SONIFICATION AND RELIABILITY – IMPLICATIONS FOR SIGNAL DESIGN

James P. Bliss and Randall D. Spain

Psychology Department – MGB 250
Old Dominion University
Norfolk, Virginia
jbliss@odu.edu

ABSTRACT

Sonifications of complex data streams represent a new way for task designers to convey important information to task operators. In recent years, researchers have applied sonification technology in a variety of task domains, including medical device monitoring, complex task instruction, and visualization of data streams and sets. The use of sonifications as emergency signals has been suggested as a way to convey continuous task state information to operators. However, researchers have focused mostly on acoustic properties of sonifications, and have not considered operator trust of them. Past research has shown predictable operator trust-driven reactions to conventional alarms. It is necessary to extend such investigations to sonifications, so that designers may know whether sonifications might represent a technological solution to foster more rapid and appropriate real-time trust calibration by task operators. In this paper, we describe prior research with alarm mistrust, and highlight the potential benefits of further research combining signal reliability and sonification.

[Keywords: Sonification, Reliability, Trust]

1. INTRODUCTION

Sonification is a form of continuous auditory display that maps sensed relations in data to an acoustic signal for the purposes of display (Watson & Sanderson, 2004). Changes in data values are associated with changes in an associated acoustic parameter, such as sound wave frequency or amplitude. Sonifications are built upon the notion of pre-attentive awareness and exploit the auditory modality's ability to recognize patterns or small changes in an auditory event. Unlike binary auditory displays, auditory icons, or auditory earcons, sonifications promote eyes-free continuous monitoring without startling or disrupting attentional focus (Watson & Sanderson, 2004). Thus, if sonifications are designed and implemented effectively, human operators may effectively monitor complex systems while adhering to additional responsibilities without having to constantly switch attention from one task to another.

1.1. Common Domains/Applications of Sonification

There are a number of continuous auditory displays in use. Perhaps the most popular use of a sonification has been the Geiger counter for hazardous material (radiation) detection. Geiger counters provide a clicking signal that is mapped in terms of frequency to the level of radioactive materials in the environment. The beneficence of the Geiger counter's auditory display is evident when operators are required to have their hands free to accomplish other tasks, or when they must be able to detect minute changes in radioactivity levels.

In medical environments, the pulse oximeter has been used as a sonification display for many years. A common auditory display used by anesthesiologists, the purpose of the oximeter is to map a continuous series of beeps to the patient's heart rate, and to map the pitch of the beep to oxygen saturation in the arterial blood (Watson & Sanderson, 2004). Parameter changes are interpreted with respect to the patient's current status, and attention is directed to the patient or the visual display as required. The pulse oximeter has been shown to be the most successful monitor in the operating room theater for detecting evolving patient incidents (Watson & Sanderson, 2004; Webb et al., 1993) The temporal nature of the auditory display promotes "eyes free" monitoring while providing information that is essential for determining when decisions have to be made as abnormal patient states approach (Watson, Sanderson, & Russell, 2004).

In addition to their use in medical environments, one of the most popular uses for sonifications in recent years has been in conjunction with vehicular backup alarms. For many such alarms, a tone will sound with increasing frequency as the vehicle approaches an object. As contact becomes imminent, the tone becomes steady-state (Williams, Online).

In their report to the National Science Foundation, Kramer et al. (1997) discuss in some detail the use of sonification to serve as an educational tool, allowing students and scientists to observe data trends that might otherwise escape visual search. In such cases, continuous auditory stimuli are mapped to mathematical data properties so that students may more easily and intuitively observe cause-and-effect relationships.

In general, researchers have shown interest in determining how continuous auditory displays can be used in aviation to enhance situation awareness and reduce reliance on auditory alarms. For instance, Kazem, Noyes, and Lieven, (2003) have demonstrated the use of a continuous auditory display to create a spatialized auditory environment that may provide pilots with an intuitive image of their aircraft in spatial context. When an agent enters the aircraft's environment, a continuous representational sound source indicates the agent's positioning within a virtual 3-D environment. Sound source manipulation allows the pilot to track the positioning of the agent until the agent vacates the environment.

