High Time for Temporal Variation: Improving Sonic Interaction with Auditory Interfaces

Article *in* IEEE Instrumentation and Measurement Magazine · October 2021

DOI: 10.1109/MIM.2021.9549230

2 authors:

Liam Foley
McMaster University

Michael Schutz
McMaster University

High Time for Temporal Variation: Improving Sonic Interaction with Auditory Interfaces

Liam Foley and Michael Schutz

Listening is a powerful, under-appreciated, mechanism for acquiring information about the world. Changes in the hum of a car engine can offer an early clue about potential problems; the rustle of leaves informs us of weather conditions even before checking the forecast; slight fluctuations in the sound of a kettle tell us the water is almost boiled. These are but a few examples of important sonic interactions we have with our environment throughout the day.

Use of Electronic Auditory Interfaces

For much of our species' evolutionary history, sounds came from physical events such as rocks dropping, animals howling, or thunder clapping. Today our sonic interactions come increasingly through electronic devices such as smartphones, smart home devices, and wearable technology. When designing consumer interfaces in a competitive market, ergonomics plays a crucial role, which has given rise to an entire industry devoted to exploring innovative approaches for building products. Although designers spend a great deal of time and attention to the specific shapes and colors used in visual interfaces, less attention and care has been given to the specific sounds used in auditory interfaces [1].

Auditory interfaces have a vital role in human computer interactions, particularly in fast-moving and high consequence situations where users need to devote visual attention elsewhere. Unfortunately, as human factors appear relatively under-considered in their design, problems with their use are widespread—particularly in areas outside the scope of consumer-facing industries (mobile phones, web design, etc.). For example, their shortcomings are widely recognized in contexts ranging from medical devices [2] to transportation systems [3], [4] and monitors related to industrial processing [5]. In many cases, these issues trace back to problems with the specific types of sounds used.

Issues with Human Perception of Alarms

The most widely known auditory interfaces are auditory alarms, a critically important form of human-computer interaction designed for rapidly attracting attention to time-sensitive signals. Unfortunately, examples of problems with alarms are well documented across many industries [6]. For example, alarms are known to be problematic in aviation [4]—a field where visual attention is needed for tasks beyond monitoring notification systems. Perceptual problems such as inattentional deafness, when an alarm sounds but is not heard, can result in dangerous outcomes for aircraft pilots [7]. With environments ever increasing in complexity and dynamicity, the number of alarms triggered in each of these interfaces increases regularly. These increases have been reported in processing (nuclear plants, refineries and chemical plants) [5] and railway industries [3], among others. As healthcare is the domain in which these issues are documented the most extensively, we will summarize some key findings related to medical device sounds, offering useful insight pertinent to auditory alarms across many domains.

Healthcare Alarms as a Case Study

Problems with the alarms used in healthcare are numerous and have been extensively researched and widely discussed [8]. Consequently, we will focus on issues that can benefit the most from using more sophisticated sounds. One such issue is alarm fatigue, a general desensitization to alarms felt by hospital staff as a result of repeated exposure [2]. This can be partly attributed to the high alarm rates in hospitals, with reports of over three hundred alarms per patient per day [9]. Unfortunately, reducing the number of alarms is neither simple nor ideal. Given the high consequences of failure, devices typically employ a "better safe than sorry" approach—with liberal criteria for alarming to ensure potentially important information is broadcasted [10]. Additionally, many devices sound continually, not just when a dangerous threshold is reached, adding more sound to an already crowded sonic environment [9].

Beyond high alarm rates, doctors and nurses often struggle with basic issues of alarm recognition and differentiation such as masking, when more than one alarm sounds and renders another inaudible [11]. This results in potentially dangerous situations where some alarms are heard, while others are missed. Even in best case scenarios when alarms are audible to

1094-6969/21/$25.00©2021IEEE

users, different alarms are easily confused. Although there are numerous issues contributing to these problems, some issues stem directly from similarities in their acoustic design [12].

