INVITED PAPER

Efficient Multimodal Cuing of Spatial Attention

In this paper, the authors detail how attentional distribution can work in the important example of multimodal spatial cuing, and ground it in several classes of applications.

By Rob Gray, Charles Spence, Cristy Ho, and Hong Z. Tan, *Senior Member IEEE*

ABSTRACT | Behavioral studies of multisensory integration and cross-modal spatial attention have identified many potential benefits of using interfaces that engage more than just a single sense in complex operating environments. Particularly relevant in terms of application, the latest research highlights that: 1) multimodal signals can be used to reorient spatial attention effectively under conditions of high operator workload in which unimodal signals may be ineffective; 2) multimodal signals are less likely to be masked in noisy environments; and 3) there are natural links between specific signals and particular behavioral responses (e.g., head turning). However, taking advantage of these potential benefits requires that interface designers take into account the limitations of the human operator. In particular, multimodal interfaces should normally be designed so as to minimize any spatial incongruence between component warning signals presented in different sensory modalities that relate to the same event. Building on this rapidly growing cognitive neuroscience knowledge base, the last decade has witnessed the development of a number of highly effective multimodal interfaces for driving, aviation, the military, medicine, and sports.

KEYWORDS | Crossmodal; interface design; multimodal; multisensory; spatial attention

Manuscript received May 21, 2012; revised August 19, 2012; accepted October 3, 2012. Date of publication January 4, 2013; date of current version August 16, 2013. This work was supported in part by the Engineering and Physical Sciences Research Council (U.K.) under Grant RRAH15977.

R. Gray is with the School of Sport and Exercise Sciences, University of Birmingham, B15 2TT Birmingham, U.K. (e-mail: r.gray.2@bham.ac.uk).

C. Spence and **C. Ho** are with the Department of Experimental Psychology, University of Oxford, OX1 3UD Oxford, U.K. (e-mail: charles.spence@psy.ox.ac.uk; cristy.ho@psy.ox.ac.uk).

H. Z. Tan is with the School of Electrical and Computer Engineering, Purdue University, West Lafayette, IN 47907 USA, and also with the Human Computer Interaction Group, Microsoft Research Asia, Beijing 100080, China (e-mail: hongtan@purdue.edu).

Digital Object Identifier: 10.1109/JPROC.2012.2225811

I. INTRODUCTION

The term "multimedia interfaces" (often referred to as "multimodal" or "multisensory") refers to human–machine interfaces that engage at least two sensory systems (or sensory modalities) in order to facilitate effective interactions between humans and complex systems. Multimedia interfaces are expected to outperform those that utilize only a single medium (or modality), due, in part, to the fact that the involvement of multiple modalities in these interfaces allows information to be presented and responded to in a variety of different ways. Higher information transmission rates may also be achieved through redundancy of modalities utilized and formats of information to be incorporated into the systems [1], and when real-life environments are more effectively emulated. Additionally, when used effectively, multimodal interfaces can reduce mental workload, improve memory, and can potentially make human–computer interactions more natural and intuitive (e.g., [2]). However, designing such interfaces presents a number of unique challenges because interactive effects between the different modalities may arise and there is always a danger of sensory and attentional overload. The factors underlying these interactive effects certainly need to be better understood in order to fully realize the benefits of multimodal interfaces. This paper provides an overview of psychophysical interrelations between the spatial senses of audition, vision, and touch, with respect to visual search, attentional capture, and related topics, with an emphasis on how they may be applied in a variety of multimodal situations.

II. NEURAL MECHANISMS UNDERLYING MULTISENSORY PERCEPTION

The last two decades have seen great advances in our understanding of both multisensory integration and cross-modal attention. A large (and growing) number of

behavioral/psychophysical studies have demonstrated extensive interactions between the spatial senses of audition, vision, and touch. Such insights have been picked up by many applied researchers interested in the design of more effective displays (e.g., in the design of warning signals to capture an interface operator's spatial attention). Numerous studies have now demonstrated that the attention of an interface operator can be captured by the presentation of a noninformative spatial cue, and that such (multisensory) warning signals can significantly enhance a human operator's behavioral performance. However, the danger is that unless such multisensory warning signals are designed with the limitations of the operator's perceptual and attentional systems in mind, they might not be as effective as they have the potential to be. In fact, there is always the worry that the inappropriate (or incongruent) use of multiple sensory modalities might potentially confuse/overload a user, with the associated performance costs that might result in minimal or even no gain in terms of performance (see [3]).

A. Distinguishing Between Response Priming and Attentional Orienting Effects

It is important here to distinguish between the facilitation of performance that results from the priming of a particular behavioral response, and the facilitation of responding that results from the attentional enhancement of the perception of the target stimulus (which may, in turn, give rise to faster and/or more accurate behavioral responses). Traditionally, human factors researchers have often failed to distinguish between the two (e.g., see [4] for a review). It is, however, important to note that the neural mechanisms underlying the facilitation of an operator's performance are likely to be different in the two cases [5]. While response priming can lead to a speeding up of an operator's responses that is often greater than that documented when cross-modal spatial attentional facilitation is the only mechanism in play, such performance enhancement often comes at the cost of a speed-accuracy tradeoff (i.e., faster but potentially less accurate responding on the part of the operator) [6]. In other words, an operator may be primed to make a particular behavioral response (be it turning their head to check the wing mirror, or hitting the brake pedal) without necessarily giving due consideration as to whether this is actually the most appropriate response in a given situation.

B. The Response of an Operator to Multimodal Signals Cannot Be Directly Predicted From Neurophysiological Findings

Much of the applied research on multisensory warning signals in recent years has been inspired by the findings of single-cell neurophysiological studies traditionally conducted in the anesthetized animal model (see [5] and [7], for reviews; see also [8]). Recording from neurons in the deeper layers of the superior colliculus (SC), a subcortical brain structure involved in the control of overt orienting movements of the eyes and/or head toward peripheral events of interest often reveals an enhanced neural response following the presentation of pairs of individually weakly effective sensory cues (note that covert attentional orienting also engages the SC [9]). The maximal response is normally seen in such neurons when the component unimodal signals are presented from more or less the same position at about the same time. These are known as the spatial and temporal rules, respectively. However, it turns out that such effects are somewhat hard to obtain in practice, at least when it comes to behavioral cuing in awake human operators [8]. Certainly, one is unlikely to see the 1400% improvement in performance that has, on occasion, been documented at the single-cell level. Two important points to bear in mind here are that the neuron (which often appears in presentations; e.g., see [10]) was apparently one of the best ever recorded by Stein, Meredith, and their colleagues. What is more, there is also a potential for sampling bias in neurophysiological research: namely, there is a tendency for only the best (or most responsive) neurons to be reported in the literature. Relevant here then is an article by Stein *et al.* [11] in which they documented the distribution of neuronal responses in the SC. Visual inspection of Fig. 1 shows that at the population level, the responses of neurons appear to fall on a normal distribution centered around additivity or linear summation.

