

# Cross-modal detection using various temporal and spatial configurations

James A. Schirillo

Published online: 16 November 2010  
© Psychonomic Society, Inc. 2010

**Abstract** To better understand temporal and spatial cross-modal interactions, two signal detection experiments were conducted in which an auditory target was sometimes accompanied by an irrelevant flash of light. In the first, a psychometric function for detecting a unisensory auditory target in varying signal-to-noise ratios (SNRs) was derived. Then auditory target detection was measured while an irrelevant light was presented with light/sound stimulus onset asynchronies (SOAs) between 0 and  $\pm 700$  ms. When the light preceded the sound by 100 ms or was coincident, target detection ( $d'$ ) improved for low SNR conditions. In contrast, for larger SOAs (350 and 700 ms), the behavioral gain resulted from a change in both  $d'$  and response criterion ( $\beta$ ). However, when the light followed the sound, performance changed little. In the second experiment, observers detected multimodal target sounds at eccentricities of  $\pm 8^\circ$ , and  $\pm 24^\circ$ . Sensitivity benefits occurred at both locations, with a larger change at the more peripheral location. Thus, both temporal and spatial factors affect signal detection measures, effectively parsing sensory and decision-making processes.

**Keywords** Cross-modal detection · Signal detection theory

Multisensory integration can influence detection and localization of an audiovisual cross-modal signal. For example, depending on their spatial and/or temporal relationships, irrelevant visual stimuli can enhance or depress detection of an auditory target stimulus (Bothe & Marks, 1970; Treisman

& Howarth, 1959). Although numerous studies have examined the spatial effects of cross-modal processes, few have explored their temporal relationships, and those often have confounded sensory and decision-making processes (London, 1954). For example, Ozbaydar (1961) showed that observers were better able to detect a 500-ms 1200-Hz tone when the lights in the room were on, attributing this peculiar result to changes in either sensory processes or a criterion bias.

This becomes relevant when multisensory information makes complex signals appear more salient than their components. Stein, London, Wilkinson, and Price (1996) had observers judge the intensity of a visual signal (a red LED), using a visual analog scale. When a spatially coincident ( $0^\circ$ ) or noncoincident ( $\pm 45^\circ$ ) auditory stimulus (a 40-ms 45-dB SPL broadband 20–20-kHz sound) was simultaneously presented, perceived visual intensity was most enhanced (a 73.5% increase). Effects weakened proportionately as the intensity of the visual stimulus increased, with only 7.9% enhancement at the highest signal intensity. This enhancement is evident not only with supraliminal stimuli, but also with cross-modal stimuli closer to threshold. In fact, these findings concur with the *law of inverse effectiveness* (Frassinetti, Pavani, & Làdavas, 2002; Lovelace, Stein, & Wallace, 2003; Meredith & Stein, 1983, 1985; Stein et al., 1996), which states that responses are most enhanced with near-threshold stimuli, as compared with suprathreshold stimuli.

However, Odgaard, Arieh, and Marks (2003) varied the time (25%, 50%, 75%) that noise accompanied a near-threshold light and showed that the enhancement disappeared at 75%, as compared with 50%, suggesting that it reflected response bias, not an early sensory level of processing. They supported their claim by repeating the study using a paired-comparison method, which is not as

---

J. A. Schirillo (✉)  
Department of Psychology, Wake Forest University,  
Winston-Salem, NC 27109, USA  
e-mail: schirija@wfu.edu

susceptible to response bias as the rating task in Stein et al. (1996). In this case, they found no evidence of cross-modal interaction. Participants showed no tendency to assess the light-plus-noise as brighter than the light alone when the luminances were equal. The methodological differences between Stein et al. (1996) and Odgaard et al. suggested our use of a signal detection, not a rating, method, which presents lights versus lights-and-sounds in an equal number of trials as the simplest way to parse sensory effects from response bias.

The fact that the effectiveness of the auditory stimulus was not limited to one location in Stein et al. (1996) contrasts with observations of some multisensory neurons, particularly those in the superior colliculus, where enhancement is very sensitive to the spatial relationship of the auditory and visual stimuli (Meredith & Stein, 1985; Wallace, Meredith, & Stein, 1992). One way to test the effects of location on enhancement might be to use auditory stimuli as the target and spatially coincident and non-coincident visual stimuli as irrelevant cues. The present study explored this approach.

Since auditory localization is best at midline and gets increasingly worse toward the periphery, varying locations may enhance detection. It improves slightly around 90° (along the interaural axis) but is still not as good as around 0° (Makous & Middlebrooks, 1990; Mills, 1958). Although visual acuity is always better than auditory acuity, it also decreases toward the periphery (Hess, Baker, May, & Wang, 2008). For example, near fixation, the minimum visual angle is only approximately 0.06°, as compared with a minimum auditory angle of approximately 1.1°. However, the visual spatial resolution advantage over audition diminishes as the two stimuli move further toward the periphery (compare Mills, 1958, Fig. 6, with Hess et al., (2008), Fig. 3). Determining whether the change in peripheral cross-modal acuity can generalize to detection is especially interesting.

In the temporal domain, an early study by Child and Wendt (1938) examined cross-modal detection by varying the temporal interval between a near-threshold auditory target and an extraneous suprathreshold light. Their yes/no detection task contained trials with both stimuli, neither stimulus, or each stimulus presented alone. They found that detection was most enhanced when the visual stimulus preceded the auditory target by 500 ms.

Fortuitously, because they included catch trials without sound (either no stimulus or light alone), Lovelace (personal communication, October 2002) was able to reanalyze their data in terms of hits and false alarms. Using signal detection theory (SDT), he showed that sensitivity, measured as  $d'$ , did not vary across time intervals, whereas criterion shifts, measured as  $\beta$ , accounted for the change in performance at longer (stimulus onset asynchrony [SOA])

intervals. Whereas Child and Wendt (1938) proposed that temporal factors regulate enhanced detection, criterion shifts illustrate how decision mechanisms might affect expectancy and subsequently influence what they claimed were sensory processes.