In each of the cases described, a critical challenge has been the proper specification of physical task parameters to be represented, and selection of auditory aspects to manipulate. In many such cases, designers make such decisions according to intuition or based on convenience or manipulability of stimulus qualities. To date, there have been few attempts to empirically determine and specify recommendations for sonification.

1.2. Research Concerning Sonification

Since the advent of sonfications, investigations of their use have been driven by applied task operators (such as surgeons, aviators, and nuclear control operators) as well as a small number of theoretically-driven researchers. The ultimate goal of these investigations has been to converge on a set of best practices that can be followed to ensure optimal design and use of sonifications.

In many sonifications, auditory variables such as fundamental frequency, amplitude, and tempo are used to represent environmental data. As such, many researchers have focused on determining the best way to convey information through these variables (Walker, 2002). For instance, Walker and Kramer (1996) used a simulated process control interface to examine how well sound attributes such as loudness, pitch, tempo, and onset time best represented different data dimensions (e.g. pressure, temperature). They found that certain sound attributes were better at representing specific types of data dimensions. Specifically, they found that amplitude was best suited for conveying temperature, whereas sound onset rate was best for representing size. Surprisingly though, tempo, which may seem to be an intuitive mapping for rate, was only moderately successful at conveying this information.

Researchers have also found that changes in sound dimensions correlate differently with perceived changes in data dimensions. For instance, Walker (2002) found that increases in tempo were positively correlated to perceptual increases in temperature, pressure, and velocity, but negatively correlated to perceptual changes in size. Walker (2002) also found that increases in sonification pitch are positively related to higher estimates of temperature, pressure, and velocity. This link between changes in sound and changes in environmental data is collectively called the polarity of sound, and is a key question that sonification researchers must address to design an effective auditory display.

In recent years, researchers have continued to investigate sonification of data for a variety of purposes. For example, researchers at Georgia Technological University have been using sonified signals to assist visually challenged individuals with navigating terrain or other environments (Walker & Lindsay, 2005). Other researchers have focused on the use of sonification as teaching aids for musical instrument skills (Ferguson, 2006) and for representing elements of complex perceptual tasks such as anesthesia monitoring (Anderson & Sanderson, 2004).

1.3. Definition of Signal Reliability

One of the more perplexing problems addressed by perceptual researchers is operator mistrust of signaling systems that have a reputation for generating false signals (Breznitz, 1983; Sorkin, 1988; Bliss, Deaton, & Gilson, 1995). As noted by Getty, Swets, Pickett, and Gonthier (1995), signal reliability is a direct function of the positive predictive value of a signal. That is, the potential of a signaling system to generate an alarm when there is indeed an operational problem. Unfortunately, designers are often faced with a dilemma because of the logic driving alarm system sensors: If the sensors are set too liberally, they will consistently generate signals when it is appropriate to do so. However, they will also generate false signals. Conversely, sensors that are set too conservatively will not generate many false signals, but will fail to generate signals at times when it is appropriate to do so. Because of the legal disposition for designers to warn users, emergency signaling systems are frequently set too liberally. As a result, the false alarm problem commonly occurs, and leads to operator mistrust of signaling systems.

1.4. Research Concerning Signal Reliability

Since the early 1960s, researchers have devoted considerable effort to the problems of alarm mistrust and concomitant alarm trust calibration by task operators. They have determined several things about alarm mistrust:

- Alarm mistrust leads directly to performance degradation, often in the form of reduced or eliminated responsiveness (Pate-Cornell, 1986).

- Individuals vary with regard to the strategies they use to react to alarm signals deemed unreliable (Xiao & Seagull, 2000).
- Trust calibration may be accomplished more quickly by advertising reliability rates before a task session (Bliss, Gilson, & Deaton, 1995)
- Individuals often respond to alarms by mirroring the stated reliability rate, even in the absence of feedback. (Craig, 1978).
- Task workload may exacerbate the cry-wolf effect (Dunn, 1995).
- Teams of individuals take longer to react, but may show gains in reaction appropriateness (Bliss & Fallon, 2003).
- Requiring operators to hold reliability information in memory before reacting leads to delayed responses (Bliss & Capobianco, 2003).
- Longer duration signals are associated with true alarms (Bliss, Fallon, & Nica, 2004).