Now, all of this should not be taken to imply that the phenomenon of superadditivity does not exist. It certainly does at the neuronal level [12]. It is just that predicting the response of the human operator to particular patterns of multisensory stimulation is a much more complex business. What is more, the kind of multisensory integration seen at the behavioral level may depend on the type of

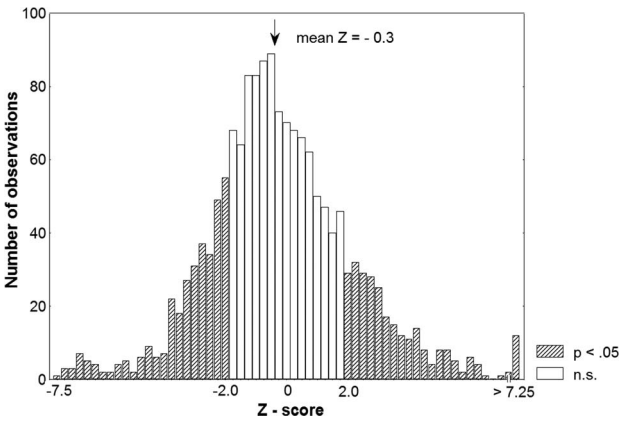

Fig. 1. ***The probability of obtaining a particular multisensory response given the prediction by simple summation of the modality-specific inputs. The crosshatched bars indicate significant subadditivity (Z-score < −1.96) and superadditivity (Z-score > 1.96). Figure reprinted from [11] with permission.***

stimuli that are presented (e.g., speech versus simple tones and light flashes; see [13]), not to mention the specific task of an operator [14]. Different models have been proposed to explain multisensory integration at the behavioral level. For example, according to the maximum-likelihood estimation (MLE) model, the integration of information across the senses is weighted as a function of the reliability of the component unisensory signals [15]. While the MLE approach has proved very successful in accounting for a wide variety of research documenting sensory dominance, we unfortunately do not have space to cover this rapidly growing literature in more detail here. The interested reader is directed to [16] for further reading on this topic.

C. The Ability of Multimodal Cues to Capture an Operator's Attention Depends on Workload

No matter whether superadditivity is seen at the behavioral level, there are still important reasons why multisensory stimulation may give rise to enhanced spatial orienting responses from an interface operator working under demanding conditions. To date, such multisensory enhancement has perhaps been demonstrated most clearly for the case of those warning signals used to capture a person's spatial attention. Now, it is often said in the literature that auditory and tactile warning signals automatically capture a person's spatial attention, even when the signal itself happens to be uninformative with regard to the likely location of a target event [17]. It turns out, though, that many of the studies in which such automatic (or exogenous) spatial cuing effects have been demonstrated were conducted under conditions in which the participant had nothing to do but respond to a series of target stimuli presented in an otherwise featureless sensory environment. Santangelo *et al.* have conducted a number of studies demonstrating that such unimodal attentional capture effects frequently dissipate under those conditions in which a participant happens to be performing a concurrent attention-demanding central task (as is obviously more likely to be the case in the majority of real-world situations; e.g., [18]; see [19] for a review).

However, the crucial result to emerge from a number of studies has been that while multisensory cues are not necessarily any more effective than unimodal cues under conditions of low perceptual load (see [20] for a review), they retain their capacity to capture an operator's spatial attention under conditions of high load, when unimodal cues may no longer be all that effective (see [19] for a review). This, at least, is the result if the auditory and tactile components of a multisensory cue (or warning signal) are presented from the same spatial location (or at least from the same direction). Multisensory cues are, however, no longer effective if exactly the same component unimodal cues are presented from different spatial locations/directions [21]. Such results are consistent with the notion that spatial coincidence facilitates both multisensory integration and spatial attentional capture.

D. When Spatial Congruency Is Not Quite so Important

Thus far, we have looked at those situations in which a warning signal (or cue) is presented *prior* to the presentation of a target stimulus. Interestingly, however, when the onset of the warning signal is timed to coincide with that of the target event, then the location from which the warning signal (especially if it is an auditory or tactile stimulus) is presented does not matter quite so much. In fact, if the auditory cue is not especially well localized in the first place, then it may make the visual target event pop out, and the temporal synchrony of the sound and visual stimulus might lead to the perceived location of the sound actually being ventriloquized toward that of the visual target [22]–[24].

While synchronous warning signals can enhance visual target performance in cluttered scenes/displays regardless of whether or not they happen to be presented from a spatially coincident position, there is evidence to suggest that presenting the auditory or tactile stimuli from closer to the relevant visual display can nevertheless still give rise to somewhat larger cross-modal attentional capture effects than when the stimuli are placed farther away (see [23]; see [25] and [26] for reviews). It remains an interesting question for future research as to what modulatory role spatial coincidence/congruency plays in the multisensory integration of various different combinations of sensory inputs (audiovisual versus audiotactile, for example). Here we have focused the discussion on the spatial aspects of multisensory integration and attentional capture. It has been argued by some researchers that audiotactile interactions are fundamentally nonspatial [27]. It should, therefore, be pointed out that other factors, such as temporal synchrony (i.e., the simultaneity of the presentation of two stimuli) [28] or semantic congruency (i.e., the correspondence in terms of semantic meanings associated with the two stimuli) [29], [30], may take precedence in determining the resultant performance gain (or loss) of multisensory integration.

