




Perspective

# Improving Human–Computer Interface Design through Application of Basic Research on Audiovisual Integration and Amplitude Envelope

Sharmila Sreetharan<sup>1</sup> and Michael Schutz<sup>2,\*</sup> 

<sup>1</sup> Department of Psychology, Neuroscience & Behaviour, McMaster University, Hamilton, ON L8S 4L8, Canada; sreet@s@mcmaster.ca

<sup>2</sup> School of the Arts, McMaster University, Hamilton, ON L8S 4L8, Canada

\* Correspondence: schutz@mcmaster.ca

Received: 20 December 2018; Accepted: 18 January 2019; Published: 22 January 2019



**Abstract:** Quality care for patients requires effective communication amongst medical teams. Increasingly, communication is required not only between team members themselves, but between members and the medical devices monitoring and managing patient well-being. Most human–computer interfaces use either auditory or visual displays, and despite significant experimentation, they still elicit well-documented concerns. Curiously, few interfaces explore the benefits of multimodal communication, despite extensive documentation of the brain’s sensitivity to multimodal signals. New approaches built on insights from basic audiovisual integration research hold the potential to improve future human–computer interfaces. In particular, recent discoveries regarding the acoustic property of amplitude envelope illustrate that it can enhance audiovisual integration while also lowering annoyance. Here, we share key insights from recent research with the potential to inform applications related to human–computer interface design. Ultimately, this could lead to a cost-effective way to improve communication in medical contexts—with significant implications for both human health and the burgeoning medical device industry.

**Keywords:** auditory alarms; human–computer interface design; auditory perception; amplitude envelope; audiovisual integration; aesthetics

---

## 1. Introduction

The appropriate design of human–computer interactions plays a crucial role in harnessing the powerful capabilities of electronic devices. Research on visual [1,2], and auditory [3,4] interfaces illustrates the importance of careful attention to the design of unimodal displays. Although relatively little research explores the efficacy of multimodal systems for human–computer interactions, psychologists and neuroscientists routinely illustrate the perceptual benefits of multimodal processing [5]. As a contribution to this special issue, we summarize current theories on these disparate but complementary areas of inquiry, highlighting recent discoveries related to acoustic properties facilitating audiovisual integration. Rather than a comprehensive overview, our goal is to contribute new ideas for interface design by illuminating points of potential interest through the practical applications of basic research.

## 2. Multimodal Processing

Our brains interpret incoming stimuli from various senses (i.e., cross-modal stimuli) to form a unified perception of our surroundings. One way of formally documenting and exploring this integration is by assessing neural responses to unisensory versus multisensory stimuli. Processing at the level of the neuron is measured by recording changes in neuronal membrane potential in response

to external events. When the membrane potential exceeds a certain threshold value, action potentials (APs) are generated. The neuronal response elicited by a stimulus is positively correlated to the firing rate and frequency of APs. Comparing the neuronal response to cross-modal stimuli (e.g., audiovisual) with the neuronal response to its unimodal components (e.g., the auditory and visual components) [6] sheds light into the nature of multimodal integration. Responses to multimodal stimuli larger than their unimodal components indicate multisensory enhancement, and neurons demonstrating such enhancements are considered “multisensory neurons” [7].

Multisensory enhancement is often inversely related to the effectiveness of the unimodal cues [8]. For example, noticing a cat approaching can involve both peripheral sight (a weak visual cue) and hearing soft footsteps (a weak auditory cue). Independently, these weak, unimodal cues are unlikely to attract interest; yet, when presented together (i.e., as a multimodal stimulus), they are more likely to capture attention. This is seen at the neuronal level as weak, unimodal cues combine together to produce a cross-modal response that is much larger than the sum of the responses to each of the unimodal cues—A super-additive response [7–9].

Conversely, strong unimodal cues evoke strong responses on their own. For instance, detecting a barking dog running towards you is easy to notice as the barking (a strong auditory cue) is sufficient in grabbing your attention; there is likely no added benefit in observing the dog (a strong visual cue). This is seen at the neuronal level when strong unimodal cues are combined together. Their cross-modal response is likely to be less than the sum of the responses to each of the unimodal cues, referred to as a sub-additive response [7–9]. This pattern of multisensory enhancement inversely related to unimodal effectiveness is known as the principle of inverse effectiveness [7]. This holds potential applications in the design of multi-modal alarms in medical settings as keeping individual alarm signals as weak as possible (while still detectable) helps to prevent sensory overload from multiple concurrent systems—An application which may be of use to some members of a medical team but not others.