For example, a warning cue can reduce uncertainty (Hawkins et al., 1990; Luck, Hillyard, Mouloua, & Hawkins, 1996). Gregg and Brogden (1952) showed that a near-threshold auxiliary contiguous light could increase sensitivity to a 200-ms 1000-Hz tone when observers were not required to report the light's presence. If asked to report on the light and the tone, sensitivity to the tone decreased, suggesting that auditory acuity may be inhibited by decisional, rather than sensory, effects. Performance also improves as a stimulus becomes more probable, which can occur by changing attentional resource allocations (Ciaramitaro, Cameron, & Glimcher, 2001); uncertainty is reduced when attentional focus is narrowed in time and space (Treisman & Howarth, 1959; Ward, McDonald, & Golestani, 1998; Ward, McDonald, & Lin, 2000).

To confirm that auditory detection of 200-ms broadband noise could be enhanced by an irrelevant visual stimulus (200-ms LED) without the confounding effects of response bias, Lovelace et al. (2003) used a signal detection paradigm with feedback. Their second experiment included blocking trials that provided light only, sound and light together, no stimuli, and sound only and reduced the window of presentation, with an LED signaling the beginning of a trial, from 1,000 ms in their first experiment to 500 ms. They argued that these factors corrected for a potential response bias in the first experiment, based on the impression that a sound was presented more frequently in trials with than without light. They also argued that making the timing of the sounds as predictable as possible (only a 500-ms presentation window) helped to minimize temporal uncertainty. As compared with their first experiment, the second experiment blocked trials and minimized the temporal window, changing  $d'$  without corresponding changes in  $\beta$ .

In the following two experiments, an auditory target and an irrelevant visual signal were used to address how timing affects attentional (response bias) versus sensory (detection) factors and how spatial factors may be affected by the law of inverse effectiveness.

### Experiment 1a: establishing a unimodal psychometric function

An auditory psychometric function determined the intensity of an auditory signal on any given yes/no trial based on stimulus set magnitudes and the observer's prior knowledge.

## Method

**Participants** Twenty-five Wake Forest University undergraduate students (8 males; all between the ages of 18 and 22 years) from introductory psychology classes participated in this experiment for course credit. The study was performed in accordance with the ethical standards of the Declaration of Helsinki, and IRB approval was obtained. An initial screening session determined that all the experimentally naïve observers' hearing was normal between 50 Hz and 8 kHz, determined by undergoing a standard pure tone audiometry performed with a GSI 61 clinical audiometer (Viasys NeuroCare,; Madison, WI) and dynamic, open, diffuse-field, studio headphones from Beyerdynamic (DT 990 Pro; Nominal frequency response, 5Hz–35kHz), and they had normal or corrected-to-normal vision.

**Stimuli and apparatus** The stimulus apparatus was located in a dimly lit sound-attenuated room (Industrial Acoustics Co., Model 1203). A GoldLine Personal White Noise Maker produced broadband noise at approximately 50 dB SPL to mask any ambient environmental noise. Observers sat at a table 70 cm in front of a CRT screen with their head supported by an adjustable chinrest. The apparatus consisted of a 36-in. Super Fine Pitch FD Trinitron CRT monitor (Sony) with two 30-W Bose speakers (Model A9319) placed 2 cm to the left and the right of the monitor. The CRT presented a fixation light, a visual flash, and instructions (white letters on a black background). Both speakers simultaneously presented the sound to produce the illusion of a central source. Both visual and auditory stimuli were located at eye level. A computer keyboard (located 30 cm between the chinrest and the CRT) was used to collect responses. A small lamp located under the table provided sufficient light to see the keyboard. A small cutout cardboard cover overlay the keyboard to prevent superfluous key presses. A Pentium class personal computer (IBM Model 4BU) utilizing E-Prime software (Version 1.1; Schneider, Eschman, & Zuccolotto, 2002) modulated the visual and auditory stimuli. An oscilloscope and a photodiode were used to ensure that the timing was accurate, which it was.

**Procedure and design** The auditory target stimulus was a 50-ms broadband noise burst (ranging from 60 to 8000 Hz) containing a 5-ms ramp on and 5-ms ramp off. The peak intensity was 65 dB SPL, and the average starting intensity was 53 dB SPL. The average ensured that the output intensities of each of the speakers produced equivalent ILDs. These values were determined by integrating the signal over 50 ms.

An adaptive staircase procedure used a modified QUEST method (Watson & Pelli, 1983) to determine the initial auditory detection intensity for each observer. Using a

termination criterion set at 64 trials, the yes/no QUEST method set thresholds to an accuracy level of approximately 50% correct. The use of 64 trials was determined by consulting the well-cited Watson and Pelli (1983) article on the modified QUEST method and by conducting a pilot study using 5 naïve observers (who were not used in the rest of the experiments) that showed that this termination point gave a reasonable threshold measure. The experimenter then measured the target sound in decibels, directly after the experiment, from the location of the chinrest. In a subsequent experimental session, the sound was randomly set to one of nine intensities from  $\pm 8$  dB in 2-dB steps of the 50% correct performance. These settings produced a psychometric function that ranged from approximately 2% to 98% correct. Each trial interval (i.e., containing either sound or no sound) was presented within a 2,000-ms stimulus interval. Figure 1 illustrates the subsequent experimental trial procedure.

