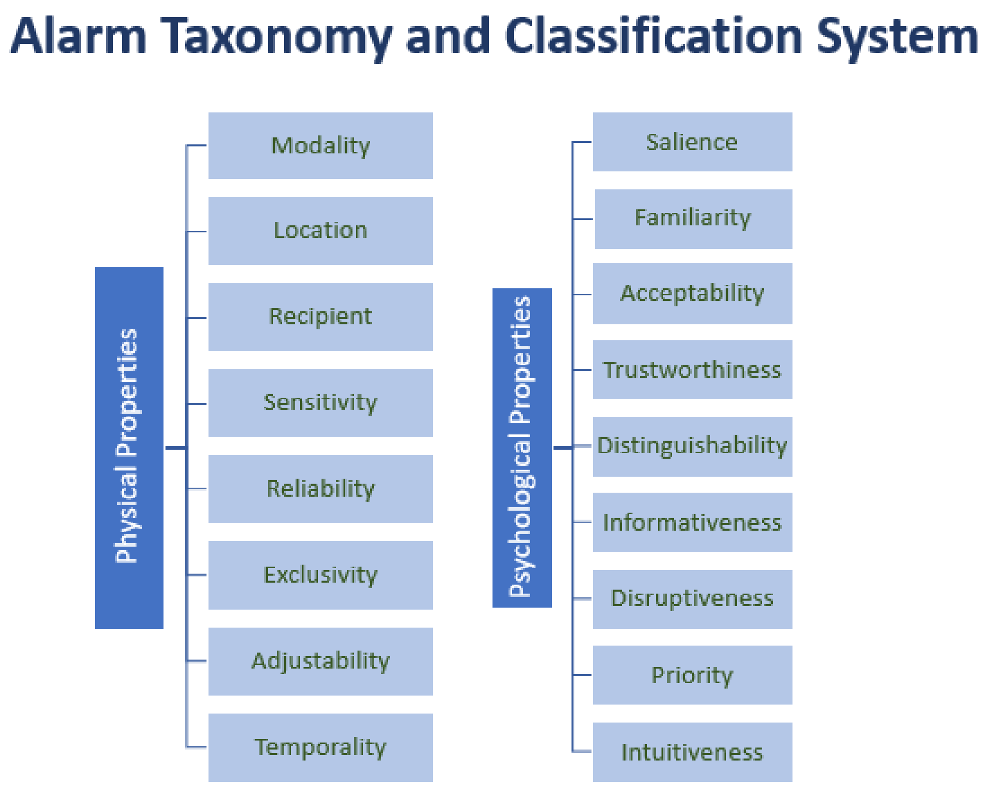
The Taxonomy

The purpose of the proposed alarm taxonomy (see Figure 1) is to comprehensively analyze a particular alarm in a quantitative language shared among users, human factors experts, engineers, and designers. While previous scales focused on the alarm as a whole (Lange et al., 2022) or focused solely on one occupation (authors blinded, under review), the current taxonomy takes a different approach. Here, the whole of the alarm is broken down into 17 different categorical characteristics generated from the literature. This allows the user to be very specific about their feedback and does not punish the characteristics that have been well-designed. For example, an alarm might be perfect if only the location differed. This taxonomy and scoring system allows for focused feedback and individual correction.

The taxonomy is divided into two main categories: Physical and Psychological Properties. Physical properties comprise the physical makeup of the alarm (Modality, Location, Recipient, Accuracy, Reliability, Exclusivity, Adjustability, and Temporality). The Psychological Properties (Salience, Familiarity, Acceptability, Trustworthiness, Distinguishability, Informativeness, Disruptiveness, Priority, and Intuitiveness) address how humans perceive what comes from the alarm. While other characteristics may be proposed, we believe they would fall under one (or more) of the categorical characteristics in Figure 1. For example, the color of an alarm may be considered a characteristic, but we argue it falls under Salience, Distinguishability, and perhaps Informativeness.

Figure 1

Alarm Taxonomy and Classification System



DOI: 10.58940/2374-6793.1794

Definitions

Table 1 provides a summary of the property terms and definitions.