### 2.1. Lower-Order Multimodal Integration (Stimulus Orientation)

Sound localization provides a useful example of multimodal signal enhancement. Stimulus orientation and stimulus localization are mediated by the superior colliculus (SC). The SC is a mid-brain structure implicated in multisensory integration and generation of spatial orienting responses. Although many brain regions are implicated in multisensory integration, the convergence of inputs from various sensory modalities (e.g., visual, auditory, somatosensory) [6,10], projections to motor areas [10,11], and the abundance of multisensory neurons [12] make the SC a natural location for illustrating multimodal processing on a neural level. Multisensory enhancement in the SC occurs in many animals, including cats [8,13], rats [14], ferrets [15], and primates [16], decreasing response times for stimulus orientation and localization behaviors. The SC’s role is preserved in humans, mediating multisensory spatial integration [17], thereby aiding faster response generations. Extending beyond simple stimulus orientation, there is some evidence that the SC facilitates primitive social behaviors such as facial mimicry in neonates [18].

### 2.2. Higher-Order Multimodal Integration (Perceptual Binding)

Lower-order multimodal processes are useful in understanding the neural basis of multimodal integration. However, understanding higher-order multimodal integration is crucial in our interpretation of the complex stimuli encountered in our daily lives and holds useful lessons for the design of multimodal interfaces. For example, verbal communication was once thought to be a purely auditory process; yet, it involves the perceptual binding of watching a speaker’s lip movements (e.g., vision) and hearing the speech sounds produced (e.g., audition) [19]. The well-known McGurk effect [20] clearly illustrates vision’s influence on speech perception, an effect that is magnified when semantic constraints are imposed. When lip movements are incongruent to the speech produced (e.g., lips pronounce ‘bows’ and speech produced is ‘goes’), an intermediate between the lip movements

and heard speech (e.g., 'doze') is perceived; however, when the visual is removed, this effect disappears and speech sounds are heard accurately [20].

Comparative studies in primates implicate the role of the superior temporal sulcus (STS) in audiovisual integration in response to face and speech sounds [21]. Similarly in humans, the STS has been implicated as the primary site for higher-order audiovisual processing [22–25]. One study using functional magnetic resonance imaging (fMRI) revealed an increase in the blood oxygen level-dependent signal in the STS in response to temporally aligned audiovisual stimuli but not to audio-only or visual-only stimuli [25], highlighting the multimodal capabilities of the STS. To establish a causal relationship between the STS and audiovisual integration as observed in the McGurk effect, one study used fMRI-guided transcranial magnetic stimulation (TMS) to create temporary virtual lesions while participants observed McGurk and control stimuli [26]. Creating a temporary lesion to the STS significantly decreased the McGurk effect [26]. These results were corroborated by another study using transcranial direct current stimulation (tDCS), a non-invasive neuromodulatory technique; cathodal tDCS (decreases excitability) applied over the STS showed a disruption to the McGurk effect, whereas anodal tDCS (increases excitability) applied over the STS showed an increase in the McGurk effect [27].

In addition to its role in speech, the STS also plays a crucial role in the audiovisual integration of non-speech sounds. For example, in the sound-induced fission illusion [28], a single flash is perceived to be multiple flashes when paired with multiple auditory beeps. When anodal tDCS is applied over the STS during the sound-induced fission illusion, there is an increase in perceived fission whereas cathodal tDCS applied over the STS results in a decrease in perceived fission [29]. These studies stress the important role of the STS in higher-order audiovisual processing for both speech and non-speech sounds.

### 2.3. Gains in Performance from Multisensory Stimulation

Basic research illustrates that multisensory stimulation alters neural and perceptual responses. This raises important questions about the potential benefits of multimodal presentations in human–computer interfaces. Although a full assessment requires explicit tests of future interfaces, previous research exploring generalized improvements in multimodal presentations offers useful insights. Each of the modalities brings different strengths, with audition offering superior temporal resolution and vision providing better spatial resolution [30]. When these modalities are used in tandem, audiovisual interfaces offer greater temporal and spatial resolution than either audio-only or visual-only interfaces. Furthermore, the use of auditory signals (both spatially correlated and non-spatial) during visual search tasks facilitates greater target saliency [31] and faster target identification [32–34].

Multisensory stimulation can also enhance the performance of a single primary modality. For example, bimodal presentation of visual and tactile motion stimuli results in faster detection of motion stimuli [35]. These kinds of effects can be used to enhance graphical user interfaces (GUIs), especially for those with disabilities. Scanning input is a visual task where users with physical disabilities (i.e., unable to operate a mouse) identify a required target by visually scanning items in the form of a grid. One study found that a sonically enhanced version of scanning input increased motivation and user engagement, with the potential to also increase scanning speed [36]. Additionally, the use of multimodal presentations in common GUI functions (e.g., drag and drop) have the potential to reduce perceived mental workload without affecting performance [37,38].