Although much of the research on problems with healthcare alarms focuses on staff issues, they hold obvious implications for patient well-being. High alarm rates contribute to excessive sound levels in hospitals well above recommended levels [13]. These loud environments impact patient sleep quality which can extend recovery time. In an even more serious sense, an FDA survey of device errors found over five hundred alarm related deaths between 2005 and 2008 [2], with anecdotes of power failure alarms being missed by staff, resulting in device malfunctioning and a patient dying [2], [9].

Human Constraints vs. Engineering Constraints

Given the widespread use of alarms, fixing all of their problems requires a multifaceted approach. One common problem across many non-speech alarms is their acoustic simplicity and uniformity. This approach is easy to describe mathematically (and therefore straightforward to synthesize) but poses significant problems from a human factors' perspective. For example, this intentionally simplistic design renders many non-speech alarms too similar for our perceptual system to safely differentiate. This simplicity also avoids taking advantage of the aspects of auditory processing that the human perceptual system evolved to handle elegantly.

Technical limitations acted as a constraint in early sound synthesis approaches; however, for many alarm applications this has not been the case for more than two decades [14]. Consequently, today's limitations are not *technical* but rather *historical* —the result of today's designers making choices that appear to be constrained by yesterday's technology. Curiously, video games regularly employ complex and intricate sound design, taking advantage of our auditory abilities to create immersive experiences [15]. Despite the critical importance of medical device alarms for patient well-being, few surpass the auditory complexity of those used in early video games such as "Pong." Consequently, the level of untapped sophistication available to alarm designers is enormous, matched only by the potential of public health and workplace safety gains due to acoustic improvements.

With well documented instances of fatalities stemming from missed auditory alarms [9], it might appear that sound is not ideal for communicating critical information. It is important to recognize, however, that the auditory system is quite adept at recognizing and understanding the meaning of a wide range of complex sounds. For example, we can quickly recognize potential problems with our vehicles from changes in their engine's hum [16]. Natural sounds are complex, and the auditory system has evolved to take advantage of this complexity to understand our environment [16]. For example, the spectrogram in Fig. 1a illustrates a variety of walkers' footsteps. Our ears are highly attuned to nuanced differences that might elude the eye in these visual representations. For example, perceptual research clarifies that through listening alone, we can reliably recognize a walkers' cadence [17], classify their gender [18], and even identify their shoe material [19]. Our remarkable ability to discriminate nuanced events by ear presents a curious contrast to the widely recognized difficulties in