Table 1

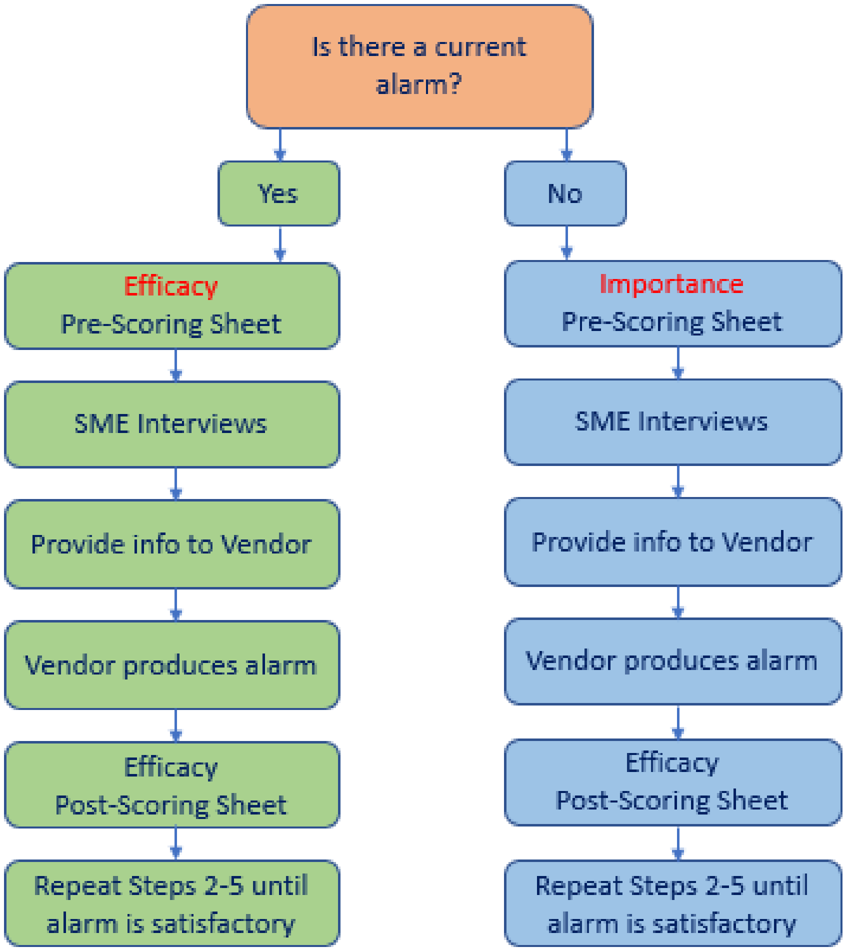
Alarm Taxonomy and Classification System Terms and Definitions

Physical Properties		**Psychological Properties**	
Term	Definition	Term	Definition
Modality	Refers to the sensory input of the alarm (e.g., visual, auditory, tactile, olfactory or gustatory) (Soto-Faraco, 2019).	*Saliency*	Refers to the conspicuousness of an alarm, and indicates how easy it is for a user to notice or detect the alarm (Baldwin & May, 2011).
Location	Describes where the alarm is physically placed (Lee & Kong, 2006).	*Familiarity*	Refers to how familiar the user is with the alarm (McKeown, et al., 2010).
Recipient	Refers to which operator(s) need to respond to the alarm (Rice, 2009).	*Acceptability*	Refers to how acceptable the alarm is to the user (Taylor & Wogalter, 2012).
Sensitivity	Refers to the avoidance of false alarms and misses (Dixon, Wickens & McCarley, 2007).	*Trustworthiness*	Describes how trustworthy the alarm is to the user (Jian, et al., 2000).
Reliability	Refers to how consistently the alarm performs over time (Breznitz, 1982).	*Distinguishability*	Refers to the uniqueness of each alarm when multiple alarms are presenting simultaneously (Edworthy & Hellier, 2006).
Exclusivity	Refers to how many different hazards the alarm is paired to (Edworthy, 1994).	*Informativeness*	Refers to how informative the alarm is (Rayo & Moffatt-Bruce, 2015).
Adjustability	Describes the ability to adjust, or even suppress, an alarm to suit the user's needs (Graham & Cvach, 2010).	*Disruptiveness*	Refers to the ability of the alarm to appropriately disrupt the user's attention (Ljungberg, et al., 2012).
Temporality	Describes when the alarm will be activated, for how long, any repeated activations, and when the alarm stops (Edworthy, et al., 1991).	*Priority*	Refers to which alarm takes precedence, and is a function of urgency and impact (Mathenge, 2020).
		Intuitiveness	Describes how intuitive the alarm is to the user, and whether it easily makes sense.