E. Are Meaningful Multimodal Signals (i.e., Icons) Better Than Abstract Signals?

There are two additional points to note here. First, the warning signals that have been used in many of the studies that have been published to date have typically involved the presentation of meaningless auditory and tactile warning signals (e.g., pure tones or white noise bursts). There is growing interest in the question of whether meaningful auditory or tactile warning signals (e.g., auditory or haptic icons) might work better, especially under more ecologically valid real-world testing conditions (i.e., under those conditions in which an interface operator may have to deal with a large number of potentially informative auditory, or tactile, warning signals [31]). For instance, Ho and Spence [32] recently demonstrated that a white noise burst can sometimes be just as effective as the sound of a car horn

(a well-known auditory icon; see [33]). It should be noted that Ho and Spence used spatially nonpredictive warning sounds (as is common in laboratory studies of spatial attention, but which may, at first glance, appear rather curious to more applied researchers). Moreover, the nature of the auditory stimulus (e.g., auditory icon versus pure tone) was not tied to different behavioral responses, as might have been expected to be the case in any plausible real-world setting. Relative to research on visual and auditory icons, less work has been published in the area of haptic icons. This is perhaps because it is harder to think of many examples of intuitive haptic icons that could be easily presented to an interface operator. However, there have been some successful developments, and it is an area of research that is growing rapidly (see [34] and [35]; reviewed in [36]).

F. There May Be Specific Links Between Particular Warning Signals and Behavioral Responses

A second area of growing interest relates to the different regions of space in which warning signals are presented. To give but one example, Ho and Spence [32] conducted a study in which they demonstrated that when auditory warning signals were presented from 40 cm behind the head they proved particularly effective in terms of orienting a driver's head (and hence gaze), as compared to those conditions in which the warning signals were presented 80 cm in front (auditory signals), or peripheral/central (visual signals), or vibrotactile signals presented on the wrist/around the waist (cf. [37] and [38]). One important point to bear in mind here is that most laboratory studies have been conducted with participants making simple button press responses. However, as Ho and Spence noted, it is possible that there may be specific links between particular warning signals and behavioral responses. Thus, while presenting a warning signal from close to the back of the head might well prove to be particularly effective at facilitating a driver's head turning responses, it might not necessarily be so effective if the driver simply has to make a speeded footpedal response (such as hitting the brake pedal). In this regard, researchers have also investigated the possibility of presenting the warning signal from the effector of the desired response, for example, vibrating the brake pedal when the driver ought to hit the brake pedal (e.g., [39]).

Multisensory warning signals offer a number of potential benefits over their unimodal (or unisensory) counterparts. First, because unimodal warning signals can be masked, auditory warning signals might be blocked out by loud background noise (or the radio) while tactile warning signals might be missed (or at the very least become less effective) for a human operator who happens to be wearing thick clothing (e.g., imagine the driver of a snow plough working in midwinter [40]) or under those conditions where there is a lot of vibration attributable to a rough road surface. The utilization of multisensory warning signals therefore makes it less likely that a warning signal will be missed. Perhaps more importantly though, multisensory warning signals also have the advantage that they appear to remain effective under conditions of high operator workload. At least that is the case when the component unisensory signals are presented from the same spatial location at more or less the same time. That said, the latest research suggests that precise spatial colocation may be somewhat less important when an auxiliary signal is presented at the same time as a visual target (e.g., in an air traffic control scenario [41], [42]).

G. Why Introducing a Slight Asynchrony Between the Signals Presented to the Different Modalities May Lead to More Effective Multimodal Warnings

One other consideration to mention here concerning the optimal design of multisensory warning signals for interface operators working under demanding conditions relates to signal synchrony. In particular, a closer reading of the cognitive neuroscience literature may lead one to question whether perfect synchrony is necessarily always optimal when it comes to the design of multisensory warning signals [41]. There is suggestive evidence that making audiovisual warning signals slightly asynchronous might, counterintuitive as it might seem, actually improve their effectiveness [8], [43]. One idea here is that slightly desynchronizing the unimodal components of a warning signal might mean that they reach the SC at same time (due to the existence of modality-specific differences in transduction latencies) and hence give rise to enhanced spatial attentional orienting. Another reason why one might consider the introduction of a slight asynchrony between the auditory and visual components of a multisensory warning signal would be in order to mimic an event presented at a specific distance from an operator (think thunder and lightning), and which the operator's brain might be able to interpret implicitly. However, additional psychophysical research, initially based in the laboratory, is most certainly needed here in order to further test the validity of these ideas. It should also be noted that other higher level cognitive processes such as emotions obviously have an impact on multisensory perception, which we do not have space to go into in any detail here [44].

III. APPLICATIONS

In this section, we discuss specific applications of multimodal interfaces in driving, aviation, the military, medicine, and sports, and the latest key developments in each of these research domains. For consistency with the rest of the paper, we restrict our review to applications designed to cue/reorient an operator's attention to a particular region of external space.

As mentioned already, laboratory-based studies of human attention indicate that additive (or, on occasion, super-additive) multisensory integration effects result in a

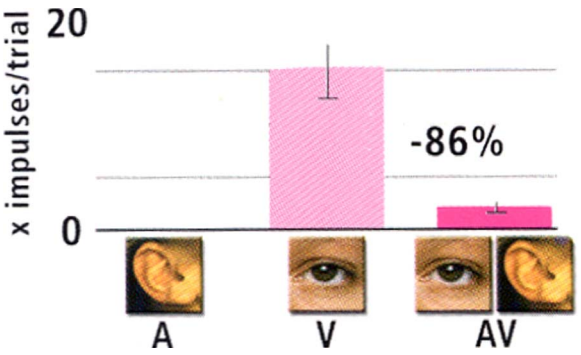

Fig. 2. ***Subadditivity: One of the rules of multisensory integration discovered at the single cell level in the superior colliculus. Subadditivity is more likely to be observed when the individual unimodal stimuli are presented from different positions and/or are presented outside the temporal window of integration.***

greater facilitation of performance for multimodal interfaces as compared to systems that present signals via only a single sensory modality. However, if the individual signals in a multisensory interface are presented in a way that creates some kind of sensory mismatch or conflict (e.g., they are presented from very different spatial locations or too far out of temporal synchrony), then subadditive effects (multisensory suppression) can occur, such that responses to the multimodal system are impaired relative to a more traditional unimodal system (see Fig. 2; see e.g., [45]). Unfortunately, as the proliferation of multimodal interfaces has often occurred without proper consideration of the basic mechanisms of human multisensory integration and cross-modal attention (e.g., [4], [8], and [46]), this lesson has often been learned the hard way, that is, through trial and error.