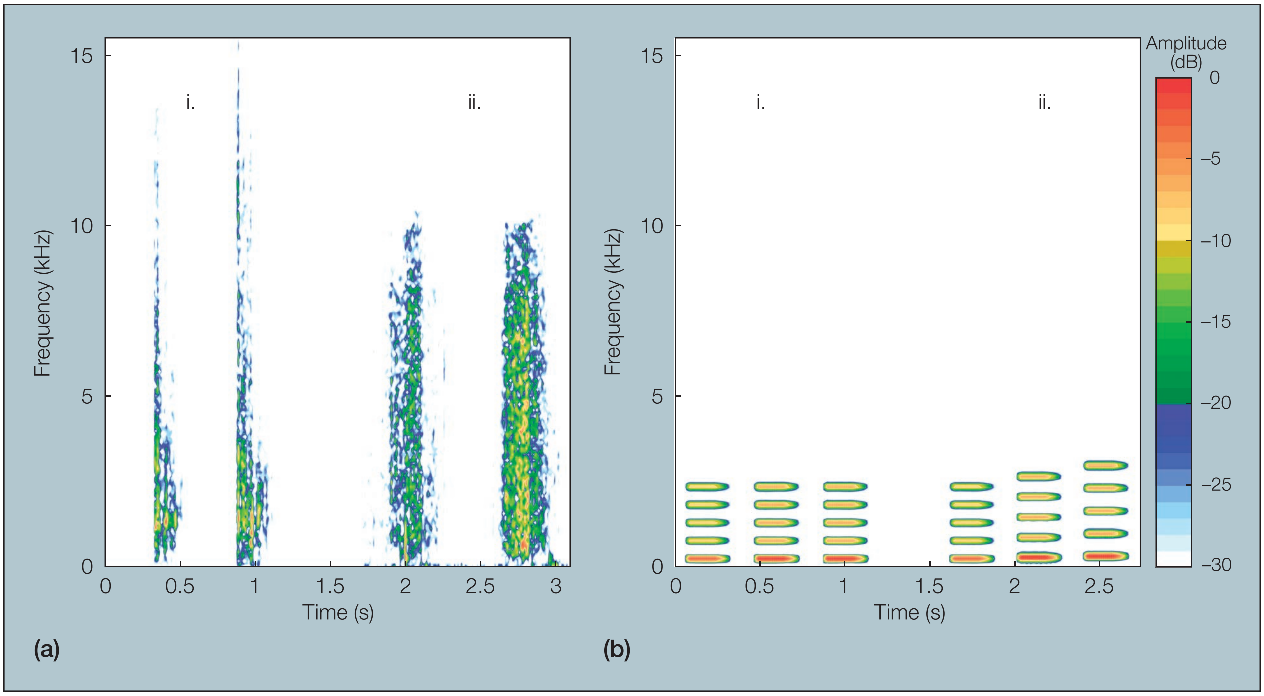

Fig. 1. (a) Spectrogram (or sonogram) of footsteps on i. gravel and ii. leaves. Note the broad range of spectral information, with numerous temporal differences and idiosyncrasies. (b) Spectrogram depicting the i. three note general alarm followed by the ii. three note temperature delivery alarm from the IEC:2006 Standard [20].

discriminating between current alarms [12]. This suggests that perhaps the problem is less with the auditory system's *general recognition and discrimination* of alarm sounds but rather recognition of the *specific types* of sounds used in auditory interfaces. In many ways we find this almost painfully ironic, given that most alarm sounds have been designed explicitly to communicate specific messages and critical information—in the absence of physical constraints on natural sound production.

Spectrogram depicting the i. three note general alarm followed by the ii. three note temperature delivery alarm from the IEC:2006 Standard [20].

Compared to the complexity of natural sounds, prototypical non-speech-alarms are surprisingly simplistic. Fig. 1b shows two subsequent alarms from a popular medical alarm standard [20]. Each alarm consists of three pitched tones creating a short three-note melody. The first alarm is a general alarm, while the second indicates an issue with patient temperature. Despite their design that is intended to signal unrelated events requiring completely different responses, the sounds are identical in timing, timbre, rhythmic pattern, and starting pitch. The clear organization and consistent tone structure seem sensible from an engineering perspective. Yet from a human-factors perspective, these "strengths" actually manifest as significant weaknesses. Their relative simplicity contrasts with the kinds of complexity inherent in natural sounds, leading to a range of problems. This is especially problematic for alarm signals, since the lack of temporal variability ends up making them more prone to masking [21] and makes them more annoying for users [22].

Lack of Temporal Variation in Auditory Sciences

Despite the simplicity of many non-speech alarms, our ability to differentiate and remember their associations is poor [12], in stark contrast to our capability for such tasks with complex sounds [23]. One clear difference between alarm sounds and natural sounds is the lack of temporal variation in most alarms—which ironically contributes to our remarkable ability to recognize natural sounds. This oversight is not constrained to alarm design—it reflects perspectives common in the auditory sciences as a whole [24]. Although temporally invariant sounds are easier for precise experimental control, this trade-off comes at the expense of taking for granted the impressive capabilities of the auditory system for parsing and processing complex sound. Extensive research illustrates that temporal variation, the very property ignored in most auditory interfaces, plays a crucial role in tasks such as integrating sight and sound, estimating event durations, and inferring physical properties of materials involved in certain events [24].