The Process

Figure 2 presents a flowchart of the process one would use to rate and design an alarm.

Figure 2
Flowchart of the Process for Designing or Redesigning Alarms



First, one must decide if they are redesigning a current alarm or designing a new one from scratch. If redesigning a new alarm, one would follow the green part of the flowchart. This would involve having SMEs (we consider seasoned users to fall into this category as well) use the Efficacy Scoring Sheet to rate the current alarm. Next, the HF expert would conduct interviews with the SMEs to elicit more detailed information about each of the ratings and answer any questions raised by the scoring process. Next, the HF expert would provide this information to the vendor. When the vendor has produced the alarm product, the HF expert can take the alarm back to the original SMEs and have them provide a post-score to ensure their expectations were met. If not, an iterative process would ensue until all parties have reached satisfaction. All information from each stage in this process should be retained by all parties so that one can keep a timeline of each action and refer back to it when necessary.

Scoring Sheets

To quantify the taxonomy, we designed a scoring sheet with the opportunity for SMEs to rate each of the 17 categorical characteristics individually (see Appendix A). Users should score 1 to 5 for each characteristic, with a maximum possible score of 85 for the entire alarm. Readers should notice there are two scoring sheets. The first one is designed to replace current alarms, and the second one is meant for new alarms where none previously existed.

The Efficacy Scoring Sheet focuses on rating a current alarm. This scoring system is intuitive; SMEs simply provide a score for how well each characteristic was designed. The Importance Pre-Scoring sheet (note there is no post-scoring) allows SMEs to determine how important each characteristic is in a new alarm design. For example, a crucial life-saving alarm must be extremely salient and appropriately disruptive. In contrast, another alarm may require placing it on the skin (location and modality). This scoring sheet allows the SMEs to tell the vendors what they consider the highest priority in the design of the new alarm**.**

To Redesign Current Alarms

The following steps detail the process to redesign a current alarm system:

1. Rate the design quality of each characteristic for the current alarm using the quantitative scoring sheet. Each characteristic is rated on a scale of 1-5 (1 = poorly designed; 5 = perfectly designed).
2. Users (SMEs) then participate in a structured interview to clarify and flesh out any characteristics of the alarm they rated.
3. The HF expert provides the scoring sheet and interview notes to the vendor.
4. The vendor designs a prototype alarm.
5. SMEs rate the alarm’s efficacy again and compare it to the original scores.
6. If the prototype is not approved, repeat Steps 2-5 until a satisfactory alarm design is achieved.

To Design a New Alarm

The same process is used, except SMEs will initially rate the importance of each characteristic for a new alarm rather than the design quality of a pre-existing alarm. Note that the scoring sheet changes to an efficacy post-scoring once the vendor delivers the first prototype.

Hypothetical Example

The stall warning is an alarm that notifies the pilot of a pending stall, where the aircraft’s wing exceeds the critical angle of attack. A stall results in a loss of lift and requires prompt corrective action from the pilot to reestablish smooth airflow over the wing. The stall warning was a constant horn in many general aviation aircraft for decades. However, in newer general aviation aircraft, the stall warning is still an audio alarm, but now is a vocal, interrupting verbal announcement of 'stall, stall, stall, stall.' Using this example, we will highlight the changes from the old constant warning alarm to the new verbal warning alarm and assess it using ATACS to demonstrate how this system could be beneficial in an alarm redesign. Table 2 highlights the characteristics that changed due to the alarm redesign.