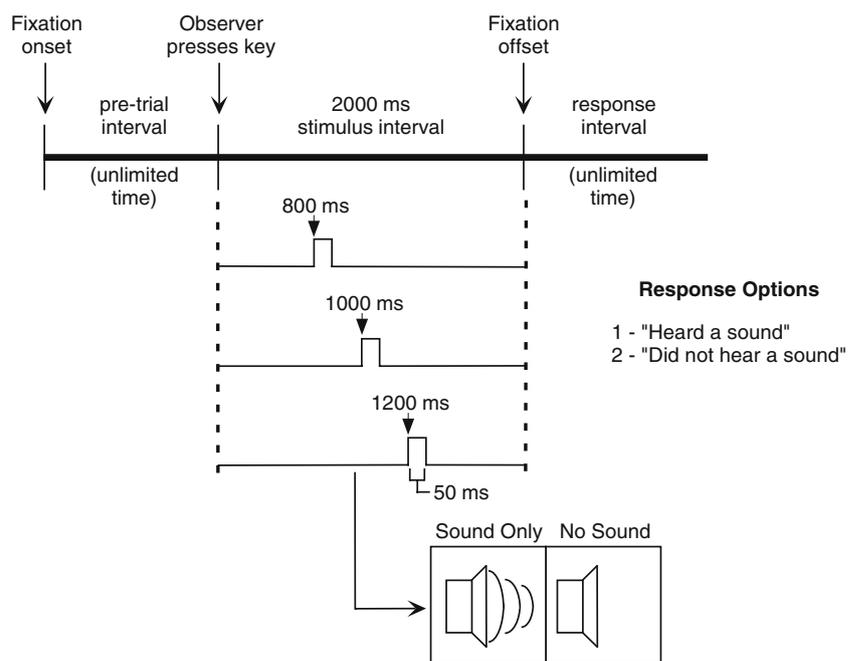
At the onset of a trial, a gray fixation ring (3.22° outer diameter, 0.16° thick) appeared in the center of the CRT screen signaling the observer to press any key to begin the trial. No sound occurred during the 2,000-ms stimulus interval on 50% of the trials. On the other half of the trials, a 50-ms sound burst was presented randomly at 800, 1,000, or 1,200 ms after the beginning of the stimulus interval. This variable onset was to minimize temporal cues to target onset. At the end of the stimulus interval, the fixation ring disappeared, and response instructions appeared, which stated that the observer should press “1” to indicate that they heard a sound or “2” to indicate that they did not hear a sound. The observers had unlimited time to respond, and there was no feedback. Once the computer recorded their response, the fixation ring reappeared, prompting the observer to begin the next trial. The observers could pause between trials if they wished.

A single session consisted of the QUEST adaptive staircase (64 yes/no trials) and two blocks of 216 experimental trials, where each block contains 4 trials of each stimulus combination (i.e., nine sound intensities, three sound delays) and an equal number of trials with no sound. All the trials within a block occurred randomly without replacement. A 2-min break separated the two experimental blocks, and the experimental session took approximately 70 min.

## Results and discussion

A psychometric function describes the data (Fig. 2). The fitted curve was produced on the average of the 25 psychometric functions (i.e., the raw proportions). SEMs represent the between-subjects variability in percentage correct measured during the experimental session. Thus, the

**Fig. 1** Trial sequence for Experiment 1

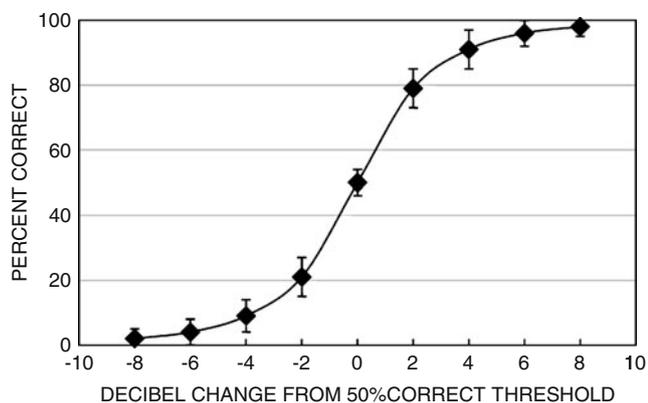


group mean psychometric function shows the range and dispersion of the measured thresholds. The individual psychometric functions that composed this function were used to set the sound intensity levels in later experiments.

### Experiment 1b: establishing a multimodal psychometric function

As in the Child and Wendt's (1938) study, the sounds we used might or might not have flashes of light presented before or after them. This allowed the signal detection paradigm to measure both sensitivity ( $d'$ ) and criterion shifts ( $\beta$ ). Instructions informed the observers that the

visual signal had no predictive value regarding the presence of the target sound. However, it was hypothesized that a light presented just before or concurrent with the target sound would elicit better detection performance, as compared with when only sound occurred. At longer SOAs, it is more likely that performance would improve due only to shifts in criterion, as opposed to improvements in sensitivity. There should be no difference in detection performance when the light followed the target sound, since, in these cases, the light is not a temporal cue for the sound. Also expected was a larger gain in performance for sounds presented at lower intensities when accompanied by a light, as compared with when the same intensity sound was presented alone, predicted by the law of inverse effectiveness (Frassinetti et al., 2002; Lovelace et al., 2003; Stein et al., 1996).



**Fig. 2** Average percentage correct psychometric function constructed from 50% correct threshold established using *Quest*. Error bars show standard errors of the means

### Method

**Participants** The same group of students participated for course credit.

**Stimuli and apparatus** The stimulus apparatus was the same one used in Experiment 1a. However, in this experiment, the dim ( $0.43 \text{ cd/m}^2$ ) gray central fixation ring ( $3.22^\circ$  outer diameter,  $0.16^\circ$  thick) surrounded a light ( $1.45^\circ$  in diameter,  $77 \text{ cd/m}^2$ ) that might flash in the middle of the ring for 50 ms.

**Procedure and design** After receiving instructions, each observer's psychometric function from Experiment 1a determined their set of intensities to generate a multimodal

psychometric function, using a method of constant stimuli. The trial procedure was identical to Experiment 1a, except that whereas in all experimental blocks there was a 50% chance that the sound would be present, in the multisensory blocks, a 50-ms light flashed within the gray fixation ring. Thus, the multisensory trials represent 25% of the total trial structure (i.e., 50% no sound, 25% sound alone, 25% multisensory). The light occurred concurrently, before (minus SOA), or after (plus SOA) the sound by 100, 350, or 700 ms. At the end of the 2,000-ms stimulus interval, the fixation ring disappeared, and response instructions appeared. The observer indicated whether or not they heard a sound by pressing “1” to indicate that they heard a sound or “2” to indicate that they did not hear a sound. Before the session, observers were told the following: “A sound will be presented on half of the trials; you are to report whether you detect it or not on each trial. In addition, a light will flash on half of the trials. Since the light occurs on only half of the trials, it has no predictive value that a sound will occur in that trial and should be ignored.”