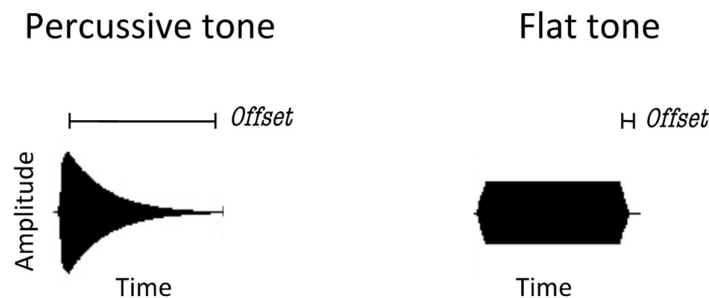
## 3. Amplitude Envelope and Alarm Design

Literature on multimodal processing illustrates the potential for improving human–computer interface design by drawing upon multiple modalities of presentation. However, this will require designers to consider the best ways to organize information across modalities in a way conducive to effective integration. To that end, we see potential for building on recent discoveries from our

team with respect to amplitude envelope. Amplitude envelope (henceforth “envelope”) refers to a sound’s shape—i.e., its amplitude profile over time. Sounds with flat envelopes (Figure 1, right panel) feature abrupt offsets offering little information about the materials (i.e., metallic/wooden) involved in sound producing events. In contrast, percussive sounds originating from impact events often exhibit relatively long, decaying offsets informing listeners about the materials involved in the event, such as the hollowness of the struck material [39]. Many auditory interfaces in medical devices, such as those mandated by the International Electrotechnical Commission (IEC), involve melodic alarm tones that use sounds with flat envelopes [40], and these alarm systems are widely recognized as problematic [40,41].

Although tradeoffs between envelope and other properties of tones have long fascinated musicians given the complexity of synthesizing musically satisfying sounds [42], psychologists have generally opted for study and manipulation of isolated parameters such as duration, frequency, and amplitude [43]. These parameters are more well-suited for careful control and manipulation, particularly with the advent of modern computers [44]. This careful focus on easily controlled parameters has been helpful in clarifying the importance of some low-level properties of sound, such as onset [45,46]. Onset plays an important role in distinguishing between musical timbres [47], to the point where removing the onset entirely renders otherwise distinct instruments indistinguishable [48]. Additionally, a lack of sensitivity to tone onset predicts deficits in reading [49].

Curiously, onset is the one temporal parameter that is relatively consistent between flat and percussive tones, which both feature abrupt rises. These tones differ in their sustain and most notably their offset segments (Figure 1), a parameter that has not often been overlooked in importance within the field of auditory psychophysics [50]. Nonetheless, differences in offset can trigger qualitatively different patterns regarding the perception of duration [51–55], loudness [56–58], and loudness change [59,60].



**Figure 1.** Percussive (left) and flat (right) tones. The shape of decay in percussive tones provides information about the sound-producing event (materials involved, nature of impact, etc.). In contrast, flat sounds possess abrupt offsets unlike those encountered in our evolutionary history. Figure taken with permission from Schutz et al. (2017).

### 3.1. A Demonstration of Envelope Affecting Audiovisual Integration

One example of envelope’s role in the assessment of event duration is particularly pertinent to multimodal interfaces. Research on audiovisual integration using flat tones has generally concluded that vision rarely influences auditory evaluations of duration (provided that the acoustic information is of sufficient quality [30,61,62]). However, a novel musical illusion illustrates that this long-standing conclusion does not hold in some contexts involving percussive sounds. For example, percussionists are able to manipulate audience perception of note duration by using long versus short motions when striking their instrument. Although these gestures fail to alter notes’ acoustic structures, audiences observing these long or short striking motions perceive these notes to sound either long or short [63]. Aside from resolving a long-running debate in the percussion community [64], this finding illustrates a novel documentation regarding a visual influence on the auditory perception of duration.

This finding contrasts markedly with previous findings that vision does not affect auditory duration assessments [30,61,62].

This novel pattern of processing led to a new understanding of audiovisual integration when using sounds with percussive envelopes. For example, this variation of this illusion illustrates it holds with other sounds produced by impact events (i.e., striking an instrument), but not sustained events such as blowing into a mouthpiece [65]. Although several factors play a role in this binding, one key factor is the amplitude envelope of notes produced by the marimba—A percussion instrument similar to a xylophone made of wooden bars struck by performers using mallets. These percussive sounds bind with striking gestures (producing impact sounds), whereas other sustained sounds such as those produced by bowing a cello do not [66]. Crucially for the design of interfaces with synthesized sounds, even simple pure tones (sine waves) shaped with decaying envelopes appear to trigger this privileged binding [67]. In contrast, pure tones shaped with flat envelopes fail to integrate with the same visual information [68].

Extensions to this research illustrate that short sound sequences (similar to those used in auditory alarms) shaped with percussive envelopes are easier to associate with everyday objects [69]. These associations appear to be both learned and retained better when using percussive instead of flat sounds. Although that study involved physical objects, the general question of which types of sounds are best associated is highly relevant to the design of effective auditory alarms as users need to learn and retain associations between sound sequences and commands.