Table 2

A Comparison of Stall Warning Alarms Using Relevant Characteristics from ATACS

Characteristic	Constant Stall Warning Horn	Verbal Stall Callout
Temporality	A constant sound	A repeating start/stop sound
Salience	Sometimes is not noticed by pilots during high workload situations	Captures attention more effectively
Familiarity	No unique identification	Unique identification could result in new pilot becoming more quickly familiar with it
Acceptability	Ambiguity may lead to lower acceptance by users	Distinct message may result in stronger acceptance by users
Distinguishability	Is not distinguishable from other similar sounds	Clearly distinguishable from other sounds
Informativeness	No vibrant information on what triggered the alarm	Vibrant information on why alarm occurred
Disruptiveness	Disruptive but could be deferred	Highly disruptive and demands attention

As summarized in Table 2, there were several noticeable differences across the ATACS characteristics. In this use case, ATACS could have been used to conduct an initial assessment and rating of the traditional stall warning horn. Based on this initial assessment, designers could identify the specific characteristics where the constant stall horn was rated poorly. As part of the alarm redesign, the ATACS could be used through an iterative process to compare the features of the old and new alarms. The value provided by using ATACS is through a quantifiable scoring system that can differentiate between the various alarm prototypes until the highest-rated alarm is produced.

Use Case

To test this taxonomy in the real world, we redesigned the landing gear warning of an in-development fictional aircraft. We hypothesized that the ATACS process would result in a significantly improved alarm version.

Methods

Participants

Four experienced pilots (1 female) were recruited to provide the SME ratings. The pilot's mean age was 52.75 (*SD* = 7.81). Their average flight hours were 7,750 (*SD* = 2,986.08). The pilots are experienced on several aircraft, including the Boeing 737, 757/767, 777, 787, the Airbus 320, 329, 330, and several military aircraft.

Procedure

Each of the SMEs were presented with the current version of a landing gear warning on the fictional in-development aircraft. The pilots were given a short training lesson on what they were being asked to do and how to use the scoring sheet. They viewed the alarm in use and provided their scores to the experimenter. One of the experimenters, who has prior experience working with NASA alarms, redesigned the alarm according to the scores provided by the SMEs. This revised alarm was then presented back to the same four SMEs, who again provided scores for the alarm. We wish to note that while only two rounds of scoring were needed for this particular study, one can repeat this process as often as needed.

Alarm Design Changes

The alarm audio was created in Audacity version 3.2.1, a free and open-source software for recording and editing sounds. The aircraft cockpit simulation video was created with Microsoft Flight Simulator X using the Cessna 172 cockpit model to represent an in-development fictional aircraft. Visual mockups and animation of the landing gear lever and warning lights were completed in Microsoft PowerPoint version 2211. The aircraft simulation video and the animated landing gear warning light display were edited in Blender version 3.3.1, a free and open source 3D creation software. The final versions of the alarms were uploaded to YouTube.com for easy distribution to the participants.

Alarm Version 1

An auditory alarm was sounded with the following repeating pattern: 456 Hz tone sounding for 0.5 seconds, followed by 0.5 seconds of silence. A visual signal was present in the aircraft cockpit. Three lights adjacent to the landing gear lever were used to indicate the status of the landing gear. The lights were labeled "GEAR DN & LOCKED" (Table 4). In the neutral state, the lights remained a dim gray color to indicate that they were not illuminated. When the aircraft entered the alarm activation criteria configuration, the auditory alarm was activated, the warning lights illuminated bright red, and a red lighted downward arrow labeled "GEAR WARN" was illuminated below the gear lever.

Alarm Version 2

An auditory alarm was sounded with the following repeating pattern: three 0.2 second repetitions of the 456 Hz tone, followed by 0.2 seconds of silence each, followed by the verbalization "*GEAR*!." The same visual signal was present in the aircraft cockpit as with Alarm Version 1. Visual depictions of the alarms are shown in Table 3.