A. Spatially Congruent, Multimodal Collision Warnings Reduce Driver Reaction Time

Multimodal interfaces designed to alert a driver and direct their spatial attention to the location of a potential collision on the roadway is one area of research that has seen rapid growth and development over the past decade or so [47], [48]. Consistent with the psychophysical findings described earlier, research in this area has revealed that multimodal systems do not always lead to better performance than unimodal systems. In one representative study, for example, Lee *et al.* [49] investigated the effectiveness of an interface designed to reengage a driver's attention during adaptive cruise control (e.g., when an emergency intervention might be necessary on the part of the driver). Unimodal visual (an icon depicting a collision), auditory (a tone presented via a speaker on the dashboard), and two different tactile (seat vibration or brake pulse) signals were compared with a combination of all four signals. The results revealed that there were no significant differences in braking reaction times (RTs) between the multimodal combination and the unimodal signals.

However, this discrepancy can be understood by considering the additive/subadditive effects mentioned earlier. As noted by Spence and Ho [40], there was a large spatial separation between the different signals (seat/foot versus eye level). Thus, when Ho *et al.* [50] combined the presentation of auditory (a car horn presented via a speaker on the dashboard) and tactile (presented in the middle of the driver's stomach) signals in a spatially congruent manner, braking RTs were significantly shorter for multimodal signals as compared to the best of the unimodal signals. This study was conducted in a highly realistic driving simulator. The key difference between the warning signals used by Lee *et al.* [49] and those used by Ho *et al.* [50] seems to have been that, in the latter case, the two cues were both presented from in front of the driver. Similar benefits for multimodal signals have been shown in other driver warning systems with spatially congruent signals [40], [50], [51].

B. Similar Benefits for Unimodal and Multimodal Threat Warnings in Ground Combat Vehicles and Military Aircraft

Similar developments have occurred in the area of threat cuing for military ground combat vehicles and aircraft. So, for example, Oskarsson *et al.* [52] recently investigated the effectiveness of a multimodal threat warning system for combat vehicles that was composed of audio (virtual 3-D sounds presented via headphones), tactile (directional signals delivered via one of 12 tactors located on the operator's belt), and visual signals (lines indicating the direction of threat presented on a small display mounted in front of the operator). A simulator test revealed that performance (assessed by combining an accuracy measure in terms of localizing threat and the mean RT to orient to it) was better when the three signals were combined in a trimodal warning signal as compared to the best of the unimodal and bimodal signals. Somewhat surprisingly, multisensory facilitation occurred with this interface even though there was some spatial incongruence between the signals (that is, the auditory and tactile signals were presented from around the body while the visual signal was always presented from in front of the operator). It is possible that subadditive integration did not occur in this case because the different signals were associated with different types of attentional orienting: exogenous for tactile and auditory versus endogenous for visual. (Exogenous orienting refers to the stimulus-driven capture of attention by peripheral cues; endogenous orienting refers to the voluntary, or top–down, deployment of spatial attention [53].) It will be an interesting question for future research to investigate how best to combine endogenous and exogenous cues in multimodal cuing systems [53]. Another possibility is that a spatial ventriloquism effect may have occurred, that is, the visual signal may have been

"captured" by the tactile and auditory signal so that all three signals appeared to come from more-or-less the same spatial location [54].

In a second experiment, Oskarsson *et al.* [52] replaced the "head-down" visual threat cue [head-down display (HDD)] with a 3-D visual cue presented via a head-mounted display [i.e., so that all three cues were spatially congruent; head-up display (HUD)]. Performance was significantly better for the trimodal display with the HUD as compared to the trimodal display with the HDD. Similar benefits for multimodal directional cuing systems have now been observed in the context of military aviation [55], [56].

Elsewhere, Ngo *et al.* [57] have examined the possibility of using auditory, tactile, and multimodal (i.e., audiotactile) warning signals in addition to localized visual alerts in order to facilitate performance in a realistic air-traffic control training scenario. Their results demonstrated a significant enhancement in operator performance when auditory or audiotactile warning signals were presented. Surprisingly, in this particular study, unimodal tactile warning signals did not facilitate participants' performance over-and-above the unimodal visual alert.

C. Informative Multimodal Signals Used in Surgery Can Shift Attention to Critical Events and Increase Situational Awareness

A relatively new area for the application of multimodal interfaces for attentional orienting is medicine. Ferris and Sarter [58] recently investigated whether adding an informative tactile cuing system to the commonly used auditory and visual warning systems could improve the performance of anesthesiologists. In a surgery simulation situation, visual (popup messages on a patient screen in front of the anesthesiologist) and auditory (alarm sounds and a periodic signal that conveyed heart rate and blood oxygenation information) signals were combined with signals delivered via a vibrotactile vest that provided continuous information concerning the patient's status. Unlike the other multimodal interfaces described so far, this system was designed to both inform the operator about the nature of the critical event and orient his/her attention to a particular location in space. For example, lung volume information was presented via signals to the anesthesiologist's back while blood pressure information was presented via the arm. The combined trimodal interface was found to elicit superior performance than the auditory–visual interface that is commonly used. However, it should be noted that performance for the tactile system alone was not measured so it is unclear whether the improved performance was the result of multisensory facilitation. It will be interesting in future research to explore how informative warnings (i.e., that convey some information about the nature of the event that the operator must respond to) are best combined with warnings designed to reorient their spatial attention. (See [59] for a similar informative/reorienting interface that utilized the context of driving.)

D. Applying Multimodal Interfaces in Sports Training Is a Potentially Fruitful Area for the Future

Although to our knowledge there are no multimodal attentional cuing systems currently being developed for sports applications, there are a number of findings suggesting that this may be a fruitful area for future research. For example, Gray [60] demonstrated that virtual visual, auditory, and tactile signals (presented as feedback after a baseball batter completes a swing) can be used to improve skill acquisition in batting. In this study, tactile feedback was presented via tactors mounted on the bat, auditory feedback was presented via loudspeakers mounted directly behind the batter, and visual feedback was presented on a screen in front of the batter. Such results suggest that it is possible to create a multisensory interface that would direct an athlete's attention to the relevant cues for successful performance (see also [61]). Consistent with this notion, it has now been reported that unimodal visual cues that direct attention to information that can be used to anticipate an opponent's action can improve performance in badminton [62] and soccer [63].