Because of the many and diverse operational problems associated with alarm mistrust, researchers have devoted effort to meliorate the affects of alarm mistrust.

In some cases, they have relied on operator-related strategies to improve alarm reaction strategies. For example, Bliss, Dunn and Fuller (1995) showed that the presence of hearsay information about alarm reliability rates influenced subsequent response tendencies.

In other cases, researchers have focused on task-related strategies for improving reaction performances. Bliss (1997) presented aviators with voice alarms, noting that the addition of voice stimuli was sufficient to bolster response rates.

A third strategy for improving reaction performance has been to implement technological changes to the alarm reaction paradigm. For example, researchers have long known that adding verbiage to alarm signals can improve responding (Pollack & Tecce, 1958).

1.5. Reliability of Sonified Signals

Even though researchers have examined the consequences of alarm mistrust for over 20 years, they have limited their investigations to traditional alarm signals that annunciate in a discrete fashion. That is, the signals indicate that dangerous conditions exist presently; however, the physical and acoustic properties of the alarm signals are not mapped to any particular task parameter and do not change over time.

Yet, as suggested by the work of researchers and signal designers, sonifications may represent a new and improved approach for emergency signal design. Their continuous nature may allow task operators to better devote attention to the task and its changes. In fact, in some cases designers have advocated the use of sonification based alarm signals specifically to help task operators avoid false alarms (Williams, Online). However, it is clear from research literature that certain sonified signals still suffer from frequent false alarms (Xiao & Seagull, 2000).

There is certainly real-world justification for the idea of examining the perceived reliability of sonified signals, and for empirically determining operator preferences and reactions to sonified signals. Reasons why reactions to differences in reliability might be different for sonifications include the following:

- It is possible that trust calibration may occur more quickly because of the continuous nature of sonifications.
- Task operators may display elevated overall trust levels to sonifications because of the perception that such signals are updated more quickly and are therefore more temporally accurate.
- Task operators may trust sonifications more (in terms of the predictability aspect) because there are more embedded data in sonfications.
- Because sonifications may allow easier creation of likelihood alarm signals (Sorkin & Woods, 1985), operators may feel more empowered to make reaction decisions.
- Because of the presence of a stimulus timeline, operators can build a more detailed schema concerning time-based fluctuations in signal reliability, and may therefore predict future reliability states and fluctuations.

One of the complexities surrounding the task of exploring the perceived reliability of sonifications is their temporal nature. Because sonifications are by definition dynamic auditory stimuli, their primary acoustic and secondary environmental correlates will fluctuate across time. In some cases, this may mean that the validity of the sonification may also vary across time. Therefore, it is possible that reliability estimates should be phrased in terms of ranges instead of absolute values.

Another potential difficulty is the acoustic complexity of the sonifications. In some cases, sonified signals may consist of auditory concoctions rather than unitary stimuli. Because the acoustic components may fluctuate independently, any estimates of signal reliability will need to be associated with specific aspects of the sonification.

A third difficulty relates to cognitive associations made by operators. Because sonifications are acoustically complex, it is quite possible that task operators may associate elements of the signal with other signals they have experienced before. If so, the reliability of the prior signals may be generalized to the sonification. Such generalization may often be inappropriate because the sonification bears little operational relationship to the prior signal domain or task.

Challenges notwithstanding, the potential benefits of using sonifications as a design strategy to overcome alarm mistrust warrant examination. The following section describes one recent project to accomplish this goal.

1.6. Empirical Research Results

To date, there have been very few empirical investigations focusing on perceived reliability of and reactions to sonifications. However, recent investigations have yielded some results.

Spain (2006) recently completed an investigation of sonification and reliability associated with simulated anesthesia alarms. He manipulated sonification pulse rate, presenting participants with sonifications within an anesthesia simulation task. For that task, participants were responsible for monitoring a simulated "patient" while performing a compensatory tracking task and a resource allocation task from the Multi-Attribute Task Battery (MATB; Comstock & Arnegard, 1992). The blood pressure of the "patient" was available for monitoring on a separate computer screen 90 degrees to the right of the MATB computer. Deviations in blood pressure were presented to patients using sonifications, where the fundamental frequencies of the auditory stimuli were mapped to the blood pressure readings (40, 60, or 80 pulses per minute). Participants were to detect changes in fundamental frequency and decide whether to respond to them, based on their knowledge of the sonification system's reliability (40% or 60% true signals).