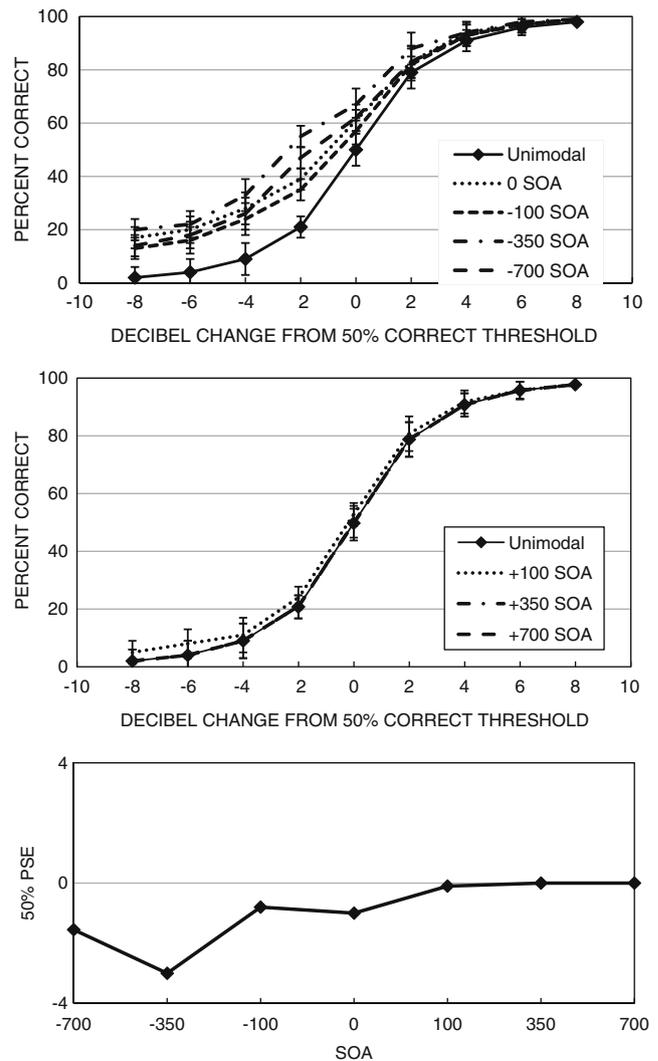
A single experimental session consisted of four blocks of 252 experimental trials each, for a total of 1,008 trials. Each block had 2 trials of each multisensory temporal combination (i.e., concurrent,  $\pm 100$  ms,  $\pm 350$  ms, and  $\pm 700$  ms) at each of nine intensity levels and an equal number of trials with no sound. All the trials within a block occurred randomly without replacement. A 5-min break separated the second and third blocks. An experimental session took approximately 2 hr.

Results and discussion

The statistical analysis was a repeated measures ANOVA. The results from the sound-only (unimodal) condition approximated 50% correct and served as a sensitivity and criterion baseline (Fig. 3a, b).

In the multisensory conditions when the light was present concurrently with the sound (i.e., 0-ms SOA) or whenever the light preceded the sound (i.e., 100-, 350-, or 700-ms SOA), there was an increase in the percentage correct at the lower intensities (Fig. 3a). It had been hypothesized that there would be a difference between negative SOAs and the unimodal condition, but not between positive SOAs and the unimodal condition. Thus, 2 two-way repeated measures ANOVAs were performed. One had negative and 0-ms SOAs and intensity as factors, and the other had positive SOA and intensity as factors. For both, there was an expected main effect of intensity.

For the positive SOA ANOVA, there was no statistically significant interaction between SOA and intensity. Therefore, this analysis was not pursued further. However, for the negative and 0-ms SOA ANOVA, there was a statistically



**Fig. 3** Average percentage correct psychometric functions established from psychometric function intensities established in Experiment 1a. The upper panel includes the sound-only (unimodal) condition (comparable to Fig. 2) and multimodal conditions (0 SOA = temporally coincident; minus SOAs indicating light occurring before the sound). The middle panel includes the sound-only (unimodal) condition (comparable to Fig. 2) and multimodal conditions (plus SOAs indicating light occurring after the sound). Error bars show standard errors of the means. The lower panel plots the 50% PSEs obtained from Experiments 3a and 3b as a function of SOA

significant interaction between SOA and intensity. To minimize possible Type I error, scores were collapsed across negative and 0-ms SOAs, and nine planned comparisons were performed (i.e., one for each intensity level) against the unimodal condition. The negative decibel changes (i.e., 2, 4, 6 and 8 dB) were all statistically significantly different from the unimodal condition, with the highest probability being  $p < .02$ ,  $F(1, 24) = 5.97$ . All the other decibel changes (i.e., 0, 2, 4, 6, and 8 dB) were not statistically significantly different from the unimodal condition, with the lowest probability being

$p > .11$ ,  $F(1, 24) = 2.73$ . Figure 3c plots the 50% PSEs obtained from Fig. 3a and b as a function of SOA. It shows the maximal effect at 350 ms, which falls off at both longer and shorter SOAs, and the fact that this function is flat at positive SOAs.

The statistical measures (i.e.,  $d'$  for sensitivity and  $\beta$  for criterion) derived from SDT (Green & Swets, 1966) describe the data. The analysis combined the hit and false alarm rates for each observer before generating receiver operating characteristic (ROC) curves. Averaging hit and false alarm rates across observers yields a measure similar to averaging sensitivity across individual observers (Haase, Theios, & Jenison, 1999). Although this may mildly underestimate the true value of  $d'$ , the reduction in variability is advantageous, since the collapsed measure is more consistent with that in situations in which a few observers participate in many trials (Macmillan & Kaplan, 1985). Macmillan and Kaplan argued that performing SDT on averaged data is often better than the alternatives. In essence, they claimed that it is prudent to collapse  $d'$  by first averaging the hits and false alarms of the observers and then calculating  $d'$  from the averaged proportions. They claim that collapsed  $d'$  is nearest to true average  $d'$  if the observers have similar sensitivities, as well as biases. This was verified in the present data set; observers were consistent in both their sensitivities and their biases. Macmillan and Kaplan claimed that computing a collapsed  $d'$  from averaged data can be a reliable, unbiased way to estimate true average  $d'$ . Given the sufficient number of observers in the present study, it is not expected that slightly unequal variances between the *noise* and *signal plus noise* distributions are an issue. Yet, one concern is that separate data points from each observer that are fit to one ROC curve might actually fit better on separate ROC curves (Markowitz & Swets, 1967; Tanner & Sorkin, 1972). Not changing the a priori probabilities across observers allays this concern. Individual-observer ROC curves (not shown) served as a check against this pooled data, ensuring that this averaging method was suitable.