E. How the Effectiveness of Multimodal Applications Depends on Operator Workload and Level of Experience Are Important Questions to Address in Future Research

Clearly, recent research has provided several examples of multimodal systems that seem to have great promise for improving operator performance and decreasing the incidence of errors/accidents in real operating environments. However, there are some important unresolved issues that should be addressed in order to optimize their effectiveness in future applications. First, as discussed in Section II, the effectiveness of multimodal signals appears to be highly dependent on the operator's mental workload. How well will the multimodal applications described above function in real operating environments in which the workload can be substantially higher than that observed in a laboratory or simulator?

In one of the few warning signal studies to have systematically varied operator workload, Mohebbi *et al.* [64] reported that the effectiveness of unimodal auditory and tactile collision warnings varied substantially with driver workload (cf. [65]). As shown in Fig. 3, auditory warning signals that were highly effective under low workload conditions (just driving) were rendered totally ineffective under moderate workload conditions (simple mobile phone conversation while driving) while tactile signals that were effective under both low and moderate workload conditions were significantly less effective in a high workload condition (complex phone conversation). Will the multimodal collision warning signals developed in

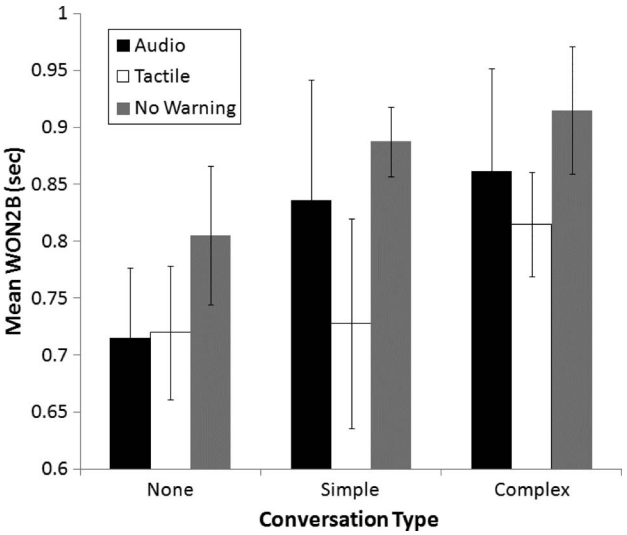

Fig. 3. ***The effect of driver workload on responses to auditory and tactile collision warnings. Workload was manipulated by varying the level of hands-free phone conversation during driving from no conversation to simple (casual conversation about weather, etc.) to complex (simulated business meeting). Note that WON2B is the braking reaction time. Figure reprinted from [64], SAGE Publications, Inc., all rights reserved ©.***

previous studies still produce significant benefits when tested under high workload conditions?

A related but unresolved issue here concerns the interplay between workload and any potential multisensory facilitation for aging operators. Intriguingly, it has been suggested that stimulating multiple sensory modalities might give rise to a greater performance gain in older people than in their younger counterparts [66], [67]. However, will such enhanced multisensory facilitation, which compensates the sensory decline experienced by the aging operators by overloading their senses, be sustained under high workload conditions? This constitutes another important question for future research.

A further important issue related to workload is whether the semantic content and delivery method of the warning should change as a function of operator workload. For example, in the anesthesiology study [58] described above, those multimodal displays that provided the operator continuous information concerning the patient's health resulted in better performance than warning signals that were only active when the situation became critical. This was presumably because the continuous signals allowed the operator to anticipate critical situations and respond more rapidly. However, the benefit of continuous signals was only observed when operator workload was low. Under high workload conditions, continuous signals tended to be ignored by operators with the result that performance was superior with noncontinuous warning signals.

Since it is common for workload to vary (i.e., between low and high values) in complex environments, it will likely prove rather difficult (if not impossible) to design a single system that will be effective under all conditions. One solution suggested by Ferris and Sarter [58] would be to develop adaptive systems that change the information presented to the operator as workload changes (i.e., continuous signals at low workload changing to critical warnings at high workload). However, it is not clear how well an operator will respond to such an adaptive attentional reorienting system. Research on adaptive systems in aviation suggests that the adaptivity may serve to increase workload and lead to problems of situational awareness [68], although these variable workload conditions remain to be tested. A similar issue, that has yet to be addressed empirically, is whether an operator will be able to pick up the semantic information delivered through auditory or haptic icons under conditions of high workload.

A second important issue that we feel has not been addressed adequately yet concerns how the experience-level/training of the operator influences the effectiveness of the multimodal interface. In many of the laboratory or simulator studies of multimodal cuing systems, participants are given limited training with the system before any experimental data are collected. Would the system be more effective with additional trainings? The amount of training/learning required for the multimodal interfaces to work effectively is rarely addressed in research. If these systems do require substantial training, it is highly likely that they will never be adopted in practice (e.g., it is unreasonable to expect that drivers will willingly accept that they must go through training before driving their new car away from the showrooms). Additionally, in most cases when these systems were tested in the laboratory, warning signals occurred systematically under controlled and simplified scenarios, in contrast to real-world scenarios in which there may just be no regularity. How well can people respond to warning signals if there are many different ones, and that they may not be presented with any degree of regularity?

At the other end of the spectrum, it is also unclear whether the types of multimodal interface described above will provide any benefit to highly experienced operators who have developed complex mental representations of their operating environment. For example, will highly experienced drivers who have learned to anticipate hazards [69] still benefit from being warned about a potential hazard? In the majority of research that has assessed the effectiveness of multimodal interfaces, the participants were relatively inexperienced operators of the systems under study. It will be an important question for future research to determine whether similar benefits of multimodal displays will also be found for experts.

IV. CONCLUDING REMARKS

It should hopefully be clear from the present review that multisensory warning signals hold the promise of

delivering more effective means of capturing and redirecting a human interface operator's spatial attention under attentionally demanding conditions. However, the most effective new multisensory warning signals will most likely need to be congruent in terms of their spatial location, timing, and semantic content in order to avoid incongruency and conflict. In the future, it is to be hoped that multimodal interface design will increasingly be based on the inherent limitations in human operators' ability to process multisensory information under demanding task conditions. ■

REFERENCES

[1] A. Gallace, M. K. Ngo, J. Sulaitis, and C. Spence, "Multisensory presence in virtual reality: Possibilities & limitations," in *Multiple Sensorial Media Advances and Applications: New Developments in MulSeMedia*, G. Ghinea, Ed. Hershey, PA: IGI Global, 2012, pp. 1–40.