Current approaches to visualizing sound not only reflect *past* oversight regarding the importance of temporal structure, but also *disincentivizes* its consideration in the future. For example, Fig. 2 shows two types of figures commonly used to explain sound—waveforms (Fig. 2a), showing a summary of temporal changes for all components; and power spectra (Fig. 2b), summarizing the components themselves. For sounds such as i. sine tones, ii. sawtooth tones, and iii. medical alarm tones, these visualizations suffice. As these sounds lack components that change independently, the critical differences are adequately summarized. Visualizing sound in this way, however, disincentivizes thinking about the use of more complex sounds with components that do not move in lockstep. We find this ironic, given that the natural sounds we are so adept at processing contain considerable complexity.

Fig. 3 shows four different representations of single notes played on three instruments. Note how some visualizations obscure the differences between these instruments—which are clearly discriminable to our ears. For example, the cello

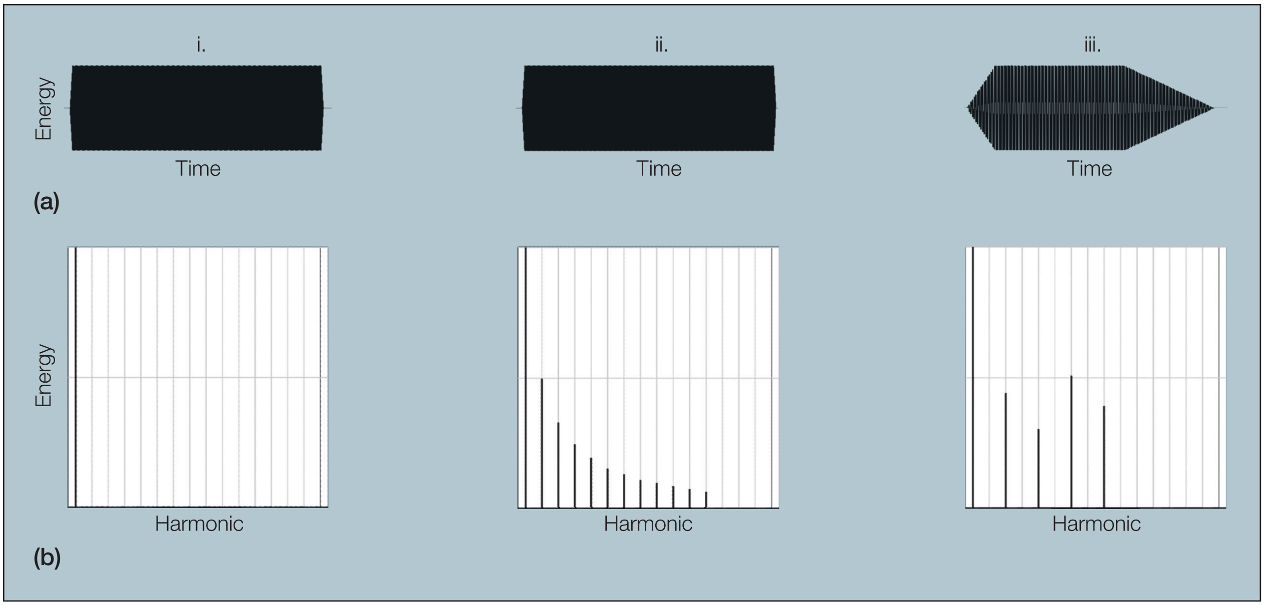

Fig. 2. (a) Waveforms and (b) Power spectra: of i. 1000 Hz sinewave; ii. 1000 Hz sawtooth tone, and iii. a single tone with a fundamental of 261 Hz from Power Failure alarm from IEC:2006 [20].

