Parncutt, R., & Sattmann, S. (Eds.) (2018). *Proceedings of ICMPC15/ESCOM10.* Graz, Austria: Centre for Systematic Musicology, University of Graz.

# Designing Effective Auditory Interfaces: Exploring the Role of Amplitude Envelope

Sharmila Sreetharan,[1] Joseph Schlesinger,[2] and Michael Schutz[3,1]

[1] *Psychology, Neuroscience & Behaviour, McMaster University, Canada*
[2] *Vanderbilt University Medical Center, Vanderbilt University, USA*
[3] *School of the Arts, McMaster University, Canada*

[1]sreets@mcmaster.ca, [2]joseph.j.schlesinger@vanderbilt.edu, [3,1]schutz@mcmaster.ca

## Abstract

A sound's shape over time or the amplitude envelope affects our cognitive and perceptual processing of sounds. Previous work has suggested that percussive tones, sounds with (i.e., naturally or exponentially decaying) amplitude envelopes, facilitate better encoding and retrieval of sounds than flat tones, sounds with (i.e., amplitude invariant) envelopes, during associative memory tasks. Flat tones are errantly used by most auditory interfaces, including the auditory alarms used by the global medical device International Electrotechnical Commission (IEC) 60601-1-8. The IEC alarms are criticized being difficult to be learnt and identified, easily confused, and are generally perceived as annoying. We manipulated the amplitude envelope of tones otherwise conforming to the IEC standard to explore its effect on learnability and annoyance. Forty participants were randomly assigned to learn alarm-referent associations composed entirely of either flat envelopes or percussive envelopes. Learning was measured by the number of training blocks required to learn the alarm-referent associations and the number of correctly identified alarms after a short break. Participants then evaluated alarm annoyance using a two-alternate forced task where trials varied by either alarm or envelope. Our results percussive tone sequences suggest reduce perceived annoyance without harming learning. These results highlight the importance of temporal considerations when creating auditory interfaces to reduce negative effects of prolonged exposure.

## Introduction

### Auditory Interfaces

Auditory interfaces rapidly convey information from machine to user using either speech or non-speech sounds and are particularly beneficial when visual interaction with devices is limited (Sodnik & Tomažič, 2015). From using Morse code during the 1800's to low battery levels and incoming email alerts on modern mobile devices, auditory interfaces have been employed for centuries using speech or non-speech sounds. Speech interfaces provide naturalness of communication and allow for multitasking during communication by requiring fewer attentional resources (Rosenfield, Olsen & Rudnicky, 2001). However, they are limited by language barriers, low robustness and accuracy in loud environments, a lack of privacy, and are slow forms of communication (Sodnik & Tomažič, 2015). Alternatively, non-speech interfaces use synthesized sounds conveying information using metaphors (i.e., auditory icons), abstract/symbolic sounds (i.e., earcons), or sped up spoken phrases no longer resembling speech (i.e., spearcons; Gaver, 1986). They can beneficially protect privacy, be universally understood, and rapidly conveyed (Gaver, 1986). However, unlike visual interfaces, it is difficult to create a set of concrete, intuitive metaphors (Sodnik & Tomažič, 2015).

### Auditory Alarms

Non-speech interfaces are ubiquitous, but one prominent example is the use of auditory alarms in hospitals. The global medical device the IEC 60601-1-8 uses a standardized set of alarms to monitor patients and machine-related problems. These alarms are composed of three-note (i.e., medium priority) or five-note (i.e., high priority) melodic tone sequences with mnemonics (i.e., auditory icons) to aid in learning alarm-referent associations. Despite careful consideration by Block, Rouse, Hakala and Thompson (2000) on tone design specifications and mnemonics and Patterson (1990) on creating ergonomic auditory warnings in the creation of the alarm set, these alarms are riddled with problems. One study evaluating critical care nurses' abilities to learn these alarm-referent associations found only one nurse able to successfully identify 100% of associations after two training sessions, and that mnemonics did not improve identification association accuracy (Wee & Sanderson, 2008).