### 3.2. Sounds Currently Used in Medical Device Alarms

The bias towards the use of flat tones in auditory interfaces is consistent with a bias towards such sounds in auditory perception research [70]. We suspect this is largely due to their high degree of experimental control [44], as their temporal structure can be easily and precisely specified using a minimum number of parameters.

However, using stimuli lacking the dynamic temporal changes (inherent in natural sounds) is problematic as they may be processed with different underlying processing strategies [71]. This poses major challenges for generalizing from controlled research studies to real-world applications, creating challenges both for theoretical and applied work. For example, theories derived from a large literature of audiovisual integration using flat tones fail to generalize to sounds with natural shapes. Experiments with flat tones have repeatedly concluded that vision does not affect auditory judgments of duration; however, vision can have a substantial influence when the sounds exhibit natural decays [65,67,68].

The use of flat tones also poses challenges in applied contexts when considering human factors related to auditory interfaces. For example, flat tones sound less pleasing than percussive tones, lowering the perceived value of products in which they are used [72]. Additionally, they are perceived as significantly more annoying than similar sequences of percussive tones [73]. These results help explain repeated previous findings that current alarms (based heavily on flat tones) are problematic for medical professionals [74] who hear hundreds of alarms throughout their workday [75]. Although the pleasantness of alarms may seem a secondary issue, it plays an important role in their use as users might disable unpleasant auditory alerts even if they are actually informative [76]—rendering them ineffective as a consequence of auditory aesthetics, rather than the appropriateness of their signals.

Problems with some current alarms are so widely recognized that those involved with their creation have issued formal apologies in the peer-reviewed literature [77]. Many of these alarms by definition require flat envelopes [40] forcing alarm designers into using sounds that are hard to learn [40,41] and not conducive to integration with visual information. In contrast, percussive sounds integrate more readily with visual information [66,68], are aesthetically preferable [72] and are perceived as significantly less annoying [73].

### 3.3. The Use of Percussive Tones in Multimodal Interfaces

Sounds with percussive envelopes integrate more strongly with visual information when assessing duration [68] and event unity [78], and are more easily associated with everyday objects [69]. Together, these findings hold important implications for the design of auditory alarms, which require users to learn and retain associations between sound sequences and commands. This is intriguing, given well-known problems with learning [40,41], retention [41,79], and confusion [41,80] of many current approaches. As the field evolves to more fully explore the possibilities of multimodal alarms for medical devices, we encourage exploration of envelope as a way to simultaneously increase audiovisual integration, lower annoyance, improve aesthetics, and offer better user experiences with human–computer interfaces.

Given the well-documented perceptual “gains” of stimulation in multiple modalities (Section 2) as well as the benefits of redundancy in important signals related to the critical issue of patient health, we see the pursuit of multimodal alarm systems as a potentially fruitful area for future research. Although these principles could be of use in any human–computer interface, they are of particular relevance for medical alarm design. Although Industry Canada values the medical device market at CAD \$6.7 billion, and USD \$336 billion annually [81], the design of auditory alarms in these devices is fraught. Yet, such devices’ instruments play an increasingly important role in patient care, with hundreds of alarms sounding per patient per day in busy hospitals [75].

To aid researchers interested in generating dynamically changing sounds, we have posted a free tool online allowing for the synthesis of percussive and flat tones of any duration and frequency useful in experimental contexts [82]. We have used this tool for several experiments in our lab as it offers a simple interface for precise stimulus generation. Although envelope is slowly gaining increased attention as a research topic for theoretical explorations [57,60,72,73,78], its applications in interface design have been relatively underexplored to date. As our interest in this property came directly from observing its surprising importance in audiovisual integration tasks [63,65,68,78], we are intrigued by its potential applications in improving multimodal interfaces in medical devices. Consequently, we are pleased to contribute to this special issue focused on raising awareness of this important topic.

**Author Contributions:** Conceptualization for manuscript provided by M.S. Writing for Section 2 primarily completed by S.S.; with other sections completed by M.S. Funding acquisition for project acquired by M.S.