[2] S. Oviatt, "Ten myths of multimodal interaction," *Commun. ACM*, vol. 42, pp. 74–81, 1999.

[3] C. Spence, M. E. R. Nicholls, and J. Driver, "The cost of expecting events in the wrong sensory modality," *Percept. Psychophys.*, vol. 63, pp. 330–336, 2001.

[4] C. Ho and C. Spence, *The Multisensory Driver: Implications for Ergonomic Car Interface Design*. Surrey, U.K.: Ashgate, 2008.

[5] S. Baillie, A. Crossan, N. Forrest, and S. May, "Developing the 'Ouch-o-Meter' to teach safe and effective use of pressure for palpation *Proceedings of EuroHaptics 2008*, vol. 5024, M. Ferre, Ed. Berlin, Germany: Springer-Verlag, 2008, pp. 912–917.

[6] C. Spence and J. Driver, "Audiovisual links in exogenous covert spatial orienting," *Percept. Psychophys.*, vol. 59, pp. 1–22, 1997.

[7] B. E. Stein and T. R. Stanford, "Multisensory integration: Current issues from the perspective of the single neuron," *Nature Rev. Neurosci.*, vol. 9, pp. 255–267, 2008.

[8] C. Spence and J. Driver, "A new approach to the design of multimodal warning signals," in *Engineering Psychology and Cognitive Ergonomics*, vol. 4, D. Harris, Ed. Surrey, U.K.: Ashgate, 1999, pp. 455–461.

[9] A. A. Kustov and D. L. Robinson, "Shared neural control of attentional shifts and eye movements," *Nature*, vol. 384, pp. 74–77, 1996.

[10] B. E. Stein and M. A. Meredith, *The Merging of the Senses*. Cambridge, MA: MIT Press, 1993.

[11] B. E. Stein, T. R. Stanford, M. T. Wallace, W. J. Vaughan, and W. Jiang, "Crossmodal spatial interactions in subcortical and cortical circuits," in *Crossmodal Space and Crossmodal Attention*, C. Spence and J. Driver, Eds. Oxford, U.K.: Oxford Univ. Press, 2004, pp. 25–50.

[12] N. P. Holmes and C. Spence, "Multisensory integration: Space, time, and superadditivity," *Current Biol.*, vol. 15, pp. R762–R764, 2005.

[13] L. A. Ross, D. Saint-Amour, V. M. Leavitt, D. C. Javitt, and J. J. Foxe, "Do you see what I am saying? Exploring visual enhancement of speech comprehension in noisy environments," *Cerebral Cortex*, vol. 17, pp. 1147–1153, 2007.

[14] S. Sinnett, S. Soto-Faraco, and C. Spence, "The co-occurrence of multisensory competition and facilitation," *Acta Psychologica*, vol. 128, pp. 153–161, 2008.

[15] M. O. Ernst and M. S. Banks, "Humans integrate visual and haptic information in a statistically optimal fashion," *Nature*, vol. 415, pp. 429–433, Jan. 2002.

[16] J. Trommershäuser, M. S. Landy, and K. P. E. Körding, Eds., *Sensory Cue Integration*. New York: Oxford Univ. Press, 2011.

[17] C. Spence and J. Driver, Eds., *Crossmodal Space and Crossmodal Attention*. Oxford, U.K.: Oxford Univ. Press, 2004.

[18] V. Santangelo and C. Spence, "Multisensory cues capture spatial attention regardless of perceptual load," *J. Exp. Psychol., Human Percept. Performance*, vol. 33, pp. 1311–1321, 2007.

[19] C. Spence, "Crossmodal spatial attention," *Ann. New York Acad. Sci. (The Year in Cognitive Neuroscience)*, vol. 1191, pp. 182–200, 2010.

[20] N. Lavie, "Distracted and confused? Selective attention under load," *Trends Cogn. Sci.*, vol. 9, pp. 75–82, 2005.

[21] C. Ho, V. Santangelo, and C. Spence, "Multisensory warning signals: When spatial correspondence matters," *Exp. Brain Res.*, vol. 195, pp. 261–272, 2009.

[22] E. Van der Burg, C. N. L. Olivers, A. W. Bronkhorst, and J. Theeuwes, "Non-spatial auditory signals improve spatial visual search," *J. Exp. Psychol., Human Percept. Performance*, vol. 34, pp. 1053–1065, 2008.

[23] M. K. Ngo and C. Spence, "Auditory, tactile, and multisensory cues facilitate search for dynamic visual stimuli," *Attention Percept. Psychophys.*, vol. 72, pp. 1654–1665, 2010.

[24] M. Ngo and C. Spence, "Crossmodal facilitation of masked visual target identification," *Attention Percept. Psychophys.*, vol. 72, pp. 1938–1947, 2010.

[25] C. Spence, "Multisensory perception, cognition, and behavior: Evaluating the factors modulating multisensory integration," in *The New Handbook of Multisensory Processing*, B. E. Stein, Ed. Cambridge, MA: MIT Press, 2012, pp. 241–264.

[26] C. Spence and M.-C. Ngo, "Does attention or multisensory integration explain the crossmodal facilitation of masked visual target identification?" in *The New Handbook of Multisensory Processing*, B. E. Stein, Ed. Cambridge, MA: MIT Press, 2012, pp. 345–358.

[27] N. Kitagawa and C. Spence, "Audiotactile multisensory interactions in human information processing," *Jpn. Psychol. Res.*, vol. 48, pp. 158–173, 2006.

[28] M. M. Murray, S. Molholm, C. M. Michel, D. J. Heslenfeld, W. Ritter, D. C. Javitt, and J. J. Foxe, "Grabbing your ear: Rapid auditory-somatosensory multisensory interactions in low-level sensory cortices are not constrained by stimulus alignment," *Cerebral Cortex*, vol. 15, pp. 963–974, 2005.