Overall, these alarms are frequently critiqued for being difficult to learn, easily confused, and poorly retained (Edworthy & Hellier, 2005; Sanderson, Wee & Lacherez, 2006). A large contributing factor in the poor discriminability and learnability of the IEC alarms is their lack of heterogeneity within the alarm set (Edworthy, Hellier, Titchener, Naweed & Roels, 2011). All tone sequences were synthesized within a narrow frequency range (262-523 Hz) in the key of C major, having similar contours, durations, rhythmic patterns, and starting on the note "C" (Edworthy et al., 2011). The large amount of similarity shared between the tone sequences is likely to contribute to alarm misidentification. Together, these identification errors contribute to the larger problem of alarm fatigue. Alarm fatigue results from sensory overload triggered by excessive alarms, lowering sensitivity and leading to missed alarms (Sendelbach & Funk, 2013). This negatively affects patient care due to the number of missed alarms and the noise pollution affecting recovery rates (Darbyshire & Young, 2013).

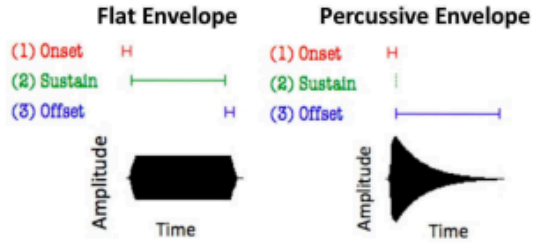


**Figure 1. Sample wave forms of flat and percussive amplitude envelopes.**

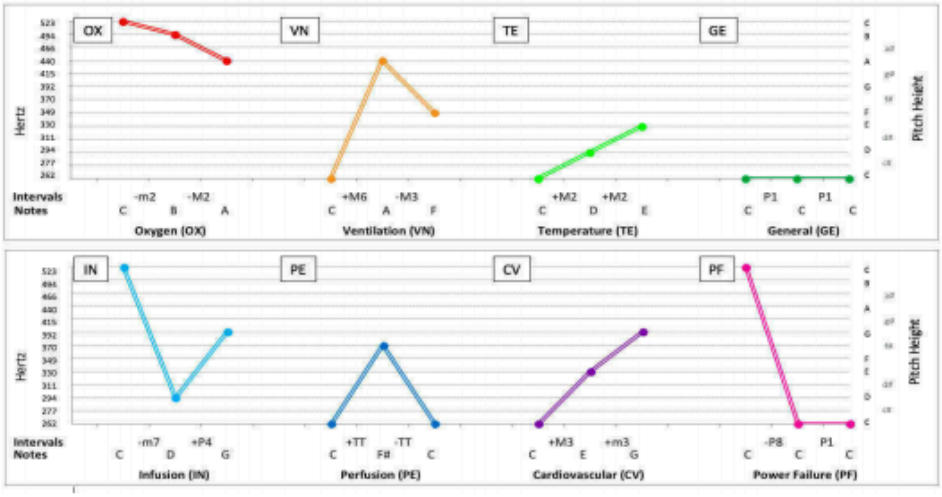


**Figure 2. Contours of the medium priority IEC 60601-1-8 tone sequences used in the experiment. M = Major, m = Minor, P= Perfect, TT = Tritone, + = Ascending, - = Descending.**

### Amplitude Envelope

Amplitude envelope refers to a sound's temporal structure, specifically in reference to changes in a sound energy (i.e., intensity or loudness) over time. Flat envelopes are characterized by a rapid onset followed by a period of sustain, and a rapid offset, exhibiting minimal changes in a sound's energy (Figure 1A). Flat envelopes are characteristic of most artificially synthesized sounds, such as the emergency broadcast tone. Conversely, naturally produced sounds have percussive envelopes characterized by a rapid onset followed by an immediate, exponential decay (Figure 1B). Percussive envelopes are characteristic of everyday sounds including impact sounds, like the sound produced by two wine glasses clinking.

Amplitude envelope (hereafter referred to as "envelope") plays an essential role in giving details about impact events. For example, Lutfi (2001) demonstrated that the physical attributes of struck objects can be detected based on the varying acoustical features of percussive envelopes. Lutfi (2001) used a forced-choice task to determine specific acoustical features used in identification of a bar's hollowness (solid or hollow) and its material composition (iron, wood, or aluminium). Listener strategies that involved using envelope information aided in optimal performance levels (Lutfi, 2001). These findings, further extended by Giordano, Rocchesso, and McAdams (2010), suggest that envelope is a salient cue in determining the object hardness perception.