**Funding:** We are grateful for support for this work provided by NSERC (Natural Science and Engineering Council of Canada), CFI-LOF (Canadian Foundation for Innovation Leaders Opportunity Fund), as well as the McMaster Arts Research Board and International Initiatives Micro Fund.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Mullet, K.; Sano, D. *Designing Visual Interfaces: Communication Oriented Techniques*; Prentice Hall PTR: Upper Saddle River, NJ, USA, 1994.
2. Marcus, A. *Human-computer Interaction*; Baecker, R.M., Grudin, J., Buxton, W.A.S., Greenberg, S., Eds.; Morgan Kaufmann Publishers Inc.: San Francisco, CA, USA, 1995; pp. 425–441, ISBN 1-55860-246-1.
3. Rocchesso, D.; Bresin, R.; Fernstrom, M. Sounding Objects. *IEEE MultiMed.* **2003**, *10*, 42–52. [[CrossRef](#)]
4. Jeon, M. Auditory User Interface Design: Practical Evaluation Methods and Design Process Case Studies. *Int. J. Des. Soc.* **2015**, *8*, 1–16. [[CrossRef](#)]
5. Calvert, G.A.; Spence, C.; Stein, B.E. *The Handbook of Multisensory Processes*; MIT Press: Cambridge, MA, USA, 2004; ISBN 0262033216.
6. Meredith, M.A.; Stein, B.E. Interactions among Converging Sensory Inputs in the Superior Colliculus. *Science* **1983**, *221*, 389–391. [[CrossRef](#)] [[PubMed](#)]
7. Stein, B.E.; Stanford, T.R.; Ramachandran, R.; Perrault, T.; Rowland, B. Challenges in quantifying multisensory integration: Alternative criteria, models, and inverse effectiveness. *Exp. Brain Res.* **2009**, *198*, 113–126. [[CrossRef](#)] [[PubMed](#)]

8. Stein, B.E.; Meredith, M.A. *The Merging of the Senses*; The MIT Press: Cambridge, MA, USA, 1993; ISBN 0-262-19331-0.
9. Stein, B.E.; Stanford, T.R. Multisensory integration: Current issues from the perspective of the single neuron. *Nat. Rev. Neurosci.* **2008**, *9*, 255–266. [[CrossRef](#)]
10. Meredith, M.A.; Stein, B.E. Visual, auditory, and somatosensory convergence on cells in superior colliculus results in multisensory integration. *J. Neurophysiol.* **1986**, *56*, 640–662. [[CrossRef](#)] [[PubMed](#)]
11. May, P.J. The mammalian superior colliculus: Laminar structure and connections. In *Neuroanatomy of the Oculomotor System*; Büttner-Ennever, J., Ed.; Elsevier: New York, NY, USA, 2006; Volume 151, pp. 321–378, ISBN 0079-6123.
12. King, A.J. The superior colliculus. *Curr. Biol.* **2004**, *14*, R335–R338. [[CrossRef](#)] [[PubMed](#)]
13. Wallace, M.T.; Meredith, M.A.; Stein, B.E. Multisensory Integration in the Superior Colliculus of the Alert Cat. *J. Neurophysiol.* **1998**, *80*, 1006–1010. [[CrossRef](#)] [[PubMed](#)]
14. Hirokawa, J.; Sadakane, O.; Sakata, S.; Bosch, M.; Sakurai, Y. Multisensory Information Facilitates Reaction Speed by Enlarging Activity Difference between Superior Colliculus Hemispheres in Rats. *PLoS ONE* **2011**, *6*, e25283. [[CrossRef](#)] [[PubMed](#)]
15. Hammond-Kenny, A.; Bajo, V.M.; King, A.J.; Nodal, F.R. Behavioural benefits of multisensory processing in ferrets. *Eur. J. Neurosci.* **2017**, *45*, 278–289. [[CrossRef](#)] [[PubMed](#)]