[29] P. J. Laurienti, R. A. Kraft, J. A. Maldjian, J. H. Burdette, and M. T. Wallace, "Semantic congruence is a critical factor in multisensory behavioral performance," *Exp. Brain Res.*, vol. 158, pp. 405–414, 2004.

[30] Y.-C. Chen and C. Spence, "When hearing the bark helps to identify the dog: Semantically-congruent sounds modulate the identification of masked pictures," *Cognition*, vol. 114, pp. 389–404, 2010.

[31] J. D. McKeown and S. Isherwood, "Mapping the urgency and pleasantness of speech, auditory icons, and abstract alarms to their referents within the vehicle," *Human Factors*, vol. 49, pp. 417–428, 2007.

[32] C. Ho and C. Spence, "Using peripersonal warning signals to orient a driver's gaze," *Human Factors*, vol. 51, pp. 539–556, 2009.

[33] W. W. Gaver, "Auditory icons: Using sound in computer interfaces," *Human-Computer Interact.*, vol. 2, pp. 167–177, 1986.

[34] I. Hwang, K. E. MacLean, M. Brehmer, J. Hendy, A. Sotirakopoulos, and S. Choi, "The haptic crayola effect: Exploring the role of naming in learning haptic stimuli," in *Proc. IEEE World Haptics Conf.*, 2011, pp. 385–390.

[35] L. A. Jones, B. Lockyer, and E. Piateski, "Tactile display and vibrotactile pattern recognition on the torso," *Adv. Robot.*, vol. 20, pp. 1359–1374, 2006.

[36] K. E. MacLean, "Foundations of transparency in tactile information design," *IEEE Trans. Haptics*, vol. 1, no. 2, pp. 84–95, Jul.–Dec. 2008.

[37] C. Brozzoli, L. Cardinali, F. Pavani, and A. Farnè, "Action-specific remapping of peripersonal space," *Neuropsychologia*, vol. 48, pp. 796–802, 2010.

[38] N. Kitagawa, M. Zampini, and C. Spence, "Audiotactile interactions in near and far space," *Exp. Brain Res.*, vol. 166, pp. 528–537, 2005.

[39] S. B. Brown, S. E. Lee, M. A. Perez, Z. R. Doerzaph, V. L. Neale, and T. A. Dingus, "Effects of haptic brake pulse warnings on driver behaviour during an intersection approach," in *Proc. 49th Annu. Meeting Human Factors Ergonom. Soc.*, 2005, pp. 1892–1896.

[40] C. Spence and C. Ho, "Tactile and multisensory spatial warning signals for drivers," *IEEE Trans. Haptics*, vol. 1, no. 2, pp. 121–129, Jul.–Dec. 2008.

[41] A. H. S. Chan and K. W. L. Chan, "Synchronous and asynchronous presentations of auditory and visual signals: Implications for control console design," *Appl. Ergonom.*, vol. 37, pp. 131–140, 2006.

[42] C. Spence and J. Driver, "A new approach to the design of multimodal warning signals," in *Engineering Psychology and Cognitive Ergonomics*, vol. 4, D. Harris, Ed. Surrey, U.K.: Ashgate, 1999, 455–461.

[43] C. Spence and S. Squire, "Multisensory integration: Maintaining the perception of synchrony," *Current Biol.*, vol. 13, pp. R519–R521, 2003.

[44] C. Ho and C. Spence, "Affective multisensory driver interface design," *Int. J. Veh. Noise Vib., Special Issue on Human Emotional Responses to Sound and Vibration in Automobiles*, to be published.

[45] G. M. Fitch, R. J. Kiefer, J. M. Hankey, and B. M. Kleiner, "Toward developing an approach for alerting drivers to the direction

of a crash threat," *Human Factors*, vol. 49, pp. 710–720, 2007.

[46] N. B. Sarter, "Multimodal information presentation: Design guidance and research challenges," *Int. J. Ind. Ergonom.*, vol. 36, pp. 439–445, 2006.

[47] S. C. de Vries, J. B. F. van Erp, and R. J. Kiefer, "Direction coding using a tactile chair," *Appl. Ergonom.*, vol. 40, pp. 477–484, 2009.

[48] J. H. Hogema, S. C. De Vries, J. B. F. Van Erp, and R. J. Kiefer, "A tactile seat for direction coding in car driving: Field evaluation," *IEEE Trans. Haptics*, vol. 2, no. 4, pp. 181–188, Oct.–Dec. 2009.

[49] J. D. Lee, D. V. McGehee, T. L. Brown, and D. Marshall, "Effects of adaptive cruise control and alert modality on driver performance," *Transp. Res. Record*, vol. 1980, pp. 49–56, 2006.

[50] C. Ho, N. Reed, and C. Spence, "Multisensory in-car warning signals for collision avoidance," *Human Factors*, vol. 49, pp. 1107–1114, 2007.

[51] J. B. F. van Erp and V. Veen, "Vibrotactile in-vehicle navigation system," *Transp. Res. F—Traffic Psychol. Behav.*, vol. 7, pp. 247–256, 2004.

[52] P.-E. Oskarsson, L. Eriksson, and O. Carlander, "Enhanced perception and performance by multimodal threat cueing in a simulated combat vehicle," *Human Factors*, vol. 54, pp. 122–137, 2012.

[53] C. Spence, "Crossmodal attention," *Scholarpedia*, vol. 5, p. 6309, 2010.

[54] C. Spence and J. J. Driver, "Attracting attention to the illusory location of a sound: Reflexive crossmodal orienting and ventriloquism," *NeuroReport*, vol. 11, pp. 2057–2061, 2000.

[55] G. L. Calhoun, J. Fontejon, M. Draper, H. A. Ruff, and B. Guilfoos, "Tactile vs. aural redundant alert cues for UAV control applications," in *Proc. HFES 48th Annu. Meeting*, 2004, pp. 137–141.

[56] J. B. F. van Erp, L. Eriksson, B. Levon, O. Carlander, J. A. Veltman, and W. K. Vos, "Tactile cueing effects on performance in simulated aerial combat with high acceleration," *Aviat. Space Environ. Med.*, vol. 78, pp. 1128–1134, 2007.

[57] M. K. Ngo, R. Pierce, and C. Spence, "Utilizing multisensory cues to facilitate air traffic management," *Human Factors*, vol. 54, pp. 1093–1103, 2012.