Envelope also plays an important role in audiovisual integration. Schutz and Lipscomb (2007) demonstrated in their "marimba illusion" that the length of a striking gesture can influence the perceived duration of percussive sounds produced by a marimba. This suggests visual information largely influences judgements of tone duration tasks, despite a popular notion that audition dominates such judgements (Fendrich & Corballis, 2001). A follow-up study by Schutz (2009) suggests this atypical pattern of integration stems from the amplitude envelope of percussive tones. When using flat tones, audition dominates tone duration judgements but when using percussive tones, visual dominates tone duration judgements. The over-reliance on flat envelopes in auditory perception research (Schutz & Vaisberg, 2014) seems to have led to conclusions from artificial tone beeps that fail to generalize to natural sounds.

Additionally, it is suggested that percussive sequences are rapidly processed compared to flat sequences during associative memory tasks (Schutz, Stefannuci, Baum & Roth, 2017). Schutz et al. (2017) examined the effect envelope on tone sequence-object associations and found that percussive tone sequence-object associations were learned more rapidly, were correctly recalled more frequently than flat tone-sequence-object associations. This suggests that amplitude envelope may play an important role in associative memory, especially during encoding and retrieval stages (Schutz et al., 2017).

### Current Experiment

Like most non-speech interfaces, the auditory alarms used by the IEC 60601-1-8 exclusively uses tone beeps (i.e., tone sequences with flat envelopes). Our experiment aims to explore the potential benefits of amplitude envelope manipulation on the IEC alarm tone sequences while preserving the current alarm-referent associations. Specifically, we examined the effect of incorporating percussive envelopes into the IEC medium priority alarm tone sequences to assess the effect on learnability and annoyance.

## Method

### Participants

Forty undergraduate McMaster University students (35 female, 5 male) ranging in age from 18-23 years ($M$ = 19.08, $SD$ = 1.40) recruited from undergraduate psychology and

linguistic research pools participated in this study and received 1 research course credit as compensation.

### Stimuli and Apparatus

We synthesized eight tone sequences according to the frequency and amplitude envelope requirements in IEC 60601-1-8 standard for medium priority alarms using SuperCollider – a real time audio synthesis software (Figure 2). All tone sequences were within the range of 261.4 Hz (C3) - 523.4 Hz (C4). Unlike the IEC standards, here we used pure tones (sine waves) rather than more complex tones (an issue we will explore in future experiments). "Flat alarms" exhibited flat envelopes conforming to the temporal constraints outlined by the IEC standards (Figure 3A). "Percussive alarms" exhibited overlapping percussive envelopes (Figure 3B); since the envelopes overlapped for percussive alarms, the peaks amplitudes were set to be equivalent as this was sufficient to achieve an approximate equivalent perceived loudness for the two tone types. The total time sweep was 1040 ms for the flat alarms and 1180 ms for the percussive alarms. The last tone for the percussive alarms was approximately 1.6 times longer (after the onset) than the last tone for the flat alarms; this accounted for the shorter perceived duration of percussive envelopes and achieved an approximate equivalent perceived duration for the two tone types (Grassi & Darwin, 2006).

Tone sequences were appended together using Audacity – a free sound-editing software and stored on a MacBook Air laptop. Participants listened to these sounds through Sennheiser HDA-300 headphones. In addition, participants completed a consent form and a short demographic survey inquiring demographic information and musical background. Participants received cue cards with each of the eight alarm names and were permitted to re-arrange them to aid them in learning associations.

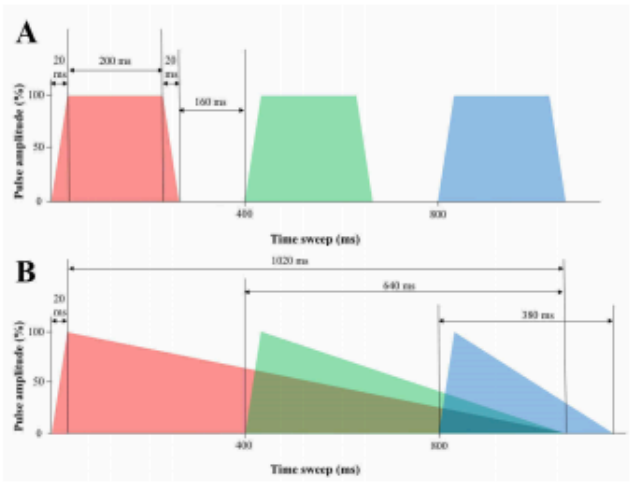


**Figure 3. Temporal characteristics of flat (A) and percussive (B) tone sequences used in the experiment. For simplicity, exponential tone onsets and offsets are depicted to be linear.**

### Procedure

The experiment consisted of two separate assessments: the first examined learning of sequence-command associations and the second annoyance with the sequences themselves (Figure 4). We randomly assigned participants to either the "flat" or "percussive" conditions for the learning paradigm, dictating whether they heard "flat" or "percussive" alarm-referent pairings throughout the entirety of the learning paradigm. The learning paradigm was split into four stages: study phase, training phase, break, and evaluation phase.