Table 3

A Depiction of the Auditory Alarm Signals

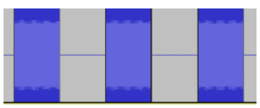
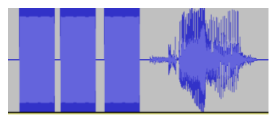
Alarm Version 1		**Alarm Version 2**	
Visualization	Sound	Visualization	Sound
	Repeated: [Beep – Beep – Beep]		Repeated: [Beep, Beep, Beep – “*GEAR!*”]

Table 4

States of the Visual Landing Gear Status Indicator

Neutral	**Gear Up, Alarm Active**
 All lights dark	 Three illuminated red lights and a red arrow labeled “gear warn”

Results

Since four SMEs were each providing 17 scores (one for each characteristic) twice over, we were able to conduct a binomial analysis on the dataset of 68 total scores. In short, we compared the initial score to the revised score and determined if it were equal, worse, or better. One would expect that if the redesign of the alarm had no effect, then the number of worse or better scores should be roughly equivalent. In fact, there were only two revised scores that were worse than the first initial score. Thirty-four (34) of the 68 scores improved, 32 remained the same, and two decreased. A binomial analysis of that data reveal this effect to be highly

DOI: 10.58940/2374-6793.1794

significant, $z(68) = 7.64$, $p < .001$. Clearly, the revised version of the alarm was superior to the initial version, indicating that the ATACS process worked to improve the design. Table 5 provides an overview of the average rating of each characteristic.

Table 5
Characteristic Scores

Characteristic	Version 1 Average	Version 2 Average
Physical Traits		
Modality	3.50	5.00
Location	3.00	3.50
Recipient	3.00	4.25
Sensitivity	4.75	5.00
Reliability	4.75	5.00
Exclusivity	4.00	5.00
Adjustability	3.50	4.25
Temporality	3.50	4.00
Psychological Traits		
Saliency	4.50	5.00
Familiarity	4.50	4.75
Acceptability	4.25	5.00
Trustworthiness	4.50	4.75
	3.00	
Distinguishability		5.00
Informativeness	4.25	5.00
Disruptiveness	4.50	4.75
Priority	3.75	4.50
Intuitive	4.00	5.00
Total Score	67.25	79.75

The authors also conducted a paired samples *t*-test to explore differences between the paired observations. Using the same scoring apparatus, participants were tested at two time points under two different conditions. The average increase for each characteristic was 0.74, with a total score change of 12.50. This was highly significant, (95% CI, 0.496 to 0.974), $t(67) = 6.145$, $p < 0.001$.

Discussion

The purpose of the Alarm Taxonomy and Classification System is to create a usable system that several stakeholders can jointly use in developing or redeveloping an alarm in any field. The system allows users, designers, vendors, engineers, and human factors experts to assess an alarm using standard quantitative scoring. This standardization helps provide specific feedback about the alarm based on two main properties and seventeen characteristics. In addition to the standardized operationalization of these characteristics, the system offers a step-by-step iterative methodology to create a new alarm or assess and, if necessary, improve an existing alarm.

ATACS offers several useful, practical applications. First, it helps to quantify the scoring of existing or new alarms. While other scales have a single overall score, ATACS has 17 characteristics. This value can allow for a targeted response during the development of an alarm to correct poorly scoring areas. Second, several stakeholders, such as engineers, human factors experts, users, and vendors, can easily deploy the system. This universal aspect is essential as the system offers a common language across these stakeholder groups. Third, the administration of the system is efficient. Rating the 17 characteristics can be completed quickly by users or SMEs. This efficiency reduces the chance of participant fatigue and can be easily administered multiple times when assessing multiple alarm options at once. Fourth, this process generates a permanent database of information along the timeline of the iterative process. Anyone involved in the project can access this database anytime during or after the project is completed. This system provides a clear record for all parties to refer to when necessary. Lastly, the system is versatile because it can be applied to any field where alarms are used.

Future Research

Up to this point, we have only discussed this classification system in the context of alarms. However, this methodology can be used for *any* product. The only changes necessary would be for the researcher to identify the categorical characteristics of the product they wish to design or redesign. The process and the scoring system would remain intact. In addition, future research should attempt to identify any new characteristics relevant to alarms or other products. ATACS is meant to be a flexible, evolving document, and future researchers should be allowed to decide what characteristics they think best describe their products.