[58] T. K. Ferris and N. Sarter, "Continuously informing vibrotacile displays in support of attention managements and multitasking in anaesthesiology," *Human Factors*, vol. 53, pp. 600–611, 2011.

[59] R. Gray, "Looming auditory collision warnings for driving," *Human Factors*, vol. 53, pp. 63–74, 2011.

[60] R. Gray, "How do batters use visual, auditory, and tactile information about the success of a baseball swing?" *Res. Quart. Exercise Sport*, vol. 80, pp. 491–501, 2009.

[61] J. B. F. Van Erp, I. Saturday, and C. Jansen, "Application of tactile displays in sports: Where to, how and when to move," in *Proc. Eurohaptics*, 2006, pp. 105–109.

[62] N. Hagemann, B. Strauss, and R. Canal-Bruland, "Training perceptual skill by orienting visual attention," *J. Sport Exercise Psychol.*, vol. 28, pp. 143–158, 2006.

[63] R. Canal-Bruland, "Guiding visual attention in decision making—Verbal instructions versus flicker cueing," *Res. Quart. Exercise Sport*, vol. 80, pp. 369–374, 2009.

[64] R. Mohebbi, R. Gray, and H. Z. Tan, "Driver reaction time to tactile and auditory read-end collision warnings while talking on a cell phone," *Human Factors: J. Human Factors Ergonom. Soc.*, vol. 51, pp. 102–110, 2009.

[65] R. Tilak, I. Xholi, D. Schowalter, T. Ferris, S. Hameed, and N. Sarter, "Crossmodal links in attention in the driving environment: The roles of cueing modality, signal timing, and workload," in *Proc. Human Factors Ergonom. Soc. Annu. Meeting*, 2008, vol. 52, pp. 1815–1819.

[66] P. J. Laurienti, J. H. Burdette, J. A. Maldjian, and M. T. Wallace, "Enhanced multisensory integration in older adults," *Neurobiol. Aging*, vol. 27, pp. 1155–1163, 2006.

[67] C. Spence and C. Ho, "Multisensory driver interface design: Past, present, and future," *Ergonomics*, vol. 51, pp. 65–70, 2008.

[68] D. B. Kaber, C. M. Perry, N. Segall, C. K. McClernon, and L. J. Prinzel, "Situation awareness implications of adaptive automation for information processing in an air traffic control-related task," *Int. J. Ind. Ergonom.*, vol. 36, pp. 447–462, 2006.

[69] C. T. Scialfa, D. Borkenhagen, D. J. Lyon, H. Horswill, and M. Wetton, "The effects of driving experiences on responses to a static hazard perception test," *Accident Anal. Prevention*, vol. 45, pp. 547–553, 2012.

ABOUT THE AUTHORS

Rob Gray received the B.A. degree in psychology from Queens University, Kingston, ON, Canada, in 1993 and the M.S. and Ph.D. degrees in psychology from York University, Toronto, ON, Canada, in 1995 and 1998, respectively.

He was a Research Acientist with Nissan Cambridge Basic Research from 1998 to 2001 before joining the Department of Applied Psychology, Arizona State University, Tempe. From 2006 to 2010, he served as Department Head of the Department of Applied Psychology and as a Research Psychologist for the United States Air Forces. He is currently a Reader in Perception & Action in the School of Sport and Exercise Sciences, University of Birmingham, Birmingham, U.K. He is the author of more than 60 published refereed journal articles and chapters.

Dr. Gray serves as an Associate Editors for the *Journal of Experimental Psychology: Human Perception and Performance* and is an editorial board member for *Human Factors*. In 2007, he was awarded the Distinguished Scientific Award for Early Career Contribution to Psychology from the American Psychological Association and the Earl Alluisi Award for Early Career Achievement in the Field of Applied Experimental and Engineering Psychology.

Charles Spence received the Ph.D. degree in experimental psychology from the Department of Psychology, University of Cambridge, Cambridge, U.K., in 1998.

Currently, he is a University Professor at the Department of Experimental Psychology, Oxford University, Oxford, U.K. He heads the Crossmodal Research Laboratory at the Department of Experimental Psychology, Oxford University (http://www.psy.ox.ac.uk/xmodal). He has published over 400 articles in top-flight scientific journals over the last decade.

Prof. Spence has been awarded the Tenth Experimental Psychology Society Prize; the British Psychology Society: Cognitive Section Award; the Paul Bertelson Award, recognizing him as the young European Cognitive Psychologist of the Year; and the Friedrich Wilhelm Bessel Research Award from the Alexander von Humboldt Foundation in Germany. He is currently an Associate Editor for the *Journal of Experimental Psychology: General and Seeing & Perceiving*.

Cristy Ho received the B.Cog.Sc. degree from The University of Hong Kong, Hong Kong, in 2001, the M.Sc. degree in human–computer interaction with ergonomics from the University College London, London, U.K., in 2003, and the D.Phil. degree in experimental psychology from the University of Oxford, Oxford, U.K., in 2006.

Currently, she is a Postdoctoral Research Scientist at the Crossmodal Research Laboratory, University of Oxford.

Dr. Ho was awarded the American Psychology Association's New Investigator Award in Experimental Psychology: Applied, in 2006.

Hong Z. Tan (Senior Member, IEEE) received the B.S. degree in biomedical engineering from Shanghai Jiao Tong University, Shanghai, China, in 1986 and the M.S. and Ph.D. degrees in electrical engineering and computer science, from the Massachusetts Institute of Technology (MIT), Cambridge, in 1988 and 1996, respectively.

She was a Research Scientist at the MIT Media Lab from 1996 to 1998 before joining the faculty at Purdue University, West Lafayette, IN. She is a Professor of Electrical and Computer Engineering, with courtesy appointments in the School of Mechanical Engineering and the Department of Psychological Sciences. She is also a Senior Researcher at Microsoft Research Asia, Beijing, China. She founded and directs the Haptic Interface Research Laboratory at Purdue University. Her research focuses on haptic human–machine interfaces in the areas of haptic perception, rendering, and multimodal performance.

Dr. Tan is currently the Editor-in-Chief of the World Haptics Conference editorial board. She served as the Founding Chair of the IEEE Technical Committee on Haptics from 2006 to 2008. She was a recipient of the National Science Foundation CAREER award from 2000 to 2004.