**Study phase.** Participants heard the eight tone sequences-referent pairings twice. The alarms were presented in a random order to prevent order effects.

**Training phase.** During the training phase, participants heard an alarm and verbally indicated the referent. Regardless of their response, participants were reminded of the correct alarm-referent pairing. The end of the training phase occurred when participants correctly identified 7/8 alarms on two consecutive blocks (each block containing all alarms in the alarm set) to a maximum of ten blocks.

**Break.** After, participants had a five-minute break when they played a silent game of mini-putt, which was implemented to prevent mental rehearsal.

**Evaluation phase.** Participants completed a final identification task on one block of alarms during the evaluation phase in addition to reporting the confidence of their response on a 6 point Likert-type scale. Up to this point in the experiment, participants only heard sounds with one type of envelope (percussive/flat).

After the learning paradigm, participants completed an annoyance task when they identified alarm they perceived to be more annoying during a two-alternative forced choice task. The task contained 16 trials—with eight trials differing only in envelope type and eight trials differing only in alarm type. For example, trials comparing the percussive-general vs. flat-general alarms forced assessment based on envelope, whereas trials comparing the percussive-temperature vs. the percussive-cardiovascular alarms forced assessment based on melodic structure. All participants heard each alarm-envelope combination twice throughout the task and all trial types were mixed and randomized during the task. As the trials comparing alarm types served primarily as distractors to reduce task transparency, here we report only analysis of trials comparing envelopes.

## Results

### Learning Paradigm

A between-subjects t-test conducted on the number of blocks to reach criterion based on condition during the training phase revealed a non-significant effect of condition, $t(38)=0.52$, $p>.05$. A between-subjects t-test conducted on the number of correctly identified alarms based on condition during the evaluation phase revealed a non-significant effect of condition $t(38)=1.94$, $p>.05$.

### Annoyance Task

**Overall.** A one-sample chi-square goodness-of-fit test revealed flat alarms are more annoying than percussive alarms, $X^2(1, N=320) = 99.01$, $p<.05$; $w=.56$ (Figure 5A).

**Percussive alarm training.** A one-sample chi-square goodness-of-fit test revealed participants training on percussive alarms found flat alarms considerably more annoying, $X^2(1, N=160) = 115.60$, $p<.05$; $w=.85$ (Figure 5B). Furthermore, a series of one-sample chi-square goodness-of-fit tests with Holm-Bonferroni corrections indicate that participants found flat versions of each alarm type (e.g., general alarm, power failure alarm) more annoying than their respective percussive counterparts (all $p$'s$<.05$).

**Flat alarm training.** A one-sample chi-square goodness-of-fit test revealed participants training on flat alarms also found flat alarms considerably more annoying $X^2(1, N=160) = 11.03$, $p<.05$; $w=.26$ (Figure 5C). A series of one-sample chi-square goodness-of-fit tests were conducted assessing whether the flat or percussive versions of each alarm type were equally perceived to be annoying for those trained on flat alarms. All tests were non-significant, suggesting that annoyance ratings for flat and percussive versions of all alarm types are not statistically different (all $p$'s$>.05$).

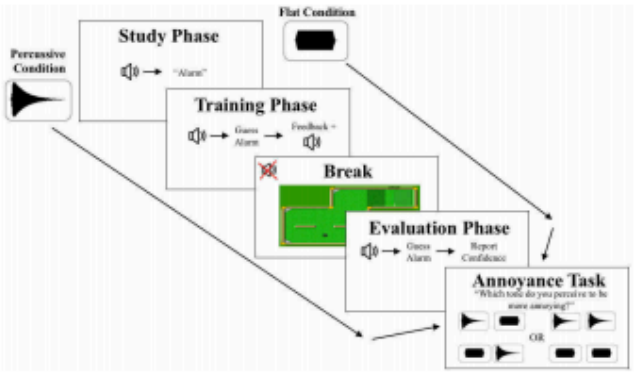


**Figure 4. Overview of experimental paradigm. Condition assignment was reflective of envelope type of the sounds heard during the learning paradigm (i.e., study phase, training phase, break, evaluation phase). The annoyance task, common to both experimental conditions, was comprised of sounds with both percussive/flat envelopes.**