Conclusions

The purpose of this study was to present an Alarm Taxonomy and Classification System (ATACS) that can be used to design or redesign any alarm in any field. This system creates a method for stakeholders in alarm development and designs a standardized format that assesses the alarm. Users, designers, vendors, engineers, and human factors experts, who often work together in alarm development, can jointly use this system to create a standardized measurable observation. The iterative process provides a format to seek continuous improvement of the alarm until it reaches a satisfactory level. In addition, the system offers 17 unique alarm characteristics so the exact area of possible deficiency in alarm development can be detected. While the use case example in this paper highlights an aviation example, ATACS can be used in any field in which alarms are used.

References

American Society of Anesthesiologists. (2018). *Statement on principles for alarm management for anesthesia professionals*. https://www.asahq.org/standards-and-guidelines/statement-on-principles-for-alarm-management-for-anesthesia-professionals#:~:text=ALARM%20SYSTEM%20SETTINGS%20must%20balance,conditions%20in%20each%20care%20environment.

Arrabito, G. R., Mondor, T. A., & Kent, K. J. (2004). Judging the urgency of non-verbal auditory alarms: A case study. *Journal of Ergonomics, 47*(8), 821-840. https://doi-org.ezproxy.libproxy.db.erau.edu/10.1080/0014013042000193282

Baldwin, C. L., & May, J. F. (2011). Loudness interacts with semantics in auditory warnings to impact rear-end collisions. *Transportation Research. Part F, Traffic Psychology and Behaviour*, *14*(1), 36–42. https://doi.org/10.1016/j.trf.2010.09.004

Bliss, J. P., & Acton, S. A. (2003). Alarm mistrust in automobiles: How collision alarm reliability affects driving. *Applied Ergonomics*, *34*(6), 499–509. https://doi.org/10.1016/j.apergo.2003.07.003

Bliss, J. P., Freeland, M. J., & Millard, J. C. (1999). Alarm related incidents in aviation: A survey of the Aviation Safety Reporting System Database. *Proceedings of the Human Factors and Ergonomics Society... Annual Meeting Human Factors and Ergonomics Society. Meeting*, *43*(1), 6–10. https://doi.org/10.1177/154193129904300102

Breznitz, S. (1982). Cry wolf: The psychology of false alarms. *Psychology Press*.

Burt, J. L., Bartolome, D. S., Burdette, D. W., & Comstock, J. R. (1995). A psychophysiological evaluation of the perceived urgency of auditory warning signals. *Ergonomics*, *38*(11), 2327-2340. https://doi.org/10.1080/00140139508925271

Bustamante, E. A., Bliss, J. P., & Anderson, B. L. (2007). Effects of varying the threshold of alarm systems and workload on human performance. *Ergonomics*, *50*(7), 1127–1147. https://doi.org/10.1080/00140130701237345

Clark, T. (2016, June 5). *HTF Update: 2016 national clinical alarm survey results*. AAMI Conference and Expo, Tampa, FL.

Dixon, S. R., Wickens, C. D., & McCarley, J. S. (2007). On the independence of compliance and reliance: Are automation false alarms worse than misses? *Human Factors*, *49*(4), 564–572. https://doi.org/10.1518/001872007X215656

Dorgo, G., Tandari, F., Szabó, T., Palazoglu, A., & Abonyi, J. (2021). Quality vs. quantity of alarm messages - How to measure the performance of an alarm system. *Chemical Engineering Research and Design*, *173*, 63–80. https://doi.org/10.1016/j.cherd.2021.06.022

Edworthy, J. (1994). The design and implementation of non-verbal auditory warnings. *Applied Ergonomics, 25*(4), 202–210. https://doi.org/10.1016/0003-6870(94)90001-9

Edworthy, J. (2013). Medical audible alarms: A review. *Journal of the American Medical Informatics Association, 20*(3), 584-589. https://doi.org/10.1136/amiajnl-2012-001061

Edworthy, J., & Hellier, E. (2006). Alarms and human behaviour: Implications for medical alarms. BJA: *British Journal of Anaesthesia, 97*(1), 12–17. https://doi.org/10.1093/bja/ael114

Edworthy, J., Loxley, S., & Dennis, I. (1991). Improving auditory warning design: Relationship between warning sound parameters and perceived urgency. *Human Factors*, *33*(2), 205–231. https://doi.org/10.1177/001872089103300206