Therefore, examination of the resulting ROC curves indicates in what multisensory conditions an increase in the percentage correct at the lower intensities was due to an increase in sensitivity ( $d'$ ), beta ( $\beta$ ), or both (Fig. 4). Figure 4 is an ROC plot. Iso-contour lines along the diagonal represent various levels of constant  $d'$  from  $d' = 0$  ( $45^\circ$  line) to  $d' = 1.8$ . The lines along the negative diagonal represent conditions with the same criterion from  $\beta = .1$  to  $\beta = 1.0$ . This allows one to readily see that Fig. 4 depicts that in the 0-ms SOA (temporally concurrent) condition (designated “0”), there was an increase in sensitivity ( $d'$ ), as compared with the unimodal sound condition (designated “1”), reflecting enhancement due to multisensory integration. However, there was no statistically significant differ-

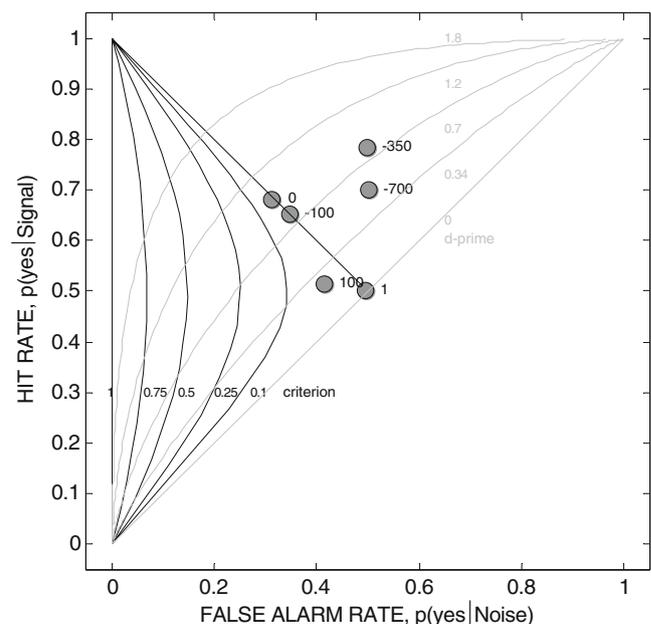
ence in beta ( $\beta$ ) (Table 1). In conditions in which the light preceded the sound by 100 ms (100-ms SOA), an increase in sensitivity ( $d'$ ) also occurred, without a significant shift in beta ( $\beta$ ) (Table 1).

However, when the light preceded the sound by either 350 or 700 ms, there was only a small increase in  $d'$  but a significant shift in criteria ( $\beta$ ) (Table 1).

This finding agrees with the reanalyzed Child and Wendt (1938) data. That is, when the light precedes the sound by a significant margin, it can serve as a cue to the observer, causing them to respond more often that a signal is probable. This increases both false alarms and hits. Notably, when  $d'$  was significant and  $\beta$  was not or vice versa, this was true for all the observers, not just a subset of the participant pool.

Yet when the sound preceded the light by either 350 or 700 ms, there was no change in sensitivity or criterion relative to the unimodal sound condition (data points not shown in Fig. 4 for +350- or +700-ms SOA, in that they overlap with the unimodal data point; see Table 1), suggesting that the light no longer served as a temporal cue.

Of interest was whether there was a correlation between unimodal and multimodal thresholds. For example, perhaps observers with high thresholds were more (or less) affected by the visual stimulus. This turned out to be statistically insignificant (e.g.,  $r^2 = .081$ ,  $p = .427$ ).



**Fig. 4** Averaged ROC curve. The gray lines indicate several  $d'$  values, and the black lines indicate iso-criterion ( $\beta$ ) shifts. 1 = unimodal; 0 = 0 SOA; minuses = (light precedes sound) 100, 350, and 700 SOAs; while plus = (sound precedes light) 100 SOA. 0 and -100 SOAs are significantly more sensitive than the unimodal condition, whereas 350 and 700 SOAs have a significantly different criterion than the unimodal condition

**Table 1** Wilcoxon statistical significance results for  $d'$  and  $\beta$  obtained from Fig. 4

SOA	Measure	Wilcoxon: W(8)	Z	P
0	$d'$	-0.010 and 0.957	-2.67	0.008
0	$\beta$	1.014 and 0.989	-0.53	0.594
-100	$d'$	-0.010 and 0.787	2.67	0.008
-100	$\beta$	1.014 and 0.991	-0.42	0.678
-350	$d'$	0.010 and 0.799	-1.36	0.173
-350	$\beta$	1.014 and 0.734	-2.19	0.028
-700	$d'$	-0.010 and 0.417	-1.01	0.314
-700	$\beta$	1.014 and 0.627	-2.31	0.021
100	$d'$	-0.010 and 0.267	-1.01	0.314
100	$\beta$	1.014 and 1.033	-0.18	0.859
350	$d'$	-0.010 and 0.016	-0.42	0.674
350	$\beta$	1.014 and 1.023	-0.14	0.889
700	$d'$	-0.010 and -0.003	-0.28	0.779
700	$\beta$	.014 and 0.997	-0.14	0.889

## Experiment 2

Observers detected sounds that either did or did not have a flash of light presented with the sound at some eccentricity. There are two interesting potential outcomes for this experiment, depending on whether a peripheral light has the same detection-enhancing effect on an auditory stimulus as a foveal cross-modal pair. Perrott (1993) showed that the reduction in a peripheral light's localization acuity does *not* diminish its localization-enhancing effect on coincident auditory targets. If this finding generalizes to detection, a peripheral cross-modal event should produce as much enhancement as the same event presented centrally. However, Stein et al. (1996) might predict that any reduction in the saliency of a cross-modal peripheral light would evoke the law of inverse effectiveness, since signals closer to threshold traditionally result in the greatest level of cross-modal enhancement. If this is the case in the present study, less salient peripheral visual distractors should be *more* likely to result in enhanced detection of the auditory target.

### Method

**Participants** The same group of students at that in Experiments 1a and 1b participated for course credit.

**Stimuli and apparatus** The stimulus array was similar to the one used in Experiments 1a and 1b. However, this experiment did not use the speakers placed to the left and right of the monitor. Instead, the 36-in. monitor had small (2.2° in diameter) translucent speakers placed at four equidistant locations spanning 80 cm. Each speaker was

directly below where a light (1.45° in diameter) could flash. An ASL Model 510 eye-tracking system ensured that observers maintained proper fixation.