## Discussion

### Learning paradigm

Performance during the learning paradigm suggests that our percussive tone sequences do not affect training required to learn alarm-referent associations nor their recall rate after a short break. We failed to replicate the results by Schutz et al. (2017) suggesting percussive tone sequences aided learning and recall of tone sequence-object associations. A distinguishing feature between our experiments is that Schutz et al. (2017) measured associative memory using concrete objects, whereas the alarm-referent associations we used were comprised of abstract commands/states. According to the dual-coding theory during encoding and retrieval, concrete words use both verbal and multi-modal sensory systems whereas abstract words only use the verbal system (Paivio, 1991). Due to the abstractness of alarms, it is possible that a greater training period would be required for any observable differences between the flat and percussive alarms.

### Annoyance task

Annoyance task performance suggests that flat alarms are more annoying than percussive—regardless of sounds heard during training. However, those trained on percussive alarms found flat alarms particularly annoying. Breaking down responses by eight alarms types, all of the flat versions were chosen as more annoying by 80%-100% of participants. Consequently, the flat version of the eight alarms appears significantly more annoying than the percussive—despite each participant providing a single rating per alarm type.

Those trained on flat alarms still found flat more annoying, although to a lesser degree than when analyzing ratings of those trained on percussive alarms. When broken down by the eight alarm types, we found no statistically significant differences based on envelope for any individual alarm. Nonetheless, flat versions were chosen to be more annoying in by 50%-70% for individual alarm sequences. Together these findings suggest percussive alarms are generally less annoying than flat, although the degree of difference varies based on exposure and training.

### Novelty effect

We believe the effect of training on annoyance ratings to reflect novelty's effect on annoyance ratings. When trained on flat alarms and exposed to the percussive alarms, participants are newly exposed to overlapping tone sequences, characteristic of our set of percussive tone sequences. It is possible that this overlapping tones may accentuate consonance and dissonance Western harmonic structure of the tone sequences. According to Western musical classifications of consonance and dissonance, simple frequency ratios (e.g., unison, major third, and perfect fifth intervals) are considered consonant whereas complex frequency ratios (e.g., minor second, tritone, minor seventh intervals) are considered dissonant (Schellenberg & Trehub, 1994). We noticed a slight trend towards percussive versions of alarms with dissonant intervals being rated as more annoying than consonant intervals. When individual notes in the tone sequence are separate, it is likely the interval relationships go unnoticed, especially for those lacking musical training. Since we did not control for musical training, it is possible that varying amounts of Western musical training may have contributed to the observed pattern of results.

### Implications

Our envelope manipulation had no effect on learnability; however, it did affect short-term annoyance ratings of the alarm tone sequences. Together these results suggest there is great potential benefit to exploring more varied envelope shapes when designing auditory alarms in medical devices. Although learnability is a concern when introducing medical

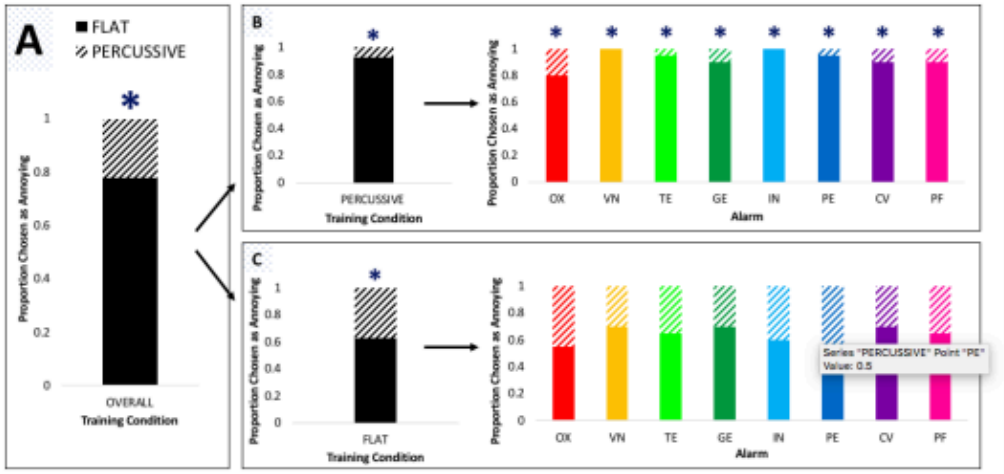


**Figure 5. Annoyance task results for all participants (A), broken down by those trained on percussive alarms (B) or flat alarms (C) during the learning paradigm. *$p < .05$**

professionals to current alarms, alarm annoyance is a long-term persisting problem affecting both clinicians across a career and patients over the duration of their hospital stay.