Endsley, M. R. (2018). Automation and situation awareness. In *Automation and human performance: Theory and applications* (pp. 163–181). CRC Press. https://www.taylorfrancis.com/chapters/edit/10.1201/9781315137957-8/automation-situation-awareness-mica-endsley

Graham, K. C., & Cvach, M. (2010). Monitor alarm fatigue: Standardizing use of physiological monitors and decreasing nuisance alarms. *American Journal of Critical Care: An Official Publication, American Association of Critical-Care Nurses*, *19*(1), 28–34; quiz 35. https://doi.org/10.4037/ajcc2010651

Hughes, J. S., Rice, S., Trafimow, D., & Clayton, K. (2009). The automated cockpit: A comparison of attitudes towards human and automated pilots. *Transportation Research. Part F, Traffic Psychology and Behaviour*, *12*(5), 428–439. https://doi.org/10.1016/j.trf.2009.08.004

International Organization for Standardization (2003). Ergonomics — Danger signals for public and work areas — Auditory danger signals (ISO Standard No. 7731:2003). https://www.iso.org/standard/33590.html

Jian, J.-Y., Bisantz, A., & Drury, C. (2000). Foundations for an empirically determined scale of trust in automated systems. *International Journal of Cognitive Ergonomics*, *4*, 53–71. https://doi.org/10.1207/S15327566IJCE0401_04

Keller, D., & Rice, S. (2009) System-wide versus component-specific trust using multiple aids. *The Journal of General Psychology, 137*(1), 114-128. https://doi.org/10.1080/00221300903266713

Lange, R., Rice, S., Severin, C., Chiu, J., Ruskin, K., & Rice, C. (2022). Creation of tonal and speech alarm efficacy scales. *The Collegiate Aviation Review*

International, *40*(2). https://ojs.library.okstate.edu/osu/index.php/CARI/article/view/8709

Lee, J.-S., & Kong, D. (2006). Investigation of hearing protection effects in an extreme noise environment with an alarm location problem. *International Journal of Industrial Ergonomics, 36*(8), 685–693. https://doi.org/10.1016/j.ergon.2006.04.008

Lewandowska, K., Weisbrot, M., Cieloszyk, A., Mędrzycka-Dąbrowska, W., Krupa, S., & Ozga, D. (2020). Impact of alarm fatigue on the work of nurses in an intensive care environment-A systematic review. *International Journal of Environmental Research and Public Health, 17*(22), 8409. https://doi.org/10.3390/ijerph17228409

Li, T., Matsushima, M., Timpson, W., Young, S., Miedema, D., Gupta, M., & Heldt, T. (2018). Epidemiology of patient monitoring alarms in the neonatal intensive care unit. *Journal of Perinatology: Official Journal of the California Perinatal Association*, *38*(8), 1030–1038. https://doi.org/10.1038/s41372-018-0095-x

Ljungberg, J. K., Parmentier, F. B. R., Hughes, R. W., Macken, W. J., & Jones, D. M. (2012). Listen out! Behavioural and subjective responses to verbal warnings. *Applied Cognitive Psychology*, *26*(3), 451–461. https://doi.org/10.1002/acp.2818

Mathenge, J. (2020). *Impact, urgency & priority: Understanding the matrix*. https://www.bmc.com/blogs/impact-urgency-priority/.

McKeown, D., Isherwood, S., & Conway, G. (2010). Auditory displays as occasion setters. *Human Factors*, *52*(1), 54–62. https://doi.org/10.1177/0018720810366861

Meyer, J. (2001). Effects of warning validity and proximity on responses to warnings. *Human Factors*, *43*(4), 563–572. https://doi.org/10.1518/001872001775870395

Meyer, J. (2004). Conceptual issues in the study of dynamic hazard warnings. *Human Factors*, *46*(2), 196–204. https://doi.org/10.1518/hfes.46.2.196.37335

Nielsen, J., & Levy, J. (1994). Measuring usability: Preference vs. performance. *Communications of the ACM*, *37*(4), 66–75. https://doi.org/10.1145/175276.175282

Parasuraman, R., Sheridan, T. B., & Wickens, C. D. (2000). A model for types and levels of human interaction with automation. *IEEE Transactions on systems, man, and cybernetics-Part A: Systems and Humans, 30*(3), 286-297. https://www.ida.liu.se/~769A09/Literature/Automation/Parasuraman,%20Sheridan,%20Wickens_2000.pdf