In addition to MATB and alarm reaction data, participants also completed a trust questionnaire, adapted from Jian, Bisantz and Drury's (2000) work.

Spain's (2006) research findings replicated prior work with alarm mistrust, showing that participants exhibited greater trust for sonifications that were more reliable. These results are similar to past research examining alarm response behaviors (see Bliss et al., 1995) and provide behavioral evidence that participants display greater trust in the more reliable system. Participants who interacted with the 60% reliable sonification system responded significantly faster to patient problems than participants who interacted with the 40% reliable system. These results are also consistent with previous literature that suggests alarm reliability significantly affects response time (Bliss et al., 1996; Getty et al., 1995).

A particularly interesting finding was that participants exhibited more trust (and less perceived workload) in the sonification stimuli that had an intermediate pulse rate (60 ppm). Such a finding may suggest that task operators are more comfortable, or more knowledgeable of signals that activate at 60 ppm. Possible reasons for this may include the fact that humans tend to regress toward the mean, or perhaps human familiarity with music that is played at an intermediate tempo.

1.7. Implications for Future Research

As indicated earlier, sonification may provide a way to realize Sorkin et al.'s (1988) dream of a usable "likelihood alarm display." However, before this can happen, more information must be gathered concerning operator perceptions of reliability where sonifications are concerned. Spain's (2006) research is only one step toward a more complete investigation of perceived reliability of sonifications. Other researchers such as Avi Harel (2006) have also acknowledged the issue, and hopefully will contribute knowledge so that designers will soon have a list of best practices concerning reliability and sonification design.

2. CONCLUSIONS

For some time, researchers have stressed the notion that auditory stimuli should be designed to convey realistic and appropriate urgency levels (Stanton & Edworthy, 1998). As Sorkin (1988) and Breznitz (1984) poignantly demonstrated, urgency is only one issue among many that must be considered when designing and implementing auditory signals. The perceived reliability of auditory signals may have a large influence on the tendency of operators to heed, cancel, minimize, or respond to auditory signals.

As complex task environments become more complex, researchers and designers will continue to look to sonifications as potential solutions for embedding greater amounts of information within auditory signals.

Researchers such as Walker (2002) and Anderson and Sanderson (2004) have already employed sonifications for a variety of tasks. In doing so, they have made recommendations concerning the proper frequency, amplitude, and timbre of those signals. It is important that researchers also consider the unique influence that sonification may have on perceived reliability of signals. If signals are properly designed and implemented, it is possible that claims such as those made by Williams (online) about reduced false alarm rates may be realized.

3. REFERENCES

Anderson, J., & Sanderson, P. (2004). Designing sonification for effective attentional control in complex work domains. Proceedings of the 48th Annual Meeting of the Human Factors and Ergonomics Society. HFES: Santa Monica, CA.

Bliss, J.P. (1997). Alarm reaction patterns by pilots as a function of reaction modality. The International Journal of Aviation Psychology, 7(1), 1-14.

Bliss, J.P., & Capobianco, G. (2003). Temporal persistence of alarm mistrust. Proceedings of the International Ergonomics Association XVth Triennial Congress. August 25-29. Seoul, Korea.

Bliss, J.P., Dunn, M., & Fuller, B.S. (1995). Reversal of the cry-wolf effect: an investigation of two methods to increase alarm response rates. Perceptual and Motor Skills, 80, 1231-1242.

Bliss, J.P., & Fallon, C. (2003). The effects of leadership style and primary task workload on team performance and follower satisfaction. International Journal of Applied Aviation Studies, 3(2), 259-276.

Bliss, J.P., Fallon, C.K., & Nica, N. (2006). The role of alarm signal duration as a cue for alarm validity. Applied Ergonomics, 38, 1-9.