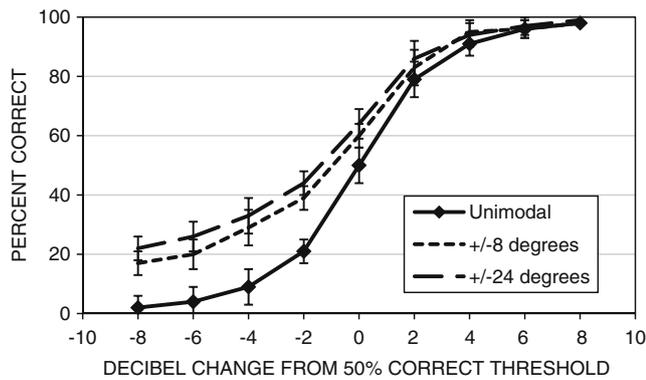
As in the earlier experiments, each observer's psychometric function from Experiment 1a was used to determine the intensities to present the auditory stimuli. The sound duration was 50 ms, and only one speaker presented the sound on a given trial. The light (1.45° in diameter, 77 cd/m<sup>2</sup>) and sound were concurrent (0-ms SOA).

**Procedure and design** After the observers had received instructions, the sound intensities from the psychometric function derived in Experiment 1a were used. Except for the different spatial locations, the stimuli were presented using a method of constant stimuli, as in Experiment 1b. Prior to the onset of a trial, the fixation ring appeared at the middle of the CRT screen signaling the observer to press any key to begin the trial. Data analyses included only those trials that had proper fixation throughout the course of a trial verified with an eyetracker (within a 3° window). On half of the trials, a 50-ms sound burst was randomly presented at one of four possible locations (i.e., ±8° or ±24° from the central fixation). In addition, a 50-ms flash of light could appear spatially just above one of the four speaker locations for 50 ms on half of the trials. A single session consisted of four blocks of 144 experimental trials each, for a total of 576 trials. Each block contains 2 trials of each multisensory spatial combination (i.e., four sound locations) at each of nine intensity levels and an equal number of trials with no sound. All the trials within a block randomly occurred without replacement. A 5-min break separated the two experimental blocks, and the experimental session took approximately 70 min.

### Results and discussion

Trials on which an observer failed to maintain fixation within a 3° window were discarded prior to analysis. This reduced the data set by 4.6%. It is possible that anticipation of impending peripheral signals caused the observers' eyes to make (irrelevant) saccades. However, the eye-tracking data did not reveal a bias toward either side of the screen. More important, the eyes did not jump significantly more toward the side that ultimately contained the target. The statistical analysis was a repeated measures ANOVA. There were no statistically significant differences across fields of presentation, so the data were pooled across the left and right fields of presentation.

The results from the sound-only (unimodal) condition approximated 50% correct and served as a sensitivity and criterion baseline (Fig. 5). In the multisensory conditions, the light was always present concurrently with the sound



**Fig. 5** Average percentage correct psychometric functions established from psychometric function intensities established in Experiment 1a. It includes the sound-only (unimodal) condition (comparable to Fig. 2) and 0-SOA multimodal conditions ( $\pm 8^\circ$  and  $\pm 24^\circ$ ). Error bars show standard errors of the means

(i.e., 0-ms SOA), and the light was located at either  $\pm 8^\circ$  or  $\pm 24^\circ$ . A two-way repeated measures ANOVA was performed with location and intensity as factors. There was an expected main effect of intensity and a statistically significant interaction between SOA and intensity. To minimize possible Type I error, scores were collapsed across locations, and nine planned comparisons were performed (i.e., one for each intensity level) against the unimodal condition. The negative decibel changes (i.e., 2, 4, 6, and 8 dB) and 0 dB were all statistically significantly different from the unimodal condition, with the highest probability being  $p < .03$ ,  $F(1, 24) = 5.28$ . All the other decibel changes (i.e., 2, 4, 6, and 8 dB) were not statistically significantly different from the unimodal condition, with the lowest probability being  $p > .26$ ,  $F(1, 24) = 1.32$ .

Figure 6 shows the ROC curve for the averaged data. Although the  $\pm 8^\circ$  multimodal stimuli were more sensitive than the unimodal sound stimuli, as was expected (compare with Fig. 4; Wilcoxon for  $\pm 8^\circ d'$ ,  $W(8) = 0.010$  and  $0.870$ ,  $Z = 2.31$ ,  $p = .021$ ; Wilcoxon for  $\pm 8^\circ \beta$ ,  $W(8) = 1.014$  and  $0.993$ ,  $Z = 0.06$ ,  $p = .953$ ), interestingly, the  $\pm 24^\circ$  multimodal stimuli were even more sensitive than the  $\pm 8^\circ$  multimodal stimuli (Wilcoxon for  $\pm 24^\circ d'$ ,  $W(8) = 0.870$  and  $1.155$ ,  $Z = 2.07$ ,  $p = .038$ ; Wilcoxon for  $\pm 24^\circ \beta$ ,  $W(8) = 0.993$  and  $1.001$ ,  $Z = 1.01$ ,  $p = .314$ ).

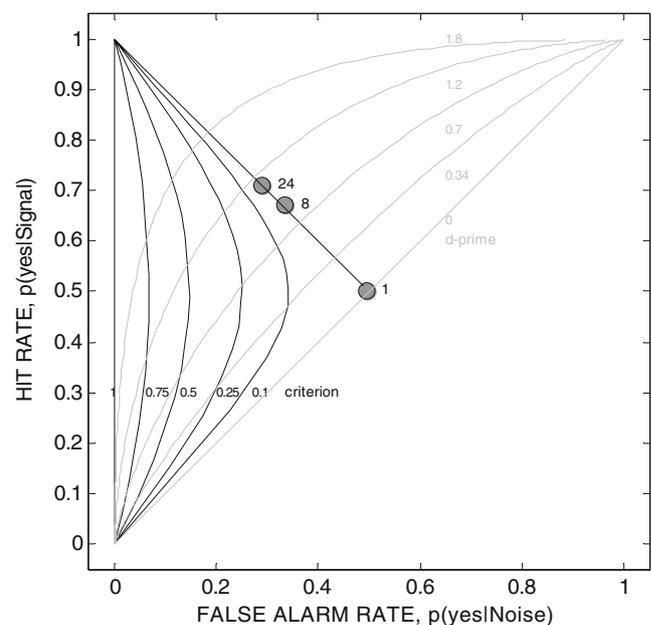
Perrott (1993) showed that as compared with the sound-only condition, peripheral light has the same impact on auditory localization, as compared with when the cross-modal pair was foveal. This occurs even though the light is harder to localize in the periphery. This presents an interesting conundrum.