Rayo, M. F., & Moffatt-Bruce, S. D. (2015). Alarm system management: Evidence-based guidance encouraging direct measurement of informativeness to improve alarm response. *BMJ Quality & Safety*, *24*(4), 282. https://doi.org/10.1136/bmjqs-2014-003373

Rice, S. (2009). Examining single- and multiple-process theories of trust in automation. *The Journal of General Psychology*, *136*(3), 303–319. https://doi.org/10.3200/GENP.136.3.303-322

Rice, S., Ruskin, K., & Clebone, A. (under review). *A framework for alarm evaluation and design*.

Singh, I. L., Molloy, R., & Parasuraman, R. (1993). Automation- induced "complacency": Development of the complacency-potential rating scale. *The International Journal of Aviation Psychology*, *3*(2), 111–122. https://doi.org/10.1207/s15327108ijap0302_2

Soto-Faraco, S. (2019). *Multisensory interactions in the real world* (1st ed.). Cambridge University Press.

Taylor, J. R. I., & Wogalter, M. S. (2012). Acceptability of evacuation instruction fire warnings. *Proceedings of the Human Factors and Ergonomics Society... Annual Meeting Human Factors and Ergonomics Society. Meeting*, *56*(1), 1753–1757. https://doi.org/10.1177/1071181312561352

Watson, M. O., & Sanderson, P. M. (2007). Designing for attention with sound: Challenges and extensions to ecological interface design. *Human Factors*, *49*(2), 331–346. https://doi.org/10.1518/001872007X312531

Wickens, C. D., & Dixon, S. R. (2007). The benefits of imperfect diagnostic automation: A synthesis of the literature. *Theoretical Issues in Ergonomics Science*, *8*(3), 201–212. https://doi.org/10.1080/14639220500370105

Wickens, C. D., Rice, S., Keller, D., Hutchins, S., Hughes, J., & Clayton, K. (2009). False alerts in air traffic control conflict alerting system: Is there a "cry wolf" effect? *Human Factors*, *51*(4), 446–462. https://doi.org/10.1177/0018720809344720

Appendix A – System Scoring Sheets

Instructions. For each of the characteristics below, please provide a score of 1 (poorly designed) to 5 (perfectly designed), indicating how *appropriately the alarm was designed* based on that particular characteristic. The final row should be the total of all the scores.

Physical Properties

The modality of the alarm is a good choice.	______________
The location of the alarm is a good choice.	______________
The alarm effectively reaches all necessary recipients.	______________
The alarm is sensitive (accurate).	______________
The alarm is reliable (consistent).	______________
The alarm is exclusive, in that it is limited to a single purpose.	______________
The alarm is suppressible by the proper authority.	______________
The alarm activates at appropriate times.	______________

Psychological Properties

The alarm is salient (easily perceived).	______________
This type of alarm is familiar to users.	______________
The alarm is acceptable to users.	______________
The alarm is trustworthy.	______________
The alarm is distinguishable from other alarms.	______________
The alarm is appropriately informative.	______________
The alarm is appropriately disruptive.	______________
The alarm reflects the appropriate priority level.	______________
The alarm is intuitive.	______________

Importance Pre-Scoring Sheet

Instructions. For each of the characteristics below, please provide a score of 1 (not important) to 5 (extremely important), indicating how important that characteristic is to the success of the alarm.

Physical Properties

The modality of the alarm. ______________
The location of the alarm. ______________
The alarm's recipients. ______________
The alarm's sensitivity (accuracy). ______________
The alarm's reliability (consistency). ______________
The alarm's exclusivity. ______________
The alarm's suppressibility. ______________
The alarm's temporality. ______________

Psychological Properties

The alarm's salience. ______________
The alarm's familiarity to users. ______________
The alarm's acceptability to users. ______________
The alarm's trustworthiness. ______________
The alarm's distinguishability. ______________
The alarm's informativeness. ______________
The alarm's disruptiveness. ______________
The alarm's priority level. ______________
The alarm's intuitiveness. ______________

DOI: 10.58940/2374-6793.1794