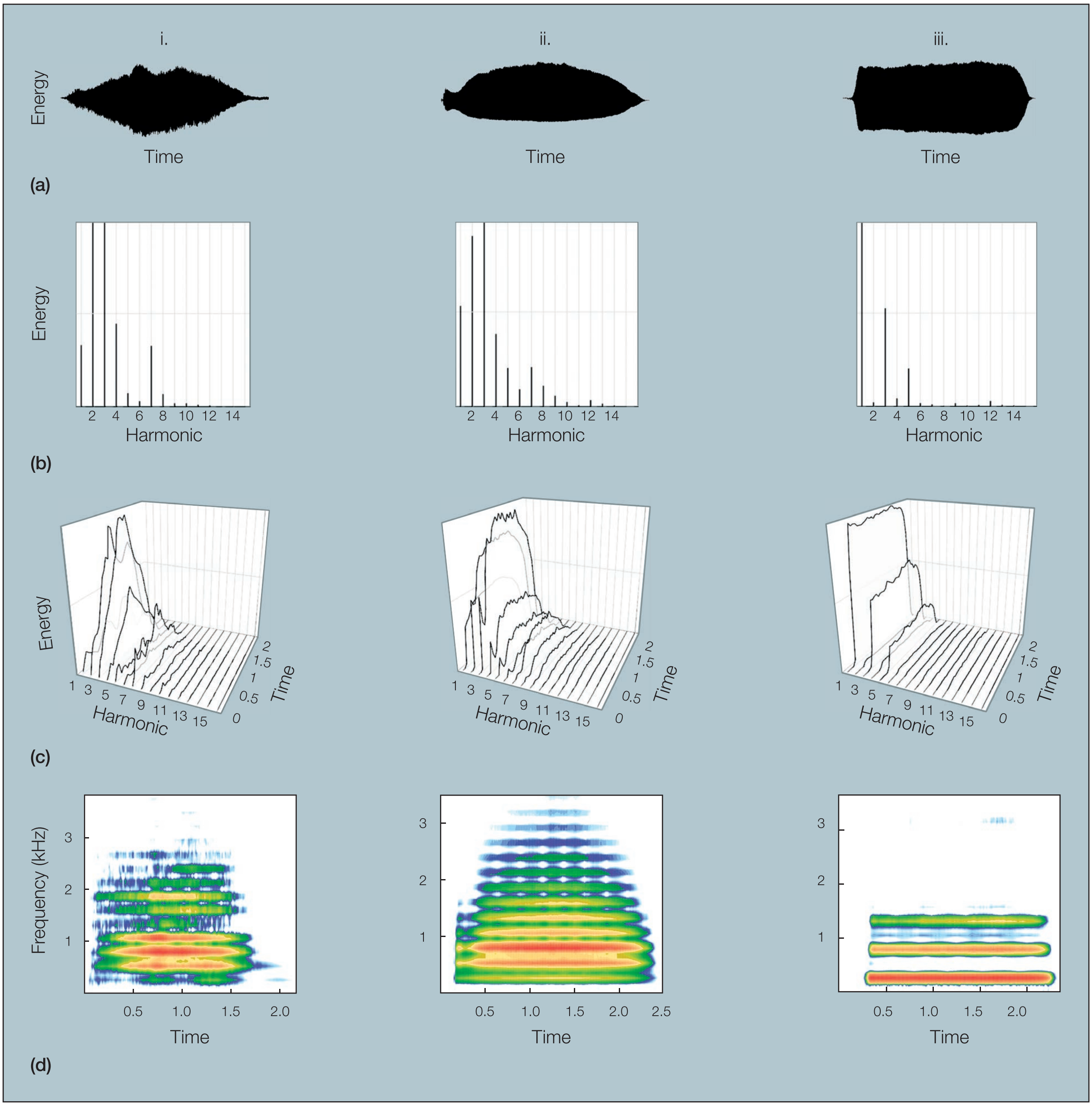

Fig. 3. Three columns are for i. cello, ii. trombone, and iii. clarinet, showing: (a) Waveform; (b) Power spectra (c) 3-D power spectra; and (d) Spectrogram.

and trombone have similar power spectra (Fig. 3b), even though the strength of each harmonic component follows a different temporal trajectory (Fig. 3c). The spectrogram (Fig. 3d) captures some aspects of this change, although they are not as clear as in the 3D visualization of each component in Fig. 3c.

Although the cello and trombone are rarely confused when listening, visualizations using power spectra are difficult to differentiate. This illustrates problems with the overuse of simplified sound visualizations—which are common in both in introductory textbooks as well as auditory research approaches in top journals [24]. We believe the challenge in visually representing the importance of sounds' temporal complexity has led to an underappreciation of its crucial role in auditory perception amongst the designers of auditory interfaces (see [25] for a discussion that advocates for more realistic sounds in auditory experiments). Consequently, greater awareness of our ability to detect and use temporal changes in acoustic structure can lead to useful new developments in a potentially valuable way to improve auditory interfaces in a range of technical devices.