In the present experiment, the more peripheral light ( $\pm 24^\circ$ ) had more of a detection-enhancing effect on an auditory target than did the  $\pm 8^\circ$  cross-modal pair. This agrees with Stein et al.'s (1996) prediction that the

reduction in a peripheral signal's saliency evokes the law of inverse effectiveness, making a cross-modal peripheral target *more* likely to show an enhancement in detection. This can be seen as a shift in sensitivity as targets are located more peripherally, but not a shift in the decision criterion.

## General discussion

Experiment 1b focused on the influence of timing on cross-modal detection. As was expected, when the light occurred slightly before (100-ms SOA) or was concurrent with the sound, multisensory enhancement resulted. These findings are in keeping with the second experiment of Lovelace et al. (2003), even though we did not block trials, and suggest that minimizing the temporal window reduced uncertainty and left the response criterion unaltered. They also agree with Diederich and Colonius (2004), who showed that physiological, not physical, synchronicity was important for bimodal enhancement when the light was presented slightly before the sound. Longer SOAs (350 and 700 ms) increase detection due to a shift in criteria, which suggests that the light can also serve as a temporal cue to relocate attention, as Gregg and Brogden (1952) would predict and in keeping with Lovelace et al.'s reanalysis of the data of Child and Wendt (1938; personal communication, October 2002),



**Fig. 6** Averaged ROC curve. The gray lines indicate several values of  $d'$ , and the black lines indicate iso-criterion ( $\beta$ ) shifts. 1 = unimodal; 8 =  $\pm 8^\circ$  0 SOA; 24 =  $\pm 24^\circ$  0 SOA. The  $\pm 24^\circ$  locations are significantly more sensitive than the  $\pm 8^\circ$  conditions, which are significantly more sensitive than the unimodal condition, whereas all five measures do not significantly differ in their criterion

attributing their greatest enhancement to a criterion shift when the irrelevant light was presented before the target sound by 500 ms. Hence, SDT can parse out how multisensory enhancement can reflect decisional processes.

However, the sensitivity shifts at lower percentage correct in Experiment 1b are most interesting. First, these 0- and 100-ms SOA data support Stein et al.'s (1996) prediction of the law of inverse effectiveness. Adding a light to a weak auditory signal showed the greatest enhancement; the present study is novel in making the auditory signal the target and the visual signal irrelevant, whereas Stein and colleagues did the reverse.

Second, in the spatial domain, by moving the irrelevant light further toward the periphery, Experiment 2 decreased its visual saliency and made auditory targets *more* detectable at low intensities. This result is unlike that in Stein et al. (1996), who found comparable enhancement effects when simultaneously presenting either a spatially coincident (0°) or a noncoincident ( $\pm 45^\circ$ ) auditory stimulus. Thus, their findings may be limited to using a visual target, and whereas the saliency of an auditory signal is unaffected in azimuth (Colonius & Diederich, 2004; Diederich & Colonius, 2009; Steenken, Colonius, Diederich, & Rach, 2008), visual signals are not.

The decrease in saliency of the visual signal as it moves to the periphery also affects the law of inverse effectiveness. In the prior literature, this effect has been shown only when the stimulus was the target (Frassinetti et al., 2002; Lovelace et al., 2003; Meredith & Stein, 1983, 1985; Stein et al., 1996). However, our spatial data set emphasizes that the law of inverse effectiveness can apply to *either* input of a cross-modal stimulus (the target or the distractor). These results go against Perrott's (1993). We found more enhancement with more peripheral targets and less for the same event presented centrally. Presenting visual signals at increasing eccentricities demonstrates the law of inverse effectiveness as supraliminal stimuli approach threshold.

Our findings may also be related to latent eye movement. Frens and van Opstal (1995) and Frens, van Opstal, and van der Willigen (1995) found that saccadic eye movements to a visual target that was spatially coincident with an irrelevant auditory signal had shorter latencies than did saccades directed toward the same visual target without an auditory co-stimulus. However, when the visual and auditory stimuli were increasingly separated in either space or time, reaction time gradually lengthened. Multimodal spatial and temporal effects on saccadic latency may be due to comparable factors that affected detection measures in the present study.

A series of elegant studies (Colonius & Diederich, 2004; Diederich & Colonius, 2009; Steenken et al., 2008) have also shown that saccadic reaction to a visual target tends to be quicker when auditory stimuli are presented in close

temporal and spatial proximity. This effect is attributed to a time-window-of-integration (TWIN) model, lasting roughly 200 ms, that comprises a peripheral stage of parallel processing in separate sensory channels, followed by a secondary stage of multisensory integration. The authors conjecture that rather distant stimuli of different modalities can temporally integrate because more eccentric stimuli, perceived to be of different saliencies, can alter the window of integration. This conjecture supports the postulates in our Experiment 2.

The enhanced detection at more peripheral locations found in the present study may well have been due to a reduction in visual signal saliency. Since previous studies typically have looked at auxiliary auditory stimuli, not auxiliary visual stimuli, the reduction in visual signal saliency illustrates that *either* cross-modal stimulus can affect detection as threshold is approached.

**Acknowledgements** I would like to thank Adam Gresko for help with data collections, Julie Edelson for editorial advice, and Mark Wallace and an anonymous reviewer for their thoughtful and careful reviews.