16. Sparks, D.L. Translation of sensory signals into commands for control of saccadic eye movements: Role of primate superior colliculus. *Physiol. Rev.* **1986**, *66*, 118–171. [[CrossRef](#)] [[PubMed](#)]
17. Leo, F.; Bertini, C.; di Pellegrino, G.; Ladavas, E. Multisensory integration for orienting responses in humans requires the activation of the superior colliculus. *Exp. Brain Res.* **2008**, *186*, 67–77. [[CrossRef](#)] [[PubMed](#)]
18. Pitti, A.; Kuniyoshi, Y.; Quoy, M.; Gaussier, P. Development of the Multimodal Integration in the Superior Colliculus and Its Link to Neonates Facial Preference BT. In *Advances in Cognitive Neurodynamics (IV)*; Liljenström, H., Ed.; Springer: Dordrecht, The Netherlands, 2015; pp. 543–546.
19. Ross, L.A.; Saint-Amour, D.; Leavitt, V.M.; Javitt, D.C.; Foxe, J.J. Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments. *Cereb. Cortex* **2007**, *17*, 1147–1153. [[CrossRef](#)] [[PubMed](#)]
20. McGurk, H.; MacDonald, J. Hearing lips and seeing voices. *Nature* **1976**, *264*, 746–748. [[CrossRef](#)] [[PubMed](#)]
21. Ghazanfar, A.A. The multisensory roles for auditory cortex in primate vocal communication. *Hear. Res.* **2009**, *258*, 113–120. [[CrossRef](#)] [[PubMed](#)]
22. Venezia, J.H.; Vaden, K.I.; Rong, F.; Maddox, D.; Saberi, K.; Hickok, G. Auditory, Visual and Audiovisual Speech Processing Streams in Superior Temporal Sulcus. *Front. Hum. Neurosci.* **2017**, *11*, 174.
23. Beauchamp, M.S.; Lee, K.E.; Argall, B.D.; Martin, A. Integration of Auditory and Visual Information about Objects in Superior Temporal Sulcus. *Neuron* **2004**, *41*, 809–823. [[CrossRef](#)]
24. Hein, G.; Knight, R.T. Superior Temporal Sulcus—It's My Area: Or Is It? *J. Cogn. Neurosci.* **2008**, *20*, 2125–2136. [[CrossRef](#)]
25. Noesselt, T.; Rieger, J.W.; Schoenfeld, M.A.; Kanowski, M.; Hinrichs, H.; Heinze, H.; Driver, J. Audiovisual Temporal Correspondence Modulates Human Multisensory Superior Temporal Sulcus Plus Primary Sensory Cortices. *J. Neurosci.* **2007**, *27*, 11431–11441. [[CrossRef](#)]
26. Beauchamp, M.S.; Nath, A.R.; Pasalar, S. fMRI-Guided transcranial magnetic stimulation reveals that the superior temporal sulcus is a cortical locus of the McGurk effect. *J. Neurosci.* **2010**, *30*, 2414–2417. [[CrossRef](#)]
27. Marques, L.M.; Lapenta, O.M.; Merabet, L.B.; Bolognini, N.; Boggio, P.S. Tuning and disrupting the brain-modulating the McGurk illusion with electrical stimulation. *Front. Hum. Neurosci.* **2014**, *8*, 533. [[CrossRef](#)]
28. Shams, L.; Kamitani, Y.; Shimojo, S. Visual illusion induced by sound. *Cogn. Brain Res.* **2002**, *14*, 147–152. [[CrossRef](#)]
29. Bolognini, N.; Rossetti, A.; Casati, C.; Mancini, F.; Vallar, G. Neuromodulation of multisensory perception: A tDCS study of the sound-induced flash illusion. *Neuropsychologia* **2011**, *49*, 231–237. [[CrossRef](#)] [[PubMed](#)]
30. Walker, J.T.; Scott, K.J. Auditory-visual conflicts in the perceived duration of lights, tones and gaps. *J. Exp. Psychol. Hum. Percept. Perform.* **1981**, *7*, 1327–1339. [[CrossRef](#)] [[PubMed](#)]