Although visualizations such as the spectrogram (Fig. 3d) do indicate where energy is most concentrated across a sound's spectrum, as can be seen comparing the spectrogram of the cello and trombone with the 3-D Power Spectra, and the characteristic changes in energy are difficult to pinpoint when looking at the heatmap of the spectrogram. Greater emphasis on visualizing the vast complexity in the temporal structure of sound will focus more attention to the vital role it plays in sound perception.

Temporal Variation in Alarms

More attention to the role that temporal variation plays in our perception and auditory cognitive processes can open new avenues for designing more distinguishable and ergonomic alarms. Although different areas of alarm implementation will necessitate careful analysis to ensure optimal usability, embracing temporal variability holds great promise—particularly with respect to the alarms used in medical devices [26]. The complexity that can be found even in single musical tones holds important lessons that can be invaluable in generating new ideas for auditory interface sounds [27].

One alarm problem that we think can benefit from temporal variability is annoyance. Investigations have shown that changing prototypical invariant alarms (Fig. 2, iii.) to temporally variable alarms, whether by adding an exponentially decaying or complex envelopes seen in musical sounds (Fig. 3), can decrease annoyance without harming learning or recognition [22], [28].

Our team illustrated this in an experiment designed to assess the benefits of introducing even slight temporal complexity to otherwise standard alarm tones. We synthesized two sets of alarm tones: one temporally variant in envelope and the other invariant. The temporally invariant sounds had envelopes similar to those seen in Fig. 2, while the temporally variable alarms had exponentially decaying envelopes—similar to those produced by musical instruments such as the marimba, piano, or guitar. We found no difference in learning or recognition of the alarm sounds; however, participants rated temporally variant alarms as significantly less annoying in each of three independent experiments. This illustrates that even slight changes in the temporal profile of a sound can significantly decrease annoyance relative to standard temporally invariant tones, without negative affects to recognition of alarm sequences [22].

Although annoyance may seem a trivial concern, it is important to remember that high alarm rates are documented in many settings, from healthcare [9] to processing industries [5]. In healthcare, high alarm rates are a leading reason for hospital staff switching off annoying alarms [29]. This has unfortunately resulted in patient deaths due to alarms being turned off or their volume turned down [2], [9]. It is unlikely that the prevalence of alarms will decrease, as more technology is continually introduced to monitor important developments. If alarms are unavoidable, reducing their annoyance holds important benefits. Embracing temporal variation offers one clear path to create alarms that are less annoying without sacrificing their communicative function.

Conclusion

We propose that future alarm sounds be designed with greater attention to the *abilities of the system processing them* (i.e., human perception) rather than the *simplicity of the devices producing them*. Although there are many unique problems with specific alarm implementations, embracing temporal variability offers an under-explored path to improvement. In many domains, the sheer number of alarms is unlikely to be reduced due to increases in the number of problems and conditions that should be monitored. Therefore, temporal variation promises a relatively simple approach to reduce annoyance, without having to change other alarm system parameters such as alarm rate, safe thresholds, etc. Additionally as introducing temporal variation to individual tones does not impair recognition of tone sequences, these more complex alarm sounds can be backwards compatible with invariant melodic alarms.

In the early days of sound synthesis, technical constraints precluded complex temporal variation. Today, those engineering constraints no longer apply. We can easily and inexpensively generate a near-infinite range of sounds for use in auditory interfaces. If birthday cards can feature musical passages, why do so many highly advanced machines communicate using sounds that appear stuck with limitations from decades past? Ironically, when it comes to sounds such as alerts on smartphones or in video games, designers routinely use a wide array of temporally varying, complex sounds. Yet, alarms critical to safety, which signal crucial information with life-or-death consequences, routinely employ sounds no more complex than a garbage truck backing-up!