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

Breznitz, S. (1984). Cry wolf: The psychology of false alarms. Hillsdale, NJ: Lawrence Erlbaum Associates.

Comstock, J.R., & Arnegard, R.J. (1992). The multi-attribute task battery for human operator workload and strategic behavior research. (NASA Technical Memorandum #104174). Langley, VA: National Aeronautic and Space Administration.

Craig, A. (1978). Is the vigilance decrement simply a response adjustment towards probability matching? Human Factors, 20(4), 441-446.

Dunn, M.C. (1995). Primary-task and secondary-task workload and the cry-wolf phenomenon. Unpublished doctoral dissertation, The University of Alabama in Huntsville.

Fallon, C.K., & Bliss, J.P., & Nica, N. (2004). The effects of alarm mistrust and signal duration on alarm reactions and perception of alarm validity. Proceedings of the 48th Annual Meeting of the Human Factors and Ergonomics Society. Santa Monica, CA: Human Factors and Ergonomics Society. September 20-24: New Orleans, LA.

Ferguson, S. (2006). Learning musical instrument skills through interactive sonification. Proceedings of the International Conference on New Interfaces for Musical Expression. Paris, France.

Getty, D.J., Swets, J.A., & 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, 19-33.

Harel, A. (2006). Alarm reliability: What if an alarm goes off and no one hears it? User Experience, 5(3), 10-14.

Kazem, M.L.N., Noyes, J.M. & Lieven, N.J.A. (2003). Design considerations for a background auditory display to aid pilot situation awareness. In Proceedings of the 9th International Conference on Auditory Display. Boston, MA: Boston University Publications.

Kramer, G., Walker, B., Bonebright, T., Cook, P., Flowers, J., Miner, N., & Neuhoff, J. (Eds.) (1997). Sonification report: Status of the field and research agenda. Prepared for the National Science Foundation by members of the International Community for Auditory Display.

Pate-Cornell, M.E. (1986). Warning systems in risk management. Risk Analysis, 6(2), 223-234.

Pollack, I., & Tecce, J. (1958). Speech annunciator warning indicator system: Preliminary evaluation. The Journal of the Acoustical Society of America, 30(1).

Sorkin, R.D. (1988). Why are people turning off our alarms? Journal of the Acoustical Society of America, 84(3), 1107-8.

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

Spain, R.D. (2006). The effect of sonification pulse rate and system reliability on perceived workload, urgency, trust, and monitoring performance. Unpublished master's thesis, Old Dominion University, Norfolk, Virginia.

Stanton, N.A., & Edworthy, J. (1998). Human factors in auditory warnings. London: Ashgate.

Walker, B. N. (2002). Magnitude estimation of conceptual data dimensions for use in sonification. Journal of Experimental Psychology: Applied, 8, 211-221.

Walker, B. N. & Kramer, G. (1996). Mappings and metaphors in auditory displays: An experimental assessment. In Proceedings of the International Conference on Auditory Display (ICAD1996). Palo Alto, CA. pp. 71-74.

Walker, B. N., & Lindsay, J. (2005). Using virtual reality to prototype auditory navigation displays. Assistive Technology Journal 17(1), 72-81

Watson, M., & Sanderson, P. (2004). Sonification helps eyes-free respiratory monitoring and task time-sharing. Human Factors, 46, 497–517.

Watson, M., Sanderson, P., & Russell, W. J. (2004). Tailoring reveals information requirements: The case of anesthesia alarms. Interacting With Computers, 16, 271–293.

Webb, R. K., van de Walt, J., Runciman, W. B., Williamson, J. A., Cockings, J., Russell, W. J., et al. (1993). Which monitor? An analysis of 2000 incident reports. Anaesthesia and Intensive Care, 21, 529–542.

Williams, H.S. (Online). Microwave motion sensors for off-road vehicle velocity data and collision avoidance. Online: http://archives.sensorsmag.com/articles/1299/34_1299/main.shtml. Retrieved April 26, 2007.

Xiao, Y., & Seagull, F.J. (2000). An analysis of problems with auditory alarms: Defining the roles of alarms in process monitoring tasks. Proceedings of the IEA/HFES 2000 Congress. Santa Monica, CA: HFES.