31. Iordanescu, L.; Grabowecky, M.; Franconeri, S.; Theeuwes, J.; Suzuki, S. Characteristic sounds make you look at target objects more quickly. *Atten. Percept. Psychophys.* **2010**, *72*, 1736–1741. [[CrossRef](#)] [[PubMed](#)]
32. Vroomen, J.; de Gelder, B. Sound enhances visual perception: Cross-modal effects of auditory organization on vision. *J. Exp. Psychol. Hum. Percept. Perform.* **2000**, *26*, 1583–1590. [[CrossRef](#)] [[PubMed](#)]
33. Van der Burg, E.; Olivers, C.N.L.; Bronkhorst, A.W.; Theeuwes, J. Pip and pop: Nonspatial auditory signals improve spatial visual search. *J. Exp. Psychol. Hum. Percept. Perform.* **2008**, *34*, 1053–1065. [[CrossRef](#)]
34. Perrott, D.R.; Sadralodabai, T.; Saberi, K.; Strybel, T.Z. Aurally Aided Visual Search in the Central Visual Field: Effects of Visual Load and Visual Enhancement of the Target. *Hum. Factors* **1991**, *33*, 389–400. [[CrossRef](#)] [[PubMed](#)]
35. Ushioda, H.; Wada, Y. Multisensory integration between visual and tactile motion information: Evidence from redundant-signals effects on reaction time. *Proc. Fechner Day* **2007**, *23*, 1.
36. Brewster, S.; Raty, V.-P.; Kortekangas, A. Enhancing Scanning Input With Non-Speech Sounds. In Proceedings of the Second Annual ACM Conference on Assistive Technologies, Vancouver, BC, Canada, 11–12 April 1996; pp. 10–14.
37. Vitense, H.S.; Jacko, J.A.; Emery, V.K. Multimodal feedback: An assessment of performance and mental workload. *Ergonomics* **2003**, *46*, 68–87. [[CrossRef](#)] [[PubMed](#)]
38. Brewster, S. Sonically-Enhanced Drag and Drop. In Proceedings of the International Conference on Auditory Display (ICAD 98), Glasgow, UK, 1–4 November 1998; pp. 1–7.
39. Lutfi, R.A. Auditory detection of hollowness. *J. Acoust. Soc. Am.* **2001**, *110*, 1010–1019. [[CrossRef](#)]
40. Wee, A.N.; Sanderson, P. Are melodic medical equipment alarms easily learned? *Anesth. Analg.* **2008**, *106*, 501–508. [[CrossRef](#)] [[PubMed](#)]
41. Sanderson, P.; Wee, A.N.; Lacherez, P. Learnability and discriminability of melodic medical equipment alarms. *Anaesthesia* **2006**, *61*, 142–147. [[CrossRef](#)] [[PubMed](#)]
42. Risset, J.-C.; Wessel, D.L. Exploration of Timbre by Analysis and Synthesis. In *The Psychology of Music*; Deutsch, D., Ed.; Gulf Professional Publishing: San Diego, CA, USA, 1999; pp. 113–169, ISBN 0122135652.
43. Gaver, W. What in the world do we hear?: An ecological approach to auditory event perception. *Ecol. Psychol.* **1993**, *5*, 1–29. [[CrossRef](#)]
44. Neuhoff, J.G. *Ecological Psychoacoustics*; Neuhoff, J.G., Ed.; Elsevier Academic Press: Amsterdam, The Netherlands, 2004; ISBN 9780125158510.
45. Gordon, J.W. The perceptual attack time of musical tones. *J. Acoust. Soc. Am.* **1987**, *82*, 88–105. [[CrossRef](#)] [[PubMed](#)]
46. Strong, W.; Clark, M. Perturbations of synthetic orchestral wind-instrument tones. *J. Acoust. Soc. Am.* **1967**, *41*, 277–285. [[CrossRef](#)]
47. Skarratt, P.A.; Cole, G.G.; Gellatly, A.R.H. Prioritization of looming and receding objects: Equal slopes, different intercepts. *Atten. Percept. Psychophys.* **2009**, *71*, 964–970. [[CrossRef](#)] [[PubMed](#)]
48. Saldanha, E.L.; Corso, J.F. Timbre cues and the identification of musical instruments. *J. Acoust. Soc. Am.* **1964**, *36*, 2021–2026. [[CrossRef](#)]
49. Goswami, U. A temporal sampling framework for developmental dyslexia. *Trends Cogn. Sci.* **2011**, *15*, 3–10. [[CrossRef](#)] [[PubMed](#)]
50. Schutz, M. Acoustic structure and musical function: Musical notes informing auditory research. In *The Oxford Handbook on Music and the Brain*; Thaut, M.H., Hodges, D.A., Eds.; Oxford University Press: Oxford, UK, 2011.
51. Schlauch, R.S.; Ries, D.T.; DiGiovanni, J.J. Duration discrimination and subjective duration for ramped and damped sounds. *J. Acoust. Soc. Am.* **2001**, *109*, 2880–2887. [[CrossRef](#)] [[PubMed](#)]
52. Grassi, M.; Pavan, A. The subjective duration of audiovisual looming and receding stimuli. *Atten. Percept. Psychophys.* **2012**, *74*, 1321–1333. [[CrossRef](#)] [[PubMed](#)]
53. Grassi, M.; Darwin, C.J. The subjective duration of ramped and damped sounds. *Percept. Psychophys.* **2006**, *68*, 1382–1392. [[CrossRef](#)] [[PubMed](#)]
54. DiGiovanni, J.J.; Schlauch, R.S. Mechanisms responsible for differences in perceived duration for rising-intensity and falling-intensity sounds. *Ecol. Psychol.* **2007**, *19*, 239–264. [[CrossRef](#)]