Rather than mere "bells and whistles" [30], auditory interfaces are integral to safe operations in many areas, often providing a first warning sign to potential hazards. The human auditory system can differentiate between numerous complex sounds and does so on a regular basis when interacting with the world. Careful attention to the types of sounds we are adept at differentiating in general, and the ways musicians have chosen to shape and create sound in particular, offers a rich source of inspiration for designing more ergonomic and efficient auditory interfaces [26], [27]. We believe greater collaboration between auditory perception researchers, sound designers, musicians, engineers, and industry-specific experts will lead to promising new approaches to creating more effective alarms and auditory interfaces—as well as safer and more pleasant environments.

References

[1] M. Jeon, "Auditory user interface design: practical evaluation methods and design process case studies," *Int. J. Des. Soc.*, vol. 8, no. 2, pp. 1–16, 2015.

[2] M. Cvach, "Monitor alarm fatigue: an integrative review," *Biomedical Instrum. Technol.*, vol. 46, no. 4, pp. 268–277, 2012.

[3] J. Edworthy, E. Hellier, K. Titchener, A. Naweed, and R. Roels, "Heterogeneity in auditory alarm sets makes them easier to learn," *Int. J. Ind. Ergon.*, vol. 41, no. 2, pp. 136–146, Mar. 2011.

[4] J. P. Bliss, "Investigation of alarm-related accidents and incidents in aviation," *Int. J. Aviat. Psychol.*, vol. 13, no. 3, pp. 249–268, 2003.

[5] J. C. Laberge, P. Bullemer, M. Tolsma, and D. V. C. Reising, "Addressing alarm flood situations in the process industries through alarm summary display design and alarm response strategy," *Int. J. Ind. Ergon.*, vol. 44, no. 3, pp. 395–406, 2014.

[6] R. D. Patterson, "Auditory warning sounds in the work environment," *Philos. Trans. R. Soc. London*, vol. 327, no. 1241, pp. 485–492, 1990.

[7] F. Dehais *et al*., "Missing critical auditory alarms in aeronautics: evidence for inattentional deafness?" *Proc. Hum. Factors Ergon. Soc*., vol. 2, pp. 1639–1643, 2012.

[8] J. Edworthy, "Medical audible alarms: a review," *J. Am. Med. Informatics Assoc*., vol. 20, no. 3, pp. 584–589, 2013.

[9] S. Sendelbach and M. Funk, "Alarm fatigue: a patient safety concern," *AACN Adv. Crit. Care*, vol. 24, no. 4, pp. 378–386, 2013.

[10] J. Edworthy and E. Hellier, "Fewer but better auditory alarms will improve patient safety," *Qual. Saf. Heal. Care*, vol. 14, no. 3, pp. 212–215, Jun. 2005.

[11] P. Lacherez, E. L. Seah, and P. Sanderson, "Overlapping melodic alarms are almost indiscriminable," *Hum. Factors*, vol. 49, no. 4, pp. 637–645, 2007.

[12] P. Sanderson, A. N. Wee, and P. Lacherez, "Learnability and discriminability of melodic medical equipment alarms," *Anaesthesia*, vol. 61, no. 2, pp. 142–147, 2006.

[13] A. J. Oleksy and J. J. Schlesinger, "What's all that noise—improving the hospital soundscape," *J. Clin. Monit. Comput*., vol. 33, no. 4, pp. 557–562, 2019.

[14] J. Edworthy and R. Hards, "Learning auditory warnings: the effects of sound type, verbal labelling and imagery on the identification of alarm sounds," *Int. J. Ind. Ergon*., vol. 24, no. 6, pp. 603–618, 1999.

[15] I. Salselas, R. Penha, and G. Bernardes, "Sound design inducing attention in the context of audiovisual immersive enviornments," *Pers. Ubiquitous Comput*., vol. 4, no. 4, pp. 1–12, 2020.

[16] W. W. Gaver, "What in the world do we hear? an ecological approach to auditory event perception," *Ecol. Psychol*., vol. 5, no. 1, pp. 1–29, 1993.