**Auditory alarms.** An observational study found that on average there about two critical alarms missed per hospital, per day in the United States (Donchin et al., 2003). Extrapolating this number across every hospital in entire country over the span of one year roughly equates to about 4.2 million alarms missed annually in the US. Edworthy (1998) suggests as alarms become increasingly irritating, there is a higher likelihood of the device being shut down or severe volume reduction. Although there are many contributing factors to missed alarms, it is likely that alarm fatigue is one (Cvach, 2012). Currently, there are implementations in practise to reduce alarm fatigue, yet there are few strategies in place to reduce alarms missed critical alarms (Rayo & Moffatt-Bruce, 2015). Our envelope manipulation offers one solution to reduce annoyance in hopes of reducing the number of missed critical alarms.

**Auditory Interfaces.** Although the advancement of technology has led to widespread use of higher quality speakers and tone generators, the sounds used in medical devices routinely fail to take full advantage of these possibilities. Presently, many auditory interfaces employ "tone beeps" that become incessant over time. The most basic example to illustrate this is a cashier who receives auditory feedback of a "beep" each time an item passes over the scanner. These sounds create a cacophony of beeps between cashiers become both annoying and confusing (Edworthy, 1998). Amplitude envelope manipulation can create more aesthetically pleasing sounds with greater heterogeneity. One study evaluating effect on envelope on product value found that participants were willing to pay more for cell phones with percussive ring tones than flat ring tones (Schutz & Stefanucci, under review). Despite the practical applications, those designing the alarms are focused on the nuances of the machine itself and the importance of the design of the sounds is lost. Our results emphasize the importance of temporal considerations of sounds in the creation of effective auditory interfaces.

## Future Directions

Our current experimental design did not take into account the effect of variable alarm exposure on annoyance due to the variable amount of blocks required to learn the alarm-referent associations. It is likely that exposure is positively correlated with annoyance; future experiments will control for amount of exposure prior to assessments of annoyance. Furthermore, we did not assess annoyance prior to the beginning of the experiment. It is possible that annoyance ratings may have been influenced by pre-existing biases. It is unclear whether individuals with high exposure to sounds with flat envelopes present reduced or greater sensitivity to such sounds affecting their annoyance ratings. We aim to account for this by assessing annoyance prior to the beginning of the experiment to understand how envelope manipulation may influence any pre-existing biases. In consideration of ecological validity, sounds with acoustically richer structure should be implemented to combat masking from environmental noise (Patterson, 1990). Our stimuli consisted of pure tones to determine how effective amplitude envelope manipulation is at the most basic level. In the future we plan on using complex tones with greater number of harmonics in subsequent experiments.

To the best of our knowledge, these results represent a novel approach to improving alarm design with great potential to assist the larger problem of auditory alarm annoyance in medical settings. As such, we invite feedback and comments from the research presented on this extended abstract.

**Acknowledgements.** None of this work have been possible without Carmen Dang, Cameron Anderson, Rebecca Benjamin, and Neerjah Skantharajah for their assistance with data collection in addition to Aimee Battcock and David Prete for their ongoing support. We would also like to acknowledge Drs. Pratik Pandharipande and Matthew Weinger from the

Vanderbilt University Medical Center Department of Anesthesiology. This wok was supported from the grants to Michael Schutz from the Natural Sciences and Engineering Research Council of Canada and the Canadian Foundation for Innovation [grant number CFI-LOF 30101].

## References

Block, F.E., Rouse, J.D., Hakala, M., & Thompson, C.L. (2000). A proposed new set of alarm sounds which satisfy standards and rationale to encode source information. *Journal of Clinical Monitoring and Computing, 16,* 541-546.

Cvach, M. (2012). Monitor alarm fatigue: An integrative review. *Biomedical Instrumentation & Technology, 46(4),* 268-277.