55. Grassi, M. Sex difference in subjective duration of looming and receding sounds. *Perception* **2010**, *39*, 1424–1426. [[CrossRef](#)] [[PubMed](#)]
56. Ries, D.T.; Schlauch, R.S.; DiGiovanni, J.J. The role of temporal-masking patterns in the determination of subjective duration and loudness for ramped and damped sounds. *J. Acoust. Soc. Am.* **2008**, *124*, 3772–3783. [[CrossRef](#)] [[PubMed](#)]
57. Stecker, G.C.; Hafter, E.R. An effect of temporal asymmetry on loudness. *J. Acoust. Soc. Am.* **2000**, *107*, 3358–3368. [[CrossRef](#)] [[PubMed](#)]
58. Teghtsoonian, R.; Teghtsoonian, M.; Canévet, G. Sweep-induced acceleration in loudness change and the “bias for rising intensities”. *Percept. Psychophys.* **2005**, *67*, 699–712. [[CrossRef](#)]
59. Neuhoff, J.G. An Adaptive Bias in the Perception of Looming Auditory Motion. *Ecol. Psychol.* **2001**, *13*, 87–110. [[CrossRef](#)]
60. Neuhoff, J.G. Perceptual bias for rising tones. *Nature* **1998**, *395*, 123–124. [[CrossRef](#)]
61. Welch, R.; Warren, D. Immediate perceptual response to intersensory discrepancy. *Psychol. Bull.* **1980**, *88*, 638–667. [[CrossRef](#)] [[PubMed](#)]
62. Fendrich, R.; Corballis, P.M. The temporal cross-capture of audition and vision. *Percept. Psychophys.* **2001**, *63*, 719–725. [[CrossRef](#)] [[PubMed](#)]
63. Schutz, M.; Lipscomb, S.D. Hearing gestures, seeing music: Vision influences perceived tone duration. *Perception* **2007**, *36*, 888–897. [[CrossRef](#)] [[PubMed](#)]
64. Schutz, M. The mind of the listener: Acoustics, perception, and the musical experience. *Percuss. Notes* **2009**, 22–28.
65. Schutz, M.; Kubovy, M. Causality and cross-modal integration. *J. Exp. Psychol. Hum. Percept. Perform.* **2009**, *35*, 1791–1810. [[CrossRef](#)] [[PubMed](#)]
66. Chuen, L.; Schutz, M. The unity assumption facilitates cross-modal binding of musical, non-speech stimuli: The role of spectral and amplitude cues. *Attent. Percept. Psychophys.* **2016**, *78*, 1512–1528. [[CrossRef](#)] [[PubMed](#)]
67. Armontrout, J.A.; Schutz, M.; Kubovy, M. Visual determinants of a cross-modal illusion. *Attent. Percept. Psychophys.* **2009**, *71*, 1618–1627. [[CrossRef](#)]
68. Schutz, M. *Crossmodal Integration: The Search for Unity*; University of Virginia: Charlottesville, VA, USA, 2009.
69. Schutz, M.; Stefanucci, J.; Baum, S.H.; Roth, A. Name that percussive tune: Associative memory and amplitude envelope. *Q. J. Exp. Psychol.* **2017**, *70*, 1323–1343. [[CrossRef](#)]
70. Schutz, M.; Vaisberg, J.M. Surveying the temporal structure of sounds used in Music Perception. *Music Percept. Interdiscip. J.* **2014**, *31*, 288–296. [[CrossRef](#)]
71. Vallet, G.; Shore, D.I.; Schutz, M. Exploring the role of amplitude envelope in duration estimation. *Perception* **2014**, *43*, 616–630. [[CrossRef](#)]
72. Schutz, M.; Stefanucci, J. Hearing value: Exploring the effects of amplitude envelope on consumer preference. *Ergon. Des. Q. Hum. Factors Appl.* in press.
73. Sreetharan, S.; Schlesinger, J.; Schutz, M. Designing Effective Auditory Interfaces: Exploring the Role of Amplitude Envelope. In Proceedings of the ICMPC15/ESCOM10, Graz, Austria, 23–28 July 2018; Parncutt, R., Sattmann, S., Eds.; pp. 426–431.
74. Rayo, M.F.; Moffatt-Bruce, S.D. Alarm system management: Evidence-based guidance encouraging direct measurement of informativeness to improve alarm response: Table 1. *BMJ Qual. Saf.* **2015**, *24*, 282–286. [[CrossRef](#)] [[PubMed](#)]
75. AAMI. *Clinical Alarms Summit Conveners*; AAMI: Arlington, VA, USA, 2011.
76. Edworthy, J. 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”. *Interact. Comput.* **1998**, *10*, 401–409.
77. Block, F.E. “For if the trumpet give an uncertain sound, who shall prepare himself to the battle?” (I Corinthians 14:8, KJV). *Anesth. Analg.* **2008**, *106*, 357–359. [[CrossRef](#)] [[PubMed](#)]
78. Grassi, M.; Casco, C. Audiovisual bounce—Inducing effect: When sound congruence affects grouping in vision. *Attent. Percept. Psychophys.* **2010**, *72*, 378–386. [[CrossRef](#)] [[PubMed](#)]
79. Edworthy, J.; Hellier, E. Alarms and human behaviour: Implications for medical alarms. *Br. J. Anaesth.* **2006**, *97*, 12–17. [[CrossRef](#)] [[PubMed](#)]

80. Gillard, J.; Schutz, M. Composing alarms: Considering the musical aspects of auditory alarm design. *Neurocase Neural Basis Cogn.* **2016**, *22*, 566–576. [[CrossRef](#)] [[PubMed](#)]
81. Canada G of Medical Devices Industry Profile. 2017. Available online: [https://www.ic.gc.ca/eic/site/lsg-pdsv.nsf/eng/h\\_hn01736.html](https://www.ic.gc.ca/eic/site/lsg-pdsv.nsf/eng/h_hn01736.html) (accessed on 27 September 2017).
82. Schutz, M. Software Tool for Tone Creation. Available online: [www.maplelab.net/software](http://www.maplelab.net/software) (accessed on 21 January 2019).



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).