[17] W. Young, M. Rodger, and C. M. Craig, "Perceiving and reenacting spatiotemporal characteristics of walking sounds," *J. Exp. Psychol. Hum. Percept. Perform*., vol. 39, no. 2, pp. 464–476, 2013.

[18] X. Li, R. J. Logan, and R. E. Pastore, "Perception of acoustic source characteristics: walking sounds," *J. Acoust. Soc. Am*., vol. 90, no. 6, pp. 3036–3049, Dec. 1991.

[19] B. L. Giordano, H. Egermann, and R. Bresin, "The production and perception of emotionally expressive walking sounds: similarities between musical performance and everyday motor activity," *PLoS One*, vol. 9, no. 12, pp. 1–23, 2014.

[20] *Medical electrical equipment — Part 1-18: General requirements for basic safety and essential performance — Collateral standard: Requirements for medical electrical equipment and medical electrical systems used in the home healthcare environment*, International Standard IEC 60601, International Electrotechnical Commission, 2006.

[21] J. Edworthy *et al*., "The recognizability and localizability of auditory alarms: setting global medical device standards," *Hum. Factors*, vol. 59, no. 7, pp. 1108–1127, Nov. 2017.

[22] S. Sreetharan, J. J. Schlesinger, and M. Schutz, "Decaying amplitude envelopes reduce alarm annoyance: exploring new approaches to improving auditory interfaces," *Appl. Ergon,* vol. 96, 2021.

[23] C. Suied, T. R. Agus, S. J. Thorpe, N. Mesgarani, and D. Pressnitzer, "Auditory gist: recognition of very short sounds from timbre cues," *J. Acoust. Soc. Am*., vol. 135, no. 3, pp. 1380–1391, Mar. 2014.

[24] M. Schutz and J. Gillard, "On the generalization of tones: a detailed exploration of non-speech auditory perception stimuli," *Sci. Rep*., vol. 10, no. 1, p. 9520, 2020.

[25] J. G. Neuhoff, *Ecological psychoacoustics*. Amsterdam, The Netherlands: Elsevier Academic Press, 2004.

[26] L. Foley, C. J. Anderson, and M. Schutz, "Re-sounding alarms: designing ergonomic auditory interfaces by embracing musical insights," *Healthcare*, vol. 8, no. 4, p. 389, Oct. 2020.

[27] M. Schutz, "Acoustic structure and musical function: musical notes informing auditory research," in *The Oxford Handbook on Music and the Brain*, M. H. Thaut and D. A. Hodges, Eds. Oxford, UK: Oxford University Press, pp. 145–164, 2019.

[28] C. Anderson, "Examining melodic alarm recognition accuracy across contrasting timbres," B.Mus. (Hons) Thesis, Department of Psychology, Neuroscience and Behaviour, McMaster University, Hamilton, Ontario, Canada, 2019.

[29] F. E. Block, L. Nuutinen, and B. Ballast, "Optimization of alarms: a study on alarm limits, alarm sounds, and false alarms, intended to reduce annoyance," *J. Clin. Monit. Comput*., vol. 15, no. 2, pp. 75–83, 1999.

[30] W. W. Gaver, "Auditory Interfaces," in *Handbook of Human-Computer Interaction*, pp. 1003–1041, 1997.

Liam Foley (ltimfoley@protonmail.com) is a Master of Science degree student in the Department of Psychology, Neuroscience, and Behaviour at McMaster University, Hamilton, ON, Canada. His research explores ways to make auditory alarms more ergonomic and intuitive. By combining research from auditory perception, music perception, and human factors, he hopes to create better sounding auditory environments.

Michael Schutz is Associate Professor of Music Cognition/Percussion at McMaster University, Hamilton, ON, Canada. He has been designated a "University Scholar" in recognition of his innovative merging of music performance and perception. His work is funded by grants from NSERC, SSHRC, and CFI, and his team's research appears in numerous textbooks. Dr. Schutz received an Early Researcher Award from Ontario, as well as the prestigious 2019 Alumni Award from the Penn State School of Music.