Darbyshire, J.L., & Young, J.D. (2013). An investigation of sound levels on intensive care units with reference to the WHO guidelines. *Critical Care, 17(5),* R187.

Donchin, Y., Gopher, D., Olin, M., Badihi, Y., Biesky, M., Sprung, C. L., … Cotev, S. (2003). A look into the nature and causes of human errors in the intensive care unit. *Quality and Safety in Health Care, 12(2),* 143–148.

Edworthy, J. (1998). Does sound help us to work better with machines? a commentary on Rauterberg's paper "About the importance of auditory alarms during the operation of a plant simulator." *Interacting with Computers, 10(4),* 401–409. https://doi.org/10.1016/S0953-5438(98)00036-8

Edworthy, J., & Hellier, E. (2005). Fewer but better auditory alarms will improve patient safety. *Quality and Safety in Health Care, 14(3),* 212–5. https://doi.org/10.1136/qshc.2004.01305

Edworthy, J., Hellier, E., Titchener, K., Naweed, A., & Roels, R. (2011). Heterogeneity in auditory alarm sets makes them easier to learn. *International Journal of Industrial Ergonomics, 41(2),* 136–146. https://doi.org/10.1016/j.ergon.2010.12.004

Fendrich, R., & Corballis, P. M. (2001). The temporal cross-capture of audition and vision. *Perception & Psychophysics, 63(4),* 719–725.

Gaver, W. W. (1986). Auditory Icons: Using Sound in Computer Interfaces. *Human–Computer Interaction, 2(2),* 167–177.

Giordano, B. L., Rocchesso, D., & McAdams, S. (2010). Integration of acoustical information in the perception of impacted sound sources: The role of information accuracy and exploitability. *Journal of Experimental Psychology: Human Perception and Performance, 36(2),* 462–476.

Grassi, M., & Darwin, C. J. (2006). The subjective duration of ramped and damped sounds. *Perception & Psychophysics, 68(6),* 1382–1392.

Lutfi, R. A. (2001). Auditory detection of hollowness. *The Journal of the Acoustical Society of America, 110(2),* 1010–1019.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology, 45(3),* 255-287.

Patterson, R. (1990). Auditory warning sounds in the work environment. *Philosophical Transactions of the Royal Society of London, 327(1241),* 485–492.

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),* 282–286. https://doi.org/10.1136/bmjqs-2014-003373

Rosenfield, R., Olsen, D., & Rudnicky, A. (2001). Universal speech interfaces. *Interactions, 8(6),* 34-44.

Sanderson, P., Wee, A. N., & Lacherez, P. (2006). Learnability and discriminability of melodic medical equipment alarms. *Anaesthesia, 61(2),* 142–147.

Schellenberg, E.G., & Trehub, S.E. (1994). Frequency ratios and the perception of tone patterns. *Psychonomic Bulletin & Review, 1(2),* 191-201.

Schutz, M. (2009). *Crossmodal integration: The search for unity.* University of Virginia.

Schutz, M., & Lipscomb, S. D. (2007). Hearing gestures, seeing music: Vision influences perceived tone duration. *Perception, 36(6),* 888–897. https://doi.org/10.1068/p5635

Schutz, M., & Vaisberg, J.M. (2014). Surveying the temporal structure of sounds used in music perception. *Music Perception, 31(3),* 288-296.

Schutz, M., Stefanucci, J., Baum, S., & Roth, A. (2017). Name that percussive tune: Associative memory and amplitude envelope. *Quarterly Journal of Experimental Psychology, 70(7),* 1323-1343.

Schutz, M., & Stefanucci, J.K. (under review). Hearing value: Exploring the effects of amplitude envelope on consumer preference. *Ergonomics in Design: The Quarterly of Human Factors Applications.*

Sendelbach, S., & Funk, M. (2013). Alarm fatigue: A patient safety concern. *AACN Advanced Critical Care, 24,* 378-386.

Sodnik, J., & Tomažič, S. (2015). *Auditory Interfaces. In Spatial Auditory Human-Computer Interfaces* (pp. 33–44). Cham: Springer International Publishing.

Wee, A. N., & Sanderson, P. (2008). Are melodic medical equipment alarms easily learned? *Anesthesia & Analgesia, 106(2),* 501–508. https://doi.org/10.1213/01.ane.0000286148.58823.6c