## References

- Bothe, G. G., & Marks, L. E. (1970). Absolute sensitivity to white noise under auxiliary visual stimulation. *Perception & Psychophysics*, *8*, 176–178.
- Child, I. L., & Wendt, G. R. (1938). The temporal course of the influence of visual stimulation upon the auditory threshold. *Journal of Experimental Psychology*, *23*, 109–127.
- Ciaramitaro, V. M., Cameron, E. L., & Glimcher, P. W. (2001). Stimulus probability directs spatial attention: An enhancement of sensitivity in humans and monkeys. *Vision Research*, *41*, 57–75.
- Colonius, H., & Diederich, A. (2004). Multisensory interaction in saccadic reaction time: A time-window-of-integration model. *Journal of Cognitive Neuroscience*, *16*, 1000–1009.
- Diederich, A., & Colonius, H. (2004). Bimodal and trimodal multisensory enhancement: Effects of stimulus onset and intensity on reaction time. *Perception & Psychophysics*, *66*, 1388–1404.
- Diederich, A., & Colonius, H. (2009). Crossmodal interaction in speeded responses: Time window of integration model. *Progress in Brain Research*, *174*, 119–135.
- Frassinetti, F., Pavani, F., & Làdavas, E. (2002). Acoustical vision of neglected stimuli: Interaction among spatially converging audio-visual inputs in neglect patients. *Journal of Cognitive Neuroscience*, *14*, 62–69.
- Frens, M. A., & van Opstal, A. J. (1995). A quantitative study of auditory-evoked saccadic eye movements in two dimensions. *Experimental Brain Research*, *107*, 103–117.
- Frens, M. A., van Opstal, A. J., & van der Willigen, R. F. (1995). Spatial and temporal factors determine auditory–visual interactions in human saccadic eye movements. *Perception & Psychophysics*, *57*, 802–816.
- Green, D. M., & Swets, J. A. (1966). *Signal detection theory and psychophysics*. New York: Wiley.
- Gregg, L. W., & Brogden, W. J. (1952). The effect of simultaneous visual stimulation on absolute auditory sensitivity. *Journal of Experimental Psychology*, *43*, 179–186.

- Haase, S. J., Theios, J., & Jenison, R. (1999). A signal detection theory analysis of an unconscious perception effect. *Perception & Psychophysics*, *61*, 986–992.
- Hawkins, H. L., Hillyard, S. A., Luck, S. J., Mouloua, M., Downing, C. J., & Woodward, D. P. (1990). Visual attention modulates signal detectability. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 802–811.
- Hess, R. F., Baker, D. H., May, K. A., & Wang, J. (2008). On the decline of 1st and 2nd order sensitivity with eccentricity. *Journal of Vision*, *8*(1, Art. 19), 1–12.
- London, I. D. (1954). Research on sensory interaction in the Soviet Union. *Psychological Bulletin*, *51*, 531–568.
- Lovelace, C. T., Stein, B. E., & Wallace, M. T. (2003). An irrelevant light enhances auditory detection in humans: A psychophysical analysis of multisensory integration in stimulus detection. *Cognitive Brain Research*, *17*, 447–453.
- Luck, S. J., Hillyard, S. A., Mouloua, M., & Hawkins, H. L. (1996). Mechanisms of visual-spatial attention: Resource allocation or uncertainty reduction? *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 725–737.
- Macmillan, N. A., & Kaplan, H. L. (1985). Detection theory analysis of group data: Estimating sensitivity from average hit and false-alarm rates. *Psychological Bulletin*, *98*, 185–199.
- Makous, J. C., & Middlebrooks, J. C. (1990). Two-dimensional sound localization by human listeners. *The Journal of the Acoustical Society of America*, *87*, 2188–2200.
- Markowitz, J., & Swets, J. A. (1967). Factors affecting the slope of empirical ROC curves: Comparison of binary and rating responses. *Perception & Psychophysics*, *2*, 91–100.
- Meredith, M. A., & Stein, B. E. (1983). Interactions among converging sensory inputs in the superior colliculus. *Science*, *221*, 389–391.
- Meredith, M. A., & Stein, B. E. (1985). Spatial factors determine the activity of multisensory neurons in cat superior colliculus. *Brain Research*, *365*, 350–354.
- Mills, A. W. (1958). On the minimal audible angle. *The Journal of the Acoustical Society of America*, *30*, 237–246.
- Odgaard, E. C., Arieh, Y., & Marks, L. E. (2003). Cross-modal enhancement of perceived brightness: Sensory interaction versus response bias. *Perception & Psychophysics*, *65*, 123–132.
- Ozbaydar, S. (1961). The effects of darkness and light on auditory sensitivity. *British Journal of Psychology*, *52*, 285–291.
- Perrott, D. R. (1993). Auditory and visual localization: Two modalities, one world. In *Proceedings of the 12th International Conference of the Audio Engineering Society* pp 221–231, Copenhagen.
- Schneider, W., Eschman, A., & Zuccolotto, A. (2002). *E-Prime user's guide*. Pittsburgh: Psychology Software Tools.
- Steenken, R., Colonius, H., Diederich, A., & Rach, S. (2008). Visual–auditory interaction in saccadic reaction time: Effects of auditory masker level. *Brain Research*, *1220*, 150–156.
- Stein, B. E., London, N., Wilkinson, L. K., & Price, D. P. (1996). Enhancement of perceived visual intensity by auditory stimuli: A psychophysical analysis. *Journal of Cognitive Neuroscience*, *8*, 497–506.
- Tanner, W. P., & Sorkin, R. D. (1972). The theory of signal detectability. In J. V. Tobias (Ed.), *Foundations of modern auditory theory* (Vol. 2). New York: Academic Press.
- Treisman, M., & Howarth, C. I. (1959). Changes in threshold level produced by a signal preceding or following the threshold stimulus. *The Quarterly Journal of Experimental Psychology*, *11*, 129–142.
- Ward, L. M., McDonald, J. J., & Golestani, N. (1998). Cross-modal controls of attention shifts. In R. D. Wright (Ed.), *Visual attention* (pp. 232–268). New York: Oxford University Press.
- Ward, L. M., McDonald, J. J., & Lin, D. (2000). On asymmetries in cross-modal spatial attention orienting. *Perception & Psychophysics*, *62*, 1258–1264.
- Wallace, M. T., Meredith, M. A., & Stein, B. E. (1992). Integration of multiple sensory inputs in cat cortex. *Experimental Brain Research*, *91*, 484–488.
- Watson, A. B., & Pelli, D. G. (1983). QUEST: A Bayesian adaptive psychometric method. *Perception & Psychophysics*, *33*, 113–120.